LA CORRELACIÓN Y REGRESIÓN EN BACHILLERATO: ANÁLISIS DE LIBROS DE TEXTO Y DEL CONOCIMIENTO DE LOS FUTUROS PROFESORES

María Magdalena Gea Serrano

Tesis Doctoral

UNIVERSIDAD DE GRANADA

Directora: Dra. Carmen Batanero Bernabeu

GRANADA, 2014

Trabajo realizado en el marco del Proyecto EDU2010-14947 y beca BES-2011-044684 (MICINN-FEDER), dentro del grupo de investigación FQM126 *Teoría de la Educación Matemática y Educación Estadística* (Junta de Andalucía).

RESUMEN

En esta investigación se aborda el estudio de la correlación y regresión en los libros de texto de Bachillerato, y la evaluación y el desarrollo del conocimiento didáctico-matemático sobre el tema en futuros profesores de Educación Secundaria y Bachillerato. El interés del tema se justifica por su importancia para la formación del estudiante y la escasez de investigaciones previas.

Nos basamos en el *enfoque ontosemiótico* de la cognición e instrucción matemática, el estudio histórico y matemático del tema, las directrices curriculares a nivel nacional y en Andalucía, así como en las investigaciones previas sobre la comprensión de la correlación y regresión y sobre el conocimiento del profesor para enseñar estadística.

A partir de estos fundamentos se presentan tres estudios interrelacionados:

- En el *Estudio 1* analizamos la presentación de la correlación y regresión en una muestra de 16 libros de texto de Matemáticas de primer curso de Bachillerato de las dos modalidades en que se incluye este tema: Humanidades y Ciencias Sociales y Ciencia y Tecnología. Se estudian los objetos matemáticos primarios definidos en el enfoque ontosemiótico, identificando las principales variables que los afectan y comparando los libros de una misma modalidad de Bachillerato y las dos modalidades. Asimismo, se analiza el uso de la tecnología y se identifican algunos conflictos semióticos potenciales. Con todo ello se determina el significado institucional pretendido de la correlación y regresión para este nivel educativo, que es la base de la construcción de las actividades e instrumentos de evaluación utilizados en los siguientes estudios.
- El *Estudio 2* estuvo dirigido a la evaluación y desarrollo del conocimiento matemático común y ampliado sobre la correlación y regresión en una muestra de futuros profesores. Los datos se recogen a partir de una actividad práctica en un curso dirigido a futuros profesores, en dos años sucesivos, y se basa en la realización de un proyecto estadístico, que ellos mismos podrían utilizar para la enseñanza del tema en primer curso de Bachillerato.
- El *Estudio 3* estuvo orientado a evaluar y desarrollar las diferentes facetas del conocimiento didáctico sobre la correlación y regresión en una submuestra del estudio anterior. Más concretamente, se centra en el análisis de las respuestas de estos futuros profesores a una actividad de análisis de la idoneidad didáctica del proyecto de análisis de datos trabajado por ellos mismos en la primera parte del estudio.

Como resultado se obtiene, en primer lugar, información detallada del significado institucional de referencia en los libros de texto, que proporciona criterios para su mejora y uso por parte del profesor. Otras aportaciones son las actividades diseñadas para el trabajo con futuros profesores, el análisis detallado de su conocimiento, y la descripción de ejemplos en los cuales los futuros profesores ponen en práctica diferentes facetas de dicho conocimiento, que pueden ser de utilidad en los cursos de formación de profesores. Las publicaciones derivadas de la tesis y el estado de la cuestión son otras aportaciones.

ABSTRACT

In this research we face the study of correlation and regression in the high school textbooks as well as in the assessment and evaluation of the related didactic-mathematical knowledge in prospective high and secondary school teachers. The interest of the topic is justified by its relevance in the education of students and by the scarce previous research.

We base on the *ontosemiotic approach* to mathematical cognition and instruction, the historical and mathematical study of the topic, the curricular guidelines at national level and in Andalusia, as well as in previous research on understanding correlation and regression and on teachers' knowledge to teach statistics.

Starting from these foundations we present three interrelated studies:

- In *Study 1*, we analyze the presentation of correlation and regression in a sample of 16 mathematics textbooks directed to high school first level in two modalities where the topic is included: Humanities and Social Sciences and Science and Technology. We study the primary mathematical objects defined in the ontosemiotic approach, identify the main variables affecting them and compare books in the same and different high school modalities. We also analyze the use of technology and identify some potential semiotic conflicts. Consequently, we determine the institutional intended meaning of correlation and regression in this educational level. This is the base for building the activities and assessment instruments used in the following studies.

- *Study 2* was intended to assess and develop the common and widened mathematical knowledge of correlation and regression in a sample of prospective teachers. Data are collected from a practical activity in a course directed to prospective teachers in two consecutive years and based on a statistical project that the participants may also use to teach the topic in the first level of high school.

- Study 3 was aimed to assess and develop the different facets of didactic knowledge of correlation and regression in a subsample taken from the previous study. Specifically we focus on the analysis of these prospective teachers' responses in an activity where they analyse the didactic suitability of the data analysis project they worked with in the first part of the study.

Some results include detailed information of the institutional intended meaning in the textbooks, which provide criteria for improving and using them in the classroom. Other contributions include the activities designed to work with prospective teachers, the detailed analysis of their knowledge and the description of examples where these prospective teachers use different facets of the aforementioned knowledge, which can be useful in the training of teachers. Finally, the publications derived from the thesis and the state of art is additional contributions for research.

AGRADECIMIENTOS

En primer lugar quiero expresar mi agradecimiento a la Dra. Carmen Batanero Bernabeu por confiar en mis posibilidades, por haberme dado la oportunidad de trabajar a su lado, compartiendo gratas experiencias, y superando con su respaldo muchos retos profesionales. También quiero agradecer su paciencia y, sobre todo, su gran ayuda y preocupación constante por mi formación como investigadora y profesora. Gracias por su sabiduría y buen consejo, que han iluminado y guiado este trabajo.

Al departamento de Didáctica de la Matemática de la Universidad de Granada, donde desarrollo mi trabajo, en especial a Juan D. Godino, por sus ideas y orientación, por estar siempre dispuesto a ayudarme. A Rafael Roa, por sus sabios consejos, buenos ratos, y sobre todo, por haberme hecho un hueco en su despacho en el que poder trabajar. También al Departamento de Didáctica de las Ciencias de la Universidad de Jaén, por su empujón en mis comienzos, un verdadero placer compartir con ellos, en especial a Antonio Estepa y Ángel Contreras.

Gracias a los alumnos del Máster universitario en Profesorado de Enseñanza Secundaria Obligatoria y Bachillerato de la Universidad de Granada, por su generosa colaboración realizando el cuestionario. Y a los profesores Indalecio Pérez-Entrena y Natalia Jiménez Luna, por su inestimable ayuda, cediéndome el espacio del IES Juan Rubio Ortiz cada vez que lo he necesitado.

A mis padres, por todo lo que me han inculcado, y a toda mi familia, por su preocupación e interés por todos los pasos que he ido dando en mi vida.

Y sobre todo agradecerte a ti, Juan, toda tu paciencia y amor con el que compartes todo lo que estoy consiguiendo. Por tu comprensión y tus palabras de aliento que, en los momentos de mayor tensión, me han dado el empuje emocional necesario para avanzar. Gracias por hacer entender a nuestros hijos (Juan y María Magdalena) que el trabajo de mamá es importante.

ÍNDICE

INTRODUCCIÓN

Explicar, controlar y predecir los acontecimientos que se suceden en nuestro día a día depende, en gran medida, de habilidades y destrezas para detectar asociaciones entre las variables implicadas. Por ello, disciplinas como la Psicología, la Sociología y la Didáctica Estadística, entre otras, se han interesado por estos procesos de razonamiento, y aportan conocimiento específico sobre este tema. En todas ellas se puede advertir la dificultad intrínseca del ser humano en la emisión de juicios de asociación correctos.

La correlación y regresión resultan de gran ayuda en cuanto a discernir si dos variables se relacionan, con qué intensidad o en qué sentido lo hacen, e incluso para predecir una de ellas a partir de la otra, cuando sea posible. De acuerdo a Burrill y Biehler (2011), estos conceptos son fundamentales en estadística, pues amplían la dependencia funcional a situaciones aleatorias, que son las más frecuentes en la investigación, política, economía y otras áreas de la actividad humana. Asimismo, permiten graduar la intensidad de la dependencia, mediante el coeficiente de correlación.

Este tema se incluye actualmente en el nivel de Bachillerato (MEC, 2007b), tanto en la modalidad de Humanidades y Ciencias Sociales, como en la de Ciencias y Tecnología. En este nivel educativo, la regresión se limita al estudio del modelo lineal, es decir, a la recta de regresión, sus parámetros y su uso en la resolución de problemas. La comprensión de este modelo, así como de su relación con la correlación, preparará a los estudiantes para el aprendizaje posterior de otros más complejos, como por ejemplo, la regresión no lineal, o la regresión múltiple, en la que intervienen varias variables independientes.

Aparentemente, este contenido debería ser sencillo, pues los estudiantes de Bachillerato ya están familiarizados con las funciones y con la ecuación de la recta. Quizás debido a esta simplicidad supuesta, no se presta un interés suficiente a su enseñanza; por ejemplo, son pocos los problemas propuestos sobre correlación y regresión en las pruebas de selectividad. Sin embargo, muchas investigaciones previas, tanto en Psicología como en Didáctica de las Matemáticas (ej., Beyth-Marrom, 1982; Estepa, 1994; Sánchez Cobo, 1998; Zieffler, 2006; Catro-Sotos et al., 2009) nos advierten de la dificultad que implica el razonamiento correlacional, así como de la persistencia de errores conceptuales y sesgos en la estimación de la correlación, o en la interpretación de propiedades de la correlación y regresión en los alumnos que ingresan en la universidad.

Con el fin de proporcionar nuevos resultados en esta línea de investigación, y contribuir a la mejora de la enseñanza y del aprendizaje de la correlación y regresión por parte de los estudiantes, en esta Memoria se abordan dos problemas relacionados:

a. El análisis del contenido sobre correlación y regresión en los libros de texto de Bachillerato en las dos modalidades en que actualmente se incluye este tema. Dicho análisis tiene como finalidad aproximarse a la descripción del significado institucional de referencia del tema en el nivel de Bachillerato;

b. La evaluación y desarrollo del conocimiento didáctico-matemático sobre correlación y regresión en una muestra futuros profesores de Educación Secundaria y Bachillerato.

La Memoria se organiza en seis capítulos relacionados, que describen y fundamentan tres estudios empíricos:

En el primero de ellos se define con más precisión el problema de investigación abordado, y se justifica su importancia, tanto para la formación del estudiante, como dentro de la propia estadística. Asimismo, se analizan con detalle los conceptos de correlación y regresión, describiendo su origen histórico y precisando el significado inicial de referencia de estos objetos matemáticos en nuestro trabajo. Seguidamente, se analizan las directrices curriculares de este tema a nivel nacional y en la Comunidad Autónoma andaluza, comparando también con el currículo americano, que ha inspirado en parte el nuestro. Se finaliza con la presentación de los objetivos de esta investigación, la forma en que se organiza la investigación, y sus características metodológicas.

En el segundo capítulo se describen los fundamentos teóricos, que constan de dos partes: En primer lugar, se analiza el marco teórico, que es el enfoque ontosemiótico de la cognición e instrucción matemática (EOS) (Godino y Batanero, 1994; 1998a y b; Godino, 2002; Godino, Batanero, y Font, 2007; 2012) cuyas componentes ontológicas, cognitivas, semióticas e instruccionales se adaptan a nuestros objetivos. Por otro lado, presentamos un análisis de diversos modelos sobre el conocimiento del profesor de matemáticas y de estadística, para seleccionar el que específicamente se utilizará en el trabajo, que es el conocimiento didáctico-matemático descrito en el enfoque ontosemiótico (Godino, 2009; 2011), que también proporciona criterios para la evaluación de sus componentes.

El tercer capítulo resume las investigaciones educativas y psicológicas más relevantes sobre correlación y regresión, enmarcándolas en el *razonamiento covariacional*. Se analizan las estrategias intuitivas en la estimación de la correlación, los sesgos en esta tarea, las concepciones descritas, la precisión en la estimación del coeficiente de correlación desde diversas representaciones, y el desarrollo del razonamiento correlacional con la enseñanza. Un segundo bloque es el análisis de la presentación de contenidos de estadística y probabilidad en los libros de texto, donde se resalta especialmente las dedicadas al estudio de la correlación y regresión en los libros de texto. Igualmente se resumen las investigaciones sobre el conocimiento del profesor en la enseñanza de la estadística, por un lado en contenidos previos (como la representación de datos, las medias de posición central, dispersión y distribución) de interés por su influencia en la enseñanza y aprendizaje de la correlación y regresión, y también en los específicos de correlación y regresión, que son escasos.

Seguidamente se describen los tres estudios empíricos llevados a cabo y sus resultados. En el Estudio 1, descrito en el Capítulo 4, analizamos la presentación de la correlación y regresión en una muestra de 16 libros de texto de Matemáticas, dirigidos a estudiantes de primer curso de Bachillerato en las dos modalidades en que se incluye este tema: Humanidades y Ciencias Sociales y Ciencia y Tecnología. Se estudian los objetos matemáticos primarios definidos en el enfoque ontosemiótico, identificando las principales variables que los afectan, y comparando entre libros de una misma modalidad de Bachillerato y entre modalidades. Asimismo, se analiza el uso de la tecnología, y se identifican algunos conflictos semióticos potenciales. Con todo ello, se determina el significado institucional pretendido de la correlación y regresión para este nivel educativo, que junto al análisis epistémico y curricular desarrollado en el Capítulo 1 se concretará en significado de referencia de estas nociones, que será la base de la construcción de las actividades e instrumentos de evaluación utilizados en los siguientes estudios.

En el Capítulo 5 se presenta un estudio exploratorio de evaluación y desarrollo del conocimiento matemático sobre la correlación y regresión en una muestra de futuros profesores (Estudio 2). Los datos se recogen a partir de una actividad práctica realizada en un curso dirigido a futuros profesores, en dos años sucesivos, y se basa en la realización de un proyecto estadístico, que ellos mismos podrían utilizar para la enseñanza del tema en primer curso de Bachillerato. Esta actividad permitirá describir y desarrollar el conocimiento común y ampliado sobre la correlación y regresión de los participantes.

En el Capítulo 6 se describe el Estudio 3, orientado a evaluar y desarrollar los conocimientos didácticos sobre la correlación y regresión en la submuestra del estudio anterior, que participó el primer año. Nos enfocamos en las facetas del conocimiento didáctico-matemático consideradas por Godino (2009, 2011). Más concretamente, se centrará en el análisis de las respuestas de estos futuros profesores a una actividad de análisis de la idoneidad didáctica (Godino, Bencomo, Font y Wilhelmi, 2006) del proyecto de análisis de datos trabajado por ellos mismos en la primera parte del estudio.

El Capítulo 7 se discuten las conclusiones en relación a los objetivos planteados, y se analizan las principales aportaciones. Entre ellas destacamos la caracterización detallada del significado institucional de referencia, a partir del análisis de los libros de texto, que proporciona criterios para su mejora y para su uso en el aula por parte del profesor. Las actividades realizadas con los futuros profesores y el análisis detallado de su conocimiento, como resultado de los estudios de evaluación, constituyen también una aportación para los cursos de formación de profesores. Resultados parciales de todas estas contribuciones se han dado a conocer a través de diversas publicaciones que se detallan a lo largo de la Memoria.

La Memoria finaliza con las referencias y dos Anexos. El primero de ellos es el proyecto estadístico utilizado en el trabajo con los futuros profesores en el Estudio 2; el segundo la pauta de análisis de la idoneidad didáctica utilizada en el Estudio 3.

CAPÍTULO 1.
EL PROBLEMA DE INVESTIGACIÓN

1.1. INTRODUCCIÓN

Como se ha indicado en la introducción, nuestra investigación se centra en el estudio de la correlación y regresión, temas que amplían el concepto de dependencia funcional, y que los estudiantes ya conocen al ingresar en el Bachillerato, que es la etapa educativa en que se inicia en España el estudio del tema y en la que nos centramos. Además, nos restringimos al estudio bivariante, es decir, se consideran sólo dos variables estadísticas.

Podemos describir, informalmente, estos temas por medio de dos preguntas, que suelen plantearse en cualquier estudio estadístico cuando se dispone de datos de dos variables *X* e *Y* (Batanero, 2001): La primera es si existe alguna relación entre estas dos variables, y si es suficientemente intensa (problema de correlación). La segunda es si existe la posibilidad de encontrar una expresión algebraica que permita calcular en forma exacta o aproximada los valores de una de las variable en función de la otra, con una finalidad predictiva (problema de regresión).

En la búsqueda de dicha expresión puede suceder que los valores que toma la variable dependiente *Y*, queden determinados, de un modo unívoco, por los valores de la variable independiente *X*. Este es el caso de la dependencia funcional o determinista; por ejemplo, la caída libre de los cuerpos (Batanero y Díaz, 2008). El caso más general es cuando a cada valor de *X* corresponde una distribución de valores de *Y* (dependencia aleatoria). La función de regresión en dichos casos es un modelo que permite aproximar a los datos; la precisión de la aproximación dependerá de la intensidad de la relación entre las variables.

Como veremos al final del capítulo, en el nivel de Bachillerato la regresión se limita al estudio del modelo lineal, es decir, la recta de regresión. La comprensión de este modelo, así como de su relación con la correlación, preparará a los estudiantes para el aprendizaje posterior de otros modelos más complejos de regresión, por ejemplo, la regresión no lineal o con varias variables independientes.

La finalidad de este capítulo es definir con más precisión el problema de investigación abordado, y justificarlo. En primer lugar, se resalta la importancia de la alfabetización estadística de los ciudadanos y se describe la emergencia reciente de la educación estadística como campo propio de investigación. En las siguientes secciones, analizamos con más detalle los conceptos de correlación y regresión, describiendo su origen histórico, resaltando su actual importancia en el conjunto de métodos estadísticos y precisando el significado inicial de referencia de estos objetos matemáticos en nuestro trabajo.

Seguidamente se analizan las directrices curriculares de este tema a nivel nacional y en la comunidad autónoma andaluza, comparando con el currículo americano, que es en la actualidad guía de otros en educación estadística. Se finaliza con la presentación de los objetivos de este trabajo, la forma en que se organiza la investigación, y sus características metodológicas. Las hipótesis correspondientes a cada uno de los tres estudios principales en que se divide la investigación se exponen con detalle en los capítulos correspondientes.

1.2. CULTURA ESTADÍSTICA E INVESTIGACIÓN EN EDUCACIÓN ESTADÍSTICA

Como señala Cabriá (1994, p.21): “*Los orígenes de la estadística se confunden con los de la humanidad; pero sólo en tiempos recientes ha adquirido la categoría de disciplina relevante*”.

Su importancia actual es resaltada por muchos autores. Por ejemplo, Hacking (1990) sugiere que el descubrimiento del carácter no determinista de nuestro mundo y la aceptación de la posibilidad de utilizar modelos aleatorios para comprender y predecir muchos de sus fenómenos ha sido el salto conceptual más importante del siglo XX. La superación de la concepción determinista del mundo y de la relación de causalidad lleva a asumir que el pasado no determina el futuro de forma unívoca. La regularidad observada en nuestro mundo, no siempre es consecuencia de leyes deterministas, sino también de otras de carácter aleatorio (Batanero, Serrano y Green, 1998).

Este cambio de perspectiva científica fue posible por el desarrollo imparable de la estadística en el pasado y presente, y como consecuencia, ha llevado a la necesidad de una cultura estadística para el ciudadano (Wallman, 1993; Gal, 2002). La estadística puede ser utilizada como método de investigación y como vehículo de transmisión de información, lo que precisa unos conceptos, terminología y un tipo de razonamiento que requieren de una enseñanza y aprendizaje específicos. Es también evidente la presencia de la estadística en la vida diaria del ciudadano (prensa, radio, TV, libros, Internet, comercio, publicidad, etc.).

Es por ello necesario fortalecer el pensamiento estadístico en todos los sectores de nuestra población. Esta necesidad ha sido observada por los organismos productores de estadística (Institutos de Estadística y otros organismos), quienes desean hacer comprender a la población las estadísticas elaboradas por ellos, y su importancia en la toma de decisiones que afectan a sus vidas profesionales y personales. Ello ha llevado a diversos autores a reclamar la alfabetización estadística, analizando sus componentes. Así por ejemplo, Wallman (1993) nos indicó que:

> Alfabetización estadística es la capacidad de comprender y evaluar críticamente los resultados estadísticos que impregnan nuestra vida cotidiana - junto con la capacidad de apreciar las contribuciones que el pensamiento estadístico puede hacer en las decisiones públicas y privadas, profesionales y personales (Wallman, 1993, p. 1).

Uno de los mecanismos principales para lograr la cultura estadística para todos ha sido la incorporación generalizada en el currículo de contenidos estadísticos desde la Educación Primaria, en la mayoría de países desarrollados. Paralelamente, los investigadores en didáctica de la matemática se interesan por los problemas de enseñanza y aprendizaje de esta materia, dando origen a una línea de investigación de gran desarrollo actual.

Educación estadística, como campo de investigación

Hace unos años, pocos investigadores se interesaban por la enseñanza y aprendizaje de la estadística, y la investigación sobre razonamiento probabilístico se desarrollaba preferentemente en el campo de la psicología (Batanero, 2004). También desde la propia estadística ha habido un gran interés por la educación, como se muestra en el establecimiento de conferencias internacionales específicas, especialmente los *ICOTS* (*International Conference on Statistical Education*) iniciadas en 1982.

Otro indicador de este interés por la educación dentro de la estadística fue la constitución de la sociedad *IASE* (*International Association for Statistical Education*), como rama del Instituto Internacional de Estadística, centrada en el desarrollo de la investigación y mejora de la enseñanza. Posteriormente, se crean revistas dirigidas a profesores como *Teaching Statistics*, *Induzioni* o *Journal of Statistics Education* y más tarde otras de investigación como *Statistics Education Research Journal* (SERJ), *Statistique et Enseignement* o *Technology Innovations in Statistics Education*.

En las últimas dos décadas, y como consecuencia de la introducción de la estadística en el currículo de Educación Primaria y Educación Secundaria Obligatoria, con la consiguiente necesidad de formar los profesores en esta materia, la didáctica de la estadística ha comenzado a interesar a los investigadores en educación matemática. Por ello encontramos actividades de educación estadística en los Congresos Internacionales de Educación Matemática (*ICME*), organizados por *ICMI* (*International Commission on Mathematical Instruction*) o *CERME* (*Congress of European Research in Mathematics Education*), promovidos por la *European Association for Research in Mathematics Education*. En España destacamos el grupo de investigación de la *Sociedad Española de*

Investigación en Educación Matemática, que el año 2013 organizó las *I Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria* (http://jvdiesproyco.es/).

Como indican Batanero y Godino (2005), el siglo XX fue el siglo de la estadística, y su final ha marcado la emergencia de la educación estadística. Nuestro trabajo quiere contribuir a este campo en un tema de escasa investigación.

1.3. CORRELACIÓN Y REGRESIÓN

La relevancia de la correlación y regresión se debe a que subyacen en multitud de métodos estadísticos (Engel y Sedlmeier, 2011). Por otro lado, investigaciones llevadas a cabo desde la Psicología, la Sociología y la Didáctica Estadística sugieren falta de razonamiento correlacional, pues el razonamiento humano en situaciones de incertidumbre está regido por valores y creencias del propio individuo. Por ello, el procesamiento de la información difiere de un proceso algorítmico que produce una solución para cualquier problema dentro de una clase dada (Batanero, 2001).

Debido a esta importancia, autores como King (2000) han tratado de inducir el razonamiento relacionado desde la Educación Primaria. Como indica Moritz:

> Una vez que los estudiantes han comenzado a tratar el contexto de las variables, se puede comenzar a investigar la covariación existente, discutir los modos de razonamiento acerca de esta covariación, y poco a poco, ir presentando las convenciones para expresar sus razonamientos en los gráficos, términos y métodos numéricos. (Moritz, 2004, p. 253).

A continuación describimos brevemente los orígenes de estas nociones, que hemos resumido en Estepa, Gea, Cañadas y Contreras (2012).

1.3.1. ORIGEN HISTÓRICO

Los juicios sobre la posible asociación de variables acompañan al ser humano desde el comienzo de su existencia. Como ejemplo podemos citar los juicios de asociación que en la civilización Egipcia se realizaron entre la crecida del río Nilo y la periodicidad del ascenso (en igual sentido que el Sol) de la estrella Sirio. Esta observación y su análisis, dieron un avance considerable al conocimiento científico en cuanto a la construcción de calendarios y medición del tiempo.

La formación de las nociones de correlación y regresión proviene, en gran parte, de estudios realizados en Biología, Biometría y Eugenesia. Según Benzecri (1982), el primer autor que se interesa en el tema fue Lambert-Adolphe-Jacques Quetelet (1796-1874), nacido en Gante, Bélgica, quien fue director del observatorio astronómico de Bruselas, y utilizó su talento y energía para crear varias instituciones internacionales.

Sus aportaciones sobre la correlación y regresión se originan desde sus estudios sobre el hombre medio, estimando empíricamente las medias y desviaciones típicas de medidas antropométricas, que suponía dependen de varias variables independientes, tales como el sexo, edad, edad, profesión o nivel de educación. En sus estudios relaciona dos o más variables, por ejemplo llega a obtener una ecuación de una hipérbola que relaciona la edad y la altura de las personas entre cero y 30 años (Hald, 1998). Su originalidad no consistió en haber calculado las medias de las magnitudes antropométricas, sino haber considerado su dispersión y descubierto que la ley normal (bien conocida en Astronomía) ofrecía una descripción aceptable de tal variabilidad, por lo que utilizó esta distribución como ajuste a sus medidas antropométricas, introduciéndola en Biometría (Seal, 1967).

Augusto Bravais (1811-1863) contribuye al desarrollo de esta teoría desde otro campo: la astronomía, al estudiar los errores en las medidas de las coordenadas de cuerpos espaciales. Fue él quien utilizó por primera vez el término correlación en un estudio presentado en 1846 en la Academia de Ciencias en Francia (Seal, 1967). Sin embargo, Pearson (1965) indicará que Bravais al estudiar la teoría de errores no consideraba variables aleatorias correlacionadas, sino consideraba errores independientes unos de otros; por tanto, no llegó a una verdadera idea de la correlación, tal como hoy la conocemos.

Aportaciones de Galton

La principal contribución al tema la encontramos en Francis Galton (1822-1917). Según Hald (1998), Galton estudió medicina y matemáticas en Londres y Cambridge. En 1844 murió su padre, dejándole una inmensa fortuna que le permitió dedicar su vida a los estudios geográficos y meteorológicos.

En el periodo 1865-1890, su principal interés fueron los estudios empíricos de las leyes de la herencia por medio de métodos estadísticos. Galton no conocía los refinados métodos estadísticos de la época, debidos a Laplace y Gauss. Sin embargo, por medio de investigaciones empíricas de las leyes de la herencia, estudia la variabilidad de características humanas y desarrolla sus propios y rudimentarios métodos para describir observaciones univariadas y bivariadas normalmente distribuidas, explicando la utilidad y el significado de la regresión y correlación no solamente en el contexto de la herencia, sino en forma general.

Galton utiliza el método de Quetelet para ajustar una distribución normal a sus datos. Este método era muy simple, ya que requiere solamente el cálculo de frecuencias relativas y la interpolación en la tabla de la binomial acumulativa. Como no domina con soltura la matemática de su tiempo, utiliza artificios mecánicos para “probar” las propiedades de la distribución binomial como el de la Figura 1.3.1 que llamó “*quincunx*” (1889), también llamado aparato de Galton.

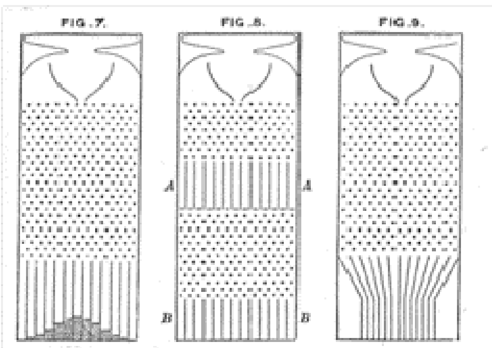


Figura 1.3.1. Versiones del quincux en “*Natural Inheritance*” (Hald, 1998, p. 605)

Su parentesco con Charles Darwin le interesó por ofrecer un sustento científico a la ley de herencia biológica presentada en *“El origen de las especies”*, escrito por Darwin en 1859. Al no disponer de suficientes datos humanos, diseñó un experimento con semillas de guisantes (Hald, 1998). Motivado por la idea de que ciertos caracteres de los padres y otros antepasados se reproducían en los hijos, planeó un método de trabajo donde, al observar que la distribución de los pesos de las semillas utilizadas se distribuía según una distribución normal, seleccionó siete grupos conteniendo cada grupo 70 semillas del mismo peso. El peso de cada semilla de cada grupo era la media más o menos 0, 1, 2 y 3, desviaciones típicas. Pidió a siete amigos de diferentes partes del país que cultivaran un grupo de semillas y que le enviarán las semillas cosechadas. Sus conclusiones, que ilustra en dibujos similares a los de un quincunx (Figura 1.3.2), fueron:

a. Para cada grupo de semillas padres, el peso de las semillas filiales estaba normalmente distribuido;

b. El peso medio de las semillas filiales es una función lineal del peso de las semillas padres con una pendiente menor que la unidad, es decir, el peso medio de la progenie se desvía menos de la población media que el de los padres, Galton llamó a esta propiedad *reversión* (después se llamará correlación);

c. La desviación probable del peso de las semillas filiales es la misma para todos los grupos y más pequeña que la desviación probable del peso de las semillas padres.

Los padres de peso $M + x$ producen hijos adultos de peso medio $M + r \cdot x$, donde $0 < r < 1$. El peso de los hijos llega a ser $M + r \cdot x + y = M + z$ por la variación aleatoria entre hijos del mismo grupo de padres, variación que se observa en el quincunx. La identidad estadística de las dos generaciones significa que $\sigma_x^2 = \sigma_z^2$, en consecuencia, $\sigma_x^2 = r^2 \cdot \sigma_x^2 + \sigma_y^2$, o lo que es lo mismo, $\sigma_y^2 = \sigma_x^2 \cdot (1 - r^2)$, que nos da la relación entre la varianza condicional (variación entre grupos), la varianza marginal (la variación entre el total de la población) y el coeficiente de reversión (Estepa, 1994).

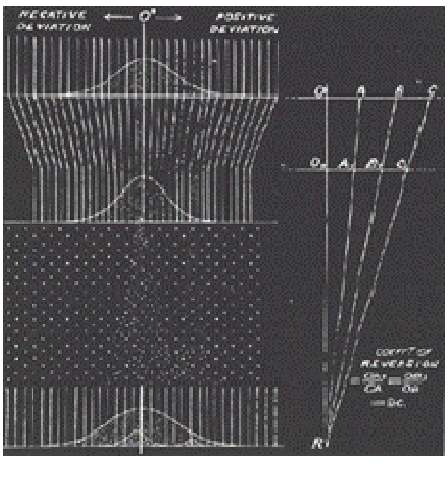


Figura 1.3.2. Ilustración de Galton de la identidad estadística de dos generaciones (Hald, 1998, p. 606)

En 1869, Galton publica "*Hereditary Genius"*, orientado a estudiar la influencia de los padres y otros antepasados en los hijos. El método para expresar estas relaciones lo describe del siguiente modo:

> Parecía evidente por observación, y había sido completamente confirmado por esta teoría, que existía un "índice de correlación"; o sea, una fracción, que ahora llamamos simplemente *r* que relaciona con la mayor aproximación cada valor de desviación (de la media) por parte del sujeto con el promedio de todas las desviaciones asociadas, del pariente, tal como ha sido descrito. Por lo tanto, la aproximación de cualquier parentesco específico puede ser hallada y expresada con un término único. Si un individuo particular se desvía mucho, el promedio de las desviaciones de todos sus hermanos será una fracción definida de esa cantidad; del mismo modo que los hijos, los padres, primos hermanos, etc. Cuando no hay relación alguna, *r* se vuelve igual a 0; cuando es tan cercana que sujeto y pariente poseen idéntico valor, entonces $r = 1$. Por lo tanto, el valor de *r* reside siempre entre los límites extremos de 0 y 1 (reproducido en Newman, 1956, pg. 239).

Galton no había considerado más que la distribución de una medida X , tomada conjuntamente sobre el padre y su descendencia, pero con posterioridad se da cuenta de la posibilidad de estudiar variaciones conjuntas de medidas biológicas diferentes sobre los mismos individuos. A pesar del enorme ingenio, y como señala Pearson: "*Galton no había aún alcanzado la idea de correlación negativa*" (Pearson, 1920, pg. 199). Se basa en la definición que propone en el artículo "*Co-relations and their measurement, chiefly from anthropometric data"*, Publicado en los *Proceedings de la Royal Society*, 45, 135-145 en 1888:

> Co-relación o correlación de estructura es una frase de gran uso en biología y no menos en la rama que estudia la herencia, y la idea es incluso más frecuente que la frase; pero no soy consciente de ningún intento anterior de definirla claramente, de trazar su forma de acción en detalle o mostrar cómo medirla. Dos órganos variables se dicen co-relacionados cuando la variación en uno se acompaña en promedio por más o menos variación del otro y en la misma dirección. Así, la longitud del brazo se dice correlacionada con la de la pierna porque una persona de brazo largo usualmente tiene pierna larga y al contrario. Si la correlación es cercana, entonces una persona de brazos muy largos tendrá piernas muy largas; si es moderadamente próxima, entonces la longitud de la pierna sólo será larga, no muy larga; y si no hubiese ninguna correlación, entonces la longitud de la pierna sería, en promedio, mediocre" (citado por Pearson, 1920, pg.199).

En su obra, *Natural Inheritance*, publicada en 1889, Galton propone un amplio programa de investigación biométrica: estudiar estadísticamente la variabilidad de las medidas físicas de distintas especies, a fin de confirmar matemáticamente el mecanismo de la evolución descrito por Darwin (Benzecri, 1982). En consecuencia, Galton fue consciente de que sus descubrimientos parecían dar lugar a un amplio campo de aplicación de problemas que caerían bajo las leyes de la correlación.

Desarrollos posteriores

La sugerencia fue recogida por prestigiosos autores como Edgeworth, Pearson, Yule, Seppard (Hald, 1998) que en tiempos posteriores desarrollaron estas ideas en profundidad. Por ejemplo, Weldon calculó empíricamente coeficientes de correlación entre varias medidas físicas de órganos de camarones, es decir correlaciones entre medidas del mismo sujeto. También contribuyó a crear la teoría sobre la distribución muestral del coeficiente de correlación. Fue el primero que publicó un artículo en el que señala el significado de un coeficiente de correlación negativo, indicando que era posible determinar una razón (coeficiente), cuyo valor se convierte en ± 1 cuando un cambio en cualquiera de órganos implica un cambio igual en la otra, y 0 cuando los dos órganos son bastante independientes (Pearson, 1920).

De acuerdo a Benzecri (1982), las ideas modernas sobre regresión se originan en los trabajo de Legendre y Gauss sobre el método de mínimos cuadrados, para ajustar los datos sobre las órbitas de cuerpos celestes. El primer estudio documentado sobre el método de mínimos cuadrados, de donde deriva la idea de regresión, es debido a Legendre en 1805.

Esta técnica de optimización intenta encontrar la función (dentro de una familia) que mejor se ajusta a los datos bivariantes, de acuerdo con el criterio de mínimo error cuadrático, y siempre que los datos cumplan algunas condiciones (como independencia). Se conocían las ecuaciones funcionales de estas órbitas, pero los errores de medida hacía que los cálculos fuesen aproximados y se ajustaban ciertas familias de funciones, usando la teoría de errores (y la distribución normal para describirlos). En 1829, Gauss fue capaz de establecer la razón por la cuál este procedimiento es muy adecuado desde el punto de vista estadístico. Lo que hoy se conoce como teorema de Gauss-Márkov muestra que los estimadores obtenidos con este método son insesgados y no se requiere una distribución específica para los datos que se ajustan. El impulso posterior lo dan los trabajos citados de Galton y Pearson sobre la herencia. Yule la utiliza para el estudio de fenómenos sociales, como las causas de pobreza alrededor de 1899.

A lo largo del siglo XIX el trabajo de los estadísticos era mayormente descriptivo; la inferencia estadística se va a desarrollar como consecuencia de la creación de la Escuela Biométrica del University College de Londres, bajo la dirección del matemático Karl Pearson (1857-1936), quien trata de aportar bases matemáticas a los descubrimientos de Galton. Pearson defiende que el método científico es esencialmente estadístico, pues sus inferencias se basan en la asociación entre antecedentes y consecuentes. Alrededor de 1895, Pearson había resuelto las propiedades matemáticas del coeficiente de correlación y la regresión simple utilizada para la predicción lineal entre dos variables continuas. También generaliza la idea de regresión de Galton a varias dimensiones, comprobando que muchas variables biométricas siguen la distribución normal multivariante, y por tanto, se podría aplicar la regresión múltiple para relacionar varias de estas medidas.

Yule, por su parte, introduce la notación e ideas para la correlación parcial de dos variables, fijando una tercera, en un artículo “On the theory of correlation for any number of variables, treated by a new system of notation” publicado también en los *Proceedings de la Royal Society* en 1907.

1.3.2. IMPORTANCIA EN EL MÉTODO ESTADÍSTICO

La correlación y regresión se extendieron rápidamente, y además constituyeron la base de otros métodos estadísticos, que describimos brevemente en este apartado.

1.3.2.1. CORRELACIÓN Y REGRESIÓN EN EL ANÁLISIS EXPLORATORIO DE DATOS

En primer lugar detacamos la importancia del tema en el *Análisis Exploratorio de datos* (en adelante AED) desarrollado por Tukey (1962; 1977) y recomendado actualmente en el currículo. Dicha corriente fue propiciada por el impulso que la tecnología dio a la estadística en aquellos años y posteriormente, permitiendo explorar grandes cantidades de datos con técnicas que antes sólo eran utilizadas en situaciones-problema con objetivos concretos (Batanero, Estepa y Godino, 1991).

Este es el caso de las nociones de correlación y regresión. Según Batanero, Estepa y Godino (1991), cuando se *analiza* la posible relación entre variables de interés, el investigador no debe basar su estudio únicamente en ajustar los datos a un modelo prefijado, como puede ser un ajuste lineal, sino que puede estudiar diversos estadísticos, comparar los modelos planteados con los residuos, estudiar la significación estadística de los parámetros utilizados, etc. Por ejemplo, aunque el coeficiente de correlación presente un valor estadísticamente significativo, la relación entre las variables puede no ajustarse bien al modelo lineal. Estas posibilidades de exploración las prestan actualmente los ordenadores y dispositivos móviles.

Batanero (2001) indica que el AED es un puente entre estadística descriptiva e inferencia; y que trata de investigar tanto la tendencia, como la dispersión en los datos. Una vez estudiada la tendencia que sigue un cierto patrón (ajuste lineal, parabólico, exponencial, etc.), la observación de las desviaciones o variaciones de los datos al modelo determinarán la bondad de los ajustes.

Desde el punto de vista educativo, las características que Batanero, Estepa y Godino (1991) atribuyen al AED son:

- *Posibilidad de generar situaciones de aprendizaje interesantes para el alumno*. Analizar datos relativos a un tema de interés del alumno dota a la situación-problema de una idoneidad afectiva (emocional) destacada para su implicación.
- *No necesita una teoría matemática compleja*. Por ejemplo, no se necesita introducir la inferencia. Las nociones que se ponen en práctica podrán ser introducidas gradualmente y se pueden ajustar a la etapa educativa del alumno.
- *Fuerte apoyo en representaciones gráficas y uso de diferentes escalas o re-expresión*. De este modo se ejercitará uno de los componentes del pensamiento estadístico, según Wild y Pfannkuch (1999), pues el cambio de representación proporciona nuevo conocimiento y crea nuevas situaciones-problema las cuales posibilitan que emerjan nuevos objetos matemáticos.

1.3.2.2. CORRELACIÓN Y REGRESIÓN EN OTRAS APROXIMACIONES A LA ESTADÍSTICA

La correlación y regresión se pueden estudiar también desde las diferentes ramas en que tradicionalmente se ha dividido la estadística, en primer lugar, en el estudio de las variables estadísticas bidimensionales o multidimensionales dentro de la estadística descriptiva, que es el tema que abordaremos en el apartado 1.4.

Aparecen como parte del cálculo de probabilidades, en el estudio de las variables aleatorias bidimensionales, ya que la distribución normal bivariante se caracteriza porque la función de densidad conjunta de la variable *(X,Y)* tiene forma acampanada, y viene dada por:

$$f(x,y)=\frac{1}{2\cdot\pi\cdot\sigma_x\cdot\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2\cdot(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2\cdot\rho\cdot x\cdot y}{\sigma_x\cdot\sigma_y}\right)\right)$$

Siendo ρ el coeficiente de correlación de las dos variables; además, el lugar geométrico de las medias marginales es la función de regresión:

$$E[Y/X=x]=\mu_y+\rho\cdot\sigma_y\frac{(x-\mu_x)}{\sigma_x}$$

$$V[Y/X=x]=\sigma_y^2\cdot(1-\rho^2)$$

En la estadística inferencial se ponen en relación estos dos aspectos, utilizándose los datos de las muestras para estimar los coeficientes de correlación y regresión. Por ejemplo, Olivo (2008) describe con detalle el método de construcción de intervalos de confianza para el coeficiente de correlación y los parámetros de la recta de regresión, tema obligado en los cursos de ingeniería.

En análisis multivariante, los datos estadísticos de un mismo sujeto se representan geométricamente por medio de un punto en un espacio n-dimensional, en que cada variable representa uno de los ejes de coordenadas en dicho espacio. El coeficiente de correlación entre dos variables cualesquiera en dicho espacio es una noción fundamental, pues corresponde al coseno del ángulo que forman las dos variables, de modo que en caso de independencia, las variables serían ortogonales, y cuando coinciden, es porque el valor absoluto del coeficiente de correlación es igual a la unidad. La regresión, como en el caso del cálculo de probabilidades, representa el lugar geométrico de los valores medios de las distribuciones condicionadas de una variable con los diferentes valores de la otra (Peña, 2002).

Métodos como el análisis discriminante o el análisis factorial, derivan directamente de esta interpretación de la correlación y regresión, por lo cual, la comprensión de las anteriores será necesaria para la de éstos. Más aún, tanto el coeficiente de correlación y la proporción de varianza explicada (cuadrado de dicho coeficiente) sirven de base a los modelos de análisis de varianza univariantes y multivariantes.

1.4. PRIMERA APROXIMACIÓN AL SIGNIFICADO DE LA CORRELACIÓN Y REGRESIÓN EN NUESTRO ESTUDIO

En las secciones anteriores se han expuesto los diferentes puntos de vista con que se pueden estudiar las nociones de correlación y regresión, cada una de las cuales constituiría un significado institucional diferenciado. Para describir con más precisión los objetivos e hipótesis de nuestra investigación, en esta sección se expone una primera aproximación del significado institucional de estos objetos en nuestro trabajo.

El estudio se realizará desde el punto de vista de la estadística descriptiva, acercándonos, cuando sea posible, a la filosofía del análisis exploratorio de datos descrito con anterioridad. Además, puesto que se trata de objetos sistémicos (compuestos a su vez de otros objetos matemáticos de gran importancia), dividimos el análisis en los siguientes apartados: dependencia aleatoria; distribuciones marginales y condicionadas; covarianza; coeficiente de correlación; ajuste de una función de regresión; y bondad de ajuste.

1.4.1. DISTRIBUCIÓN BIDIMENSIONAL Y DEPENDENCIA ALEATORIA

En primer lugar, como hemos dicho, tratamos el problema de la búsqueda de posible relación entre dos variables. Para el caso de fenómenos donde al observar pares de valores correspondientes a las variables estadísticas implicadas, no sea posible encontrar una fórmula que relacione, de un modo funcional,podemos encontrar diferentes grados y tipos de relación.

Para realizar un estudio estadístico de este tipo tomamos para cada individuo valores de dos variables estadísticas: *X* que toma los valores $x_1, x_2, \ldots, x_r$ e *Y*, que toma los valores $y_1, y_2, \ldots, y_s$. Podemos representar por f_{ij} a la frecuencia absoluta con que aparece el par (x_i, y_j), y mediante h_{ij} a su frecuencia relativa, esto es:

$$h_{ij} = \frac{f_{ij}}{n}$$

para $i = 1, \ldots, r$ y $j = 1, \ldots, s$, donde *n* representa el tamaño de individuos estudiados. Llamaremos distribución bidimensional el conjunto de pares $\{(x_i, y_j): i = 1, \ldots, r; j = 1, \ldots, s\}$,junto con sus frecuencias f_{ij} o h_{ij}, que llamamos frecuencias dobles o bidimensionales.

Al representar la distribución bidimensional en un sistema cartesiano, donde cada variable se hace corresponder con un eje de coordenadas, los puntos no siempre se ajustan de un modo exacto a una función, sino que se obtiene un conjunto de puntos más o menos dispersos. Una representación de ese tipo recibe el nombre de *nube de puntos o diagrama de dispersión* (Ver ejemplos en la Figura 1.4.1). En ambos la relación es directa, el primero lineal, y en el segundo curvilínea.

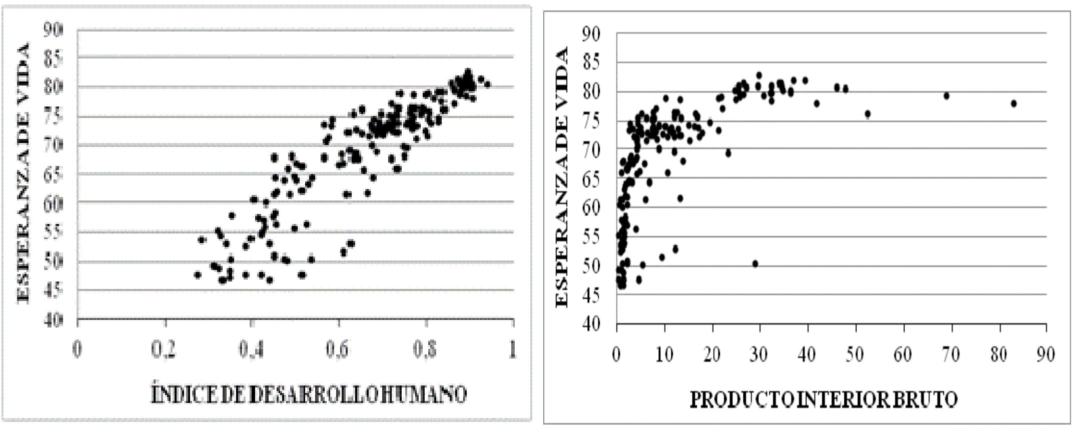


Figura 1.4.1. Esperanza de vida del hombre en función de dos variables

El estudio de la posible relación entre dos variables cuantitativas suele iniciarse mediante la observación del correspondiente diagrama de dispersión o nube de puntos. La presencia de una relación entre las variables se pondrá de manifiesto en el diagrama por una cierta tendencia de los puntos a acumularse en las proximidades de una línea, como hemos visto en los ejemplos anteriores.

1.4.2. DISTRIBUCIONES MARGINALES Y CONDICIONADAS

A partir de las frecuencias bidimensionales pueden obtenerse diferentes distribuciones unidimensionales. Podemos representar con $f_{\cdot j}$ el número de individuos con un valor de la variable $Y = y_j$, independientemente del valor *X*, y por consiguiente obtener la *distribución marginal* de la variable *Y*. De forma análoga podemos definir la *distribución marginal* de la variable *X* (Johnson y Kuby, 2004).

Otro tipo de distribución unidimensional es la que puede obtenerse fijando un valor $Y = y_j$, que se conoce como *distribución de X condicionada* para $Y = y_j$, para la cual, la frecuencia absoluta de la distribución de *X* condicionada por un valor de $Y = y_j$ coincide con f_{ij}, es decir, con la de la variable bidimensional. Si representamos por $h(x_i|y_j)$ la frecuencia relativa condicional del valor x_i entre los individuos que presentan el carácter y_j, obtenemos la igualdad:

$$h(x_i|y_j) = \frac{f_{ij}}{f_{\cdot j}}$$

En el caso anterior hemos hallado la distribución condicional de la variable *X* en función de uno de los valores de *Y*. Podríamos intercambiar los papeles de filas y columnas y obtener la distribución condicional de *Y* en función de alguno de los valores de *X*, y obtener la frecuencia relativa de y_j condicionada por $X = x_i$ mediante la expresión $h(y_j|x_i) = \frac{f_{ij}}{f_{i\cdot}} = \frac{h_{ij}}{h_{i\cdot}}$. Como consecuencia se verifica la igualdad siguiente, que nos permite obtener la frecuencia relativa doble a partir de las condicionales y marginales.

$$h_{ij} = h(x_i|y_j) \cdot h_{\cdot j} = h(y_j|x_i) \cdot h_{i\cdot}$$

El mayor interés del estudio de las distribuciones condicionadas es que, a partir de ellas estamos en condiciones de definir el concepto de dependencia aleatoria. Se dice que la variable *X* es *independiente* de *Y* si todas las distribuciones de frecuencias relativas que se obtienen al condicionar *X* por diferentes valores de $Y = y_j$ son iguales entre sí, e iguales a la distribución marginal de la variable *X*, es decir, cuando se verifica para todo par de valores *i, j:*

$$h(x_i|y_j) = h_{i\cdot}$$

Esta propiedad indica que todas las distribuciones condicionales por columna coinciden con la distribución marginal de la variable *X*, o lo que es lo mismo, la distribución de *X* no cambia cuando se condiciona por cualesquiera valor de *Y*. En el caso de independencia, se cumplen, además, las propiedades siguientes (Johnson y Kuby, 2004):

$$h_{ij} = h_{i\cdot} \cdot h_{\cdot j} \quad \text{para todo } i,j$$
$$h\left(y_j \middle| x_i\right) = h_{\cdot j} \quad \text{es decir, } Y \text{ no depende de } X$$

1.4.3. COVARIANZA

Para variables cuantitativas podemos emplear algunos coeficientes cuyo valor nos indica el tipo de relación entre las variables. El primero de ellos es la covarianza S_{xy} cuya fórmula de cálculo viene dada en la expresión

$$S_{xy} = \frac{1}{n} \sum_{i,j=1}^{r,s} (x_i - \bar{x}) \cdot (y_j - \bar{y})$$

Es decir, para calcular la covarianza, para cada uno de los puntos (x_i, y_j) restamos a cada valor x_i su media $\bar{x}$, y el resultado lo multiplicamos por la diferencia entre y_j y su media $\bar{y}$. La covarianza tiene la propiedad de ser igual a cero si las variables son independientes, positiva si las variables tienen dependencia directa, y negativa en el caso de dependencia inversa.

La covarianza y la correlación pueden ser comprendidas de modo intuitivo considerando patrones en el diagrama de dispersión, para posteriormente utilizar las fórmulas. Para tal efecto, es interesante el razonamiento de Holmes (2001). Partiendo del diagrama de dispersión y dividiendo éste en cuatro cuadrantes (mediante las rectas perpendiculares referidas a las medias de cada variable unidimensional, ver Figura 1.4.2) es posible distinguir aquellos puntos que varían conjuntamente tanto en sentido positivo como en sentido negativo.

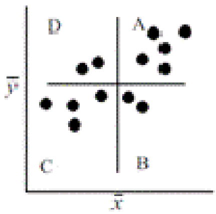


Figura 1.4.2. Diagrama de dispersión dividido en cuatro cuadrantes (Holmes, 2001, p.68)

Supongamos que existen *n* puntos y denotamos por *n(A), n(B), n(C)* y *n(D)* al total de puntos que caen en la región A, B, C y D, respectivamente. Definimos el siguiente coeficiente, estudiando sus propiedades

$$cor = \frac{n(A) + n(C) - n(B) - n(D)}{n}$$

- Si todo punto se encuentra en las regiones A y C, *cor* toma el valor 1.
- Si todo punto se encuentra en B y D, *cor* toma el valor -1.
- Si hay puntos en 3 ó 4 regiones *cor* se encuentra entre -1 y 1.
- Si los puntos son predominantes en A y C *cor* será positivo.
- Si los puntos son predominantes en B y D *cor* será negativo.

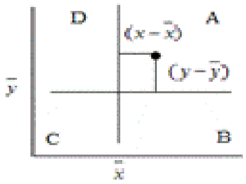


Figura 1.4.3 Distancia a la media de cada dato (Holmes, 2001, p.68)

Si se asigna más peso a los puntos que se encuentren más lejos de las líneas divisorias, llegaremos a una fórmula más útil. Las distancias de las líneas son dadas por las diferencias: $(x-\bar{x})$ e $(y-\bar{y})$ mostradas en la Figura 1.4.3. El promedio de estos productos es precisamente la covarianza, donde observamos que:

1. Si la dependencia entre las variables es directa, la mayor parte de los puntos del diagrama se sitúan en los cuadrantes (1) y (3). Ahora bien, si un punto está en el cuadrante (1) su valor x_i es superior al de la media $\bar{x}$ y su valor y_j es superior a la media $\bar{y}$, luego el producto $(x_i-\bar{x})(y_j-\bar{y})$ tiene signo positivo. Igualmente, para los puntos situados en el cuadrante (3) el producto $(x_i-\bar{x})(y_j-\bar{y})$ tiene signo positivo. Por tanto, en el caso de dependencia directa el signo de la covarianza será positivo, puesto que la mayoría de los sumandos son positivos.
2. Si la dependencia entre las variables es inversa podemos mostrar de forma análoga que el signo de la covarianza es negativo.
3. El caso restante, de independencia, corresponde a la covarianza nula, donde los productos positivos y negativos se compensarían, resultando su suma próxima a cero.

1.4.4. COEFICIENTE DE CORRELACIÓN

Un problema con la covarianza es que no hay un máximo para el valor que puede tomar, por lo cual, no nos sirve para comparar la mayor o menor intensidad de la relación entre las variables. Un coeficiente que permite estudiar, no sólo la dirección de la relación, sino también su intensidad, es el *coeficiente de correlación lineal* o coeficiente de Pearson, que se define por la relación siguiente, siendo S_x y S_y las desviaciones típicas de las variables *X* e *Y*, respectivamente, en la muestra analizada:

$$r=\frac{S_{xy}}{S_xS_y}=\frac{1}{n}\frac{\sum(x-\bar{x})\cdot(y-\bar{y})}{S_xS_y}=\frac{\sum(x-\bar{x})\cdot(y-\bar{y})}{\sqrt{\sum(y-\bar{y})^2\sum(x-\bar{x})^2}}$$

Puesto que las desviaciones típicas son siempre positivas, *r* tiene el mismo signo que la covarianza, y por tanto (Batanero y Díaz, 2008):

- Si $r > 0$ la relación entre las variables es directa;
- Si $r < 0$ la relación entre las variables es inversa;
- Si $r = 0$ las variables son independientes.

Además, el coeficiente de correlación *r* es siempre un número real comprendido entre -1 y 1.

- Cuando existe una relación lineal funcional, esto es, todos los puntos se encuentran sobre una recta (que es el caso de máxima asociación) el valor de *r* será 1 si la recta es creciente (relación directa) o -1 si la recta es decreciente (relación inversa);
- Cuando las variables son independientes, $r = 0$, pues la covarianza es igual a cero;
- Los casos intermedios son aquellos en que existe dependencia aleatoria entre las variables. Esta dependencia será más intensa cuanto más se aproxime a 1 o -1 el coeficiente de correlación.

1.4.5. AJUSTE DE UNA FUNCIÓN DE REGRESIÓN A LOS DATOS

En caso que exista una correlación significativa entre dos variables numéricas, podemos plantearnos un nuevo problema que consiste en tratar de determinar la ecuación de una función matemática que nos permita predecir una de las variables *(Y)* cuando conocemos la otra variable *(X)*. Esta función será el *modelo de regresión de Y en función de X*. Esto puede ser útil cuando la variable *Y* se refiere a un acontecimiento futuro, mientras que *X* se refiere al presente o pasado, o si la variable *X* es más fácil de medir que la *Y* (Batanero y Díaz, 2008).

Para cualquier tipo de función de regresión que sea necesario ajustar a una cierta nube de puntos, el problema que se plantea es determinar los parámetros de la curva particular, perteneciente a una familia de funciones posibles, que mejor se adapte a la muestra de datos de que se disponga. Cuando la forma de la nube de puntos sugiere que una recta de ecuación $Y = a + b \cdot X$ puede ser apropiada como línea de regresión, será necesario calcular las constantes *a* y *b*. Si, por el contrario, es más apropiada una parábola de ecuación $Y = a + b \cdot X + c \cdot X^2$, se precisa determinar tres constantes: *a, b* y *c*.

El principio general que se utiliza para calcular dichas constantes se conoce con el nombre de *criterio de los mínimos cuadrados*. Está basado en la idea de que a medida que una curva se ajusta mejor a una nube de puntos, la suma de los cuadrados de las desviaciones d_i de todos los puntos, es más pequeña. Esta propiedad se visualiza en forma dinámica en el applet de regresión de la Consejería de Educación de Navarra (http://docentes.educacion.navarra.es/msadaall/geogebra/figuras/e3regresion.htm; Figura 1.4.4).

La desviación o residuo del punto (x_i, y_j) respecto de la curva, es la diferencia entre la ordenada y_j del punto y la ordenada de un punto de la curva que tiene la misma abscisa x_i. Es decir $d_i = y_i - (a + b \cdot x_i)$. El procedimiento de obtención de las constantes es hacer mínima la cantidad $D = \sum [y_i - f(x_i)]^2$ siendo $f(x_i) = a + b \cdot x_i$, si lo que se trata de ajustar es una línea recta.

Como consecuencia se obtienen las cantidades a y b que vienen dadas por:

$$b = \frac{S_{xy}}{S_x^2} \quad ; \quad a = \bar{y} - b\bar{x}$$

siendo S_x^2 la varianza de la variable *X*, $\bar{x}$, $\bar{y}$ las medias de *X* e *Y* respectivamente y S_{xy} la covarianza de *X* e *Y*.

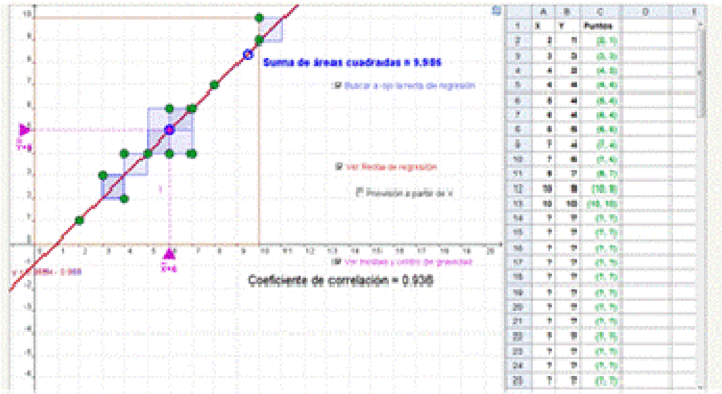


Figura 1.4.4. Desviaciones a la recta de regresión (Applet del servidor de Educación de Navarra)

Como puede apreciarse, el par $(\bar{x} - \bar{y})$, satisface la ecuación de regresión; esto es, la recta pasa por el *centro de gravedad* de la nube de puntos, formado por las dos medias: $(\bar{x}, \bar{y})$. La constante *b*, pendiente de la recta de regresión, recibe el nombre de *coeficiente de regresión* de *Y* sobre *X*.

Nótese que la ecuación de la recta de regresión de *Y* sobre *X* expresa los valores medios de la variable *Y* para cada valor fijo de *X*. También puede plantearse el problema de hallar la recta que determine los valores medios de *X* en función de cada valor de *Y*, es decir, el cálculo de la *recta de regresión* de *X* sobre *Y*. En este caso, intercambiando en la expresión de la recta de regresión los papeles de las variables, obtenemos:

$$(x_i - \bar{x}) = (y_i - \bar{y})\frac{S_{xy}}{S_y^2}$$

Esta recta es, en general, diferente de la recta de regresión de *Y* sobre *X*, aunque si *r* se aproxima a 1 son similares. También pasa por el punto $(\bar{x}, \bar{y})$.

1.4.6. BONDAD DE AJUSTE: COEFICIENTE DE DETERMINACIÓN

La cantidad D/n o *varianza residual* representa la fracción de la varianza de *Y* que es debida al azar, o sea, a las desviaciones de las observaciones y_j respecto de la recta de regresión, y puede demostrarse que es igual a $1 - r^2$, siendo *r* el coeficiente de correlación. El cuadrado del coeficiente de correlación r^2, llamado *coeficiente de determinación*, representa la fracción de la varianza de *Y* debida o explicada por la regresión. En el caso de que se desee ajustar a la nube de puntos una ecuación polinómica de grado n ($P_n(X) = a + b_1 \cdot X + b_2 X^2 + ... + b_n X^n$), será preciso minimizar la expresión siguiente, obteniendo un sistema de ecuaciones que nos permite determinar los parámetros $a, b_1, b_2, ..., b_n$:

$$S_r^2(X) = \sum[y_i - (a + b_1 \cdot x_i + b_2 x_i^2 + ... + b_n x_i^n)]^2$$

En caso de regresión no lineal, el coeficiente de determinación (R^2) representa la proporción de varianza en la variable dependiente que es explicada por el modelo de regresión y no coincidirá con r^2.

1.4.7. SÍNTESIS DEL SIGNIFICADO INSTITUCIONAL

Del análisis realizado en las secciones anteriores, podemos identificar algunos objetos matemáticos asociados al estudio descriptivo de la correlación y regresión, que se presentan en las Tablas 1.4.1 y 1.4.2. Posteriormente, estas tablas serán refinadas con el análisis de los libros de texto (Capítulo 4), uno de cuyos productos será la caracterización más completa del significado institucional de referencia para el tema en el Bachillerato.

Tabla 1.4.1. Problemas, conceptos y procedimientos en el estudio de la correlación y regresión

La situación problema	
	P0.Organización/Reducción de datos
	P1.Analizar la existencia de relación entre variables
	P2.Predecir una variable en función de otra
Conceptos	
	C1.Variable estadística bidimensional; distribución
	C2. Distribución marginal y condicional
	C3.Dependencia funcional/estadística/independencia
	C4.Covarianza y/o correlación
	C5.Regresión
	C6.Coeficiente de determinación
Procedimientos	
	PC1.Representación tabular
	PC2.Representación gráfica
	PC3.Traducción entre representaciones
	PC4.Cálculo e interpretación covarianza y/o correlación
	PC5.Cálculo e interpretación de modelos de regresión
	PC6.Predicción a través de modelos de regresión.
	PC7.Uso de la tecnología
	PCI1.Lectura de gráficos
	PCI2.Lectura de tablas.

Tabla 1.4.2. Propiedades en el estudio descriptivo de la correlación y regresión

Distribución bidimensional	
	PP1.Relaciones entre tipos de frecuencias
	PP2.Centro de gravedad
Dependencia funcional/estadística/independencia	
	PP3.Tipos de covariación
	PP4.Dependencia estadística/correlación y causalidad
Covarianza y correlación	
	PP5.Covarianza: signo; puntos en cuadrantes; datos atípicos
	PP6.Correlación: signo y rango de variación
	PP7.Intensidad de la dependencia
Regresión	
	PP10.Recta de mínimos cuadrados
	PP11.Regresión lineal y no lineal.
	PP13.Dos rectas diferentes; diferencia v. dependiente/independiente
	PP14.Centro de gravedad y recta de regresión
	PP15.Significado de los coeficientes
Relaciones entre conceptos	
	PP16.Signo de correlación, signo covarianza y pendiente rectas de regresión.
	PP17.Regresión y correlación
	PP18. Dispersión de la nube, correlación y regresión

Observamos, en primer lugar, que podemos considerar tres campos principales de problemas, a partir de los cuáles emergen el resto de objetos matemáticos: la organización de los datos bivariantes, el estudio de la relación entre variables (problema de correlación) y la determinación de una función de ajuste (problema de regresión).

Nuestro análisis también nos ha permitido determinar algunos conceptos, propiedades y procedimientos importantes en el tema, que posteriormente se ampliarán con el estudio de los libros de texto. En estas tablas no hemos introducido el lenguaje ni los argumentos, pues preferimos observar los que se empleen en la enseñanza del tema.

1.5. ENSEÑANZA DE LA CORRELACIÓN Y REGRESIÓN

La exposición realizada justifica la importancia que la estadística en general, y la correlación y regresión en particular, poseen en la formación integral de nuestros estudiantes. En lo que sigue, primero haremos una breve referencia a los contenidos de estadística y su secuenciación en las orientaciones nacionales y en el sistema educativo andaluz, y seguidamente analizamos con más detalles los contenidos concretos sobre correlación y regresión.

1.5.1. CONSIDERACIONES GENERALES SOBRE LA ENSEÑANZA DE LA ESTADÍSTICA

La enseñanza de la estadística comienza ya en la Educación Infantil en Andalucía (Consejería de Educación, 2008a, p. 22), donde se alude al desarrollo de la capacidad de "*comprender y representar algunas nociones y relaciones lógicas y matemáticas referidas a situaciones de la vida cotidiana*" y se considera relevante que "*los niños y las niñas constaten la existencia en nuestras vidas de situaciones con interrogantes o incógnitas cuya resolución exige la reflexión sobre ellas y la aplicación de esquemas de pensamiento.*".

Mediante situaciones vinculadas a su entorno y vivencias cotidianas, se plantea la necesidad de propuestas educativas que impliquen la recogida de datos y la organización de los mismos. Con la ayuda del maestro, se intenta que los niños describan los datos recogidos, y se pueda discernir desde una terminología cercana y comprensible, si una situación es probable o improbable. También la encontramos en el segundo ciclo de Educación Infantil, dentro del área de *Conocimiento del entorno*, concretamente en el *Bloque I: Medio físico: elementos, relaciones y medidas*, en la sección de *Elementos y relaciones. La representación matemática* (Consejería de Educación , 2008a, p. 33).

Desde una temprana edad debe tratarse la situación aleatoria (Consejería de Educación, 2008a), pues los alumnos perciben el mundo que les rodea desde una concepción determinista, por lo que el razonamiento estadístico debiese hacerse notar, proporcionando a los niños ricas experiencias que favorezcan el reconocimiento de los aspectos aleatorios y deterministas de nuestro entorno.

La Educación Primaria y Educación Secundaria Obligatoria consideran la estadística como parte integrante del área de matemáticas, si bien, la formulación, las ideas, y las técnicas estadísticas en la Educación Secundaria Obligatoria será, lógicamente, más compleja, respondiendo además a la diferencia de contenidos relativos a la opcionalidad de la asignatura de Matemáticas (opción A y opción B) en el cuarto curso de esta etapa (MEC, 2006; 2007a). En cuanto a la educación secundaria postobligatoria, destacamos la presencia de la estadística en dos de las tres modalidades en que se desarrolla el Bachillerato (MEC, 2007 b): *Ciencias y Tecnología* (asignaturas de *Matemáticas I* y *II*) y *Humanidades y Ciencias Sociales* (asignaturas de *Matemáticas aplicadas a las Ciencias Sociales I* y *II*).

A todas estas consideraciones, hemos de añadir el aspecto multidisciplinar del conocimiento estadístico, ya que podemos relacionar la estadística con las Ciencias Sociales, Biología, Geografía e Historia. Batanero (2001) señala la importancia del aprendizaje de la estadística, con una doble funcionalidad:

- Ayudan al alumno a *entender su entorno.*
- Preparan el *conocimiento posterior*.

Debido al escaso tiempo de que se dispone en el aula, algunos profesores justifican la necesidad de ceñir la enseñanza de los contenidos estadísticos a la transmisión de una serie de fórmulas, o bien, la omiten. Por otro lado, la idea de que la enseñanza de la estadística debe posponerse a los cursos próximos a la Universidad o la Escuela Politécnica, es contraria a los principios educativos actuales, pues ello implicaría que una parte importante de la población quedase privada de la adquisición de conceptos y técnicas de gran importancia para su vida. Así, la investigación en Educación Estadística existente, como por ejemplo los pioneros de Fischbein (1975), demuestran la posibilidad de iniciar a los niños en el estudio de las situaciones aleatorias desde edades tempranas.

Dadas estas consideraciones sobre la enseñanza de los contenidos estadísticos, nos centramos a continuación en las nociones de correlación y regresión. También se hará una breve comparación con el currículo americano.

1.5.2. LA CORRELACIÓN Y REGRESIÓN EN LAS ORIENTACIONES CURRICULARES

Aunque la correlación y regresión no son tratadas como contenidos de enseñanza en la educación obligatoria, encontramos una breve mención a las tablas de contingencia ya en la Educación Primaria, donde se indica: *"lectura e interpretación de tablas de doble entrada de uso habitual en la vida cotidiana"* (MEC, 2006, p.43099).

También implícitamente se presenta en la asignatura de Matemáticas, en cuarto curso de Educación Secundaria Obligatoria, en sus dos opciones: A y B. Encontramos el siguiente criterio de evaluación: "*Se pretende, además, que los resultados obtenidos se utilicen para la toma de decisiones razonables en el contexto de los problemas planteados*" (MEC, 2007a, pp. 758 y 760), relacionado con a la aplicación de los conceptos y técnicas de cálculo de probabilidades para resolver diferentes situaciones y problemas de la vida cotidiana. Además, es presente la relevancia y el sentido educativo implícito de estas nociones en el currículo correspondiente a la Educación Secundaria Obligatoria en Andalucía, que para el núcleo temático *Interpretación de fenómenos ambientales y sociales a través de las matemáticas* se indica:

> El estudio de las relaciones entre las variables y su representación mediante tablas, gráficas y modelos matemáticos contribuirá a describir, interpretar, predecir y explicar fenómenos económicos, sociales o naturales (Consejería de Educación, 2007, p.55).

Centrándonos en la enseñanza de Bachillerato, la correlación y regresión se incluyen en los contenidos de las asignaturas *Matemáticas I* y *Matemáticas aplicadas a* *las Ciencias Sociales I* de las modalidades de *Ciencias y Tecnología* y *Humanidades y Ciencias Sociales*, respectivamente (MEC, 2007b), donde se concretan estos contenidos en el bloque 4: "*Estadística y Probabilidad: Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal"* (p. 45449). También se especifican los criterios de evaluación a seguir, entre los que destacamos los dedicados concretamente a estas nociones:

> Asignar probabilidades a sucesos correspondientes a fenómenos aleatorios simples y compuestos y utilizar técnicas estadísticas elementales para tomar decisiones ante situaciones que se ajusten a una distribución de probabilidad binomial o normal.
> En este criterio se pretende medir la capacidad para determinar la probabilidad de un suceso, utilizando diferentes técnicas, analizar una situación y decidir la opción mas conveniente. También se pretende comprobar la capacidad para estimar y asociar los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden. (MEC, 2007b, p. 45450).

En las *Matemáticas aplicadas a las Ciencias Sociales I*, se concretan estos contenidos en el bloque 3 "*Probabilidad y estadística*" (MEC, 2007b, p. 45475):

> Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación grafica de una nube de puntos.
> Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados.

Igualmente se especifican los criterios de evaluación a seguir, entre los que destacamos los dedicados concretamente a estas nociones:

> Distinguir si la relación entre los elementos de un conjunto de datos de una distribución bidimensional es de carácter funcional o aleatorio e interpretar la posible relación entre variables utilizando el coeficiente de correlación y la recta de regresión.
> Se pretende comprobar la capacidad de apreciar el grado y tipo de relación existente entre dos variables, a partir de la información grafica aportada por una nube de puntos; así como la competencia para extraer conclusiones apropiadas, asociando los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden. En este sentido, más importante que su mero cálculo es la interpretación del coeficiente de correlación y la recta de regresión en un contexto determinado (MEC, 2007b, p. 45475).

Al desarrollar los bloques de contenidos propuestos en el *Real Decreto 1467/2007*, la Consejería de Educación de Andalucía (2008b) indica que debe tenerse en cuenta núcleos transversales: La resolución de problemas; Aprender de y con la Historia de las Matemáticas, e Introducción a los métodos y fundamentos matemáticos; incluyendo la Modelización matemática para la asignatura *Matemáticas I*.

1.5.3. PERSPECTIVA INTERNACIONAL

La correlación y regresión se incluyen igualmente en los documentos curriculares de otros países en la etapa de secundaria. Como ejemplo, analizamos el currículo americano.

NCTM (2000)

Los Estándares del *National Council of Teachers of Mathematics* justifican el bloque de *Análisis de Datos y Probabilidad* por la cantidad de datos disponibles para la toma de decisiones en los diversos campos profesionales y en la vida diaria.

Indican que con frecuencia, la información estadística se utiliza de manera incorrecta para orientar la opinión pública en una determinada dirección, o dar una imagen sesgada de la efectividad de algunos productos comerciales. Esto lleva a la necesidad de que los ciudadanos tengan una base suficiente sobre el análisis de datos y algunas nociones elementales de probabilidad, para convertirse en ciudadanos informados y consumidores inteligentes. Respecto a la correlación y la regresión, se mencionan contenidos relacionados en los siguientes estándares:

1. Formular preguntas que pueden ser atendidas con datos y recolectar, organizar y mostrar datos relevantes para responderlas:
   - En los grados 6-8 todos los estudiantes deben formular preguntas, diseñar estudios y recoger datos acerca de una característica compartida por dos poblaciones, o características diferentes dentro de una población; también han de seleccionar, crear, y usar adecuadamente representaciones gráficas de datos, incluyendo histogramas, diagramas de caja, y diagramas de dispersión;
   - En los grados 9-12 todos los estudiantes deben comprender el significado de los datos de medida y categóricos, univariados y bivariados, y el término variable; asimismo comprender los histogramas, diagramas de caja paralelos y diagramas de dispersión y utilizarlos para representar los datos;
2. Seleccionar y usar métodos apropiados para analizar los datos:
   - En los grados 6-8 todos los estudiantes deben discutir y comprender la correspondencia entre los conjuntos de datos y su representación gráfica, especialmente con los histogramas, tallo y hojas de gráficos, diagramas de caja, y diagramas de dispersión.
   - En los grados 9-12 todos los estudiantes deben ser capaces de representar datos bivariados en un diagrama de dispersión, describir su forma, y determinar los coeficientes de regresión, las ecuaciones de regresión y coeficientes de correlación con herramientas tecnológicas; deben representar y discutir datos de dos variables, donde al menos una variable es categórica; también identificar las tendencias en los datos de dos variables, y encontrar funciones que ajusten los datos o transformar los datos para que puedan ser modelados.
3. Desarrollar y evaluar inferencias y predicciones basadas en los datos:
   - En los grados 6-8 los estudiantes deben hacer conjeturas sobre las relaciones entre dos características de una muestra en base a un diagrama de dispersión de los datos, y aproximar la recta de mejor ajuste; deben usar estas conjeturas para plantear nuevas preguntas y diseñar estudios que permitan responderlas.

Proyecto GAISE

Este proyecto (Franklin y cols., 2005) completa en Estados Unidos al anterior, y está dirigido a dos grupos de estudiantes: para la educación K-12 y para estudiantes en cursos preuniversitarios. Fueron elaborados por la Asociación Americana de Estadística y el National Council of Teachers of Mathematics. En relación al periodo K-12 (desde preescolar hasta final de la secundaria), se indica que cualquier curso de estadística debe tener como principal objetivo ayudar a los estudiantes en la adquisición de los elementos básicos del razonamiento estadístico, que son los siguientes:

- *La necesidad de datos.* Reconocer la necesidad de basar las decisiones personales en la evidencia (datos), además de reconocer los problemas y peligros que pueden surgir al actuar sobre supuestos que no están respaldados por datos.

- *La importancia de generar buenos datos*. Reconocer la dificultad de conseguir datos de calidad, resaltando que en ocasiones es necesario emplear bastante tiempo en la obtención de datos de buena calidad, y este tiempo no debe considerarse como perdido.

- *La presencia de la variabilidad.* Reconocer que la variabilidad está presente en muchos fenómenos cotidianos. La variabilidad es la esencia de la estadística como disciplina, y no puede ser entendida solamente a través de su estudio, sino que debe ser experimentada.

- *La cuantificación y explicación de la variabilidad.* Reconocer que la variabilidad puede ser medida y explicada, tomando en consideración la aceptación de la idea de aleatoriedad y variable aleatoria. Para ello se estudian y analizan los promedios y estadísticos de dispersión de dichas variables, junto a una variedad de modelos matemáticos y estadísticos.

- *Más datos y conceptos y menos fórmulas.* Cualquier curso en estadística puede ser mejorado poniendo más énfasis en los datos y conceptos, y menos en los algoritmos. En la medida de lo posible, se debe dedicar más tiempo a actividades de interpretación de cálculos estadísticos y gráficos.

- *Fomento del aprendizaje activo.* Los profesores de estadística deberían basarse mucho menos en las lecciones magistrales y mucho más en otras opciones tales como proyectos estadísticos, ejercicios de laboratorio y resolución de problemas en equipo, así como discusión y debates sobre las actividades y resultados de las mismas. Unos de los objetivos sería que los estudiantes tuviesen una participación más activa.

En este informe se contemplan tres niveles de aprendizaje de los contenidos. En relación al tema que nos ocupa, los contenidos sugeridos en cada nivel son los siguientes:

- *Nivel A:* Observar la asociación entre dos variables; estudiar la asociación a partir de representaciones. Por ejemplo, estudiar la asociación entre una variable numérica y otra cualitativa a partir de diagramas de puntos y entre dos variables cuantitativas a partir de diagramas de dispersión.

- *Nivel B:* Cuantificar de algún modo la asociación con modelos simples; notar diferencias de intensidad de la asociación; realizar interpretaciones básicas de modelos de la asociación; comenzar a diferenciar entre asociación y causalidad. Los estudiantes debieran interpretar la asociación en tablas de contingencia a partir de la comparación de proporciones; la asociación entre una variable cuantitativa y otra cualitativa a partir de diagramas de cajas o comparando resúmenes estadísticos; y entre dos variables cuantitativas mediante el número de puntos que cae en las cuatro zonas determinadas por los valores medios de las variables, según sugiere Holmes (2001), y ajustar una línea (en caso de regresión lineal) en forma aproximada, con ayuda de software.

- *Nivel C:* Cuantificar la asociación ajustando modelos a los datos. Interpretar medidas de intensidad de la asociación y modelos de asociación; distinguir las conclusiones que pueden obtenerse de estudios experimentales y correlacionales. Se sugiere como ejemplo que los estudiantes calculen la recta de regresión con ayuda de software, dibujen los residuos para evaluar la bondad de ajuste e incluso determinen los intervalos de confianza del 95% para la predicción del valor de la variable dependiente en función de la independiente.

1.6. OBJETIVOS DEL TRABAJO

Finalizado el análisis conceptual y curricular del tema, pasamos a detallar con mayor precisión los objetivos generales, que se desglosarán en objetivos específicos en los diversos estudios que componen la Memoria. Estos objetivos generales se apoyan en nuestro marco teórico que es el enfoque ontosemiótico para la Didáctica de la Matemática (Ver Capítulo 2) y se concretan en dos:

O1. Realizar un análisis detallado de la presentación de la correlación y regresión en una muestra de libros de texto de Bachillerato, con la finalidad de caracterizar el significado institucional pretendido en las dos modalidades en que se contempla su enseñanza.

El interés de este objetivo se justificará por la inexistencia de estudios comprensivos de libros de texto publicados con posterioridad a la actual normativa, y la escasez de investigaciones sobre la presentación de la correlación y regresión en los libros de texto. Sus resultados permitirán caracterizar el significado institucional pretendido en las modalidades en que se contempla su enseñanza, y contribuirán a destacar puntos a mejorar en esta presentación. Podrán también ser la base para la construcción del significado institucional de referencia, que se utilizará en la construcción de instrumentos de evaluación en el Estudio 2.

Para cumplir este objetivo general, se organiza el Estudio 1, descrito en el Capítulo 4, donde se realiza un análisis profundo de la correlación y regresión en una muestra de 16 textos de Bachillerato. Se categorizan y analizan los diferentes tipos de objetos primarios considerados en nuestro marco teórico, considerando para cada uno de ellos diferentes variables relacionadas. Asimismo, se estudian las referencias a la tecnología, y se proporciona una lista de conflictos semióticos encontrados en los textos analizados.

O2. Realizar un estudio exploratorio de evaluación del conocimiento didáctico-matemático sobre correlación y regresión, según el modelo propuesto en nuestro marco teórico (Godino, 2009; 2011), reinterpretando algunos de sus componentes y facetas en términos del modelo MKT (*Mathematical Knowledge for Teaching*) de Hill, Ball, y Schilling (2008).

Pretendemos acercarnos a los conocimientos de los estudiantes, que se preparan como profesores de matemáticas de Secundaria y Bachillerato, para evaluar algunos de sus conocimientos sobre el tema. Nos apoyaremos en el estudio previo de análisis de libros de texto, y en la metodología propuesta por Godino (2009; 2011).

La importancia de este objetivo se debe a la escasez de estudios sobre el conocimiento didáctico-matemático sobre la correlación y regresión en futuros profesores (como se expone en el Capítulo 3), y la necesidad de aportar una formación adecuada para su tarea docente. Para lograrlo, se diseñan dos actividades formativas, que permiten evaluar y desarrollar algunos elementos de este conocimiento, siguiendo la metodología de investigación basada en diseño.

En el Capítulo 5 se describe la primera de estas actividades formativas, consistente en un proyecto estadístico en que los futuros profesores analizan diversas variables relacionadas con la esperanza de vida, utilizando datos del servidor de las Naciones Unidas. A lo largo del proyecto se les plantean preguntas que permiten evaluar y desarrollar sus conocimientos matemáticos sobre los prerrequisitos (gráficos estadísticos y resúmenes univariantes), la correlación (interpretación de diagramas de dispersión, estimación de la correlación y explicación de la misma) y regresión (elección de un modelo y comprensión del concepto bondad de ajuste). Esta actividad se experimenta en dos cursos sucesivos (Estudio 2).

En el Capítulo 6 se describe el Estudio 3, llevado a cabo con la primera de las muestras participantes del Estudio 2, y que consistió en el análisis de la idoneidad didáctica del proyecto realizado por los mismos participantes. Los resultados describen algunos componentes de las diferentes facetas del conocimiento didáctico de los profesores sobre la correlación y regresión.

1.7. ORGANIZACIÓN DE LA INVESTIGACIÓN Y CARACTERÍSTICAS METODOLÓGICAS GENERALES

La investigación se organiza en una primera parte de fundamentos, y una segunda parte que recoge los tres estudios empíricos citados, cada uno de los cuales corresponde a uno de los objetivos generales anteriormente descritos (Ver Figura 1.7.1). La parte de fundamentación a su vez se descompone en los siguientes apartados:

- *Análisis del contenido matemático*, expuesto anteriormente, que nos ha permitido una primera aproximación al significado institucional de referencia.
- *Análisis curricular*, presentado en este capítulo, que ofrece una perspectiva adecuada al análisis posterior de los libros de texto.
- *Marco teórico*, que se presenta en el Capítulo 2, donde se describen los elementos del enfoque ontosemiótico que se utilizarán en la parte empírica del trabajo.
- *Modelos e investigaciones sobre el conocimiento del profesor de matemáticas y de estadística*, donde analizamos los antecedentes y especificamos el modelo utilizado en este trabajo (Capítulo 2).
- *Antecedentes de investigación sobre la correlación y regresión*, que permite ubicar nuestro trabajo y discutir sus resultados, validando las conclusiones (Capítulo 3).

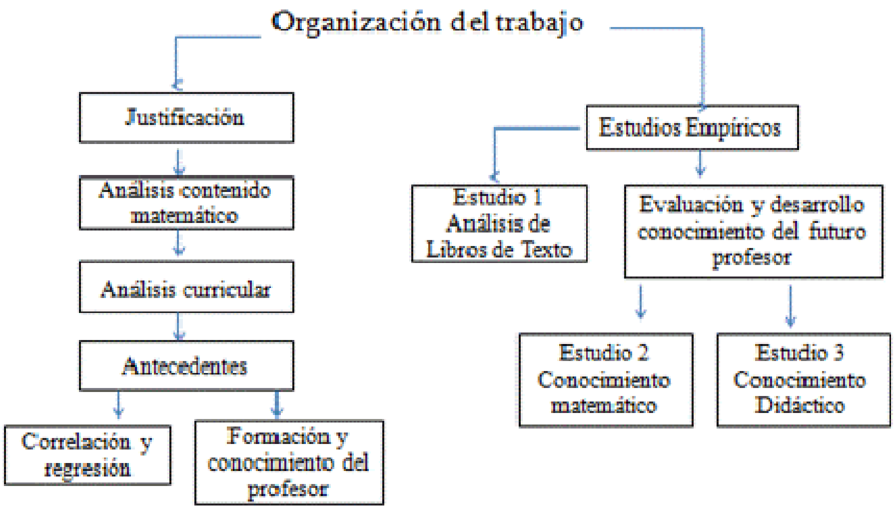


Figura 1.7.1. Organización de la investigación

A continuación resumimos muy brevemente las características metodológicas generales de los estudios empíricos, que se describen con mayor detalle en los capítulos correspondientes.

El primer estudio empírico (Estudio 1, Capítulo 4) es un análisis de 16 libros de texto de Bachillerato (8 de la modalidad de *Ciencia y Tecnología* y 8 de la modalidad de *Humanidades y Ciencias Sociales*) en que estudiamos con detalle los diferentes tipos de objetos matemáticos que se consideran en nuestro marco teórico, y la forma en que se presentan e interrelacionan.

Se trata de una investigación cualitativa de tipo documental. Según las clasificaciones de Bisquerra (1989), el proceso de investigación seguido es inductivo, se trata de una investigación aplicada y descriptiva. La muestra de textos es intencional, aunque elegida de forma que represente adecuadamente los textos utilizados en Andalucía para este nivel de enseñanza. El método de análisis es de tipo cualitativo, y sigue el empleado en trabajos previos, por ejemplo por Cobo (2003). Nos apoyamos en el enfoque ontosemiótico pues se consideran como categorías primarias los tipos de objetos elementales considerados en este enfoque. Estas categorías se desarrollan posteriormente en forma inductiva utilizando la técnica de análisis de contenido (Cook y Reichardt, 2000).

Los resultados principales consisten en ejemplos prototípicos de las diferentes categorías, junto a tablas resumen de presencia de las mismas en los textos, que permiten comparar los textos entre sí, así como unos objetos con otros; y las características de los textos en las dos modalidades de Bachillerato, tanto globalmente, como dentro de la misma editorial.

Los dos estudios de evaluación del conocimiento didáctico-matemático son también básicamente cualitativos, con algunos elementos cuantitativos. Se pueden clasificar como investigación basada en diseño, pues utiliza el diseño y el análisis sistemático de estrategias y herramientas instruccionales (Godino et al., 2013). Para ello se incluye el diseño de dos situaciones didácticas dirigida a evaluar y desarrollar el conocimiento didáctico-matemático sobre la correlación y regresión en futuros profesores. Al mismo tiempo se experimenta la situación didáctica diseñada en el aula en contexto natural, y se evalúan los resultados de la experimentación, proponiendo acciones de mejora de la misma (DBRC, 2003). De este modo, una finalidad esperada es que la investigación resulte, en teoría, compartida por profesores y diseñadores curriculares para apoyarles en su trabajo.

Más concretamente, para los Estudios 2 y 3 se diseñan actividades didácticas que se realizarán en varias sesiones prácticas, que combinan el trabajo con un proyecto estadístico, la discusión colectiva de las soluciones en este proyecto (Estudio 2) y el análisis de la idoneidad didáctica del mismo (Estudio 3). Se diseña un entorno de aprendizaje completo con tareas, materiales, herramientas y medios para secuenciar y apoyar el aprendizaje (Reimann, 2011). Los medios específicos están formados por datos reales tomados del servidor de las Naciones Unidas, la hoja de cálculo Excel que facilita los cálculos, algunos recursos interactivos, así como una guía de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas (Godino, 2011). Todos ellos se describen con detalle en los Capítulos 5 y 6.

La recogida de datos en la secuencia de actividades se realiza a través de dos cuestionarios, que han sido construidos siguiendo la metodología usual recomendada por APA, AERA y NCME (1999) para reforzar su validez de contenido y su fiabilidad y generalizabilidad. La definición semántica de las variables objeto de medición se apoya en fundamentos adecuados para definir "el conocimiento didáctico-matemático sobre la correlación y regresión", que es un constructo inobservable (León y Montero, 2002). Dichos fundamentos son los expuestos anteriormente.

Finalizado el diseño del proyecto y los instrumentos de recogida de datos, se realizó una evaluación exploratoria con el mismo, con la finalidad de mejorar el diseño y revisar las hipótesis previstas sobre el proceso de aprendizaje. Tanto el diseño como la experimentación están apoyados en teorías instruccionales específicas (Cobb y Gravejemeir, 2008), en este caso, el enfoque ontosemiótico, así como los antecedentes de investigación sobre aprendizaje de la correlación y regresión y sobre formación de profesores.

Tomaron parte en la experimentación una muestra de 23 futuros profesores, participantes en el Máster de Formación del Profesorado de Educación Secundaria y Bachillerato, que participó en ambos estudios (Estudio 2 y 3) en el curso 2012-2013 y una segunda muestra de 42 estudiantes que participaron en el Estudio 2 en el curso 2013-2014, donde no se consideró necesario repetir el Estudio 3 en cuanto que el análisis de la primera experimentación del mismo aportó suficiente información. Reconociendo las limitaciones debidas al tamaño moderado de muestra, y al carácter voluntario de la participación de los estudiantes, pensamos que los resultados proporcionan una primera información sobre los conocimientos de estos profesores y los puntos en que se debiera reforzar su enseñanza.

Asimismo, se trata de compensar este tamaño limitado de muestra con una mayor profundidad de análisis de los datos. Por otro lado, los instrumentos construidos pueden utilizarse como base en otros talleres similares con profesores en formación y en ejercicio, con objeto de reforzar su conocimiento didáctico-matemático para la enseñanza de la correlación y regresión.

CAPÍTULO 2.
FUNDAMENTOS

2.1. INTRODUCCIÓN

Finalizada la presentación y justificación del interés del problema de investigación abordado, pasamos en este capítulo a analizar su fundamentación, que consta de dos partes:

1. En primer lugar se analiza el marco teórico, que es el enfoque ontosemiótico de la cognición e instrucción matemática (EOS), cuyas componentes ontológicas, cognitivas, semióticas e instruccionales, se adaptan a nuestros objetivos. De este marco teórico se describen los principales elementos que se utilizarán en el trabajo.

2. Por otro lado, presentamos un análisis de diversos modelos sobre el conocimiento del profesor de matemáticas y de estadística, para seleccionar el que específicamente se utilizará en el trabajo.

2.2. MARCO TEÓRICO

En primer lugar presentamos un resumen de algunos elementos del EOS (Godino, 2002), que considera la matemática desde tres puntos de vista complementarios:

1. Un quehacer humano, para dar solución a cierta clase de situaciones problemáticas, del mundo real o de la propia matemática.
2. Un lenguaje simbólico con función comunicativa e instrumental.
3. Un sistema conceptual lógicamente organizado, donde cualquier nuevo conocimiento se establece dentro de una estructura propia.

El considerar la actividad matemática como una actividad humana, mediada por el lenguaje en los procesos de comunicación y desarrollada en una institución, genera el interés por equilibrar el dominio de lo personal y de lo institucional, evidenciar el papel determinante de la situación-problema, y precisar las nociones de *práctica*, *objeto* y *significado*.

2.2.1. OBJETO MATEMÁTICO, PRÁCTICA Y SIGNIFICADO

En el EOS se considera *objeto matemático* a cualquier entidad a la cual nos referimos, y que interviene de algún modo en la actividad matemática (Font, Godino y D'Amore, 2007). Los problemas planteados, las notaciones utilizadas, los procedimientos llevados a cabo, las proposiciones formuladas, o los argumentos utilizados, entre otros, son entendidos como "objetos matemáticos".

Esta variedad de objetos matemáticos, se ponen en juego y emergen de la actividad matemática realizada en los procesos de enseñanza y aprendizaje. Como parte de esta actividad, los objetos matemáticos son manipulados, nombrados y descritos mediante las *prácticas* asociadas (Godino y Batanero, 1994), que pueden ser personales e institucionales y se describen como:

> Toda actuación o expresión (verbal, gráfica, etc.) realizada por alguien para resolver problemas matemáticos, comunicar a otros la solución obtenida, validarla o generalizarla a otros contextos y problemas (Godino y Batanero, 1994, p. 334).

Godino y Batanero (1994) indican que, asociado a un campo de problemas, existe un sistema de prácticas prototípicas; en nuestro caso, nos interesamos por las relacionadas con los problemas ligados a la correlación y regresión. Los autores entienden por *práctica institucional*, el conjunto de prácticas compartidas en el seno de la institución, que dependen de los instrumentos disponibles en la misma, sus reglas y modos de funcionamiento (Godino y Batanero, 1994; Godino, Batanero y Font, 2007). Por otra parte, el sujeto realiza prácticas personales para alcanzar la solución a la tarea propuesta (que podrían variar respecto a las propuestas en la institución).

En la enseñanza y aprendizaje de las matemáticas, más que una práctica particular ante un problema concreto, interesa considerar los sistemas de prácticas de las personas en su actuación ante tipos de situaciones problemáticas. Así, a la pregunta, ¿qué es un objeto matemático?, se propone como respuesta:

> El sistema de prácticas que realiza una persona (significado personal), o compartidas en el seno de una institución (significado institucional) para resolver un tipo de situaciones-problemas ligadas con el objeto (Godino y Batanero, 1994, p. 335).

Objetos emergentes e intervinientes en los sistemas de prácticas

El EOS considera, en un primer nivel, una tipología de objetos, y en segundo nivel, distintos modos de ver, hablar, operar, etc. sobre los mismos (Godino, Font, Contreras y Wilhelmi, 2006; Godino, Batanero y Font, 2007).

Desde el primer punto de vista, el EOS ofrece una clasificación de objetos en *situación-problema*, *lenguaje*, *conceptos*, *proposiciones*, *procedimientos*, y *argumentos*. Utilizaremos esta clasificación en el análisis de textos presentado en el Estudio 1, y en el análisis de las actividades propuestas a los futuros profesores en el Estudio 2.

- La *situación-problema* son todas las aplicaciones extra-matemáticas o matemáticas. En el caso de la enseñanza, corresponden a las tareas, ejercicios ó actividades planteadas que inducen una actividad matemática. Suelen estar agrupados en tipos o clases de problema; en la enseñanza de un tema deben permitir contextualizar los conocimientos pretendidos (Godino, Font, Contreras y Wilhelmi, 2006). Un ejemplo en nuestro estudio sería determinar la existencia, dirección e intensidad de la relación entre dos variables estadísticas.

- Los elementos del *lenguaje* son los términos, expresiones, notaciones, gráficos, que se utilizan para representar los datos del problema, las operaciones que hacemos con ellos, los objetos matemáticos que se utilizan, y la solución encontrada. Ejemplos en nuestro estudio serán los términos, símbolos y expresiones matemáticas utilizadas al trabajar con la correlación y regresión, así como las representaciones específicas (diagrama de dispersión, o representación gráfica de la función de ajuste).

- L*os conceptos* son evocados implícita o explícitamente por el alumno cuando realiza cualquier acción, con la intención de resolver las cuestiones planteadas. Vienen especificados por su definición en el tema, que el alumno habrá de recordar, en caso de tratarse de conceptos iniciales, o asimilar, si se trata de conceptos nuevos. Entre otros, en el tema que nos ocupa encontramos el de variable y distribución bidimensional, distribución marginal y condicionada, correlación, covarianza o regresión.

- Las *proposiciones, atributos o propiedades* son los enunciados que se realizan sobre el objeto matemático que se trata. Regulan las relaciones de este objeto con otros objetos, y con ello se producirá un enriquecimiento en su significado. Son ejemplos en nuestro estudio la relación entre el signo de la covarianza y de la correlación; o del valor de la correlación y el ángulo que forman las dos rectas de regresión.

- Los *procedimientos* son el conjunto de operaciones, algoritmos, técnicas de cálculo, que permiten resolver la tarea propuesta, y que llegan a automatizarse para tipos específicos de problemas. En nuestro caso podemos destacar los procedimientos de cálculo de diferentes coeficientes, la representación gráfica, o los de predicción a partir de la recta de regresión.

- Los *argumentos* son los enunciados emitidos para validar o explicar los resultados obtenidos, o propiedades utilizadas en el transcurso de la resolución, así como las propiedades o teoremas introducidos por el profesor.

Las relaciones o redes que se establecen en la enseñanza y aprendizaje de un tema, o en la solución de problemas entre varios de estos objetos, dan lugar a entidades más complejas, que el EOS denomina *configuraciones*, que pueden ser *epistémicas*, si son propias de la institución o *cognitiva*s, si son específicas de una persona. Por otro lado, los objetos, en un momento dado, pueden ser emergentes (si aparecen por primera vez) o intervinientes (si los posee el sujeto previamente). En la enseñanza, o de la evaluación, se pueden describir estas configuraciones para analizar el aprendizaje y su relatividad al contexto institucional (D'Amore y Godino, 2007). En la Figura 2.2.1 se muestra las relaciones entre los diferentes tipos de objetos de una configuración, que surgen en una práctica matemática.

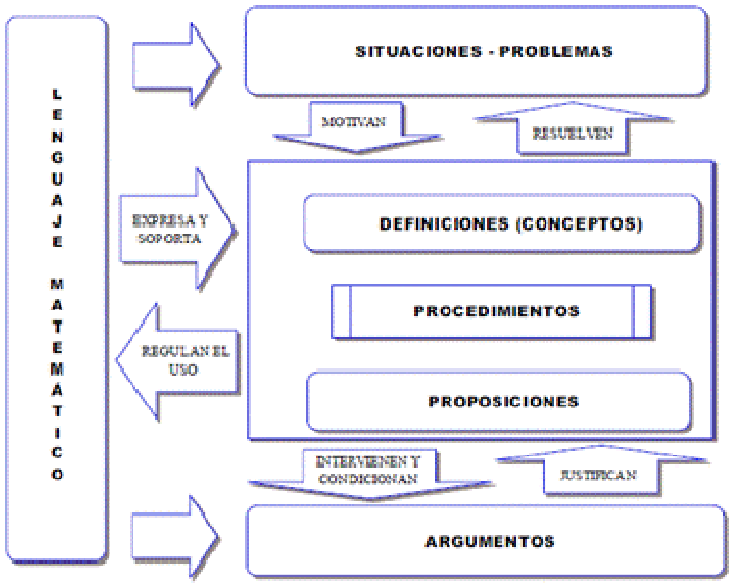


Figura 2.2.1. Configuración de objetos primarios (Godino y Batanero, 2009, p. 7)

2.2.2. FACETAS DUALES DEL CONOCIMIENTO MATEMÁTICO

Los anteriores objetos matemáticos pueden ser vistos desde diferentes perspectivas, que serán tenidas en cuenta a lo largo de nuestro trabajo, y que se describen a continuación.

Faceta institucional y personal

Según se centre la atención en el sujeto, la institución, o la interacción entre ambos, estamos considerando esta faceta del conocimiento matemático. Si se centra en el sujeto, hablaremos de los *objetos personales* (faceta personal del conocimiento matemático). Si por el contrario se considera la institución (documentos curriculares, libros de texto, explicación del profesor, etc.) se habla de *faceta institucional* del conocimiento matemático. En esta dualidad, es importante tener conciencia de los tipos de significado institucional o personal del objeto matemático (Figura 2.2.2).

En cuanto a los significados institucionales, podemos distinguir los significados de *referencia*, *pretendido*, *implementado* y *evaluado*. Cuando el investigador fija el significado que tendrá en cuenta del objeto matemático, se genera el significado institucional de referencia. Para ello, se vale del saber científico (e incluso del desarrollo histórico) y del análisis de las orientaciones curriculares, delimitando con ello las prácticas operativas y discursivas inherentes al objeto que se tomarán como referencia en su trabajo.

Figura 2.2.2. Significados institucionales y personales (Godino, Batanero y Font, 2007, p. 6**)**

Una vez establecido el significado de referencia del objeto matemático, se puede considerar un significado restringido propuesto a los estudiantes, atendiendo a diversos factores: nivel educativo, temporalización, conocimientos previos de los estudiantes, etc. Con ello se concreta el *significado institucional pretendido* del objeto tratado. El profesor basará la enseñanza en el significado pretendido, pero según los condicionantes en el proceso de instrucción, se puede variar este significado, llegando al *significado institucional implementado*. En base a ello, el profesor diseña, de un modo objetivo, la evaluación del aprendizaje; Por tanto, en esa colección de tareas, cuestiones o problemas, se refleja *el significado evaluado*, que se intenta sea representativo del significado implementado.

De igual modo, se distinguen tres tipos de significado personal de un objeto matemático: el significado *global*, *declarado* y *logrado*. El significado global es el conjunto de sistema de prácticas significativas que es capaz de desempeñar el alumno, relativas al objeto matemático tratado. Con el significado declarado nos referimos a las prácticas que el alumno realiza en una determinada tarea propuesta, y que son manifestadas en la evaluación o la investigación. Dentro de ellas, el significado personal logrado sería el conjunto de prácticas que son consideradas correctas en la institución.

Comprensión

En un análisis dirigido al estudio de los cambios producidos por la enseñanza, interesará también tener en cuenta los *significados iniciales* o previos de los estudiantes y los que *finalmente* alcancen. En los puntos en que el significado declarado no se ajuste al significado de referencia, serán manifestados los errores de aprendizaje. Las relaciones dialécticas que surgen en el proceso de enseñanza y aprendizaje (Figura 2.2.2) propician el ajuste entre los significados personales e institucionales.

Aceptando la dualidad personal e institucional del conocimiento, según el EOS, *la comprensión persona*l será considerada como la *apropiación* del significado institucional del objeto matemático por el individuo. El EOS establece la *comprensión de un objeto matemático* como un proceso progresivo. En el modelo de comprensión propuesto por Godino (1996), se tienen en cuenta dos ejes: uno descriptivo, que indica los aspectos o componentes de los objetos de los cuales queremos evaluar su comprensión, y otro procesual que indica las fases o niveles en la comprensión.

Godino (1996) también resalta que la comprensión personal de un determinado objeto por parte de un sujeto, no puede ser observada directamente, pero lo que sí puede observarse directamente son las prácticas personales (significado personal). Por lo tanto, la evaluación de la comprensión sería el estudio de la correspondencia entre los significados personales e institucionales. Las prácticas observables son los indicadores empíricos que nos permiten realizar esta evaluación.

Faceta ostensiva y no ostensiva

Todo objeto matemático tiene una faceta ostensiva, esto es perceptible, y otra no ostensiva. La faceta no ostensiva de un objeto matemático permite ser "*capaz de pensar e imaginar uno de estos objetos sin necesidad de mostrarlo externamente*" (Godino, 2002, p. 9,). El hecho de poder nombrar o representar (mediante el lenguaje oral o escrito, numérico, verbal, gráfico ó simbólico), e incluso reconocer situaciones en que sea necesaria la aplicación de un objeto, evidencian su dimensión ostensiva.

El matiz que aporta el EOS con esta dimensión ostensiva y no ostensiva de un objeto matemático es que puede aplicarse a los distintos objetos. El motivo es que un objeto ostensivo (una palabra escrita, un gráfico, etc.) puede ser también pensado, imaginado, o estar implícito (por ejemplo, el signo de multiplicar en la notación algebraica). Análogamente, un cálculo puede ser realizado por una persona de manera ostensiva, o mentalmente (Godino, 2002).

Faceta extensiva e intensiva

En el trabajo matemático, un aspecto de gran interés es la generalización progresiva de los objetos matemáticos. De este modo, desarrollando procedimientos, estrategias y técnicas generales, conseguimos situar a cada objeto en un estatus de representante de una clase más amplia de objetos.

Evidenciar lo concreto y particular que acontece a un determinado objeto en una situación determinada, y la abstracción que conlleva su generalización, ha de ser reflejado como una dualidad extensiva e intensiva. Para cada uno de los elementos primarios, permitirá distinguir si el objeto interviene como caso particular o como una clase más general o abstracta de objetos. Por ejemplo, una recta de regresión lineal, particular: $y = 3x - 2$, es un elemento extensivo, y cuando la generalizamos a la ecuación general de la regresión: $y = a \cdot x + b$, obtenemos la faceta intensiva.

Faceta unitaria y sistémica

Los objetos matemáticos, concebidos como elementos unitarios, son unidades elementales de significado para el sujeto que las utiliza como herramientas para su trabajo matemático. Al establecer relaciones entre ellos, se ofrece un carácter sistémico a los objetos matemáticos. Por ejemplo, en el estudio de la adición y sustracción en los últimos niveles de primaria, el sistema de numeración decimal (valor posicional: decenas, centenas,…) se considera como algo conocido y que se puede utilizar, y en consecuencia, como entidades unitarias. Estos mismos objetos, en el primer curso, tienen que ser considerados de manera sistémica para su estudio.

Faceta expresión y contenido. Conflicto semiótico.

Dada la diversidad de objetos, así como la dimensión ostensiva y no ostensiva de los mismos, es importante contemplar los procesos de semiosis que se realizan en el desarrollo de la actividad matemática. El *análisis ontológico-semiótico* propuesto por el EOS ofrece un instrumento para reflejar las relaciones entre objetos matemáticos (*función semiótica*) y los procesos interpretativos en las prácticas matemáticas (Godino, 2002; Godino, Batanero y Font, 2007). Cuando entre dos objetos se establece una dependencia en la que uno de los objetos se pone en lugar del otro o bien uno es usado por otro, un objeto matemático adquiere el estatus de significante y el otro de significado en la tarea matemática (Godino y Font, 2007).

El EOS postula que, la expresión y el contenido pueden ser cualquier tipo de objeto. Además, el tipo de relación expresión-contenido no se limita a la acepción representación (Font, Godino y D'Amore, 2007). Atendiendo a las relaciones de dependencia entre la expresión y el contenido se pueden distinguir tres tipos de funciones semióticas: Cuando un objeto se pone en lugar de otro para un cierto propósito hablamos de *función semiótica representacional*. Si un objeto utiliza a otro u otros como instrumento hablamos de *función semiótica instrumental*, y si dos o más objetos componen un sistema del cual emergen nuevos objetos hablamos de *función semiótica estructural* (Godino, Font, Contreras y Wilhelmi, 2006; D'Amore y Godino, 2007).

El alumno activa su propia trama o red/redes de funciones semióticas pudiendo, por una parte, reducir o aumentar la complejidad de dicha trama, y por otra, facilitar la realización, de un modo efectivo, de la tarea propuesta (Godino y Batanero, 1994). Las prácticas que desarrolla el alumno serán consideradas correctas si se conforman a la institución. Surge así la noción de *conflicto semiótico* para tratar cualquier

> Disparidad o discordancia entre los significados atribuidos a una expresión por dos sujetos (personas o instituciones) en interacción comunicativa y pueden explicar las dificultades y limitaciones de los aprendizajes y las enseñanzas implementadas (Godino, 2002, p.250).

Tendremos en cuenta especialmente la idea de conflicto semiótico en el análisis de los libros de texto presentados en el Estudio 1.

Síntesis

A modo de resumen, en la Figura 2.2.3 se muestran los objetos y facetas duales expuestas anteriormente, donde se reflejan los procesos que tienen lugar en cada una de ellas. De la emergencia de los objetos que configuran el sistema de prácticas (lenguaje, situaciones-problemas, definiciones, proposiciones, procedimientos y argumentos) surgen los procesos de comunicación, problematización, definición, enunciación, elaboración de procedimientos/algoritmización y argumentación. Asimismo, los procesos cognitivos/epistémicos implícitos en cada una de las facetas duales que han sido explicitadas anteriormente son:

- Institucionalización – personalización (faceta institucional y personal);
- Generalización – particularización (faceta extensiva e intensiva);
- Análisis/descomposición – síntesis/reificación (faceta unitaria y sistémica);
- Materialización/concreción – idealización/abstracción (faceta ostensiva y no ostensiva);
- Expresión/representación – significación (faceta expresión y contenido).

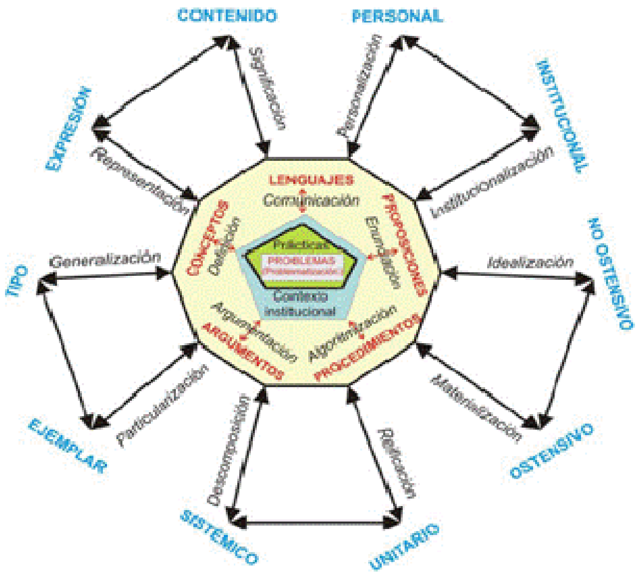


Figura 2.2.3. Configuración de objetos y procesos (Godino, Batanero y Font, 2007, p. 10)

2.2.3. LA INSTRUCCIÓN MATEMÁTICA

Los significados institucionales implementados en el aula son el resultado de diferentes interacciones entre el docente, el discente, el saber y el medio. El EOS aborda la enseñanza y aprendizaje de las matemáticas con nuevas herramientas teóricas, algunas de las cuáles tendremos en cuenta en los Estudios 2 y 3.

2.2.3.1. CONFIGURACIONES DIDÁCTICAS

La resolución de una tarea en el aula implica en el docente, en el alumno, en los recursos didácticos y las propias situaciones, un conjunto de estados interconectados que, en su conjunto, constituyen una unidad de análisis del proceso de instrucción. Una *configuración didáctica* lleva asociada una *configuración epistémica* determinada por el saber de referencia a cargo del profesor, de los estudiantes o distribuidas entre ambos. Asociada a ella existe una *configuración instruccional*, constituida por la red de objetos docentes, discentes y mediacionales que se presentan en la tarea; y una *configuración cognitiva*, constituida por los objetos intervinientes y emergentes de los sistemas de prácticas personales (Godino, Batanero y Font, 2007).

Los autores distinguen, atendiendo a la acción desempeñada por el alumno y el docente, cuatro tipos de configuraciones didácticas:

- *Configuración magistral*, donde el docente presenta el significado de los objetos matemáticos implicados en el tema, y a continuación, plantea a los alumnos ejercicios donde se apliquen. Corresponde al alumno dar sentido al discurso del profesor, ayudándose de los ejemplos y ejercicios que resuelve.

- *Configuración a-didáctica*, que se produce cuando el alumno asume la responsabilidad de realizar tareas que impliquen explorar, en la búsqueda de la solución, y formular conjeturas e hipótesis que puedan ser validadas por la propia situación-problema. Nosotros tendremos en cuenta este tipo de configuración globalmente en el Proyecto planteado a los estudiantes en el Estudio 2, y más específicamente en las actividades de ampliación propuestas en el mismo.

- En el momento de regulación del significado los objetos emergentes, puede surgir un espacio de *diálogo contextualizado* entre el docente y los alumnos (Godino, Contreras y Font, 2006). Esta configuración aparece en los debates organizados por el formador de profesores para la corrección de las tareas propuestas, una vez han sido completadas por los futuros profesores, en los Estudio 2 y 3.

- Si la tarea se realiza por el alumno sin que exista una interacción expresa del profesor, se habla de configuración didáctica personal; donde básicamente predomina el *estudio personal* (Godino, Contreras y Font, 2006). La lectura de documentos por parte de los futuros profesores correspondería a este tipo de configuración.

2.2.3.2. TRAYECTORIAS DIDÁCTICAS

En el EOS, se modeliza un proceso de instrucción mediante un proceso estocástico, ya que siempre existen elementos aleatorios respecto a lo planeado. Los estados de transición son conectados mediante diferentes trayectorias dependientes del *tiempo*, que se describen a continuación.

- *Trayectoria epistémica*. Sería la distribución del significado implementado en la enseñanza a lo largo del tiempo (Godino, Contreras y Font, 2006). De este modo, a lo largo del proceso de instrucción, el alumno se situará en distintos estados, dependiendo del tipo de objeto que trabaja: actuativo, situacional, lingüístico, conceptual, proposicional o argumentativo.

- *Trayectoria docente*. Secuenciación de las tareas docentes en el proceso de instrucción. Como aproximación a una categorización de las tareas o funciones docentes se destacan: planificación, motivación, asignación de tareas, regulación o evaluación.

- *Trayectoria discente*. Distribución de las acciones del alumno en el proceso de instrucción. Una categorización de las tareas o funciones discentes que desarrolla el alumno incluye: exploración, recuerdo, formulación de soluciones, argumentación o recepción de información, entre otras.

- *Trayectoria mediacional.* Distribución de los recursos o medios utilizados en el proceso de instrucción.
- *Trayectoria cognitiva.* Distribución temporal del proceso de formación del significado personal de un objeto matemático.
- *Trayectoria emocional.* Distribución temporal de los estados emocionales de cada alumno en relación a los objetos matemáticos en el proceso de instrucción.

En nuestro trabajo no se analizan separadamente cada una de estas trayectorias, pero sí se describe globalmente la trayectoria de estudio planteada en las actividades propuestas a los futuros profesores en el Estudio 2.

2.2.4. DIMENSIÓN NORMATIVA

Como cualquier actividad social, la clase de matemáticas se encuentra regulada, de modo explícito e implícito, mediante normas (Godino, Contreras y Font, 2006). Estas normas pueden ser clasificadas atendiendo a diferentes criterios, algunos de los cuáles se representan en la Figura 2.2.4.

Si se clasifican atendiendo a los *grados de coerción o rigidez* se distinguirán convenciones de cumplimiento obligatorio, o no. Teniendo en cuenta su *origen* las normas pueden venir de la administración educativa, la sociedad, la escuela, el aula o la propia disciplina. Si se considera el *momento* o fase del proceso de enseñanza y aprendizaje en que se manifiestan, se pueden distinguir normas que se desarrollan en el diseño curricular, en la planificación de la enseñanza, en su implementación, o bien en la evaluación de la enseñanza y el aprendizaje.

Por último, las normas que regulan los procesos de estudio matemáticos se pueden categorizar según la faceta del proceso de estudio a que se refieren:

- *Normas epistémicas*, que rigen las matemáticas susceptibles de ser enseñadas y aprendidas en una institución.
- *Normas cognitivas,* regulan el comportamiento matemático del alumno y pueden concordar o no con las normas epistémicas (Godino, Font, Wilhelmi, Castro, 2009).
- *Normas interaccionales*, que rigen las interacciones docente-discente y discente-discente.
- *Normas mediacionales*, que rigen el uso de los recursos humanos, materiales y temporales. Son ejemplos la duración de la tarea, el uso de determinados recursos y espacios, etc.
- *Normas afectivas,* que se refieren a la motivación, a la actitud positiva de los estudiantes, la autoestima, y a la “obligación” de transmitirlas a los alumnos.
- *Normas ecológicas*, que rigen la relación con el entorno institucional (social, político y económico) en que se desarrolla el proceso de instrucción.

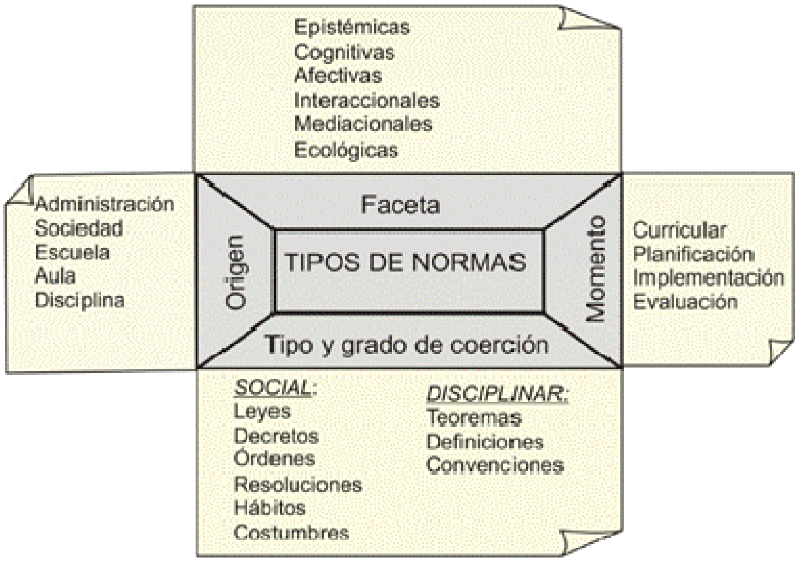


Figura 2.2.4.Tipos de normas (Godino, Batanero y Font, 2007, p. 14)

2.2.5. IDONEIDAD DIDÁCTICA

La caracterización de posibles criterios que optimicen el aprendizaje matemático se aborda en el EOS con el concepto de idoneidad didáctica (Wilhelmi, Font y Godino, 2005; Godino, Contreras y Font, 2006):

> Se trata de una guía de orientación para la mejora de los procesos de instrucción, no de un criterio que produzca la frustración del profesor «normal» al no poderlo alcanzar. Los criterios de idoneidad son reglas de corrección útiles en dos momentos de los procesos de estudio matemáticos. A priori, los criterios de idoneidad son principios que orientan «cómo se deben hacer las cosas». A posteriori, los criterios sirven para valorar el proceso de estudio efectivamente implementado (Godino, Font, Wilhelmi y Castro, 2009, p.61).

Los autores distinguen los siguientes criterios de idoneidad:

- *Idoneidad epistémica*. Representatividad del significado institucional implementado en relación al significado de referencia. Supone la elaboración de una transposición didáctica (Chevallard, 1991) capaz de acomodar por un lado el significado implementado al pretendido, y por otro, el significado pretendido al de referencia.
- *Idoneidad cognitiva*. Referida a la relación existente entre el significado personal del alumno previo a la instrucción y el implementado. Es decir, que exista el máximo ajuste entre estos significados, teniendo en cuenta además las restricciones cognitivas de los alumnos.
- *Idoneidad emocional*. Predisposición e implicación del alumno ante la actividad matemática a desarrollar. Una ayuda para potenciar la percepción de la utilidad de la matemática es enfatizar las conexiones de las matemáticas con la realidad, y presentar los conocimientos dentro de contextos en que se les dé sentido.
- *Idoneidad mediacional*. Disponibilidad y adecuación de los recursos materiales y

temporales necesarios para el proceso de instrucción. A modo de ejemplo, las situaciones de juego o de aprendizaje por proyectos a veces quedan sin realizar por falta de tiempo.

- *Idoneidad interaccional*. Referida por un lado a la interacción del alumno, a favor de su autonomía en el aprendizaje, y por otro, al proceso de enseñanza-aprendizaje, ya que debe permitir identificar y resolver los conflictos semióticos potenciales (Godino Batanero y Font, 2009).
- *Idoneidad ecológica*. Adaptación del proceso de estudio al proyecto educativo de centro, a las directrices curriculares, a las condiciones del entorno social y cultural, etc.

Todas ellas conforman la noción de *idoneidad didáctica*, que pretende abordar de modo integral la complejidad de factores que intervienen en el diseño, desarrollo y evaluación de cualquier proceso de estudio matemático (Godino, Batanero y Font, 2007). Nosotros utilizaremos el concepto de idoneidad didáctica, sus componentes y descriptores, en el Estudio 3, en el que se pedirá a los futuros profesores valorar la idoneidad didáctica de un proceso de instrucción vivido por ellos mismos. Esta actividad permitirá valorar y desarrollar las diferentes facetas de su conocimiento didáctico sobre el tema de la correlación y regresión.

2.2.6. EL ENFOQUE ONTOSEMIÓTICO EN NUESTRA INVESTIGACIÓN

En las secciones anteriores se han descrito las diferentes herramientas teóricas del marco teórico, EOS, algunas de las cuáles utilizaremos en nuestro trabajo. En primer lugar, consideramos las ideas de significado personal e institucional, con sus diferentes tipos, así como las configuraciones de objetos y procesos matemáticos, y sus diferentes facetas.

En el análisis de libros de texto presentado en el Capítulo 4, se categorizarán y analizarán con detalle los objetos matemáticos elementales considerados en el marco teórico, y se comparará su presentación e interrelación en los diferentes textos de la muestra. Tendremos también en cuenta la idea de función semiótica y conflicto semiótico, pues uno de nuestros objetivos será identificar la existencia de posibles conflictos semióticos en los textos.

En los Estudios 2 y 3, al analizar los componentes del conocimiento didáctico y matemático de los futuros profesores respecto a la correlación y regresión, se tendrán también en cuenta los siguientes elementos.

- Configuraciones epistémicas de objetos y procesos matemáticos implícitos en la resolución de las tareas matemáticas propuestas (Capítulo 5).
- Descripción de las trayectorias y configuraciones didácticas en la experiencia realizada con los futuros profesores (Capítulos 5 y 6).
- Valoración de la idoneidad didáctica del proceso formativo realizado con los futuros profesores (Capítulos 5 y 6).

2.3. CONOCIMIENTO DEL PROFESOR

2.3.1. INTRODUCCIÓN

La investigación sobre la formación del profesor de matemáticas es muy amplia. Podemos ver esta extensión en los capítulos que sobre el tema, se incluyen en diferentes "Handbooks" de investigación en educación matemática (Brown y Borko, 1992; Thompson, 1992; Llinares y Krainer, 2006; Ponte y Chapman, 2006; Hill, Sleep, Lewis y Ball, 2007; Wood, 2008; Kieran, Krainer y Shaughnessy, 2013; y White, Jaworski, Agudelo-Valderrama y Gooya, 2013), en revistas como *Journal of Mathematics Teacher Education*, el *ICMI Study 15* sobre "La formación y desarrollo profesional de los profesores de matemáticas" (Even y Ball, 2009), o la serie *Mathematics Teacher Education* de Springer.

En España, destaca la investigación sobre el tema realizada desde los grupos de investigación coordinados por Salvador Llinares, Victoria Sánchez, José Carrillo y Lorenzo Blanco, entre otros, potenciada desde el Grupo de Investigación, "Conocimiento y desarrollo profesional del profesor de matemáticas" de la Sociedad Española de Investigación en Educación Matemática (SEIEM).

Todo este esfuerzo por analizar los componentes de la formación del profesor, y las actividades para llevarlas a cabo, ha tenido un menor peso en educación estadística. En España encontramos algunas tesis doctorales centradas en la formación de profesores para enseñar probabilidad o estadística, como las de Azcárate (1995), Cardeñoso (1998), Pinto (2010), Arteaga (2011), Contreras (2011), Mohamed (2012), o Gómez (2014). La necesidad de impulsar este tipo de investigación fue reconocida por *ICMI* e *IASE*, que promovieron un "Joint ICMI/IASE Study" específicamente orientado a promover la investigación a nivel internacional sobre la educación y desarrollo profesional de profesor para enseñar estadística. Tras la conferencia del estudio, (Batanero, Burrill, Reading y Rossman, 2008), sus resultados se publicaron en Batanero, Burrill y Reading (2011). Dicho estudio ha promovido la investigación relacionada, como se refleja en las sesiones específicas de trabajos invitados sobre este tema en los congresos *ICOTS* celebrados en 2010 y 2014.

Debido a la amplitud de la investigación sobre formación de profesores de matemáticas, sólo nos centraremos en la específica sobre formación de profesores para enseñar estadística. A continuación, comenzamos analizando algunos modelos del conocimiento del profesor de matemáticas y de estadística que hemos tenido en cuenta para nuestro estudio. Seguidamente se resumen los resultados de las investigaciones sobre actitudes y creencias, conocimientos estadísticos y conocimientos didácticos de los profesores o futuros profesores de estadística. Finalmente, se presenta el modelo de conocimiento del profesor que utilizaremos en nuestro trabajo.

2.3.2. MODELOS DE CONOCIMIENTO DEL PROFESOR DE MATEMÁTICAS

En los trabajos sobre conceptualización del conocimiento necesario para enseñar matemáticas, aprender a enseñar se considera un proceso continuo, en el que profesor construye su conocimiento tomando como referencia su conocimiento previo y el contexto en el que desarrolla su labor (Llinares, 1998). En lo que sigue, describimos el modelo propuesto por Shulman, que ha sido muy influyente en la investigación posterior, y otros modelos que consideramos de interés para nuestro trabajo.

2.3.2.1. MODELO DE SHULMAN

El interés por el conocimiento didáctico específico para enseñar una determinada materia es iniciado por Shulman (1986; 1987). Este autor, además del conocimiento de la materia impartida (a), reconoce la necesidad de que los profesores aprendan otros contenidos, tales como: (b) conocimientos pedagógicos generales, como los relacionados con la gestión y organización de la clase; (c) conocimiento del currículo, materiales de enseñanza y directrices curriculares; (d) conocimiento didáctico del contenido (o conocimiento pedagógico del contenido), que sería el conocimiento didáctico específico de la materia, "*conocimiento que va más allá del conocimiento de la materia en sí, hacia un conocimiento de la materia para enseñar*" (Shulman, 1986, p.9); (e) conocimiento de los estudiantes y de sus características; (f) conocimiento del contexto educativo; y (g) conocimiento de los objetivos, valores y fines educativos junto a sus fundamentos filosóficos e históricos.

Especial influencia en la investigación posterior ha tenido el conocimiento didáctico del contenido (CDC) que incluye, según Shulman (1986), las estrategias educativas válidas para una materia específica, es decir, formas productivas de representación de las ideas a enseñar, analogías, ilustraciones, ejemplos, explicaciones y demostraciones de la materia para hacerla comprensible a los estudiantes. Conceptualiza la manera en que el contenido específico puede ser interpretado en una situación de enseñanza, y se basa en la experiencia e ideas del profesor acerca de las matemáticas, su enseñanza y aprendizaje (Cooney, 1994). Llinares lo describe en la forma siguiente:

> Conocimiento del profesor del contenido que tiene que enseñar, considerándolo desde la perspectiva de la enseñanza-aprendizaje. Se centra en el conocimiento del profesor de las potencialidades y limitaciones de los diferentes modos de representación del contenido matemático como medios para hacer comprensible a los estudiantes dicho contenido y el conocimiento de las dificultades y errores más comunes de los alumnos en relación a un contenido específico (Llinares, 1998, p. 161).

Pinto (2010) resume las características que diferentes autores han asignado al CDC, y las clasifica en cuatro aspectos: (1) la contextualización del CDC tiene en cuenta tanto la naturaleza del contenido que se enseña como la instrucción; (2) consiste en la transformación, transferencia o transposición didáctica del contenido para enseñarlo en el ámbito escolar, (3) es diferente al conocimiento de la materia y no es una simple mezcla de pedagogía y contenido, por lo cual, requiere de características especiales para su formación y desarrollo en los profesores; y (4) para la formación de los profesores, se requiere la reflexión y aplicación sobre la acción, la integración de psicología y contenido, la investigación en didáctica de la disciplina y el estudio de los diferentes modos de representar el contenido a enseñar.

Por otro lado, si queremos analizar el CDC del profesor, se requiere una descripción más detallada de los componentes de dicho conocimiento. Varios autores han propuesto modelos de estos componentes, que pueden ser útiles en nuestro trabajo. Ya Shulman (1986) propuso tres componentes esenciales del CDC, que están ligados, ya que la mayoría de las representaciones y estrategias de enseñanza tienen relación tanto con la forma de comprender el proceso de aprendizaje de los estudiantes, como con el conocimiento de la materia:

1. *El conocimiento del contenido o conocimiento de la materia* (que va a ser enseñada en un contexto escolar). Una condición necesaria es que el profesor tenga conocimientos suficientes de aquello que va a enseñar. La investigación sobre el conocimiento del

contenido matemático del profesor se orienta a analizar su naturaleza conceptual y epistemológica, sus componentes, características, y el grado de conocimiento matemático (genérico o específico) que tienen los profesores, así como sus relaciones con la enseñanza y el aprendizaje (Even, 1990, 1993).

2. *Las formas de representación o estrategias para enseñar el tema*. Este componente combina el conocimiento de las representaciones de la materia en cuestión y las estrategias educativas, tales como uso de analogías, ilustraciones, ejemplos, explicaciones y demostraciones. Todos ellos son elementos utilizados por el profesor para ayudar en la generación del conocimiento en los alumnos (Llinares, Sánchez y García, 1994) y un profesor debiera usar un amplio repertorio de representaciones y estrategias en la enseñanza, según Shulman (1986).

3. *Conocimiento del aprendizaje del estudiante,* incluyendo los errores, dificultades, creencias y concepciones de sus estudiantes y las condiciones necesarias para lograr transformar estas concepciones de manera adecuada y correcta.

2.3.2.2. CONOCIMIENTO MATEMÁTICO PARA LA ENSEÑANZA

Posteriormente a Shulman, muchos investigadores han tratado de refinar las categorías propuestas por este autor o describir otras nuevas. No pretendemos ser exhaustivos, debido a la amplitud de la bibliografía, tan sólo haremos mención a algunas propuestas útiles en nuestro trabajo.

Por ejemplo, Even (1990) sugiere que el conocimiento de la materia por parte del profesor debe integrar cuatro aspectos: (a) comprender el rol e importancia del tópico en matemáticas y en el currículo, (b) conocer las aportaciones de la investigación educativa y los marcos teóricos sobre el aprendizaje para el tema dado, (c) comprender los conceptos matemático ligados al tópico específico, y (d) conocer la investigación y marcos teóricos sobre el conocimiento de la materia del profesor y su rol en la enseñanza.

Por otro lado, Ball y sus colaboradores proponen el marco del *Conocimiento Matemático para la Enseñanza* (MKT, Mathematical Knowledge for Teaching), que dividen en los siguientes componentes: *Conocimiento Común del Contenido (CCK), Conocimiento Especializado del Contenido (SCK), Conocimiento del Contenido y la Enseñanza (KCT), Conocimiento del Contenido y los Estudiantes (KCS)* (Ball, Thames y Phelps, 2005). Hill, Ball, y Schilling (2008) proponen tener en cuenta el *Conocimiento en el Horizonte Matemático* y *Conocimiento del currículum*.

El *Conocimiento Común del Contenido* se refiere al conocimiento puesto en juego por una persona instruida para resolver problemas matemáticos, para lo cual, un sujeto con suficiente conocimiento está capacitado. El *Conocimiento Especializado del Contenido* (SCK), indica un conocimiento especial del profesor que lo habilita para planificar y desarrollar secuencias de enseñanza del contenido. Con el "*Conocimiento en el Horizonte Matemático*" se refieren a aspectos más avanzados, que aportan perspectivas al profesor, por ejemplo, en la detección de posibles confusiones respecto a las ideas matemáticas que subyacen en el tema, aspectos históricos, etc.

Respecto al conocimiento didáctico del contenido, el "*Conocimiento del Contenido y los Estudiantes* (*KCS*) se refiere al conocimiento de cómo los estudiantes piensan, saben, o aprenden un objeto matemático. Incluye el conocimiento de los errores y dificultades comunes, las concepciones erróneas, las estrategias utilizadas, el ser capaz de valorar la comprensión del alumno y saber cómo evoluciona su razonamiento matemático. El *Conocimiento del Contenido y la Enseñanza* (*KCT*) resulta de la integración del contenido matemático con el conocimiento de la enseñanza de dicho contenido. Incluye saber construir, a partir del razonamiento de los estudiantes y las estrategias utilizadas por ellos, procesos pertinentes para tratar y corregir sus errores y concepciones erróneas. El *Conocimiento del Currículo* se refiere a las directrices curriculares, orientaciones, fines y motivaciones de las mismas, materiales curriculares y secuenciación del tema en los diferentes ciclos formativos.

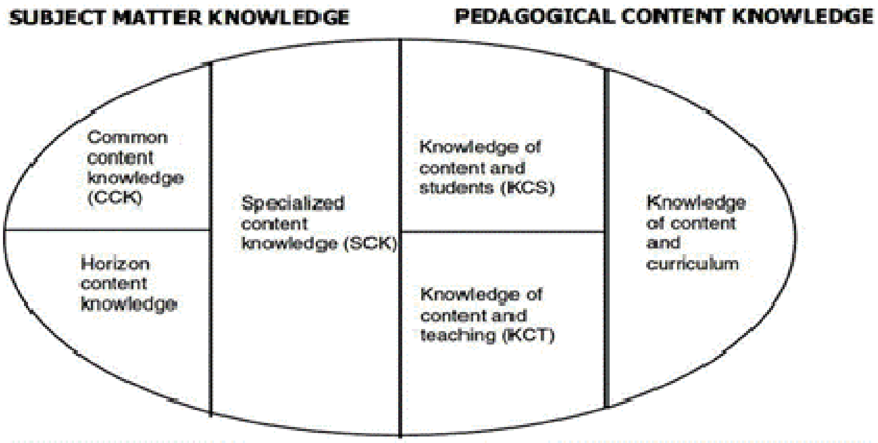


Figura 2.3.1. Componentes del MKT (Ball, Thames y Phelps, 2008, p. 403)

Este mismo modelo se puede tener en cuenta para el caso de la estadística, considerando las características específicas de la materia. En la Figura 2.3.1 se muestra un esquema de los diferentes tipos de *Conocimiento matemático para la enseñanza* en el modelo de Hill, Ball y Schilling, algunos de los cuales serán evaluados en nuestro trabajo (Capítulos 5 y 6).

2.3.3. MODELOS DEL CONOCIMIENTO DEL PROFESOR DE ESTADÍSTICA

2.3.3.1. PRIMERAS REFERENCIAS

La literatura sobre formación de profesores para enseñar estadística es mucho más escasa que la relacionada con la formación de profesores para enseñar matemáticas. Una de las primeras referencias al estudio del conocimiento del profesor para explicar estadística se hace en Godino, Batanero y Flores (1999), quienes describen las siguientes competencias básicas que deben formar parte de dicho conocimiento:

- *La reflexión epistemológica* sobre el significado de los objetos matemáticos que se pretenden enseñar, la naturaleza del conocimiento estocástico, su desarrollo y evolución.
- *Análisis de las transformaciones del conocimiento* para adaptarlo a los distintos niveles de enseñanza. Se trata de reflexionar sobre los diversos niveles de comprensión posibles respecto a un objeto matemático, y valorar el nivel y forma particular en que

un determinado concepto podría ser enseñado a una persona en particular.

- *Estudio de las dificultades, errores y obstáculos* de los alumnos en el aprendizaje junto a sus estrategias en la resolución de problemas, que permitirá orientar la tarea de enseñanza a sus necesidades, y adecuar la evaluación a su aprendizaje.
- *Análisis del currículo, situaciones didácticas, metodología de enseñanza para temas específicos, y recursos didácticos específicos*, que junto al contexto educativo condicionan el proceso de enseñanza y aprendizaje.

Batanero, Garfield, Ottaviani y Truran (2000) proponen la evaluación del conocimiento del profesor de estadística como tema prioritario de investigación. Al delimitar los dominios de conocimiento que debe tener el profesor de estadística, los autores se refieren al "conocimiento pedagógico en estadística", el cual está integrado por conceptos de pedagogía, psicología, conocimiento de las concepciones erróneas, e intuiciones, así como conocimiento de la epistemología, currículo y materiales en estadística.

Otra mención temprana al CDC en estadística la realiza Watson (2001), quien basándose en los estudios de Shulman, diseñó un instrumento para medir los perfiles de competencia de los profesores en estadística. El cuestionario contiene ítems que permiten evaluar los tipos de conocimiento propuestos por Shulman, incluyendo el conocimiento de factores significativos en la enseñanza, planificación de la enseñanza, prácticas de enseñanza y evaluación de las dificultades de los estudiantes, creencias acerca de la estadística en la vida diaria, evaluación de respuestas de estudiantes, y cómo usarlas en la clase.

Batanero (2002) sugirió que la formación de profesores debe considerar tanto el conocimiento estadístico, como el conocimiento didáctico del contenido, y describió los mismos componentes básicos descritos en Godino, Batanero y Flores (1999). Estos componentes son ampliados en Batanero, Godino y Roa (2004), quienes incluyen las siguientes facetas en la formación de profesores para enseñar estadística, sugiriendo también la necesidad de utilizar situaciones de tipo constructivista en la formación de los profesores:

- Reflexión epistemológica sobre el significado de los objetos matemáticos que se enseñan (por ejemplo, sobre los significados diferenciados de la probabilidad). Esta reflexión incluiría conocimiento de tipo histórico, filosófico y cultural, así como las relaciones con otros dominios de las ciencias.
- Experiencias para adaptar el conocimiento estadístico según diferentes niveles de enseñanza y la capacidad de los estudiantes, organizando e implementando proyectos estadísticos, simulaciones y gráficos, no sólo como ayudas metodológicas sino como formas útiles de aprender y comprender la estadística.
- Capacidad crítica para analizar libros de texto y materiales curriculares.
- Predicción de las dificultades, errores, estrategias y obstáculos de los estudiantes al resolver problemas.
- Capacidad para desarrollar y analizar ítems de evaluación e interpretar las respuestas de los estudiantes a los mismos.

- Experiencia con buenos ejemplos de situaciones didácticas, materiales y recursos.

Sorto (2004) usa la noción de CDC de Shulman como un modelo teórico para analizar el conocimiento del profesor, y valora diferentes documentos que en Estados Unidos se utilizan para la formación de profesores (estándares nacionales y estatales para la acreditación de profesores). Para la autora, el CDC sintetiza los tres tipos de conocimiento que todo profesor deber tener: de la matemática, de los estudiantes y de las prácticas de instrucción.

Garfield y Ben-Zvi (2008) proponen otro modelo de conocimiento profesional para enseñar estadística, que incluye cinco competencias que se han de desarrollar en la formación de profesores: (a) ideas estadísticas fundamentales; (b) uso de datos reales; (c) uso de actividades para el aula; (d) integración de las herramientas tecnológicas; (e) implementación del discurso en el aula y (f) uso de métodos alternativos de evaluación. Estos componentes, sin embargo, no se relacionan con los descritos en la literatura de formación de profesores, ya que más bien, serían aspectos a tener en cuenta en la formación estadística tanto de alumnos como de profesores.

2.3.3.2. MODELO DE BURGESS

Burgess (2008; 2011) compara varios marcos teóricos de las investigaciones en educación matemática y estadística, y propone un modelo que puede ser usado para estudiar el conocimiento del profesor y el conocimiento estadístico para la enseñanza. Para ello cruza los cuatro componentes definidos por Ball, Thames y Phelps (2005) con las categorías que definen los modos esenciales de razonamiento estadístico (Wild y Pfannkuch, 1999), y construye un modelo bidimensional (ver Figura 2.3.2), que es útil para proporcionar distintos perfiles de profesores según su conocimiento estadístico para la enseñanza: reconocer la necesidad de datos; transnumeración (cambio de representaciones para extraer nueva información de los datos); consideración de la variación (reconocer la variación en los datos, describir patrones acerca de la variación y tratar de comprender los datos en relación al contexto); razonamiento con modelos estadísticos (desde los simples, como tablas y gráficos, a los complejos); e integración de la estadística y el contexto.

			Statistical knowledge for teaching			
			Content knowledge		Pedagogical content knowledge	
			Common knowledge of content (ckc)	Specialised knowledge of content (skc)	Knowledge of content and students (kcs)	Knowledge of content and teaching (kct)
Statistical thinking in empirical enquiry	Types of thinking	Need for data				
		Transnumeration				
		Variation				
		Reasoning with models				
		Integration of statistical and contextual				
	Investigative cycle					
	Interrogative cycle					
	Dispositions					

Figura 2.3.2. Modelo de conocimiento profesional de Burgess (2008, p.3)

Además, añade otros elementos del modelo de razonamiento estadístico de Wild y Pfannkuch (1999): ciclo de investigación (problema, plan, datos, análisis y conclusión), el ciclo interrogativo (generar, buscar, interpretar, criticar y juzgar) y las disposiciones, tales como escepticismo, perseverancia, compromiso e imaginación, entre otras.

2.3.3.3. MODELO DE GROTH

Groth (2007) propone un modelo de conocimiento estadístico para la enseñanza, utilizando la diferencia entre conocimiento común y especializado de Ball y sus colaboradores, En este último incluye los componentes didácticos. Su modelo está basado en las recomendaciones de GAISE (Franklin et al., 2007) con los componentes de formular preguntas, recoger datos, analizar datos y formular conclusiones. Propone ejemplos de conocimientos matemáticos y no matemáticos requeridos en las categorías implicadas, que presentamos traducidos en la Tabla 2.3.1.

Tabla 2.3.1. Aspectos del conocimiento estadístico para la enseñanza (Groth, 2007, p. 430)

Marco de GAISE	Tipo de conocimiento	Ejemplo de tareas que requieren conocimiento matemático	Ejemplo de tareas que requieren conocimiento no matemático
Formular preguntas	Común	Leer correctamente un gráfico para plantear preguntas sobre los datos	Comprender la diferencia entre situación aleatoria y determinista
	Especializado	Comprender las diferencias entre la forma que los estudiantes interpretan el diagrama de puntos y de caja	Valorar la potencialidad de las preguntas planteadas por los estudiantes
Recoger datos	Común	Medir cantidades o construir una simulación correctamente	Construir cuestionarios y diseñar experimentos
	Especializado	Comprender las estrategias y dificultades de los estudiantes al construir un algoritmo de simulación	Anticipar las dificultades de los estudiantes para distinguir el propósito del muestreo aleatorio
Analizar datos	Común	Calcular estadísticos como media, mediana o moda	Utilizar metáforas como “señal y ruido” en estadística
	Especializado	Identificar las propiedades de la media que son difíciles para el estudiante	Darse cuenta que el alumno puede calcular la media y no tener en cuenta el contexto de los datos
Interpretar resultados	Común	Interpretación correcta del *p* valor	Juzgar la conveniencia del nivel de significación elegido por el alumno
	Especializado	Comprender la interpretación que hace un alumno del *p* valor	Anticipar la generalización abusiva del término “significativo” por el alumno

Así observamos, en dicha tabla, que la mayoría de ejemplos que propone para el conocimiento especializado corresponde a lo que Ball y colaboradores denominan conocimiento del contenido y los estudiantes. No obstante, la idea de proponer ejemplos de conocimientos específicos del profesor para enseñar estadística nos parece útil, y se puede concretar más proponiendo ejemplos de conocimientos específicos requeridos para la enseñanza de la correlación y regresión, lo que haremos en el Capítulo 6.

2.3.3.4. MODELOS QUE CONSIDERAN LA TECNOLOGÍA

Debido al interés actual de la tecnología en la enseñanza de la estadística, algunos autores tienen en cuenta los conocimientos específicos que necesita el profesor para enseñar estadística con apoyo de la tecnología. Así, Lee y Hollebrands (2011) analizan el modelo propuesto por Niess (2005), que describe los conocimientos necesitados en la enseñanza de la matemática y las ciencias con apoyo de la tecnología. Niess parte del modelo de PCK de Shulman, que incluye el conocimiento del contenido (Content Knowledge) y de la Pedagogía (Pedagogical Knowledge) y añade un tercer componente (Technological Knowledge), teniendo en cuenta todas las posibles intersecciones de estos tres componentes (Ver Figura 2.3.3).

En el corazón del modelo estaría el *Technological Pedagogical Content Knowledge*, es decir, el conocimiento específico didáctico para enseñar un contenido particular con ayuda de la tecnología. Incluiría el conocimiento de las estrategias de enseñanza y las representaciones para enseñar temas particulares con apoyo de la tecnología; sobre la comprensión, razonamiento y aprendizaje de los estudiantes con la tecnología o del currículo y materiales curriculares que integran la tecnología en el aprendizaje. Para el caso de la estadística, un ejemplo sería conocer el modo de usar Applets para la enseñanza de la correlación y regresión.

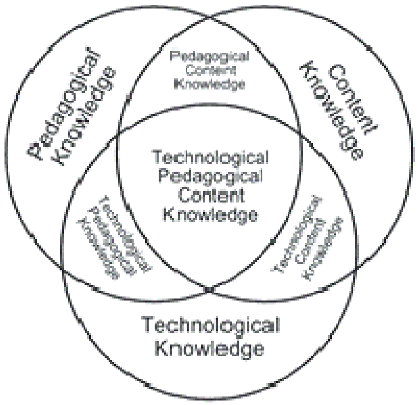


Figura 2.3.3. Modelo de conocimiento de Niesss (Lee y Hollebrands, 2011 p.361)

Lee y Hollebrands (2011) adaptan el modelo anterior, teniendo en cuenta la especificidad del conocimiento estadístico. Las autoras indican cómo la tecnología puede usarse como amplificador de los procesos de razonamiento estadístico, y como reorganizador cognitivo para promover la comprensión conceptual. Simplifican también el esquema presentado por Niess, y proponen el que mostramos en la Figura 2.3.4, formado por tres tipos de conocimiento, cada uno de los cuáles es más amplio que el anterior. Así, la capa más externa (conocimiento estadístico; SK) tiene en cuenta los componentes del modelo de pensamiento estadístico de Wild y Pfannkuch (1999), que hemos descrito anteriormente. Las autoras lo presentan en el círculo exterior, porque es un pre-requisito para los otros tipos de conocimiento. No se puede tener, por ejemplo, un conocimiento pedagógico de cómo enseñar estadística, si no se tiene un conocimiento suficiente de la estadística.

A continuación, incluyen el conocimiento tecnológico sobre la estadística (TSK), que implica el conocimiento de los recursos tecnológicos y de su uso; por ejemplo, el trabajo con instrumentos de visualización o de análisis exploratorio de datos. Finalmente, el conocimiento tecnológico-pedagógico sobre la estadística (TPSK), es el corazón del modelo, e implica comprender cómo los alumnos razonan y trabajan con la tecnología, qué dificultades tienen con la misma; como plantear tareas adecuadas con la tecnología, etc.

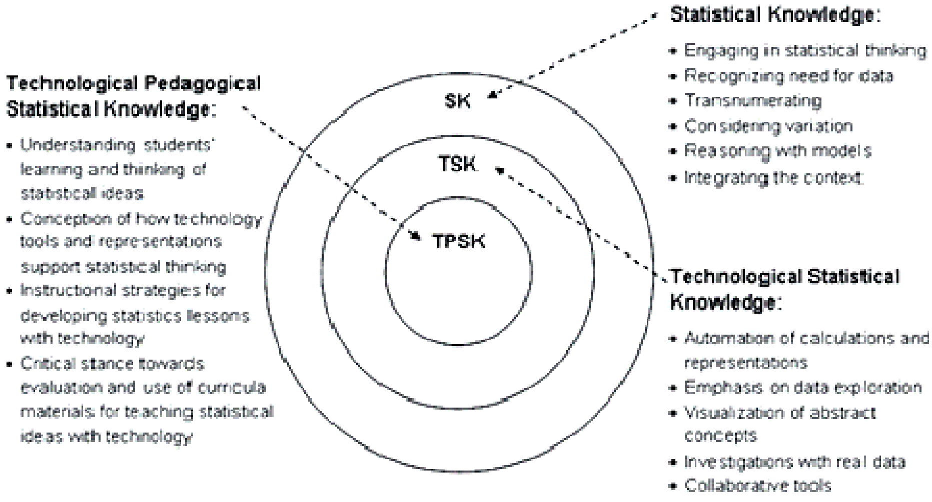


Figura 2.3.4. Modelo de conocimiento de Niesss (Lee y Hollebrands, 2011 p.362)

Otra adaptación del modelo de Niess es propuesta por Wasson y Biehler (2010) (Figura 2.3.5). En este caso loa autores tienen en cuanta las tres categorías de conocimiento del contenido de Ball y colaboradores: común (CCK), especializado (SCK) y ampliado (HK), que sitúan en el anillo exterior, junto con el conocimiento pedagógico (PK) y el tecnológico (TK). En el anillo intermedio, estos tres dominios de conocimientos se combinan en parejas generando las categorías de conocimiento pedagógico-tecnológico (TPK), tecnológico del contenido (TCK) y conocimiento didáctico (pedagógico, PCK) Finalmente, en el centro del esquema aparece el conocimiento pedagógico-tecnológico del contenido (TPCK).

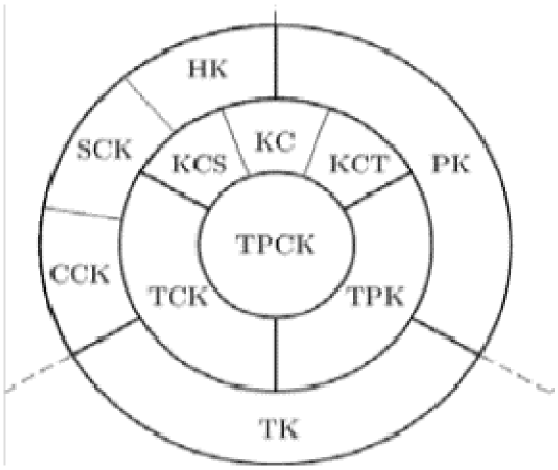


Figura 2.3.5. Modelo de conocimiento del profesor (Wasson y Biehler, 2010 p.2)

Aunque nosotros no evaluaremos específicamente el conocimiento matemático o pedagógico en relación a la tecnología, nos ha parecido importante presentar resumidamente estos modelos, ya que en las actividades planteadas a los profesores en el Capítulo 5 se trabaja con la tecnología, y por tanto, ocasionalmente se puede hacer referencias al conocimiento tecnológico sobre enseñanza de la estadística, aunque no sea nuestro objetivo principal su evaluación o desarrollo.

2.3.4. MODELOS PARA LA ENSEÑANZA DE LA CORRELACIÓN Y REGRESIÓN

Puesto que nuestro trabajo se centra en estos objetos matemáticos, resaltamos en este apartado un modelo específico para la enseñanza de la correlación y regresión.

Casey (2008; 2010) propone un marco de referencia sobre el conocimiento que el profesor de secundaria necesita para la enseñanza de la correlación, utilizando la conceptualización del conocimiento matemático para la enseñanza de Ball y colaboradores anteriormente descrito. Para formularlo, analiza previamente las concepciones de los profesores de secundaria sobre la noción de asociación estadística y sus ideas acerca de la relevancia en el currículum y su tratamiento en el aula.

Tabla 2.3.2. Categorías del conocimiento de la asociación (Casey, 2008, p. 89)

Categoría	Nº de cuestiones de enseñanza observadas
Mejor ajuste lineal (incluyendo el ajuste lineal por mínimos cuadrados)	116
Coeficiente de correlación	100
Contraste y estadístico Chi-cuadrado	96
Modelización matemática (excluyendo el mejor ajuste lineal)	83
Contraste de independencia Chi-cuadrado	64
Gráfico de residuos	61
Calculadora	60
Residuo	54
Predicción	52
Coeficiente de determinación	46
Álgebra (excluyendo las rectas y la pendiente)	43
Contraste T para la pendiente de la Regresión lineal	38
Asociación	37
Diagrama de dispersión	36
Contexto	35
Test de hipótesis (en general)	35
Pendiente	34
Recta	27
Terminología	26
Razonamiento proporcional	25
Datos	21
Ordenador	19
Probabilidad	18
Tabla de doble entrada	15
Intervalos de confianza para la pendiente y la regresión lineal	10
Puntos Outliers y puntos influyentes	8
Gráficos estadísticos (excluyendo el diagrama de dispersión, la tabla de doble entrada y el gráfico de barras)	7
Promedios (incluida la media)	6

Indica que los profesores tienen un conocimiento moderado de cómo resolver problemas de asociación estadística; utilizan sus teorías previas en la resolución de los problemas; para detectar la asociación, algunos comparten la concepción causal; no entienden la necesidad de mostrar a los estudiantes problemas donde exista independencia; y además, la mayoría de los profesores de su muestra se sentían incómodos resolviendo problemas de asociación, o no tenían confianza en sus respuestas. Esta evaluación no trata de documentar el conocimiento de los profesores sobre la noción de asociación estadística, sino describir el conocimiento que se emplea en la planificación, la implementación y la gestión del proceso de instrucción.

Como resultado de su investigación presenta una categorización del conocimiento necesario para la enseñanza de la asociación estadística, junto con el número de veces que observa el empleo de cada categoría, por ejemplo, en cuestiones/reflexiones del estudiante, explicaciones proporcionadas por el profesor, acciones determinadas de los estudiantes en torno a problemas planteados. Considera el estudio inferencial, no sólo el descriptivo, por lo cual, algunas de sus categorías no han aparecido en nuestro significado de referencia.

En total se distinguen veintinueve categorías (Tabla 2.3.2), de las cuales, la mejor línea de ajuste ocupa un lugar destacado, seguida del coeficiente de correlación. Entre las nociones referidas a datos categóricos se destacan las de estadístico Chi-cuadrado y Test Chi-cuadrado en general y Test Chi-cuadrado para el estudio de la independencia. Tres de estas categorías están presentes tanto en datos categóricos como en datos cuantitativos y sirven de base para el desarrollo de dos resultados principales en este estudio: Por una parte se utilizan como eje para la elaboración de un mapa conceptual (Figura 2.3.6), y por otra parte sirven como principios para concretar los requisitos esenciales para la enseñanza de la noción de asociación estadística, que se resumen a continuación:

- Conocer el significado de la asociación estadística como correspondencia de variables estadísticas, diferenciando asociación y causación.
- Fluidez y precisión en el lenguaje estadístico es crucial para enseñar asociación estadística.
- Conocimiento del contexto es otra componente del conocimiento esencial para enseñar asociación, coincide con Franklin et al. (2007).
- Conocer los datos con los que están trabajando, es decir, los diferentes tipos de datos que podemos encontrar en el trabajo estadístico, y su tratamiento: representar, analizar e interpretar gráficos (diagrama de dispersión, histograma, diagrama de caja, entre otros) así como conceptualizar adecuadamente resúmenes estadísticos, como la media.
- Conocer características de diferentes modelos de ajuste (exponenciales, logarítmicos, etc.) desde la tabla o el diagrama de dispersión, especialmente el modelo lineal.
- Distinguir correlación y coeficiente de determinación.
- Dominio de recursos tecnológicos.

El análisis de Casey es más limitado que el realizado al presentar nuestro significado de referencia (excluyendo la parte inferencial, que no tenemos en cuenta). Por un lado, no tiene en cuenta elementos como el lenguaje, propiedades o argumentos. Además, incluso en la parte conceptual, nuestro análisis es mucho más completo y se ampliará aún más cuando se realice el análisis de libros de texto en el Capítulo 4.

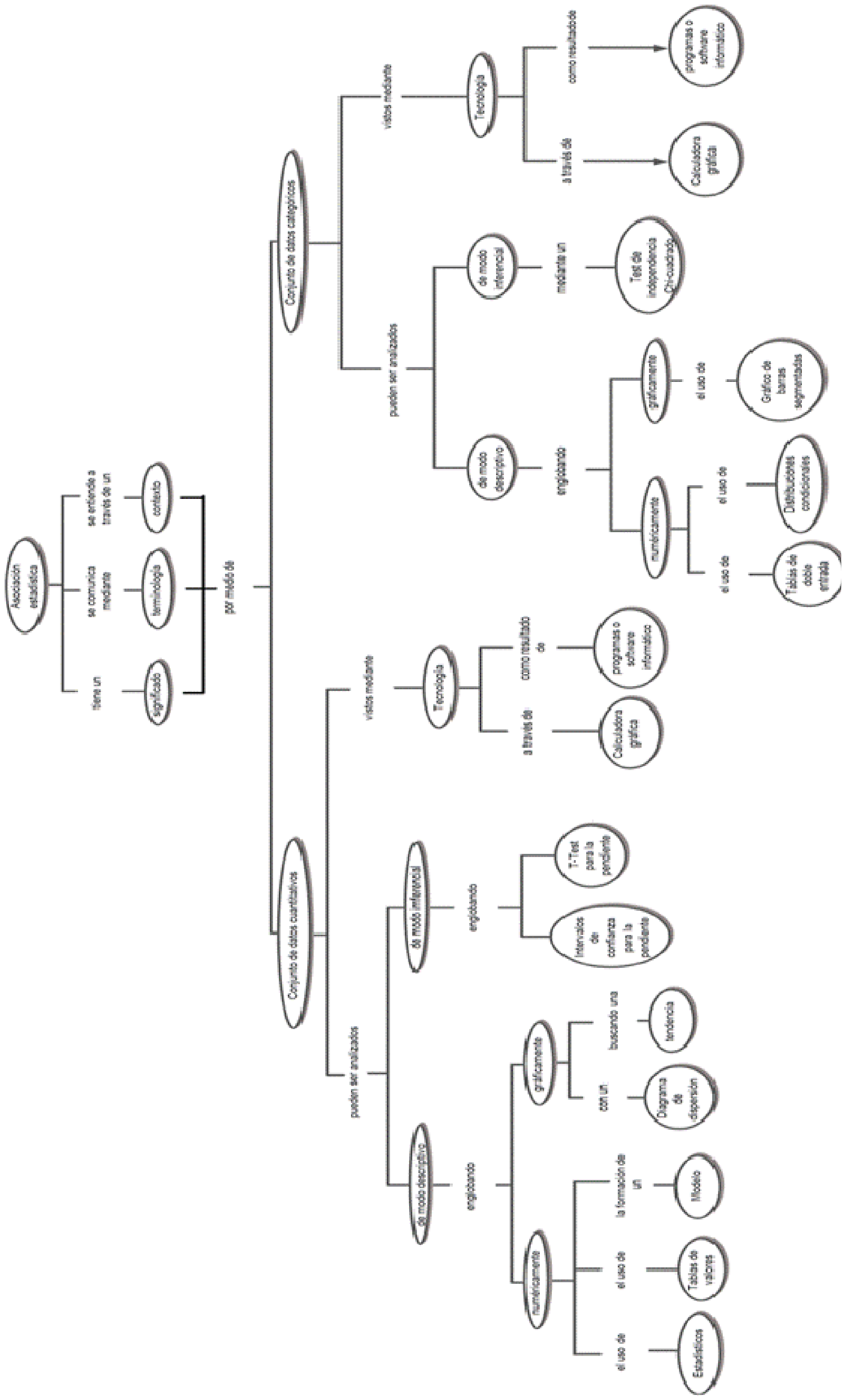


Figura 2.3.6. Mapa conceptual de la noción de asociación estadística (Casey, 2008, p.94)

2.3.5. EL CONOCIMIENTO DEL PROFESOR DESDE EL ENFOQUE ONTOSEMIÓTICO

Para finalizar nuestro análisis de modelos del conocimiento del profesor, presentamos en esta sección el que ofrece el EOS, como herramienta de gran utilidad para la investigación, la formación del profesor y el análisis de la propia práctica docente. Este modelo es denominado por Godino (2009) *Conocimiento Didáctico-Matemático* del profesor, y en él incluye seis facetas o dimensiones, que son:

- *Faceta epistémica:* conocimiento matemático, en nuestro caso de la correlación y regresión, según los objetos matemáticos (problemas, lenguaje, conceptos, propiedades, argumentos y procedimientos) y procesos asociados. Según el autor, incluye el conocimiento común, especializado y ampliado del contenido (en el modelo de Ball y colaboradores).

- *Faceta cognitiva:* conocimiento del razonamiento, aprendizaje y dificultades de los estudiantes, es decir, la evolución de sus significados personales.

- *Afectiva:* conocimiento del grado de implicación (interés y/o motivación) del alumnado en el proceso de estudio. La unión de las facetas cognitivas y afectivas corresponden al conocimiento del contenido y el estudiante en el modelo de Ball.

- *Mediacional:* conocimiento del uso de recursos tecnológicos, manipulativos y de todo tipo, apropiados para la enseñanza-aprendizaje del tema según el nivel de formación o el grado al cual se imparte la enseñanza.

- *Interaccional:* conocimiento de modelos de comunicación entre los actores del proceso de instrucción (en particular, reconocimiento de conflictos semióticos potenciales y su resolución). Estas dos facetas (mediacional e interaccional) cubren el conocimiento del contenido y la enseñanza.

- *Ecológica:* conocimiento del grado en que el proceso de enseñanza y aprendizaje se ajusta al proyecto educativo de la institución (o comunidad) y a los condicionamientos del entorno en que se desarrolla, que cubre el conocimiento del contenido y el currículo en el modelo de Ball y colaboradores.

Además, para cada una de las anteriores facetas, proponen promover la capacidad de análisis didáctico del profesor. Para ello, serán útiles cada uno de los niveles de análisis y estudio de los procesos de enseñanza y aprendizaje de las matemáticas propuestos por el EOS (Godino, Font y Wilhelmi, 2008), permitiendo al profesor tomar decisiones fundamentales sobre cómo realizar de un modo efectivo su labor profesional:

1. Análisis de prácticas matemáticas y didácticas: se identifican problemas planteados, prácticas operativas y discursivas realizadas para resolver una tarea propuesta con enfoque constructivista, y de contextualización del contenido. Incluye las líneas generales de actuación del docente y discentes.

2. Descripción de las configuraciones de objetos y procesos matemáticos: Se identifican los objetos y procesos que intervienen o emergen en la realización de las prácticas, con énfasis en la complejidad, de contenido y didáctica, y sus significados. Funciona como factor explicativo de los conflictos y de la progresión del aprendizaje, en nuestro estudio, de la correlación y regresión.

3. Identificación del sistema de normas y metanormas: Se estudian los fenómenos de índole social a través de una dimensión normativa, la cual condiciona y hace posible el proceso de enseñanza y aprendizaje de la correlación y regresión.

4. Valoración de la idoneidad didáctica del proceso de estudio: sintetiza los análisis previos con el fin de valorar el proceso de estudio como un todo, e identificar potenciales mejoras para futuras implementaciones.

2.3.6. MODELO DEL CONOCIMIENTO DEL PROFESOR EN NUESTRO TRABAJO

En este trabajo utilizaremos una versión reciente del modelo de Conocimiento Didáctico-Matemático de Godino (2014), aún en elaboración, que recoge algunos componentes del conocimiento matemático para la enseñanza del modelo MKT de Ball y colaboradores, expuesto anteriormente, mientras que realiza una división diferente del conocimiento didáctico, teniendo en cuenta las propuestas de Godino (2009; 2011).

Sobre el conocimiento matemático, se consideran dos de las categorías del MKT: el Conocimiento Común, que el profesor comparte con los estudiantes a quienes va dirigida la enseñanza, y el Avanzado (que podríamos asimilar a algunos aspectos del Conocimiento en el Horizonte Matemático) y que corresponde a etapas educativas posteriores (Ver Figura 2.3.7). Nosotros consideraremos estas dos categorías en el Estudio 2, en el que se lleva a cabo un estudio exploratorio de evaluación de estos dos componentes para el estudio específico de la correlación y regresión, y de conceptos previos requeridos, en una muestra de futuros profesores.

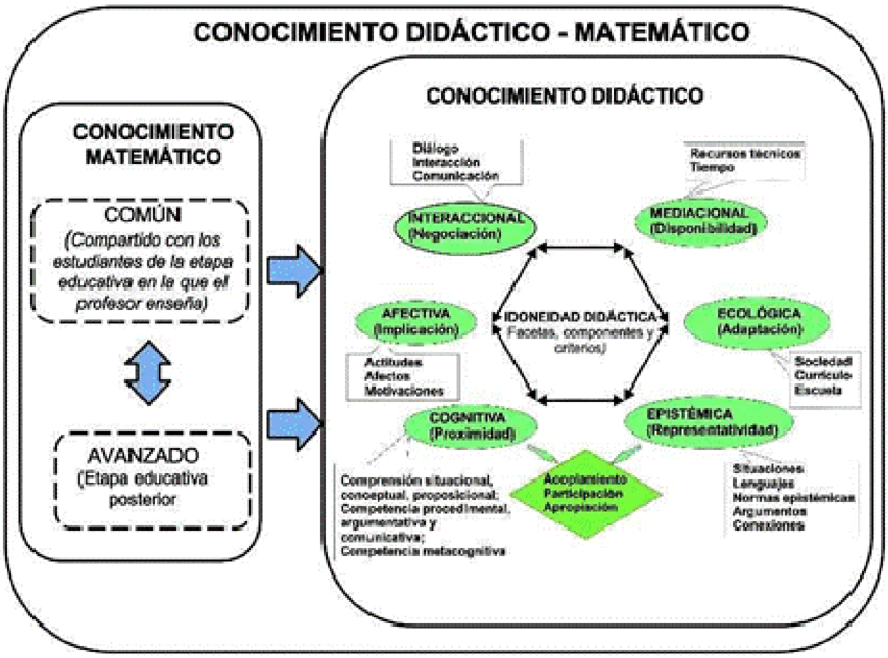


Figura 2.3.7. Conocimiento Didáctico-Matemático (Godino, 2014)

En el conocimiento didáctico se diferencian las facetas epistémicas, cognitiva, afectiva, interaccional, mediacional y ecológica, cada una de las cuáles refleja las correspondientes componentes, ya descritas, de la idoneidad didáctica. En el Capítulo 6 se lleva a cabo un estudio exploratorio de evaluación (Estudio 3) de las diferentes facetas del conocimiento didáctico propuestas por Godino, entre ellas, la faceta epistémica incluiría, para nosotros, el Conocimiento Especializado del Contenido del MKT, la cognitiva y afectiva reflejarían el Conocimiento del Contenido y el Estudiante, y el resto, el Conocimiento del Contenido y la Enseñanza y Conocimiento del Contenido y el Currículo. Sin embargo, las facetas propuestas por Godino son, en general, más amplias, y además permiten aplicar las herramientas EOS en su análisis.

Además, en los Estudios 2 y 3 adoptaremos la metodología propuesta por Godino (2009), que consiste en dos pasos:

1. Elegir una serie de tareas matemática cuya solución ponga en juego los principales objetos matemáticos que forman parte del significado institucional pretendido (en este caso correlación y regresión);
2. Formular consignas que cubran las distintas facetas del conocimiento didáctico-matemático del profesor.

Para evaluar el conocimiento del contenido (común y avanzado), la consigna sería resolver tareas que forman parte del campo de problemas asociado a la correlación y regresión, y que se definirán con detalle en el Estudio 1. Para ello, propondremos a los futuros profesores una serie de tareas que conjuntamente constituyen un proyecto estadístico, pues la enseñanza por proyectos es actualmente recomendada en la estadística (Batanero y Díaz, 2011; MacGillivray, y Pereira Mendoza, 2011).

En dicho proyecto se comenzará con tareas relacionadas con conceptos previos a la idea de correlación y regresión, como interpretación de tablas, gráficos y estadísticos. Posteriormente se presentarán tareas específicas del estudio de la correlación y regresión. Puesto que la mayoría son adecuadas para estudiantes de secundaria, consideramos que se evalúa el conocimiento común del contenido, mientras que las tareas finales relacionadas con la tecnología, evalúan el conocimiento avanzado porque, aunque pudieran usarse en Bachillerato, en la actualidad no es frecuente su uso.

Los diferentes componentes del conocimiento didáctico en el modelo de Godino (2014) se evalúan mediante la aplicación, en una parte de la misma muestra de estudiantes que completa el Estudio 2, de una nueva tarea de análisis del sistema de componentes y descriptores de la idoneidad didáctica de un proceso de estudio (que se aplicará a la experiencia vivida por los mismos estudiantes en el proyecto estadístico previamente realizado), siguiendo el mismo método de Arteaga (2011).

CAPÍTULO 3.
INVESTIGACIONES PREVIAS

3.1. INTRODUCCIÓN

El presente capítulo, constituye una síntesis de las investigaciones más relevantes sobre las nociones de correlación y regresión. Para enmarcar el estudio desde un punto de vista más amplio, consideramos la asociación entre variables cualesquiera, por lo que se comienza el capítulo con un análisis de lo que se ha denominado *razonamiento covariacional*, y los pasos que se han considerado para resolver este tipo de tareas.

Una vez centrados en el estudio de la dependencia en variables cuantitativas, se comienza describiendo las estrategias intuitivas de los estudiantes en la detección de la correlación, a partir de diagramas de dispersión. Seguidamente, se analizan los sesgos y concepciones que se describen para esta tarea; la precisión en la estimación del coeficiente de correlación desde diversas representaciones; y el desarrollo del razonamiento correlacional con la enseñanza. Se completa con el análisis de la presentación de la correlación y regresión en los libros de texto diseñados para su enseñanza. Para ello comenzamos con unas consideraciones sobre la transposición didáctica y la necesidad derivada de análisis de los textos, describiendo brevemente algunas investigaciones sobre la presentación de contenidos de estadística y probabilidad.

Un segundo bloque incluye el análisis del conocimiento del profesor de diferentes objetos matemáticos (gráficos estadísticos, medidas de posición central, dispersión y distribución); su preparación didáctica para la enseñanza de la estadística, dando paso al estudio de la comprensión que posee de la correlación y regresión . Se finaliza con las conclusiones sobre las investigaciones previas y el modo en que estas conclusiones influyen en nuestro trabajo.

3.2. INVESTIGACIÓN SOBRE CORRELACIÓN Y REGRESIÓN

3.2.1. RAZONAMIENTO COVARIACIONAL

Explicar, controlar y predecir los sucesos que se presentan en nuestro día a día, depende de habilidades y destrezas para detectar covariaciones entre los acontecimientos que se suceden (Alloy y Tabachnik, 1984). En consecuencia, el razonamiento covariacional se encuentra presente en la vida cotidiana del ser humano, como actividad cognitiva fundamental (Moritz, 2004; Zieffler, 2006; McKenzie y Mikkelsen, 2007), aunque no exenta de dificultades.

Disciplinas como la psicología, sociología y la didáctica de la estadística, entre otras, proporcionan líneas de investigación sobre este tema (Moritz, 2004). Desde la psicología, Alloy y Tabachnik plantean una importante cuestión: "*¿Bajo qué condiciones son los organismos precisos en la detección de la covariación entre sucesos?*" (Alloy y Tabachnik, 1984, p. 113). Los autores reducen a dos las fuentes de información relevantes para percibir el grado de covariación entre eventos:

- Las *expectativas previas* o *creencias;* ya que el "individuo" posee ideas o juicios sobre la asociación de los fenómenos tratados, debidas bien a la experiencia previa directa con dichos fenómenos en situaciones similares o bien por transmisión cultural o biológica.
- La *información objetiva* de la situación aleatoria, o datos que el individuo dispone acerca de las relaciones de los fenómenos en el entorno en que se encuentra.

Atendiendo a estas dos fuentes, McKenzie y Mikkelsen (2007) proponen un punto de vista bayesiano al razonamiento covariacional. Según estos autores, nuestro sistema cognitivo se aproxima a la tarea de covariación situándose en un marco inferencial más amplio, que depende del entorno. Aunque el ser humano no sea un procesador bayesiano óptimo de la información, sugieren que utiliza el procedimiento bayesiano dado que las personas son sensibles a las creencias previas y a los datos. Los autores indican, que las condiciones en el laboratorio igualmente influyen en el razonamiento de los sujetos. Es decir, pese a los intentos de los experimentadores de descontextualizar las tareas o de eliminar las influencias del mundo real, los participantes en su investigación recurren a hipótesis previas sobre parámetros de la tarea y, además, estas hipótesis parecen coincidir con lo que razonablemente cabría esperar en el mundo real.

3.2.2. PASOS EN UNA TAREA COVARIACIONAL

Diversos autores como Crocker (1981) o Moritz (2004) distinguen pasos o subtareas secuenciales en que se puede descomponer un estudio covariacional, que parten del diseño y la necesidad de recoger datos para estudiar la asociación estadística, hasta el uso de los resultados obtenidos para producir juicios y predicciones de interés para el investigador. Crocker (1981) presenta esta subdivisión de tareas del siguiente modo:

1. Decidir los tipos de datos a recolectar;
2. Elegir la muestra de la población de estudio;
3. Diferenciar o interpretar los casos, (codificar los datos recolectados);
4. Recodificar y estimar las frecuencias de los casos determinados;
5. Integrar resultados; y por último,
6. Utilizar las estimaciones como fundamento para hacer predicciones o emitir juicios.

Moritz (2004) describe una secuenciación equivalente, resumida en cuatro pasos, donde el inicial es la generación de hipótesis sobre la asociación de las variables estadísticas de estudio. Una sugerencia didáctica que Moritz señala es la importancia de la construcción de gráficas estadísticas en el diseño y secuenciación del proceso de enseñanza. Por ejemplo, considerar las gráficas de series temporales permitirá a los estudiantes considerar la asociación bivariada de modo natural, cuando se contemple la variación de la variable de interés en cuanto al tiempo.

Zieffler (2006) resume los diferentes estudios relativos al razonamiento covariacional desarrollados en Psicología, Educación, Didáctica de la Matemática y Didáctica de la Estadística, e indica que la mayoría de los sujetos se centran en el paso (6) anterior, es decir, determinar si existe relación entre dos variables presentadas en una tabla de contingencia para después justificar esta decisión. Aunque las tareas varíen ligeramente, hay un esquema común de investigación, que abarca cuatro categorías de codificación:

- *Covariación mínima.* Los sujetos utilizan la información de una sola celda.
- *Covariación inadecuada.* Los sujetos utilizan la información de dos celdas.
- *Covariación adecuada.* Los sujetos utilizan las cuatro celdas para justificar la covariación.
- *Covariación avanzada.* Los sujetos basan sus juicios en cálculos elaborados atendiendo a las cuatro celdas.

Los principales resultados de estas investigaciones son los siguientes (Ver exposición más detallada en Estepa, 1994; 2004 y Cañadas, 2010):

- Las creencias o teorías previas de los individuos sobre la relación entre las variables de estudio tienen gran influencia en sus juicios de covariación; fenómeno conocido como *correlación ilusoria* (Chapman y Chapman, 1967);
- Los juicios de asociación parecen estar más influenciados por la presencia conjunta de las variables de estudio que por la ausencia conjunta de éstas;
- Existe una gran dificultad para el razonamiento covariacional negativo, esto es, para identificar la asociación inversa entre las variables de estudio, efecto que fue descrito por Estepa (1994);
- Los juicios de asociación tienden a evidenciar una asociación inferior a la que realmente presentan las variables de estudio (Alloy y Tabachnik, 1984); y
- Existe una tendencia a establecer relaciones de causalidad en el estudio de la asociación de entre variables (también descrito por Estepa, 1994).

3.2.3. LAS TAREAS DE CORRELACIÓN

En el desarrollo de habilidades y destrezas de razonamiento covariacional, se requiere utilizar implícitamente objetos matemáticos relacionados con el análisis de datos bivariados. Los más importantes son los referidos a la correlación y regresión entre variables numéricas:

> Los problemas a que se enfrentan los estudiantes referidos al razonamiento covariacional aparecen a lo largo del plan de estudios de introducción a la estadística. Estos a menudo incluyen examinar la relación entre dos variables cuantitativas (por ejemplo, la correlación y la regresión) o dos variables categóricas (por ejemplo, las tablas de doble entrada, el test de chi-cuadrado) (Zieffler, 2006, pp.6-7).
> Todo esto, de algún modo, conlleva al profesorado a interesarse por el razonamiento de los estudiantes en cuanto a la lectura diagramas de dispersión, a la interpretación de la correlación y otras destrezas que son utilizadas en el estudio e interpretación de los datos bivariados (Zieffler, 2006, p. 7).

Es por ello de nuestro interés, centrarnos en el estudio de la correlación y regresión, para fundamentar el diseño del proceso de enseñanza y su implementación en el aula, y la evaluación del alumnado en la comprensión de estos objetos matemáticos.

La investigación desarrollada en Psicología en torno al razonamiento covariacional, llevada a cabo en un amplio rango de edad (desde estudiantes de primaria hasta estudiantes universitarios), ha producido resultados muy sólidos (Estepa, 2004; Zieffler, 2006). Y aunque las investigaciones didácticas se interesen más por cómo los estudiantes razonan en un contexto educativo específico, esto es "*la ciencia para resolver problemas*" (Zieffler, 2006, p.14), las tareas dadas a los sujetos en la investigación didáctica son similares a las descritas en Psicología.

La mayor parte de esta investigación ha sido enfocada al estudio de la evaluación de la covariación de variables binarias (presencia-ausencia de cierto carácter en la población), evidenciándose en estos estudios la pobreza de razonamiento del ser humano en el desempeño de estas tareas (Pérez Echeverría, 1990; Estepa, 2004, Zieffler, 2006; McKenzie y Mikkelsen, 2007; Cañadas, 2012). En las siguientes secciones, se recogen los resultados más relevantes que, en torno a la enseñanza y aprendizaje de la correlación y regresión, se ha investigado.

3.2.4. ESTRATEGIAS EN LA ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DE DIAGRAMAS DE DISPERSIÓN

Diversas investigaciones (Sánchez Cobo, 1999; Sánchez Cobo, Estepa y Batanero, 2000; Cañadas, 2012; Cañadas, Batanero, Contreras y Arteaga, 2011) llevan a cabo un estudio pormenorizado de la resolución de las tareas de covariación propuestas a estudiantes, que podemos dividir en tres tipos fundamentales (Sánchez Cobo, 1999):

a. *Juicios de asociación en tablas de contingencia.* Donde se pide analizar la asociación entre dos variables cualitativas y el sujeto debe utilizar las frecuencias de dicha tabla.

b. *Juicios de asociación en diagramas de dispersión.* Donde se trata de analizar la correlación entre dos variables numéricas a partir de la forma y dispersión de la nube de puntos.

c. *Juicios de asociación en la comparación de muestras.* Donde se pide analizar la relación entre una variable cuantitativa y otra variable cualitativa.

En los estudios llevados a cabo en torno a cada uno de estos bloques de tareas, se presentan diferentes estrategias de resolución por parte de los estudiantes (Estepa, 1994; Batanero, Estepa, Godino y Green, 1996; Sánchez Cobo, 1999; Cañadas, 2012). Nosotros nos restringimos al estudio de las estrategias empleadas en la estimación de la asociación entre dos variables numéricas, es decir, la estimación de la correlación a partir de diagrama de dispersión.

Estepa y Batanero (1996) estudian dichas estrategias en una muestra de 213 estudiantes del curso preuniversitario. Los autores indican que en estos diagramas puede determinarse visualmente la asociación entre las dos variables. Clasifican las estrategias encontradas en correctas, parcialmente correctas e incorrectas:

- Estrategias correctas:

    1. *Comparación global.* Cuando la respuesta del alumno se basa en una comparación global de la relación entre las dos variables.

    2. *Crecimiento.* Se usa como argumento el crecimiento, decrecimiento o constancia de la forma de la nube de puntos para justificar el tipo de dependencia.

- Estrategias parcialmente correctas:

    3. *Comparación con un patrón.* El alumno compara la forma de la nube de puntos con una función conocida, por ejemplo, lineal o cuadrática, para argumentar la relación entre las variables. Esta respuesta es parcialmente correcta ya que los datos podrían ajustarse a un patrón diferente del lineal.

    4. *Interpretación correcta de puntos aislados* que cumplen el tipo de relación existente entre las variables. Esta respuesta es parcialmente correcta porque otros puntos pudieran contradecir el tipo de correlación estimada al no utilizar toda la información aportada por el diagrama de dispersión.

- Estrategias incorrectas:

    5. *Interpretación incorrecta de puntos aislados*. El alumno emplea pares aislados de valores para justificar, de forma incorrecta, la relación entre las variables.

6. *Teorías previas.* Se usan las teorías previas sobre el contexto para justificar la asociación, y no tanto la evidencia empírica de los datos, es decir, se manifiesta la correlación ilusoria.
7. *Otras variables.* Cuando la existencia de otras variables que puedan influir en la dependiente es considerada como motivo para la no existencia de asociación.
8. *Uniformidad.* Se argumenta que no existe dependencia porque a un solo valor de la variable dependiente pueden corresponder varios de la dependiente, ya que la relación no es de tipo funcional.
9. *Causalidad.* A pesar de la evidencia que muestren empíricamente los datos, se argumenta que no existe asociación entre las variables ya que no existe relación causa-efecto entre las mismas.

La investigación desarrollada por Sánchez Cobo (1999) con estudiantes universitarios, corrobora la existencia de factores que explican las relaciones existentes entre las tareas planteadas y las estrategias empleadas:

- En primer lugar, se encuentra una oposición entre el razonamiento numérico y gráfico covariacional: Ante la tarea de estimar el coeficiente de correlación a partir de la descripción verbal, se asocia la estrategia que pone en juego el marco gráfico. "*Los alumnos recurren a la representación gráfica para estimar el coeficiente de correlación cuando ésta no aparece en el enunciado del problema*" (Sánchez Cobo, 1999, p. 252).
- Se muestra también un empleo de argumentaciones conjuntas (numéricas, gráficas, y teorías previas) donde tan solo se requiere un argumento gráfico. Esta estrategia se presenta principalmente en situaciones cercanas al alumno.

3.2.5. ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DE DIFERENTES REPRESENTACIONES

Sánchez Cobo (1999) hizo notar que los alumnos estiman de un modo más preciso el coeficiente de correlación a partir de un diagrama de dispersión, así como la tarea inversa, haciéndose mayor esta precisión cuando la correlación es más intensa. El autor sugiere que esto puede ser debido al protagonismo que los diagramas de dispersión adquieren en la enseñanza de la correlación y regresión.

El razonamiento sobre datos bivariados implica procesos de traducción entre los datos, sus representaciones gráficas, y las descripciones verbales sobre las variables. También procesos de cálculo estadístico, uso de modelos matemáticos para el ajuste de datos, y traducción desde y hacia las expresiones algebraicas y gráficas utilizadas; además, la explicación de la posible existencia de relaciones causales. Moritz (2004), presenta un resumen (Figura 3.2.1) donde se reflejan las formas más comunes de representar la correlación (que él denomina covariación estadística) junto a las destrezas o habilidades requeridas.

El estudio desarrollado por Sánchez Cobo y cols. (2000), nos acerca un poco más a comprender el modo en que los alumnos universitarios adquieren la noción de correlación y llevan a cabo estos procesos de traducción, todo ello cimentado en los estudios previos (Batanero, Estepa y Godino, 1997; Batanero y Godino, 1998; Batanero, Godino y Estepa, 1998; Estepa 1994; Estepa y Batanero, 1995, 1996; Estepa y Sánchez Cobo, 1998, entre otros), que se centran en el análisis de las concepciones de los estudiantes, el efecto sobre las estimaciones de variables aisladas, o el análisis de la dificultad de tareas particulares (Sánchez Cobo y cols., 2000).

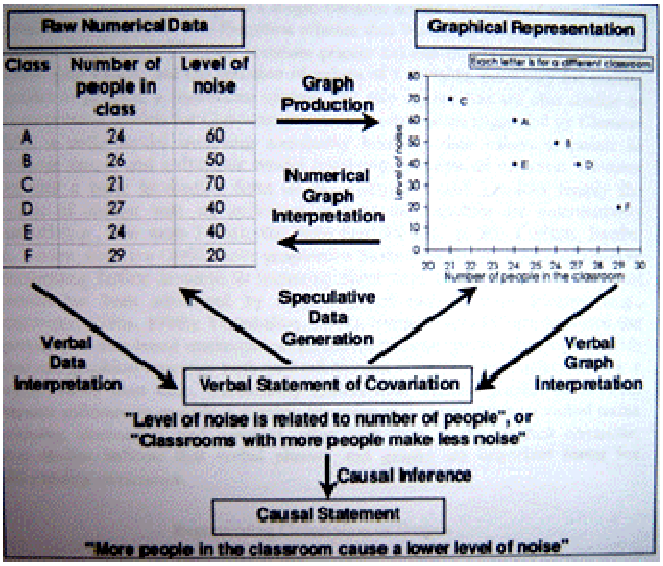


Figura 3.2.1. Representaciones de la correlación y destrezas requeridas (Moritz, 2004, p. 230).

Dada la necesidad de un análisis sistemático de las variables que influyen en las tareas de traducción de la correlación de una representación a otra, y de su efecto sobre la estimación de la correlación y las estrategias empleadas por alumnos, Sánchez Cobo y cols. (2000), abordan una investigación que considera cuatro formas de representar la correlación entre dos variables cuantitativas:

a. *Descripción verbal*, cuando describimos una distribución bivariada mediante el lenguaje natural

b. *Tabla de valores*, o presentación de un conjunto de pares de valores numéricos de una distribución bivariada,

c. *Diagrama de dispersión*, cuando el conjunto de pares de valores de una distribución bivariada se presentan mediante un diagrama cartesiano, y

d. *Coeficiente de correlación*, cuando se da el valor del coeficiente de correlación existente entre las variables de una distribución bivariada.

En las seis tareas propuestas (cada una de ellas con cinco apartados) se presentan diferentes tipos de covariación (Barbancho, 1973):

- *Dependencia causal unilateral*: Cuando la ocurrencia de una variable *X* (causa) influye en la ocurrencia de *Y* (efecto), y no al contrario.

- *Interdependencia*: Cuando la ocurrencia de una variable *X* influye en la ocurrencia de una variable *Y*, y viceversa.

- *Dependencia indirecta*: Dos variables pueden mostrar cierta covariación debido a la variación de una tercera variable que está correlacionada con ambas..
- *Concordancia*: Correlación producida por la ordenación de un conjunto de datos por dos personas de forma independiente.
- *Covariación casual*: Cuando en la covariación de dos variables hay cierta sincronía, lo cual podría interpretarse como la existencia de asociación entre ambas; pero ésta es casual o accidental.

Se incluyen, además, como variables de tarea la intensidad y dirección de la correlación, la posible linealidad del ajuste de los datos, y las teorías previas de los alumnos. Aunque las tareas no siempre resultan sencillas, los alumnos mostraron una buena capacidad de estimación de la correlación, haciéndose más viable cuando mayor es la intensidad de ésta, y cuando se presenta causalidad unilateral o interdependencia. Se presenta una mayor dificultad al construir un diagrama de dispersión a partir de una descripción verbal, o estimar el coeficiente de correlación a partir de una tabla de valores. Como mostró Sánchez Cobo (1999), el alumno podría tener un conflicto semiótico entre el crecimiento/decrecimiento no uniforme de las dos variables mostradas mediante una tabla de datos y el crecimiento/decrecimiento uniforme de la dependencia funcional.

Moritz (2004) se centra en el estudio de tres destrezas importantes como son: (a) generar datos especulativos, mediante el desarrollo de gráficos que reflejen los juicios de asociación textuales; (b) interpretar gráficos tales como diagramas de dispersión con la correspondiente emisión de juicios de asociación; y por último, (c) interpretar tablas de frecuencias de datos covariados. En cuanto a las dificultades generales que presenta un adecuado razonamiento covariacional, destaca las siguientes:

- Considerar la tendencia global de los datos bidimensionales en lugar de considerar tan sólo la correspondencia de datos aislados;
- Considerar las correspondencia de las dos variables implicadas en el estudio de modo conjunto, en lugar de considerar cada una de las variables de modo aislado;
- La existencia de teorías o creencias previas en cuanto a las variables de estudio y su asociación.

Algunas sugerencias para solventar estas dificultades son:

- La lectura progresiva, punto a punto, hasta una posterior generalización con los datos disponibles;
- La utilización de un enfoque de variación temporal, que permita a los estudiantes centrarse en el cambio de una variable, mediante la variación implícita de la variable tiempo, para una posterior correspondencia entre variables no temporales;
- Alentar la aparición de las creencias y teorías previas que serían poco a poco equilibradas mediante la información proporcionada por el estudio. Añadir tareas que impliquen un razonamiento covariacional contraintuitivo, donde se cuestione de modo natural la fiabilidad del conjunto de datos de que se dispone.

Como indican McKenzie y Mikkelsen (2007), aunque las creencias y teorías previas son imprecisas e incluso inexactas, son de interés en sí mismas, dado que cuestionan cómo debiese ser la relación de dichas teorías y la información de la tarea desde una perspectiva bayesiana, categorizando éstas creencias previas no como un error, sino como útil de enseñanza. Una actividad curiosa que Moritz (2004) propone, es el uso de gráficos manuales desarrollados de modo anónimo por los estudiantes, para su estudio en el aula.

3.2.6. SESGOS EN EL RAZONAMIENTO CORRELACIONAL

Otro punto en que se ha centrado la investigación es la presencia de errores o sesgos en los juicios de correlación, o en la estimación de su signo e intensidad. Igualmente se han descrito otros en la detección de la covariación, en general, que son debidas a existir en los sujetos expectativas o esquemas referidos a los estímulos presentes en la situación a que se enfrentan (Crocker, 1981; Alloy y Tabachnik, 1984).

El sesgo que más interés ha suscitado es el denominado *correlación ilusoria* (Chapman y Chapman, 1967) que hace referencia a la presencia de errores sistemáticos producidos por variables inherentes a los estímulos que el sujeto percibe, y que conllevan a juicios erróneos de covariación. Como sus autores indican:

> Correlación ilusoria es un error sistemático en el informe de tales relaciones. Es definida según el informe dado por un observador de una correlación entre dos clases de eventos que en realidad (a) no están correlacionados, o (b) están correlacionados en menor medida que lo manifestado, o (c) están correlacionados en sentido contrario del que se manifiesta en el informe. (Chapman y Chapman, 1967, p. 194).

Es clara la vulnerabilidad del ser humano a creer lo que ve, intuye o siente frente a lo que cualquier informe le pueda proporcionar (Chapman y Chapman, 1967). Es por ello que muchos autores proponen advertir a nuestros estudiantes de la dificultad que conlleva disponer de un razonamiento covariacional correcto, con el fin de sensibilizar ante el peligro que acarrea sus sesgos o errores (Chapman y Chapman, 1967,1969; Moritz, 2004). Engel y Sedlmeier (2011), describen los siguientes sesgos complementarios:

- *No tener en cuenta las variables extrañas.* La paradoja de Simpson ocurre cuando se olvida una variable extraña que influye en la correlación entre otras dos. Dicha variable podría hacer cambiar el sentido de la correlación o su intensidad.

- *El efecto de regresión.* Es habitual en una situación de test- retest que los sujetos con puntuaciones atípicas en la primera prueba vuelvan hacia el valor medio en la segunda. Esto se considera, por algunas personas, indicativo de un cambio debido al tratamiento, cuando es una propiedad de la regresión.

3.2.7. CONCEPCIONES SOBRE LA CORRELACIÓN

En su estudio, Estepa (1994) observa que, en algunos estudiantes, una estrategia correcta o parcialmente correcta no se corresponde con una decisión correcta. Considera que ello se debe a que el estudiante tiene concepciones erróneas sobre la correlación. A lo largo de sus trabajos describe las siguientes (Estepa, 1994; Estepa y Batanero, 1996;

Batanero, Estepa y Godino, 1997):

- *Concepción determinista de la correlación*. Los alumnos tienden a asignar un único valor de la variable independiente a cada uno de los valores de la variable dependiente. Esto es, la relación de las variables sólo es considerada desde un punto de vista funcional. En casos extremos, los estudiantes exigen la existencia de una fórmula que ligue las dos variables. Los autores denominan a esta creencia *concepción algebraica* de la asociación.
- *Concepción local de la correlación.* Los alumnos utilizan parte de los datos del estudio, y se limitan a confirmar la asociación según este subconjunto pues, de algún modo justifica algún tipo de patrón, obviando la tendencia global de los datos.
- *Concepción unidireccional:* En este caso el estudiante no admite la asociación inversa, considerándose la intensidad de la asociación, pero no su signo. Se presentan casos en los que los alumnos consideran la dependencia inversa como independencia.
- *Concepción causal:* Cuando el sujeto sólo considera la dependencia entre variables si puede adjudicarse a la presencia de una relación causal entre las mismas. Por ejemplo, si hay terceras variables que afecten la relación, piensan que no hay relación entre las dos primeras.
- *Transitividad.* El coeficiente de correlación no tiene propiedad transitiva; es decir, dos variables *A* y *B* pueden estar correlacionadas, así como *B* y *C*, sin estarlo *A* y *C*. Esta propiedad no se comprende por parte de los estudiantes (Castro-Sotos, Van Hoof, Van den Noortgate y Onghena, 2009).

3.2.8. DESARROLLO DEL RAZONAMIENTO CORELACIONAL EN LA ENSEÑANZA

Algunos autores han analizado el aprendizaje de la correlación y regresión con experiencias específicas docentes. A continuación describimos los principales.

3.2.8.1. ESTUDIO DE ESTEPA

Estepa (1994) estudia el efecto de la enseñanza basada en un uso intensivo de ordenadores, en un curso de análisis exploratorio de datos, sobre el aprendizaje del concepto de asociación con 22 futuros maestros, incluida la correlación.

En Batanero, Estepa y Godino (1997) realizan un estudio de los cambios después de la instrucción, para mejorar los conocimientos de los estudiantes. Para esto, se propuso un test posterior a la instrucción a la muestra experimental, observándose como había influido la instrucción en sus estrategias y concepciones. En los ítems relacionados con los juicios de correlación en diagramas de dispersión hubo un aumento fuerte en el número de estrategias correctas (casi el doble), aunque el cambio no fue homogéneo. La mayoría de los estudiantes de la muestra superaron la concepción determinista de la correlación, y la concepción local fue eliminada al comprender la importancia de tener en cuenta los datos completos. Los autores informan que, en general, no se observa mejora respecto a la concepción causal de la correlación. Por consiguiente, creen los autores que hay una necesidad de encontrar las nuevas actividades prácticas que ayuden a los estudiantes a reflexionar sobre este tema.

Estepa (1994) también propuso a los alumnos una prueba final para ser resuelta con ayuda del ordenador. Esta evaluación era abierta (en el sentido que el alumno podía usar cualquier programa del software proporcionado) e incluyó una pregunta sobre la correlación numérica. Aunque algunos alumnos lo resolvieron correctamente mediante el estudio del coeficiente de correlación, o a partir del diagrama de dispersión, otros lo trataron incorrectamente como un problema de comparación de las dos variables (y no de asociación) aplicando procedimientos inapropiados, por ejemplo, comparar promedios de las dos variables. En consecuencia, como en otras investigaciones se deduce que el ordenador por sí sólo no es suficiente para que los estudiantes adquieran competencia en el análisis de datos.

Actos de comprensión de la asociación

Estepa (1994) analizó mediante grabaciones, entrevistas y análisis de las tareas escritas y resueltas por ordenador, el proceso de enseñanza para dos estudiantes, con objeto de describir actos de comprensión de algunos elementos de significado de la asociación (Sierpinska, 1994), describiendo los siguientes relacionados con la correlación:

1. *Para estudiar la correlación se ha de tener en cuenta la distribución completa*. Encontrar las diferencias locales no es suficiente, puesto que la asociación debe ser deducida de los datos completos.

2. *De la misma frecuencia absoluta se pueden ser calculadas dos frecuencias relativas condicionales diferentes, dependiendo de cuál es la variable condicionada*. El papel de la condición y lo que se condiciona en la frecuencia relativa condicional no es intercambiable. Falk (1986) y otros autores han señalado que los estudiantes tienen dificultades en la interpretación de las probabilidades condicionales, porque no discriminan entre la probabilidades *P(A/B)* y *P(B/A).* Varios estudiantes de Estepa mostraron una confusión similar, aunque lo solucionaron con ayuda del profesor.

3. *Dos variables son independientes si la distribución de una de estas variables no cambia cuando la condicionamos por valores de la otra variable*. Hasta la sesión 5, los estudiantes no descubrieron que una condición para la independencia era que la distribución fuese invariable respecto de la distribución de frecuencia condicional, cuando cambia el valor de la variable condicional.

4. *Cuando se estudia la asociación (correlación), ambas variables juegan un papel simétrico. Sin embargo, cuando investigamos la regresión el papel jugado por las variables no es simétrico.* El hecho es que la correlación ignora la distinción entre explicativas y variables de respuesta, mientras en la regresión esta diferencia es esencial (Moore, 1995), lo cual causa mucha confusión a los estudiantes.

5. *Una correlación positiva señala a una asociación directa entre las variables*. Aunque, en la sesión 6, algunos estudiantes podían interpretar el tamaño del coeficiente de correlación, no hablaron del tipo de la asociación (directa o inversa). Al final de la sesión, notaban que cuando el coeficiente de correlación es positivo, y hay una relación lineal, las variables son asociadas positivamente, pero no usaron el término "Asociación directa" explícitamente.

6. *Una correlación negativa señala a una asociación inversa entre las variables*. Cuando, en la sesión 6, los estudiantes encontraron un coeficiente de correlación negativo por primera vez, se extrañaron tanto que preguntaron a su profesor si esto era posible. También tenían problema cuando comparaban dos coeficientes de correlación negativos. Los alumnos no usaron el término "asociación inversa" explícitamente, ni diferenciaron entre los dos tipos de la relación (directa e inversa) al final de su aprendizaje.
7. *El valor total del coeficiente de correlación indica la intensidad de la asociación*. Aunque los estudiantes relacionaron el valor total del coeficiente de correlación con la intensidad de la relación, no relacionaron esta idea con la dispersión de los datos.

3.2.8.2.ESTUDIO DE SÁNCHEZ COBO

Sánchez Cobo (1999) intenta caracterizar el significado personal que los alumnos universitarios dan a la correlación y regresión al finalizar un curso de introducción de estadística, mediante la descripción de sus errores conceptuales y procedimentales, la estimación del coeficiente de correlación a través de diferentes representaciones de la correlación (verbal, tabla, numérica y gráfico) y la capacidad de traducción entre dichas representaciones.

Su muestra estuvo compuesta de 193 estudiantes de Diplomatura de Empresariales (104) y Diplomatura de Enfermería (89) de la Universidad de Jaén. Sus resultados indican que la mayoría de los alumnos conocen que el signo de la covarianza denota la dirección de la correlación existente entre las componentes de la variable bidimensional. Así mismo, un pequeño porcentaje de alumnos no toma en consideración el decrecimiento de la intensidad de la dependencia al disminuir el valor absoluto de la covarianza, además de la posibilidad de que la relación entre las variables sea no lineal (aunque se presente una covarianza positiva).

Sánchez Cobo (1999) hizo notar que los alumnos, al finalizar la enseñanza, comprenden con facilidad la adimensionalidad del coeficiente de correlación, así como la relación entre el signo de la correlación y el sentido en que covarían los valores de una variable bidimensional. No obstante, presentan dificultades en la concepción de una covariación negativa, captada por menos del cincuenta por ciento de los sujetos. La estimación es más precisa al estimar el coeficiente de correlación a partir de un diagrama de dispersión, así como la tarea inversa, haciéndose mayor esta precisión cuando la correlación es más intensa:

> Es pertinente manifestar que los estudiantes tienen más facilidad en conectar la dispersión de la nube de puntos con la intensidad de la dependencia que con el coeficiente de correlación. Además, parece más sencillo de comprender la correspondencia entre la intensidad de la dependencia y la dispersión de la nube de puntos que la existencia entre el signo de la correlación y la pendiente de la recta de regresión. (Sánchez Cobo, 1999, p.210).

El autor sugiere que esto puede ser debido, a la familiaridad con los diagramas de dispersión. Los errores son mayores al construir una nube de puntos a partir de una descripción verbal y estimar el coeficiente de correlación desde una tabla de valores numéricos. En cuanto a la capacidad de proponer situaciones factibles a un coeficiente de correlación dado, los alumnos en su mayoría proponen variables bidimensionales consistentes. El 63,5% de los alumnos propone variables cuya correlación tiene el mismo signo que el dado en el problema y se observan dificultades en la identificación de la dependencia funcional y, en menor medida, de la independencia. Otros resultados de este estudio son:

- Aunque los alumnos tienen presente que la intensidad de la dependencia se obtiene a partir del coeficiente de correlación, presentan dificultades al comparar diferentes valores del coeficiente de correlación distintos de: -1, 0 y 1.
- Presentan dificultad al diferenciar la variable explicativa de la explicada en cuanto al cálculo de la recta de regresión. Este mismo resultado es encontrado por Ruiz (2006) en la interpretación de la distribución de probabilidad de una variable aleatoria.
- Dos de cada cinco estudiantes, aproximadamente, relacionan que ambas rectas de regresión son perpendiculares cuando el coeficiente de correlación es nulo, evidenciándose que se considera casi en exclusiva, la modelización de ajuste lineal. En concreto, casi la mitad de los sujetos del estudio consideran que si existe correlación positiva, ésta se deduce en una dependencia lineal.

3.2.8.3.ESTUDIO DE ZIEFFLER

Algunos factores que podrían explicar el razonamiento covariacional son la secuenciación de los temas en un curso de estadística, o el desarrollo del razonamiento sobre temas anteriores. A este respecto, Zieffler (2006) nos plantea las siguientes cuestiones:

1. ¿Cuál es la naturaleza o el patrón de razonamiento de los estudiantes en cuanto a los datos bivariados?
2. La secuenciación de conocimientos en un curso de introducción a la Estadística, ¿influye en el razonamiento covariacional?
3. Los cambios del razonamiento de los estudiantes al adquirir nociones estadísticas fundamentales como la de distribución estadística, ¿influyen en los patrones de razonamiento covariacional?

Para responder estas preguntas, el autor realiza una investigación en un curso universitario de introducción a la estadística con 113 estudiantes. Parte de evaluaciones del razonamiento de los estudiantes sobre la distribución de una variable estadística, previo análisis de diseño de muestras (muestreo) y AED, y en ese momento, dos profesores ordenaron al azar los temas referentes al análisis bidimensional. Uno comenzó su programa con este tema y concluyó con los temas relacionados con distribuciones muestrales, probabilidad e inferencia y el otro invirtió el orden, dejando el tema de variables bidimensionales para final de curso.

Zieffler observó un desarrollo del razonamiento covariacional, obviamente marcado por la idiosincrasia de cada individuo, donde la mayor parte del desarrollo ocurre al inicio de la instrucción, previo a la enseñanza formal de este tema, evidenciando la existencia del significado personal de este objeto matemático en sí mismo, más que ser producto de una instrucción formal. Lo describe con la metáfora de “modelo cuadrático” porque, aunque los alumnos muestran inicialmente un gran avance en su razonamiento covariacional, con el tiempo decrece su intensidad, incluso retrocediendo. Zieffler explica una posible causa del pobre desarrollo posterior de este razonamiento por la brevedad de la unidad en el curso.

En cuanto a la segunda cuestión planteada, no se observó una influencia en el desarrollo del razonamiento covariacional marcada por la secuenciación del tema. De cualquier modo, Zieffler remarca la influencia, obvia, de la enseñanza de este tema, en otros, como por ejemplo la inferencia. Por último, en cuanto a la tercera cuestión planteada, no se encuentra una marcada influencia del estudio de la variable estadística unidimensional en el desarrollo del razonamiento covariacional, aunque la mayor parte del crecimiento en el razonamiento sobre datos bivariados, según el autor, parecía ocurrir durante la instrucción de la distribución univariante, luego tal vez estos dos tipos de razonamiento estén íntimamente conectados. Respecto a estas dos últimas cuestiones, Sánchez Cobo indica:

> Que no se haga una aproximación al concepto de variable estadística bidimensional a partir de dos variables estadísticas unidimensionales, como se efectuaba en algunos manuales de bachillerato, ya que podría favorecer el que los alumnos sean poco conscientes de que dichas variables unidimensionales tienen que estar referidas a la misma unidad estadística (Sánchez Cobo, 1999, p. 291).

Y en cuanto a la secuenciación de enseñanza, defiende trabajar en primer lugar la correlación y luego abordar la regresión, indicando además:

> Aunque a toda nube de puntos podamos hacerle corresponder una curva de regresión, carecería de interés hallarla si las características de la variable estadística no estuvieran correlacionadas o si nos diera predicciones poco fiables e información de escasa trascendencia (Sánchez Cobo, 1999, p. 291).

3.2.9.ANÁLISIS DE LIBROS DE TEXTO

3.2.9.1.INTRODUCCIÓN

El estudio del proceso de enseñanza y aprendizaje del saber matemático debe ser concebido desde un enfoque sistémico, cuya base es la noción de sistema didáctico:

> El didacta de las matemáticas se interesa en el juego que se realiza -tal como lo puede observar, y luego reconstruir, en nuestras clases concretas- entre un docente, los alumnos y un saber matemático. Tres lugares, pues: es el sistema didáctico. Una relación ternaria: es la relación didáctica. (Chevallard, 1991, p.15).

El saber matemático (conjunto de resultados admitidos como verdaderos por la comunidad científica de referencia) o *saber sabio*, se encuentra sometido a un proyecto social de enseñanza y aprendizaje; y en consecuencia, requiere de ciertas modificaciones para poder ser enseñado (constituyéndose así en saber a enseñar). Así es que surge la noción de transposición didáctica como herramienta que da respuesta a las necesidades del sistema de enseñanza:

> Para que la enseñanza de un determinado elemento de saber sea meramente posible, ese elemento deberá haber sufrido ciertas deformaciones, que lo harán apto para ser enseñado. El saber-tal-como-es-enseñado, el saber enseñado, es necesariamente distinto del saber-inicialmente-designado-como-el-que-debe-ser-enseñado, el saber a enseñar (…). El saber enseñado debe

aparecer conforme al saber a enseñar. (Chevallard, 1991, pp.16-17).

Así, cuando se construye el saber sabio, este se encuentra personalizado por el investigador, y contextualizado por las situaciones y problemas que se han estudiado. Sin embargo, el productor del saber, cuando lo comunica, lo despersonaliza y lo descontextualiza, quedando oculto este proceso de generación, pues se comunica de manera limpia, secuenciada. Estas transformaciones, reformulaciones, generalizaciones y aplicaciones que sufre, algunas veces pueden hacer que pierda su identidad. El trabajo del profesor es, en cierta medida, inverso al del productor del saber: debe producir una recontextualización y una repersonalización de los conocimientos.

En este sentido, desde el currículo propuesto en las directrices curriculares al implementado en el aula, una fase importante es el currículo escrito, que se plasma en los libros de texto (Herbel, 2007). Cordero y Flores (2007) indican que el discurso matemático escolar es determinado con frecuencia por el libro de texto, que regula las acciones de enseñanza y aprendizaje, junto con las creencias de los profesores. Reys, Reys y Chavez (2004) sugieren que los libros de texto presentan las ideas matemáticas en diferentes contextos, a la vez que permiten a los estudiantes explorar diferentes ideas y facilitan el aprendizaje.

En base a una revisión de la literatura, Ortiz (1999) enumera además, los siguientes puntos adicionales, que justifican la importancia del análisis de libros de texto: son una fuente de datos y actividades para el aula; resultan de un gran esfuerzo de planificación y síntesis; se asumen como un conocimiento que hay que transmitir; y el alumno lo considera como una autoridad del conocimiento y guía del aprendizaje.

En consecuencia, se hace necesario el estudio del resultado de la transposición didáctica, que se plasma en los libros de texto, para asegurar que el proceso de transposición no lleva a desajustes entre el significado institucional de los objetos matemáticos y el plasmado en dichos textos.

3.2.9.2. ESTUDIOS SOBRE LA PRESENTACIÓN DE LA ESTADÍSTICA Y PROBABILIDAD EN LOS LIBROS DE TEXTO

Varios autores en nuestro grupo de investigación han analizado los libros de texto, utilizando el enfoque ontosemiótico, con la finalidad de establecer un significado de referencia en su trabajo. A nivel universitario, destacamos los trabajos de Tauber (2001) sobre la distribución normal, Alvarado (2007) sobre el teorema central del límite y Olivo (2008) sobre intervalos de confianza. En lo que sigue describimos muy brevemente las investigaciones sobre la presentación de conceptos de estadística y probabilidad en libros de texto de educación secundaria.

Las medidas de tendencia central

Encontramos un estudio de la presentación del tema en los libros de texto en Cobo (2003) y Mayén (2009). El estudio de Cobo, realizado en una muestra de 22 libros de texto de Educación Secundaria Obligatoria, presenta gran diversidad de acepciones de la idea de media, donde se otorga una mayor importancia a las definiciones y al cálculo de la media que al estudio de sus propiedades. Una posible consecuencia es que, aunque los estudiantes consigan manejar perfectamente los procedimientos de cálculo, pueden no alcanzar una comprensión completa de este concepto, sus posibilidades de uso, y la conveniencia, por tanto, de elegir unas u otras según la situación.

La importancia que tiene el lenguaje matemático en la enseñanza y aprendizaje se pone de manifiesto en la gran variedad de representaciones empleadas por los distintos libros de texto en este estudio. En cuanto a los campos de problemas, Cobo (2003) constata que hay situaciones de las que emerge la idea de media, como es la de estimar una medida a partir de diversas mediciones en presencia de errores, que no se recogen en los libros de texto.

En los contenidos de los libros mexicanos analizados, Mayén (2009) encuentra un tratamiento breve al tema de medidas de tendencia central y mayor énfasis en los algoritmos de cálculo. Por otro lado, los textos se centran más en explicar la definición y cálculo de la media aritmética, que el de la moda y mucho menos el de la mediana, aunque sí presentan distintas formas de resolución. Se utiliza lenguaje y simbología básicos para conocer estos conceptos. También contienen ejemplos para datos agrupados en intervalos y representados en tablas de frecuencias. No se consideran para este tema las representaciones gráficas ni diagramas. Por otra parte, las propiedades de las medidas de tendencia central no se consideran en la enseñanza para este nivel de estudios, aunque sólo algunas se encuentran en los ejemplos de manera implícita. Lo mismo sucede con los argumentos, que son escasos para justificar alguna respuesta.

Probabilidad

Ortiz (1999) analiza el significado de los conceptos probabilístico en 11 libros de texto del primer curso de bachillerato. Siguiendo un procedimiento cualitativo, Ortiz analiza las definiciones, tipos de problemas y representaciones asociadas a los conceptos de aleatoriedad, frecuencia, probabilidad y variable aleatoria. A partir de este análisis, define los elementos de significado asociados cada uno de ellos, siguiendo el marco teórico de Godino y Batanero (1994).

Un punto a destacar en su estudio es que la mayoría de los libros analizados se basaban en el esquema teoría-práctica en la presentación de los contenidos, pues los ejemplos y ejercicios se introducen sólo una vez que se ha estudiado las definiciones y propiedades. Se daba también un énfasis excesivo a la teoría, mientras que el tipo de situaciones problemáticas presentadas es limitado, respecto a las potencialmente adecuadas para este nivel de enseñanza. Hay, por tanto, una falta de ejercicios interpretativos, ya que la mayoría de los ejercicios se reducen al cálculo; poco análisis experimental de los fenómenos aleatorios y la convergencia, es decir, no se presentaba la aproximación frecuencial a la probabilidad. Informa también de algunas confusiones, por ejemplo, entre los conceptos de variable aleatoria y variable estadística, o en la definición de un “suceso condicionado” (que no tiene contra partida en matemáticas).

Así mismo, indica que las actividades presentes sólo se manejan referidas a espacios muestrales finitos en contexto únicamente de juego, que la asignación de probabilidades predominante es mediante la regla de Laplace, y que la forma de presentación más usada es la verbal, en ocasiones con tablas de datos. Ortiz recomienda diversificar las actividades en los libros de texto, sin olvidar que los conceptos deben introducirse acordes al nivel de enseñanza y edades de los alumnos.

3.2.9.3.LA CORRELACIÓN Y REGRESIÓN EN LOS LIBROS DE TEXTO

Una primera investigación sobre el tema es la de Sánchez Cobo (1999), quien analiza 11 libros de texto de tercer curso de Bachillerato, nivel en que se enseñaba el tema. Para el análisis de la presentación teórica se tiene en cuenta los objetivos de los libros, la metodología usada en la exposición del tema, contenidos matemáticos presentados, número de ejercicios y ejemplos, y presencia de consideraciones históricas. Se ofrece una clasificación de definiciones y un análisis de las demostraciones, tanto desde el punto de vista de la función que realizan como de las componentes que la integran.

Sólo dos de los libros analizados presentan los objetivos; dos de ellos comienzan el tema con exposiciones teóricas y el resto con ejemplos. Existe una presentación basada en el esquema teoría-práctica, reforzada por la ubicación de los ejemplos con relación al concepto que ejemplifican. Las definiciones son principalmente de tipo instrumental, y apenas se presentan de tipo instrumental-relacional. Además, las demostraciones se presentan con una marcada función explicativa y de convicción, obviando la necesidad de desarrollar en nuestros alumnos la capacidad de argumentar de un modo lógico. Con todo ello, se pudiera "*transmitir una visión de las matemáticas como disciplina conformada por una colección de reglas y hechos que deben ser recordados y que se refieren sobre todo al cálculo*" (Sánchez Cobo, 1999, p. 287).

Analiza con detalle los contenidos expuestos, indicando que sólo tres libros incluyen la diferenciación entre dependencia funcional, aleatoria e independencia, y sólo uno aborda la covariación. La mayoría de los textos incluyen la correlación y la covarianza, pero hay cuatro que no diferencian entre correlación positiva y negativa. La mayoría de los textos incluye la diferenciación de las dos rectas de regresión.

Respecto al análisis de ejercicios, se estudia la distribución de las siguientes variables de tarea: contextos utilizados, contenido matemáticos, tipo de dependencia y valor absoluto del coeficiente de correlación. Entre los tipos de tarea se diferencia: cálculo, interpretación, representación gráfica, predicción, comprobación de propiedades, comparación de grados de asociación, y recogida y análisis de datos. Hay un fuerte sesgo en estos ejercicios y diagramas de dispersión hacia la asociación directa y de fuerte intensidad. Se destaca el aspecto del ajuste de la recta de regresión, olvidando la problemática de la predicción a partir de la misma. Asimismo, es escasa la discusión de los diferentes tipos de covariación, falta de contextualización en los ejercicios que, además, parecen pensados para ser resueltos con papel y lápiz y no mediante el uso de las nuevas tecnologías.

Sánchez Cobo (1999) concluye que es de gran importancia ofrecer a nuestros alumnos situaciones de aprendizaje que muestren la diversidad de tipos de covariación (dependencia, intensidad, signo), y que de algún modo contribuyan a eliminar las concepciones erróneas que manifiestan nuestros estudiantes. De ello se deduce el interés de que las tareas que se presenten a los alumnos sean dadas en contexto y con datos reales.

Como complemento a la investigación desarrollada por Sánchez Cobo y cols. (2000), Lavalle y cols. (2006) pretenden acercarse un poco más al estudio de la enseñanza de las nociones de correlación y regresión. Las autoras analizan el contenido, precisando las nociones "*que necesitan ser incorporados al tratamiento de la regresión y la correlación lineal para favorecer la comprensión*" (Lavalle y cols., 2006, p.387), así como los procedimientos con los que se vincula, y sus relaciones entre ellos. Estas nociones, junto con los procedimientos que implican cada una de ellas, son:

- *Variable estadística*, tanto unidimensional como bidimensional: variable explicativa (*X*) y variable explicada/respuesta (*Y*); como procedimientos se distinguen: Identificar las variables en estudio; Distinguir entre una variable respuesta y una variable explicativa; Identificar las unidades de medida; y Calcular la media y la desviación típica.

- *Datos* bidimensionales que se tratan para ser comparados, analizados e interpretados, es decir, pares ordenados que pueden ser representados mediante puntos del plano cartesiano. Los procedimientos asociados a su tratamiento son: Identificar los valores de las variables como pares ordenados; Graficar la nube de puntos/diagrama de dispersión; Identificar si hay relación entre las variables a partir del gráfico (existe o no relación, es lineal o no, etc.); y Observando el gráfico de dispersión de una relación lineal, indicar aproximadamente qué tipo y grado de relación lineal existe (directa, inversa)

- *Diagrama de dispersión o nube de puntos*, representación en un sistema cartesiano de los datos.

- *Coeficiente de correlación lineal*, el coeficiente de correlación lineal cuantifica la asociación o relación lineal de dos variables; los procedimientos son: Calcular la covarianza; Calcular el coeficiente de correlación; e Interpretar valores del coeficiente de correlación.

- *Recta de regresión*, para obtener una relación funcional entre las variables en la que se contempla la componente aleatoria y determinista del fenómeno de estudio. En particular, el propósito del análisis de regresión lineal, es obtener un ajuste de la variable respuesta/explicada (*Y*) mediante una función lineal. Los procedimientos asociados son: Verificar a partir del gráfico la conjetura de relación lineal entre las variables de estudio. Calcular los coeficientes de la recta de regresión. Graficar la recta; Interpretar los coeficientes de la recta; y Observando la nube de puntos y la recta, indicar aproximadamente el grado de relación lineal existente.

- *Estimación de los valores*, una vez hallada la recta de ajuste mínimo cuadrática, serán estimados mediante la sustitución en la recta de regresión obtenida, los correspondientes valores asociados en la variable explicativa (*X*). Si el valor de *X* que se sustituye se encuentra en el rango de valores observados de dicha variable, ocurre una interpolación; y si el valor de *X* se encuentra fuera de dicho rango, ocurre una extrapolación. Los procedimientos asociados son: Calcular valores estimados; Indicar gráficamente el error de estimación; e Interpretar el error como desviación.

Las autoras analizan siete libros de texto, centrándose en: el enfoque con que se presentan las nociones anteriores y su nivel de profundidad; si se deducen las fórmulas; el tipo de situaciones problemáticas; y si se utiliza ó se sugiere el uso de herramientas tecnológicas. Seis de los siete manuales analizados introducen las nociones mediante problemas:

> Muestran una intención explícita por desarrollar los conceptos, haciendo énfasis en la construcción crítica de las habilidades del pensamiento, no en el enfoque tradicional de la enseñanza de la estadística, donde se pone de manifiesto el interés por el manejo de fórmulas y ecuaciones (Lavalle y cols., 2006, pp.390-391).

Las situaciones problemáticas que proponen los textos analizados presentan mayor cantidad de actividades donde se utiliza la relación lineal directa que la inversa. Y en cuanto al uso de los ordenadores como herramienta de trabajo, ningún texto propone actividades para el uso de ordenadores. En cuanto al nivel de profundidad, se distinguen cinco niveles:

1. Se trabaja de modo intuitivo sin formalizar ningún concepto ni cálculo. La aproximación se realiza mediante gráficos y tablas de datos.
2. Se lleva a cabo un desarrollo profundo del tema, realzando los aspectos significativos para la comprensión
3. Igual profundidad que el nivel anterior, pero sólo utilizando ejemplos y fórmulas.
4. Desarrolla el concepto de correlación y de modo intuitivo la regresión lineal
5. Desarrollan los conceptos de correlación, regresión lineal y coeficiente de correlación con buena selección de problemas y enfocados a la comprensión. Este nivel se presenta en la mayoría de los textos analizados.

En cuanto a la deducción de fórmulas, sólo uno de los textos deduce la fórmula de cálculo de la covarianza aplicando propiedades de sumatoria, explicando que el método para obtener la pendiente y la ordenada en el origen de la recta de regresión tiene como premisa minimizar la expresión del error del modelo. La deducción de la fórmula de cálculo no se desarrolla, sino que el método se sustenta en un desarrollo gráfico. Este texto, obviamente se encuentra en el nivel 5 de profundidad.

3.3. INVESTIGACIONES SOBRE EL CONOCIMIENTO DEL PROFESOR

3.3.1. INTRODUCCIÓN

En el Capítulo 2 se analizaron diferentes marcos teóricos o conceptuales, que intentan describir el conocimiento del profesor en la enseñanza de las matemáticas, y en la estadística en particular. Encontramos investigaciones donde se trata de evaluar con profesores o futuros profesores algunas categorías de estos conocimientos, con la finalidad de obtener criterios para su desarrollo (Llinares, 1998). En esta sección, analizamos las más relevantes en relación con nuestro trabajo. Apenas existen sobre el tema específico de correlación y regresión; pero podemos citar las que estudian el conocimiento del profesor en otros temas de estadística cuyo conocimiento se requiere para la enseñanza de la correlación y regresión.

3.3.2. INVESTIGACIONES SOBRE LOS CONOCIMIENTOS ESTADÍSTICOS DEL PROFESOR

Un primer grupo de trabajos se orienta a la evaluación de los conocimientos estadísticos (común o ampliado) de los profesores o futuros profesores. El interés de analizar los conocimientos matemáticos de los profesores se debe a que muchas tareas cotidianas del profesor en la clase de matemáticas, tales como "*indagar lo que los estudiantes conocen, elegir y manejar representaciones de las ideas matemáticas,* *seleccionar y modificar los libros de texto, decidir entre modos posibles de acción"* (Ball, Lubienski y Mewborn, 2001, p. 453) dependen de su razonamiento y conocimientos matemáticos. Por tanto, asegurar un conocimiento matemático es el primer paso para desarrollar su conocimiento didáctico-matemático.

En el caso de la estadística, la investigación que resumimos a continuación se ha realizado básicamente con futuros profesores de educación primaria y revela una variedad de dificultades y errores conceptuales entre los mismos. Clasificamos esta investigación en varios apartados, dependiendo del objeto estadístico considerado.

3.3.2.1.GRÁFICOS ESTADÍSTICOS

Es claro que la enseñanza de la correlación y regresión requiere unas habilidades mínimas de construcción e interpretación de gráficos, siendo éstos un vehículo importante en la enseñanza del tema.

La investigación sobre el conocimiento de los gráficos estadísticos por profesores es relativamente amplia y se describe con detalle en la tesis de Arteaga (2011). Como ejemplos, citamos los trabajos de Bruno y Espinel (2005) quienes describen los errores de los futuros profesores de educación primaria al construir o interpretar histogramas y polígonos de frecuencia. Espinel (2007) también compara los anteriores resultados con los de estudiantes universitarios americanos encontrando mayor dificultad en los primeros, sobre todo al predecir la forma de un gráfico a partir de la descripción verbal de variables conocidas, o al leer los histogramas.

Monteiro y Ainley (2006, 2007) proponen a un grupo de futuros profesores en Brasil interpretar una serie de gráficos tomados de la prensa. Los resultados obtenidos en una de las preguntas sugieren que muchos participantes no tenían conocimientos matemáticos suficientes para llevar a cabo dicha lectura, pues no tuvieron formación específica en la lectura de gráficos estadísticos durante sus estudios en la universidad

Sin embargo, se observó que la mayoría de los futuros profesores mostraron habilidades para pensar críticamente sobre los datos representados en los gráficos, y que al interpretarlos, ponían en juego distintos tipos de conocimientos. Los autores concluyen que, para realizar una lectura completa de un gráfico, se necesita conectar las relaciones cuantitativas expresadas en el mismo, con los conocimientos estadísticos y del contexto social de donde se han tomado los datos que aparecen representados.

Gitirana, Guimarães, Magina y Cazorla (2008) proponen un cuestionario sobre lectura, interpretación y construcción de gráficos y tablas a 244 futuros profesores y 43 profesores de educación primaria en Brasil. Las autoras indican que, a medida que avanza en sus estudios, el participante mejora sus respuestas. No obstante, se encontraron resultados contradictorios; por ejemplo, los alumnos de los primeros cursos construyen mejor un gráfico de barras que los estudiantes que acababan sus estudios. Asimismo, estos últimos presentaron una tendencia a construir gráficos utilizando ejes cartesianos con pares ordenados, no así los alumnos de los primeros años.

Arteaga (2011) evalúa los gráficos construidos por 200 futuros profesores de educación primaria al trabajar con un proyecto abierto de análisis de datos. Define un nivel de complejidad, según el gráfico es más completo, comenzando desde la representación de datos aislados (nivel 1), representación de un listado de datos, sin ordenar y sin formar la distribución (nivel 2), formar y representar la distribución de frecuencias (nivel 3), o comparar dos distribuciones de frecuencias en un mismo gráfico (nivel 4).

Sus resultados sugieren que la mayoría de los futuros profesores construyen gráficos de complejidad adecuada para resolver el problema planteado. No obstante, se observan errores de escala, o de elección de un gráfico inapropiado en la resolución de la tarea. Por otro lado, no todos los participantes son capaces de leer correctamente los gráficos que han construido, o si lo hacen, el nivel de lectura es muy superficial, limitándose a una lectura literal, o bien a realizar comparaciones en el gráfico. Pero no llegan a obtener una conclusión completa sobre la pregunta de investigación planteada en el proyecto.

3.3.2.2.MEDIDAS DE POSICIÓN CENTRAL

Otras investigaciones identifican dificultades de futuros profesores de Educación Primaria con las medidas de posición central. Por ejemplo, Batanero, Godino y Navas (1997) analizan las respuestas de 273 futuros profesores a un cuestionario escrito, que incluye cuatro ítems sobre distintos aspectos interpretativos de la media aritmética. Sus resultados muestran la existencia de errores conceptuales, y dificultad de aplicación práctica de los conocimientos sobre los promedios. Entre los más frecuentes, encuentran que el 40% de la muestra no desecha los valores atípicos para calcular la media, el 30% sobrevalora la dispersión de los datos y el 60% no conoce la relación entre media, mediana y moda en distribuciones simétricas. Estas dificultades se reproducen en otro trabajo con 387 futuros profesores en la Universidad de Lleida (Estrada, Batanero y Fortuny, 2004). Además, en este estudio, aproximadamente la mitad de los futuros profesores confundieron correlación y causalidad; y sólo el 28% invirtieron adecuadamente el algoritmo de la media.

Groth y Bergner (2006) investigaron sobre la comprensión de los conceptos de media, mediana y moda por parte de 46 futuros profesores de Educación primaria y secundaria. En su estudio, los participantes tuvieron que explicar las diferencias y similitudes entre los conceptos estadísticos de media, mediana y moda. Los autores distinguieron entre 4 diferentes categorías sobre la comprensión de dichos conceptos estadísticos por parte de los futuros profesores: (a) uniestructural/concreto simbólico; (b) multiestructural/concreto simbólico; (c) relacional/concreto simbólico; (d) abstracto.

Hubo 8 futuros profesores que trabajaron en el nivel uniestructural/ concreto simbólico, debido a que sus respuestas sólo usaban la definición literal de los conceptos de media, mediana y moda para obtener similitudes y diferencias entre los distintos promedios. Veintiún futuros profesores respondieron dentro del nivel multiestructural/concreto simbólico, ya que sus respuestas mostraban que los promedios que tenían que estudiar representaban más un objeto matemático que un simple procedimiento de cálculo. Trece de los participantes mostraron en sus respuestas el nivel relacional/concreto simbólico, pues sus respuestas iban más allá de procedimientos para encontrar determinadas medidas, para así concluir a cerca de características “típicas” sobre un conjunto particular de datos. Finalmente, 3 futuros profesores trabajaron dentro del nivel superior, o nivel abstracto, ya que las respuestas que proporcionaron iban más allá del conocimiento procedimental, e incluían una discusión sobre cuándo una determinada medida de posición central sería más útil para un cierto conjunto de datos.

3.3.2.3.DISPERSIÓN

La importancia del concepto de dispersión es que está incluida en el de variación, considerada por Wild y Pfannkuch (1999) como el núcleo de su modelo sobre razonamiento estadístico. En nuestro estudio, permite comprender la intensidad de la correlación y la bondad del ajuste. A pesar de ello, la investigación de Silva y Coutinho (2008) muestra que el razonamiento predominante de los profesores brasileños sobre la variación es verbal, lo que les impide enseñar a sus estudiantes el significado de medidas tales como la desviación típica, limitándose a la enseñanza de algoritmos.

Estos investigadores pidieron a los participantes en su estudio crear una distribución a partir de un promedio dado sobre edad de un grupo de personas. Después de organizar los datos en una tabla de frecuencias y representarlos en un histograma, se pidió a los profesores pensar sobre distintos modos de representar el conjunto de las edades. Los autores analizaron el razonamiento de los profesores cuando estos analizaron la distribución de las edades. Ninguno fue capaz de integrar el razonamiento sobre la media con el significado de la desviación respecto a la media, ni estimar la frecuencia en el intervalo de *k* desviaciones típicas desde la media.

Makar y Confrey (2005) realizaron entrevistas a 17 futuros profesores de Educación Secundaria de matemáticas y ciencias, para analizar cómo estos usaban la dispersión cuando comparaban dos distribuciones El estudio mostró cómo los futuros profesores expresaban importantes ideas sobre este concepto, a través de un lenguaje tanto estándar como no estándar. Algunos ejemplos de expresiones estándar en estadística, de las usadas por los futuros profesores, fueron: proporción, media, máximo/mínimo, tamaño muestral, valores atípicos, rango, forma y desviación típica. En las entrevistas emergieron dos categorías de términos no estándar para expresar ideas sobre la variación, que correspondían a ideas intuitivas sobre dispersión (agrupado, disperso) y distribución (triadas, agrupaciones modales). El uso de estos términos también creció a lo largo de las entrevistas.

3.3.2.4.DISTRIBUCIÓN

En el estudio de la correlación y regresión, los estudiantes analizan distribuciones bivariantes, por lo que es importante comprender la idea de distribución. Canada (2008) analizó el razonamiento de estudiantes de secundaria y futuros profesores cuando comparan conjuntos de datos con la misma media y diferente dispersión. El autor indica que la comprensión de la idea de distribución emergerá cuando el sujeto sea capaz de razonar simultáneamente respecto a los promedios y la dispersión para realizar dicha comparación. A pesar de ello, aunque el grupo de futuros profesores tuvo mejor rendimiento que el de los estudiantes en la tarea propuesta, todavía un 35% de los futuros profesores pensaba que dos conjuntos de datos con la misma media eran iguales, aunque la dispersión fuese muy diferente. Por tanto, su comprensión de la idea de distribución no era completa.

Mickelson y Heaton (2004) estudiaron cómo una profesora de tercer curso de primaria razonaba con la idea de distribución, al aplicar sus conocimientos en unas clases donde enseñaba a sus alumnos mediante una investigación estadística. Dependiendo de los distintos contextos, la profesora aplicaba razonamientos estadísticos sobre distribuciones, tanto de manera adecuada, como demasiado ingenua. Parecía también que era menos capaz de mostrar un conocimiento profundo sobre la idea de distribución, cuando la investigación llevada a cabo en la clase implicaba el trabajo con tareas abiertas, basadas en datos reales, que al trabajar con tareas estándares tomadas de libros de texto.

Pfannkuch (2006) estudió el razonamiento llevado a cabo por un profesor de secundaria, cuando este realizaba inferencias informales al comparar gráficos de cajas e interpretar las distribuciones representadas por dichos gráficos. La investigación permitió desarrollar un modelo compuesto de diez elementos diferenciados, que categorizan el pensamiento del profesor, y permiten describir la naturaleza y el tipo de razonamiento inferencial informal cuando los estudiantes trabajan comparando distribuciones.

En resumen, todas estas investigaciones apuntan a la falta de conocimiento sobre conceptos estadísticos básicos, aunque han sido llevada a cabo generalmente con futuros profesores o profesores de educación primaria. También muestran la necesidad de potenciar los contenidos estadísticos en la formación de profesores. Nuestra investigación trata de analizar el conocimiento matemático de futuros profesores de secundaria, en torno a la correlación y regresión, que ha sido escasamente investigado.

3.3.3. CONOCIMIENTO DIDÁCTICO DE LA ESTADÍSTICA

En esta sección, analizamos las escasas investigaciones relacionadas con el conocimiento didáctico de los profesores para enseñar estadística, que, de acuerdo a Batanero, Godino y Roa (2004), tiene características específicas. Los clasificamos de acuerdo a las facetas del modelo de Godino (2009), aunque no todas han sido investigadas.

Faceta epistémica

Respecto a la faceta epistémica del conocimiento didáctico (conocimiento especializado del contenido estadístico en el modelo MKT de Ball y colaboradores), algunas investigaciones estudian a profesores en ejercicio durante su enseñanza, observándolos a lo largo de un periodo de tiempo, y deduciendo su conocimiento a partir de esta observación, complementada con entrevistas o a través de otros medios. Las pocas investigaciones al respecto indican que este conocimiento es también escaso.

Así, Jacobbe (2008) entrevista a tres profesores a lo largo de un periodo dilatado, observando sus clases y las tareas propuestas a los alumnos durante las mismas. A pesar de tratarse de profesores muy motivados, y tener una amplia experiencia de enseñanza, el estudio mostró que no poseían suficiente conocimiento de la mediana para enseñarla. Por ejemplo, algunos hacían errores de cálculo elementales, como no ordenar los datos para calcular la mediana; otros no sabían decidir en qué situaciones es necesario calcular la media o la mediana, ni eran capaces de describir verbalmente la diferencia entre los dos conceptos.

Stohl (2005) por su parte, analiza el conocimiento de los profesores sobre la enseñanza de la estadística por medio de proyectos. En su estudio, los profesores observados proponían tareas que no permitían implicar a los estudiantes en la investigación o para profundizar en sus razonamientos. Especialmente, se perdieron estas oportunidades en las fases de análisis de los datos, e interpretación de los resultados obtenidos.

Facetas cognitiva e interaccional

Cai y Gorowara (2002) llevaron a cabo una investigación en la que participaron 12 profesores noveles y 11 profesores experimentados en la práctica docente. En dicha investigación se estudiaron las concepciones y la construcción de representaciones para la enseñanza del concepto de media, por parte de los dos grupos de profesores. Se recogieron datos por medio de las siguientes tareas, que tuvieron que realizar los profesores: (1) planificar una lección centrada en el concepto de media; (2) tratar los posibles modos en que los estudiantes de 12- 13 años pudieran responder a una serie de preguntas que estaban relacionadas con la media; (3) evaluar las respuestas de los estudiantes a algunas de las preguntas anteriores.

Los dos grupos de profesores (con o sin experiencia) fueron capaces de responder a las preguntas que se les propuso en el apartado (2) anteriormente citado. Del total de la muestra, 8 de los profesores experimentados y 2 de los profesores noveles fueron capaces de proporcionar múltiples estrategias para resolver los problemas. Dentro del conjunto de profesores noveles, sólo uno de ellos fue capaz de resolver las tareas sin usar un enfoque algorítmico.

Aunque se pidió a los profesores que indicaran los modos en los que los estudiantes podrían resolver las preguntas, los profesores sin experiencia no trataron los posibles errores conceptuales que podrían tener los estudiantes, mientras que los profesores experimentados sí que tuvieron en cuenta los posibles errores conceptuales al resolver casi todas las tareas. El resto de tareas, que estaban relacionadas con la planificación de una lección y la evaluación de las respuestas de los estudiantes, proporcionaron resultados similares en ambos grupos. A modo de síntesis, los resultados del estudio mostraron que según los profesores ganan experiencia docente y adoptan mayores roles de liderazgo, son más capaces de reflexionar sobre los posibles errores conceptuales de sus estudiantes para mejorar su propia comprensión, así como para mejorar su habilidad al hacer frente a los errores mostrados por sus estudiantes.

Watson (2001) examinó el conocimiento de profesores de Educación Primaria y Secundaria sobre las dificultades de sus alumnos con la probabilidad y la estadística. Sólo cuatro de ellos mencionaron haber encontrado dificultades, sin especificarlas, mientras que trece indicaron dificultades concretas en aspectos procedimentales o conceptuales. Watson mostró que algunos profesores tenían una apreciación muy limitada de cómo sus alumnos podrían contestar preguntas tales como encontrar un error en un gráfico o identificar un error en un informe relacionado con la media en el muestreo de una población. Asimismo, los profesores, a menudo se limitaban a sugerir cómo ellos podrían utilizar las respuestas de sus alumnos en sus clases.

Así mismo, Calligham y Watson (2011) diseñaron un cuestionario para medir el conocimiento matemático para la enseñanza de profesores de estadística, donde pedían a los profesores predecir respuestas de los alumnos, como usar los ítems en sus clases, o cómo resolver las posibles dificultades de los alumnos. En otras preguntas les pidieron clasificar respuestas de niños según niveles de razonamiento.

Utilizando el cuestionario, entrevistaron a 42 profesores australianos de primaria y secundaria que cursaban un programa de formación en estadística. El análisis con métodos estadísticos mostró tres perfiles de profesores según su habilidad en la enseñanza de la estadística. El nivel bajo, con 14 profesores, se caracterizaba porque podía empezar a predecir respuestas de los estudiantes y usar materiales en clase. El nivel medio, con 19 profesores, se caracterizaba por la capacidad de sugerir respuestas correctas e incorrectas para algunos temas y tratar problemas de razonamiento proporcional, involucrando contenido matemático. El nivel alto, con 9 profesores, se caracterizó por un buen desempeño para dirigir y orientar la matemática involucrada en problemas de razonamiento proporcional, y sugerir respuestas correctas e incorrectas para ítems de dificultad moderada, aunque su desempeño no fue bueno con ítems más complejos.

Faceta mediacional

Es también patente la dificultad de algunos profesores al trabajar con proyectos y problemas abiertos en matemáticas (Jaworski, 1994; Ponte, 2001). En el caso de la estadística, un tema fundamental es que los proyectos se usen para enseñar razonamiento estadístico y no cálculos rutinarios. Por ello, es importante ver si los profesores son capaces de reconocer qué conceptos pueden ser estudiados a partir de un proyecto, o un conjunto de datos dado, e implementar una enseñanza efectiva en una clase utilizando proyectos estadísticos.

En la investigación de Chick y Pierce (2008), los profesores participantes no hicieron un uso adecuado de los datos y proyectos al planificar sus lecciones. A pesar de la riqueza de conceptos de la situación didáctica planteada, fallaron en sacar a la luz conceptos latentes. Por el contrario, se limitaron a pedir cálculos o nuevos gráficos, con pocas actividades de interpretación.

Burgess (2008) se interesó en explorar los conocimientos del profesor para enseñar estadística usando su modelo de conocimiento descrito anteriormente. Para ello, comparó el conocimiento de dos profesores de educación primaria, en una experiencia de enseñanza que usaba varios conjuntos de datos multivariados. Los profesores desarrollaron su propia secuencia de enseñanza del tema, cada una de ellas grabada en video. El autor analiza el conocimiento puesto en práctica (o que no pudo poner en práctica) por cada profesor en las diferentes categorías de su marco teórico, mostrando, con episodios, los diferentes tipos de conocimiento.

En algunos casos, las observaciones indicaron oportunidades perdidas de usar uno de estos tipos de conocimiento. Así, mientras que uno de los profesores sólo perdió cuatro oportunidades cuando podía haber usado el conocimiento, el otro perdió catorce. Ejemplos de estas oportunidades perdidas serían no verificar la corrección de la interpretación de un ejercicio por parte de los alumnos; perder oportunidad de usar conocimiento relacionado con la transnumeración; o bien, las relacionadas con el conocimiento del contenido y los estudiantes.

Trabajos globales

El trabajo más completo que hemos encontrado es llevado a cabo por Arteaga (2011), quien contempla las diferentes facetas del conocimiento didáctico de nuestro modelo de investigación. El autor propuso a 207 futuros profesores de Educación Primara un proyecto abierto en el que se pedía recoger y analizar datos para llegar a una conclusión. Como continuación de la actividad, una parte de los participantes (109) analizaron los diferentes componentes y descriptores de la idoneidad didáctica del proyecto, utilizando una guía de análisis adaptada de Godino (2009).

Sus resultados mostraron poca capacidad de análisis en los futuros profesores, variando según la faceta. Especialmente fueron pobres los resultados en el análisis epistémico, lo que el autor explica según el poco conocimiento matemático que mostraron sobre los contenidos del proyecto. En el Estudio 3 que realizamos en nuestra investigación, también aplicamos la guía de análisis de la idoneidad didáctica, administrada a un grupo de participantes del Estudio 2, y compararemos nuestros resultados con los de Arteaga.

En resumen, las pocas investigaciones sobre las diversas facetas del conocimiento didáctico del profesor para enseñar estadística son incompletas, y dejan puntos por investigar. En nuestro caso, nos centramos en un tema poco tratado, y con profesores de secundaria, para lo cual, abordamos las diferentes facetas del conocimiento didáctico en una misma investigación.

3.3.4. COMPRENSIÓN DE LA CORRELACIÓN Y REGRESIÓN POR LOS PROFESORES

Para finalizar el resumen de las investigaciones previas con profesores, queremos resaltar las realizadas en el tema objeto de nuestro trabajo, donde son prácticamente inexistentes los estudios con profesores o futuros profesores. Las excepciones más notables son las siguientes:

- El trabajo de Estepa y sus colaboradores (Estepa, 1994; Batanero, Estepa, y Godino, 1997; Batanero, Godino y Estepa, 1998), en cuyo experimento de enseñanza participó una muestra pequeña de futuros profesores de educación primaria. Puesto que se ha descrito con detalle este trabajo, no se insiste más en el mismo en esta sección. Este trabajo es posiblemente el más completo sobre el aprendizaje de la correlación y regresión, no sólo con futuros profesores, sino en general. Remitimos al lector a las secciones anteriores para un análisis detallado.
- El trabajo de Casey (2008; 2010) dirigido a evaluar el conocimiento de profesores en ejercicio sobre la correlación. Este trabajo también se describió con detalle en el Capítulo 2, y estuvo orientado a presentar un marco teórico sobre los conocimientos requeridos en la enseñanza del tema. Nuestro trabajo difiere de la citada autora en que, por un lado, aquel incluye el conocimiento inferencial, pero por otro, su significado de referencia es mucho más restringido que el nuestro. Además, nos orientamos más hacia la evaluación y desarrollo del conocimiento didáctico que a la construcción de un modelo teórico.

Finalmente, resaltamos el trabajo de Quintas, Oliveira y Freitas (2013), quienes observan el trabajo de dos profesoras de secundaria con amplia experiencia (15 y 20 años) mientras explican la estadística bivariante en un curso de secundaria. Se basan tanto en los modelos descritos de Batanero, Godino y Roa (2004), como en el del pensamiento estadístico de Wild y Pfannkuch (1999) y el de Burgess (2011). Una de ellas trabaja con la calculadora gráfica, y otra con un software didáctico disponible en un portal de internet denominado “*escola virtual*”.

Siguiendo a Garfield y Ben-Zvi (2008), el razonamiento con datos covariados (quieren decir bivariados) debe implicar la comprensión de las ideas de estructura (modo en que denominan la tendencia) y fuerza (refiriéndose a la intensidad), así como la comprensión del papel que desempeña la relación entre las variables en la predicción. Estos autores no diferencian explícitamente la dependencia funcional, independencia y dependencia aleatoria, pues únicamente hablan de "dependencia bivariada", que podría englobar a las anteriores.

Analizan, dentro del conocimiento didáctico, el conocimiento de la estadística, del currículo, del aprendizaje y de la enseñanza. Dentro del conocimiento estadístico incluyen las ideas fundamentales estadísticas definidas por Burrill y Biehler (2011), la comprensión de la problemática detrás de estas ideas, el uso de la transnumeración, la percepción de la variación, el trabajo con modelos estadísticos (Wild y Pfannkuch, 1999) y las competencias descritas por Godino, Batanero y Flores (1999).

La primera profesora desarrolla el tema en 3 sesiones de 90 minutos, utilizando conjuntos de datos sencillos y asequibles a los alumnos; y apoya su enseñanza en la resolución de tareas y el uso de la calculadora gráfica. En una primera tarea, a partir de datos de equipos de baloncesto, pide dibujar a mano y con calculadora el diagrama de dispersión. Pasa a la interpretación de la tendencia, insistiendo en ésta, aún cuando una niña observa que hay puntos que no la siguen. Recuerda a los alumnos sus conocimientos sobre la recta, y trata de forzar que ellos mismos aproximen "a ojo" una recta que se acerque a los puntos, incluso que encuentren su expresión algebraica, aunque no da criterios claros de qué significa tal aproximación.

Finaliza su enseñanza pidiendo a los alumnos comparar esta recta con la de regresión, calculada con la calculadora, pero no queda claro a los alumnos por qué las rectas son diferentes; aunque la profesora insiste que la de regresión ha de pasar por el centro de gravedad. Esto es comprobado por los alumnos, viendo que el centro cumple la ecuación de la recta proporcionada por la calculadora. Ayuda a los alumnos a comprender la relación entre dispersión e intensidad, y crecimiento de la nube y signo de la correlación.

La segunda profesora comienza por la presentación de conceptos de correlación, centro de gravedad y recta de regresión, trabaja las representaciones de datos bivariados, así como los pasos para obtener los cálculos del coeficiente de correlación y la recta de regresión, finalizando con la aplicación de lo presentado en tareas resueltas con la calculadora gráfica. Analiza el significado del signo y valor del coeficiente de correlación, en relación con el tipo de correlación. No explica la relación entre el centro de gravedad y la recta de regresión.

Uno de los puntos a destacar en este trabajo, que pasa desapercibido por las autoras del mismo, es que en él se refleja la falta de conocimiento estadístico de las profesoras participantes, a pesar de su amplia experiencia. En ambos casos, las profesoras consideran una única recta de regresión, y piden a los chicos despejar el valor de X, dado un valor Y, a partir de la recta de regresión de Y sobre X. Asimismo, piden hacer predicciones, en un rango mucho más amplio que el considerado por los datos recogidos, extrapolación que contaría, sin que las profesoras lo noten, con un gran error de estimación.

Estos errores, que quizás podrían verse como parte del conocimiento ampliado de la correlación y regresión, muestran la necesidad de este conocimiento ampliado, ya que la investigación muestra claramente cómo las dos profesoras insisten en su enseñanza, transmitiendo estos errores a sus estudiantes.

3.4. CONCLUSIONES SOBRE LAS INVESTIGACIONES PREVIAS

Para finalizar el estado de la cuestión, recogemos las conclusiones más importantes extraídas de su estudio. Observamos, en primer lugar, que el estudio de la correlación y regresión, e incluso el más amplio de covariación (que incluye también las tablas de contingencia), no ha tenido mucha relevancia en didáctica de la matemática, pues el número de investigaciones, así como los temas tratados, son limitados, y no agotan toda la problemática relacionadas. Sin embargo, encontramos una amplia literatura en psicología, donde se describe, en general, el razonamiento covariacional, como una competencia fundamental para la toma de decisiones.

El análisis de las estrategias de los estudiantes en la detección de la correlación a partir de diversas representaciones muestra, por un lado, la existencia de numerosas estrategias incorrectas; por otro, la de concepciones incorrectas y sesgos en la detección de la correlación. Algunas de estas concepciones y sesgos son resistentes a experimentos de enseñanza tradicionales, e incluso a otros basados en el uso de tecnología, por lo que varios autores realizan sugerencias para una mejor enseñanza del tema.

Son pocos los estudios específicamente centrados en la presentación del tema en los libros de texto, y ninguno de los encontrados hace un uso sistemático del enfoque ontosemiótico. De ello, deducimos el interés de realizar un trabajo en que se profundice sobre los tipos de problemas, lenguaje, conceptos, propiedades, procedimientos y tipos de argumentos en el estudio de la correlación y regresión, que es el objeto del Estudio 1, descrito en el Capítulo 4, donde además se analiza el uso de la tecnología y la presencia de posibles conflictos semióticos.

Otra segunda conclusión es la poca atención que en la investigación didáctica se ha dado al estudio de las actitudes, creencias y conocimientos de los profesores de estadística, en comparación con la abundante literatura existente para el caso de la matemática. Solo con el reciente *Joint ICMI/IASE Study "Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education"* (Batanero, Burrill y Reading, 2011) ha surgido el interés por la investigación en este tema y el estudio de algunos modelos sobre el conocimiento del profesor en el campo de la estadística.

Respecto al estudio de las actitudes, creencias y conocimientos estadísticos y didácticos, los trabajos existentes se centran, en su mayor parte, en futuros profesores de educación primaria, e indican unos conocimientos deficientes. Por ello, sería interesante continuar el estudio de las actitudes y creencias de los profesores de secundaria, junto a sus conocimientos estadísticos y didácticos y con el diseño de planes para su mejora.

Este interés es todavía mayor en relación a los temas de correlación y regresión, donde apenas se encuentran trabajos, aislados e incompletos, que tratamos de completar en nuestra investigación por medio de los Estudios 2 y 3, el primero de los cuáles aborda la evaluación y desarrollo del conocimiento estadístico, y el segundo del conocimiento didáctico, en sus diferentes facetas.

CAPÍTULO 4.

LA CORRELACIÓN Y REGRESIÓN EN LOS LIBROS DE TEXTO

4.1. INTRODUCCIÓN

En este capítulo se presenta el Estudio 1, en el que analizamos la presentación de la correlación y regresión en una muestra de libros de texto de Matemáticas, dirigidos a estudiantes de primer curso de Bachillerato en las dos modalidades en que se incluye este tema: *Humanidades y Ciencias Sociales* y *Ciencia y Tecnología* (MEC, 2007b). Se estudian con detalle los tipos de objetos matemáticos definidos en el enfoque ontosemiótico, identificando las principales variables que influyen en ellos y se compara los libros de una misma modalidad de Bachillerato y entre modalidades. Asimismo, se analiza el uso de la tecnología y se identifican algunos conflictos semióticos potenciales.

Con ello, junto al análisis curricular desarrollado en el Capítulo 1, se determina el significado institucional pretendido de la correlación y regresión para este nivel educativo, que será la base de la construcción de las actividades e instrumentos de evaluación utilizados en los Estudios 2 y 3. Se relacionan, asimismo, los resultados obtenidos con los antecedentes (Capítulo 3), donde el único trabajo sobre textos españoles (Sánchez Cobo, 1999) se realizó hace 15 años, y es anterior a la normativa vigente. Nuestro estudio complementa y extiende el anterior, analizando en profundidad cada tipo de objeto matemático considerado en nuestro marco teórico.

4.2. OBJETIVOS DEL ANÁLISIS DE LIBROS DE TEXTO

Ya en el Capítulo 3 se justificó con detalle la importancia del libro de texto, como recurso didáctico en la enseñanza y aprendizaje de las matemáticas. También se analizaron las investigaciones sobre libros de texto en probabilidad y estadística, y en particular, los antecedentes directos del análisis que se realiza en este capítulo, que son los estudios de Sánchez Cobo (1999) y Lavalle, Micheli y Rubio (2006). Conscientes de esta importancia, en el Capítulo 1 se presentó el siguiente objetivo general del estudio que presentamos en este capítulo.

O1. Realizar un análisis detallado de la presentación de la correlación y regresión en los libros de texto de Bachillerato, con la finalidad de caracterizar el significado institucional de referencia en las dos modalidades en que se contempla su estudio.

Este objetivo se justifica, porque una adecuada idoneidad epistémica del libro de texto requiere una representatividad en el mismo de los diferentes objetos matemáticos que conforman el significado institucional de referencia del tema, que se describió brevemente en el Capítulo 1. Además, servirá para identificar el *significado institucional pretendido* en el Bachillerato, y puede ser de utilidad tanto para identificar puntos a mejorar en los textos, como para la elaboración de futuros instrumentos de evaluación del aprendizaje en relación al tema. Asimismo, será la base de elaboración de actividades e instrumentos de evaluación en los Estudios 2 y 3, sobre conocimientos de los futuros profesores, que se describen en los capítulos 5 y 6.

Dentro de este objetivo general podemos concretar, para el Estudio 1, los siguientes objetivos específicos:

O1.1. Caracterizar los campos de problemas sobre correlación y regresión presentados sobre el tema en los textos de Bachillerato.

Las situaciones-problema son un elemento fundamental del proceso de enseñanza y aprendizaje de un objeto matemático en el enfoque ontosemiótico, por lo que las analizamos con detalle, determinando los contextos empleados, el tipo e intensidad de dependencia presentada, la actividad que se plantea respecto a los mismos, y si se utiliza la tecnología para su resolución. Todas estas variables se identificaron, en estudios previos, como potencialmente influyentes en el aprendizaje del estudiante (Sánchez Cobo, 1999).

O1.2. Caracterizar la diversidad de lenguaje matemático utilizado en los textos para presentar la correlación y regresión, y los procesos de traducción entre los mismos.

La importancia de este objetivo se deduce del papel dado a la diversidad de representaciones de los objetos matemáticos por Duval (1993) y la doble función, representacional e instrumental, del lenguaje matemático (Godino, Batanero y Font, 2007). Se analizan los términos, expresiones simbólicas, y representaciones tabulares y gráficas, mostrando su uso en el tema. Asimismo, se deduce del uso inapropiado del lenguaje en algunos textos en trabajos previos sobre otros temas, por ejemplo por Ortiz (1999).

O1.3. Caracterizar los conceptos y propiedades relacionados con la correlación y regresión en los textos.

De igual modo se analizan los conceptos y propiedades, que categorizamos según se refieran a la distribución bidimensional, la dependencia estadística, la covarianza, la correlación y la regresión. Para cada concepto se analiza el tipo de definición, lugar y uso en el tema, y propiedades asignadas, así como las relaciones entre conceptos.

O1.4. Caracterizar los procedimientos empleados en el estudio de la correlación y regresión.

En muchas ocasiones, los algoritmos, operaciones y/o técnicas de cálculo utilizados en la resolución de una tarea se llegan a automatizar, y se hacen específicos del tipo de problema que resuelven, constituyéndose en sí mismos objetos de enseñanza. Por ello, se establece una conexión entre los tipos de procedimientos que se emplean en el tema, y las situaciones problema a las que dan respuesta.

O1.5. Determinar los tipos de argumentación empleados para justificar propiedades, procedimientos o soluciones a los problemas.

Los argumentos que se emplean en el tema permiten validar y explicar el por qué de un resultado o propiedad, o incluso el cómo llegar a nuevos resultados y comunicar los mismos, por lo que su uso es esencial en Bachillerato (MEC, 2007b). Analizamos las demostraciones formales deductivas, características de las matemáticas, que se incluyen en el tema, junto a todas aquellas justificaciones que, aunque informales y no deductivas, aportan rasgos característicos y recursos expresivos característicos del tema, utilizando algunas de las categorías establecidas por Recio (1999).

O1.6. Analizar el uso que se hace de la tecnología.

Analizamos, además, el uso que se hace de la tecnología en el tema, considerando para nuestro análisis el contenido del CD que acompaña a algunos libros de texto, así como los enlaces a páginas web que algunos textos incluyen en el tema. Se trata de mostrar y clasificar aquellas técnicas de resolución que, mediante el uso de la tecnología, se presenta en los diferentes textos; pues aportan elementos de significado a un mismo objeto matemático sobre una misma situación problema (Godino, 2002).

O1.7. Identificar los principales conflictos semióticos en el tema.

En los diferentes análisis realizados, se encuentran algunas asignaciones imprecisas de significado a determinados objetos matemáticos, que pudieran provocar un conflicto semiótico en el estudiante si el profesor no está atento al uso que de ellos se pueda llevar a cabo. Se presenta una clasificación de los mismos, según el objeto matemático que se trate.

4.3. HIPÓTESIS DEL ESTUDIO 1

Nuestra hipótesis general es que *los libros de texto presentarán un tratamiento diferenciado del tema, dependiendo de la modalidad de Bachillerato a la que se dirigen*, pues así se sugiere en el Decreto de Enseñanzas Mínimas (MEC, 2007b).

Así, en este documento, cuando se describe la asignatura de *Matemáticas aplicadas a las Ciencias Sociales I* se indica el menor peso en esta asignatura a la abstracción simbólica y la formalización, con respecto a la asignatura de *Matemáticas 1* (dirigida a la modalidad de *Ciencia y Tecnología*), junto a la importancia del apoyo de la tecnología para facilitar el trabajo con las matemáticas. Asimismo, se diferencia, tanto el uso de fórmulas, como la importancia dada a la demostración.

También se esperan diferencias, en cada especialidad dependiendo de la editorial, pues estas diferencias se han señalado en estudios anteriores como los de Ortiz (1999) o Sánchez Cobo (1999). Esta hipótesis general se divide en cuatro específicas:

H1. Se prevé mayor formalización en los textos dirigidos a la modalidad de Bachillerato de Ciencia y Tecnología.

H2. Se espera que los contextos utilizados en unos y otros textos favorezcan la apreciación de las aplicaciones de la correlación y regresión en ambas modalidades (Humanidades y Ciencias Sociales y Ciencia y Tecnología).

H3. Por otro lado, *se espera una variabilidad en la presentación, dentro de la misma modalidad, de diferentes editoriales,* en consonancia con lo observado en el estudio de Sánchez Cobo (1999), y en otros estudios sobre libros de texto.

H4. Asimismo, *podrían presentarse algunos sesgos en cuanto a las características de la correlación (intensidad, signo o tipo) en estos libros,* pues ya se ha descrito anteriormente la presencia de conflictos semióticos respecto a otros conceptos matemáticos en los libros de texto (por ejemplo en Mayén, 2009).

4.4. METODOLOGÍA DEL ANÁLISIS

Puesto que trabajamos con textos, se trata de una investigación cualitativa que, según Kirk y Miller (1986), incluye *"la inducción analítica, el análisis de contenido, la semiótica y ciertas manipulaciones de archivos, informáticas y estadísticas"* (p. 10). Se mantiene una concepción global fenomenológica, que trata de identificar la naturaleza profunda de las realidades analizadas, en este caso, el libro de texto (Cook y Reichardt, 2000). Según las clasificaciones recogidas en Bisquerra (1989), el proceso de investigación seguido es inductivo, pues partimos del examen de casos particulares; y el objetivo es descubrir generalizaciones a partir de observaciones sistemáticas de la realidad. Es una investigación aplicada, ya que está encaminada a obtener criterios para el desarrollo curricular. Es descriptiva, puesto que no se manipula ninguna variable, sino que se limita a observar y describir los fenómenos.

De acuerdo a López (2002), nuestro análisis se sitúa entre los *métodos intensivos*, que estudian con detenimiento algunos documentos, en lugar de recurrir a una muestra más amplia pero analizada someramente (métodos extensivos). Siguiendo a este mismo autor, se trata de un *análisis interno* de los documentos, procurando destacar su sentido y características fundamentales.

Muestra utilizada

La muestra utilizada está formada por dieciseis libros de textos de primer curso de Bachillerato; ocho de ellos en la modalidad de *Humanidades y Ciencias Sociales* y otros ocho de la modalidad de *Ciencia y Tecnologia*, publicados recién implantado el currículo actual de Bachillerato (MEC, 2007b). Los libros se eligieron por ser los más utilizados en la enseñanza pública en la Comunidad Autónoma de Andalucía, y estar publicados en editoriales de gran tradición y prestigio, siendo analizados para cada editorial los textos correspondientes a las dos modalidades citadas. Resaltamos el hecho de que todos han estado vigentes hasta final de 2013 (no se reeditaron).

En la Tabla 4.4.1 se asigna un código a cada libro, que será utilizado a lo largo del capítulo para referirse al mismo. Los códigos se han asignado por orden alfabético según las editoriales, utilizando [H] para los textos de la modalidad de *Humanidades y Ciencias Sociales* y [T] para la modalidad en *Ciencias y Tecnología*.

Al tratarse de una investigación cualitativa, se elige una muestra intencional con una serie de criterios convenientes para los fines de la investigación, procurando que los elementos se complementen y equilibren recíprocamente. Es decir, se trata de buscar una muestra que sea comprehensiva y, a su vez, haciendo énfasis en los casos más representativos y paradigmáticos, para capturar la mayor riqueza posible de la realidad analizada (Martínez, 2006).

Por tanto, no se aspira a generalizar los resultados del análisis a otros textos; sino que se busca la comparabilidad y traducibilidad (Goetz y Lecompte, 1998), con lo que la responsabilidad de la generalización de los resultados no está en el investigador, sino en el lector del informe resultante:

- La *comparabilidad* exige que el investigador use una terminología y un marco analítico normalizado. Las características de la muestra y de los constructos generados se definen con todo detalle para hacer posible la comparación de resultados con los de otros estudios relacionados.

- La *traducibilidad* es el grado en que los marcos teóricos y técnicas de investigación resultan comprensibles para otros investigadores de la misma disciplina o de otras relacionadas. Por este motivo, se explicitó con todo detalle el marco teórico en el Capítulo 2, y se explicitarán igualmente las categorías analizadas en este estudio, utilizando ejemplos de las mismas cuando sea necesario.

Tabla 4.4.1. Libros de texto utilizados en el análisis

Código	Referencia
H1	Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Grupo Anaya.
H2	Arias, J. M. y Maza, I. (2011). *Matemáticas aplicadas a las Ciencias Sociales 1*. Madrid: Grupo Editorial Bruño.
H3	Anguera, J., Biosca, A., Espinet, M. J., Fandos, M. J., Gimeno, M. y Rey, J. (2008). *Matemáticas I aplicadas a las Ciencias Sociales*. Barcelona: Guadiel - Grupo Edebé.
H4	Monteagudo, M. F. y Paz, J. (2008). *1º Bachillerato. Matemáticas aplicadas a las Ciencias Sociales.* Zaragoza: Edelvives (Editorial Luis Vives).
H5	Martínez, J. M., Cuadra, R., Heras, A. (2008). *Matemáticas aplicadas a las Ciencias Sociales. 1.º Bachillerato*. Madrid: McGraw-Hill.
H6	Bescós, E. y Pena, Z. (2008). *Matemáticas aplicadas a las Ciencias Sociales.* Vizcaya: Oxford University Press España.
H7	Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santos, D. y de Vicente, M. (2009). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Santillana Educación.
H8	Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales 1*. Madrid: Ediciones SM.
T1	Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). *Matemáticas I*. Madrid: Grupo Anaya.
T2	Arias, J. M. y Maza, I. (2011). *Matemáticas 1*. Madrid: Grupo Editorial Bruño.
T3	Biosca, A., Doménech, M., Espinet, M. J., Fandos, M. J. y Jimeno, M. (2008). *Matemáticas I*. Barcelona: Guadiel - Grupo Edebé.
T4	Monteagudo, M. F. y Paz, J. (2008). *1º Bachillerato. Matemáticas. Ciencias y Tecnología.* Zaragoza: Edelvives (Editorial Luis Vives).
T5	Martínez, J. M., Cuadra, R., Barrado, F. J. (2007). *Matemáticas 1º Bachillerato*. Madrid: McGraw-Hill.
T6	Bescós, E. y Pena, Z. (2009). *Matemáticas. 1 Bachillerato.* Navarra: Oxford University Press España.
T7	Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santos, D. y de Vicente, M. (2008). *Matemáticas I. 1 Bachillerato*. Madrid: Santillana Educación.
T8	Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008). *Matemáticas 1*. Madrid: Ediciones SM.

Método de análisis

Utilizamos el análisis de contenido, que asume que un texto puede dividirse en unidades que pueden clasificarse en un número reducido de categorías, en función de variables subyacentes, y que permiten realizar inferencias sobre su contenido (Krippendorff, 1997). Este análisis difiere de otras técnicas de estudio documental (por ejemplo, del método histórico), porque sustituye en lo posible las interpretaciones y subjetividad del estudio de documentos o de comunicaciones por procedimientos estandarizados, con el fin de convertir en datos los contenidos analizados en los documentos (León y Montero, 2002).

De acuerdo a Ghiglione y Matalón (1989), se trata de un análisis de contenido temático, donde se recurre a la lógica y al conocimiento del investigador sobre el tema para resumir el contenido del texto, definir categorías, y verificar su validez. Este conocimiento, en nuestro caso, lo hemos adquirido a través de la revisión bibliográfica y el estudio histórico y matemático previo (Estepa y Gea, 2010; Estepa, Gea, Cañadas y Contreras, 2012). El marco teórico del enfoque ontosemiótico nos proporciona, además, una categorización de objetos matemáticos que orienta y aporta concreción al análisis de contenido. Seguimos el método utilizado por Cobo (2003), que consiste en los siguientes pasos:

1. Seleccionados los libros y el tema correspondiente a la correlación y regresión, se efectuaron varias lecturas cuidadosamente para determinar los párrafos que constituirían la primera unidad de análisis.
2. Mediante un proceso cíclico e inductivo, se comparó el contenido de dichos párrafos con los objetos matemáticos identificados en el significado de referencia, para determinar su presencia en los libros de texto. Estos elementos constituirían nuestras unidades secundarias de análisis.
3. Una vez que se llegó a una lista de los principales campos de problemas, definiciones, propiedades, representaciones, procedimientos y argumentos presentes en los libros, se procedió a analizar la forma en que se presentan, y a buscar y describir el ejemplo más característico para cada posible presentación, seleccionando imágenes para ejemplificar cada elemento de significado hallado. Paralelamente, se realizó una clasificación de los posibles conflictos semióticos observados, según cada objeto matemático analizado.
4. Elaboración de tablas para resumir los resultados, y obtener conclusiones sobre el significado de referencia en los libros analizados.

Finalmente, se realiza una discusión de los resultados del análisis en cada uno de los elementos de significado, para contrastarlos con el significado institucional pretendido descrito en el Capítulo 1. El análisis de libros de texto también se llevó a cabo por Lavalle, Micheli y Rubio (2006), y más profundamente por Sánchez Cobo (1999), quien analizó los contextos, el contenido matemático implicado, el tipo de tarea, el tipo de covariación según las categorías definidas por Barbancho (1973), y el tipo e intensidad de la dependencia. A medida que se presentan los resultados obtenidos en nuestro análisis, se irán estableciendo comparaciones con los de ambas investigaciones.

A continuación, se describen los resultados del análisis realizado, tanto en el desarrollo de la lección en los textos, donde se incluyen enlaces a Internet como complemento del tema, como en los CDs que algunos libros de texto incluyen.

4.5. ANÁLISIS DE LAS SITUACIONES-PROBLEMAS

En el enfoque ontosemiótico (EOS) (Godino, Batanero y Font, 2007) se da una gran importancia a la situación-problema, pues postula que los objetos matemático emergen de las prácticas, personales o institucionales, al resolver problemas.

Dependiendo de aspectos específicos, se pueden agrupar en tipos o clases, donde, *"el paso de un tipo puntual a otro más amplio es el determinante del progreso o avance del conocimiento matemático, tanto individual como institucional*" (Godino, 2002, p.

109). En el contexto educativo, las situaciones-problemas que se plantean al estudiante deben, por un lado, ser representativas de las incluidas en el significado institucional de un concepto y, por otro lado, permitir contextualizar los conocimientos pretendidos, ejercitarlos y aplicarlos a situaciones relacionadas (Godino, Font, Contreras y Wilhelmi, 2006).

Por todo ello, el primer elemento analizado en los libros de texto son las situaciones-problema. Algunos resultados previos del análisis se han publicado en Gea, Batanero y Cañadas (2013) y Gea, Batanero, Cañadas y Contreras (2013a, 2013b, en prensa). No incluimos en el análisis los campos de problemas contenidos en los CDs, sino que se realiza un análisis específico de los mismos en la Sección 4.10.3.

Variables consideradas en el análisis

El primer paso fue identificar en los textos los principales campos diferenciados de problemas. Seguidamente, diferenciamos, siguiendo a Ortiz (1999), ejemplos, ejercicios resueltos o a resolver por el estudiante. En tercer lugar, nos interesamos por los contextos que se utilizan en las situaciones para mostrar la aplicación de los diferentes objetos matemáticos y dotarles de sentido. Por último, se analizan las características de la correlación (intensidad, signo y tipo de función de ajuste) que se emplean en el diseño de las tareas.

4.5.1. CAMPOS DE PROBLEMAS

La clasificación teórica de los principales campos de problemas que dotan de sentido a la correlación y regresión, se basa en un estudio previo del significado de estas nociones en la Historia (Estepa y Gea, 2010; Estepa, Gea, Cañadas y Contreras, 2012), donde se observaron dos preguntas principales que dieron origen a estos conceptos:

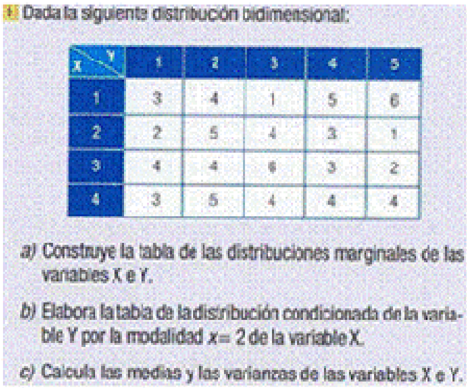
- ¿Hay alguna relación entre las variables? ¿Es intensa o moderada? ¿Directa o inversa?
- ¿Puedo usar una variable para predecir la otra?

De las anteriores preguntas se derivan los campos de problemas que se proponen en los libros de texto, añadiendo otro (que denominamos P0), relacionado con la organización y reducción de los datos. Estos tres campos principales de problemas, algunos de los cuales hemos subdividido, se analizan a continuación.

P0. Organización/representación de datos bidimensionales

La organización de la información es un primer paso en el estudio de la correlación y regresión (Gea, Batanero, Cañadas y Arteaga, 2013a). Incluimos aquí también la representación tabular y gráfica de un conjunto de datos bivariantes (Figura 4.5.1), lectura de las mismas y las traducciones entre dichas representaciones, como la que se muestra en la Figura 4.5.2.

Estas actividades requieren, en algunos casos, un nivel “leer entre los datos” (Curcio, 1989), ya que el alumno no se limita a leer directamente un dato de la tabla o gráfico, sino que ha de realizar comparaciones o traducciones. Este campo de problemas incluiría a las categorías “construir la tabla de frecuencias de la distribución bidimensional" y "representación gráfica del diagrama de dispersión", del estudio de Sánchez Cobo (1999).



Dada la siguiente distribución bidimensional:

X \ Y	1	2	3	4	5
1	3	4	1	5	6
2	2	5	4	3	1
3	4	4	6	3	2
4	3	5	4	4	4

a) Construye la tabla de las distribuciones marginales de las variables X e Y.

b) Elabora la tabla de la distribución condicionada de la variable Y por la modalidad *x* = 2 de la variable X.

c) Calcula las medias y las varianzas de las variables X e Y.

Figura 4.5.1. Representación tabular y gráfica de datos bidimensionales ([H4], p. 220)

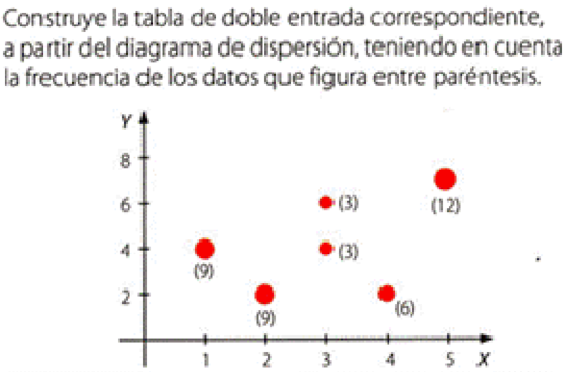


Figura 4.5.2. Tarea de traducción entre diferentes representaciones ([T7], p.317)

P1. Analizar la existencia de relación entre variables

Organizados los datos bivariantes, el segundo paso sería analizar la relación entre las variables del estudio, que da origen a la correlación (Estepa, et al. (2012). Este campo de problemas fue también considerado por Lavalle, Micheli y Rubio (2006), pudiendo equipararse al denominado "interpretación" por Sánchez Cobo (1999). Su finalidad es reflexionar sobre la dependencia funcional o estadística de las variables, y si existe esta última, estudiar el tipo de dependencia (principalmente lineal), su intensidad y sentido. Podemos subdividir este campo de problemas en los siguientes:

P11. Análisis de las variables que conforman la variable estadística bidimensional. Cuando se analiza la variable estadística bidimensional, es necesario delimitar cada una de las variables que la constituyen; por ejemplo, en el texto [H7] (p.244) se describe la variable bidimensional: "*Tamaño de un archivo informático y tiempo que se tarda en copiarlo*" y se pide a los estudiantes indicar las variables estadísticas unidimensionales que la conforman, y proponer a tres pares de valores.

También consideramos en esta categoría el análisis descriptivo de cada variable y los pasos requeridos para completar una tabla de doble entrada conocidas las medias marginales, o bien determinar el centro de gravedad (Figura 4.5.3). Este campo de problemas no fue considerado por Sánchez Cobo (1999), pero sí por Lavalle, Micheli y Rubio (2006). Al igual que lo sugerido por estas autoras, nuestros textos lo tratan a nivel elemental, y lo relacionan con conceptos previos, como el cálculo de promedios.

Se quiere construir una escuela a la que acudan los niños y niñas de 6 pequeños núcleos de población de una comarca. La posición sobre el plano y el número de niños de cada pueblo se dan en la tabla:

Pueblo	*A*	*B*	*C*	*D*	*E*	*F*
Niños	30	15	10	35	8	5
Posición	(3, 4)	(2, 5)	(5, 4)	(2, 2)	(6, 6)	(9, 4)

a) Determina el pueblo más adecuado para construir la escuela, sin tener en cuenta el número de niños.
b) Haz lo mismo teniendo en cuenta su número.

Figura 4.5.3. Tarea de análisis unidimensional en las variables de estudio ([T5], p.375)

P12. Análisis del tipo de dependencia. Definidas las variables, habría que decidir si a cada valor de la variable independiente corresponde un único valor de la variable dependiente (dependencia funcional), varios (dependencia aleatoria o estadística), o no hay relación (independencia). Encontramos un ejemplo en los textos [H3] y [T3], en el que se pide caracterizar la dependencia que presentan los datos en un diagrama de dispersión que muestra dependencia aleatoria exponencial inversa y alta (Figura 4.5.4). Este campo de problemas se encuentra en todos los textos analizados, y motiva el análisis posterior de regresión. Principalmente se usa el diagrama de dispersión para analizar la dependencia, y proponer alguna función de ajuste. Su inclusión en el tema ayudará a controlar la concepción determinista (Estepa, 1994), pues con frecuencia los estudiantes usan un procedimiento determinístico en situaciones en que es adecuada la regresión, debido a la naturaleza aleatoria de los datos (Agnelli *et al.*,2009).

Observa el diagrama de dispersión de la figura y caracteriza la relación que existe entre las variables *X* e *Y*.

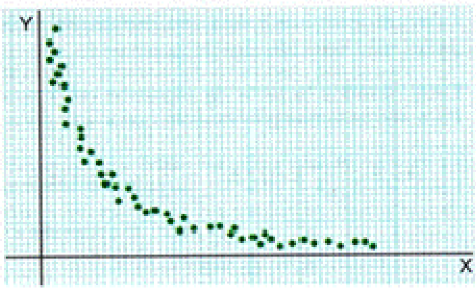


Figura 4.5.4. Tarea de análisis del tipo de dependencia entre las variables de estudio ([T3], p.282)

8. Las puntuaciones en Matemáticas y Física de siete alumnos han sido las siguientes.
PAU

Matemáticas	8	8	6	7	8	6	2
Física	7	7,5	5	7	7,5	5	7

a) Halla el coeficiente de correlación de las calificaciones en Matemáticas y Física de los seis primeros alumnos.

b) Calcula el coeficiente de correlación de esas dos variables para los siete alumnos.

c) Explica la diferencia entre los resultados obtenidos.

Figura 4.5.5. Tarea de análisis de dependencia con presencia de outliers ([T8], p.323)

Incluimos también en esta categoría las tareas sobre discriminación de correlación y causalidad (que sólo se presentan en [H8]). Otras tareas poco frecuentes plantean observar el efecto de los datos atípicos (Figura 4.5.5). Estas actividades no aparecen en ninguno de los textos de la investigación de Sánchez Cobo (1999), a pesar de su importancia, ya que los datos atípicos influyen en el valor del coeficiente de correlación, y pueden dar una falsa apariencia del tipo de dependencia.

P13. Medida de la intensidad de la relación entre las variables. Una vez detectada la relación, interesa cuantificar su intensidad, que variará desde la independencia hasta la dependencia funcional. El objeto matemático más usado para determinarla, en este nivel educativo, es el coeficiente de correlación lineal, con el cálculo previo de la covarianza. Otras tareas incluidas en este campo de problemas son las de estimación de la correlación a partir de representaciones de datos, principalmente mediante el diagrama de dispersión, y aquellas en que se pide asignar unos coeficientes de correlación dados a determinados diagramas de dispersión (Figura 4.5.6).

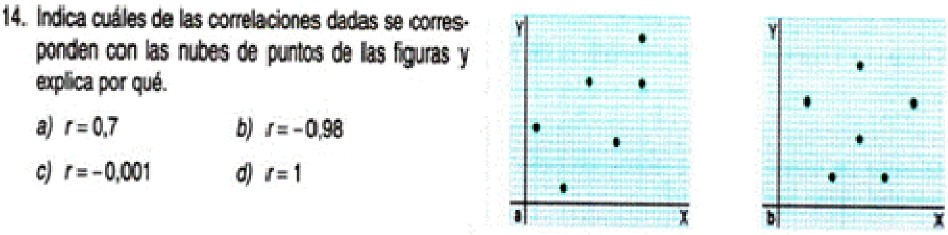


Figura 4.5.6. Elección del coeficiente de correlación a partir del diagrama de dispersión ([H3], p.229)

P14. Determinar la dirección de la relación entre variables. El último problema en este apartado es estudiar el sentido de la dependencia, donde se pretende que el estudiante distinga si la relación es directa (al crecer/decrecer una variable crece/decrece la otra) o inversa (al crecer/decrecer una variable decrece/crece la otra). Hacemos notar que, en caso que los datos presenten relación curvilínea (exponencial o polinómica, por ejemplo), ésta se ha caracterizado en directa o inversa siempre que los datos mostrasen una tendencia creciente/decreciente de dependencia evidente.

P2. Predicción de una variable en función de otra

Una vez aceptada la existencia de una relación entre las dos variables de estudio (problema de correlación), interesa encontrar la función que permita obtener una de las variables a partir de la otra (problema de regresión). Se pide al estudiante que de la expresión algebraica de la función de ajuste, usualmente lineal. Este problema se puede descomponer en dos, como se describe a continuación.

P21. Analizar el ajuste lineal entre variables. Usualmente, la enseñanza de la regresión en los textos analizados se reduce al modelo de regresión lineal, que es el más sencillo y asequible para los estudiantes. Generalmente, los problemas proponen el ajuste matemático o el trazado "a ojo" de las rectas de regresión a partir de un diagrama de dispersión; en el primer caso, es necesario el cálculo de los parámetros de las rectas. En menor medida, encontramos algunas situaciones donde se pide la comprobación de propiedades del modelo de regresión lineal, el cálculo del centro de gravedad sin aplicar cálculos de medias marginales, asignar una recta a una nube de puntos, o estimar el ángulo entre las rectas, como el ejemplo presentado en la Figura 4.5.7.

56 Investiga sobre las siguientes cuestiones.

a) ¿Es cierto que el signo de las pendientes de las dos rectas de regresión de una variable bidimensional es siempre igual?

b) ¿Qué sucede si las dos rectas de regresión tienen la misma pendiente? ¿Cómo es la correlación?

57 El ángulo que forman las dos rectas de regresión de una distribución bidimensional es mayor cuanto menor sea el coeficiente de correlación.

Vamos a comprobarlo estudiando las dos magnitudes en estas distribuciones.

10	12	14	16	18
3	8	1	9	2

10	12	14	16	18
3	6	8	6	7

10	12	14	16	18
5	6	6,5	8,5	9

Figura 4.5.7. Análisis de regresión lineal ([H7], p. 263)

Destacamos los textos [H8] y [T8] por explicar (un ejercicio resuelto) y proponer tareas sobre la recta de Tukey. Además, estos textos introduce también el procedimiento de cambio de variable para determinar un ajuste no lineal a los datos. También se pide comparar estos modelos de ajuste con el lineal y justificar la elección del mejor modelo, tarea que ayudará al estudiante a comprender mejor la regresión.

P22. Realizar estimaciones mediante el modelo de regresión. Una vez calculada la función de ajuste a la distribución bidimensional, los textos resaltan su utilidad planteando ejercicios de estimación como los que se muestran en la Figura 4.5.8. En dicho ejemplo, una vez calculada la función de ajuste, se pide realizar extrapolaciones en valores de la variable independiente cercanos y alejados al rango de datos. Este campo de problemas es similar a la categoría "predicción" de Sánchez Cobo (1999).

El alcalde de un pueblo ha constatado una reducción del número de nacimientos de niños, y ha encargado realizar un estudio.

Año	86	89	92	95	98	01	04	07
Nacimientos	50	54	40	33	34	23	21	17

a) ¿Puede establecerse, de forma fiable, una fórmula que relacione el año con el número de nacimientos?

b) ¿Cuántos nacimientos pueden estimarse en 2008? ¿Y en 2010? ¿Qué puede estimarse para 2050?

c) ¿Es fiable esta última estimación? Razona la respuesta.

En una población la media de los pesos es de 70 kg y la de las estaturas 175 cm. Las desviaciones típicas son 5 kg y 10 cm, respectivamente y la covarianza de ambas variables es 40.

a) Estima el peso de una persona de esa población que mide 185 cm de estatura.

b) Usando el coeficiente de correlación lineal, explica hasta qué punto confías en la estimación que ha hecho en el apartado a).

Figura 4.5.8. Estimación mediante el ajuste lineal ([H7], p.262 y [T5], p.374)

Advertimos algunas imprecisiones en los textos [H6], [T6], [H4] y [T4] al igual que Lavalle, Micheli y Rubio (2006), porque el cálculo de la estimación lleva asociado una reflexión en torno al valor esperado o promedio y al valor real, ya que puede haber diferentes valores observados. Cabe matizar que, a diferencia del campo de problemas P13, el cálculo del coeficiente de correlación o determinación, en este caso, se utilizan más para valorar las predicciones realizadas que para valorar la bondad del ajuste, como se muestra en la Figura 4.5.8.

Además, se incluyen en esta categoría las tareas que piden al estudiante analizar la bondad del ajuste del modelo de regresión, que en algunos textos se plantea mediante el cálculo del coeficiente de determinación. Se trata de cuantificar la intensidad de la dependencia entre las variables mediante el coeficiente de determinación. A modo de ejemplo, en [T6] se plantea la siguiente tarea: "*¿Cuál sería la fiabilidad de un ajuste bidimensional con r = 0,7? ¿Y con r = -0,8? ¿Y con r = 0,9*?" ([T6], p. 187). Como señala Sánchez Cobo:

> Nos da una interpretación alternativa del coeficiente de correlación, al indicar que un valor grande del mismo denota que una de las variables puede dar cuenta de un alto porcentaje de variabilidad de la otra. (Sánchez Cobo, 1999, p. 109).

Síntesis de los campos de problemas

En la Tabla 4.5.1 se resume el análisis realizado de los campos de problemas. Observamos que se plantean todos los campos de problemas descritos, aunque con diferente frecuencia, predominando el campo P1 (determinación de la existencia de relación entre las variables), con aproximadamente, entre el 56% y el 65% en ambas modalidades de Bachillerato. Dentro del mismo, hay tendencia a determinar la intensidad de la dependencia (P13) y su signo (P14), más que en analizar el tipo de dependencia aleatoria o estadística (P12) o definir las variables (P11), a pesar de que, como se indicó anteriormente, estos campos de problemas son necesarios para un desarrollo adecuado de la noción de regresión. En este sentido, destacamos el texto [H8] por dar mayor importancia a dicho campo de problemas (P21).

Tabla 4.5.1. Frecuencias (y porcentaje) de campos de problemas por texto

Campo de problemas		H1	H2	H3	H4	H5	H6	H7	H8
P0		18(6,7)	24(10,9)	45(17,4)	38(16,9)	28(8,8)	8(4,5)	70(17,4)	23(7,7)
P1	P11	27(10,1)	26(11,8)	16(6,2)	45(20)	21(6,6)	27(15,3)	34(8,4)	38(12,8)
	P12	39(14,6)	31(14)	51(19,8)	26(11,6)	44(13,8)	24(13,6)	45(11,2)	44(14,8)
	P13	56(20,9)	53(24)	49(19)	45(20)	76(23,9)	39(22,2)	92(22,8)	61(20,5)
	P14	52(19,4)	21(9,5)	36(14)	27(12)	51(16)	16(9,1)	54(13,4)	40(13,5)
P2	P21	44(16,4)	26(11,8)	26(10,1)	26(11,6)	50(15,7)	28(15,9)	59(14,6)	60(20,2)
	P22	32(11,9)	40(18,1)	35(13,6)	18(8)	48(15,1)	34(19,3)	49(12,2)	31(10,4)
Total		268	221	258	225	318	176	403	297

		T1	T2	T3	T4	T5	T6	T7	T8
P0		18(6,7)	24(10,9)	36(19,4)	38(16,9)	27(8,8)	8(4,5)	50(13,5)	15(5,3)
P1	P11	27(10)	26(11,8)	14(7,5)	45(20)	21(6,8)	27(15,3)	34(9,2)	28(9,9)
	P12	39(14,5)	31(14)	38(20,4)	26(11,6)	41(13,4)	24(13,6)	41(11,1)	46(16,2)
	P13	56(20,8)	53(24)	29(15,6)	45(20)	72(23,5)	39(22,2)	91(24,5)	70(24,6)
	P14	52(19,3)	21(9,5)	23(12,4)	27(12)	48(15,6)	16(9,1)	53(14,3)	45(15,8)
P2	P21	45(16,7)	26(11,8)	22(11,8)	26(11,6)	53(17,3)	28(15,9)	56(15,1)	47(16,5)
	P22	32(11,9)	40(18,1)	24(12,9)	18(8)	45(14,7)	34(19,3)	46(12,4)	33(11,6)
Total		269	221	186	225	307	176	371	284

La representación de datos bidimensionales (P0), tiene poca presencia, principalmente en [H6] y [T6], siendo los textos [H3], [H4], [H7] y [T3], [T4], [T7] los que mayor número de tareas proponen en este campo. Este tipo de problemas facilitará al estudiante alcanzar un nivel de abstracción en la organización de los datos. Por ejemplo, en la tarea que se mostró en la Figura 4.5.8, el estudiante debe reflexionar sobre la representación gráfica de los datos del problema, y advertir la continuidad de la variable tiempo, para interpretar el coeficiente de correlación y la recta de regresión.

En cuanto al campo de problemas P1, no hay diferencias significativas entre las dos modalidades, planteándose principalmente tareas en las que se pide cuantificar la intensidad de la relación entre las variables (P13), seguidas de aquellas en que se pide determinar su signo (P14) o analizar el tipo de dependencia (P12), a excepción de [H4], [H6] y [T4], [T6], donde se plantean más tareas referidas al análisis de las variables que conforman la variable estadística bidimensional (P11). Dentro del campo P2 no hay mucha diferencia en cada editorial, entre las dos modalidades de Bachillerato, donde unas editoriales proponen más problemas tipo P21 (ajustar una recta a los datos): [H1], [H4], [H5], [H7], [H8], [T1], [T4], [T5], [T7], [T8], y el resto al contrario.

4.5.2. TIPO DE ACTIVIDAD

Una segunda variable considerada, al igual que en el estudio de Ortiz (1999), es si la situación planteada es un ejemplo o ejercicio, y en este caso, si se resuelve o no en el texto. Como el autor, consideramos como ejemplo las descripciones de situaciones en que se aclara el significado de los conceptos introducidos teóricamente; y ejercicios, todas aquellas actividades, resueltas o no, que se presentan al estudiante para adquirir o reforzar el aprendizaje de un objeto matemático (independientemente de que en el libro su título sea problema, actividad, ejercicio o ejemplo). Detallamos además, si los ejercicios propuestos al estudiante se encuentran en el desarrollo del tema o al final, ya que, si se presentan al comienzo se puede suponer que es un verdadero problema para el estudiante, mientras que si se presenta al final se podría considerar como un ejercicio de refuerzo del aprendizaje. Sánchez Cobo (1999) manifiesta que los ejercicios generalmente aparecen al final del tema, a excepción de dos casos en los que aparecen en el desarrollo de la lección y al final.

Tabla 4.5.2 Frecuencia y porcentaje de tipos de tareas por texto

	Tarea	H1	H2	H3	H4	H5	H6	H7	H8
	Ejemplo	38(14,2)	12(5,4)	42(16,3)	20(8,9)	79(24,8)	20(11,4)	24(6)	66(22,2)
Ejer.	Resuelto	43(16)	14(6,3)	31(12)	38(16,9)	55(17,3)	17(9,7)	60(14,9)	33(11,1)
	En el tema	30(11,2)	43(19,5)	85(32,9)	29(12,9)	11(3,5)	4(2,3)	84(20,9)	50(16,8)
	Al final	157(58,6)	152(68,8)	100(38,8)	138(61,3)	173(54,4)	135(76,7)	235(58,3)	148(49,8)
	Total	268	221	258	225	318	176	403	297

		T1	T2	T3	T4	T5	T6	T7	T8
	Ejemplo	38(14,1)	12(5,4)	42(22,6)	20(8,9)	79(25,7)	20(11,4)	24(6,5)	60(21,1)
Ejer.	Resuelto	43(16)	14(6,3)	15(8,1)	38(16,9)	49(16)	17(9,7)	44(11,9)	35(12,3)
	En el tema	30(11,2)	43(19,5)	70(37,6)	29(12,9)	11(3,6)	4(2,3)	71(19,1)	44(15,5)
	Al final	158(58,7)	152(68,8)	59(31,7)	138(61,3)	168(54,7)	135(76,7)	232(62,5)	145(51,1)
	Total	269	221	186	225	307	176	371	284

En la Tabla 4.5.2 podemos observar que se plantean gran cantidad de actividades al estudiante al final del tema; en los textos [H6] y [T6] prácticamente todas. Esta tendencia también es observada por Sánchez Cobo (1999), y por Ortiz (1999) en su trabajo sobre probabilidad, quien indica que este exceso de ejercicios al final del tema puede corresponder a una orientación teoría-práctica en los libros de texto, que se daría también en nuestro estudio. Esta orientación parece menor en el texto [T3], por el menor número de ejercicios al final del tema, aunque todavía éstos superan a los propuestos a lo largo del tema.

En cuanto a los ejemplos y ejercicios resueltos que se incluyen en el tema, las frecuencias son variadas: se incluyen entre un 5% ([H2], [T2]) y un 25% ([H5], [T5]) de ejemplos, con un tratamiento similar en cada modalidad; y entre un 6% ([H2], [T2]) y un 17% ([H4], [H5], [T4]) de ejercicios resueltos, siendo mayor su representatividad en la modalidad de *Humanidades y Ciencias Sociales*. Como señala Ortiz (1999), la presentación de ejemplos influye positivamente en la enseñanza de los conceptos tratados, y aunque aparentemente no son muchos, si añadimos a los ejemplos los ejercicios resueltos, la diferencia entre ejemplos y ejercicios se hace menos acusada. Aún así, se mantienen los casos extremos de los textos [H2], [T2], [H6] y [T6], que muestran una orientación teoría-práctica muy grande, con la mayor parte de ejercicios al final del tema. También destacan los textos [H3] y [T3] por ser de los presentan un mayor equilibrio en la distribución de tareas según su tipo.

La distribución de los ejercicios propuestos en el tema varía según editorial, no tanto entre modalidad, encontrando entre un 2% ([H6] y [T6]) y un 33% en [H3] o 38% en [T3], aunque en algunos casos, superan a los ejemplos o los ejercicios resueltos, según textos. Por ello, si se compara con los ejemplos y ejercicios resueltos como si fuesen una única categoría, se mejora mucho la situación ya que, estos juntos superan a los ejercicios propuestos en todos los casos (el caso de [H7] y [T7] es muy aproximado), salvo en los casos extremos de [H2], [H3], [T2] y [T3].

4.5.3. CONTEXTOS

Chevallard (1991) indica que, en el proceso de transposición didáctica, una vez introducido un tema en el sistema de enseñanza, el dispositivo didáctico pretende progresivamente buscarle aplicaciones, que pueden no tener relación con aquellas en que originariamente se inició el concepto. La función que tienen los contextos es permitir finalmente la recontextualización del saber.

Con objeto de analizar este aspecto, también tenido en cuenta por Sánchez Cobo (1999), se han estudiado los contextos de aplicación que fundamentan las tareas que se presentan en los libros de texto. Concretamente, clasificaremos los contextos según las situaciones, que podrán ser, de acuerdo a la caracterización del estudio PISA (Ministerio de Educación, 2009):

- *Personales*: relacionadas con las actividades diarias de los alumnos. Hemos encontrado dos tipos diferentes:
  - (a) *Fenómenos biológicos* (como la estatura de hijos – estatura de padres, precisamente el problema que históricamente dio origen a la idea de regresión);
  - (b) *Deportivos* (distancia del jugador – número de encestes al jugar al baloncesto);
- *Científicas:* que implican la comprensión de un proceso tecnológico o una interpretación teórica. Resaltaremos estos especialmente, ya que consideramos el Bachillerato de *Ciencias y Tecnología*. Algunos ejemplos son el aumento de peso de un animal y la cantidad de mg diarios de un fármaco; alargamiento de una barra

metálica y la temperatura a la que se expone; o altura que alcanza una piedra y la fuerza con que se lanza;

- *Públicas:* se refieren a la comunidad local, u otra más amplia, en la que los estudiantes observan determinados aspectos sociales de su entorno, o que aparezcan en los medios de comunicación. Hemos encontrado dos tipos:
  (c) *Economía* (consumo de energía per cápita y renta per cápita; kg de capturas de pescado y precio de subasta en la lonja);
  (d) *Sociología y demografía* (renta per cápita e índice de natalidad);
- *Educativa, ocupacional o laboral*: son las que encuentra el alumno en el centro escolar o en un entorno de trabajo futuro como por ejemplo las notas de exámenes: física y matemáticas; matemáticas y filosofía;
- Finalmente, hemos considerado también una séptima categoría "sin contexto".

Tabla 4.5.3. Frecuencia (y porcentaje) de contextos por texto

Contexto		H1	H2	H3	H4	H5	H6	H7	H8
Personal	Biológico	22(8,2)	24(10,9)	18(7)	10(4,4)	44(13,8)	31(17,6)	28(6,9)	37(12 ,5)
	Deportes	12(4,5)		14(5,4)	5(2,2)	7(2,2)			
Científico		35(13,1)	73(33)	38(14,7)	30(13,3)	42(13,2)	29(16,5)	38(9,4)	27(9,1)
Educativo		28(10,4)	10(4,5)	5(1,9)	32(14,2)	54(17)	27(15,3)	23(5,7)	35(11,8)
Público	Economía	38(14,2)	18(8,1)	35(13,6)	12(5,3)	22(6,9)	26(14,8)	14(3,5)	30(10,1)
	Sociología	17(6,3)	38(17,2)	28(10,9)	32(14,2)	29(9,1)	22(12,5)	40(9,9)	29(9 ,8)
Sin contexto		116(43,3)	58(26,2)	120(46,5)	104(46,2)	120(37,7)	41(23,3)	260(64,5)	139(46,8)
	Total	268	221	258	225	318	176	403	297

		T1	T2	T3	T4	T5	T6	T7	T8
Personal	Biológico	22(8,2)	24(10,9)	21(11,3)	10(4,4)	46(15)	31(17,6)	25(6,7)	24(8,5)
	Deportes	12(4,5)		17(9,1)	5(2,2)	4(1,3)			
Científico		35(13)	73(33)	31(16,7)	30(13,3)	39(12,7)	29(16,5)	38(10,2)	33(11,6)
Educativo		28(10,4)	10(4,5)	6(3,2)	32(14,2)	44(14,3)	27(15,3)	23(6,2)	33(11,6)
Público	Economía	38(14,1)	18(8,1)	13(7)	12(5,3)	22(7,2)	26(14,8)	11(3)	21(7 ,4)
	Sociología	17(6,3)	38(17,2)	17(9,1)	32(14,2)	23(7,5)	22(12,5)	30(8,1)	26(9 ,2)
Sin contexto		117(43,5)	58(26,2)	81(43,5)	104(46,2)	129(42)	41(23,3)	244(65,8)	147(51,8)
	Total	269	221	186	225	307	176	371	284

En la Tabla 4.5.3 clasificamos las tareas según el contexto utilizado. Observamos un alto número de ejercicios descontextualizados, en especial en [H7] y [T7], seguido de [H8] y [T8], superando a lo obtenido por Sánchez Cobo (1999), que obtuvo un 37,9% de ejercicios descontextualizados; nuestro porcentaje es mayor (entre un 42% y un 66% en la modalidad científico-tecnológica y entre un 38% y un 65% en la modalidad de humanidades y ciencias sociales), a excepción de [H2] y [T2] con un 26% y [H6] y [T6] un 23%. Esta alta presencia de ejercicios sin contexto es contraria a las recomendaciones actuales en la enseñanza, pues uno de los modos fundamentales de razonamiento estadístico, de acuerdo a Wild y Pfannkuch(1999), es la integración de la estadística con el contexto, y este tipo de ejercicios no facilitan este proceso.

El resto de contextos tienen igual representatividad, destacndor algo el contexto científico, principalmente en [H2] y [T2], y menos, el deportivo. Este resultado es preocupante, pues los alumnos tienen mejores resultados cuando el contexto les resulta familiar que si se trata de un nuevo contexto o situaciones abstractas.

4.5.4. TIPO DE DEPENDENCIA

Es importante que los textos presenten una distribución equilibrada de diferentes tipos dependencia. Por ello, y al igual que Sánchez-Cobo (1999), hemos analizado la intensidad, el sentido y el tipo de dependencia que son variables muy relevantes en el estudio del tema (Gea, Batanero, Cañadas, Contreras y Arteaga, 2013b).

Intensidad de la dependencia

Hemos clasificado las situaciones según su grado de dependencia. Cuando los datos presentan dependencia lineal, clasificamos su intensidad en alta si el coeficiente de correlación corresponde al intervalo [0,8; 1), media para el intervalo [0,5; 0,8) y baja si toma un valor [0,1; 0,5). Consideramos "independencia" en las correlaciones menores, y "funcional" cuando el ajuste de los datos a la recta es perfecto (valor absoluto del coeficiente de correlación igual a 1 para la dependencia lineal). La intensidad de la dependencia no lineal se considera alta si el ajuste de los datos a la función que los modeliza es muy bueno, y, en caso de poder calcular el coeficiente de determinación, cuando sea mayor o igual a 0,98; media para valores comprendidos entre 0,8 y 0,98 y baja en otros casos. Añadimos la categoría "descripción verbal" para cuando en el problema se dan dos variables, generalmente descritas verbalmente, y no es posible deducir la intensidad de la relación.

Tabla 4.5.4. Frecuencia (y porcentaje) de tareas según intensidad de la correlación

	H1	H2	H3	H4	H5	H6	H7	H8
Independencia	2(0,7)	5(2,3)	24(9,3)	12(5,3)	5(1,6)	6(3,4)	30(7,4)	21(7,1)
Baja	25(9,3)	19(8,6)	20(7,8)	52(23,1)	31(9,7)	15(8,5)	89(22,1)	27(9,1)
Media	53(19,8)	23(10,4)	34(13,2)	45(20)	55(17,3)	21(11,9)	65(16,1)	50(16,8)
Alta	127(47,4)	168(76)	124(48,1)	88(39,1)	169(53,1)	107(60,8)	114(28,3)	146(49,2)
Funcional	15(5,6)	4(1,8)	22(8,6)	14(6,2)	7(2,2)	3(1,7)	35(8,7)	24(8,1)
D. Verbal	46(17,2)	2(0,9)	34(13,2)	14(6,2)	51(16)	24(13,6)	70(17,4)	29(9,8)
Total	268	221	258	225	318	176	403	297

	T1	T2	T3	T4	T5	T6	T7	T8
Independencia	2(0,7)	5(2,3)	16(8,6)	12(5,3)	5(1,6)	6(3,4)	27(7,3)	21(7,4)
Baja	25(9,3)	19(8,6)	21(11,3)	52(23,1)	31(10,1)	15(8,5)	73(19,7)	20(7)
Media	53(19,7)	23(10,4)	46(24,7)	45(20)	64(20,8)	21(11,9)	61(16,4)	51(18)
Alta	128(47,6)	168(76)	48(25,8)	88(39,1)	148(48,2)	107(60,8)	111(29,9)	133(46,8)
Funcional	15(5,6)	4(1,8)	15(8,1)	14(6,2)	7(2,3)	3(1,7)	32(8,5)	30(10,6)
D. Verbal	46(17,1)	2(0,9)	40(21,5)	14(6,2)	52(16,9)	24(13,6)	67(18,1)	29(10,2)
Total	269	221	186	225	307	176	371	284

En la Tabla 4.5.4 se presentan las frecuencias y porcentajes de tareas según la intensidad de la correlación de los datos. Observamos que la mayoría tienen una correlación alta (entre un 26% y un 76% en la modalidad científico-tecnológica y entre un 28% y un 76% en *Humanidades y Ciencias Sociales*) sobre todo en [H2] y [T2]. Siguen las tareas con intensidad media, siendo menos frecuentes las de dependencia baja (excepto en [H4], [H7], [T4] y [T7]), independencias o dependencia funcional. Los resultados coinciden con los de Sánchez Cobo (1999), aunque sus porcentajes son más altos ya que, el 52,7% de las tareas tienen un valor absoluto del coeficiente de correlación superior a 0,9 (23,5% igual o superior a 0,99).

Sánchez Cobo, Estepa y Batanero (2000) indican que la precisión de las estimaciones del coeficiente de correlación es mayor cuando la correlación es más intensa; es posible que este hecho lleve a los autores de los textos a incidir tanto en la correlación alta. Sin embargo, en los estudios estadísticos, especialmente en las ciencias sociales, las correlaciones suelen ser moderadas o bajas; por lo que sería conveniente una mejor distribución de la intensidad de la correlación en los ejercicios de los textos.

Dirección de la dependencia

Se ha estudiado también el sentido de la dependencia, que se clasifica en directa si el crecimiento (decrecimiento) de una de las variables se acompaña del crecimiento (decrecimiento) de la otra, e inversa en caso contrario. Hemos añadido, como en la sección anterior, las categorías independencia y descripción verbal. En "otras" se incluye una diagrama de dispersión que se aproxima a una circunferencia ([H3], p.232; [T3], p.282), otro que lo hace a $Y=2$, por lo cual sería adecuado usar una recta de regresión con pendiente nula ([H3], p.229); el resto son funciones polinómicas de segundo grado, dos de ellas con dependencia funcional ([H7], p. 258; [T7], p. 316) y el resto aleatoria ([H7], p.260; [T7], p.318; [T8], p.323).

Tabla 4.5.5. Frecuencia (y porcentaje) de tareas según sentido de la correlación

	H1	H2	H3	H4	H5	H6	H7	H8
Independencia	2(0,7)	5(2,3)	24(9,3)	12(5,3)	5(1,6)	6(3,4)	30(7,4)	21(7,1)
Directa	160(59,7)	188(85,1)	123(47,7)	153(68)	147(46,2)	109(61,9)	208(51,6)	169(56,9)
Inversa	60(22,4)	26(11,8)	73(28,3)	46(20,4)	115(36,2)	37(21)	92(22,8)	78(26,3)
Otras			4(1,6)				3(0,7)	
D. Verbal	46(17,2)	2(0,9)	34(13,2)	14(6,2)	51(16)	24(13,6)	70(17,4)	29(9,8)
Total	268	221	258	225	318	176	403	297

	T1	T2	T3	T4	T5	T6	T7	T8
Independencia	2(0,7)	5(2,3)	16(8,6)	12(5,3)	5(1,6)	6(3,4)	27(7,3)	21(7,4)
Directa	161(59,9)	188(85,1)	87(46,8)	153(68)	150(48,9)	109(61,9)	195(52,6)	151(53,2)
Inversa	60(22,3)	26(11,8)	40(21,5)	46(20,4)	100(32,6)	37(21)	79(21,3)	81(28,5)
Otras			3(1,6)				3(0,8)	2(0,7)
D. Verbal	46(17,1)	2(0,9)	40(21,5)	14(6,2)	52(16,9)	24(13,6)	67(18,1)	29(10,2)
Total	269	221	186	225	307	176	371	284

En la Tabla 4.5.5 observamos que la mayoría de tareas corresponden a asociación directa, destacando [H2] y [T2] con un 85%. Le sigue la asociación inversa, y aquellas en que no es posible determinar su sentido. Destacamos [H3] y [T3] por el equilibrio en cuanto a los tipos de dependencia. Los resultados coinciden con los de Sánchez Cobo (1999), aunque menos de la mitad de los textos analizados por el autor presentan situaciones de independencia, y en cuatro textos ni se incluye como contenidos la diferencia entre correlación directa e inversa. Como el autor, pensamos que es importante una distribución de tareas más equilibrada en cuanto al signo de la correlación, para corregir (caso que se diese) la concepción unidireccional de la asociación, descrita por Estepa (1994), consistente en no aceptar la correlación inversa.

Función de ajuste

Una última clasificación de las tareas es considernado la función de ajuste, no analizada por Sánchez Cobo (1999) ni Lavalle, Micheli y Rubio (2006), por lo que constituye una aportación de nuestra investigación. Para cada tarea se determina la función de ajuste a los datos, considerando tanto su tendencia, como el coeficiente de determinación. Los resultados (Tabla 4.5.6) incluyen la categoría “otras” para las tareas donde no hemos podido especificar la función de ajuste. Observamos la enorme cantidad de tareas diseñadas para estudiar la dependencia lineal, por ser la más comprensible por los estudiantes y estar explícitamente recogida en las directrices curriculares. Sin embargo, ya que los alumnos conocen otros tipos de funciones, sería bueno incluirlas también, al menos como ejemplo.

Tabla 4.5.6. Frecuencias (y porcentajes) de tareas según función de ajuste

		H1	H2	H3	H4	H5	H6	H7	H8
Independencia		2(0,7)	5(2,3)	24(9,3)	12(5,3)	5(1,6)	6(3,4)	30(7,4)	21(7,1)
Funcional	Lineal	12(4,5)	4(1,8)	17(6,6)	14(6,2)	6(1,9)	3(1,7)	31(7,7)	18(6,1)
	Polinómica	3(1,1)		2(0,8)				4(1)	
	Exponencial								6(2)
	Potencia/Radical			3(1,2)		1(0,3)			
Aleatoria	Lineal	205(76,5)	128(57,9)	160(62)	149(66,2)	242(76.1)	141(80.1)	257(63.8)	185(62.3)
	Tukey								6(2)
	Exponencial			14(5,4)	12(5,3)	4(1,3)			16(5,4)
	Logarítmica					3(0,9)			
	Polinómica		82(37,1)	1(0,4)	24(10,7)	6(1,9)	2(1,1)	11(2,7)	16(5,4)
	Otras			3(1,2)					
D. Verbal		46(17,2)	2(0,9)	34(13,2)	14(6,2)	51(16)	24(13,6)	70(17,4)	29(9,8)
Total		268	221	258	225	318	176	403	297

		T1	T2	T3	T4	T5	T6	T7	T8
Independencia		2(0,7)	5(2,3)	16(8,6)	12(5,3)	5(1,6)	6(3,4)	27(7,3)	21(7,4)
Funcional	Lineal	12(4,5)	4(1,8)	10(5,4)	14(6,2)	6(2)	3(1,7)	28(7,5)	24(8,5)
	Polinómica	3(1,1)		2(1,1)				4(1)	
	Exponencial								6(2,1)
	Potencia/Radical			3(1,6)		1(0,3)			
Aleatoria	Lineal	206(76,6)	128(57,9)	106(57)	149(66.2)	231(75.2)	141(80.1)	234(63.1)	170(59.9)
	Tukey								5(1.8)
	Exponencial			5(2,7)	12(5,3)	4(1,3)			13(4.6)
	Logarítmica					3(1)			
	Polinómica		82(37,1)	1(0,5)	24(10,7)	5(1,6)	2(1,1)	11(2,9)	16(5.6)
	Otras			3(1,6)					
D. Verbal		46(17,1)	2(0,9)	40(21.5)	14(6,2)	52(16,9)	24(13,6)	67(18,1)	29(10,2)
Total		269	221	186	225	307	176	371	284

La siguente categoría de modelo de ajuste a los datos (salvo en [H2] y [T2]) es la descripción verbal, (Ver tarea 56 de la Figura 4.5.7), y aquella en que se tratan variables independientes. En cuanto al tratamiento del tipo de dependencia no lineal, destacamos los textos [H8], [H3], [H5], [H4], [T8], [T3], [T5] y [T4], por su variedad, frente a los textos [H1] y [T1], donde prácticamente ni se hace referencia a otro tipo de dependencia que no sea lineal. En general, se suelen proponer como situaciones de dependencia no lineal, las que se refieren a funciones exponenciales o polinómicas (grados 2 y 3).

4.6. ANÁLISIS DEL LENGUAJE

Desde el enfoque ontosemiótico (EOS), se postula que las prácticas matemáticas están mediadas por el lenguaje, que es, la parte ostensiva del resto de objetos matemáticos, luego tiene carácter representacional y operativo (Godino, Batanero y Font, 2007). En consecuencia, los estudiantes deben adquirir un dominio del lenguaje para entender los problemas que se les plantean, resolver las tareas, comunicar y justificar sus resoluciones.

Orton (1990) indica que el vocabulario, símbolos, y otras representaciones, pueden afectar al aprendizaje de las matemáticas. En este sentido, y como indican Cordero y Flores (2007), el discurso matemático escolar viene determinado por el lenguaje del libro de texto, el cual llega a regular, en determinadas ocasiones, las acciones de enseñanza y aprendizaje sobre un tópico. Los resultados que a continuación se presentan se refieren a variables utilizadas en investigaciones previas como la desarrollada por Ortiz, Batanero y Serrano (2001): términos y expresiones verbales; notación simbólica y expresiones algebraicas; y representaciones tabulares y gráficas. Parte de ellos se han publicado en Gea, Contreras, Arteaga y Cañadas (2012) y Gea et al., (en prensa).

4.6.1. LENGUAJE VERBAL

El primer componente analizado es el lenguaje verbal, es decir, las palabras y expresiones empleadas en los libros para denotar los conceptos, propiedades, y plantear o describir problemas, procedimientos y justificaciones o argumentos. Su interés fue resaltado por Orton (1990), quien señala la importancia de que sea apropiado, para que los alumnos le atribuyan un significado acorde con el que pretendemos. En nuestro análisis, encontramos una variedad de términos asociados a la correlación y regresión, que dividimos en dos grupos. Por un lado, hay *términos básicos*, que debiera conocer el alumno al iniciar el tema, como por ejemplo, *intervalo*, que se usa en el estudio de las tablas estadísticas de datos agrupados. Por otro lado, se muestran los *términos específicos* utilizados en el desarrollo del tema, como por ejemplo, *covarianza*.

Presentamos en la Tabla 4.6.1 los términos básicos y específicos encontrados en los textos analizados. De cada tipo se ha encontrado una amplia variedad, que muestran la riqueza conceptual y la complejidad con que se desarrolla la enseñanza de estos tópicos. Además, entre los términos no específicos, encontramos algunos que hacen referencia a conceptos geométricos (como baricentro, bisectriz, ángulo), algebra y funciones (función, ecuación), pensamiento numérico (proporcionalidad, porcentaje), probabilidad o estadística, por lo que observamos la implicación de diferentes bloques de contenidos estudiados por el estudiante previamente para la adquisición de las nociones de correlación y regresión.

A esta variedad, cabe añadir aquellos términos o expresiones que se utilizan con sentido matemático diferente a su uso en la vida ordinaria. En este sentido, Rothery (1980) diferencia tres tipos de expresiones en la enseñanza de las matemáticas: (a) Términos matemáticos específicos que, normalmente, no forman parte del lenguaje cotidiano; (b) Palabras usadas en matemáticas y el lenguaje ordinario, aunque no con el mismo significado; y (c) Palabras con significados iguales o muy próximos en ambos contextos. Indica que los problemas de aprendizaje tiendan a aparecer, sobre todo, con las dos primeras categorías. Un desafío para el profesor sería alcanzar una buena conexión entre ambos, ya que los términos matemáticos tienen mayor precisión que el lenguaje ordinario, ya que proporcionan definiciones necesarias y suficientes, mientras que el lenguaje ordinario es simplemente descriptivo (Schleppegrell, 2007).

Tabla 4.6.1. Términos básicos y específicos en los textos

Términos no específicos del tema	Términos específicos del tema
Amplitud de intervalo; ángulo de dos rectas; área, baricentro, bisectriz; azar; cambio de variable; carácter, característica; coeficiente de variación; coordenada; crecimiento; cuadrante; datos agrupados; desviación típica, media; diagrama de barras y barras adosadas; diagrama de sectores; dispersión; distancia; distribución; ecuación; ecuación punto-pendiente; elemento; expresión algebraica; ejes cartesianos; encuesta; escala; estadística; estimación; extrapolación; extremos de intervalo; fiabilidad; frecuencia (absoluta, acumulada, relativa, relativa acumulada); función; histograma; individuo, interpolación; intervalo (de clase, modal, mediano); marca de clase; máximo; media aritmética; medidas de centralización; mediana; mínimo; moda; muestra; ordenada; origen de coordenadas; paralelepípedo; parámetro; pendiente de una recta; perpendicular; población; polígono de frecuencias; porcentaje; prisma; probabilidad; proporcionalidad; punto medio; raíz cuadrada; rango, recorrido; recta, signo, sentido, subíndice; sumatorio; tabla de datos (de frecuencias); tangente; tendencia; valor absoluto; variable estadística (cualitativa, cuantitativa continua, discreta); varianza; volumen.	Bondad del ajuste; centro de gravedad; coeficiente de correlación de Pearson; coeficiente de determinación, coeficiente de regresión; correlación (curvilínea, espuria, perfecta, lineal); covarianza; dependencia (funcional, estadística); desviación típica marginal; diagrama de barras tridimensional; diagrama de dispersión; distribución (conjunta, marginal, condicional); error de estimación; estadística bidimensional; frecuencia conjunta (absoluta, relativa, condicionada, marginal); grado de dependencia; histograma tridimensional; incorrelada; independencia; intensidad, media marginal; método de mínimos cuadrados; nube de puntos; pictograma tridimensional; recta (de regresión; mínimos cuadrados; Tukey); regresión (lineal, exponencial, logarítmica, cuadrática, parabólica, potencial); tabla de doble entrada, valor (esperado, estimado, observado), variable (dependiente, independiente); variable estadística bidimensional; variación conjunta, varianza conjunta, varianza marginal.

Tabla 4.6.2. Términos coloquiales con significado matemático en los textos

Expresión coloquial	Término matemático al que alude
Estatura normalita ([H1], p.225) ([T1], p.331)	Estatura media
Según lo apretados que estén los puntos ([H1], p. 227) ([T1], p.333)	Dispersión
Rectas que "se acoplan bien" a la nube de puntos ([H1], p. 230) ([T1], p.336); la recta de regresión se amolda a la nube de puntos ([H1], p. 231) ([T1], p.337); la nube de puntos se condensa en torno a ([H8], p. 252) ([T8], p.322)	Ajuste lineal a la nube de puntos casi perfecto
Hinchar los puntos proporcionalmente a su frecuencia ([H1], p. 233) ([T1], p.339)	Representar circunferencias con diámetro proporcional a la frecuencia
Los puntos de la nube están completamente en desorden. ([H3], p.222) ([T3], p.272); los puntos del diagrama están esparcidos al azar ([H8], p.252) ([T8], p.322)	Están muy dispersos
Cómo se apartan a la vez las dos coordenadas de un dato respecto de la media ([H8], p. 251)	Covariación / variación conjunta
La nube de puntos es estrecha/ancha ([H7], p.247) ([T7], p.305)	Los puntos presentan más o menos dispersión
Se puede estimar (apostar, suponer) su estatura, con una certeza probable ([H5], p.250) ([T5], p.357)	Se puede estimar su estatura con una cierta probabilidad
Una nube de puntos alargada indica correlación lineal.	Si la nube de puntos se distribuye en torno a una recta existe correlación lineal
La estrechez de la nube expresa que la correlación es fuerte ([H5], p. 252) ([T5], p.359)	La dispersión en la nube de puntos informa de la intensidad de la correlación
Siempre que no se exagere en la extrapolación de resultados ([H5], p. 264) ([T5], p.370)	Siempre que la estimación se realice en valores próximos a la media
los datos no están demasiado concentrados ([T7], p.312)	Los datos presentan gran dispersión

Términos coloquiales con significado matemático. En segundo lugar, en la Tabla 4.6.2 mostramos una clasificación de los términos coloquiales más destacados, utilizados en algunos textos bajo un lenguaje impreciso, y en algunos casos inapropiado, que se usan como equivalentes a algunos conceptos. Se han encontrado términos y expresiones coloquiales utilizadas para aludir a conceptos u objetos matemáticos, en su mayoría con la intención de disminuir la formalización del enunciado. Este uso podría llevar, de acuerdo a Thompson y Rubenstein (2000), a imprecisiones en el uso de estas nociones por parte del estudiante, o incluso a generar conflictos semióticos.

4.6.2. LENGUAJE SIMBÓLICO

Las notaciones simbólicas y expresiones algebraicas permiten una comunicación comprimida entre individuos, pudiendo trabajar a un alto nivel de complejidad. Skemp (1980) diferencia las siguientes funciones de los símbolos matemáticos: comunicación, registro del conocimiento, formación de nuevos conceptos, confección de clasificaciones múltiples correctas, explicación, facilitar la actividad reflexiva, mostrar estructuras matemáticas, automatizar las manipulaciones rutinarias, recuperar información de la memoria y generar una actividad creativa. Al igual que Ortiz (1999), hemos encontrado una gran variedad de símbolos, que incluyen la notación funcional, subíndices y superíndices, con frecuencias variables (Tablas 4.6.3 y 4.6.4).

Tabla 4.6.3. Notación simbólica (modalidad Humanidades y Ciencias Sociales)

Notación	Concepto representado	H1	H2	H3	H4	H5	H6	H7	H8
$\sum$	Sumatorio sin subíndices	x	x		x	x	x		x
$\sum_{i=1}^{n} \sum_{i}$; $\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{ij}$	Sumatorio, doble sumatorio con subíndices			x	x			x	x
x_i ; n_i , f_i $\bar{x}$; $(\bar{x}, \bar{y})$	Valor de una variable *X*; frecuencia absoluta ;media; centro de gravedad	x	x	x	x	x	x	x	x
F_i ; h_i ; H_i	Frecuencia absoluta acumulada; relativa y relativa acumulada						x	x	
n_{ij} f_{ij} ; h_{ij}	Frecuencia absoluta y relativa de (x_i, y_j)			x	x			x	x
(x_i, y_i)	Valor de la variable bidimensional	x	x			x	x	x	x
(x_i, y_j)	Valor de la variable bidimensional			x	x			x	x
σ_x S_x ; σ_x^2 S_x^2	Desviación típica y varianza de *X*	x	x	x	x	x	x	x	x
σ_{xy} S_{xy} ; r r_{xy}	Covarianza y coeficiente de correlación lineal de *X* e *Y*	x	x	x	x	x	x	x	x
r^2	Coeficiente de determinación					x	x		
$\lvert \cdot \rvert$	Valor absoluto	x	x			x			x
m_{yx}	Pendiente recta de regresión de *Y* sobre *X*	x	x						
d_i ; d_i' ; e_i	Distancia entre ordenada (abcisa) de un punto y una recta	x	x			x		x	x
$\hat{x}(y_0)$ $\hat{x}_{y_0}$ $\hat{x}$	Valor estimado de X para y_0	x		x		x			x
$\approx$	Aproximadamente igual a	x		x		x	x	x	x
$<, >; \leq, \geq$	Menor, mayor menor (mayor) o igual		x	x	x	x	x	x	x
%	Porcentaje			x	x	x	x	x	X
$\ln y$, e^x , 2^x , t^2	Función logarítmica; exponencial, cuadrática					x			X

Llamamos la atención al uso que se lleva a cabo de las letras griegas en los textos analizados ya que, en estadística, las letras griegas están restringidas al trabajo con variables aleatorias (estudio de poblaciones o inferencias sobre ellas), mientras que las variables estadísticas y sus resúmenes se denotan con letras latinas. En el primer curso de Bachillerato, el estudio de la estadística es descriptivo, y no hay una intención inferencial, por lo que los símbolos debieran utilizar letras latinas. Este punto no es intranscendente, pues en el segundo curso del Bachillerato de Ciencias Sociales, los mismos alumnos estudiarán inferencia, y encontrarán expresiones conocidas del curso anterior, con diferente notación (letras griegas para referirse a las características de las poblaciones y latinas para nombrar las mismas características en las muestras). Este es el caso de los textos [H2], [H4], [H5], [H8], [T2], [T4], [T5] y [T8], en los que se usa la letra griega *S* para denotar a la desviación típica, varianza o covarianza (con diferentes subíndices), aunque en las tablas de resultados no se ha considerado esta distinción.

Como se puede observar (Tablas 4.6.3 y 4.6.4) se presenta gran cantidad de notación científica en la correlación y regresión, que requiere de un dominio del lenguaje simbólico específico. Del análisis realizado, encontramos diferencias de unos textos a otros, ya que hay textos muy completos como [H8] y [T8] y otros que se limitan al uso de la notación necesaria y suficiente para el desarrollo del tema, como por ejemplo [H2], [H4], [T2] y [T4].

Tabla 4.6.4. Notación simbólica (modalidad Ciencias y Tecnología)

Notación	Concepto representado	T1	T2	T3	T4	T5	T6	T7	T8
$\sum$	Sumatorio sin subíndices	x	x		x	x	x		x
$\sum_{i=1}^{n}$ $\sum_{i1}$; $\sum_{i=1}^{n}\sum_{j=1}^{m}$ $\sum_{ij}$	Sumatorio o doble o con subíndices			x	x			x	x
x_i ; n_i , f_i $\bar{x}$; $(\bar{x},\bar{y})$	Valor de una variable *X*; frecuencia absoluta; Media; Centro de gravedad	x	x	x	x	x	x	x	x
F_i ; h_i ; H_i	Frecuencia absoluta acumulada; relativa y relativa acumulada						x	x	
n_{ij} f_{ij} ; h_{ij}	Frecuencia absoluta y relativa			x	x			x	x
(x_i, y_i)	Valor de la variable bidimensional	x	x			x	x	x	x
(x_i, y_j)	Valor de la variable bidimensional			x	x				x
σ_x S_x ; σ_x^2 S_x^2	Desviación típica, varianza de *X*	x	x	x	x	x	x	x	x
σ_{xy} S_{xy} ; r r_{xy}	Covarianza y coeficiente de correlación lineal de *X* e *Y*	x	x	x	x	x	x	x	x
r^2	Coeficiente de determinación					x	x		x
$\lvert\cdot\rvert$	Valor absoluto	x	x			x		x	x
m_{yx}	Pendiente de recta de regresión de *Y* sobre *X*	x	x						x
d_i ; d_i' ; e_i	Distancia entre ordenada (abcisa) de un punto y una recta	x	x			x		x	x
$\hat{x}(y_0)$ $\hat{x}_{y_0}$ $\hat{x}$	Valor estimado de *X* para y_0	x		x		x			x
$\approx$	Aproximadamente igual a…	x		x		x	x	x	x
$<, >$; $\leq, \geq$	Menor, mayor que		x	x	x	x	x	x	x
%	Porcentaje			x	x	x	x	x	x
$\ln y$, e^x , 2^x , t^2	Función logarítmica; exponencial, cuadrática					x			x

En algunos textos, y principalmente en la modalidad científico-tecnológica, se incluye el tratamiento unidimensional de una variable estadística. Así aparece la notación $x_{\max}$ y $x_{\min}$ en [H3] y [T3] para referirse a los valores máximo y mínimo de la variable X, con el fin de agrupar los datos en intervalos; $P(x_0, y_0)$ en [H8], para referirse a un punto en el diagrama cartesiano, al recordar la ecuación de una recta mediante el procedimiento punto-pendiente; *Mo* y *Me* para referirse a la moda y mediana en los textos [H6], [T6], [T7] y [T8]; o CV_x para referirse al coeficiente de variación de la variable X en [H6], [T6] y [T7]. En cuanto a la modalidad de Bachillerato, las principales diferencias son los textos [T7] y [T8]. En [T7] encontramos un tratamiento más escaso de las frecuencias en la distribución bidimensional, por lo que se prescinde de la notación (x_i, y_j), mientras que en [T8] se incluye el tratamiento del coeficiente de determinación y las pendientes de las rectas de regresión.

Se presentan en las Tablas 4.6.5 y 4.6.6, las expresiones algebraicas que se incluyen en los textos analizados. Se refieren principalmente al cálculo de la media, varianza, desviación típica, covarianza, coeficiente de determinación, rectas de regresión y pendiente de dichas rectas, encontrando pocas diferencias entre los textos. Se debiera añadir dos que se refieren al análisis unidimensional de una variable aleatoria como es la expresión $L_i + a\dfrac{\frac{n}{2} - F_{i-1}}{f_i}$ que se incluyen en [H6] y [T6], y la expresión $\dfrac{\sum_{i=1}^{n} f_i \left| x_i - \bar{x} \right|}{N}$ que se incluye en [T7].

Tabla 4.6.5. Expresiones algebraicas (modalidad Humanidades y Ciencias Sociales)

Expresión algebraica	H1	H2	H3	H4	H5	H6	H7	H8
$\dfrac{\sum x_i}{n}$	x				x	x		x
$\dfrac{\sum n_i x_i}{N}$	x	x	x	x	x	x		x
$\sqrt{\dfrac{\sum x_i^2}{n} - \bar{x}^2}$; $\sqrt{\dfrac{\sum (x_i - \bar{x})^2}{N}}$	x				x	x		x
$\sqrt{\dfrac{\sum n_i (x_i - \bar{x})^2}{N}}$; $\sqrt{\dfrac{\sum n_i x_i^2}{N} - \bar{x}^2}$		x	x	x		x		
$\dfrac{\sigma}{\bar{x}}$						x		
$\dfrac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$; $\dfrac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}$	x				x	x	x	x
$\dfrac{\sum n_{ij} (x_i - \bar{x}) \cdot (y_j - \bar{y})}{N}$; $\dfrac{\sum x_i y_i n_{ij}}{N} - \bar{x} \cdot \bar{y}$ $\dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$	x	x	x	x	x	x	x	x
$y - \bar{y} = \dfrac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$; $x - \bar{x} = \dfrac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$	x	x	x	x	x	x	x	x
$\dfrac{\sigma_{xy}}{\sigma_x^2}$; $\dfrac{\sigma_{xy}}{\sigma_y^2}$	x	x	x	x	x		x	x

Respecto a cada una de las modalidades, destacamos los textos [T7] y [T8], porque presentan más expresiones algebraicas debido a que incluyen secciones relativas al tratamiento unidimensional de una variable aleatoria, que las respectivas editoriales en la modalidad de *Humanidades y Ciencias Sociales*.

Tabla 4.6.6. Expresiones algebraicas (modalidad Ciencias y Tecnología)

Expresión algebraica	T1	T2	T3	T4	T5	T6	T7	T8
$\frac{\sum x_i}{n}$	x				x	x		x
$\frac{\sum n_i x_i}{N}$ $\frac{\sigma_{xy}}{\sigma_x \sigma_y}$	x	x	x	x	x	x	x	x
$\sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$; $\sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$	x				x	x		x
$\sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{N}}$; $\sqrt{\frac{\sum n_i x_i^2}{N} - \bar{x}^2}$		x	x	x		x	x	x
$\frac{\sigma}{\bar{x}}$						x	x	
$\frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$; $\frac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}$	x				x	x	x	x
$\frac{\sum n_{ij} (x_i - \bar{x}) \cdot (y_j - \bar{y})}{N}$; $\frac{\sum x_i y_i n_{ij}}{N} - \bar{x} \cdot \bar{y}$	x	x	x	x	x	x	x	x
$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$; $x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$	x	x	x	x	x	x	x	x
$\frac{\sigma_{xy}}{\sigma_x^2}$; $\frac{\sigma_{xy}}{\sigma_y^2}$	x	x	x	x	x		x	x

4.6.3. REPRESENTACIÓN TABULAR

Una representación característica de la estadística es la tabla de frecuencias, que ofrecen una estructura particular de la información, presentando, no sólo números, sino las relaciones entre los valores de la variable y sus diferentes tipos de frecuencia (Ortiz, 1999). Se requiere conocer sus convenios de construcción, pues no todos los números incluidos en el cuerpo de la tabla tienen el mismo significado; por ejemplo, hay que diferenciar los totales de los valores particulares. Todos los libros analizados reconocen su importancia en la organización de datos, aunque su tratamiento varía.

Tabla bidimensional simple (Listado de datos)

La representación tabular más utilizada es el listado de datos en el que cada fila (o columna) de la tabla representa los datos de cada una de las variables observadas en cada individuo de la muestra. Es denominada en algunos manuales *tabla de frecuencias bidimensional simple* ([H8] y [T8]). Es común, avanzado el tema, que se vayan añadiendo columnas según la necesidad de los cálculos. Por ejemplo, en la Figura 4.6.1 se incluye una tabla bidimensional simple en la que se han añadido columnas con los cuadrados del valor de las variables, y con el producto de las mismas, que son útiles en el cálculo posterior de la covarianza y el coeficiente de correlación.

x_i	y_i	x_i^2	y_i^2	x_iy_i
1	10	1	100	10
2	17	4	289	34
3	30	9	900	90
4	28	16	784	112
5	39	25	1 521	195
6	47	36	2 209	282
21	171	91	5 803	723

Figura 4.6.1. Tabla bidimensional simple con cálculos auxiliares ([H1], p. 235)

La siguiente tabla representa los pesos y alturas de 20 alumnos de 1.º de Bachillerato:

N.º alumnos	4	3	2	5	4	2
Peso (kg)	73	76	73	78	80	82
Altura (m)	1,65	1,68	1,70	1,72	1,76	1,80

a) ¿Cómo están correlacionados los datos?

b) ¿Cuál será la altura estimada para un alumno de este colectivo que pese 75 kg?

Figura 4.6.2. Datos en representación tabular con frecuencias ([T6], p. 320)

Tabla bidimensional simple con frecuencias

Es una extensión del listado de datos, donde se añade una nueva columna/fila con la frecuencia que corresponde a cada par de valores (x_i, y_i), representados en la tabla (Figura 4.6.2).

Tabla de doble entrada

Es la representación tabular más habitual para organizar los datos de un estudio bidimensional, principalmente cuando se dispone de una muestra de gran tamaño. Consiste en una tabla con tantas filas y columnas como modalidades presenten las variables que conforman la distribución bidimensional; y en cada una de las intersecciones que se realicen entre fila (por ejemplo $Y = y_j$) y columna (por ejemplo, $X = x_i$) se encuentra la frecuencia del dato representado por dicha modalidad (x_i, y_j) (Ver ejemplo en la Figura 4.5.1). Su dificultad de interpretación es debida a que, mientras que en una tabla bidimensional simple, de cada celda solo se deduce una frecuencia relativa, en el caso de las tablas bidimensionales, se deducen varios tipos de frecuencias:

- *Frecuencia relativa doble:* $h_{ij} = \dfrac{f_{ij}}{N}$

- *Frecuencia relativa respecto a su columna:* $h(y_j / X = x_i) = \dfrac{f_{ij}}{f_{i\bullet}} = \dfrac{h_{ij}}{h_{i\bullet}}$

- *Frecuencia relativa respecto a su fila:* $h(x_i / Y = y_j) = \dfrac{f_{ij}}{f_{\bullet j}} = \dfrac{h_{ij}}{h_{\bullet j}}$

También es posible calcular dos tipos de frecuencias relativas marginales. Todo ello hace compleja la interpretación de estas tablas (Estepa ,1994; Cañadas, 2012). En las Tabla 4.6.7 observamos que el tratamiento de estas representaciones difiere según la modalidad de Bachillerato, así como en los textos en cada una de ellas. En general, se presenta la tabla de doble entrada al comienzo del tema, como medio de organización de la información de los datos bidimensionales. Se suele acompañar a la definición de variable y/o distribución bidimensional, y a los diferentes tipos de frecuencias (Ver Sección 4.7.1). Sólo en [H1], [H6] y [T1], [T6] se muestra la tabla de doble entrada al final del tema, dentro de algunos ejercicios resueltos.

Tabla 4.6.7. Representación tabular en los textos

	Presencia en el tema	H1	H2	H3	H4	H5	H6	H7	H8
Tabla bidimensional	Desarrollo teórico y práctico del tema	x				x	x		
simple	Uso eminentemente práctico		x	x	x			x	x
Tabla bidimensional	Desarrollo teórico y práctico del tema		x		x				x
simple con frecuencias	Uso eminentemente práctico						x		
	Presencia anecdótica			x		x		x	
Tabla de doble entrada	Desarrollo teórico y práctico del tema		x	x	x			x	x
	Presencia anecdótica	x				x	x		

		T1	T2	T3	T4	T5	T6	T7	T8
Tabla bidimensional	Desarrollo teórico y práctico del tema	x				x	x		
simple	Uso eminentemente práctico		x	x	x			x	x
Tabla bidimensional	Desarrollo teórico y práctico del tema		x		x				
simple con frecuencias	Uso eminentemente práctico						x		
	Presencia anecdótica			x		x		x	x
Tabla de doble entrada	Desarrollo teórico y práctico del tema		x	x	x			x	x
	Presencia anecdótica	x				x	x		

Poco más de la mitad de los textos utilizan esta representación para el desarrollo del tema ([H2], [H3], [H4], [H7], [H8], [T2], [T3], [T4], [T7] y [T8]). Los textos [H3], [H7], [T3] y [T7] tratan, además, la agrupación de datos en intervalos en la tabla de doble entrada (Sección 4.7). Por el contrario, varios textos ([H1], [H5], [H6], [T1], [T5] y [T6]) basan su enseñanza de la correlación y regresión en tablas bidimensionales simples, haciendo un uso anecdótico de la tabla de doble entrada o la tabla bidimensional con frecuencias.

Este resultado muestra baja idoneidad epistémica de la enseñanza, pues no se presenta una muestra representativa de este objeto matemático del lenguaje. Además, de acuerdo a Arteaga (2011), las representaciones de datos que se limitan a un listado de los mismos, no llegan a representar explícitamente la distribución de la variable bidimensional, y tendrían menor complejidad semiótica que aquellas en que se han resumido las frecuencias, como la tabla bidimensional simple con frecuencias o la tabla de doble entrada. Ello impide responder a preguntas de nivel alto en la lectura de datos según Curcio (1999), pues en un listado de datos no se detectan con facilidad las tendencias de los datos.

En las tareas propuestas, los datos se muestran principalmente en tablas simples, suponemos que la intención es agilizar los cálculos. En algunos casos, para el cálculo de la covarianza o el coeficiente de correlación lineal, se utiliza la tabla bidimensional simple con frecuencias, pero por lo general, los textos presentan un único ejercicio resuelto con el uso de esta representación ([H3], [H5], [H7], [T3], [T5], [T7] y [T8]).

Resaltamos el hecho de que, como se verá en el análisis de los CDs (Sección 4.11), algunos libros ofrecen al alumno software para el tratamiento de la correlación y regresión, como Hojas de cálculo Excel u otros recursos. Por ello, el uso de las columnas auxiliares de cálculo es útil para comprender el sentido de los cálculos, pero innecesario, y también lo es restringir los datos a listados simples, ya que la tecnología facilitaría el cálculo en todos los casos.

Por último, los textos no muestran diferencias significativas en la representación tabular, en cuanto al tipo de modalidad de Bachillerato al que se refieren, a pesar de que sería lógico un mayor uso de la tabla de doble entrada en el Bachillerato de *Humanidades y Ciencias Sociales*, por el uso abundante que se hace de ella en estas ciencias. Tan sólo [H8] muestran un tratamiento diferenciado en cuanto a la tabla bidimensional simple con frecuencias, siendo paradójico que la modalidad científico tecnológica posea un tratamiento anecdótico de ésta.

4.6.4. REPRESENTACIÓN GRÁFICA

La representación gráfica alcanza un estatus privilegiado en tema, pues el modo habitual de definir la correlación es mediante un diagrama de dispersión; y es usual tratar la regresión a partir de ejemplos gráficos, mediante el trazado de la recta que mejor se ajuste a dicha nube de puntos. Se presentan a continuación las diferentes representaciones gráficas que hemos encontrado.

Diagrama de barras tridimensional

Esta representación permite visualizar en el espacio la distribución de la variable estadística bidimensional. En el plano *XY*, se representan cada uno de los datos bidimensionales, y sobre cada uno de ellos se levanta una barra de altura proporcional a su frecuencia (Figura 4.6.3). Generaliza el diagrama de barras que ya conocen los estudiantes de los cursos anteriores.

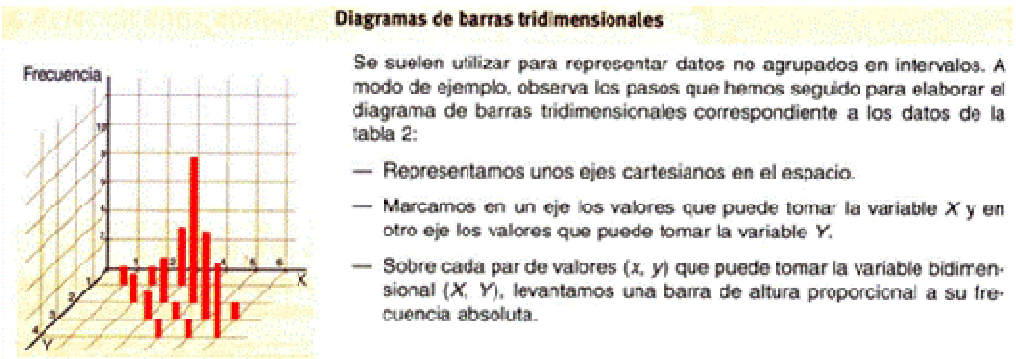


Figura 4.6.3. Diagrama de barras tridimensional ([H3], p.219)

Diagrama de dispersión y gráfico de burbujas

La representación gráfica más utilizada para mostrar la distribución bidimensional es el diagrama de dispersión o nube de puntos. En el plano, y mediante un eje de coordenadas, se representan los datos bidimensionales de la variable. Cuando los datos poseen frecuencia absoluta igual a la unidad se corresponden a puntos en el plano, y en otro caso, se puede optar por dibujar circunferencias con área proporcional a la frecuencia de cada dato (gráfico de burbujas) o bien, alrededor de donde se sitúa dicho dato, dibujar tantos puntos como indique su frecuencia absoluta (Figura 4.6.4).

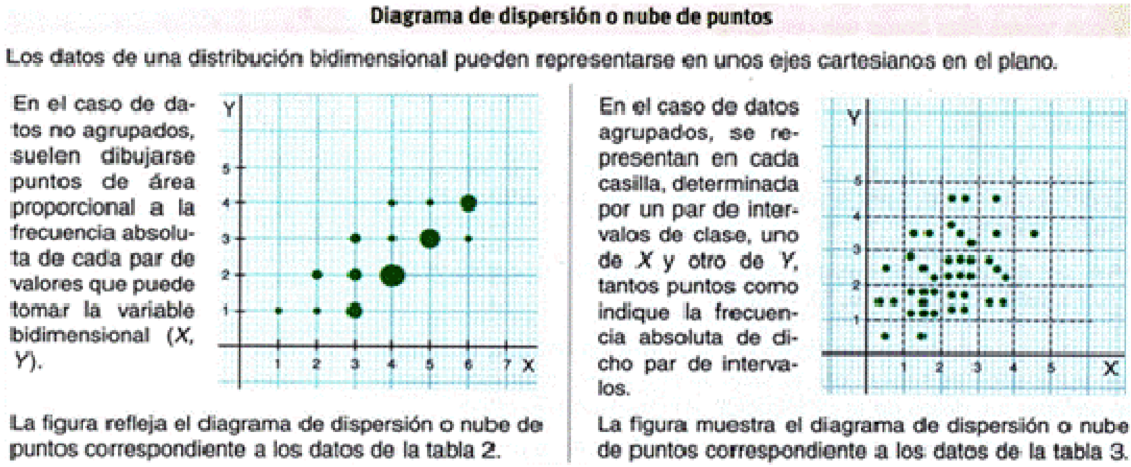


Figura 4.6.4. Diagrama de burbujas y diagrama de dispersión ([T3], p.270)

Estas representaciones son muy útiles para interpretar la relación entre las variables, ya que permite visualizar su intensidad (a través de la mayor o menor dispersión de la nube de puntos), su sentido (si la relación es directa o inversa) y el tipo (lineal o no), observando su tendencia (Sánchez Cobo, 1999). Además, el diagrama de burbujas permite visualizar en el plano, simultáneamente, hasta tres variables (el diámetro) o incluso cuatro, si mediante el color pudiera representarse una cuarta variable cualitativa.

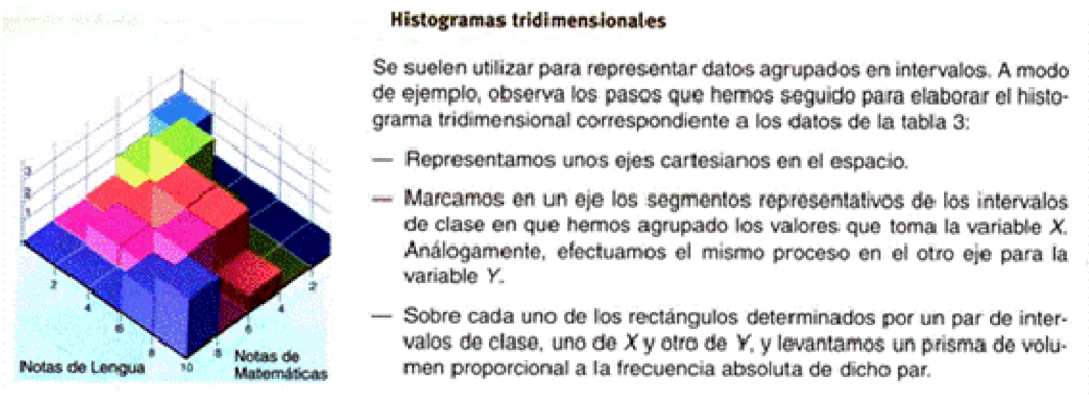


Figura 4.6.5. Histograma tridimensional ([H3], p. 220)

Histograma tridimensional

Se utiliza principalmente cuando las variables son continuas, y generaliza el histograma que los alumnos estudiaron en cursos anteriores. Permite visualizar en el espacio la distribución de la variable estadística bidimensional, situando en el plano *XY* cada pareja de intervalos de clase, y levantando sobre cada cruce un prisma recto con base rectangular, de volumen proporcional a su frecuencia absoluta (Figura 4.6.5). Como en el histograma simple, los intervalos pueden ser semiabiertos o cerrados.

Pictograma tridimensional

Permite visualizar la distribución de la variable estadística bidimensional en el espacio, a la vez que informa de su significado mediante la figura con la cual se representa. Es una variante del diagrama de barras e histograma tridimensional ya que, cada barra o prisma se sustituye por una figura cuyo volumen es proporcional a la frecuencia de los datos que representa (Figura 4.6.6). Es importante resaltar que la frecuencia se representa por el volumen, pues este punto se olvida a veces.

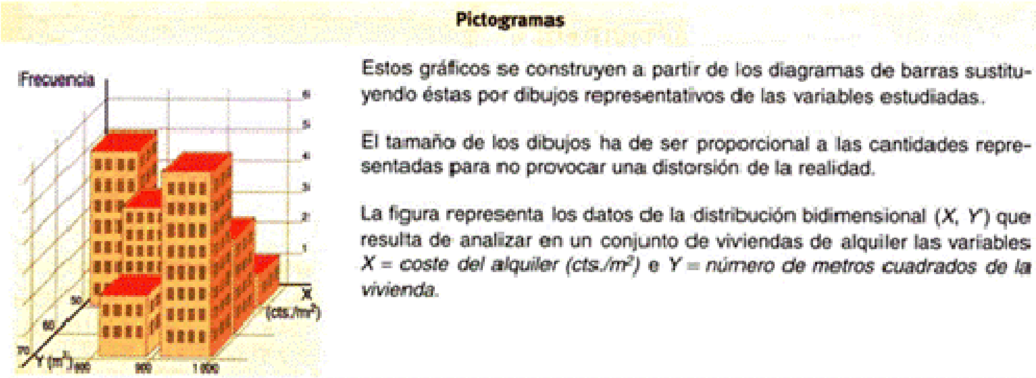


Figura 4.6.6. Diagrama de barras tridimensional ([H3], p.219)

En la Tabla 4.6.8 se muestra la relevancia de las representaciones gráficas en los textos analizados. Todos los textos presentan al menos un tipo de representación gráfica, aunque hay mucha variabilidad, desde los que presentan sólo el diagrama de dispersión, al que incluye todos los gráficos descritos. Sin embargo, hay poca precisión, en general, al describir las representaciones gráficas. Por ejemplo, no se suele distinguir el diagrama de barras del histograma ([H1], [H7], [T1] y [T7]). Por otro lado, ninguno de los textos que utilizan el gráfico de burbuja lo identifican como tal, siendo tratado como una extensión del diagrama de dispersión (Figura 4.6.7). Además, [H7] y [T7] señalan la importancia del grosor de los puntos dependiendo de la frecuencia de los datos, pero no se indica que este grosor debe ser proporcional al radio con que se representan. Estas imprecisiones debieran ser tenidas en cuenta por el profesor.

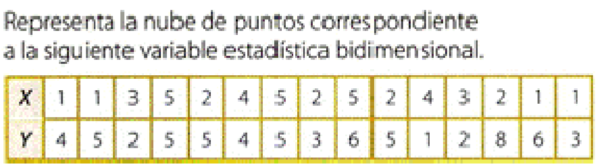
Podemos observar la destacada presencia del diagrama de dispersión en todos los libros, siendo en algunos la única representación gráfica utilizada ([H5], [H6], [H8] y [T5], [T6], [T8]). Menos de la mitad de los textos presentan el gráfico de burbujas, y los que lo hacen, es sólo con un ejemplo o tarea de aplicación. Este diagrama está especialmente indicado para distribuciones en que los datos posean frecuencia distinta a la unidad, que son las más comunes. Por este motivo se incluyó, en el análisis de la presencia del diagrama de dispersión, el descriptor "*Utiliza con frecuencias mayores a 1*", pues nos permite valorar si los textos contemplan este tipo de situaciones problemáticas en el tema, aunque no utilicen el gráfico de burbujas. Así es que, algunos textos como [H4], [H8], [T4] y [T8] omiten el tratamiento del gráfico de burbujas, pero en su lugar representan tantos puntos como frecuencia presenten los datos alrededor de las coordenadas de éstos.

Tabla 4.6.8. Representación gráfica en los textos

	Presencia en el tema	H1	H2	H3	H4	H5	H6	H7	H8
Diagramas de	Desarrollo teórico y práctico	x	x	x	x	x	x	x	x
dispersión	Utiliza con frecuencias mayores a 1			x	x			x	x
Gráfico de	Utiliza		x	x	x				
barras	Presencia anecdótica	x						x	
Gráfico de	Utiliza							x	
burbujas	Presencia anecdótica	x		x					
Histograma	Utiliza			x					
	Presencia anecdótica	x						x	
Pictograma	Presencia anecdótica			x					

		T1	T2	T3	T4	T5	T6	T7	T8
Diagramas de	Desarrollo teórico y práctico	x	x	x	x	x	x	x	x
dispersión	Utiliza con frecuencias mayores a 1			x	x			x	x
Gráfico de	Utiliza		x	x	x				
barras	Presencia anecdótica	x						x	
Gráfico de	Utiliza							x	
burbujas	Presencia anecdótica	x		x					
Histograma	Utiliza			x					
	Presencia anecdótica	x						x	

Por otro lado, los textos [H7] y [T7] plantean una tarea de representación del diagrama de dispersión (Figura 4.6.7), sin haber explicado al estudiante cómo afrontar el caso en que los datos posea frecuencia distinta a la unidad. No es hasta el final del tema que se muestra un ejemplo del gráfico de burbuja. Una situación parecida la encontramos en [T8], ya que no explica cómo representar datos con frecuencia mayor a la unidad, algo que [H8] sí hace, mientras que incluye una tarea en que el estudiante debe representar en un diagrama de dispersión una distribución bidimensional donde uno de los datos posee frecuencia dos.


Representa la nube de puntos correspondiente a la siguiente variable estadística bidimensional.

X	1	1	3	5	2	4	5	2	5	2	4	3	2	1	1
Y	4	5	2	5	5	4	5	3	6	5	1	2	8	6	3

Figura 4.6.7. Tarea de representación de diagrama de dispersión ([T7], p.305)

El tratamiento de la representación de la variable en tres dimensiones es pobre, reducida a un ejemplo o un ejercicio práctico. El gráfico de barras es el que más se utiliza, seguido del histograma, siendo [H3] el único que incluye el pictograma a modo de ejemplo. En cuanto a la poca rigurosidad con que se describen estas representaciones, destacamos los textos [H2] y [T2], que muestran el gráfico de barras sin explicar a qué se refiere la altura de cada barra, y luego proponen tareas al alumno.

No encontramos diferencias significativas entre las dos modalidades de Bachillerato, aunque los textos de la modalidad de *Ciencias y Tecnología* tienen un uso más pobre en cuanto a la representación de datos bidimensionales. Quizá porque algunos de ellos ([T7] y [T8]) incluyen el estudio de la variable estadística unidimensional, y el espacio que hay que sacrificar en el tema sea el del tratamiento de datos bidimensionales, algo que quizá también ocurre en [H6] y [T6].

4.7. ANÁLISIS DE LAS DEFINICIONES (CONCEPTOS)

Un elemento fundamental en la construcción del conocimiento matemático son los conceptos y propiedades involucrados en la resolución de los problemas. Conocimiento conceptual y procedimental se pueden considerar como polos de un continuo, siendo el primero flexible y más generalizable, ya que no está ligado a un tipo específico de problema; sino que incluye la comprensión implícita o explícita de los principios de un dominio dado y sus interrelaciones (Rittle-Johnson y Alibali, 1999; Rittle-Johnson, Siegler y Alibali, 2001).

Godino (2002) indica que un objeto matemático se caracteriza por su definición y el enunciado de sus propiedades (teoremas, proposiciones); pero éstas pueden variar según las distintas instituciones en que se trate, y por tanto, hemos de concederles un carácter relativo. Puesto que las *definiciones de conceptos* son evocadas por el estudiante cuando se enfrenta a la resolución de una situación problema, es importante analizar el tratamiento que se realiza de éstas en la enseñanza, ya que, la progresiva construcción de su significado depende directamente de los conceptos que se describan y utilicen.

Sfard (1991) describe un concepto como una idea matemática en su forma "oficial", es decir, constructo teórico correspondiente al universo matemático formal. Indica que se pueden definir de forma estructural (describiendo sus condiciones o propiedades) u operacional (cuando se describe mediante una operación o fórmula). En nuestro análisis tendremos en cuenta los dos tipos de definiciones.

Por otra parte, Skemp (1993) formuló dos principios para el aprendizaje de los conceptos matemáticos: los nuevos conceptos no pueden ser comunicados simplemente a través de una definición. En primer lugar, sería necesario proporcionar una colección apropiada de ejemplos, apoyados en los conceptos que ya se conocen. Así, Joyce y Weil (1996) consideran varios aspectos a tener en cuenta en la enseñanza de un concepto: (a) su nombre o etiqueta que se asocia al concepto, y que puede ser una palabra o símbolo; (b) sus atributos esenciales, o propiedades específicas, que lo diferencian de otros conceptos; (c) los ejemplos y contraejemplos del concepto; y (d) la regla de definición, que es una afirmación sintética y precisa elaborada a partir de los atributos esenciales que lo caracterizan. Vergnaud (1982), por su parte, considera el concepto matemático como una tripleta formada por el conjunto de situaciones que lo hacen significativo; el conjunto de invariantes que lo constituyen; y las representaciones simbólicas asociadas. En esta Sección nos limitamos a describir los invariantes asociados al concepto, dados por su definición, describiendo en la siguiente las propiedades que se asocian a éstos; las representaciones y las situaciones fueron analizadas anteriormente.

Sánchez Cobo (1999) estudió las definiciones relacionadas con la correlación y regresión en textos de Bachillerato, clasificándolas según se definan en forma estructural (Sfard, 1991), operacional o mezcla de las anteriores. En su estudio no diferencia entre conceptos y propiedades, sino que engloba las dos categorías como "contenidos". En nuestro trabajo los diferenciamos y clasificamos los primeros en varias formas:

- En primer lugar se categorizan todos los conceptos implicados en el tratamiento de la correlación y regresión hallados en el tema, en cada uno de los textos analizados.
- Seguidamente analizamos la *forma* en que se describe o presenta el concepto, que puede ser mediante ejemplos (E), usualmente utilizando implícitamente una

definición operacional, o mediante una definición explícita formal (F); que de acuerdo a Sfard (1991) puede ser operacional (O) o bien estructural (S), o bien mezclando varios de estos tipos de definiciones. Se ha tenido también en cuenta si los ejemplos se proponen antes de la definición, como es sugerido por Skemp (1993).

- A continuación se analiza el *momento* en que se formaliza una definición, diferenciando si se lleva a cabo al inicio (I), bien después de haber sido utilizado, al final del tema (F), o en diferentes momentos del tema (VM). En estos dos casos, el libro presentaría una orientación práctica-teoría, y al contrario en el primer caso. Esta variable se considera indicador de la orientación del libro, bien al enfoque teoría-práctica, dando más predominio a la primera, o al contrario (Lavalle et al., 2006).
- Por último se considera su *uso* en el tema, esto es, si su uso es continuado (S), si se usa poco (P), o más bien nada (N), para tener en cuenta la importancia dada al concepto en el desarrollo del tema.

Estas dos últimas variables no fueron consideradas por Sánchez Cobo (1999). Hemos agrupado las definiciones con objeto de facilitar su lectura en varios apartados, comenzando con los conceptos involucrados en el tratamiento de la variable estadística y distribución bidimensional (Estepa, 2007; Gea, Batanero, Cañadas y Arteaga, 2013). Seguidamente se dedican dos apartados a los conceptos relacionados con el análisis de la dependencia entre dos variables, y finalmente, los relativos al estudio de la regresión.

4.7.1. DISTRIBUCIÓN BIVARIANTE, MARGINAL Y CONDICIONAL

El trabajo con la correlación y regresión parte de la definición de variable estadística bidimensional, respecto a la cual hemos encontrado los conceptos que se analizan a continuación.

C1.Variable estadística y distribución bidimensional

En esta sección presentamos las definiciones de variable estadística bidimensional (C11), frecuencia doble (C12), y distribución bidimensional (C13). Resultados parciales de esta sección se han publicado en Gea, Contreras, Arteaga y Cañadas (2012) y Gea, Batanero, Fernandes, y Gómez (en prensa).

C11. Variable estadística bidimensional

Cinco textos en cada modalidad de Bachillerato ([H3], [H4], [H6], [H7], [H8], [T3], [T4], [T6], [T7] y [T8]) presentan una definición de variable estadística bidimensional, incluyendo el resto la definición de distribución bidimensional. Todos los textos introducen uno de estos dos conceptos al inicio del tema, aunque el modo de definirla varía de un texto a otro, siendo frecuente introducirla con ejemplos.

De acuerdo a Vergnaud (1990) y Joyce y Weil (1996), consideramos correctas estas definiciones cuando describen con detalle sus características, incluyendo también la notación *(X, Y)*, como el ejemplo mostrado en la Figura 4.7.1. En este sentido, todas las definiciones encontradas en los textos son correctas, salvo la que ofrecen los textos

[H8] y [T8], que limita el concepto de variable estadística bidimensional a variables estadísticas unidimensionales que sean cuantitativas ([T8], p.320):

> Una variable estadística bidimensional (X,Y) es el resultado del estudio de dos características cuantitativas X e Y en los individuos de una población.

Si A y B son dos caracteres que se han de estudiar, con el carácter A queda definida una variable unidimensional X de valores $x_1, x_2, ..., x_n$, y con el B, otra variable unidimensional Y de valores $y_1, y_2, ..., y_m$.

Partiendo de ellas, se define la **variable bidimensional** (X , Y) que toma los valores:

$$(x_i, y_j) \text{ con } 1 \le i \le n \text{ y } 1 \le j \le m$$

Así, a cada individuo de la población se le asignará un par (x_i, y_j); con estos pares se obtiene una serie estadística doble, que permite la elaboración de una **tabla de frecuencias absolutas de doble entrada**, como la que tienes en el margen.

X \ Y	y_1	y_2	y_3	...	y_m	$\sum_{j=1}^{m} f_{ij}$
x_1	f_{11}	f_{12}	f_{13}	...	f_{1m}	A_1
x_2	f_{21}	f_{22}	f_{23}	...	f_{2m}	A_2
x_3	f_{31}	f_{32}	f_{33}	...	f_{3m}	A_3
...	...	...	...	...	...	...
x_n	f_{n1}	f_{n2}	f_{n3}	...	f_{nm}	A_n
$\sum_{i=1}^{n} f_{ij}$	B_1	B_2	B_3	...	B_m	N

Figura 4.7.1. Definición de variables estadística bidimensional con ejemplo ([H4], p. 218)

C12. Frecuencias dobles

Introducida la variable estadística bivariada, todos los textos definen los diferentes tipos de frecuencia, salvo [H6], [T6], y [T7], siendo habitual introducirla mediante la forma operacional del concepto (Sfard, 1991), como se muestra a continuación:

> Hallamos la frecuencia absoluta de cada par de valores de la variable (X,Y). Para ello contamos el número de veces que se repite ese par de valores en la distribución y lo anotamos en la casilla correspondiente ([H3], p.217).

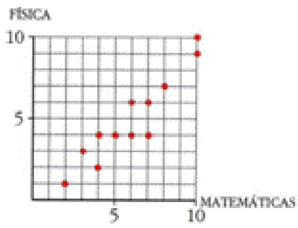
Las definiciones suelen ser parcialmente correctas ([H1], [H2], [H3], [H5], [T1], [T2], [T3] y [T5]) al no incluir la notación, aunque se describan sus propiedades o su forma de cálculo. Un ejemplo de definición correcta se presenta en [H4], que se muestra a continuación, siendo este texto junto con [H7] los únicos que definen también la frecuencia conjunta relativa de datos bidimensionales. Además, en este ejemplo se presenta la relación entre frecuencias absolutas y relativas dobles ([H4], p.218):

> Frecuencia absoluta conjunta, f_{ij}: Es el número de veces que se pr*ese*nta el par (x_i,y_j)
>
> Frecuencia relativa del par (x_i,y_j): Es el valor del cociente $(f_r)_{ij} = \frac{f_{ij}}{N}$, y se cumple que la suma de las frecuencias relativas de todos los pares de observaciones es 1.

Advertimos de la omisión de esta definición en el texto [T7] de la modalidad científico-tecnológica, que sí se presenta correctamente en [H7], incluso estableciendo la relación entre frecuencias absolutas y relativas dobles. Suponemos que esta omisión se debe a que [T7] incluye conceptos relativos al tratamiento de datos unidimensionales y presta menos atención a la distribución de datos bidimensionales, quizá por falta de espacio. Sin embargo, la utiliza en un ejercicio resuelto ([T7], p.314) sobre representación gráfica de una variable bidimensional mediante un histograma (ejemplo que también aparece en [H7]), luego parece considerarse transparente para el alumno.

C13. Distribución bidimensional

Todos los textos analizados la usan explícitamente a pesar de que la mitad no la definan, posiblemente por considerarla equivalente al concepto de variable estadística bidimensional ([H4], [H6], [H7], [H8], [T4], [T6], [T7] y [T8]). Esta hipótesis fue sugerida por Sánchez Cobo (1999), quien encontró algunos textos que no diferenciaban estos conceptos Puesto que no son totalmente equivalentes, pensamos que sería útil diferenciarlos, ya que dar la variable solo implica conocer su rango de variación y el significado de cada variable, mientras la distribución requiere también la frecuencia de cada par de valores.



■ Estas son las notas de 12 estudiantes en Matemáticas y en Física:

ALUMNO	a	b	c	d	e	f	g	h	i	j	k	l
MATEMÁTICAS	2	3	4	4	5	6	6	7	7	8	10	10
FÍSICA	1	3	2	4	4	4	6	4	6	7	9	10

Es una **distribución bidimensional**, porque a cada individuo le corresponden los valores de dos variables. Si tomamos esos dos valores como las coordenadas de un punto, la distribución puede ser representada mediante 12 puntos: **nube de puntos**.

Se aprecia una relación entre las dos variables: a mejor nota en *Matemáticas*, mejor nota en *Física*, y a peor nota en *Matemáticas*, peor nota en *Física*, pero solo a grandes rasgos, *grosso modo*. Se dice que existe **correlación** entre esas dos variables.

Figura. 4.7.2. Definición de distribución bidimensional ([T1], p.332)

Las definiciones son parcialmente correctas porque no establecen relaciones apropiadas con la variable estadística unidimensional, pues sólo se considera que el número de categorías diferentes de las dos variables unidimensionales que la forman sea el mismo. Es decir, no se trata el caso de un conjunto de $n \times m$ modalidades (*m* en la variable *X* y *n* en la variable *Y*) siendo *m* y *n* diferentes. Posiblemente este hecho se deba al escaso uso de la tabla de doble entrada. A veces, la definición viene acompañada de ejemplos, aprovechando para introducir el diagrama de dispersión, o una tabla, como por ejemplo [T1] (Figura 4.7.2), que combina tres representaciones del concepto, a la vez que los define. Posteriormente se ofrece una definición estructural (Sfard, 1991) como conjunto de valores de dos variables estadísticas unidimensionales. La misma complejidad semiótica se presenta en [H1], mostrando los mismos ejemplos.

C2. Distribución marginal y condicional

De la distribución doble pueden deducirse diferentes distribuciones unidimensionales. En esta categoría se incluye la definición de la frecuencia y distribución marginal y de la frecuencia y distribución condicional.

C21. Frecuencia marginal

La frecuencia marginal de un valor $X = x_i$ se obtiene sumando la frecuencia absoluta o relativa doble de todos los pares de valores en la distribución bidimensional que tengan este valor de *X*. Sánchez Cobo (1999) encuentra este concepto en sólo cuatro de los textos que analizó. En nuestro análisis, se presenta en cinco textos en la modalidad de humanidades ([H3], [H4], [H5], [H7] y [H8]) y cuatro de la modalidad científico-tecnológica ([T4], [T5], [T7] y [T8]), y es definida en su mayoría formalmente, al inicio de su uso. Aunque se utiliza poco y en algunos casos, como en [H3], prácticamente nada, es conveniente dar su definición. Por ello mencionamos [T3], pues la omite hasta el punto de eliminar de la tabla de doble entrada los totales de fila y columna, que sí se incluyen en [H3], y sirven de base para definir este concepto.

Las definiciones son parcialmente correctas, porque no ofrecen la notación asociada, aunque [H4], [T4], [H7], [H8] y [T8] se aproximan a ella, bien con el uso de sumatorios ([H7]), o de letras ([H4] y [T4]). En este último caso (Figura 4.7.3), la notación de la frecuencia marginal (A_i) no es habitual; además, no se explica el significado del subíndice, que se refiere a las modalidades de la variable, ni tampoco el rango de variación de la variable X. El caso más llamativo son los textos [H8] y [T8], que incluyen la notación adecuada: $f_{\bullet m}$ y $f_{n\bullet}$ en la representación tabular de la variable, y al definir el concepto la cambian por esta otra: f_i donde además, no se emplea terminología verbal "*frecuencia absoluta marginal*", sino lenguaje simbólico.

- Frecuencia absoluta marginal de la variable X_i, A_i

$$A_i = \sum_{j=1}^{m} f_{ij}$$

que representa el número de elementos en los que X toma el valor x_i

- Frecuencia absoluta marginal de la variable Y_j, B_j

$$B_j = \sum_{i=1}^{n} f_{ij}$$

que representa el número de elementos en los que Y toma el valor y_j

Figura 4.7.3. Definición de frecuencia marginal ([T4], p. 332)

C22. Distribución marginal

El conjunto de todos los valores de una de las variables, junto con su frecuencia marginal, constituye la distribución marginal. Sánchez Cobo (1999) no la incluye entre los conceptos que analiza, y la considera equivalente a la noción de frecuencia marginal, lo que no es totalmente correcto, pues una distribución es un conjunto de valores de la variable, junto con sus frecuencias, mientras que una frecuencia corresponde a un valor aislado.

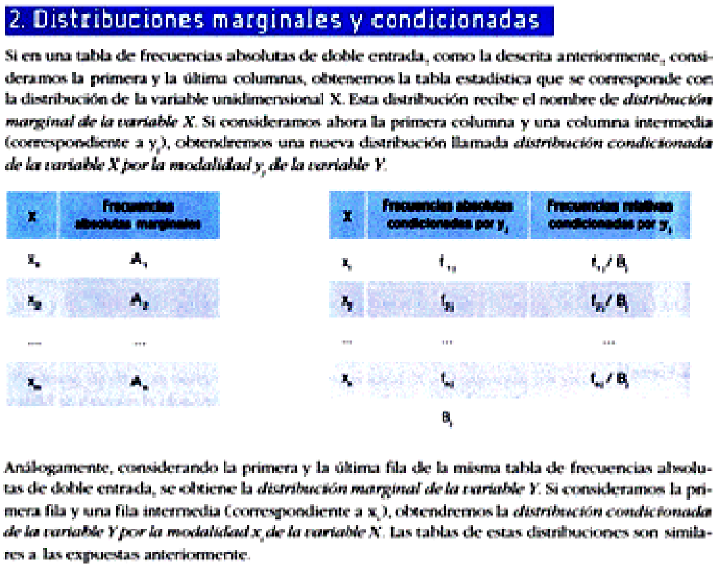
Las definiciones de distribución marginal suelen ser parcialmente correctas, ya que no se ofrece la notación asociada, siendo tan sólo correcta en [H4] y [T4]. En general, vienen acompañadas de ejemplos que clarifican su construcción. Así, en [H1] y [T1] se usa sin una definición previa, como etiquetado en una tabla de frecuencias, pudiendo inducir un conflicto semiótico en cuanto al concepto y su representación tabular (Ver Sección 4.10). Por su parte, el texto [H3] la define al margen, y como complemento a la definición de frecuencia marginal:

> Cuando se estudian por separado las variables unidimensionales X e Y que forman la variable bidimensional *(X, Y),* se habla de distribuciones marginales ([H3], p. 217).

C23. Frecuencia condicional

La frecuencia absoluta de un valor $X = x_i$ condicionado a que la variable Y tome un determinado valor, como por ejemplo $Y = y_j$ se corresponde con la frecuencia absoluta doble del dato (x_i , y_j). La relativa sería igual al cociente entre la frecuencia absoluta doble y la marginal del valor que sirve de condición. Es importante que el estudiante entienda que la variable que condiciona aporta información nueva, y por ello la frecuencia relativa condicional es, en general, diferente de la frecuencia doble relativa. Esta comprensión será de gran ayuda para dar sentido a la regresión, como modelo de ajuste a los datos, con objeto de predecir una variable a partir de la otra.

Sánchez Cobo (1999) sólo encuentra este concepto en uno de los textos que analiza, y no precisa el modo en que se presenta o cómo se usa. En nuestro caso, la definición tan sólo se incluye en [H4] y [T4], que también presentan la definición operacional de las frecuencias relativas condicionales, apoyándose en la representación tabular de la distribución bidimensional (Figura 4.7.4).



2. Distribuciones marginales y condicionadas

Si en una tabla de frecuencias absolutas de doble entrada, como la descrita anteriormente, consideramos la primera y la última columnas, obtenemos la tabla estadística que se corresponde con la distribución de la variable unidimensional X. Esta distribución recibe el nombre de *distribución marginal de la variable X*. Si consideramos ahora la primera columna y una columna intermedia (correspondiente a y_j), obtendremos una nueva distribución llamada *distribución condicionada de la variable X por la modalidad y_j de la variable Y*.

X	Frecuencias absolutas marginales
x_1	A_1
x_2	A_2
...	...
x_n	A_n

X	Frecuencias absolutas condicionadas por y_j	Frecuencias relativas condicionadas por y_j
x_1	f_{1j}	f_{1j}/B_j
x_2	f_{2j}	f_{2j}/B_j
...	...	...
x_n	f_{nj}	f_{nj}/B_j
	B_j	

Análogamente, considerando la primera y la última fila de la misma tabla de frecuencias absolutas de doble entrada, se obtiene la *distribución marginal de la variable Y*. Si consideramos la primera fila y una fila intermedia (correspondiente a x_i), obtendremos la *distribución condicionada de la variable Y por la modalidad x_i de la variable X*. Las tablas de estas distribuciones son similares a las expuestas anteriormente.

Figura 4.7.4. Definición de distribución marginal y condicionada ([H4], p. 220)

C24. Distribución condicional

La distribución condicional sería el conjunto de valores de una de las variables junto con las frecuencias condicionadas respecto a un valor de la otra variable. Sólo [H4] y [T4] (Figura 4.7.4) la definen, englobando la definición de frecuencia condicional, aunque, como hemos indicado anteriormente, sería importante diferenciar estos conceptos. Su definición se apoya en la representación tabular de la distribución bidimensional, y aunque con imprecisiones, es correcta, No menciona que se toma la primera columna (fila) porque contiene las categorías de la variable que se condiciona. Previendo esta posible dificultad, después la definición, presenta un ejemplo que clarifica su notación y funcionalidad.

Síntesis de la presentación de la distribución bidimensional en el tema

En la Tabla 4.7.1 se resume la forma en que se introducen los conceptos descritos anteriormente en los textos analizados. El que más se define es el de frecuencia doble, seguida por la variable estadística bidimensional y la distribución marginal. Por otro lado, apenas se define la distribución o frecuencia condicional (sólo [H4] y [T4]), a pesar de su importancia para el desarrollo de la regresión, ya que, para poder comprender bien el significado de la recta de regresión, es importante entender que la función de regresión es el lugar geométrico de la media de las distribuciones condicionales.

Encontramos diferencias según editorial, donde algunos textos son muy completos como [H4] y [T4], que incluyen casi todos los conceptos estudiados, hasta los que sólo se incluyen definiciones de dos conceptos como [H6] y [T6]. Además, se observa inconsistencia en cuanto al tratamiento de algunos conceptos pues, a veces, se define la frecuencia de un cierto tipo (doble, marginal y condicional), pero no la correspondiente distribución, o viceversa. También observamos diferencias en los textos de distintas modalidades de Bachillerato, principalmente debidas al tratamiento de las frecuencias asociadas a la distribución bidimensional.

Tabla 4.7.1. Forma de introducción de los conceptos asociados a la distribución bidimensional

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C11.Variable estadística bidimensional			ES	ESO		SO	SE	ESO
C12.Frecuencia doble	E	E	E	SE	E		OE	EO
C13.Distribución bidimensional	OES	SE	S		SO			
C21.Frecuencia marginal			E	OS	S		OSE	O
C22.Distribución marginal	E		E	OE		E		O
C23.Frecuencia condicional				OE				
C24.Distribución condicional				OE				
	T1	T2	T3	T4	T5	T6	T7	T8
C11.Variable estadística bidimensional			ES	ESO		SO	SE	ESO
C12.Frecuencia doble	E	E	E	SE	E			EO
C13.Distribución bidimensional	OES	SE	S		SO			
C21.Frecuencia marginal				OS	S		ES	O
C22.Distribución marginal	E			OE		E		O
C23.Frecuencia condicional				OE				
C24.Distribución condicional				OE				

E = mediante ejemplos; O = definición operacional; S = definición estructural

Los conceptos se suelen definir acompañados de ejemplos, en incluso en muchos casos sólo se presentan ejemplos del concepto, sin proponer una definición formal del mismo. Es decir, se realiza una definición implícita (en el sentido de Ortiz, 1999) por lo que es posible que el estudiante, que suele seguir exclusivamente el texto, no llegue a formalizarlo convenientemente. Los textos [H1] y [T1] suelen presentan los ejemplos antes de la definición, como recomienda Skemp (1993). Por el contrario [H4], [H7], [T4] y [T7] comienzan por una definición operacional o estructural (Sfard, 1991) a la que siguen ejemplos. El resto de los textos no sigue un patrón a este respecto. Lo ideal sería la presentación que comienza por ejemplos, seguida de la definición formal, pero son pocos los textos que siguen este procedimiento en todos estos conceptos; sería también admisible la presentación formal seguida por ejemplos, que aparece en otros textos. No se observan grandes diferencias en los textos de la misma editorial dirigidos a las dos modalidades de Bachillerato, aunque sí entre editoriales sobre todo en el número de conceptos considerados.

En cuanto al lugar del texto donde se introducen las definiciones de los conceptos analizados, suele ser la primera vez que se utiliza el concepto dentro del texto (Tabla 4.7.2). Ello indica una orientación teoría-práctica en estos textos para el tema y contradice los principios expresados por Skemp (1993), pues indica la concepción del autor de que primero hay que introducir teóricamente un concepto, antes de proponer ejemplos o problemas que permitan aplicarlo.

Tabla 4.7.2. Lugar de introducción de los conceptos asociados a la distribución bidimensional

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C11.Variable estadística bidimensional			I	I		I	I	I
C12.Frecuencia doble	I	I	I	I	I		I	F
C13.Distribución bidimensional	VM	I	I		I			
C21.Frecuencia marginal			F	I	I		I	I
C22.Distribución marginal	I		F	I		I		I
C23.Frecuencia condicional				I				
C24.Distribución condicional				I				

	T1	T2	T3	T4	T5	T6	T7	T8
C11.Variable estadística bidimensional			I	I		I	I	I
C12.Frecuencia doble	I	I	I	I	I			F
C13.Distribución bidimensional	VM	I	I		I			
C21.Frecuencia marginal				I	I		F	I
C22.Distribución marginal	I			I		I		I
C23.Frecuencia condicional				I				
C24.Distribución condicional				I				

I = al presentar el concepto; F = después de su uso; VM = en varios momentos

Esta es una tendencia general en los textos analizados, y para todos los conceptos que hemos estudiado. Únicamente en [H1] y [T1] se define la distribución bidimensional en varios momentos del desarrollo del tema, y en otros, como [H3], [H8], [T7] y [T8] se define la frecuencia (doble o marginal) o la distribución marginal después de haberlas utilizado, al final del tema. En estos casos, primero se ha propuesto ejercicios o ejemplos en que se utiliza el contexto y se define como síntesis de lo aprendido al finalizar el tema. Cabe mencionar una editorial donde la definición de frecuencia marginal en las dos modalidades de Bachillerato es diferente: en [H7] se lleva a cabo al inicio del tema, mientras que en [T7] la frecuencia marginal se define tras haber sido usada en un ejemplo; en este último caso, la definición se realiza simplemente con una nota al margen.

Se ha analizado igualmente la intensidad del uso de cada concepto a lo largo del tema, como un indicador de la importancia que los autores dan al mismo (Tabla 4.7.3). Observamos que se suelen utilizar en forma continua todos los conceptos definidos en la mayoría de los libros (salvo [H1] y [T1]), no encontrando diferencias en los textos de una misma editorial, dirigidos a las dos modalidades de Bachillerato (salvo [T3] que no incluye la frecuencia y distribución marginal).

Sí encontramos diferencias en los textos de cada editorial, ya que por ejemplo, en los textos [H1] y [T1], los pocos conceptos que define se usan poco o nada, mientras [H7], [H8], [T7] y [T8] hacen mucho mayor uso. O la variable estadística bidimensional, a pesar de que tenga un uso destacado en los libros, en los textos [H4] y

[T4] se utiliza poco, y por contra, se usa la noción de distribución marginal, frecuencia y distribución condicionada sobre todo en los enunciados de tareas.

Tabla 4.7.3. Uso en el tema de los conceptos asociados a la distribución bidimensional

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C11.Variable estadística bidimensional			S	P		S	S	S
C12.Frecuencia doble	P	S	S	S	N		S	S
C13.Distribución bidimensional	P	S	S		S			
C21.Frecuencia marginal			N	S	P		S	P
C22.Distribución marginal	N		N	S		S		S
C23.Frecuencia condicional				S				
C24.Distribución condicional				S				

	T1	T2	T3	T4	T5	T6	T7	T8
C11.Variable estadística bidimensional			S	P		S	S	S
C12.Frecuencia doble	P	S	S	S	N			S
C13.Distribución bidimensional	P	S	S		S			
C21.Frecuencia marginal				S	P		S	P
C22.Distribución marginal	N			S		S		S
C23.Frecuencia condicional				S				
C24.Distribución condicional				S				

S = se usa en el tema; P = se usa poco; N = no se usa

Algunos textos que definen la distribución bidimensional apenas la usan. En general, la frecuencia y distribución marginal se definen menos que la frecuencia o distribución doble, aunque, cuando se hace, se suele utilizar bastante. La frecuencia y distribución condicional se emplea en los textos que la definen ([H4] y [T4]).

4.7.2. ANÁLISIS DE LA DEPENDENCIA ENTRE DOS VARIABLES

Como se indicó en el Capítulo 1, una de las principales preguntas a que trata de responder el tema es si existe alguna relación entre las variables incluidas en el estudio. A continuación, describimos el tratamiento de los conceptos asociados.

C3. Dependencia funcional /estadística o aleatoria / independencia

En este apartado consideramos tanto la dependencia funcional (C31), como la estadística (C32) y la independencia (C33).

C31. Dependencia funcional

En primer lugar se analiza si los textos presentan una definición, explícita o no, de la dependencia funcional. Al contrario que en el estudio de Sánchez Cobo (1999), todos nuestros textos la describen, diferenciándola de la dependencia estadística, aunque su uso es reducido, prácticamente al introducir la correlación, covarianza y el coeficiente de correlación lineal. Es frecuente presentarla a partir de ejemplos, a veces como contrapunto de otros ejemplos de dependencia aleatoria. Por ejemplo, en el texto [T1] (p. 331), se incluye el ejemplo de altura que alcanza una piedra al lanzarla hacia arriba con diferente fuerza, y otros ejemplos de dependencia funcional, resaltando que a cada valor de la variable dependiente se da sólo uno de la dependiente.

Cabe decir, que algunos de estos textos realizan un tratamiento inapropiado de la dependencia funcional, llegando incluso a definirlo incorrectamente. Así es que en [H2], [H8] y [T2], [T8] se consideran sinónimos los términos: relación, dependencia y correlación; de hecho, los textos [H2] y [T2] definen *correlación funcional* del siguiente modo: "*Correlación funcional: todos los puntos están situados sobre una recta o una curva. Existe una relación funcional entre las variables X e Y*" ([H2], p.248). El caso más preocupante se encuentra en [H7] y [T7], que a partir de un ejemplo, identifica relación funcional con relación no lineal, pudiendo inducir un conflicto semiótico en cuanto a estas nociones (un análisis más profundo se presenta en la Sección 4.12).

C32. Dependencia estadística o aleatoria

Con la dependencia estadística se amplía el concepto de dependencia funcional, considerada hasta el momento de introducir la correlación en la enseñanza (Sánchez Cobo, 1999). Además, puesto que puede haber diferentes intensidades, se crea la necesidad de cuantificar la dependencia, y definir el coeficiente de correlación.

Esta noción se define en todos los textos, salvo en [H2] y [T2], siempre acompañada de ejemplos para facilitar su comprensión, y por lo general, en forma parcialmente correcta, pues, en ocasiones se identifica con correlación, aunque el concepto de dependencia estadística es más amplio. La correlación se limita a variables cuantitativas (Sección 4.12), mientras que también existe la dependencia aleatoria o asociación entre variables cualitativas o mixtas. Aunque reconocemos la dificultad de desligar dependencia funcional y aleatoria, debido a que en este nivel educativo su análisis se realiza exclusivamente entre variables numéricas, creemos necesario diferenciar ambas nociones, algo que sólo encontramos en [H4] y [T4]:

> En ocasiones se observa que existe una relación entre las variables, pero dicha relación no puede expresarse como una función matemática. En este caso se dice que entre las variables X e Y existe una dependencia estadística, que podrá ser fuerte o débil." ([T4], p. 338).

C33. Independencia

Este concepto es de gran importancia, y no siempre es bien comprendido por los estudiantes, como se mostró en Cañadas (2012). Sirve de apoyo para el desarrollo de la noción de correlación y regresión, y de manera implícita o explícita, llega a fundamentar el diseño de muchas tareas (Sección 4.3). A pesar de su uso generalizado, las definiciones que se presenta en los textos suelen ser parcialmente correctas, ya que se identifica correlación nula e independencia, aunque en caso de dependencia no lineal (por ejemplo parabólica) pudiera darse un coeficiente de correlación igual, o cercano a cero, siendo las variables dependientes.

Así, [H1], [H2], [H8], [T1], [T2] y [T8] no precisan la diferencia entre variables independientes e incorreladas (un análisis más detallado se presenta en la Sección 4.12). Los casos más extremos son [H7] y [T7], que no definen la independencia, utilizando en su lugar expresiones como "dependencia débil", e incluyen tareas como la que se muestra en la Figura 4.7.5. Suponemos que es considerada como una noción sencilla para el estudiante, por la facilidad con que se reconoce en el diagrama de dispersión. Creemos que se debiera definir, como hacen [H4] y [T4]: "*Cuando no existe* *dependencia funcional ni estadística, se dice que hay independencia estadística entre las variables*” ([H4], p.224). Además, la comprensión de la independencia es base de muchos temas estadísticos posteriores; por ejemplo, en inferencia, un supuesto básico de aplicación de la mayor parte de contrastes estadísticos es admitir la independencia estadística de los datos de la muestra.

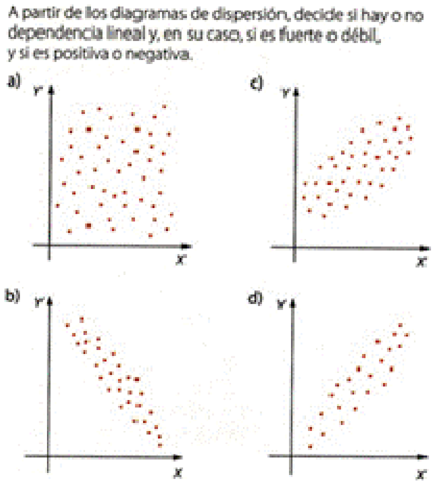


Figura 4.7.5. Tarea de interpretación de dependencia a partir del diagrama de dispersión ([T7], p.316)

C4. Covarianza y/o correlación

Con objeto de medir la intensidad de la dependencia entre dos variables estadísticas, a continuación se analizan las definiciones de correlación (C41), covarianza (C42) y coeficiente de correlación lineal de Pearson (C43).

C41. Correlación

Esta noción se encuentra en todos los textos analizados, y se define en todos menos en [H7] y [T7], que sin definirlo, plantean el cálculo del coeficiente de correlación: “*Estudia la correlación entre estas variables, utilizando la calculadora para realizar las operaciones.* […]” ([H7], p. 253). Es común encontrar la definición de correlación como sinónimo de dependencia estadística, sin requerir que las variables sean numéricas. Esta situación se presenta igualmente en cuatro de los textos que analiza Sánchez Cobo (1999), aunque en nuestro caso es más frecuente.

Destacamos los textos [H5] y [T5] por definir la correlación espuria o casual (Barbancho, 1973), aproximándose a otras tipologías de covariación, como interdependencia y dependencia causal unilateral. Aun así, el tratamiento es puramente anecdótico, tan sólo [H8] propone ejemplos de correlación espuria; [H3] y [T3] como [H5] y [T5] proponen ejemplos de correlación indirecta, como el siguiente:

> No siempre que dos variables generen una nube de puntos alargada existe correlación entre ellas. Son muchos los casos en los que dos caracteres varían a la vez, sin que por ello estén correlacionados. Por ejemplo: las canas y la miopía de las personas. Es posible que entre los canosos haya más miopes, pero no por ser canosos sino por ser mayores. Este tipo de falsas correlaciones se llaman espurias ([H5], p. 252).

[H3], [T3] y [H8] definen la correlación espuria incorrectamente, pues la consideran contraria a la causalidad, como determinante de la covariación, sin considerar otros tipos de dependencia, como la dependencia indirecta, la interdependencia o la concordancia (Barbancho, 1973), Un ejemplo se muestra a continuación:

> Se dice que existe *correlación espuria* entre dos variables estadísticas cuando éstas aumentan o disminuyen de manera conjunta sin que exista una relación causa-efecto entre ellas. Por ejemplo, es muy posible que exista una cierta correlación entre el número de restaurantes de una ciudad y el número de profesores que trabajan en ella. Esto se debe a que ambas variables están relacionadas con el número total de habitantes de la ciudad ([H3], p.221).

C42. Covarianza

La covarianza es un coeficiente de asociación para variables cuantitativas, y se define formalmente en todos los textos analizados, al igual que en la mayoría de textos del estudio de Sánchez Cobo (1999). Por lo general, se acompaña de ejemplos que faciliten su comprensión. En este sentido, es de gran utilidad el tratamiento que los textos [H1], [H2] o [T1], [T2] ofrecen, mediante la división en cuatro cuadrantes de la nube de puntos, por las rectas correspondientes a las medias de cada variable (Figura 4.7.6), ya que permite al estudiante desarrollar una comprensión más significativa de la misma. Recordemos que este es el modo en que se razona el signo de la correlación y la covarianza en la propuesta didáctica de Holmes (2001) (Véase Capítulo 1).

Otros textos como [H6], [H8] y [T6] ofrecen esta interpretación, como se expone a continuación, aunque no presentan la gráfica citada; creemos que esta omisión dificulta la interpretación:

> El valor de la covarianza indica cómo se apartan a la vez las dos coordenadas de un dato respecto de la media. Si el resultado es positivo, quiere decir que los productos son positivos, esto es, los valores de x_i y de y_i se alejan en el mismo sentido de sus respectivas medias. Y al contrario si el resultado es negativo. En caso de que el valor sea cero o próximo a cero, la covarianza informa de que no hay relación entre ambas variables ([H8], p.251).

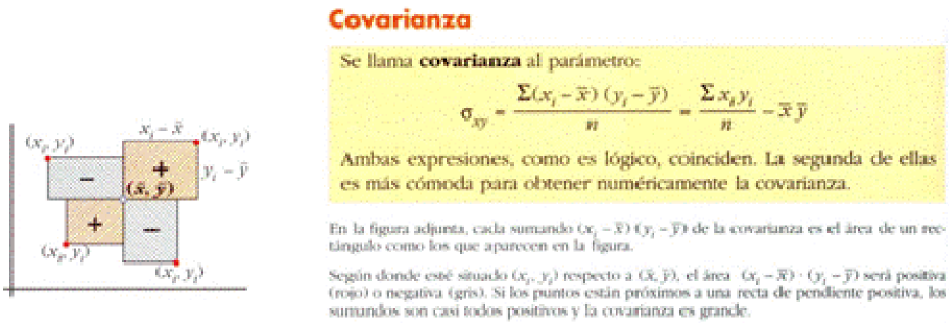
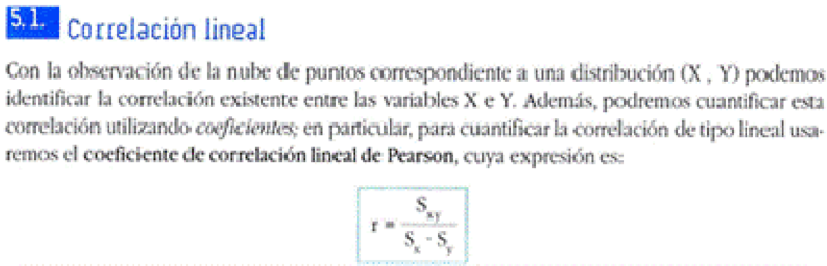


Figura 4.7.6. Definición e interpretación de la covarianza ([H1], p. 228)

La covarianza es muy utilizada en el cálculo del coeficiente de correlación lineal; principalmente desde datos organizados en tablas simples, o con frecuencias añadidas, y sólo en casos aislados, mediante el uso de tablas de doble entrada. Quizá ello explica la poca precisión en la notación asociada a la fórmula de la covarianza. En la Sección 4.12, se describe un conflicto semiótico relacionado con este sentido.

C43. Coeficiente de correlación de Pearson

La principal utilidad de este coeficiente es informar de la intensidad y sentido de la dependencia lineal de las variables de estudio. Aun así, siempre es útil que venga acompañado de la representación gráfica los datos, preferentemente con el diagrama de dispersión. Todos los textos lo definen, mayoritariamente desde el punto de vista estructural y simbólico-operacional (Figura 4.7.6), acompañado generalmente de ejemplos.


5.1. Correlación lineal

Con la observación de la nube de puntos correspondiente a una distribución (X , Y) podemos identificar la correlación existente entre las variables X e Y. Además, podremos cuantificar esta correlación utilizando *coeficientes*; en particular, para cuantificar la correlación de tipo lineal usaremos el **coeficiente de correlación lineal de Pearson**, cuya expresión es:

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

Figura 4.7.6. Fórmula del índice de correlación lineal ([H4], p. 225)

Las definiciones suelen ser correctas, aludiendo a su aplicabilidad en el análisis de la dependencia estadística lineal. Sólo los textos [H1], [H2], [T1] y [T2] pudieran inducir la confusión de que este coeficiente pudiera informar de otro tipo de dependencia; por ejemplo, curvilínea (Sección 4.12). Por el contrario, [H5] y [T5] lo definen en la siguiente forma: "*El valor del coeficiente de correlación lineal es el criterio que se utiliza para medir la fuerza de la correlación entre dos variables.*" ([H5], p.255), y, tras presentar la fórmula y las propiedades de este coeficiente, completan la definición añadiendo:

> Este coeficiente [coeficiente de correlación lineal] mide exclusivamente la correlación lineal entre variables. Por tanto, puede haber otro tipo de correlación no detectada por r. Por ejemplo, r no detectaría la correlación exponencial perfecta que hay entre los puntos (-1, 0.5), (0,1), (1,2), y (4,16), que pertenecen todos a la gráfica de $y = 2^x$ ([H5], p.255).

Síntesis de la presentación de la dependencia

En las tablas siguientes se resume la forma de introducción, el lugar y uso en el tema de las definiciones sobre la dependencia entre dos variables en los libros analizados. Prácticamente todos los conceptos considerados en este apartado se contemplan en la mayoría de los libros (Tabla 4.7.4). Los textos de la misma editorial, dirigidos a las dos modalidades de Bachillerato, introducen los mismos conceptos, e incluso los presentan del mismo modo. Por tanto, no se tiene muy en cuenta las recomendaciones de las orientaciones curriculares de disminuir la formalización en el Bachillerato de *Humanidades y Ciencias Sociales*.

Se suelen completar las definiciones relativas a la dependencia con ejemplos, sobre todo en [H1] y [T1]. Dentro de las editoriales, la principal diferencia es que [H1], [H2], [H3], [T1], [T2] y [T3] presentan los ejemplos antes de la definición, mientras que el resto lo hace al contrario, es decir, muestran mayor orientación teoría-práctica. No hay una tendencia clara en el tipo de presentación, estructural u operacional, ya que algunos textos introducen los dos tipos.

Tabla 4.7.4. Forma de introducción de los conceptos ligados a la dependencia

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C31.Dependencia funcional	E	ES	ES	SOE	SE	OE	SOE	OE
C32.Dependencia estadística o aleatoria	E		ES	SOE	SE	OE	OE	OE
C33.Independencia	E	ES	ES	SE		SE		SO
C41.Correlación	ESO	SOE	ESO	SOE	ESO	SOE		SOE
C42.Covarianza	OSE	OSE	O	SOE	OSE	OSE	OE	OS
C43.Coeficiente de correlación lineal	OS	OS	OS	SOE	SO	SOE	OSE	SO
	T1	T2	T3	T4	T5	T6	T7	T8
C31.Dependencia funcional	E	ES	ES	SOE	SE	OE	SOE	OE
C32.Dependencia estadística o aleatoria	E		ES	SOE	SE	OE	OE	OE
C33.Independencia	E	ES	ES	SE		SE		SO
C41.Correlación	ESO	SOE	ESO	SOE	ESO	SOE		SOE
C42.Covarianza	OSE	OSE	O	SOE	OSE	OSE	OE	OSE
C43.Coeficiente de correlación lineal	OS	OS	OS	SOE	SO	SOE	OSE	SO

E = mediante ejemplos; O = definición operacional; S = definición estructural

Tabla 4.7.5. Lugar de introducción de los conceptos ligados a la dependencia

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C31.Dependencia funcional	I	F	I	I	I	I	I	VM
C32.Dependencia estadística o aleatoria	I		I	I	I	I	I	VM
C33.Independencia	F	F	I	I		I		VM
C41.Correlación	VM	I	I	I	F	F		I
C42.Covarianza	I	I	I	I	I	I	I	I
C43.Coeficiente de correlación lineal de Pearson	I	I	I	I	I	I	I	I
	T1	T2	T3	T4	T5	T6	T7	T8
C31.Dependencia funcional	I	F	I	I	I	I	I	VM
C32.Dependencia estadística o aleatoria	I		I	I	I	I	I	VM
C33.Independencia	F	F	I	I		I		VM
C41.Correlación	VM	I	I	I	F	F		I
C42.Covarianza	I	I	I	I	I	I	I	I
C43.Coeficiente de correlación lineal de Pearson	I	I	I	I	I	I	I	I

Tampoco hay prácticamente diferencias en cuanto al lugar en que se definen (Tabla 4.7.5) ni entre modalidad de Bachillerato, ni entre editoriales. Es habitual encontrar la definición de estos conceptos al comienzo de su tratamiento, aunque a veces la dependencia funcional, estadística, independencia o correlación se definen mucho después de haber sido utilizadas en ejemplos. En este sentido, destacamos los textos [H8] y [T8], que presentan la correlación en varios momentos, en primer lugar, con términos como *correlación funcional*, y posteriormente, de modo más apropiado, al introducir el coeficiente de correlación lineal, utilizando ya el término dependencia. Estos resultados, junto a los mostrados en la Tabla 4.7.5 evidencian, en general, una orientación teoría-práctica en la exposición del tema.

Podemos observar que estos objetos son muy utilizados (Tabla 4.7.6), principalmente la covarianza, correlación y coeficiente de correlación. Las relativas a la dependencia y sus tipos se usan relativamente poco, aunque este aspecto hay que matizarlo. El uso de estos objetos se ve sustituido en gran medida por el de correlación, planteando tareas al estudiante en las que se utiliza la idea de dependencia y sus tipos, aunque no aparezcan como tales en sus enunciados, como en los textos [H6], [T6] y [H8]. Por ello, aunque se plantean ejercicios en que se deben interpretar los datos en contexto, el profesor deberá estar atento a la adquisición de la idea de dependencia y sus tipos por parte del estudiante, junto a su uso apropiado y lenguaje preciso, ya que el libro de texto se centra más en la medida de la intensidad de la dependencia, que en su interpretación.

Tabla 4.7.6. Uso en el tema de los conceptos ligados a la dependencia

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C31.Dependencia funcional	S	P	S	P	P	S	S	P
C32.Dependencia estadística o aleatoria	S		S	P	P	P	S	P
C33.Independencia	P	P	S	P		P		P
C41.Correlación	S	S	S	P	S	S		S
C42.Covarianza	P	S	P	S	S	P	S	S
C43.Coeficiente de correlación lineal de Pearson	S	S	S	S	S	S	S	S

Conceptos	T1	T2	T3	T4	T5	T6	T7	T8
C31.Dependencia funcional	S	P	S	P	P	S	S	P
C32.Dependencia estadística o aleatoria	S		S	P	P	P	S	P
C33.Independencia	P	P	S	P		P		P
C41.Correlación	S	S	S	P	S	S		S
C42.Covarianza	P	S	P	S	S	P	S	S
C43.Coeficiente de correlación lineal de Pearson	S	S	S	S	S	S	S	S

4.7.3. ANÁLISIS DE LA REGRESIÓN

En esta sección analizaremos las definiciones de variable dependiente e independiente, regresión, modelo de regresión, y regresión lineal, según la noción de recta de regresión, coeficientes del modelo lineal, y bondad del ajuste.

C51. Variable dependiente e independiente

La diferencia entre variable dependiente e independiente ocupa un lugar central en el análisis de la regresión ya que, una vez evidenciada la dependencia entre las variables, y con objeto de expresar en forma de ecuación o modelo una variable en función de otra, se necesita seleccionar qué variable servirá como dependiente o predictora. Estas nociones se suelen utilizar implícitamente en los textos analizados, al exponer que existen dos rectas de regresión, como se muestra a continuación:

> Si *X* se considera la variable independiente e *Y* la variable dependiente, la ecuación de la recta de regresión es: $y-\bar{y}=\frac{S_{xy}}{S_x^2}\left(x-\bar{x}\right)$.Esta recta se denomina *recta de regresión de Y sobre X*. A partir de ella, conocidos los valores de *X*, y sustituyéndolos en la ecuación, se pueden calcular con u*na cierta aproximació*n los *va*lores de *Y*.
> Si se considera *Y* como variable independiente y *X* como variable dependiente, se obtiene la *recta de regresión de X sobre Y*, cuya ecuación es: $x-\bar{x}=\frac{S_{xy}}{S_y^2}\left(y-\bar{y}\right)$.Igual que en el caso anterior, conocidos los valores de *Y*, y sustituyéndolos en la ecuación de la recta, se obtienen con cierta aproximación los valores de *X* ([H4], p.226).

La definición explícita de variable dependiente e independiente sólo la encontramos en [H5] y [T5] mediante una anotación al margen, y en [H8] del siguiente modo: "*La variable dependiente es aquella que se quiere estimar, y la variable que se utiliza para ello se denomina variable independiente.*" ([H8], p.254).

C52. Regresión

Al contrario que en la investigación de Sánchez Cobo (1999), hay un escaso tratamiento de la regresión, en forma general, lo que nos lleva a pensar que la enseñanza se centra más en el análisis de la regresión lineal. El concepto de regresión suele tratarse de modo implícito, y sólo [H3], [H4], [T3] y [T4] incluyen una definición explícita y correcta de la misma. Por ejemplo, en [H3] y [T3] se define como ajuste de una función a un conjunto de datos: "*Al análisis que pretende determinar la curva que mejor aproxima un diagrama de dispersión se le llama regresión.*" ([H3], p. 226), y en [H4] y [T4] se resalta su utilidad predictiva:

> La regresión consiste en tratar de encontrar una función matemática que relacione las variables X e Y de una distribución bidimensional, de forma que, si se conoce el valor de una variable, se puede calcular el correspondiente de la otra, con mayor o menor aproximación ([T4], p. 340)

En todos los textos que definen la regresión, se incluye la definición de regresión lineal, aunque en forma imprecisa; otros como [H5], [H6], [H8], [T5], [T6] y [T8], aluden al término regresión lineal sin definirlo. No consideramos adecuado este tratamiento tan superficial de la noción de regresión, limitándose, sin mucha profundidad, al significado institucional pretendido (Véase Capítulo 1) ya que es importante que el estudiante considere otros modelos de regresión distintos del lineal, y comprenda su significado, como se describe a continuación.

C53. Modelos no lineales de regresión

El concepto de modelo es fundamental en regresión. Como señala Moore (2005), la regresión es un método general para comprender las relaciones entre variables. Aunque Galton lo dotó de significado, anteriormente el método de mínimos cuadrados, desarrollado por Legendre, se usaba sólo para ajustar una recta a datos de dos variables. Los textos analizados raramente definen el modelo no lineal de ajuste a los datos, limitando así el significado de la regresión (una descripción más detallada se presenta en la Sección 4.12). En algunos casos queda implícito en la definición de regresión, por ejemplo:

> Si en una variable (*X,Y)* existe una correlación fuerte entre las variables *X* e *Y*, el análisis de la regresión permite encontrar la ecuación de la función matemática que mejor se ajusta a la nube de puntos. Esta puede ser una recta, una parábola, una exponencial, una cúbica, etc. A partir de aquí nos centraremos en el estudio de la regresión lineal ([H4], p.226).

[H4], [H5], [H6], [T4], [T5] y [T6] definen la regresión no lineal, situación parecida a la descrita en la investigación de Lavalle, Micheli y Rubio (2006), donde sólo dos de los siete textos aluden a la regresión no lineal. Un ejemplo de definiciones de modelo de regresión no lineal sería el siguiente:

> Aunque aquí nos limitaremos a la regresión lineal, existen otros modelos de regresión para ajustar otros tipos de correlación. Por ejemplo, regresión cuadrática si la nube de puntos tiene forma parabólica, o regresión exponencial cuando los puntos se ajustan a funciones exponenciales. ([T5], p. 364)

Por otro lado, los textos analizados plantean tareas donde el modelo de ajuste más indicado es curvilíneo (Sección 4.5). Los casos más extremos son [H2] y [T2], donde no se definen los modelos de regresión, el concepto de variable dependiente e independiente, o la regresión propiamente dicha, y en el 37% de las tareas los datos se ajustan a un modelo polinómico de segundo o tercer grado. Esta situación es un tanto mejor en la investigación de Sánchez Cobo (1999), donde sólo cinco libros incluyen ejemplos de regresión no lineal, aunque no se definan y sólo se haga alusión de modo implícito. Aún así, el apartado sobre la regresión se conforma principalmente alrededor de la determinación de las dos rectas, que describimos a continuación.

C54. Rectas de regresión

En todos los textos se define la recta de regresión como la de mejor ajuste a los datos de una distribución bidimensional, correctamente, salvo [T8] donde se presenta del siguiente modo: "*La recta que mejor se aproxima a la nube de puntos se llama recta de regresión de Y sobre X.*" ([T8], p. 324). Por otro lado, todos los textos precisan el método de mínimos cuadrados, como aquel que permite obtener aquella recta que minimiza los cuadrados de las diferencias entre los datos teóricos y los reales, pudiendo obtenerse así dos rectas: *Y* sobre *X* y *X* sobre *Y*, según qué variable se considere como dependiente e independiente.

A pesar de esto, aunque la mayoría presenta el método de mínimos cuadrados para motivar los cálculos de su pendiente y ordenada en el origen ([H1], [H2], [H5], [H6], [H7], [H1], [T2], [T5], [T6] y [T7]) se encuentran enunciados imprecisos como el siguiente, donde se habla de "la recta de regresión", pero no se indica cuál es la variable dependiente e independiente en el modelo.

> La tabla adjunta da los alargamientos de una barra metálica por efecto de cambios en la temperatura. Calcular la recta de regresión y hacer algunas estimaciones ([H1], p.231).

Sólo [H8] y [T8] aluden al tratamiento de datos atípicos, y la metodología de Tukey para obtener la recta de regresión respecto a la mediana. En cuanto a la utilidad predictiva de la recta de regresión, sólo la mitad de los textos ([H1], [H2], [H4], [H6], [T1], [T2], [T4] y [T6]) la precisan. Estos resultados coinciden con los de Sánchez Cobo (1999), quien señala que se suele aludir al ajuste a los datos (siete textos) más que a su carácter predictivo (tres textos).

C55. Coeficientes de regresión

En el ajuste del modelo de regresión lineal existen dos coeficientes de regresión, dependiendo de qué variable se considere como dependiente o independiente. Se corresponden a las pendientes de las rectas de regresión asociadas, y son útiles para el cálculo de la correlación, así como para relacionar correlación y regresión lineal.

La recta que hace mínima la suma Σd_i^2 tiene por ecuación:

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

Se llama **recta de regresión de *Y* sobre *X***.

A la pendiente, $\frac{\sigma_{xy}}{\sigma_x^2}$, se la llama **coeficiente de regresión**.

Figura 4.7.7. Definición de coeficiente de regresión de *Y* sobre *X* ([H1], p. 230)

Tan sólo los textos [H1], [H2], [H4], [T1], [T2] y [T4] definen estos coeficientes, siendo la definición parcialmente correcta en [H1] y [T1], pues sólo se trata el coeficiente de regresión de *Y* sobre *X* (Figura 4.7.7). Sánchez Cobo (1999) no analiza cómo se definen los coeficientes, tan sólo si se presentan propiedades asociadas a éstos.

C56. Coeficiente de determinación. Bondad de ajuste.

El coeficiente de determinación sólo se define en [H5], [H6], [T5], [T6] y [T8] (Figura 4.7.8). Los otros textos lo tratan implícitamente cuando valoran las estimaciones que se realizan con la recta de regresión, a pesar de su gran utilidad, para dar mayor sentido al análisis de la regresión. En otros casos, se define la bondad de ajuste de modo parcialmente correcto, como las definiciones que se incluyen en [H6] y [T6], al no aludir a la noción de coeficiente de determinación:

> En ocasiones, con el fin de calcular la calidad *o bondad del ajuste* realizado mediante la recta de regresión y, por tanto, la fiabilidad de las predicciones que con ella se puedan realizar, se utiliza la expresión ($r^2 \cdot 100$)%, que nos da el porcentaje en el que la variable Y se justifica por el valor de la variable X ([H6], p.185).

También se suelen describir las nociones de fiabilidad de la predicción y fiabilidad del modelo:

> Con el fin de calcular la calidad o bondad del ajuste realizado mediante la recta de regresión y, por tanto, la fiabilidad de las predicciones que con ella se puedan realizar, se utiliza la expresión ($r^2 \cdot 100$)%), que nos da el porcentaje en el que la variable *Y* se justifica por el valor de la variable *X* ([H6], p. 185).

Obviamente, la fiabilidad del modelo influye en la fiabilidad de la predicción, pero también influye que el valor utilizado en la predicción se encuentre en el intervalo de valores utilizados para el modelo, así como su proximidad a la media.

B. Coeficiente de determinación

Cuando *r* está próximo a 1 (o a −1), la correlación lineal es fuerte. Esto significa que los cambios en la variable *Y* se explican, en gran medida, por los cambios de la variable *X*. En consecuencia, se pueden hacer estimaciones fiables de *Y* a partir de *X*.

Una medida de esta fiabilidad la da r^2, siendo *r* el coeficiente de correlación, pues su valor indica la proporción de la variación en la variable *Y* que puede ser explicada por los cambios de la variable *X*. A r^2 se le llama **coeficiente de determinación**.

Si multiplicamos r^2 por 100, se obtiene el porcentaje de cambio de *Y* explicado por *X*. Así, si $r = 0$, los cambios en la variable *X* explican el 0% de los cambios en la *Y*, o sea, nada: las variables *X* e *Y* son linealmente independientes. Y si $r = 1$ (o $r = -1$), la variación en la *Y* se explica totalmente, al 100%, por la variación de la *X*; en este caso, las variables *X* e *Y* son linealmente dependientes. Fuera de estos casos límites, el porcentaje explicado es 100 r^2. Veamos algunos ejemplos:

Figura 4.7.8. Definición de coeficiente de determinación ([H5], p. 256)

Síntesis de la presentación de la regresión en el tema

En la Tabla 4.7.7 vemos que todos los libros definen las rectas de regresión, mientras el resto de conceptos analizados aparece tan sólo en textos aislados. En cuanto a las modalidades de Bachillerato, encontramos diferencias sólo en los textos [H8] y [T8], ya que en este último no se incluye la definición de variable dependiente e independiente, aunque sí la del coeficiente de determinación.

Tabla 4.7.7. Forma de introducción de los conceptos ligados a la regresión

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C51.Variable dependiente e independiente					O			O
C52.Regresión			ES	SE				
C53.Modelos no lineales de regresión					SOE	SOE		
C54.Rectas de regresión	ESO	EOS	SO	SOE	SOE	SOE	SOE	SO
C55.Coeficientes de regresión	O	SO		O				
C56.Bondad de ajuste. Coef. de determinación					SOE	OE		
Conceptos	T1	T2	T3	T4	T5	T6	T7	T8
C51.Variable dependiente e independiente					O			
C52.Regresión			ES	SE				
C53.Modelos no lineales de regresión					SOE	SOE		
C54.Rectas de regresión	ESO	EOS	SO	SOE	SOE	SOE	SOE	SO
C55.Coeficientes de regresión	O	SO		O				
C56.Bondad de ajuste. Coef. de determinación					SOE	OE		SOE

Se puede observar una mayor cantidad de definiciones acompañadas de ejemplos, que en las definiciones relacionadas con la dependencia y el tratamiento de datos bidimensionales, sobre todo en [H5], [H6], [T5] y [T6]. Destacamos el coeficiente de regresión, ya que aquellos textos que lo definen lo hacen formalmente, al igual que los de variable dependiente e independiente.

La definición de la recta de regresión se suele acompañar de ejemplos; en la mitad de los textos anteriores y en el resto posteriores a la definición. Se lleva a cabo de una manera práctica, utilizando diagramas de dispersión que añaden la recta de mínimos cuadrados mediante ejemplos, que muestran la distancia de los puntos de la nube a la recta. También se muestra la tendencia de la variación conjunta, evidenciando la diferencia entre el valor observado o dato real (ordenadas del punto) y el predicho o dato teórico (punto sobre la recta). Esto mismo ocurre en la investigación de Sánchez Cobo (1999), donde por lo general, se introduce mediante ejemplos en que el diagrama de dispersión muestra la recta de ajuste a los datos. Además, se muestran los dos tipos de definición, estructural y operacional, mientras en otros conceptos se mezclan diferentes tipos de definiciones.

Tabla 4.7.8. Lugar de introducción de los conceptos ligados a la regresión

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C51.Variable dependiente e independiente					F			I
C52.Regresión			I	I				
C53.Modelos de regresión					F	VM		
C54.Rectas de regresión	F	I	I	F	VM	I	I	I
C55.Coeficientes de regresión	I	I		I				
C56.Bondad de ajuste. Coeficiente de determinación					F	F		
	T1	T2	T3	T4	T5	T6	T7	T8
C51.Variable dependiente e independiente					F			
C52.Regresión			I	I				
C53.Modelos de regresión					F	VM		
C54.Rectas de regresión	F	I	I	F	VM	I	I	I
C55.Coeficientes de regresión	I	I		I				
C56.Bondad de ajuste. Coeficiente de determinación					F	F		F

Es habitual que las nociones se definan en el momento en que se presentan, al comienzo de su tratamiento (Tabla 4.7.8), salvo las nociones de modelo de regresión, e incluso a veces de recta de regresión, siendo en [H6] y [T6] definidos en varios momentos, porque se incluyen al tratar el concepto de correlación y posteriormente con la regresión. En este sentido, destacamos los textos [H5] y [T5] por presentar todas sus definiciones tras su uso.

Tabla 4.7.9. Uso en el tema de los conceptos ligados a la regresión

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C51.Variable dependiente e independiente					N			N
C52.Regresión			N	N				
C53.Modelos de regresión					S	P		
C54.Rectas de regresión	S	S	S	S	S	S	S	S
C55.Coeficientes de regresión	P	N		N				
C56.Bondad de ajuste. Coeficiente de determinación					S	S		
	T1	T2	T3	T4	T5	T6	T7	T8
C51.Variable dependiente e independiente					N			
C52.Regresión			N	N				
C53.Modelos de regresión					S	P		
C54.Rectas de regresión	S	S	S	S	S	S	S	S
C55.Coeficientes de regresión	P	N		N				
C56.Bondad de ajuste. Coeficiente de determinación					S	S		S

Podemos observar (Tabla 4.7.9) cómo las definiciones que más se usan son las referidas a las rectas de regresión y la bondad del ajuste. Así es que, en general, otros conceptos como el de modelo de regresión, variable dependiente e independiente o coeficientes de regresión, tienen un tratamiento muy pobre. En este sentido, algunos textos plantean la regresión desde un enfoque algorítmico, donde el estudiante se limita a calcular las rectas de regresión (principalmente la de Y sobre X) y hacer predicciones a través de ellas, pero no se incluyen actividades interpretativas. Más aún, algunos textos plantean tareas donde el diagrama de dispersión muestra una clara tendencia curvilínea de los datos, y se pide realizar un análisis de regresión lineal.

4.8. ANÁLISIS DE LAS PROPOSICIONES

Los libros de texto suelen presentar proposiciones, que describen *atributos o propiedades* de los objetos matemáticos a los que se refieren. Su naturaleza es epistémica, y por tanto institucional; enriquecen el significado del objeto, y se constituyen en objetos explícitos de enseñanza (Godino, Font, Contreras y Wilhelmi, 2006). En esta sección se analizan las proposiciones en los textos, que se estructuran como se hizo con las definiciones (Sección 4.7), según se refieran a la variable estadística y distribución bidimensional, análisis de la dependencia entre variables o análisis de la regresión, así como las relativas a las relaciones entre conceptos.

4.8.1.DISTRIBUCIÓN BIVARIANTE, MARGINAL Y CONDICIONAL

En relación a los conceptos de variable estadística, distribución (bidimensional, marginal y condicionada) y frecuencias asociadas, se han encontrado propiedades de las frecuencias o distribución marginal o condicional, su organización tabular o gráfica, así como del centro de gravedad, que se describen a continuación.

PP11. Los datos bidimensionales

En esta categoría incluimos las siguientes propiedades, que sirven para identificar las variables que conforman la variable estadística bidimensional, su análisis unidimensional, así como reglas para su organización tabular y gráfica.

PP11a. Los datos bidimensionales se forman con pares de valores de cada una de las variables estadísticas de estudio. Con esta propiedad se establece una relación entre variable estadística bidimensional y las variables estadísticas unidimensionales que la conforman, por ejemplo:

> Los datos de una distribución bidimensional son pares (x_1,y_1), (x_2,y_2),…,(x_n,y_n), donde x_1, x_2,…,x_n son valores de la variable X, y donde y_1, y_2,…,y_n son los valores de la variable Y ([H2], p.244).

PP11b. Organización de datos en tablas de frecuencias. En algunos textos encontramos propiedades que permiten organizar la información en las tablas estadísticas, generalmente mediante un ejercicio resuelto o un ejemplo:

> Cuando el número de datos de una distribución bidimensional es pequeño, se trabaja con los datos ordenados, pero cuando el número de datos es grande, se trabaja con tablas de frecuencias. Dichas tablas pueden darse de dos maneras:
> a) Tablas simples: se recogen en fila o columna las frecuencias de los datos.
> b) Tablas de doble entrada: se recogen en cada casilla la frecuencia correspondiente a cada fila y cada columna de los valores de cada variable ([T2], p.317).

Rara vez encontramos propiedades necesarias para la propia construcción de la tabla, ya que, en su mayoría, se describe tan sólo su procedimiento de construcción. A modo de ejemplo, citamos algunas:

> Ten en cuenta que si la variable X toma n valores, x_1, x_2,…,x_n, y la variable Y toma m valores y_1, y_2,…,y_n, en la tabla de frecuencias absolutas habrá $m{\cdot}n$ elementos ([H4], p.218).
> En la última fila de la tabla aparece el resultado del sumatorio de cada columna ([H4], p.223).
> La frecuencia de cada par se indica en el cruce de filas (x) y columnas (y) ([H5], p.250).

PP11c. Organización de datos en gráficos. Algunos textos incluyen propiedades de los gráficos que presentan, aunque en su mayoría, se limitan a presentar su procedimiento de construcción. Algunas de estas propiedades son, por ejemplo:

> Se suele utilizar [diagrama de barras tridimensional] para representar datos no agrupados en intervalos. Se suelen utilizar [histograma tridimensional] para representar datos agrupados en intervalos ([H3], p.219).
> El tamaño de los dibujos [pictograma tridimensional] ha de ser proporcional a las cantidades representadas para no provocar una distorsión de la realidad ([H3], p.219).

PP11d. Distribuciones marginales de datos. Así, en [H4] y [T4] se indica que se puede realizar un análisis unidimensional de cada variable, con lo que se obtienen los resúmenes estadísticos correspondientes. En nuestra opinión, esta propiedad se podría haber extendido también a las distribuciones condicionadas.

PP12. Relaciones entre tipos de frecuencias

Se incluyen en esta categoría las propiedades que relacionan los diferentes tipos de frecuencias de los datos en una distribución bidimensional.

PP12a. La suma de las frecuencias absolutas relativas de todos los datos es 1. Esta propiedad, que es una extensión de la correspondiente en una distribución unidimensional al caso bidimensional, aparece sólo en [H4] y [T4].

PP12b. La suma de las frecuencias absolutas de una fila (columna) es la frecuencia absoluta del valor de Y (X) correspondiente a esa fila (columna). Esta proposición permite al estudiante comprender los conceptos de frecuencia marginal y distribución marginal, y en la mayoría de los textos se acompaña de ejercicios resueltos o ejemplos. Se generaliza en [H3] del siguiente modo:

> Por tanto, la última fila y la última columna de la tabla de doble entrada contienen, respectivamente, las frecuencias absolutas de las variables X e Y, consideradas por separado ([H3], p.217).

PP13. Centro de gravedad

En la mayoría de los textos se describe el centro de gravedad de la distribución bidimensional, acompañado de un ejemplo o ejercicio resuelto para mostrar el procedimiento para calcularlo, formado por las medias de las distribuciones marginales. Además, encontramos en [H5] y [T5] el centro medio ponderado, concepto que no existe en matemáticas. Se trata en realidad de un algoritmo de cálculo abreviado de la media cuando se repiten valores de la variable, o para calcular la media de un grupo cuando se conocen las medias de los subgrupos de diferente tamaño.

Consideramos la propiedad de que el centro de gravedad no tiene por qué ser un punto de la distribución, incluida en [H1] y [T1]. Esta propiedad de la media aritmética (no es una operación interna en el conjunto de datos), supone una dificultad para algunos estudiantes (Cobo, 2003), y en raras ocasiones se encuentra en los textos.

Síntesis de las propiedades referidas a la distribución de datos

En la Tabla 4.8.1 presentamos un resumen de las propiedades que se incluyen en el tema, referidas a la distribución de datos bidimensionales. La mayoría de los textos incluyen la propiedad del centro de gravedad, la interpretación de los datos bidimensionales como compuestos por valores de la variable unidimensional, y propiedades de la representación tabular. Con menos frecuencia se presentan propiedades de las distribuciones de datos. Sólo una editorial presenta propiedades de la representación gráfica, siendo más habitual en el resto de los libros presentar el procedimiento de construcción de representaciones, sin explicar las reglas al respecto. También encontramos un sólo caso en el que se introduce propiedades de frecuencias.

No encontramos diferencias significativas entre los textos de la misma editorial dirigidos a cada modalidad de Bachillerato, salvo en [H3] y [T3]. Y destacamos los textos [H4] y [T4] por ser los que más propiedades incluyen, relativas a la distribución de datos bivariados.

Tabla 4.8.1. Propiedades de la variable estadística y distribución bidimensional

Propiedades	H1	H2	H3	H4	H5	H6	H7	H8
PP11.Datos bidimensionales.								
PP11a. Los datos se refieren a cada una de las variables		x	x	x		x	x	x
PP11b. Organización de datos en tablas de frecuencias	x	x		x	x			x
PP11c. Organización de datos en gráficos			x					
PP11d. Distribuciones de datos				x	x	x	x	
PP12 Relaciones entre tipos de frecuencias								
PP12a. La suma de frecuencias absolutas relativas dobles es 1				x				
PP12b. Frecuencia doble - frecuencia marginal			x					
PP13.Centro de gravedad	x	x		x	x	x	x	

	T1	T2	T3	T4	T5	T6	T7	T8
PP11.Datos bidimensionales.								
PP11a. Los datos se refieren a cada una de las variables		x	x	x		x	x	x
PP11b. Organización de datos en tablas de frecuencias	x	x	x	x	x			x
PP11c. Organización de datos en gráficos			x					
PP11d. Distribuciones de datos				x	x	x	x	
PP12 Relaciones entre tipos de frecuencias								
PP12a. La suma de frecuencias absolutas relativas dobles es 1				x				
PP12b. Frecuencia doble - frecuencia marginal								
PP13.Centro de gravedad	x	x		x	x	x	x	

4.8.2. DEPENDENCIA ENTRE DOS VARIABLES

Incluimos en esta categoría las propiedades relativas a la dependencia funcional o estadística e independencia, covarianza y correlación. Generalmente van acompañadas de ejemplos, apoyados en un diagrama de dispersión para justificar su validez. Las relacionadas con la dependencia a veces se presentan de manera conjunta, como se muestra en la Figura 4.8.1. La intención es que el estudiante pueda comparar gráficamente estas propiedades.

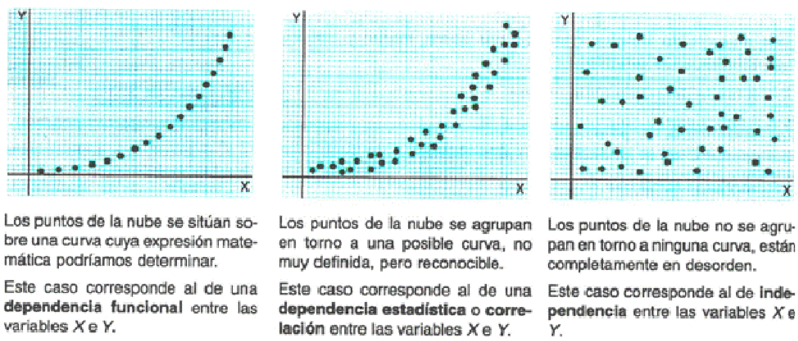


Figura 4.8.1. Interpretación del coeficiente de correlación lineal de Pearson ([H3], p. 222)

PP21. Propiedades de la dependencia funcional

Los textos suelen presentar criterios para identificar la dependencia funcional a través del diagrama de dispersión (Figura 4.8.1). Observamos que en esta descripción se podría inducir la concepción algebraica de la asociación (en este caso de la dependencia funcional) descrita por Estepa (1994), pues es posible la existencia de una relación funcional, sin que pueda expresarse mediante una expresión algébrica. Encontramos variantes de esta propiedad en otros textos, que en su mayoría manifiestan la exigencia de que todos los puntos verifiquen una misma función: *"La relación funcional se* *cumple siempre: globalmente y para cada valor particular*” ([H5], p.250).

PP22. Propiedades de la dependencia aleatoria o estadística

No es muy habitual encontrar propiedades de la dependencia estadística, pues la mayoría de los textos identifican dependencia estadística con correlación. En la Figura 4.8.1 se puede observar una propiedad de este tipo de dependencia en los textos, aunque expresada en forma imprecisa: “*Los puntos de la nube se agrupan en torno a una posible curva, no muy definida, pero reconocible*” ([H3], p.222). En el caso de los textos [H5] y [T5] se indica: “*La relación estadística se cumple sólo de una manera global, para cada valor particular la respuesta puede ser múltiple.*” ([H5], p.250).

PP23. Propiedades de la independencia

En la mayoría de los libros analizados encontramos propiedades de la independencia, que se identifica con el caso de variables incorreladas . Por ejemplo, en [H4] y [T4] se indica: “*la correlación es nula cuando no existe ninguna relación entre las variables. En este caso se dice que las variables están incorreladas.*” ([H4], p.224). Un caso curioso es el de los textos [H5] y [T5], ya que no definen la independencia, pero sí ofrecen una propiedad para interpretarla: “*En este caso, la nube no adopta una forma definida: no hay correlación (o es muy débil).*” ([H5], p.251). Otro ejemplo de esta propiedad se muestra en la Figura 4.8.1.

PP24. Propiedades de la covarianza

En algunos textos encontramos propiedades de la covarianza, en términos de los productos $(x_i - \overline{x}) \cdot (y_i - \overline{y})$ y de su posición en el diagrama de dispersión Así, la covarianza será positiva o negativa, dependiendo de la disposición de los puntos en el diagrama de dispersión, que se divide en cuatro cuadrantes, mediante las rectas paralelas a los ejes de coordenadas que pasan por las medias marginales. Utilizando esta representación, y el hecho de que las coordenadas de los puntos estén por encima o debajo (derecha o izquierda) de dichas rectas, se razona el signo del producto de las diferencias $(x_i - \overline{x})$ e $(y_i - \overline{y})$. De este modo, y aplicando la regla de los signos en el producto de números enteros, los sumandos en la fórmula de la covarianza corresponderán a productos positivos o negativos, según al cuadrante al que pertenezcan (Figura 4.9.6).

PP25. Propiedades de la correlación

En la mayoría de los textos encontramos propiedades de la correlación, principalmente, su grado (intensidad) y signo (sentido); y en menor medida, el tipo de función de ajuste. Cuanto mayor es la dispersión en los datos en el diagrama de dispersión, menor es la intensidad de la correlación, y al contrario. Además, se podrá interpretar el sentido de la correlación según la tendencia de la nube de puntos sea creciente o decreciente; así como el tipo de correlación, según se distribuyan los puntos del gráfico en torno a una línea o a una curva conocida.

PP25a. Correlación y causalidad. Incluimos en esta categoría la diferenciación entre correlación y causalidad, que se incluye al final del tema en [H8], casi como una sección de ampliación de conocimiento, pero que no se incluye en el libro de la misma editorial editado para el Bachillerato científico-tecnológico.

PP26. Coeficiente de correlación lineal

Son varias las propiedades que encontramos en los textos analizados, como se muestra a continuación.

PP26a. El coeficiente de correlación mide la correlación lineal entre las variables, es decir, no es útil para medir una correlación curvilínea. En algunos textos como [H3] y [T3] encontramos esta propiedad, una vez introducida la fórmula de la covarianza. En otros casos, como [H5] y [T5], se presenta al introducir la regresión: "*Este coeficiente mide exclusivamente la correlación lineal entre variables.*" ([H5], p.255).

PP26b. Interpretación del signo y valor absoluto del coeficiente de correlación lineal. Todos los textos, mediante ejemplos y formalmente, incluyen la interpretación del grado y sentido de la correlación, según los valores que tome el coeficiente de correlación. Coinciden estos resultados con la investigación de Sánchez Cobo (1999), donde diez textos (de once) la incluyen. Se ofrece el rango de valores del coeficiente, y su relación con una intensidad, alta o baja, mediante el diagrama de dispersión (Figura 4.8.2). Así, por ejemplo, en [H4] y [T4] se indica: "*El coeficiente r es un número comprendido entre -1 y 1, y su signo decide el comportamiento de la correlación*" ([H4], p.225). En algunos casos, encontramos recomendaciones para determinar cuándo se considera que la intensidad de la correlación es fuerte, como el siguiente ejemplo: "*Se considera que la correlación es fuerte si |r|>0,85*" ([H2], p.249).

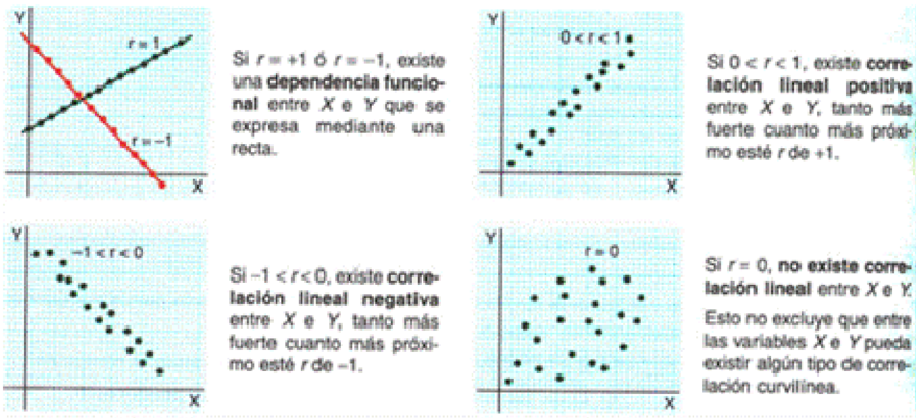


Figura 4.8.2. Interpretación del coeficiente de correlación lineal de Pearson ([H3], p. 224)

PP26c. Adimensionalidad del coeficiente de correlación. Sánchez Cobo (1999) encontró en su investigación, estudiantes que pensaban que al realizar un cambio de origen o escala en las variables, el coeficiente de correlación podría cambiar. Es importante resaltar la adimensionalidad de este coeficiente, que explica su uso preferente, frente a la covarianza que no es adimensional. Sin embargo, esta propiedad aparece poco en los textos analizados. Mostramos un ejemplo:

El coeficiente de correlación de Pearson es un número. No depende de las unidades en las que están expresadas las variables *x* e *y* ([H2], p.249).

Síntesis de las propiedades referidas a la dependencia entre variables

En la Tabla 4.8.2 observamos que, cuando los textos ofrecen propiedades relativas al tipo de dependencia entre las variables, suelen abarcarlas todas.

Tabla 4.8.2. Proposiciones referidas a la dependencia

Propiedades	H1	H2	H3	H4	H5	H6	H7	H8
PP21.Dependencia funcional			x	x	x	x		x
PP22.Dependencia estadística o aleatoria			x	x	x	x		x
PP23.Independencia			x	x	x	x		x
PP24.Covarianza	x	x				x		x
PP25.Correlación	x	x	x	x	x	x		x
PP26.Coeficiente de correlación lineal	x	x	x	x	x	x		x
PP26a. Mide la correlación lineal entre las variables			x	x	x	x	x	x
PP26b. Interpretación del coeficiente de correlación lineal	x	x	x	x	x	x	x	x
PP26c. Adimensionalidad	x	x			x			
	T1	T2	T3	T4	T5	T6	T7	T8
PP21.Dependencia funcional			x	x	x	x		x
PP22.Dependencia estadística o aleatoria			x	x	x	x		x
PP23.Independencia			x	x	x	x		x
PP24.Covarianza	x	x				x		
PP25.Correlación	x	x	x	x	x	x		x
PP26.Coeficiente de correlación lineal	x	x	x	x	x	x		x
PP26a. Mide la correlación lineal entre las variables			x	x	x	x	x	x
PP26b. Interpretación del coeficiente de correlación lineal	x	x	x	x	x	x	x	x
PP26c. Adimensionalidad	x	x			x			

No se encuentran diferencias entre la modalidad de Bachillerato, aunque sí entre los textos de cada modalidad, siendo [H1], [H2], [H7], [T1], [T2], y [T7] los menos completos. Podemos apreciar la gran cantidad de propiedades referidas al coeficiente de correlación lineal, en especial la interpretación del valor absoluto y signo incluida en todos los textos Podríamos decir que nuestros resultados son mejores que en la investigación de Sánchez Cobo (1999), en el sentido de que uno de los textos que investiga el autor no presenta ninguna propiedad referida a este coeficiente. Además, en casi todos los textos se incluyen propiedades de la correlación (en general) y de la dependencia estadística, funcional, e independencia; en menor medida de la covarianza.

4.8.3. ANÁLISIS DE REGRESIÓN

Las propiedades que se presentan en los textos analizados en cuanto al análisis de regresión, se refieren a los modelos de ajuste, el método de mínimos cuadrados, el coeficiente de determinación y la predicción.

PP31. Modelos de regresión

Puesto que los textos no suelen incluir modelos de regresión no lineal, estas propiedades se refieren más a aspectos generales de la fundamentación del análisis de regresión. Un ejemplo, referido a un modelo no lineal es el siguiente:

> Si en una variable (*X,Y)* existe una correlación fuerte entre las variables *X* e *Y*, el análisis de la regresión permite encontrar la ecuación de la función matemática que mejor se ajusta a la nube de puntos. Esta puede ser una recta, una parábola, una exponencial, una cúbica, etc. A partir de aquí nos centraremos en el estudio de la regresión lineal ([H4], p.226).

PP32. Ajuste de la recta de regresión

Las propiedades relacionadas se refieren principalmente al método de mínimos cuadrados, pues sólo [H8] y [T8] incluyen además otros métodos de ajuste a los datos.

PP32a. La recta de regresión hace mínima la suma de los cuadrados de las diferencias de los puntos (abscisas u ordenadas) de la distribución bidimensional a la recta. Todos los textos utilizan el método de mínimos cuadrados para determinar la recta de regresión, y en la mayoría se precisa esta propiedad. Se generaliza la propiedad equivalente de la media (la media es el valor que minimiza la suma de distancias de los datos), puesto que para cada valor de la variable independiente, la recta de regresión proporciona el valor medio teórico de los valores de la variable dependiente. Esta justificación no se aporta en ningún texto, y se argumenta con gráficos (Figura 4.10.3).

PP32b. Dos rectas de regresión diferentes. En todos los textos analizados se diferencian implícitamente las dos rectas de regresión, aunque hay textos que la hacen explícita. Como señala Sánchez Cobo (1999), resulta necesario resaltar esta propiedad, (en su investigación sólo cuatro textos, de once, lo hacen) no siendo suficiente presentar las dos rectas de regresión. De este modo, se podría prevenir que los estudiantes obtuvieran una de las rectas despejando la variable de la otra recta calculada previamente. Un ejemplo se da en [H5] y [T5]:

> Las rectas de regresión de *Y* sobre *X* y de *X* sobre *Y* son distintas, por eso hay que saber qué variable es la dependiente, pues *X* e *Y* no son intercambiables. (Es posible que tenga sentido deducir la variable *Y* a partir de la *X*, mientras que deducir la *X* a partir de la *Y* carezca de significado) ([H5], p.258).

PP33. Coeficiente de determinación

Pocos los textos definen el coeficiente de determinación, por lo que las propiedades referidas al mismo son escasas. La principal es que el coeficiente de determinación informa de la fiabilidad del ajuste lineal. Un ejemplo es:

> La utilidad de la recta de regresión para predecir el valor de la variable dependiente, conocido el valor de la variable independiente, está condicionada a que la relación de linealidad de ambas variables sea adecuada. El coeficiente de correlación lineal indica el grado de linealidad entre las dos variables, pero para analizar la bondad del ajuste de la recta de regresión se utiliza un parámetro nuevo llamado coeficiente de determinación. ([T8], p.326).

Incluimos en esta categoría la propiedad de simetría, que el texto [T8] confiere a este coeficiente ([T8], p.326). Igualmente aquellas propiedades del coeficiente de determinación que se formulan sin haberlo definido. Este es el caso de [H6] y [T6]:

> En ocasiones, con el fin de calcular la calidad o bondad del ajuste realizado mediante la recta de regresión y, por tanto, la fiabilidad de las predicciones que con ella se puedan realizar, se utiliza la expresión ($r^2 \cdot 100$)% que nos da el porcentaje en el que la variable *Y* se justifica por el valor de la

variable *X* ([H6], p.185).

PP34. Predicción

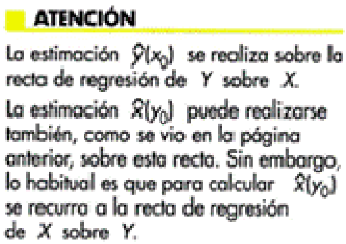
En esta categoría se encuentran las propiedades que regulan el cálculo de las predicciones a través de la recta de regresión.

PP34a. La recta de regresión permite realizar estimaciones de los valores de la variable dependiente a partir de la independiente. Pocos textos precisan la propiedad de la recta de regresión de proporcionar la mejor estimación, dado un valor de la variable independiente. Así, en [H3] y [T3] se precisa: "*La recta de regresión nos permite predecir valores de una variable a partir de los de la otra*" ([H3], p.227).

PP34b. Las estimaciones obtenidas con la recta de regresión son aproximaciones, en términos de probabilidad, al valor real. Esta propiedad no se suele incluir en los textos, aunque en su mayoría se trata de modo implícito, pues se indica que, al contrario del caso de la dependencia funcional, para cada valor de la variable independiente, corresponden varios valores de la dependiente, siendo el valor proporcionado por la recta de regresión el promedio de todos ellos. Así, en los textos [H1] y [T1] se precisa:

> Las estimaciones siempre se realizan aproximadamente y en términos de probabilidad: es probable que si $x = x_0$ entonces *y* valga, aproximadamente, $\hat{y}(x_0)$. ([H1], p.230).

PP34c. La estimación de cada variable se realiza con su correspondiente recta de regresión. Algunos textos señalan que si el coeficiente es próximo a 1 o -1, se puede utilizar una única recta de regresión para predecir cualquiera de las variables del estudio (Figura 4.8.3), aunque no consideramos adecuado establecer estos procedimientos sin una reflexión al respecto, pues el estudiante pudiera no considerar el riesgo de utilizar la misma recta de regresión para estimar una u otra variable en otros casos.



ATENCIÓN

La estimación $\hat{y}(x_0)$ se realiza sobre la recta de regresión de Y sobre X.
La estimación $\hat{x}(y_0)$ puede realizarse también, como se vio en la página anterior, sobre esta recta. Sin embargo, lo habitual es que para calcular $\hat{x}(y_0)$ se recurra a la recta de regresión de X sobre Y.

Figura 4.8.3. Propiedades de la estimación ([H1], p.232)

Síntesis de las propiedades referidas al análisis de regresión

En la Tabla 4.8.3 presentamos un resumen de las propiedades que se incluyen en el tema en el análisis de regresión. En su mayoría se refieren a la recta de regresión, como aquella que minimiza la suma de cuadrados, propiedad que igualmente se presenta en la investigación de Sánchez Cobo (1999) en diez de los once textos que analiza. Igualmente, la mayoría de los textos indican la posibilidad de realizar predicciones a partir de ella, y la existencia de dos rectas diferentes de regresión.

No encontramos diferencias en los textos dirigidos a cada modalidad, salvo en [H8] y [T8], siendo éste último mucho más completo en cuanto a las propiedades que incluye. La razón principal es que este texto define el coeficiente de determinación, por lo que incluye más propiedades referidas a la regresión.

Tabla 4.8.3.1. Propiedades referidas a la regresión

Propiedades	H1	H2	H3	H4	H5	H6	H7	H8
PP31.Modelos de regresión				x	x	x		x
PP32.Ajuste de la recta de regresión								
PP32a. Propiedad de mínimos cuadrados	x	x	x		x	x	x	x
PP32b. Dos rectas de regresión diferentes			x		x	x		
PP33.Coeficiente de determinación					x	x		
PP34.Predicción								
PP34a. La recta de regresión permite realizar estimaciones		x	x	x	x	x	x	
PP34b. Una predicción es una aproximación al valor real	x		x	x			x	
PP34c. Uso de la recta específica para realizar predicciones	x		x	x	x	x	x	

	T1	T2	T3	T4	T5	T6	T7	T8
PP31.Modelos de regresión				x	x	x		x
PP32.Ajuste de la recta de regresión								
PP32a. Propiedad de mínimos cuadrados	x	x	x		x	x	x	x
PP32b. Dos rectas de regresión diferentes			x		x	x		
PP33.Coeficiente de determinación					x	x		x
PP34.Predicción								
PP34a. La recta de regresión permite realizar estimaciones		x	x	x	x	x	x	x
PP34b. Una predicción es una aproximación al valor real	x		x	x			x	
PP34c. Uso de la recta específica para realizar predicciones	x		x	x	x	x	x	

Destacamos los textos [H5], [H6], [T5] y [T6] por ser los más completos en cuanto a las propiedades que incluyen de regresión, matizando que la propiedad que no incluyen sí se presenta de modo implícito ya que, en ambos, se precisa que existe un error entre los valores reales y estimados con la recta de regresión.

4.8.4. RELACIONES ENTRE CONCEPTOS

Cada objeto matemático se puede relacionar con otros por medio de propiedades, que enriquecen su significado. Se describen a continuación las incluidas en los textos.

PP41. Relación distribución - regresión

Se han encontrado algunas propiedades que relacionan la recta de regresión, el centro de gravedad y la valoración de las predicciones que se realizan con la recta según la distribución de los datos.

PP41a. La recta de regresión pasa por el centro de gravedad de la distribución. Esta propiedad suele aparecer como anotación al margen, o bien en ejercicios resueltos y ejemplos, donde se muestra la intersección de las dos rectas de regresión. En el estudio de Sánchez Cobo (1999), siete textos citan esta propiedad, aunque dos lo hacen de modo implícito. El resto no la mencionan, aunque expongan la ecuación de la recta de regresión en la forma de punto-pendiente: $y - \overline{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \overline{x})$ donde sí se evidencia.

PP41b. Las estimaciones con la recta de regresión son mejores en valores cercanos a la media de la variable independiente. Esta regla, para validar la bondad de la predicción con la recta de regresión, aparece en pocos textos. Por lo general, la fiabilidad de la predicción se juzga por la proximidad del coeficiente de correlación a 1 o -1, aunque esta condición es insuficiente. En los textos [H3] y [T3] se señala: "*Las predicciones obtenidas para valores próximos al punto medio de la distribución son más fiables que las obtenidas para valores muy alejados*" ([H3], p.227). Una variante de esta propiedad es la conveniencia de realizar estimaciones para valores contenidos en el intervalo utilizado para determinar la recta de regresión.

PP41c. La predicción con la recta de regresión es más fiable si la recta se calcula a partir de muchos datos. Esta propiedad se presenta en pocos de los textos, a pesar de su utilidad para comprender la incertidumbre que caracteriza a la predicción, a través de la recta de regresión. Como ejemplo, en los textos [H5] y [T5] se señala: "*La fiabilidad aumenta al aumentar los datos. Una recta obtenida a partir de pocos datos genera grandes riesgos, aunque r sea muy alto*" ([H5], p.260).

PP42. Relación dependencia - covarianza/correlación

Describimos a continuación relaciones que aparecen en algunos textos de los tipos de dependencia con la covarianza y/o la correlación, que no se incluyen en el estudio de Sánchez Cobo (1999).

PP42a. La dependencia funcional e independencia son los casos extremos de dependencia estadística o correlación. Tan sólo encontramos esta propiedad en [H3] y [T3], donde se relaciona la dependencia estadística o aleatoria con la correlación:

> Entre los casos extremos de dependencia funcional e independencia existe una amplia gama de situaciones en que se da dependencia estadística o correlación ([H3], p.222).

PP42b. El coeficiente de correlación informa del tipo de dependencia entre las variables. Muchos textos incluyen esta propiedad. Así, los textos relacionan el coeficiente de correlación lineal cuando toma los valores 1 ó -1 con la dependencia funcional, como muestra la siguiente propiedad: "*Si r = 1 o r =-1, la correlación es perfecta o funcional.*" ([H2], p.249); y en menor medida, la relación entre independencia y coeficiente de correlación con valor 0 (Figura 4.8.2), o discutir los valores en los intervalos (-1,0) y (0, 1). En este sentido, destacamos esta proposición:

> Un coeficiente de correlación lineal próximo a cero indica que entre las variables no existe relación lineal, pero puede existir otro tipo de dependencia funcional no lineal. ([T8], p.323).

PP43. Relación dependencia - regresión

Esta propiedad analiza la posición de las rectas de regresión según el tipo de dependencia. Así, en [H3] y [T3] se indica que, las rectas de regresión son coincidentes o perpendiculares dependiendo de si las variables presentan dependencia funcional o independencia, respectivamente (Figura 4.8.2).

PP44. Relación covarianza - correlación - coeficiente de correlación

En la mayoría de los textos encontramos dos propiedades que relacionan estos conceptos que, aunque muy similares, en una influye principalmente la definición operacional de la covarianza mientras que en la otra se destaca su definición estructural (Sfard, 1991):

PP44a. La covarianza informa del sentido de la correlación. Esta propiedad no aparece en el análisis de textos de Sánchez Cobo (1999), y con poca frecuencia en nuestra investigación, siempre en términos de los productos que aparecen en la fórmula (Figura 4.7.6), o bien según el signo de la misma:

> La covarianza es un número del que debes saber:
> - Si $S_{xy} > 0$, la correlación es directa.
> - Si $S_{xy} < 0$, la correlación es inversa ([H5], p.254).

PP44b. El signo de la covarianza coincide con el del coeficiente de correlación. En esta propiedad se compara los signos de la covarianza y el coeficiente de correlación lineal. Se presenta en mayor medida que la propiedad anterior. Esta propiedad no se menciona en los textos que analiza Sánchez Cobo (1999). En nuestra investigación, por ejemplo, en [H4] y [T4] se indica: "*Como las desviaciones típicas son siempre positivas, el signo de r viene dado por el signo de la covarianza*." ([H4], p.225).

PP45. Relación covarianza/correlación - regresión

En el estudio de Sánchez Cobo (1999) ningún texto incluye relaciones entre la intensidad y el sentido de la dependencia y la pendiente de la recta. A continuación se describen las que se presentan en nuestro estudio.

PP45a. La recta de regresión informa sobre el signo de la covarianza. Esta propiedad se refiere a la interpretación de la fórmula de la covarianza según la proximidad de los puntos del diagrama de dispersión a la recta de regresión (Figura 4.7.6), así como la relación entre el signo de la covarianza y los coeficientes de regresión, o la pendiente de las rectas de regresión.

PP45b. La recta de regresión informa de la intensidad y sentido de la correlación. Se puede deducir de la pendiente de la recta de regresión la dirección de la correlación, y en caso de ser horizontal o vertical, la ausencia de interdependencia entre las variables. Esta propiedad se formula en los textos analizados de diversas formas:

> Si la pendiente de la recta de regresión es positiva o negativa, la correlación se llama positiva o negativa, respectivamente ([H1], p.227).
> Si la correlación es nula, los puntos del diagrama de dispersión están distribuidos al azar, sin aproximarse a ninguna línea ([H4], p.224).

PP45c. El producto de los dos coeficientes de regresión es r^2. Esta proposición se encuentra en la mayoría de los textos, y en algunos casos, se incluye su demostración. También se incluye en los textos de la investigación de Sánchez Cobo (1999); el autor menciona la precaución que hay que tener con el uso de esta propiedad, ya que al calcular la raíz cuadrada, los estudiantes pudieran considerar el doble signo (Ver Sección 4.12). Una variante de esta propiedad la presenta [T8], que relaciona el coeficiente de determinación con r^2: "*El coeficiente de determinación es el producto de las pendientes*." ([T8], p.329).

PP45d. Los signos del coeficiente de correlación y regresión coinciden. Debido a su expresión matemática, el coeficiente de correlación tiene el mismo signo que los coeficientes de regresión. En algunos textos, si no se definen los coeficientes de regresión, esta propiedad se formula en términos de la pendiente de la recta. Además, intuitivamente es sencillo razonar que si la relación es directa (inversa), como se deduce del signo de la correlación, la recta es creciente (decreciente), por lo que su pendiente ha de ser positiva (negativa). En el estudio de Sánchez Cobo (1999) tres libros incluyen esta propiedad.

PP45e. Relación entre las estimaciones realizadas con la recta de regresión y el coeficiente de correlación. Es habitual utilizar el coeficiente de correlación para valorar la bondad del ajuste de la predicción con la recta de regresión. Las estimaciones serán mejores cuanto más próximo a uno, en valor absoluto, se encuentre el coeficiente de correlación. Incluimos en esta categoría los textos que indican la conveniencia de hacer estimaciones según la proximidad del coeficiente de correlación a 1 o -1, por lo que esta propiedad se hace implícita. Mostramos un ejemplo de esta propiedad:

> La fiabilidad de los cálculos obtenidos mediante las rectas de regresión será tanto mayor cuanto mayor sea el valor absoluto del coeficiente de correlación lineal r.
> - Si r está muy próximo a cero, no tiene sentido realizar ninguna estimación o previsión.
> - Si r está próximo a 1 o a -1, los valores reales serán, probablemente, próximos a nuestras estimaciones.
> - Si r=1 o r=-1, los valores reales coincidirán con las estimaciones efectuadas ([T4], p.340).

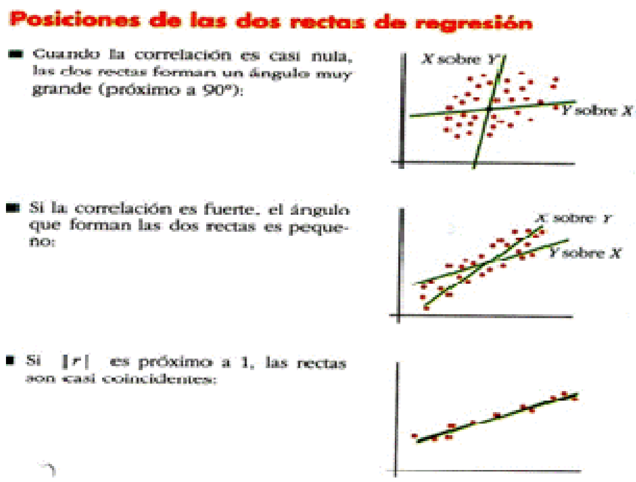


Figura 4.8.4. Coeficiente de correlación y ángulo de las rectas de regresión ([H1], p. 232).

PP45f. Posición relativa de las rectas de regresión y coeficiente de correlación. No es muy habitual encontrar propiedades geométricas referidas a la correlación y regresión en los textos analizados. En [H1] y [T1] encontramos la relación que se muestra en la Figura 4.8.4, que es una propiedad de interés ya que, el coseno del ángulo de las dos rectas es el coeficiente de correlación. Sánchez Cobo (1999) indicó que los estudiantes tienen dificultad para comprender esta propiedad, que por lo general, se acompaña de ejemplos que la justifican.

Síntesis de las propiedades que relacionan conceptos

En la Tabla 4.8.4 observamos, que las relaciones más frecuentes son que las rectas de regresión pasan por el centro de gravedad, y su relación con la bondad de la predicción (PP41a y PP41b), así como con el coeficiente de correlación (PP45e).

Tabla 4.8.4. Relaciones entre conceptos

Proposiciones	H1	H2	H3	H4	H5	H6	H7	H8
PP41.Distribución - regresión								
PP41a. Recta de regresión - centro de gravedad	x	x	x		x	x	x	x
PP41b. Predicción - centro de gravedad	x	x	x		x	x	x	x
PP41c. Tamaño de la muestra - recta de regresión - predicción			x		x			
PP42.Dependencia - covarianza/correlación								
PP42a. Tipos de dependencia - correlación			x					x
PP42b. Tipos de dependencia - coeficiente de correlación		x	x	x		x		x
PP43.Dependencia - regresión			x	x				
PP44.Covarianza - correlación - coeficiente de correlación								
PP44a. Signo de la covarianza - sentido de la correlación					x	x	x	x
PP44b. Signo de la covarianza - signo c. de correlación				x	x	x	x	x
PP45.Covarianza/correlación - regresión								
PP45a. Covarianza - recta de regresión	x				x	x		
PP45b. Correlación - recta de regresión	x		x		x		x	
PP45c. Producto c. regresión - coeficiente de correlación	x		x					
PP45d. Coeficiente de correlación - coeficientes de regresión	x		x	x				x
PP45e. Coeficiente de correlación - predicción	x	x	x	x	x	x	x	
PP45f. Posición rectas de regresión - coef. de correlación	x		x				x	x

	T1	T2	T3	T4	T5	T6	T7	T8
PP41.Distribución – regresión								
PP41a. Recta de regresión - centro de gravedad	x	x	x		x	x	x	x
PP41b. Predicción - centro de gravedad	x	x	x		x	x	x	x
PP41c. Tamaño de la muestra - recta de regresión - predicción			x		x			
PP42.Dependencia - covarianza/correlación								
PP42a. Tipos de dependencia - correlación			x					x
PP42b. Tipos de dependencia - coeficiente de correlación		x	x	x		x		x
PP43.Dependencia – regresión			x	x				
PP44.Covarianza - correlación - coeficiente de correlación								
PP44a. Signo de la covarianza - sentido de la correlación					x	x	x	x
PP44b. Signo de la covarianza - signo c. de correlación				x	x	x	x	x
PP45.Covarianza/correlación – regresión								
PP45a. Covarianza - recta de regresión	x				x	x		
PP45b. Correlación - recta de regresión	x		x		x		x	
PP45c. Producto c. regresión - coeficiente de correlación	x		x					x
PP45d. Coeficiente de correlación - coeficientes de regresión	x		x	x				x
PP45e. Coeficiente de correlación - predicción	x	x	x	x	x	x	x	
PP45f. Posición rectas de regresión - coef. de correlación	x		x				x	x

Las relaciones que menos se establecen se refieren a la discusión de la dependencia funcional e independencia, como casos extremos de la correlación (PP42a y PP43); o la importancia del tamaño de la muestra en el ajuste del modelo predictivo (PP41c). Podemos decir que no existen diferencias importantes entre las modalidades de Bachillerato, salvo en [H8] y [T8], pues al incluir este último la definición del coeficiente de determinación, en un ejercicio resuelto que no se presenta en [H8], se indica que el producto de los coeficientes de regresión coincide con el coeficiente de determinación.

En cuanto los textos de cada modalidad, encontramos diferencias, sobre todo en el tratamiento de la covarianza, ya que son las que menos se presentan en los textos. Además, el tratamiento de los tipos de dependencia en relación a la covarianza, correlación, coeficiente de correlación y regresión es muy pobre. Destacamos así los textos [H3] y [T3] por las propiedades referidas a la relación entre conceptos, siendo [H2] y [T2] los que menos conexiones establecen.

4.9. ANÁLISIS DE LOS PROCEDIMIENTOS

Para dar respuesta a una determinada situación-problema, el sujeto aplica diversas operaciones, algoritmos, técnicas de cálculo, que llegan a automatizarse y son específicas del tipo de problema. Por este motivo, en algunos casos se constituyen en objetos de enseñanza.

El análisis de los procedimientos asociados a la enseñanza y aprendizaje de la correlación y regresión en investigaciones precedentes es escaso. Lavalle, Micheli y Rubio (2006) establecen una clasificación de procedimientos, atendiendo a los conceptos que movilizan, junto a otra complementaria con las interrelaciones entre los conceptos que involucran. En nuestra investigación conectamos las situaciones-problemas descritas en la Sección 4.5 con los procedimientos. El análisis realizado en esta sección se refiere sólo a los procedimientos descritos en los libros, que no incorporan el uso de tecnología. Dedicamos una sección completa (Sección 4.11) al empleo de la tecnología en los textos, incluyendo los procedimientos.

4.9.1. ORGANIZACIÓN DE DATOS BIDIMENSIONALES

Todos los procedimientos relativos a la organización y representación de datos bidimensionales, o la determinación de las distribuciones marginales y condicionadas, se suelen presentar en los textos, como se analiza a continuación.

PC01. Construir una tabla a partir de los datos. Para organizar los datos de la distribución, en la mayoría de los textos (salvo [H1], [H6], [H8], [T1], [T6] y [T8]) se plantean tareas de recuento de datos para obtener sus frecuencias absolutas, y a partir de estas frecuencias, la construcción de la tabla asociada.

La descripción de los procedimientos de construcción es muy variada, y en algunos casos errónea. Se suelen describir mediante ejemplos (como en H4), donde se indica cómo construir una tabla de frecuencias relativas: "*Dividiendo cada valor f_{ij} de la tabla entre 100, que es el número total de individuos de la población, se obtiene la siguiente tabla de frecuencias relativas*" ([H4], p.219). Otros textos muestran el proceso de construcción en un ejercicio resuelto (Figura 4.9.1). Además, encontramos algunos textos como [H3], [H7], [T3] y [T7], que explican el procedimiento de agrupación de los datos en intervalos. Al respecto, mencionamos el texto [H7], ya que incluye un ejercicio resuelto que se titula: "*Cómo se agrupan los datos de variables bidimensionales en intervalos*", aunque no considera que los datos de las variables puedan ser proporcionados por un mismo individuo (Ver Sección 4.12).

Ejemplo

Dada la siguiente distribución bidimensional:

Variable X	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3	3	4	4	4	4
Variable Y	1	1	3	3	3	3	4	4	4	1	1	2	4	4	2	2	1	3	3	3

se obtienen las tablas de frecuencias.

Tabla simple

X	Y	Frecuencia
1	1	2
1	3	4
1	4	3
2	1	2
2	2	1
2	4	2
3	2	2
4	1	1
4	3	3

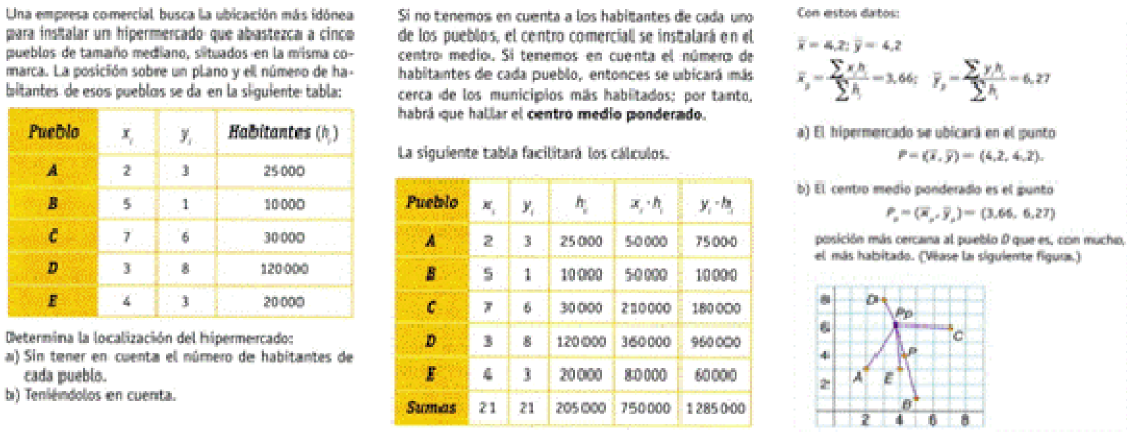
Tabla de doble entrada

Y \ X	1	2	3	4	
1	2	2	0	1	5
2	0	1	2	0	3
3	4	0	0	3	7
4	3	2	0	0	5
	9	5	2	4	20

Figura 4.9.1. Representación tabular de distribución bidimensional ([T2], p. 317)

PC02. Determinar la distribución marginal o condicionada. Son procedimientos variados. Hay textos como [H1] y [T1], que se limitan a etiquetar las tablas de frecuencias de cada distribución marginal. Otros, como [H6] y [T6] explican brevemente cómo obtener estas distribuciones, aunque siguen apoyando sus procedimientos en la tabla de doble entrada, en este caso, mostrándolas con otro color para diferenciarlas del resto de celdas. Mientras que otros incluyen explicaciones más detalladas como la siguiente ([H3], p. 217):

- La suma de las frecuencias absolutas de una columna es la frecuencia absoluta del valor de *X* correspondiente a esa columna.
- La suma de las frecuencias absolutas de una fila es la frecuencia absoluta del valor de *Y* correspondiente a esa fila.
- Por tanto, la última fila y la última columna de la tabla de doble entrada contienen, respectivamente, las frecuencias absolutas de las variables *X* e *Y*, consideradas por separado.



Una empresa comercial busca la ubicación más idónea para instalar un hipermercado que abastezca a cinco pueblos de tamaño mediano, situados en la misma comarca. La posición sobre un plano y el número de habitantes de esos pueblos se da en la siguiente tabla:

Pueblo	x_i	y_i	Habitantes (h_i)
A	2	3	25 000
B	5	1	10 000
C	7	6	30 000
D	3	8	120 000
E	4	3	20 000

Determina la localización del hipermercado:
a) Sin tener en cuenta el número de habitantes de cada pueblo.
b) Teniéndolos en cuenta.

Si no tenemos en cuenta a los habitantes de cada uno de los pueblos, el centro comercial se instalará en el centro medio. Si tenemos en cuenta el número de habitantes de cada pueblo, entonces se ubicará más cerca de los municipios más habitados; por tanto, habrá que hallar el **centro medio ponderado**.

La siguiente tabla facilitará los cálculos.

Pueblo	x_i	y_i	h_i	$x_i \cdot h_i$	$y_i \cdot h_i$
A	2	3	25 000	50 000	75 000
B	5	1	10 000	50 000	10 000
C	7	6	30 000	210 000	180 000
D	3	8	120 000	360 000	960 000
E	4	3	20 000	80 000	60 000
Sumas	21	21	205 000	750 000	1 285 000

Con estos datos:

$\bar{x} = 4,2$; $\bar{y} = 4,2$

$\bar{x}_p = \frac{\sum x_i h_i}{\sum h_i} = 3,66$; $\bar{y}_p = \frac{\sum y_i h_i}{\sum h_i} = 6,27$

a) El hipermercado se ubicará en el punto
$P = (\bar{x}, \bar{y}) = (4,2, 4,2)$.

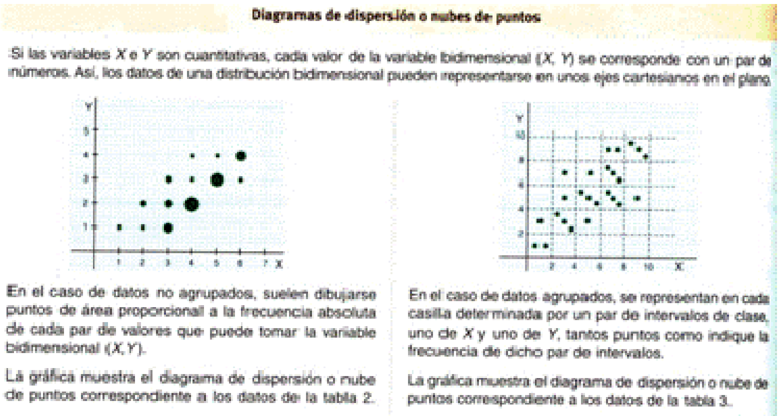
b) El centro medio ponderado es el punto
$P_p = (\bar{x}_p, \bar{y}_p) = (3,66, 6,27)$
posición más cercana al pueblo *D* que es, con mucho, el más habitado. (Véase la siguiente figura.)

Figura 4.9.2. Ejercicio resuelto de cálculo del centro de gravedad ponderado ([T5], p.371)

En algunos libros se pide determinar las distribuciones marginales, a partir de su definición, sin haber presentado previamente ni el procedimiento ni un ejemplo para obtenerlas, como [H8]. En cuanto a las distribuciones condicionales, sólo [H4] y [T4] presentan un procedimiento de cálculo, aunque impreciso. No se especifica que se toma la primera fila/columna y una intermedia porque en la primera están las categorías de la variable que se condiciona, por lo que no es una elección arbitraria. Incluimos en esta categoría el cálculo del centro de gravedad de una distribución bidimensional, que se muestra en [H5] y [T5] mediante un ejercicio resuelto (Figura 4.9.2).

PC03. Construir una gráfica. En todos los textos encontramos tareas de representación gráfica de datos bidimensionales, con procedimientos muy variados. La representación gráfica más utilizada (diagrama de dispersión) tiene una descripción más simple en [H2], [T2], [H5] y [T5]. Por ejemplo, en [H2] y [T2] se indica: "*Una nube de puntos o diagrama de dispersión es la representación en unos ejes cartesianos de los datos (x_i,y_i) de una distribución bidimensional*." ([H2], p.244). Le siguen otras descripciones más completas, por ejemplo, en [H1] y [T1]: "*Si interpretamos cada par de valores como las coordenadas de un punto, el conjunto de todos ellos se llama nube de puntos o diagrama de dispersión*" ([H1], p.227). Otros presentan su construcción detallada, como se muestra en la Figura 4.9.3.

Los textos suelen describir brevemente la construcción del diagrama de barras o histograma. Por ejemplo, en [H1] y [T1] se indica: “*levantando barras de alturas proporcionales a las frecuencias de las correspondientes casillas*” ([H1], p.233), que en el caso de un histograma serían prismas. El procedimiento de construcción se acompaña, en general, de un ejercicio resuelto, con el que se muestra íntegramente el proceso, aunque hay casos, como [H2] y [T2], donde no se indica el procedimiento de construcción, aunque se plantea una tarea de construcción. Creemos, que esto es insuficiente para el aprendizaje significativo del estudiante, ya que hay reglas de construcción difíciles de inferir, como ocurre con la proporcionalidad del volumen de cada prisma según la frecuencia de los datos. Otros textos presentan la construcción del diagrama de barras e histograma de modo detallado, como se mostró anteriormente. El proceso de construcción del pictograma sólo se incluye en [H3].


Diagramas de dispersión o nubes de puntos

Si las variables X e Y son cuantitativas, cada valor de la variable bidimensional (X, Y) se corresponde con un par de números. Así, los datos de una distribución bidimensional pueden representarse en unos ejes cartesianos en el plano.

En el caso de datos no agrupados, suelen dibujarse puntos de área proporcional a la frecuencia absoluta de cada par de valores que puede tomar la variable bidimensional (X, Y).

La gráfica muestra el diagrama de dispersión o nube de puntos correspondiente a los datos de la tabla 2.

En el caso de datos agrupados, se representan en cada casilla determinada por un par de intervalos de clase, uno de X y uno de Y, tantos puntos como indique la frecuencia de dicho par de intervalos.

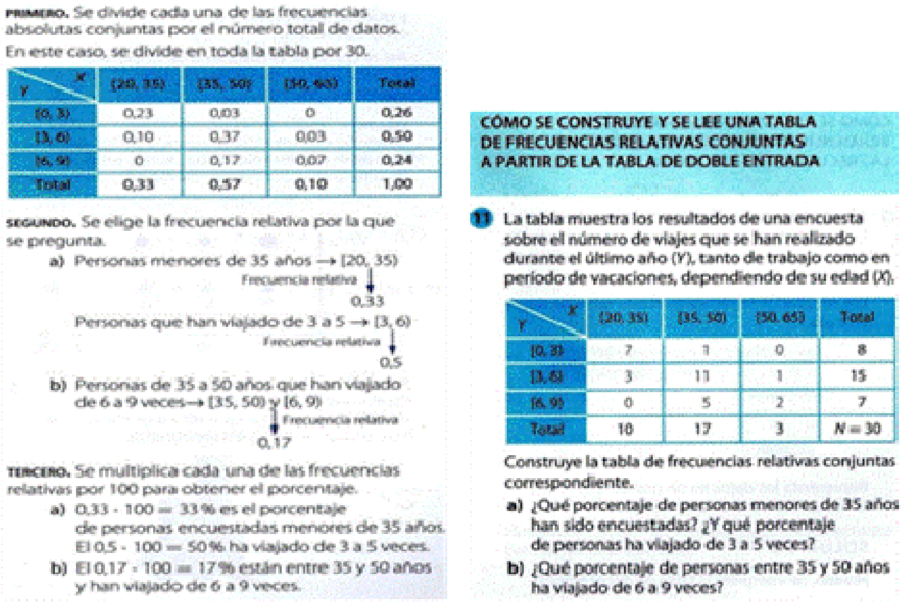
La gráfica muestra el diagrama de dispersión o nube de puntos correspondiente a los datos de la tabla 3.

Figura 4.9.3. Construcción de una nube de puntos/diagrama de dispersión ([H3], p.220)

Finalmente, pocos textos explican la construcción del diagrama de burbuja; en general, se alude al final del tema, o como nota al margen, al modo de representar un diagrama de dispersión para datos con frecuencia distinta de uno. Por ejemplo, en [H1] y [T1] se incluye una sección en la que se describe la tabla de doble entrada, y junto a ella, la construcción del diagrama de burbuja, para el que se precisa su construcción del modo: “*Hinchando los puntos proporcionalmente a su frecuencia*” ([H1], p.233).

PC04. Leer o interpretar representaciones gráfica o tabular. En algunos textos se pide al alumno que identifique ó interprete frecuencias, bien en un gráfico, bien en una tabla, como se mostró en la Figura 4.5.2. Sólo el texto [H7] muestra un ejercicio resuelto (Figura 4.9.4) en que se describe cómo identificar estas frecuencias.

Encontramos textos que proponen tareas de traducción entre representaciones, sin explicar el procedimiento, por ejemplo, [H2] y [T2] (Figura 4.9.1). En [H4] y [T4] se pide al estudiante que elabore la tabla de doble entrada (con frecuencias absolutas y relativas) a partir de una tabla bidimensional con frecuencias.



PRIMERO. Se divide cada una de las frecuencias absolutas conjuntas por el número total de datos.
En este caso, se divide en toda la tabla por 30.

Y \ X	[20, 35)	[35, 50)	[50, 65)	Total
[0, 3)	0,23	0,03	0	0,26
[3, 6)	0,10	0,37	0,03	0,50
[6, 9)	0	0,17	0,07	0,24
Total	0,33	0,57	0,10	1,00

SEGUNDO. Se elige la frecuencia relativa por la que se pregunta.

a) Personas menores de 35 años → [20, 35) $\xrightarrow{\text{Frecuencia relativa}}$ 0,33

Personas que han viajado de 3 a 5 → [3, 6) $\xrightarrow{\text{Frecuencia relativa}}$ 0,5

b) Personas de 35 a 50 años que han viajado de 6 a 9 veces → [35, 50) y [6, 9) $\xrightarrow{\text{Frecuencia relativa}}$ 0,17

TERCERO. Se multiplica cada una de las frecuencias relativas por 100 para obtener el porcentaje.

a) 0,33 · 100 = 33 % es el porcentaje de personas encuestadas menores de 35 años.
El 0,5 · 100 = 50 % ha viajado de 3 a 5 veces.

b) El 0,17 · 100 = 17 % están entre 35 y 50 años y han viajado de 6 a 9 veces.

CÓMO SE CONSTRUYE Y SE LEE UNA TABLA DE FRECUENCIAS RELATIVAS CONJUNTAS A PARTIR DE LA TABLA DE DOBLE ENTRADA

11 La tabla muestra los resultados de una encuesta sobre el número de viajes que se han realizado durante el último año (Y), tanto de trabajo como en periodo de vacaciones, dependiendo de su edad (X).

Y \ X	[20, 35)	[35, 50)	[50, 65)	Total
[0, 3)	7	1	0	8
[3, 6)	3	11	1	15
[6, 9)	0	5	2	7
Total	10	17	3	N = 30

Construye la tabla de frecuencias relativas conjuntas correspondiente.

a) ¿Qué porcentaje de personas menores de 35 años han sido encuestadas? ¿Y qué porcentaje de personas ha viajado de 3 a 5 veces?

b) ¿Qué porcentaje de personas entre 35 y 50 años ha viajado de 6 a 9 veces?

Figura 4.9.4. Lectura de frecuencias en tabla de doble entrada ([H7], p.255)

Síntesis de procedimientos en la organización de datos bidimensionales

En la Tabla 4.9.1 observamos que todos los textos presentan procedimientos referidos a la organización de datos bidimensionales. Los únicos que no describen la construcción de la tabla de frecuencias son [H6] y [T6].

Tabla 4.9.1. Procedimientos de tratamiento de datos bidimensionales

Procedimientos	H1	H2	H3	H4	H5	H6	H7	H8
PC01.Construir una tabla a partir de los datos	x	x	x	x	x		x	x
PC02.Determinar la distribución marginal o condicionada			x	x	x	x	x	
PC03.Construir una gráfica	x	x	x	x	x	x	x	x
PC04.Leer o interpretar frecuencias en representaciones		x					x	
	T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**
PC01.Construir una tabla a partir de los datos	x	x	x	x	x		x	x
PC02.Determinar la distribución marginal o condicionada				x	x	x	x	
PC03.Construir una gráfica	x	x	x	x	x	x	x	x
PC04.Leer o interpretar frecuencias en representaciones		x						

Encontramos diferencias tanto en cada modalidad de Bachillerato ([H3], [H7], [T3] y [T7]) como entre los textos de cada modalidad, siendo [H7] el más completo. La mayoría se refieren a los procedimientos de construcción de representaciones gráfica y tabular, seguido de la determinación de la distribución marginal y condicional.

En cuanto a la lectura e interpretación de frecuencias en representaciones de los textos, en [T7] no se incluye el ejercicio resuelto en el que se describe cómo interpretar las frecuencias de la tabla. Y aunque se propongan tareas de traducción entre representaciones, no se especifican procedimientos para ello. En la investigación de Lavalle, Micheli y Rubio (2006), sólo analizan la construcción del gráfico del diagrama de dispersión, y encontramos resultados similares a los nuestros.

4.9.2. ANÁLISIS DE LA DEPENDENCIA ENTRE DOS VARIABLES

En este tema se encuentran implicados procedimientos como la interpretación del diagrama de dispersión, la covarianza, o el coeficiente de correlación y su cálculo.

PC11. Interpretación del diagrama de dispersión. En todos los textos se encuentran tareas que requieren estimar el grado, sentido y tipo de la dependencia, mediante la observación de la tendencia y dispersión de los datos en el diagrama de dispersión. Algunos textos utilizan la recta de regresión como medio para interpretar el diagrama de dispersión. Incluso en algunos casos, como [H6] y [T6], se ajusta una curva a los datos para mostrar un ejemplo de correlación curvilínea. En [H1] y [T1] se describe cómo interpretar la correlación del siguiente modo ([H1], p.227):

> Puede ser más o menos fuerte según lo apretados que estén los puntos de la nube en torno a una recta que marca la tendencia y se llama recta de regresión. Si la pendiente de la recta de regresión es positiva o negativa, la correlación se llama positiva o negativa respectivamente.

Se suelen describir estos procedimientos mediante ejemplos, que evidencian las propiedades de la correlación, como se muestra en la Figura 4.9.5. Algunos textos son imprecisos ya que, por ejemplo, en [H2] y [T2] se indica que se presenta correlación directa cuando al aumentar una variable, aumenta la otra, e inversa cuando al aumentar una variable la otra disminuye y no se indica qué ocurre cuando disminuye la variable.

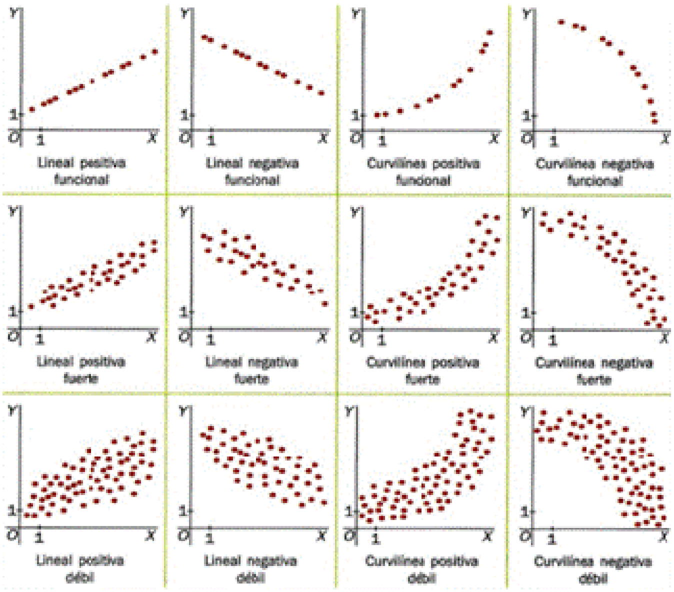


Figura 4.9.5. Interpretación del diagrama de dispersión ([T8], p.322)

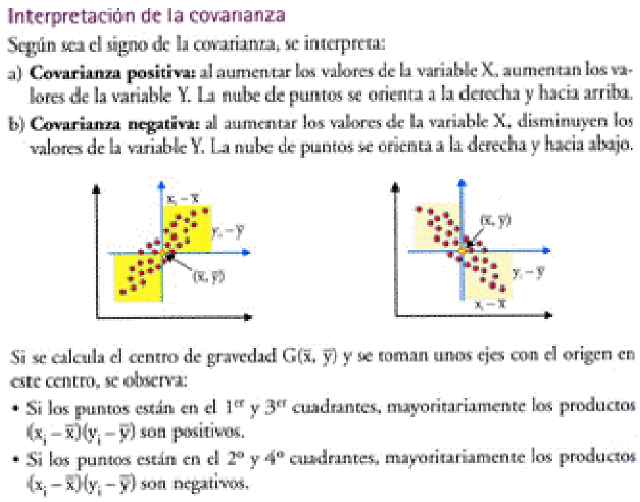
PC12. Cálculo de la covarianza. Todos los textos resaltan la utilidad de la covarianza para medir el grado de dependencia entre dos variables estadísticas, y la fórmula de cálculo, principalmente utilizando la expresión:

$$\sigma_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}$$

En algunos textos se describe el significado de los términos implicados en la fórmula ([H3], [H4], [T3] y [T4]) indicando que x_i e y_i se refieren, bien a valores de las variables *X* e *Y*, bien a las marcas de clase, en caso de datos agrupados. Son cálculos laboriosos, que añaden la dificultad de tener que asociar en un sumatorio el producto de tres elementos: los datos de cada variable (*X* e *Y*) restados a sus correspondientes medias, y la correspondiente frecuencia doble. Pero sólo algunos textos muestran estrategias para facilitar este cálculo (y del coeficiente de correlación). Además, [H1] y [T1] no precisan el rango de variación del sumatorio, lo que pudieran inducir un conflicto semiótico en el estudiante, que podría suponer que los datos tienen un solo índice de variación, $\{(x_1,y_1), (x_2,y_2),\ldots, (x_n,y_n)\}$ aunque los datos provienen de una tabla de doble entrada (Ver Sección 4.12).

En este sentido, en [H6] y [T6] encontramos la estrategia consistente en disponer los datos en un listado, repitiéndolos tantas veces como indique su frecuencia. En el siguiente ejemplo se hace referencia a las marcas de clase, pues se propone en un ejercicio resuelto donde los datos se encuentran en una tabla de datos agrupados:

> Para calcular la recta de regresión, lo primero que hay que hacer es convertir la tabla de doble entrada en una tabla simple y sustituir luego cada intervalo de clase por su marca, tal como se observa en la tabla que figura al pie de la página ([H6], p.190; [T6], p.318).


Interpretación de la covarianza

Según sea el signo de la covarianza, se interpreta:

a) **Covarianza positiva:** al aumentar los valores de la variable X, aumentan los valores de la variable Y. La nube de puntos se orienta a la derecha y hacia arriba.

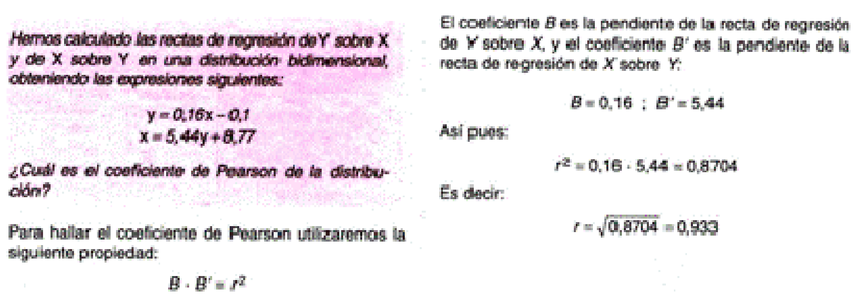
b) **Covarianza negativa:** al aumentar los valores de la variable X, disminuyen los valores de la variable Y. La nube de puntos se orienta a la derecha y hacia abajo.

Si se calcula el centro de gravedad $G(\bar{x}, \bar{y})$ y se toman unos ejes con el origen en este centro, se observa:

- Si los puntos están en el 1er y 3er cuadrantes, mayoritariamente los productos $(x_i - \bar{x})(y_i - \bar{y})$ son positivos.
- Si los puntos están en el 2º y 4º cuadrantes, mayoritariamente los productos $(x_i - \bar{x})(y_i - \bar{y})$ son negativos.

Figura 4.9.6. Interpretación de la covarianza ([H2], p.246)

PC13. Interpretación de la covarianza. Son pocos los textos que describen un procedimiento para interpretar la covarianza, que principalmente se enfoca para discriminar su signo. En su mayoría se acompañan de ejemplos gráficos, como medio de justificación a los argumentos que se ofrecen, En la Figura 4.9.6 se muestra un ejemplo de este tipo de interpretaciones, que se aplica en un ejercicio resuelto.


Hemos calculado las rectas de regresión de Y sobre X y de X sobre Y en una distribución bidimensional, obteniendo las expresiones siguientes:

$$y = 0{,}16x - 0{,}1$$
$$x = 5{,}44y + 8{,}77$$

¿Cuál es el coeficiente de Pearson de la distribución?

Para hallar el coeficiente de Pearson utilizaremos la siguiente propiedad:

$$B \cdot B' = r^2$$

El coeficiente B es la pendiente de la recta de regresión de Y sobre X, y el coeficiente B' es la pendiente de la recta de regresión de X sobre Y:

$$B = 0{,}16 \;;\; B' = 5{,}44$$

Así pues:

$$r^2 = 0{,}16 \cdot 5{,}44 = 0{,}8704$$

Es decir:

$$r = \sqrt{0{,}8704} = 0{,}933$$

Figura 4.9.7. Cálculo del coeficiente de correlación lineal ([H3], p.230)

PC14. Calculo del coeficiente de correlación lineal. Todos los textos presentan la fórmula del coeficiente de correlación lineal como cociente de la covarianza y las desviaciones típicas de las variables. Aunque menos, algunos textos presentan el procedimiento alternativo de cálculo del coeficiente de correlación lineal como producto de los coeficientes de regresión. Destacamos [H3] y [T3], donde se plantean ejercicios, dos de ellos resueltos, que usan este procedimiento (Figura 4.9.7).

PC15. Interpretación del coeficiente de correlación lineal. La interpretación del coeficiente de correlación lineal se incluye en todos los textos analizados. Se indica cómo deben compararse los valores del mismo con los valores extremos 1 y -1, el valor 0 de independencia, junto a la interpretación de su signo. La relevancia de este procedimiento se evidencia, sobre todo, en la cantidad de tareas que se plantean para su interpretación, y en algunos casos, se pide comparar los resultados de este procedimiento con la interpretación previa del diagrama de dispersión (Figura 4.9.8).

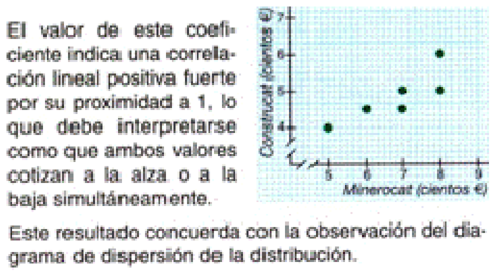


Figura 4.9.8. Interpretación del coeficiente de correlación lineal ([H3], p.225)

Síntesis de procedimientos en el análisis de la dependencia entre dos variables

En la Tabla 4.9.2 podemos observar, que todos los procedimientos se encuentran presentes en los textos analizados, salvo la interpretación que se refiere a la covarianza, que no se incluye en los textos [H3], [H4], [T3], [T4] y [T8]. No se encuentran diferencias significativas entre las modalidades de Bachillerato. Tan sólo en los textos [H8] y [T8], pues en este último se prescinde de la interpretación de la covarianza.

Tabla 4.9.2.1. Procedimientos de análisis de la dependencia

Procedimientos	H1	H2	H3	H4	H5	H6	H7	H8
PC11.Interpretación del diagrama de dispersión	x	x	x	x	x	x	x	x
PC12.Cálculo de la covarianza	x	x	x	x	x	x	x	x
PC13.Interpretación de la covarianza	x	x			x	x	x	x
PC14.Cálculo del coeficiente de correlación lineal	x	x	x	x	x	x	x	x
PC15.Interpretación del coeficiente de correlación lineal	x	x	x	x	x	x	x	x
	T1	T2	T3	T4	T5	T6	T7	T8
PC11.Interpretación del diagrama de dispersión	x	x	x	x	x	x	x	x
PC12.Cálculo de la covarianza	x	x	x	x	x	x	x	x
PC13.Interpretación de la covarianza	x	x			x	x	x	
PC14.Cálculo del coeficiente de correlación lineal	x	x	x	x	x	x	x	x
PC15.Interpretación del coeficiente de correlación lineal	x	x	x	x	x	x	x	x

En relación a otras investigaciones, como la que presentan Lavalle, Micheli y Rubio (2006), nuestros resultados son mejores ya que en su estudio se analizan siete libros de texto, y la covarianza se presenta (no especifican si su cálculo o interpretación) en tres de ellos; el cálculo del coeficiente de correlación lineal en cinco; y la interpretación del mismo en seis. Aún así, la interpretación del diagrama de dispersión la incluyen todos los textos que analizan.

4.9.3. ANÁLISIS DE LA REGRESIÓN

Se incluyen en esta categoría los procedimientos relacionados con el ajuste y predicción, que se describen a continuación.

PC21. Ajuste de la recta de mínimos cuadrados. Todos los textos basan el análisis de la regresión en el ajuste lineal, y el criterio para elegir la mejor recta de ajuste es el método de mínimos cuadrados. Los procedimientos de cálculo se presentan de dos modos, para lo que se requiere haber calculado previamente las medias y varianzas (desviaciones típicas) marginales, y la covarianza. Un procedimiento sigue el formato de ecuación punto-pendiente, que consiste en calcular la pendiente de la recta de regresión mediante las fórmulas de los coeficientes de regresión:

$$\text{Coeficiente de regresión de Y sobre X:} \quad b_{yx} = \frac{\sigma_{yx}}{\sigma_x^2} = r\frac{\sigma_y}{\sigma_x}$$

$$\text{Coeficiente de regresión de X sobre Y:} \quad b_{xy} = \frac{\sigma_{yx}}{\sigma_y^2} = r\frac{\sigma_x}{\sigma_y}$$

Y sustituir en las expresiones:

$$\text{Recta de regresión de Y sobre X:} \quad y - \overline{y} = b_{yx} \cdot (x - \overline{x})$$

$$\text{Recta de regresión de X sobre Y:} \quad x - \overline{x} = b_{xy} \cdot (y - \overline{y})$$

En otros textos, se sustituyen los valores que se obtienen con las fórmulas de la pendiente de la recta de regresión (coeficientes de regresión) y la ordenada en el origen en la ecuación general de la recta:

$$\text{Coeficientes para el cálculo de la recta de regresión de Y sobre X:} \quad b_{yx} = \frac{\sigma_{yx}}{\sigma_x^2} = r\frac{\sigma_y}{\sigma_x} \qquad a = \overline{y} - b_{yx} \cdot \overline{x}$$

$$\text{Coeficientes para el cálculo de la recta de regresión de X sobre Y:} \quad b_{xy} = \frac{\sigma_{yx}}{\sigma_y^2} = r\frac{\sigma_x}{\sigma_y} \qquad a' = \overline{x} - b_{xy} \cdot \overline{y}$$

$$\text{Recta de regresión de Y sobre X:} \quad \hat{y} = a + b_{yx} \cdot x$$

$$\text{Recta de regresión de X sobre Y:} \quad \hat{x} = a' + b_{xy} \cdot y$$

PC22. Ajuste de otros modelos de regresión. Algunos textos mencionan otros modelos de ajuste distintos al lineal, sin precisar su cálculo, aunque indican someramente cómo debiese ser dicho ajuste:

> En algunos casos, sin embargo, la nube de puntos no se distribuye alrededor de una recta, sino de una parábola (regresión parabólica), de una función logarítmica (regresión logarítmica), de una función exponencial (regresión exponencial) o de una función potencial (regresión potencial). ([H6], p.186)

Tan sólo [H8] y [T8] incluyen procedimientos para ajustar otros modelos de regresión. El primero de ellos es el procedimiento desarrollado por Tukey, que se utiliza cuando los datos muestran una tendencia lineal fuerte, y aparecen datos atípicos (outliers) que influyen en los cálculos. El procedimiento se describe mediante un ejemplo (Figura 4.9.9), y se pide en ejercicios su uso, El otro procedimiento es el cambio de variable, y como en el caso anterior, se muestra mediante un ejercicio resuelto. En [H7] los datos presentan una dependencia exponencial, y en [T7] polinómica de grado dos, luego, mediante el cambio de variable y el cálculo de la recta de regresión por el método de mínimos cuadrados, se obtiene el ajuste a los datos correspondiente.

Ejemplo. *Ajusta un modelo lineal a la distribución bidimensional dada por la siguiente tabla.*

X	1	2	3	5	6	7	8	9	10	12	14	21
Y	9	11	13	13	15	14	16	1	16	14	19	2

La variable Y toma dos valores, 1 y 2, que están muy alejados del resto. En la figura se ha representado la recta de regresión de Y sobre X y se aprecia su mal ajuste de la nube de puntos.

La recta de Tukey se calcula del siguiente modo:

1. Se ordenan los datos en orden creciente de las abscisas.
2. Se divide el conjunto ordenado de los datos en tres grupos:
   $G_1 = \{(1, 9)\ (2, 11)\ (3, 13)\ (5, 13)\}$ $G_2 = \{(6, 15)\ (7, 14)\ (8, 16)\ (9, 1)\}$
   $G_3 = \{(10, 16)\ (12, 14)\ (14, 19)\ (21, 2)\}$
3. Para cada grupo G_i se halla el punto $P_i(x_i, y_i)$, donde x_i e y_i son, respectivamente, las medianas de las abscisas y de las ordenadas del grupo G_i, es decir:
   Abscisas de G_1: (1, 2, 3, 5) $\Rightarrow x_1 = 2{,}5$; Ordenadas de G_1: (9, 11, 13, 13) $\Rightarrow y_1 = 12$ $\Rightarrow P_1(2{,}5;\ 12)$
   Abscisas de G_2: (6, 7, 8, 9) $\Rightarrow x_2 = 7{,}5$; Ordenadas de G_2: (1, 14, 15, 16) $\Rightarrow y_2 = 14{,}5$ $\Rightarrow P_2(7{,}5;\ 14{,}5)$
   Abscisas de G_3: (10, 12, 14, 21) $\Rightarrow x_3 = 13$; Ordenadas de G_3: (2, 14, 16, 19) $\Rightarrow y_3 = 15$ $\Rightarrow P_3(13;\ 15)$
4. La recta de Tukey pasa por el baricentro del triángulo P_1, P_2, P_3 y tiene la pendiente de la recta que pasa por P_1 y P_3.
   Baricentro: $x_0 = \frac{2{,}5 + 7{,}5 + 13}{3} = 7{,}67$ $\quad y_0 = \frac{12 + 14{,}5 + 15}{3} = 13{,}83$
   El baricentro tiene coordenadas $G(7{,}67;\ 13{,}83)$.
   La pendiente de la recta que pasa por P_1 y P_3 es: $m = \frac{15 - 12}{13 - 2{,}5} = 0{,}286$
   La ecuación de la recta de Tukey es:
   $$y - 13{,}83 = 0{,}286(x - 7{,}67) \Rightarrow y = 0{,}286x + 11{,}64$$
   Observa que el número de datos en este caso es $n = 12$, múltiplo de 3, y, por tanto, cada grupo está formado por 4 datos. Si el número de datos n no es múltiplo de 3, puede ocurrir que:
   - Sea múltiplo de 3 más 1: en este caso, el grupo G_2 se deja con un dato más.
   - Sea múltiplo de 3 más 2: en este caso se deja el grupo G_2 con un dato menos.

   En el caso de que dos datos tengan la misma abscisa, se dejan en el mismo grupo.

Figura 4.9.9. Procedimiento de ajuste mediante la recta de Tukey a los datos ([H1], p. 231)

PC23. Cálculo del coeficiente de determinación. Pocos textos indican el procedimiento de cálculo del coeficiente de determinación, que acompaña, en todos los casos, a la definición del mismo. Por ello, las definiciones incluidas en los textos ([H5], [H6], [T5], [T6] y [T8]) tienen un carácter operacional, y salvo en [H6] y [T6], estructural (Sfard, 1991). En [H5] y [T5] se define antes del tratamiento de la regresión, como un complemento al análisis de la dependencia entre las variables, e implícitamente se hace alusión a la fiabilidad del ajuste lineal. En cuanto al procedimiento de cálculo, en [H5] y [T5] se señala que:

> Si multiplicamos r^2 por 100, se obtiene el porcentaje de cambio de Y explicado por X. Así, si r = 0, los cambios en la variable X explican el 0% de los cambios en la Y, o sea, nada: las variables X e Y son linealmente independientes. Y si r = 1 (o r = -1), la variación en la Y se explica totalmente, al 100%, por la variación de la X; en este caso, las variables X e Y son linealmente dependientes. Fuera de estos casos límite, el porcentaje explicado es $100 \cdot r^2$ ([T5], p.363).

PC24. Valorar la bondad del ajuste del modelo de regresión. El coeficiente de determinación permite valorar la bondad del ajuste del modelo de regresión, aunque en algunos textos se valora el ajuste tan sólo con interpretar el coeficiente de correlación lineal, por ejemplo:

> En ocasiones, con el fin de calcular la calidad o bondad del ajuste realizado mediante la recta de regresión y, por tanto, la fiabilidad de las predicciones que con ella se puedan realizar, se utiliza la

expresión ($r^2 \cdot 100$)%, que nos da el porcentaje en el que la variable Y se justifica por el valor de la variable X. ([H6], p.185).

PC25. Determinar una predicción. En todos los textos se plantean tareas de estimación a través de la recta de regresión, y los procedimientos que se presentan se refieren a la sustitución de valores en dicha recta. En su mayoría no se explicita de modo formal, sino mediante un ejercicio resuelto, como en la Figura 4.9.10.

Estimamos y para $x = 55 \rightarrow \hat{y}(55) = 0{,}06 + 0{,}119 \cdot 55 = 6{,}6$

(A 55 °C se alarga 6,6 mm).

Estimamos x para $y = 4 \rightarrow \hat{x}(4) = (4 - 0{,}06) : 0{,}119 = 33{,}1$

(Para un alargamiento de 4 mm, hace falta una temperatura de 33,1 °C).

Figura 4.9.10. Predicción de valores a partir de las correspondientes rectas de regresión ([H1], p. 231)

En algunos textos, como [H1], [H2], [T1] y [T2], se indica que, si el coeficiente de correlación es alto, podría utilizarse una sola recta para establecer predicciones (las pendientes de las rectas de regresión son prácticamente las mismas, y poseen un punto en común: el centro de gravedad) Aunque esta propiedad sea cierta, su uso pudiera confundir al estudiante en otros casos (este punto se estudia en la Sección 4.12).

Incluimos en esta categoría la descripción de un procedimiento para estimar coeficientes de un modelo funcional determinado, que tan sólo muestra el texto [T8]. La actividad que se plantea ([T8], p. 335) es estimar el coeficiente de rozamiento entre dos superficies. En las condiciones iniciales de velocidad inicial nula, y con una inclinación de 30º desde la que se desliza un móvil, se ajusta el modelo de regresión lineal, y tras un cambio de variable, se obtiene la estimación de dicho coeficiente.

PC26. Valorar la bondad de la predicción. Cuando se calculan estimaciones bajo el modelo de regresión, es importante que se razone sobre los resultados obtenidos, ya que para algunos casos pueden no tener sentido, como ocurre en el siguiente ejemplo:

> La recta de regresión debe usarse para hacer estimaciones en valores próximos a los considerados. Pretender una estimación en puntos lejanos puede conducir a soluciones absurdas. Por ejemplo, si con la recta anterior estimamos el número de errores que cometería una persona que tardase 30 minutos en teclear las 40 líneas, se obtendría:
> $y = 32{,}251 - 1{,}959 \cdot 30 = -26{,}519$: ¡-27 errores! Absurdo. ([H5], p.260).

El procedimiento que se suele plantear para valorar las predicciones calculadas es comprobar si los valores que se utilizan para realizar la estimación son cercanos a la media, y si el valor del coeficiente de correlación está próximo a 1, en valor absoluto:

- Las predicciones realizadas a partir de una recta de regresión no son fiables si entre *X* e *Y* no hay un alto grado de correlación lineal, es decir, si *r* no es, en valor absoluto, cercano a 1.
- La fiabilidad de una recta de regresión es mayor cuanto mayor sea el número de datos considerados para calcularla. Las predicciones obtenidas para valores próximos al punto medio de la distribución son más fiables que las obtenidas para valores muy alejados. ([H3], p.227).

Síntesis de los procedimientos asociados al análisis de regresión y predicción

En la Tabla 4.9.3 observamos que en todos los textos se incluyen los procedimientos de ajuste de la recta de regresión por el método de mínimos cuadrados, así como los relativos al cálculo de predicciones y la valoración de su bondad. En cuanto a otro tipo de procedimientos, encontramos diferencias principalmente en los textos de cada editorial, ya que entre modalidades, tan sólo el texto [T8] presenta procedimientos relativos al cálculo del coeficiente de determinación y la bondad del ajuste del modelo de regresión, mientras que en [H8] no se incluyen.

Tabla 4.9.3. Procedimientos de análisis de regresión

Procedimientos	H1	H2	H3	H4	H5	H6	H7	H8
PC21.Ajuste de la recta de regresión	X	x	x	x	x	x	x	x
PC22.Ajuste de otros modelos de regresión								x
PC23.Cálculo del coeficiente de determinación					x			
PC24.Valorar la bondad del ajuste del modelo de regresión					x	x		
PC25.Determinar una predicción	x	x	x	x	x	x	x	x
PC26.Valorar la bondad de la predicción	x	x	x	x	x	x	x	x
	T1	T2	T3	T4	T5	T6	T7	T8
PC21.Ajuste de la recta de regresión	x	x	x	x	x	x	x	x
PC22.Ajuste de otros modelos de regresión								x
PC23.Cálculo del coeficiente de determinación					x			x
PC24.Valorar la bondad del ajuste del modelo de regresión					x	x		x
PC25.Determinar una predicción	x	x	x	x	x	x	x	x
PC26.Valorar la bondad de la predicción	x	x	x	x	x	x	x	x

La valoración de la bondad del ajuste del modelo, así como el cálculo del coeficiente de determinación no se suele incluir en los textos analizados. Tan sólo [H5], [T5] y [T8] tratan ambos procedimientos, y en [H6] y [T6] la bondad del ajuste del modelo, sin el cálculo del coeficiente de determinación. En comparación con otras investigaciones relacionadas, como la desarrollada por Lavalle, Micheli y Rubio (2006), observamos que nuestros resultados son mejores que aquellos ya que en sus análisis, de siete libros analizados la ecuación de la recta de regresión aparece en cuatro, la estimación de valores en cinco y la valoración de la bondad de las mismas tan sólo en tres de ellos. En cuanto a procedimientos de ajuste de otros modelos de regresión no lineal, los resultados son similares a los nuestros.

4.10. ARGUMENTOS

La demostración es esencial en matemáticas y debiera ser transmitida a los alumnos a partir de la etapa secundaria (12-13 años) (Crespo y Farfán, 2005). Su papel en el currículo se destaca en los estándares del NCTM (2000), que indican la necesidad de contemplarlos a lo largo de la escolaridad, para desarrollar, con diferentes contextos, la capacidad de argumentación de los estudiantes. También en el Decreto de Enseñanzas Mínimas de Bachillerato (MEC, 2007b) se establece como objetivo considerar las argumentaciones razonadas, y la existencia de demostraciones rigurosas sobre las que se basa el avance de la ciencia y la tecnología, mostrando una actitud flexible, abierta y critica ante otros juicios y razonamientos. Se matiza que:

> Las definiciones formales, las demostraciones (reducción al absurdo, contraejemplos) y los encadenamientos lógicos (implicación, equivalencia) dan validez a las intuiciones y confieren solidez a las técnicas aplicadas (MEC, 2007b, p. 45449).

En esta sección analizamos los argumentos presentados a los estudiantes en los libros de texto, algunos de cuyos resultados hemos publicado en Gea, Batanero, Arteaga, y Cañadas (2013). Siguiendo a Godino y Recio (2001), consideramos necesario ampliar la visión de la demostración, teniendo en cuenta, por tanto, no sólo las demostraciones formales deductivas características de la matemática. Estos autores indican que en los textos de matemáticas, las propiedades y teoremas se consideran necesariamente verdaderos y las argumentaciones que se usan para justificarlos suelen ser informales, no deductivas, e incluso basadas en criterios de autoridad. Por ello, consideraremos otros tipos de justificaciones descritas por estos autores:

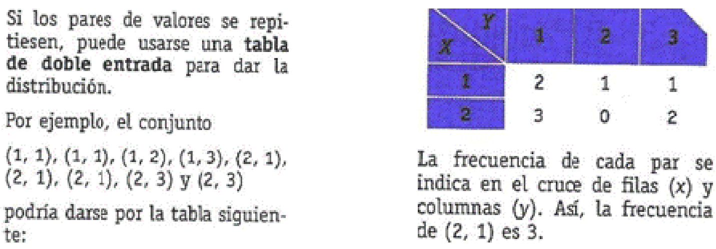
> La palabra *demostración* se utiliza en distintos contextos con diversos sentidos. A veces estos diversos sentidos y matices se reconocen mediante el uso de términos tales como *explicación*, *argumentación*, *prueba*, etc. […] Estas diferencias en las situaciones y prácticas argumentativas indican sentidos distintos del concepto de *demostración* –o bien diversos *objetos demostración* según el modelo ontosemántico adoptado (Godino y Recio, 2001, p. 406),

Una característica de los libros de texto, no sólo universitarios sino también de Bachillerato o Secundaria, es la presencia de justificaciones, que se usan con fines de verificación o convicción (para establecer la verdad de una afirmación), explicación del por qué de un resultado o propiedad, sistematización (para organizar resultados), descubrimiento (llegar a nuevos resultados) o comunicación de los mismos (de Villiers, 1993). Estos argumentos son útiles para validar y hacer comprensibles a los estudiantes los procedimientos, propiedades, definiciones, así como las representaciones que se enlazan en la resolución de problemas sobre las nociones de correlación y regresión.

En el trabajo de Sánchez Cobo (1999) se recogen en todos los textos, al igual que en nuestra revisión. Seguimos la categorización de Recio (1999), quien realiza una revisión de los estudios más importantes relacionados con el tema de la demostración.

A1. Ejemplos o contraejemplos. Es frecuente encontrar ejemplos concretos para validar una determinada propiedad, así como para justificar la solución de un problema. El interés de este tipo de prueba es que ayuda a desarrollar el pensamiento inductivo (examen de ejemplos) para posteriormente generalizar. La combinación de razonamiento inductivo y deductivo en la clase de matemáticas es recomendada por el NCTM (2000). Según Crespo y Farfán (2005), en la mayoría de las ciencias, y en la matemática en particular, se parte de la inducción como método para enunciar proposiciones.

En algunos textos encontramos justificada la necesidad de organizar y representar los datos de una distribución bidimensional, sobre todo la representación tabular. Para justificar esta importancia, principalmente se usan ejemplos como el que se muestra en la Figura 4.10.1, donde el estudiante deberá generalizar el caso particular que se explica, al resto de celdas de la tabla. Este tipo de argumento se usa para validar las definiciones de las distribuciones marginales, como por ejemplo ocurre en los textos [H7] y [T7].


Si los pares de valores se repitiesen, puede usarse una **tabla de doble entrada** para dar la distribución.

Por ejemplo, el conjunto

(1, 1), (1, 1), (1, 2), (1, 3), (2, 1), (2, 1), (2, 1), (2, 3) y (2, 3)

podría darse por la tabla siguiente:

X \ Y	1	2	3
1	2	1	1
2	3	0	2

La frecuencia de cada par se indica en el cruce de filas (*x*) y columnas (*y*). Así, la frecuencia de (2, 1) es 3.

Figura 4.10.1. Justificación mediante ejemplo de las ventajas de la representación tabular ([H5], p. 250)

En el estudio de la existencia de dependencia aleatoria, funcional o independencia, varios textos proporcionan ejemplos para validar o justificar las definiciones que incluyen. Por ejemplo, [H5] y [T5] al definir la dependencia funcional:

> Si se dejan caer dos piedras desde 1 metro de altura ambas tardan el mismo tiempo en llegar al suelo. Y también, si otra piedra se deja caer desde 1,5 metros siempre tarda más en llegar al suelo que las anteriores. La ley de la gravitación se cumple siempre. Por tanto, conocida la altura desde la que se deja caer un objeto se puede saber cuánto tardará en llegar al suelo, con una certeza absoluta ([T5], p.357).

En [H4] y [T4] se utilizan tres diagramas de dispersión (Figura 4.10.2) para justificar las definiciones de independencia y dependencia estadística, así como el grado de dependencia entre las variables. La distinción del grado de dependencia entre las variables se realiza exclusivamente a través de los ejemplos:

> En ocasiones se observa que existe una relación entre las variables, pero dicha relación no puede expresarse como una función matemática. En este caso se dice que entre las variables X e Y existe una dependencia estadística, que podrá ser fuerte o débil. Cuando no existe dependencia funcional ni estadística se dice que hay independencia estadística entre las variables ([H4], p.224).

Este tipo de argumentos se utilizan frecuentemente para justificar propiedades de la correlación. En el caso de la correlación curvilínea, en [T6] y [H6] se muestra un diagrama de dispersión con los datos ajustados por una curva exponencial:

> En algunos casos, sin embargo, la nube de puntos no se distribuye alrededor de una recta, sino de una parábola (regresión parabólica), de una función logarítmica (regresión logarítmica), de una función exponencial (regresión exponencial) o de una función potencial (regresión potencial). El estudio de estas posibilidades escapa a los propósitos de este libro, pero el procedimiento para calcular la línea de regresión consiste también en la aplicación del método de los mínimos cuadrados ([H6], p.186).

De igual modo, [H5] y [T5] incluyen un ejemplo para validar que el coeficiente de correlación lineal tan sólo cuantifica la dependencia lineal entre las variables:

> Por ejemplo, r no detectaría la correlación exponencial perfecta que hay entre los puntos (-1, 0,5), (0, 1), (1, 2) y (4, 16), que pertenecen todos a la gráfica de $y = 2^x$ ([H5], p.255).

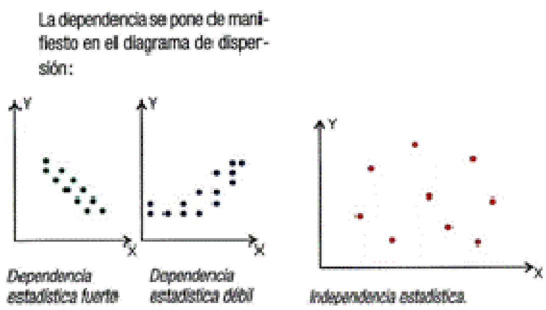


Figura 4.10.2. Justificación de la dependencia estadística y el grado de la relación ([T4], p. 338)

A2. Uso de representaciones gráficas para apoyar una argumentación verbal o simbólica. Cuando por medio de un gráfico se argumenta la verdad o falsedad de una afirmación o de una propiedad; es muy frecuente en los textos. También en el estudio de

Sánchez Cobo (1999) se indica que se suele utilizar la representación gráfica en la argumentación. Este tipo de prueba se incluiría, según Godino y Recio (2011), en las pruebas deductivas informales que incluyen argumentaciones lógicas apoyadas en analogías. Suelen emplearse para el desarrollo de explicaciones y "demostraciones" informales de propiedades de la correlación y regresión. Son muy frecuentes en el análisis de la dependencia entre dos variables aleatorias (tipo, intensidad y signo).

Por ejemplo, [H3] y [T3] analizan la existencia de dependencia mediante los ejemplos que se muestran en la Figura 4.8.1, indicando: "*La relación existente entre dos variables queda reflejada en los diagramas de dispersión o nubes de puntos*". Observamos que se pretende que el alumno generalice, de estos ejemplos, la propiedad presentada; además, se incluye una dependencia curvilínea. Igualmente encontramos ejemplos que justifican características de la correlación (intensidad, signo y tipo) mediante el uso del diagrama de dispersión, así como aquellos en que se relaciona el coeficiente de correlación, las rectas de regresión, y el ángulo que forman (Figura 4.8.4).

Como se ha indicado, algunos textos usan la división del plano en cuatro cuadrantes como medio de apoyo gráfico para justificar gráficamente la relación entre el signo de la covarianza, la forma de la nube, y la dispersión de los puntos respecto a la recta de regresión, mostrando gráficamente la posición de diferentes puntos del diagrama de dispersión respecto a dos rectas paralelas a los ejes, y que pasan por el centro de gravedad de la distribución. Por ejemplo, [H1] y [T1] relacionan estos tres conceptos por medio del soporte gráfico de forma similar a la realizada en Holmes (2001) (Figura 4.7.6). El gráfico se acompaña de una demostración deductiva informal, con poco apoyo algebraico, que comprueba los signos del producto de las diferencias de cada punto al centro de gravedad, dependiendo de su posición en los mencionados cuadrantes. Finalmente, la mayoría de los textos justifican la pertinencia del método de mínimos cuadrados para el ajuste de la recta de regresión en forma gráfica, utilizando un diagrama de dispersión en el que se visualizan las distancias entre los valores reales y los estimados en la recta de regresión como se muestra en la Figura 4.10.3.

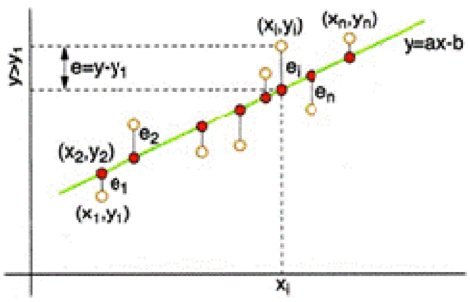


Figura 4.10.3. Relación entre el valor estimado y el valor real en el diagrama de dispersión ([T5], p.364)

A3. Razonamientos verbales deductivos. Se demuestra una propiedad en forma deductiva utilizando propiedades, axiomas o teoremas que el alumno conoce previamente. Este tipo de argumentación es la más utilizada en los textos analizados, como ocurre en el trabajo de Sánchez Cobo (1999). Se trata, generalmente, de razonamientos poco formalizados que, de forma deductiva, justifican una propiedad utilizando principalmente el lenguaje verbal.

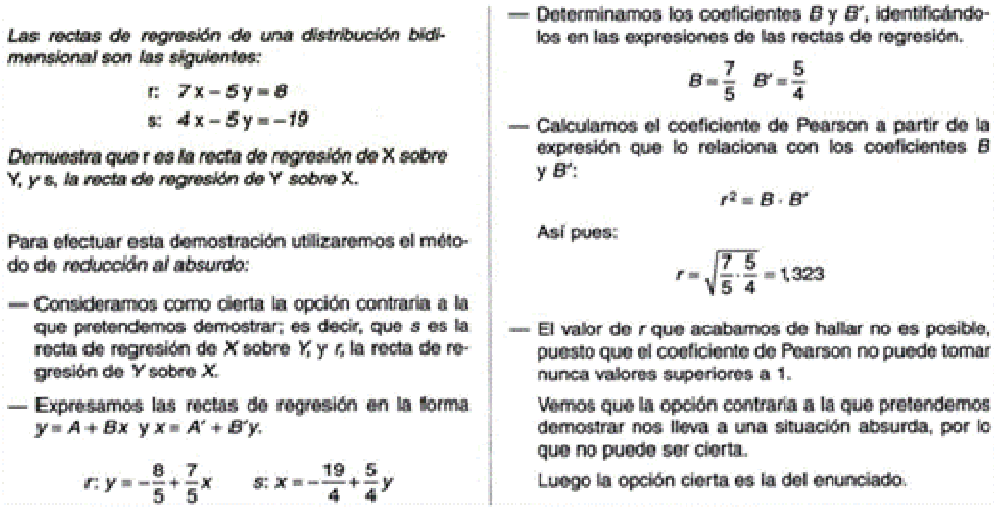
Es muy común encontrar este tipo de argumentación para relacionar la covarianza y el coeficiente de correlación. Por ejemplo, en [H7] y [T7] cuando se indica: "*Como* σ_x *y* σ_y *son siempre positivos, el signo del coeficiente de correlación viene determinado por el signo de* σ_{xy}" ([H7], p. 248). Vemos que se justifica la igualdad del signo de correlación y covarianza sin necesidad de realizar cálculos algebraicos. Otro ejemplo que encontramos de justificación de la misma propiedad es el siguiente:

> Como las desviaciones típicas son siempre positivas, el signo de r viene dado por el signo de la covarianza. ([H4], p.225)

También se encuentra este tipo de argumentaciones para justificar la utilidad del método de mínimos cuadrados. Por ejemplo, en el texto [H3] (p.226) se menciona la necesidad de utilizar un método alejado del subjetivismo de un trazado por ajuste manual. En algunos casos, se utilizan para la justificación de la bondad de las estimaciones realizadas con la recta de regresión:

> Las estimaciones son buenas porque la correlación r = 0,9994 es muy fuerte. Además, x_0 = 55 ºC está entre los valores manejados (entre 0 ºC y 75 ºC) y lo mismo le ocurre a y_0 = 4 mm. No sería buena la estimación para x_0 = 100 ºC, y mucho menos para x_0 = 200 ºC ([H4], p.226; [T4] p.340).

A4. Demostración por reducción al absurdo. Este tipo de demostración parte de la situación contraria a la que se quiere demostrar, y por medio de un razonamiento deductivo, se muestra una contradicción, que lleva a la necesidad de negar la premisa inicial. El fundamento lógico básico de la misma consiste en que, al no poder ser cierta la negación de la tesis, ya que conduce a una contradicción, a un absurdo, se infiere la necesidad de que la tesis sea verdadera (Crespo y Farfán, 2005). En el análisis realizado tan sólo encontramos este tipo de demostración en los textos [H3] y [T3]. Un ejemplo aparece en la Figura 4.10.4, donde se supone que la recta de regresión de *Y* sobre *X* es la de *X* sobre *Y;* a partir de este supuesto, se calculan los coeficientes de regresión, cuyo producto es el coeficiente de correlación, llegándose a un absurdo pues se obtiene un coeficiente cuyo valor absoluto es mayor que la unidad.



Las rectas de regresión de una distribución bidimensional son las siguientes:

$$r: \; 7x - 5y = 8$$
$$s: \; 4x - 5y = -19$$

Demuestra que r es la recta de regresión de X sobre Y, y s, la recta de regresión de Y sobre X.

Para efectuar esta demostración utilizaremos el método de *reducción al absurdo:*

— Consideramos como cierta la opción contraria a la que pretendemos demostrar; es decir, que *s* es la recta de regresión de *X* sobre *Y*, y *r*, la recta de regresión de *Y* sobre *X*.

— Expresamos las rectas de regresión en la forma $y = A + Bx$ y $x = A' + B'y$.

$$r: y = -\frac{8}{5} + \frac{7}{5}x \qquad s: x = -\frac{19}{4} + \frac{5}{4}y$$

— Determinamos los coeficientes *B* y *B'*, identificándolos en las expresiones de las rectas de regresión.

$$B = \frac{7}{5} \quad B' = \frac{5}{4}$$

— Calculamos el coeficiente de Pearson a partir de la expresión que lo relaciona con los coeficientes *B* y *B'*:

$$r^2 = B \cdot B'$$

Así pues:

$$r = \sqrt{\frac{7}{5} \cdot \frac{5}{4}} = 1{,}323$$

— El valor de *r* que acabamos de hallar no es posible, puesto que el coeficiente de Pearson no puede tomar nunca valores superiores a 1.

Vemos que la opción contraria a la que pretendemos demostrar nos lleva a una situación absurda, por lo que no puede ser cierta.

Luego la opción cierta es la del enunciado.

Figura 4.10.4. Resolución de una tarea por el método de reducción al absurdo ([T3], p. 279)

A5. Argumento algebraico deductivo. Rara vez algún texto incluye también argumentos deductivos, básicamente, a partir de lenguaje algebraico. Se trata de manipulaciones algebraicas para tratar de argumentar una propiedad, o mostrar la equivalencia de dos expresiones algebraicas. En el tratamiento de la correlación y regresión se suelen utilizar para demostrar algunas fórmulas de cálculo de la covarianza, coeficiente de correlación lineal y coeficientes de regresión. Un ejemplo se muestra en la Figura 4.10.5, donde, mediante operaciones simbólicas, se prueba la equivalencia entre dos fórmulas de cálculo de la covarianza. El mismo ejemplo aparece en [H8] (p. 251).

Además de desarrollar el producto de dos expresiones en paréntesis, descompone en suma de fracciones algebraicas, desarrollando el producto del factor común en cada sumando. Se vuelve a sacar factor común las medias de *X* e *Y*, y utiliza la propiedad de que la suma de frecuencias es igual a *N* para simplificar en el último término. Finalmente, se ha de identificar las fórmulas de las dos medias, para simplificar de nuevo y llegar a la expresión.

$$s_{xy} = \frac{\sum f_i(x_i - \bar{x})(y_i - \bar{y})}{N} \quad \text{ó} \quad s_{xy} = \frac{\sum f_i x_i y_i}{N} - \bar{x}\,\bar{y}$$

El paso de la primera a la segunda expresión se hace de la siguiente forma:

$$s_{xy} = \frac{\sum f_i(x_i - \bar{x})(y_i - \bar{y})}{N} = \frac{\sum f_i(x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\,\bar{y})}{N} =$$

$$= \frac{\sum f_i x_i y_i}{N} - \bar{y}\cdot\frac{\sum f_i x_i}{N} - \bar{x}\cdot\frac{\sum f_i y_i}{N} + \bar{x}\,\bar{y} = \frac{\sum f_i x_i y_i}{N} - \bar{y}\,\bar{x} - \bar{x}\,\bar{y} + \bar{x}\,\bar{y} = \frac{\sum f_i x_i y_i}{N} - \bar{x}\,\bar{y}$$

Figura 4.10.5. Demostración de la equivalencia de dos expresiones de la covarianza ([T8], p. 321)

En cuanto a la relación entre el coeficiente de correlación lineal y los coeficientes de regresión, habitualmente se encuentra la siguiente cadena de igualdades, como ocurre en los textos [H1] (p.232) y [T1] (p.338):

$$\frac{\sigma_{xy}}{\sigma_x^2}\cdot\frac{\sigma_{xy}}{\sigma_y^2} = \left(\frac{\sigma_{xy}}{\sigma_x\sigma_y}\right)^2 = r^2$$

Síntesis de la presentación de argumentos

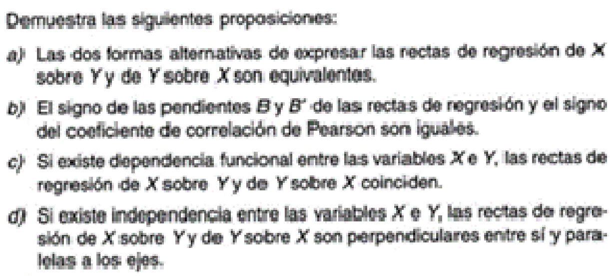
En las Tabla 4.10.1 se resumen los resultados del análisis de los argumentos realizado, donde únicamente se muestra la presencia de cada uno de los tipos de argumentos anteriormente descritos. Observamos que todos los textos utilizan ejemplos y contraejemplos, o bien gráficos, como soporte de argumentación. Igual ocurre en otros trabajos sobre libros de texto, aunque sean de nivel universitario, como los de Alvarado (2007), sobre el teorema central del límite, y el de Olivo (2008), sobre intervalos de confianza. Encontramos también que en todos los textos se hace uso de argumentaciones verbales deductivas, en general, con poca formalización.

Destacamos los textos [H3] y [T3], dado que son los únicos en tratar el método de reducción al absurdo, además, proponiendo a los estudiantes que realicen alguna demostración, utilizándolo y explicando los pasos en que se llevaría a cabo de un modo general. Además, un número de textos utilizan también argumentos algebraicos deductivos; siendo [T3] el único texto que utiliza todas las formas de argumentación que hemos analizado, y es el más formalizado en sus argumentos.

Tabla 4.10.1. Argumentos en los textos analizados

Argumentos	H1	H2	H3	H4	H5	H6	H7	H8
A1.Ejemplos/contraejemplos	x	x	x	x	x	x	x	x
A2.Gráficos auxiliares	x	x	x	x	x	x	x	x
A3.Verbales deductivos	x	x	x	x	x	x	x	x
A4.Reducción al absurdo			x					
A5.Algebraicos deductivos	x					x	x	x
	T1	T2	T3	T4	T5	T6	T7	T8
A1.Ejemplos/contraejemplos	x	x	x	x	x	x	x	x
A2.Gráficos auxiliares	x	x	x	x	x	x	x	x
A3.Verbales deductivos	x	x	x	x	x	x	x	x
A4.Reducción al absurdo			x					
A5.Algebraicos deductivos	x		x		x	x	x	x

Encontramos algunas diferencias en el tipo de argumentación utilizada por la misma editorial, dependiendo de la modalidad de Bachillerato para el cual se diseñen cada uno. Así es que, algunos textos de la modalidad de ciencias y tecnología tratan mucho más la argumentación de tipo simbólico que los correspondientes a la misma editorial de la modalidad de humanidades y ciencias sociales.


Demuestra las siguientes proposiciones:

a) Las dos formas alternativas de expresar las rectas de regresión de *X* sobre *Y* y de *Y* sobre *X* son equivalentes.

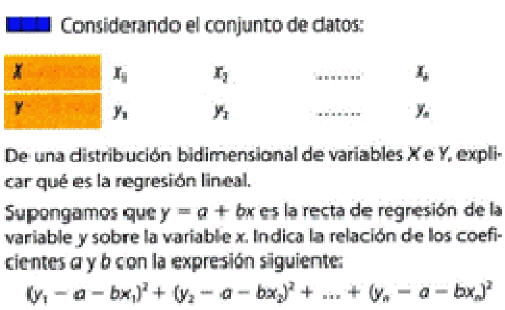
b) El signo de las pendientes *B* y *B'* de las rectas de regresión y el signo del coeficiente de correlación de Pearson son iguales.

c) Si existe dependencia funcional entre las variables *X* e *Y*, las rectas de regresión de *X* sobre *Y* y de *Y* sobre *X* coinciden.

d) Si existe independencia entre las variables *X* e *Y*, las rectas de regresión de *X* sobre *Y* y de *Y* sobre *X* son perpendiculares entre sí y paralelas a los ejes.

Figura 4.10.6. Tarea de demostración de propiedades ([T3], p.278)

Los textos [T3] y [T5] se diferencian de sus respectivos [H3] y [H5] por incluir tareas donde se pide al estudiante realizar alguna demostración utilizando un tipo específico de argumento, a pesar de no incluir argumentos simbólicos deductivos desarrollados en el tema (un ejemplo se presenta en la Figura 4.10.6). En el caso de [T5], se plantea un ejercicio al final del tema donde se pide: “*Demuestra que las dos fórmulas dadas para la covarianza son equivalentes*” ([T5], p. 372).


Considerando el conjunto de datos:

X	x_1	x_2	....	x_n
Y	y_1	y_2	....	y_n

De una distribución bidimensional de variables *X* e *Y*, explicar qué es la regresión lineal.

Supongamos que $y = a + bx$ es la recta de regresión de la variable *y* sobre la variable *x*. Indica la relación de los coeficientes *a* y *b* con la expresión siguiente:

$$(y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + \ldots + (y_n - a - bx_n)^2$$

Figura 4.10.7. Tarea para justificar mediante lenguaje simbólico una relación ([T6], p.321)

Este tipo de tareas no son exclusivas de la modalidad científico-tecnológica, ya que los textos [H6] y [H7] las incluyen, al igual que sus respectivos [T6] y [T7]. Un ejemplo se presenta en la Figura 4.10.7, que se muestra tanto en [H6] como en [T6], donde el estudiante no sólo ha de argumentar utilizando un lenguaje algebraico deductivo, sino incluso ha de descubrir la propiedad que se quiere demostrar. En [H7] y [T7] se pide al estudiante que investigue sobre dos tareas, para las que necesitará hacer uso de lenguaje simbólico, sobre todo en la tarea (b). Como en el anterior ejemplo, en estas dos actividades el alumno ha de descubrir qué es lo que tiene que demostrar:

> Investiga sobre las siguientes cuestiones.
> a) ¿Es cierto que el signo de las pendientes de las dos rectas de regresión de una variable bidimensional es siempre igual?
> b) ¿Qué sucede si las dos rectas de regresión tienen la misma pendiente? ¿Cómo es la correlación? ([H7], p.263)

4.11. RECURSOS TECNOLÓGICOS EN LA CORRELACIÓN Y REGRESIÓN

Actualmente existe una gran variedad de recursos tecnológicos, que facilitan la enseñanza y aprendizaje de la estadística como la calculadora, hoja de cálculo Excel, applets, recursos multimedia y programas de ordenador específicos. El uso de la tecnología se recomienda en España en las orientaciones curriculares de Bachillerato (MEC, 2007b), sugiriendo que pueden servir de ayuda, tanto para la comprensión de conceptos, como para la resolución de problemas, agilizando los cálculos y favoreciendo la interpretación de los resultados. Mas aún, la Junta de Andalucía recomienda:

> Formación para la utilización de las tecnologías de la información y la comunicación, estimulando su uso en los procesos de enseñanza y aprendizaje de todas las materias y en el trabajo del alumnado (Consejería de Educación, 2008b, pp. 9-10).

La tecnología se tiene en cuenta en los textos de tres modos diferentes: 1) En el planteamiento o resolución de problemas, o descripción de procedimientos dentro del texto; 2) Mediante referencias a recursos en Internet; y 3) Mediante un anexo en forma de CD que se incluye en muchos de los textos con diverso contenido. En esta sección analizamos estos modos de uso de la tecnología.

4.11.1. USO DE LA TECNOLOGÍA AL PRESENTAR LOS CAMPOS DE PROBLEMAS Y PROCEDIMIENTOS

En las tareas que se plantean en el tema, tan sólo se ha encontrado el uso (o su propuesta de uso) de la calculadora y la Hoja de Cálculo Excel. Observamos, en la Tabla 4.11.1, un uso explícito escaso de recursos tecnológicos en estas tareas (que se amplía poco en el CD). En particular hay poco uso de Excel, a pesar de que se recomiende su uso en los diseños curriculares, y de su utilidad, ya que dispone de herramientas muy completas y sencillas de usar para el análisis de correlación y regresión, incluyendo el cálculo de los parámetros de diferentes funciones de ajuste; representación de los diagramas de dispersión y otros gráficos; interpolación y extrapolación. Hay diferencias notables en los libros, pues los textos [H2] y [T2] apoyan el 24% de situaciones en la tecnología. Destacamos también [H4], [T4], [H8] y [T8], por el tratamiento que se lleva a cabo de Excel dentro del tema.

Algunos textos hacen referencia a la tecnología al presentar los procedimientos.

Por ejemplo, en [H2], [H4], [H8], [T2], [T4] y [T8] se incluye una descripción del programa Excel, junto a tareas donde se describe el modo de obtener el diagrama de dispersión o la covarianza. Lo curioso es que, ninguno menciona la posibilidad de realizar otros gráficos mediante la hoja de cálculo. Los textos [H2], [H3], [H8], [T2], [T3] y [T8] describen el procedimiento de predicción con Excel (Figura 4.11.1); En [H2] y [T2] se incluyen los procedimientos utilizando Windows Calc (Linux).

Tabla 4.11.1 Tareas con uso de tecnología

Recurso	H1	H2	H3	H4	H5	H6	H7	H8
Excel		48(21.7)	2(0.8)	10(4.4)				17(5.7)
Calculadora	13(4.9)	5(2.3)			15(4.7)	4(2.3)	17(4.2)	
No usa	255(95.1)	168(76)	256(99.2)	215(95.6)	303(95.3)	172(97.7)	385(95.8)	280(94.3)
Total	268	221	258	225	318	176	402	297

	T1	T2	T3	T4	T5	T6	T7	T8
Excel		48(21.7)	2(1.1)	10(4.4)				17(6)
Calculadora	13(4.8)	5(2.3)			15(4.9)	4(2.3)	17(4.6)	
No usa	256(95.2)	168(76)	184(98.9)	215(95.6)	292(95.1)	172(97.7)	354(95.4)	267(94)
Total	269	221	186	225	307	176	371	284

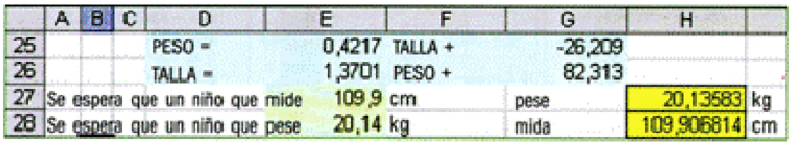


Figura 4.11.1. Ejemplo de predicción a través de la Hoja de Cálculo Excel ([T8], p. 407)

4.11.2.USO DE INTERNET

Encontramos sugerencias de uso de Internet en los textos [H2], [H3], [H4], [H5], [T2], [T3], [T4] y [T5]. En su mayoría conducen a páginas web con unidades didácticas sobre la correlación y regresión, aunque en otros casos se describe también el uso de algún applet, alguna reflexión didáctica sobre una investigación basada en el análisis de la correlación, o una prueba de autoevaluación. En los textos [H3], [H4], [T3] y [T4] encontramos enlaces a una unidad sobre correlación y regresión lineal "Descartes 2D Estadística y probabilidad", que incluye actividades y aplicaciones interactivas (descartes.cnice.mec.es/materiales_didacticos/Variables_estadisticas_bidimensionales_regresion_correlacion/Indice.htm).

También se señala ([H3] y [T3]) un enlace a una aplicación de autoevaluación: http://www.xtec.net/aulanet/ud/mates/estadistica/tr3/index.htm y a un applet para analizar la correlación lineal, así como la regresión lineal y cuadrática, en el que se permite visualizar los datos en un diagrama de dispersión, pero no interactuar con ellos.

En [H5] y [T5] encontramos además los siguientes enlaces:

- CICA (http://thales.cica.es/rd/Recursos/rd99/ed99-0151-01/capitulos/cap2.htm) que se incluye en un ejercicio que estudia la relación entre la latitud y la temperatura media anual de máximas de un conjunto de ciudades. El objeto es que el estudiante complete el estudio con otras variables que puedan influir en la temperatura.

- En http://www.fisterra.com/mbe/investiga/var_cuantitativas/var_cuantitativas.asp se describe una investigación sobre la relación entre la talla y el peso de 20 niños. Se presenta el cálculo del coeficiente de correlación lineal de Pearson, y la realización de inferencias sobre el mismo.

- Una página web destinada al estudio de la cinemática es: http://www.sc.ehu.es/sbweb/fisica/cinematica/regresion/regresion.htm, donde se incluye un applet interactivo que proporciona los valores de la pendiente y la ordenada en el origen de la recta de regresión por el método de mínimos cuadrados, el error que se comete, y el coeficiente de correlación lineal.

- http://www.iescarrus.com/edumat/taller/regresion/regresion_01.htm Enlace a un documento que trata sobre el origen de la teoría de la correlación y regresión.

4.11.3.CD CON MATERIAL TECNOLÓGICO

Algunos libros de texto se acompañan de un CD para complementar aspectos procedimentales en su formación, y servir de apoyo a su estudio. A continuación se clasifica el contenido de los mismos:

Documentos que describen el uso de recursos tecnológicos. Se suelen inlcuir documentos de lectura en los que se describen la forma de llevar a cabo los procecimientos asociados a la correlación y regresión con el uso de diferente tecnología, que puede ser la calculadora, la hoja Excel u otros programas. Sólo [H1] y [T1] describen el uso de la calculadora gráfica; los textos [H1], [H6], [T1] y [T6] tambien describen la hoja Excel. Además [H1] describe el uso de Derive. Y finalmente, en [H1] y [T1] el estudiante dispone de una hoja de cálculo, que se muestra en la Figura 4.11.1, con la que puede trabajar las tablas de doble entrada, la distribucion bidimensional, la covarianza, la correlación, y la regresión.

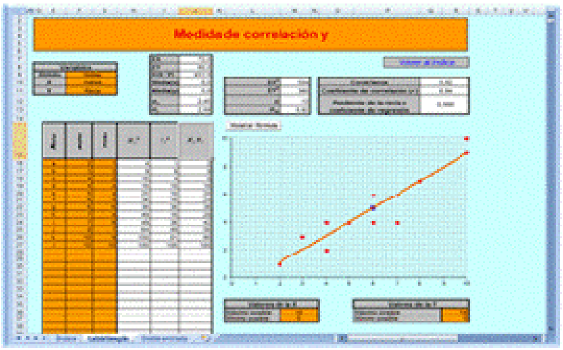


Figura 4.11.1. Hoja de cálculo para trabajar la correlación y regresión incluida en el CD de [H1] y [T1]

Complementos teóricos. En algunos CDs se encuentran enlaces a unidades didácticas de apoyo ([H5], [H8], [T5] y [T8]), documentos biográficos, o para la reflexión (sólo en [H6] y [T6]). Además, en [H6] y [T6] se incluyen documentos de lectura y algunas tareas en relación a la prueba de selectividad. Cabe mencionar en este sentido, que [H6] y [T6] incluyen en el CD el texto idéntico incluido en el libro, aunque utiliza la animación.

Actividades. Algunos textos incluyen actividades en el CD con diversa finalidad. En [H1] y [T1], además de los documentos ya citados, se incluye la resolución a la autoevaluación del tema. [H6] y [T6] proponen dos tipos de tareas: unas incluidas en el documento, que describe el uso de la Hoja de Cálculo Excel, y otras interactivas, como medio de evaluación de conocimientos; se diseñan para que el estudiante elija en un cuadro de diálogo la opción correcta; después el programa validará su solución. Se distribuyen en tres niveles:

1. Asociar un diagrama de dispersión a valores del coeficiente de correlación, y análisis completo de correlación y regresión, incluyendo el tratamiento unidimensional de las distribuciones marginales;
2. Covarianza, correlación, regresión y estimación;
3. Estimación, regresión y análisis unidimensional.

[H5] y [T5] incluyen tareas de matemáticas recreativas como un crucigrama, un juego de memoria, una sopa de letras. [H7] y [T7] incluyen tareas resueltas sobre construcción de una tabla de doble entrada, cálculo del coeficiente de correlación a partir de la tabla, determinación de la media de una variable, signo del coeficiente de correlación, y estimaciones a partir de las dos rectas de regresión. El CD incluye un vídeo donde un profesor las resuelve en una pizarra y las describe oralmente. [H8] y [T8] incluyen tres tareas interactivas. En la primera se debe elegir una recta de regresión bajo indicaciones dadas, en la segunda se trata la organización de datos, y en la tercera se debe asignar un valor del coeficiente de correlación a diferentes diagramas de dispersión que se presentan (Figura 4.11.2).

Enlaces a programas y descarga de los mismos. Algunos CDs incluyen enlaces a la descarga de determinados programas como Derive, Ecuv 2.3.33, Graph ([H5] y [T5]). En [H2] y [T2] se incluye el programa Wiris, con una rutina ya programada para tratar la estadística unidimensional y la regresión lineal.

Otros enlaces. En el CD anexo a los textos [H8] y [T8], se incluye una sección denominada “en la red”, desde la que se accede a diferentes páginas de Internet para leer sobre el origen de la estadística (en Wikipedia), talleres de estadística, tipos de encuestas (en la web de estadística para todos), y enlaces a la descarga de datos (en el Ministerio de Empleo y Seguridad Social).

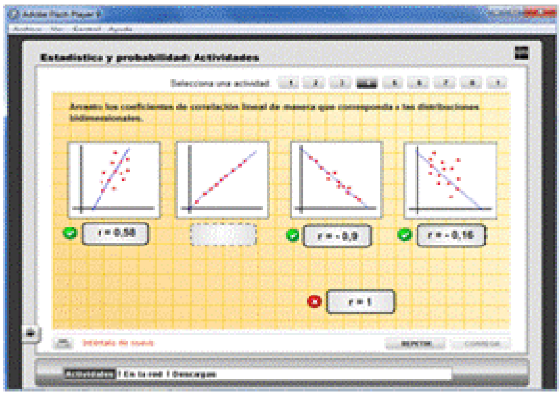


Figura 4.11.2. Tareas incluidas en el CD de [H8] y [T8]

Síntesis del uso de la tecnología

En la Tabla 4.11.1 se resume el análisis realizado del contenido del CD y el uso de Internet, donde se puede observar las pocas indicaciones al uso de applets dentro del texto, y en general, el poco uso de Internet. Cabe matizar que, en [H3] y [T3] o [H5] y [T5] se incluyen más enlaces a Internet que en [H4] y [T4], pues el único enlace que estos textos incluyen es a la unidad didáctica "Descartes", que además presenta applets.

Tabla 4.11.1. Recursos tecnológicos en los libros analizados

	H1	H2	H3	H4	H5	H6	H7	H8
Enlaces a Internet								
Unidades Didácticas y documentos teóricos		x	x	x	x			
Applets			x	x				
Contenido del CD								
Explicación del análisis bidimensional con Excel	x					x		
Explicación del análisis bidimensional con Derive	x							
Análisis bidimensional con calculadora gráfica	x							
Enlaces a páginas web sobre correlación y regresión					x			x
Actividades interactivas					x	x	x	x
Hoja de cálculo. Tabla de doble entrada	x							
Hoja de cálculo. Correlación y regresión	x							
Enlaces a programas de cálculo		x			x			

	T1	T2	T3	T4	T5	T6	T7	T8
Enlaces a Internet								
Unidades Didácticas y documentos teóricos		x	x	x	x			
Applets			x	x				
Contenido del CD								
Explicación del análisis bidimensional con Excel	x					x		
Explicación del análisis bidimensional con Derive								
Análisis bidimensional con calculadora gráfica	x							
Enlaces a páginas web sobre correlación y regresión					x			x
Actividades interactivas					x	x	x	x
Hoja de cálculo. Tabla de doble entrada	x							
Hoja de cálculo. Correlación y regresión	x							
Enlaces a programas de cálculo		x			x			

En cuanto a las tareas incluidas en los CDs, las de [H6] y [T6] presentan sobre todo intensidad alta, mientras que en [H8] y [T8] hay más variedad, incluyendo incluso dependencia funcional. Igualmente [H6] y [T6] resaltan la dependencia directa, mientras que en [H8] y [T8] hay tantas tareas con dependencia directa como inversa. Mayoritariamente se trata el ajuste lineal, aunque [H6] y [T6] incluyen una tarea bajo un ajuste polinómico de grado dos, y [H8] y [T8] una bajo un ajuste funcional (lineal).

Podemos observar gran variabilidad en las editoriales, porque en cuanto a la modalidad no hay grandes diferencias ([T1] no incluye la descripción del tratamiento de datos bidimensional con el programa Derive). Se prima más la inclusión de actividades de "evaluación" que sugerencias para que el estudiante trabaje el tratamiento de datos mediante la manipulación interactiva con applets. En ese aspecto, destacamos a los textos [H1] y [T1] pues incluyen Hojas de Cálculo programadas para el tratamiento de la correlación y regresión.

4.12.CONFLICTOS SEMIÓTICOS

Como se indicó en el Capítulo 2, el alumno activa su propia trama de funciones semióticas, que serán correctas si se conforman a la institución; en otro caso, se considera que existe un *conflicto semiótico*. Estos conflictos pueden explicar las dificultades y limitaciones de los aprendizajes (Godino, 2002; Godino, Font, Contreras y Wilhelmi, 2006).

Para finalizar nuestro análisis de los libros de texto, presentamos en esta sección las asignaciones imprecisas de significado que se han encontrado a lo largo de nuestro estudio. Se trata de funciones semióticas realizadas por los escritores de los textos, que no corresponden exactamente con las esperadas desde el punto de vista institucional. Por ello, podrían provocar en el alumno un conflicto semiótico si las generaliza, en forma excesivamente amplia, o realiza una interpretación inadecuada.

Conflictos semióticos potenciales ligados al lenguaje

Hemos encontrado identificación entre un concepto y una representación del mismo, utilización de representaciones no adecuadas de la distribución bivariante, y por tanto, un uso impreciso del lenguaje, como se describe a continuación.

Confusión de un concepto con su representación tabular y/o gráfica. El carácter no ostensivo de los objetos matemáticos, hace necesario el uso de representaciones de los mismos para su trabajo en el aula. Duval (1993) indica el interés didáctico de manejar diferentes representaciones, pero también recuerda que los objetos matemáticos nunca deben ser confundidos con su representación. En nuestro estudio, hemos encontrado algunos libros que inducen esta confusión, es decir, no separan claramente los planos expresión y contenido en una función semiótica.

Así, en [H1] y [T1] se define inicialmente la distribución bidimensional (Figura 4.7.2) mediante un ejemplo, que se presenta con una representación gráfica y otra tabular. El estudiante podría confundir el objeto (distribución) con su representación, pues no es hasta la siguiente página cuando se da una definición más precisa del concepto: *"El conjunto de pares de valores (x_1,y_1), (x_2,y_2)…(x_n,y_n) se llama una distribución bidimensional"* ([T1], p.333). Desde nuestro punto de vista, hubiese sido preferible completar el ejemplo con la definición apropiada de distribución bidimensional, en el mismo punto del texto, y diferenciar más claramente concepto y representación. Algo parecido ocurre con la definición de distribución marginal en estos mismos textos, donde el concepto se introduce como simple etiqueta de una tabla de frecuencias ([H1], p.237) y, a diferencia del caso anterior, no llega a definirse, a pesar de su relevancia.

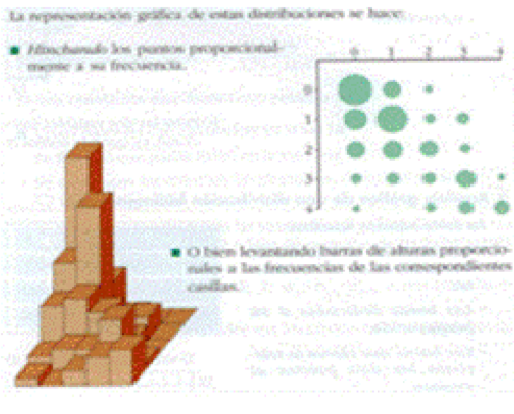


Figura 4.12.1. Representación gráfica de una distribución bidimensional ([H1], p. 233)

Utilizar una representación gráfica inadecuada de una distribución bivariante. Algunos textos intercambian, para el mismo tipo de datos, histogramas y diagramas de barras tridimensionales, independientemente de si las variables son continuas o categóricas. Así, [H1] y [T1] describen cómo representar gráficamente los datos de una tabla de doble entrada, pero no utilizan la expresión histograma o diagrama de barras, que permita identificar y diferenciar estas representaciones (Figura 4.12.1). Esto puede llevar al estudiante a no identificar correctamente cada gráfico. Así, en el ejemplo, aparentemente se utiliza un histograma tridimensional, pues no hay distancia entre las barras, pero la variable que se representa es cuantitativa discreta, y se debiera haber utilizado un diagrama de barras tridimensional. Esta situación se repite en [H7] y [T7].

Lenguaje simbólico inapropiado o impreciso. Hemos encontrado ejemplos en los que el lenguaje simbólico podría inducir conflicto en los estudiantes. Un primer caso lo encontramos en [H4] y [T4], al definir el concepto de frecuencia marginal, utilizando tablas en que aparecen estas frecuencias (Figura 4.12.2). En la tabla se podría generar un conflicto semiótico en la interpretación de la notación, pudiendo el estudiante entender las expresiones *(f_r)A_n* ó *(f_r)B_m* como producto de dos elementos.

X \ Y	y_1	y_2	y_3	...	y_m	$\sum_{j=1}^{m}(f_r)_{ij}$
x_1	$(f_r)_{11}$	$(f_r)_{12}$	$(f_r)_{13}$	...	$(f_r)_{1m}$	$(f_r)A_1$
x_2	$(f_r)_{21}$	$(f_r)_{22}$	$(f_r)_{23}$	...	$(f_r)_{2m}$	$(f_r)A_2$
x_3	$(f_r)_{31}$	$(f_r)_{32}$	$(f_r)_{33}$	...	$(f_r)_{3m}$	$(f_r)A_3$
...	...	...	...	...	...	...
x_n	$(f_r)_{n1}$	$(f_r)_{n2}$	$(f_r)_{n3}$	...	$(f_r)_{nm}$	$(f_r)A_n$
$\sum_{i=1}^{n}(f_r)_{ij}$	$(f_r)B_1$	$(f_r)B_2$	$(f_r)B_3$	...	$(f_r)B_m$	1

Figura 4.12.2. Tabla de doble entrada con la notación asociada a las frecuencias ([H4], pp. 218 y 219)

y_i \ x_i	0	1	2	3	4
0	24	6	1	0	0
1	11	19	2	3	0
2	7	8	6	2	0
3	2	3	3	7	1
4	1	0	2	4	5

COVARIANZA

Observamos en la tabla inicial los productos $x_i y_i$ y sus frecuencias:

$$\Sigma x_i y_i f_i = 1 \cdot 1 \cdot 19 + 1 \cdot 2 \cdot 2 + 1 \cdot 3 \cdot 3 + 2 \cdot 1 \cdot 8 + 2 \cdot 2 \cdot 6 + 2 \cdot 3 \cdot 2 + 3 \cdot 1 \cdot 3 + 3 \cdot 2 \cdot 3 + 3 \cdot 3 \cdot 7 + 3 \cdot 4 \cdot 1 + 4 \cdot 2 \cdot 2 + 4 \cdot 3 \cdot 4 + 4 \cdot 4 \cdot 5 = 330$$

$$\sigma_{xy} = \frac{\Sigma x_i y_i f_i}{\Sigma f_i} - \bar{x}\,\bar{y} = \frac{330}{117} - 1{,}16 \cdot 1{,}5 = 1{,}08$$

COEFICIENTE DE CORRELACIÓN

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1{,}08}{1{,}22 \cdot 1{,}3} = 0{,}68$$

Figura 4.12.3. Cálculo de la covarianza ([H1], p. 237)

Persona	P_1	P_2	...	P_n
Tiempo	t_1	t_2	...	t_n
Errores	e_1	e_2	...	e_n

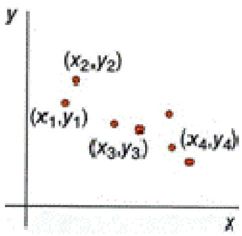


Figura 4.12.4. Proceso de traducción entre la representación gráfica y tabular ([H5], p. 250)

En [H1], [H5], [H6], [T1], [T5] y [T6] no se aclaran los símbolos utilizados al introducir la fórmula de la covarianza, como se muestra en la Figura 4.12.3. Es particularmente confuso el ejemplo que se presenta, que muestra los datos en una tabla de doble entrada, donde sería necesario utilizar dos índice *i*, *j* dado que los datos poseen frecuencia distinta a la unidad. Sin embargo, en la fórmula se utiliza únicamente el índice *i*. [H2], [H8] y [T2] previenen este conflicto implícitamente, cuando se indica al alumno la utilidad de transformar la tabla de doble entrada en una tabla simple o listado de datos para trabajar con ellos más fácilmente. Finalmente, en [H5] y [T5] el uso del lenguaje podría dificultar al estudiante para conectar la representación tabular y gráfica de una distribución bidimensional, pues se oscurece al presentar ambas representaciones bajo una notación diferente (Figura 4.12.4).

Conflictos semióticos potenciales ligados a la definición de conceptos

Igualmente advertimos imprecisiones, relacionadas con las definiciones de conceptos que describimos a continuación:

Variable estadística bidimensional. Los textos realizan una reducción del significado del concepto cuando limitan la definición a aquellas variables estadísticas unidimensionales que sean cuantitativas, como ocurre en [H8] y [T8]: "*Una variable estadística bidimensional (X, Y) es el resultado del estudio de dos características cuantitativas X e Y en los individuos de una población.*" ([H8], p.248).

Dependencia funcional. Los criterios introducidos por algunos textos para identificar la dependencia funcional a través del diagrama de dispersión pueden llevar a un conflicto. Así, el criterio "*Los puntos de la nube se sitúan sobre una curva cuya expresión matemática podríamos determinar*" ([H3], p.222) podría inducir la concepción algebraica de la asociación (en este caso de la dependencia funcional) descrita por Estepa (1994).

Dependencia y correlación. Algunos textos analizados identifican correlación con relación estadística/aleatoria entre las variables. Por ejemplo, en [H6] y [T6] se define la dependencia estadística del siguiente modo: "*La relación entre dos variables también puede ser estadística, y recibe entonces el nombre de correlación.*" ([H6], p. 180), definiendo más adelante la correlación como: "*La correlación es la relación que existe entre dos variables estadísticas unidimensionales que constituyen una variable estadística bidimensional*". En [H8] y [T8] se indica: "*Se llama correlación a la "relación o dependencia" que existe entre las dos variables que intervienen en una distribución bidimensional*" ([H8], p.253).

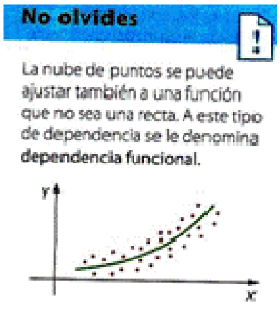


Figura 4.12.5. Definición de dependencia funcional ([H7], p. 247)

Como se ha señalado, el término correlación sólo se aplica a variables numéricas, por lo que un uso generalizado podría producir un conflicto semiótico en el estudio posterior de la asociación con variables cualitativas. Este conflicto se llega a extender hasta el punto en que se ofrecen definiciones incorrectas como en la Figura 4.12.5, donde el diagrama de dispersión muestra una dependencia claramente aleatoria, asignándole de modo incorrecto una definición de dependencia funcional. Por otro lado, [H2] y [T2] usan el término "*correlación funcional*" para referirse a la dependencia funcional, llegando a enunciar la siguiente propiedad ([H2], p. 249): "*Si r = -1 o r = 1, la correlación es perfecta o funcional*".

Independencia y correlación nula. Encontramos textos como [H1], [H2], [H8], [T1], [T2] y [T8], que tratan como equivalentes los términos *independencia* y *correlación nula* (*variables incorreladas*), cuando en realidad se debiera utilizar la noción de independencia. Al identificar conceptos no equivalentes se produce un conflicto semiótico. En los textos [H8] y [T8] se alude al término independencia, ya que ofrecen una definición del concepto en dos momentos (Figura 4.12.6).

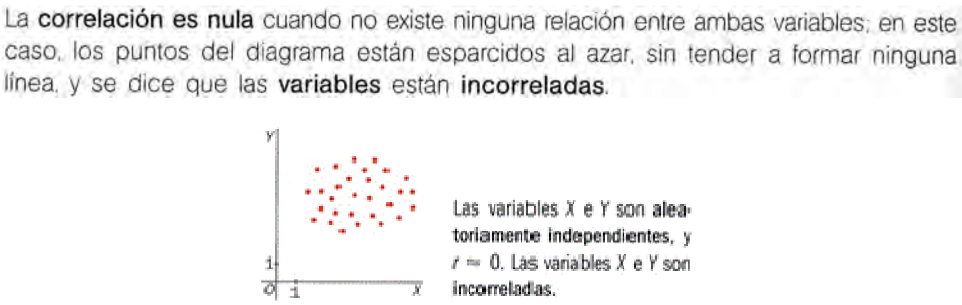

La **correlación es nula** cuando no existe ninguna relación entre ambas variables; en este caso, los puntos del diagrama están esparcidos al azar, sin tender a formar ninguna línea, y se dice que las **variables** están **incorreladas**.

Figura 4.12.6. Definición de independencia en varios momentos ([H8], pp. 252 y 253)

Suponemos que esta imprecisión proviene del uso generalizado de la noción de correlación, como sinónimo de dependencia. El uso generalizado del término incorrelación, en situaciones de independencia, conduce a algunos textos a incluir imprecisiones como también ocurre en [H5] y [T5], donde se señala: "*La correlación puede calificarse como fuerte cuando el grado de dependencia es alto; y como débil en caso contrario*" ([T5], p.358). En estos textos, no se define la independencia, ni tampoco variables incorreladas, a pesar de que implícitamente se hace uso de ellas.

Correlación y coeficiente de correlación lineal. Encontramos textos ([H1], [H2], [T1] y [T2]) donde no se diferencia correlación y coeficiente de correlación lineal, y cuando las variables no se encuentran relacionadas linealmente, se indica que las variables están incorreladas, en una situación claramente de independencia: "*la nube de puntos es amorfa y no sugiere ninguna recta: no hay correlación entre las variables (se dice que son incorreladas)*" ([H1], p.227). Estos libros pueden inducir confusión en el tratamiento de la correlación curvilínea, haciendo suponer que el coeficiente de correlación lineal pudiera informar de este otro tipo de dependencia, como por ejemplo en [H5] y [T5], donde se indica:

> Hasta ahora nos hemos limitado, y no es poco, a decir cuándo dos variables están correlacionadas. Primero, representábamos el diagrama de dispersión y decíamos que, cuando una recta se ajustaba bien a la nube de puntos, entonces la correlación era fuerte. Después, con el coeficiente r, dábamos una medida del sentido y fuerza de la correlación ([H5], p. 257).

Aunque el caso más destacado son [H2] y [T2], con la siguiente proposición:

> El coeficiente de correlación de Pearson indica la correlación que existe entre las dos variables; es decir, si los puntos están muy próximos o alejados del centro de gravedad. ([H2], p. 249)

Correlación y regresión. Encontramos varios textos en los que el concepto de correlación se confunde, en ocasiones, con el de regresión. Por ejemplo, [H4] define la regresión, pero no los modelos de regresión. Respecto a la correlación, se define como: "*la dependencia que existe entre las dos variables de una distribución bidimensional*", y al tratar sus características, se incluye como cualidad el tipo de función de ajuste del siguiente modo:

> Se pueden distinguir los siguientes casos:
> · La correlación es de tipo funcional si existe una función tal que la satisfacen todos los valores de la distribución.
> · La correlación es lineal o curvilínea cuando la nube de puntos se agrupa en torno a una línea recta o a una curva, respectivamente. ([H4], p.226)

Incluso textos en los que encontramos una definición clara del concepto de modelo de regresión, como [H5], el sentido dado a la correlación incluye al de regresión. Destacamos también [H7] y [T7], que no incluyen otra definición en el análisis de la regresión que las rectas de ajuste mínimo cuadráticas, y plantea la tarea que se muestra en la Figura 4.12.7.

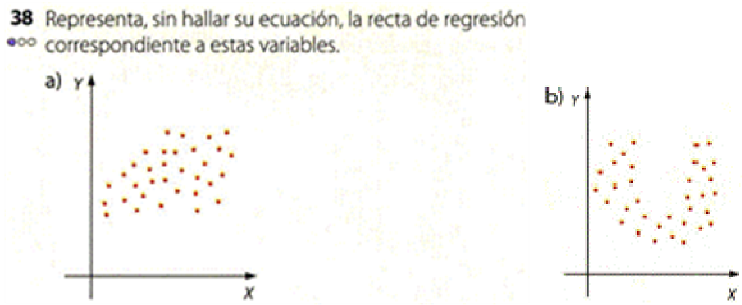


Figura 4.12.7. Tarea de regresión lineal ([H7], p. 260)

Conflictos semióticos potenciales ligados a las proposiciones

En ocasiones, se atribuyen propiedades que no tienen ciertos conceptos, o bien se realiza una generalización muy amplia de otras, lo que podría influir en la posterior adquisición, por parte de los estudiantes, de conflictos o concepciones erróneas sobre la correlación como la unidireccional o causal (Estepa, 1994; Estepa y Batanero, 1995).

Sentido de la correlación. Algunos textos asumen correlación entre las variables únicamente cuando el aumento de una de ellas implique el aumento de la otra, llegando a considerar independencia en cualquier otro caso. Esta descripción podría incidir en la concepción unidireccional (Estepa, 1994). Ninguno de los textos de la muestra advierten de la importancia de observar el decrecimiento conjunto de las variables en el estudio de correlación. En algunos textos, como [H2] y [T2], se advierte de la importancia que tiene la posición de los datos (cuatro cuadrantes en que el plano se divide mediante el trazado de las dos medias de las variables *X* e *Y*) para valorar el signo de la covarianza y la correlación. Aún así, las proposiciones que presenta invitan a confusión ([H2], p. 246):

> Según sea el signo de la covarianza, se interpreta:
> a) Covarianza positiva: al aumentar los valores de la variable X, aumentan los valores de la variable Y. La nube de puntos se orienta a la derecha y hacia arriba.
> b) Covarianza negativa: al aumentar los valores de la variable X, disminuyen los valores de la variable Y. La nube de puntos se orienta a la derecha y hacia abajo.

Correlación y causalidad. Algunos textos como [T5] (p. 363) indican, al estudiar la bondad de ajuste, que la varianza de la variable dependiente que no es explicada por la variable independiente será debida a otras causas, como muestra el siguiente texto: "*Por tanto, el coeficiente de determinación $r^2=0{,}88$. Esto significa en los niños de nuestro ejemplo, el 88% de la altura se explica por la edad. El resto hasta el 100% será debido a otras causas; altura de sus padres, dieta, …*". Este tipo de explicación podría inducir la concepción causal, que consiste en considerar asociación entre variables estadísticas solamente cuando exista causa y efecto entre ellas.

Interpretación de medidas estadísticas en un análisis unidimensional. En algunos ejercicios, como el que se muestra en la Figura 4.12.8, se interpretan incorrectamente algunos estadísticos univariantes. En este caso se confunde el coeficiente de variación, y se le considera como un coeficiente de concentración. Por otro lado, se indica que al ser el valor de la desviación típica grande (de hecho no es mucho mayor que la media), la muestra no es representativa. Pero la representatividad de la muestra no tiene nada que ver son su dispersión, sino con el hecho de que los valores centrales sean próximos a los de la población, y que tenga tamaño suficiente, así como dispersión y forma similar a la población**.**

Estas son las medidas estadísticas de un estudio sobre el número de roturas que sufrieron unos alfileres a los que se les sometió a una prueba.

$\bar{x} = 0{,}7$ $\quad Me = 0$ $\quad Mo = 0$

$\sigma = 0{,}96$ $\quad CV = 1{,}37$

Interpreta estas medidas estadísticas.

PRIMERO. Se compara $\bar{x}$ con la mediana y la moda.

La media, $\bar{x} = 0{,}7$, indica que el número medio de roturas es casi 1. Sin embargo, la mayoría de los alfileres no ha sufrido roturas ($Mo = 0$), y lo mismo indica la mediana ($Me = 0$), es decir, más de la mitad de los alfileres no ha sufrido ninguna rotura durante el estudio.

SEGUNDO. Se estudia el valor del coeficiente de variación.

$CV = 1{,}37$ es un valor muy grande, por lo que los datos no están demasiado concentrados.

TERCERO. Se estudia el valor de la desviación típica y se compara con el valor de la media.

El valor de la desviación típica es mayor que el de la media. Esto explica por qué mientras la mediana y la moda indican que el mayor número de alfileres no ha tenido roturas, la media de roturas ha sido casi de 1.

Al tener una desviación típica tan grande, la media no es muy representativa.

Figura 4.12.8. Tarea de interpretación de estadísticos en variable unidimensional ([T7], p.312)

Conflictos semióticos potenciales ligados a los procedimientos

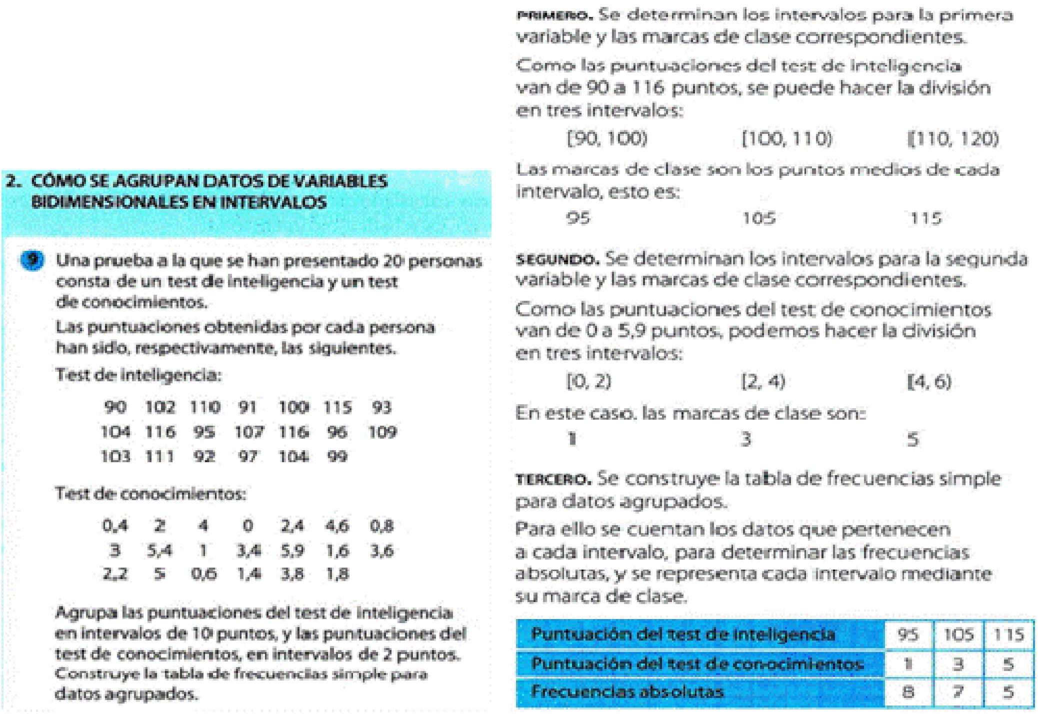
Se han encontrado las siguientes descripciones confusas o imprecisas:

Descripción confusa del cálculo de la covarianza o correlación. Cuando se introduce la covarianza, sería recomendable que la fórmula de cálculo venga acompañada de todas las expresiones equivalentes que se pretendan utilizar en el texto.

Así lo hace [H3], que utiliza la fórmula de la covarianza acompañada, tanto al margen como en el texto, de unas anotaciones para hacerla más comprensible. Por el contrario, en [H1], al rehusar del formalismo del sumatorio, se potencia un conflicto en el alumno en cuanto a la aplicación de la fórmula de la covarianza.

Explicación confusa de la construcción de tablas bidimensionales. En algunos textos, por ejemplo [H7] y [T7], se incluye una explicación completa de la construcción de tablas bidimensionales con datos agrupados (Figura 4.12.9). Sin embargo, observamos varios aspectos confusos. Por un lado, se presentan los datos inicialmente en forma de listado, y el alumno ha de construir, sin indicaciones, la variable estadística bidimensional. Por otro, se le deja clasificar los puntos obtenidos en intervalos bidimensionales. La tabla final es un listado. Al menos, como posibilidad, podría aparecer un estudiante con puntuación (95, 3), por ejemplo, posibilidad que no se contempla en la descripción. Otra complicación se produce al sustituir los intervalos por su marca de clase en la tabla final. Pensamos que esta descripción tiene un alto nivel de complejidad para el

Explicación confusa de representaciones gráficas de datos bidimensionales. Es difícil que un estudiante aprenda a construir adecuadamente un diagrama de barras, con las descripciones que aparecen en textos como [H2] y [T2], donde tan sólo se muestra un ejemplo de éste (Figura 4.12.10); además, la construcción no viene acompañada de alguna aclaración en cuanto a la proporcionalidad de la altura y la frecuencia de los datos. Además, no se indica que el gráfico no es adecuado para representar una variable estadística bidimensional continua, siendo más adecuado un histograma; aunque para este caso, el tratamiento que encontramos en algunos textos es similar (Figura 4.12.1).



2. CÓMO SE AGRUPAN DATOS DE VARIABLES BIDIMENSIONALES EN INTERVALOS

9 Una prueba a la que se han presentado 20 personas consta de un test de inteligencia y un test de conocimientos.

Las puntuaciones obtenidas por cada persona han sido, respectivamente, las siguientes.

Test de inteligencia:

90	102	110	91	100	115	93
104	116	95	107	116	96	109
103	111	92	97	104	99	

Test de conocimientos:

0,4	2	4	0	2,4	4,6	0,8
3	5,4	1	3,4	5,9	1,6	3,6
2,2	5	0,6	1,4	3,8	1,8	

Agrupa las puntuaciones del test de inteligencia en intervalos de 10 puntos, y las puntuaciones del test de conocimientos, en intervalos de 2 puntos. Construye la tabla de frecuencias simple para datos agrupados.

PRIMERO. Se determinan los intervalos para la primera variable y las marcas de clase correspondientes.

Como las puntuaciones del test de inteligencia van de 90 a 116 puntos, se puede hacer la división en tres intervalos:

[90, 100) [100, 110) [110, 120)

Las marcas de clase son los puntos medios de cada intervalo, esto es:

95 105 115

SEGUNDO. Se determinan los intervalos para la segunda variable y las marcas de clase correspondientes.

Como las puntuaciones del test de conocimientos van de 0 a 5,9 puntos, podemos hacer la división en tres intervalos:

[0, 2) [2, 4) [4, 6)

En este caso, las marcas de clase son:

1 3 5

TERCERO. Se construye la tabla de frecuencias simple para datos agrupados.

Para ello se cuentan los datos que pertenecen a cada intervalo, para determinar las frecuencias absolutas, y se representa cada intervalo mediante su marca de clase.

Puntuación del test de inteligencia	95	105	115
Puntuación del test de conocimientos	1	3	5
Frecuencias absolutas	8	7	5

Figura 4.12.9. Ejercicio resuelto de representación tabular de datos bidimensionales ([H7], p. 254)

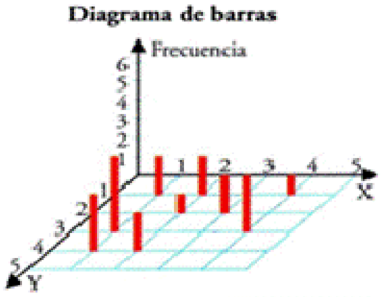


Figura 4.12.10. Representación gráfica de datos bidimensionales ([T2], p. 317)

Procedimientos confusos en el análisis de regresión. En algunos casos no se diferencia, en el cálculo de las rectas de regresión, la variable dependiente e independiente (con el posible riesgo ya comentado para la estimación).

Destacamos un ejercicio en el texto [H1], donde se estima un valor de y mediante la recta de regresión de Y sobre X, y un valor de x con la inversa de dicha recta. Es cierto que se lleva a cabo bajo una correlación casi perfecta (r = 0'9994), y que se advierte que las estimaciones de la variable independiente se suelen realizar sobre "*la otra recta de regresión*" ([H1], p. 231), pero pensamos que el alumno podría generalizar en forma incorrecta este procedimiento. En el siguiente caso, se muestran técnicas de linealización mediante el cambio de variable, pero no se explica correctamente su aplicabilidad:

> En algunas ocasiones se tiene un conjunto de datos que no se pueden ajustar con una recta, pero entre los que sí existe una dependencia funcional. Esta dependencia puede transformarse en lineal utilizando diferentes técnicas algebraicas. De esta manera se puede aplicar la regresión lineal para predecir el comportamiento de una variable en función de la otra. Entre estas técnicas de linealización está el cambio de variable. ([H8], p.257)

Conflictos semióticos potenciales ligados a los argumentos

Finalmente se ha observado que algunos textos presentan argumentos que podrían inducir conflictos semióticos en los estudiantes. Un ejemplo es: *"Si los puntos siguen, aunque sea aproximadamente, una configuración rectilínea, diremos que hay dependencia entre ellos"* ([H7], p.247), que añade descripciones y ejemplos en que se aprecia el signo e intensidad de la dependencia lineal. Suponemos que haya sido un error de imprenta ya que en la modalidad científico-tecnológica ([T7]) este argumento aparece pero utilizando el término *dependencia lineal* en lugar de dependencia. Lo mismo ocurre en [H8] y [T8] cuando se introducen otros modelos de regresión: "*En algunas ocasiones se tiene un conjunto de datos que no se pueden ajustar con una recta, pero entre los que sí existe una dependencia funcional*" ([H8], p.257).

Síntesis de conflictos semióticos observados

Como resumen de los conflictos semióticos potenciales observados, presentamos la Tabla 4.12.1, donde no se aprecian grandes diferencias en los textos de cada modalidad, salvo [T5] y [T7], donde se trata la correlación y causalidad o la interpretación de estadísticos, respectivamente, con presencia de conflictos potenciales.

En estos casos, la editorial no incluye la enseñanza de estas proposiciones en el Bachillerato de *Humanidades y Ciencias Sociales*, en el primer caso, porque el coeficiente de determinación no se define en [H5], y en el segundo, pues no se incluye el análisis unidimensional de modo tan extenso.

Tabla 4.12.1. Conflictos semióticos en los textos analizados

	H1	H2	H3	H4	H5	H6	H7	H8
Ligados al lenguaje								
Confusión objeto-representación	x							
Representación inadecuada	x						x	
Lenguaje impreciso	x			x	x	x		
Ligados a la definición								
Variable bidimensional								x
Dependencia funcional			x		x			
Dependencia y correlación		x				x	x	x
Independencia y correlación	x	x			x			x
Correlación y coeficiente de correlación	x	x			x			
Correlación y regresión			x	x	x	x	x	
Ligados a la proposiciones								
Sentido de la correlación		x						
Correlación y causalidad								
Interpretación de estadísticos								
Ligados a los procedimientos								
Fórmulas de cálculo	x							
Construcción de tablas							x	
Construcción de gráficas		x						
Procedimientos confusos en regresión	x							x
Ligados a los argumentos							x	x

	T1	T2	T3	T4	T5	T6	T7	T8
Ligados al lenguaje								
Confusión objeto-representación	x							
Representación inadecuada	x						x	
Lenguaje impreciso	x			x	x	x		
Ligados a la definición								
Variable bidimensional								x
Dependencia funcional			x		x			
Dependencia y correlación		x				x	x	x
Independencia y correlación	x	x			x			x
Correlación y coeficiente de correlación	x	x			x			
Correlación y regresión			x	x	x	x	x	
Ligados a la proposiciones								
Sentido de la correlación		x						
Correlación y causalidad					x			
Interpretación de estadísticos							x	
Ligados a los procedimientos								
Fórmulas de cálculo	x							
Construcción de tablas							x	
Construcción de gráficas		x						
Procedimientos confusos en regresión	x							x
Ligados a los argumentos							x	x

Son también más los conflictos potenciales debido a definiciones imprecisas de los conceptos; al tratar de ponerlos al alcance de los estudiantes de este nivel, la transposición didáctica ocasiona cambio del significado de dichos conceptos. Respecto a los procedimientos y argumentos no encontramos tantas diferencias entre las dos modalidades de Bachillerato, pues en general, cuando se produce un conflicto potencial en una editorial se suele reproducir en las dos modalidades. El lenguaje matemático se presenta también con poco rigor.

4.13. SIGNIFICADO DE REFERENCIA

Finalizado el análisis de los libros de texto, podemos fijar el significado de referencia del contenido matemático para nuestro trabajo. Dicho significado se presenta resumido en las Tablas 4.13.1, 4.13.2 y 4.13.3, organizadas según los objetos matemáticos considerados en nuestro marco teórico. Incluye la colección de objetos matemáticos identificadas en los textos analizados para cada categoría,.

Respecto a los campos de problemas, partiendo de los identificados en el estudio histórico, el análisis de los textos nos ha permitido subdividirlos. Aparecen asimismo los lenguajes verbal, gráfico, numérico, tabular y simbólico, que se detallaron a lo largo del capítulo, cada uno de ellos de una gran riqueza y complejidad; por ejemplo combinando subíndices o superíndices, sumatorias simples o dobles, variables y parámetros en sus expresiones algebraicas.

Tabla 4.13.1. Significado de referencia de la correlación y regresión (Procedimientos y argumentos)

Procedimientos	PC0. Organización de datos bidimensionales
	PC01.Construir una tabla a partir de los datos
	PC02.Determinar la distribución marginal o condicionada
	PC03.Construir una gráfica
	PC04.Leer o interpretar frecuencias en representaciones
	PC1. Análisis de la dependencia entre variables
	PC11.Interpretación del diagrama de dispersión
	PC12.Cálculo de la covarianza
	PC13.Interpretación de la covarianza
	PC14.Cálculo del coeficiente de correlación lineal
	PC15.Interpretación del coeficiente de correlación lineal
	PC2. Análisis de regresión
	PC21.Ajuste de la recta de regresión
	PC22.Ajuste de otros modelos de regresión
	PC23.Cálculo del coeficiente de determinación
	PC24.Valorar la bondad del ajuste del modelo de regresión
	PC25.Determinar una predicción
	PC26.Valorar la bondad de la predicción
Argumentos	A1.Ejemplos/contraejemplos
	A2.Gráficos auxiliares
	A3.Verbales deductivos
	A4.Reducción al absurdo
	A5.Algebraicos deductivos

Tabla 4.13.2. Significado de referencia de la correlación y regresión (Situaciones, lenguaje y conceptos)

Situaciones-problema	P0. Organización/Representación de datos bidimensionales
	P1. Análisis de la dependencia entre dos variables estadísticas
	P11. Análisis de las variables que conforman la variable estadística bidimensional.
	P12. Análisis del tipo de dependencia.
	P13. Medida de la intensidad de la relación entre las variables.
	P14. Determinar la dirección de la relación entre variables.
	P2. Análisis de regresión. Predicción de una variable en función de otra.
	P21. Analizar el ajuste lineal entre variables.
	P22. Realizar estimaciones mediante el modelo de regresión.
Lenguaje	L1.Verbal
	L2.Simbólico
	L3.Tabular
	L4.Gráfico
	L5.Numérico
Conceptos	C1.Variable estadística y distribución bidimensional
	C11.Variable estadística bidimensional
	C12.Frecuencia doble
	C13.Distribución bidimensional
	C2. Distribución marginal y condicional
	C21.Frecuencia marginal
	C22.Distribución marginal
	C23.Frecuencia condicional
	C24.Distribución condicional
	C3.Dependencia funcional/estadística/independencia
	C31.Dependencia funcional
	C32.Dependencia estadística o aleatoria
	C33.Independencia
	C4.Covarianza y/o correlación
	C41.Correlación
	C42.Covarianza
	C43.Coeficiente de correlación lineal de Pearson
	C5.Regresión
	C51.Variable dependiente e independiente
	C52.Regresión
	C53.Modelos no lineales de regresión
	C54.Rectas de regresión
	C55.Coeficientes de regresión
	C56. Bondad de ajuste. Coeficiente de determinación

El número de conceptos diferentes introducido en el tema no es muy grande; aparentemente entonces la complejidad del estudio de la correlación y regresión es moderada. Esta primera impresión desaparece, por un lado, al tener en cuenta que algunos conceptos son muy próximos en su significado, como ocurre en las frecuencias o distribuciones condicionales, marginales y conjuntas, y el alumno los puede confundir. Además, algunos de estos conceptos están definidos mediante una condición (formalmente sería una probabilidad condicional, concepto difícil para los estudiantes).

Tabla 4.13.3. Significado de referencia de la correlación y regresión (Proposiciones)

Proposiciones	PP1. Distribución bivariante, marginal y condicional
	PP11.Datos bidimensionales.
	PP11a. Los datos se refieren a cada una de las variables
	PP11b. Organización de datos en tablas de frecuencias
	PP11c. Organización de datos en gráficos
	PP11d. Distribuciones de datos
	PP12 Relaciones entre tipos de frecuencias
	PP12a. La suma de frecuencias absolutas relativas dobles es 1
	PP12b. Frecuencia doble - frecuencia marginal
	PP13.Centro de gravedad
	PP2. Dependencia entre variables
	PP21.Dependencia funcional
	PP22.Dependencia estadística o aleatoria
	PP23.Independencia
	PP24.Covarianza
	PP25.Correlación
	PP26.Coeficiente de correlación lineal
	PP26a. Mide la correlación lineal entre las variables
	PP26b. Interpretación del coeficiente de correlación lineal
	PP26c. Adimensionalidad
	PP3. Análisis de regresión
	PP31.Modelos de regresión
	PP32.Ajuste de la recta de regresión
	PP32a. Propiedad de mínimos cuadrados
	PP32b. Dos rectas de regresión diferentes
	PP33.Coeficiente de determinación
	PP34.Predicción
	PP34a. La recta de regresión permite realizar estimaciones
	PP34b. Una predicción es una aproximación al valor real de la variable que se estima
	PP34c. Uso de la recta específica para realizar predicciones
	PP4. Relaciones entre conceptos
	PP41.Distribución - regresión
	PP41a. Recta de regresión - centro de gravedad
	PP41b. Predicción - centro de gravedad
	PP41c. Tamaño de la muestra - recta de regresión - predicción
	PP42.Dependencia - covarianza/correlación
	PP42a. Tipos de dependencia - correlación
	PP42b. Tipos de dependencia - coeficiente de correlación
	PP43.Dependencia - regresión
	PP44.Covarianza - correlación - coeficiente de correlación
	PP44a. Signo de la covarianza - sentido de la correlación
	PP44b. Signo de la covarianza - signo coef. de correlación
	PP45.Covarianza/correlación - regresión
	PP45a. Covarianza - recta de regresión
	PP45b. Correlación - recta de regresión
	PP45c. Producto c. regresión - coef.de correlación lineal
	PP45d. Coeficiente de correlación - coeficientes de regresión
	PP45e. Coeficiente de correlación - predicción
	PP45f. Posición rectas de regresión - coef. de correlación

Así ocurre en la distribución condicional o incluso en la propia definición de recta de regresión, en que el alumno ha de diferenciar el condicional y condicionado o la variable dependiente e independiente; dificultad que aumenta al ver que algunos libros no explicitan claramente la necesidad de hacer esta distinción. Recordamos que Falk (2006) nos advierte de que muchos estudiantes confunden condicional y condicionado en una probabilidad condicional. Esta confusión puede extenderse en la interpretación de los conceptos descritos mediante una condición en el estudio de la correlación y regresión. Por otro lado, la complejidad conceptual crece si se tiene en cuenta el número distinto de propiedades y relaciones entre conceptos que han aparecido a lo largo de nuestro análisis. Cada una de ellas es importante para la resolución de los problemas asociados y cada una de ellas podría dar lugar a conflictos semióticos por parte de los estudiantes.

El número de procedimientos que el alumno debe estudiar en el tema es, aparentemente también pequeño (Tabla 4.11.3). Sin embargo, hacemos notar que algunos de estos procedimientos son en realidad familias de procedimientos que el estudiante ha de adaptar en cada problema particular. Así ocurre en la construcción de una tabla de frecuencias, en que a veces es necesario agrupar una de las variables y no hay una regla fija de agrupación, sino el alumno ha de decidir con su propio criterio.

Otro ejemplo es la construcción de un gráfico; elección del tipo de gráfico y escalas; asimismo ocurre en la interpretación del diagrama de dispersión, elección de un modelo de regresión y determinación de la bondad de ajuste.

Finalmente, encontramos diferentes tipos de argumento, no sólo algebraicos o verbales deductivos; la mayoría son argumentos basados en ejemplos y contraejemplos, y se apoyan en gráficos auxiliares. Este significado de referencia será utilizado en la evaluación y desarrollo del conocimiento de futuros profesores, descrito en los Capítulos 5 (Estudio 2) y 6 (Estudio 3). En el primero de dichos estudios, se diseña una situación didáctica, que recoge una muestra representativa del significado de referencia identificado en el presente estudio (Estudio 1), que puede ser utilizada con estudiantes de Bachillerato, y permitirá evaluar y desarrollar el conocimiento matemático de los futuros profesores. En el Estudio 3, se diseña una situación didáctica específica, que permite evaluar y desarrollar el conocimiento didáctico de los mismos.

4.14.CONCLUSIONES SOBRE EL ESTUDIO DE LOS LIBROS DE TEXTO

En este capítulo se ha descrito el Estudio 1, cuyo objetivo fue realizar un análisis detallado de la presentación de la correlación y regresión en los libros de texto de Bachillerato, con la finalidad de caracterizar el significado institucional de referencia en las dos modalidades en que se contempla su estudio.

Para cumplir el citado objetivo, se ha llevado a cabo un estudio detallado de la presentación de la correlación y la regresión en dieciséis libros de texto, publicados el año siguiente a la publicación del *Real Decreto de Enseñanzas Mínimas* (MEC, 2007b), ocho de ellos dirigidos a la modalidad de *Humanidades y Ciencias Sociales* y otros ocho, de las mismas editoriales, dirigidos al *Bachillerato en Ciencias y Tecnología*.

Nuestro análisis parte de las premisas de que un objeto matemático adquiere significado en el sujeto por medio de las prácticas significativas (operativas y discursivas) que realiza de modo activo, y de que estas prácticas podrían estar relacionadas con las propuestas en los libros de texto. El marco teórico EOS nos ha permitido describir la presentación de estas nociones en los libros de texto, analizados de manera precisa, utilizando para ello la clasificación de elementos básicos de significado. Asimismo, se realiza un estudio del uso de la tecnología y de los conflictos semióticos potenciales encontrados.

Se puede valorar, en los libros analizados, una alta idoneidad epistémica ya que se incluyen los principales objetos matemáticos ligados a la correlación y regresión; por ejemplo, todos los campos de problemas identificados en el Capítulo 1, los conceptos fundamentales, y una amplia lista de propiedades, procedimientos y argumentos. A continuación exponemos nuestras principales conclusiones, organizadas respecto a los objetivos específicos del Estudio 1 y a las hipótesis planteadas para el mismo.

Conclusiones respecto a los objetivos

O1.1. Caracterizar los campos de problemas sobre correlación y regresión presentados sobre el tema en los textos de Bachillerato.

Encontramos en nuestro análisis los dos principales campos de problemas que dieron origen a la correlación y regresión: a) Estudio de la dependencia entre dos variables; y b) determinación de una función de ajuste (Estepa, Gea, Cañadas y Contreras, 2012), al que se añade otro previo consistente en la organización de datos bivariantes. Sobre estos campos, que hemos subdividido en varios, se han analizado el tipo de actividad propuesta, contexto y tipo de dependencia. Es evidente el protagonismo del análisis de la dependencia entre las variables (P1), que no suele venir precedido de un análisis previo descriptivo bidimensional (P0). También son escasas las tareas que impliquen el manejo de la representación tabular (frecuencias absolutas/relativas, distribución marginal o condicional) que ayuden a comprender la naturaleza bidimensional de los datos.

Asimismo hay un sesgo hacia la presentación de correlaciones intensas y de sentido directo, al igual que lo observado por Sánchez Cobo (1999), que podrían inducir una concepción unidireccional de la correlación en los estudiantes (Estepa, 1994). Son pocos los textos que plantean dependencias de tipo no lineal, lo cual dificulta que los estudiantes se planteen si este modelo es el mejor posible. Destacamos igualmente la necesidad de un uso más equilibrado de situaciones en cuanto al sentido, intensidad y tipo de dependencia ya que es importante presentar al estudiante variedad de situaciones de correlación y regresión, para que se problematice el modelo de mejor ajuste a los datos, cuestionando si el modelo lineal sería el único o el mejor posible (Batanero, Díaz y Gea, 2011).

No se encuentran problemas que utilicen datos reales (tomados de Internet o de otras fuentes), a pesar de las recomendaciones en este sentido de Franklin et al. (2007), ni se proponen proyectos completos en que el alumno parta de una pregunta y recorra todos los pasos del proceso de razonamiento estadístico. Además, conviene proponer tareas con datos reales, a partir de bases de datos y anuarios, que permitirán al estudiante crear una concepción más ajustada a lo que se refiere un estudio estadístico de fenómenos sociales y/o económicos, tal y como indican las concreciones curriculares, que además mostrarían mejor su aplicabilidad (MEC, 2007b).

En el caso de la intensidad de la dependencia, la mayoría de situaciones propuestas que encontramos en los textos analizados poseen una intensidad alta, lo que podría ser adecuado en el Bachillerato de *Ciencia y Tecnología*, pero no en el de *Humanidades y Ciencias Sociales* pues en las aplicaciones de Ciencias Sociales, las correlaciones suelen ser moderadas. Y por último, y no menos importante, el notorio predominio de la dependencia directa frente a la inversa o independencia en la mayoría de los textos analizados, que puede incidir posteriormente en la presencia, en algunos estudiantes, de la concepción unidireccional de la asociación descrita por Estepa (1994).

O1.2. Caracterizar la diversidad de lenguaje matemático utilizado en los textos para presentar la correlación y regresión, y los procesos de traducción entre los mismos.

Se identificaron y clasificaron una variedad de expresiones verbales y simbólicas específicas de este tema, que evidencian su complejidad semiótica, pues cada una de estas expresiones remite a conceptos, propiedades o procedimientos. Asimismo, hay una gran variedad de representaciones tabulares y gráficas; aunque algunas de ellas están utilizadas de forma poco precisa, y no diferencian correctamente datos discretos y continuos. Los ejemplos mostrados a lo largo de esta sección proporcionan modelos al profesor de cómo utilizar los diferentes tipos de representación en el aula, y cómo complementarlos.

Se echa en falta, en ocasiones, la explicación del procedimiento de construcción de la tabla o gráfico, que se considera elemental o transparente, punto en el que no coincidimos. Por otro lado, la presentación de los temas de estadística bidimensional en los libros de texto, en que tratan la correlación y regresión, ofrecen una visión sesgada de sus diferentes representaciones, tratando mayoritariamente el registro gráfico. Seguiría la representación simbólica y verbal; esta última con usos a veces inapropiados de metáforas, donde el lenguaje cotidiano se usa en forma imprecisa para describir conceptos o propiedades matemáticas. Respecto al cambio de representación (procesos de traducción entre representaciones), generalmente se pide al estudiante pasar de un listado de datos a un gráfico, haciendo pocas actividades de traducción de otros tipos de representaciones consideradas por Sánchez Cobo (1999) o Lavalle, Micheli y Rubio (2006).

Más aún, el lenguaje en algunos textos podría inducir conflictos semióticos, como confundir un objeto con su representación gráfica, confundir gráficos entre sí, imprecisión del lenguaje simbólico, o generalización abusiva de conceptos. No es menos importante destacar el uso sesgado de la tabla de doble entrada en la mayoría de los textos analizados, a favor del uso casi generalizado del listado de datos, cuya complejidad semiótica, según Arteaga (2011), es insuficiente para visualizar las tendencias en los datos.

O1.3. Caracterizar los conceptos y propiedades relacionados con la correlación y regresión en los textos.

Se han analizado y clasificado conceptos y propiedades respecto a cada uno de los tres campos principales de problemas que hemos citado anteriormente, siendo escaso el tratamiento de los conceptos asociados a la organización de datos bidimensionales.

Para cada concepto se estudia la forma de introducción, lugar y uso a lo largo del tema. Los conceptos se suelen presentar de forma práctica, tanto a partir de su fórmula, como indicando su utilidad, que se trataría de una definición instrumento-relacional según Sánchez Cobo (1999). En su estudio aparecía sólo en una minoría de los textos, por lo que pensamos que podría haberse dado una evolución en el tipo de definiciones presentadas en los últimos años. Encontramos inconsistencia, en el sentido que a veces los conceptos se usan sin definición previa.

Al analizar las definiciones que se incluyen en los textos analizados, en primer lugar, se ha notado la escasa presencia de definiciones relativas a la regresión, exceptuada la de la recta de regresión, en comparación con las que se incluyen referidas a la dependencia entre las variables. Es también escaso el tratamiento de las definiciones asociadas a la distribución bidimensional, como son la distribución marginal y condicional. Las definiciones más frecuentes son las de la covarianza y correlación, el tipo de dependencia, variable estadística y distribución bidimensional y regresión lineal en ese orden.

Los conceptos incluidos en los textos se suelen definir acompañados de ejemplos, aunque no siempre presentan los ejemplos antes de la definición, como recomienda Skemp (1993). Sería también admisible la presentación formal seguida por ejemplos, que aparece en otros textos, aunque es preferible lo contrario. En los casos en que hay una definición explícita de los conceptos, observamos un predominio de la presentación operacional (Sfard, 1991) de los conceptos relacionados con la distribución marginal y condicional, que es pocas veces seguida por la definición estructural, con la que el estudiante podría adquirir un significado más completo del concepto. Por el contrario, en los conceptos relacionados con la variable estadística y distribución bidimensional, o bien se presentan los dos tipos de definición formal, o se da predominio a la definición estructural (Sfard, 1991).

Del análisis realizado sobre las propiedades incluidas en los textos analizados, podemos decir que, en general, los textos cubren varias propiedades importantes relativas a la correlación y regresión, lo que también muestra la gran complejidad semiótica del tema, por la riqueza conceptual y de propiedades de la configuración epistémica asociada.

Es más pobre la inclusión de propiedades referidas al tipo de dependencia entre las variables, seguido del tratamiento tabular y gráfico de los datos. En este sentido, se incluyen muy pocas propiedades referidas a las relaciones que se establecen entre las frecuencias de los datos, por lo que cabe esperar una preparación deficiente de estos alumnos para el tratamiento posterior de la probabilidad, donde, principalmente en la modalidad de Bachillerato de *Humanidades y Ciencias Sociales* se trabajará con experimentos compuestos, y por tanto con distribuciones marginales y condicionales.

Podemos apreciar también un tratamiento menor de la covarianza, seguido del asociado a las rectas de regresión, donde la mayoría de los textos presentan los cálculos para determinar cada una de las rectas de regresión, pero no las propiedades que relacionan la regresión con la correlación. Además, la relación de la recta de regresión con el coeficiente de correlación lineal se suele establecer en torno a la predicción, y no tanto para interpretar la conexión entre correlación y regresión lineal, propiamente dicha.

O1.4. Caracterizar los procedimientos empleados en el estudio de la correlación y regresión.

Encontramos un gran número de procedimientos, asociados tanto a la representación de los datos en tablas y gráficas, al cálculo de estadísticos, como a la covarianza, correlación, centro de gravedad, rectas de regresión y a la interpretación de los resultados de todos los anteriores. Todo esto señala, de nuevo, la complejidad del tema; sin embargo, esta complejidad trata de disminuirse en los textos, pues el análisis de los procedimientos realizado muestra que, en su mayoría, se encuentran descritos a través de ejercicios resueltos o ejemplos, más que de un modo formal. De algún modo, esta orientación se encuentra implícita en las concreciones curriculares, por ejemplo, en cuanto a la modalidad científico-tecnológica se señala que:

> No se trata de que los estudiantes posean muchas herramientas matemáticas, sino las estrictamente necesarias y que las manejen con destreza y oportunidad, facilitándoles las nuevas fórmulas e identidades para su elección y uso. (MEC, 2007b, 45448).
> Lo importante es que el estudiante encuentre en algunos ejemplos la necesidad de la existencia de este lenguaje para dotar a las definiciones y demostraciones matemáticas de universalidad, independizándolas del lenguaje natural. (MEC, 2007b, 45449)

Son muchas las actividades sobre interpretación de la correlación y cálculo de diversos coeficientes o resúmenes estadísticos. Pero no tantas las relativas a la aplicación o interpretación del coeficiente de determinación. Por otro lado, y debido al escaso uso de la tecnología (hecho incomprensible hoy día) la aplicación no se realiza con datos reales, perdiéndose la oportunidad de que el estudiante llegue a valorar la utilidad de la estadística.

O1.5. Determinar los tipos de argumentación empleados para justificar propiedades, procedimientos o soluciones a los problemas.

En general, los textos poseen un nivel de argumentación aceptable, ya que incluyen argumentaciones gráficas y mediante ejemplos o contraejemplos (que a veces coinciden, pues una misma demostración a través de un ejemplo es de tipo gráfico), el uso de razonamientos verbales deductivos, y aunque menos frecuentes, los algebraicos deductivos, principalmente en la modalidad de *Ciencias y Tecnología*.

Para justificar propiedades o resultados, hay una preferencia por argumentos informales, basados en el uso de ejemplos-contraejemplos y leguaje gráfico, igual que en otras investigaciones descritas (Capítulo 2), lo que nos parece adecuado, dado el nivel de enseñanza.

No se encuentran apenas demostraciones deductivas verbales, mucho menos algebraicas, con la excepción de la demostración de la equivalencia de las fórmulas de cálculo de la covarianza y otros estadísticos. Son más escasos los argumentos simbólicos o por reducción al absurdo, o las propuestas para que sea el mismo alumno el que construya una argumentación, que sería particularmente interesante para los alumnos del Bachillerato de *Ciencia y Tecnología*. En todo caso, coincidimos con Ortiz (1999) en la necesidad de dar más atención a la actividad de argumentar, y dar oportunidad a los alumnos de realizar actividades que impliquen la producción de sus propios argumentos matemáticos.

O1.6. Analizar el uso que se hace de la tecnología.

Se ha hecho un estudio detallado de las referencias en los textos a tecnología, bien proponiendo problemas o procedimientos a realizar con calculadora gráfica u hoja de cálculo, bien incluyendo actividades para realizar con applets u otros recursos en Internet. Por otro lado, la mayoría de los textos se acompañan de un CD en el que se explica el uso de programas y recursos multimedia, que facilitan el aprendizaje de la correlación y regresión.

A pesar de ello, el tratamiento de la tecnología que se realiza en algunos textos sigue siendo pobre. Esta situación no es exclusiva en España, pues en la investigación de Lavalle, Michelli y Rubio (2006) ningún libro de texto proponía actividades con uso de tecnología, mientras que en la nuestra encontramos el uso de la calculadora, la hoja de cálculo, y la mayoría de los textos se acompañan de un CD en el que se explica el uso de programas y recursos multimedia que facilitan el aprendizaje de la correlación y regresión.

Pensamos que es importante aumentar el uso de esta tecnología, pues ayudaría al estudiante a disponer, en segundos, de una o varias representaciones gráficas, del valor del coeficiente de correlación, y mucho mejor aún, probar diferentes modelos de ajuste a los datos. Los campos P0, P13, P14 y P21 se descargarían del acusado componente procedimental que acarrean, y el estudiante podría dedicarse a otras tareas tan necesarias como las de argumentación.

O1.7. Identificar los principales conflictos semióticos en el tema.

El análisis de los puntos anteriores dio como resultado la identificación de algunos aspectos en que la descripción de los diferentes objetos matemáticos podría llevar a ocasionar conflictos semióticos en los estudiantes, debido al uso o interpretación inadecuada de representaciones de dichos objetos (funciones semióticas que no concuerdan plenamente con el significado institucional).

Se describieron y clasificaron estos conflictos potenciales, dependiendo de si se referían al lenguaje, conceptos, propiedades, procedimientos o argumentos. Para cada uno de ellos se analizan las razones por las cuáles consideramos que se establecen funciones semióticas inadecuadas al describir un tipo de objeto, y se sugieren formas en que el profesor puede evitar dichos conflictos en los estudiantes.

Conclusiones respecto a las hipótesis

Al describir inicialmente el Estudio 1, planteamos como hipótesis general que *los libros de texto presentarán un tratamiento diferenciado del tema, dependiendo de la modalidad de Bachillerato a la que se dirigen,* hipótesis que se desglosó en las que ahora discutimos.

H1. Se prevé mayor formalización en los textos dirigidos a la modalidad de Bachillerato de Ciencia y Tecnología.

En general, esta hipótesis se ha contradicho en el trabajo, pues es usual que los textos de la misma editorial, para las dos modalidades, sean muy similares en el tratamiento del tema, aunque los de Humanidades y Ciencias Sociales suelen hacer un estudio más detallado de la organización de datos bivariantes.

En la única variable en que la hipótesis se confirma es respecto a los tipos de argumentación. Son más escasos los argumentos simbólicos o por reducción al absurdo y aparecen con mayor frecuencia en los textos de la modalidad de *Ciencia y Tecnología*, preferentemente como tarea propuesta. Pensamos que este mayor énfasis se debe a las directrices relacionadas con el uso de la demostración para *Matemáticas I* y *Matemáticas aplicadas a las Ciencias Sociales I* como se muestra en el Decreto de Enseñanzas Mínimas:

> Las definiciones formales, las demostraciones (reducción al absurdo, contraejemplos) y los encadenamientos lógicos (implicación, equivalencia) dan validez a las intuiciones y confieren solidez a las técnicas aplicadas. Sin embargo, este es el primer momento en que el alumno se enfrenta con cierta seriedad al lenguaje formal, por lo que el aprendizaje debe ser equilibrado y gradual. El simbolismo no debe desfigurar la esencia de las ideas fundamentales, el proceso de investigación necesario para alcanzarlas, o el rigor de los razonamientos que las sustentan (*Matemáticas I*, MEC, 2007b, p. 45449)
>
> En este contexto, la fuerte abstracción simbólica, el rigor sintáctico y la exigencia probatoria que definen el saber matemático, deben tener en esta materia una relativa presencia. Las fórmulas, una vez que se las ha dotado de significado, adoptan un papel de referencia que facilita la interpretación de los resultados pero, ni su obtención, ni su cálculo y mucho menos su memorización, deben ser objeto de estudio. (*Matemáticas aplicadas a las Ciencias Sociales I*, MEC, 2007b, p. 45474).

H2. Se espera que los contextos utilizados en unos y otros textos favorezcan la apreciación de las aplicaciones de la correlación y regresión en ambas modalidades (Humanidades y Ciencias Sociales y Ciencia y Tecnología).

Igualmente, esta hipótesis se contradice, pues, por un lado, no observamos diferencias de contextos en los dos tipos de textos. Pero además, se observa una proporción importante de ejercicios descontextualizados propuestos en los libros de texto, que en uno de los libros se acerca al 40%. Se ignora así la recomendación desde la educación estadística de dotar de sentido a las tareas, y no presentar la estadística como una colección de recetas o fórmulas de cálculo.

H3. Por otro lado, *se espera una variabilidad en la presentación, dentro de la misma modalidad, de diferentes editoriales,* en consonancia con lo observado en el estudio de Sánchez Cobo (1999) y en otros estudios sobre libros de texto.

Nuestros datos apoyaron esta hipótesis, habiendo encontrado una variabilidad muy grande entre editoriales en cada uno de los tipos de objetos analizados.

Por ejemplo, respecto a los conceptos, destacamos los textos [H4] y [T4], por ser los que más definiciones incluyen, y aunque su tratamiento de la regresión sea escaso, abarca todas las categorías de conceptos. Y en otro extremo, los textos [H7] y [T7], que aunque con diferencias respecto al tratamiento de la noción de frecuencia doble, son los que menos definiciones incluyen, seguidos de [H2] y [T2]. En este sentido, recordamos que para conseguir una comprensión conceptual adecuada de cualquier tópico, se deben incluir los principios del dominio dado y sus interrelaciones (Rittle-Johnson et al., 2001); en este caso, los diferentes tipos de frecuencias, distribuciones y variables, junto a sus distintas representaciones.

En relación a los procedimientos cabe mencionar que sólo en dos textos ([H3] y [T3]) se plantea una actividad al estudiante en la que se pida tan sólo estimar una variable en función de otra. En esta actividad se puede apreciar la motivación por la que, en la mayoría de los casos, se estructura un análisis de correlación y regresión. Mencionamos esta actividad, porque en la resolución que se presenta se plantea una estructura que podría considerarse casi como un procedimiento. Este *mini proyecto* establece su estructura comenzando con una comprensión del enunciado, después se fija su planificación, ejecución, y finalmente, se da la respuesta, junto a su justificación.

También encontramos diferencia entre editoriales en relación a los conflictos semióticos potenciales. Así, las editoriales que publican los libros [H3], [T3], [H4], [T4] presentan pocos conflictos potenciales, le sigue la que publica [H6], [T6]; siendo la editorial de [H1] y [T1] la que presenta un número alto de conflictos en las dos modalidades, seguida de las que publican [H5], [T5], [H7] y [T7].

Finalmente se destacan las diferencias en relación a la tecnología que han sido analizadas con detalle en puntos anteriores.

H4. Asimismo, *podrían presentarse algunos sesgos en cuanto a las características de la correlación (intensidad, signo o tipo) en estos libros*.

Como hemos indicado anteriormente, se encontraron una gran variedad de conflictos semióticos potenciales, al igual que en otros estudios previos de libros de texto, respecto a otros conceptos matemáticos (por ejemplo en Mayén, 2007). Todos estos conflictos se han analizado, ejemplificado y clasificado en el estudio.

Como consecuencia, observamos que la presentación de algunas propiedades (o la ausencia de dicha presentación) podría inducir conflictos semióticos en los estudiantes. Algunos ejemplos de los mismos sería confundir un objeto con su representación gráfica, confusión entre diagrama de barras e histograma, o el diagrama de dispersión y el gráfico de burbujas, generalización demasiado amplia del concepto de correlación o no diferenciar convenientemente las dos rectas de regresión.

Otros ejemplos son el uso de simbolizaciones inadecuadas o imprecisas, la descripción incompleta de procedimientos, o la equiparación de conceptos no equivalentes. En otros casos, las relaciones que se establecen entre la recta de regresión y el coeficiente de correlación debiesen ser reflexionadas en clase por el profesor, pues podrían llevar a resultados equivocados. Este es el caso de la relación entre el coeficiente de correlación y la recta de regresión para establecer su aproximación, indicando que con una misma recta se pueden realizar predicciones para cada una de las variables si el valor de $|r|$ es próximo a 1.

CAPÍTULO 5.

EVALUACIÓN Y DESARROLLO DEL CONOCIMIENTO MATEMÁTICO DE FUTUROS PROFESORES

5.1. INTRODUCCIÓN

En este capítulo se presenta un estudio de evaluación y desarrollo del componente matemático del conocimiento didáctico-matemático (Godino, 2009; 2011), sobre la correlación y regresión, en muestras de futuros profesores de Educación Secundaria y Bachillerato de dos cursos sucesivos. Los datos se recogen a partir de una actividad práctica, realizada en un curso del Máster de esta denominación, obligatorio para los futuros profesores de estos niveles educativos; y se basa en la realización de un proyecto estadístico, que ellos mismos podrían utilizar para la enseñanza del tema en primer curso de Bachillerato.

Este proyecto se desarrolló por completo, aunque, como veremos, los datos individuales de evaluación, recogidos mediante un cuestionario escrito, sólo incluyen una parte de las actividades del mismo, ya que el resto de tareas se evalúa sólo en una muestra pequeña de futuros profesores que las realizan con carácter opcional en el segundo curso. En sesiones posteriores, los futuros profesores analizaron la idoneidad didáctica del proyecto, también en forma individual y por escrito. Los informes de esta segunda actividad realizados por los participantes en el primer año se han utilizado para analizar las diversas facetas del componente didáctico de su conocimiento (Estudio 3).

En lo que sigue se describe, en primer lugar, el contexto en que se llevan a cabo los dos estudios de evaluación (Estudio 2 y 3), y las muestras participantes en los mismos. También se analiza el proyecto estadístico, que es la base de los dos estudios; las actividades que lo conforma y se presentan y discuten los resultados de la evaluación realizada con el mismo.

5.2. OBJETIVOS E HIPÓTESIS DEL ESTUDIO 2

El objetivo general de este segundo estudio es *realizar un estudio exploratorio del conocimiento didáctico-matemático sobre la correlación y regresión, en estudiantes que se preparan para ser profesores de Matemáticas de Educación Secundaria y Bachillerato*. El interés de este objetivo se debe a la escasez de investigaciones previas sobre los conocimientos de futuros profesores sobre este tema, y la necesidad de aportar una formación adecuada para su tarea docente. De modo más específico, con este Estudio 2 pretendemos alcanzar los siguientes objetivos:

O2.1.Elaborar una situación didáctica e instrumentos de evaluación, dirigidos a futuros profesores, que permita contextualizar la correlación y regresión, nos proporcione información sobre sus conocimientos matemáticos (comunes y avanzados), y contribuya a desarrollarlos.

Este objetivo se justifica por la falta de materiales didácticos específicos, dirigidos a futuros profesores sobre este tema, dado que el Máster Universitario tiene una implantación reciente en España. Igualmente, se avala por la necesidad de completar su formación sobre el conocimiento matemático para la enseñanza de la correlación y regresión, pues algunas investigaciones como la de Quintas, Oliveira y Freitas (2013) muestran la existencia de errores conceptuales sobre la correlación y regresión en profesores en ejercicio, con amplia experiencia en la enseñanza del tema.

O2.2.Realizar un estudio exploratorio de evaluación de conocimientos matemáticos (comunes y avanzados) en futuros profesores, utilizando la situación didáctica e instrumento de evaluación construidos para tal efecto.

El estudio de los antecedentes ha mostrado las pocas investigaciones que evalúan el conocimiento estadístico de la correlación y regresión en profesores o futuros profesores, y la más completa que conocemos (Estepa, 1994) está realizada con futuros profesores de educación primaria. Nuestro estudio trata de aportar información sobre los conocimientos de los futuros profesores de Educación Secundaria y Bachillerato. Por otro lado, los resultados se utilizarán para describir algunos elementos necesarios en el conocimiento estadístico para la enseñanza de la correlación y regresión.

En relación a estos objetivos, se han planteado también algunas hipótesis previas, entendidas como expectativas sobre los resultados del estudio, pues al ser exploratorio, no se pretende hace contraste formal de estas hipótesis. A continuación las exponemos.

H1. Al evaluar los conocimientos de los futuros profesores *se prevé encontrar algunas dificultades en conceptos estadísticos elementales requeridos para el trabajo con la correlación y regresión.*

Como hemos señalado, Quintas, Oliveira y Freitas (2013) muestran la existencia de errores conceptuales sobre la correlación y regresión en profesores en ejercicio; pero además, diversas investigaciones previas muestran dificultades de comprensión de gráficos (Arteaga, 2011) o resúmenes estadísticos (Jacobbe, 2008) en profesores y futuros profesores de educación primaria. Sería posible que algunos de estos errores se encuentren en nuestra muestra, puesto que la formación inicial previa no es siempre la licenciatura de matemáticas.

H2. Se espera que los futuros profesores se interesen por el trabajo con datos reales y tecnología en la actividad planteada, y ésta permita desarrollar su conocimiento matemático (común y avanzado) sobre la correlación y regresión.

Esta hipótesis se fundamenta en los escritos en educación estadística que coinciden en señalar la importancia de la tecnología y el uso de datos reales, tanto como fuente de motivación, como para el aprendizaje de la estadística.

5.3. CONTEXTO

Los estudios 2 y 3 se llevan a cabo dentro de la asignatura *Innovación Didáctica e Investigación en Matemáticas* del *Máster Universitario en Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas*, en la especialidad de matemáticas, a lo largo de dos cursos sucesivos: 2012-2013 y 2013-2014, en la Universidad de Granada. En lo que sigue se describen las características generales de este Máster y de la asignatura, pasando a continuación a describir con detalle el material utilizado en el Estudio 2.

5.3.1. EL MÁSTER DE FORMACIÓN DEL PROFESORADO

Este Máster, implantado recientemente, está orientado a los estudiantes que se preparan profesionalmente para ejercer la docencia en dichos niveles educativos. Responde a la obligatoriedad de cursar estudios de máster para ejercer la docencia, dispuesta en la Ley Orgánica de Educación 2/2006 de 3 de mayo y la regulación establecida para estos estudios de posgrado en el Real Decreto 1393/2007 de 29 de octubre, las indicaciones específicas establecidas en la Resolución de 17 de diciembre de 2007, y la Orden Ministerial ECI/3858/2007 de 27 de diciembre.

Más de 40 universidades españolas lo imparten, con una duración de un año académico (60 créditos), e incide en el desarrollo de competencias relacionadas con la formación psicológica, pedagógica, sociológica y didáctica de los futuros docentes. Destacamos, a modo de ejemplo, una de estas competencias:

> 1. Conocer los contenidos curriculares de las materias relativas a la especialización docente correspondiente, así como el cuerpo de conocimientos didácticos en torno a los procesos de enseñanza y aprendizaje respectivos. Para la formación profesional se incluirá el conocimiento de las respectivas profesiones. (MEC, 2007c, p. 53751)

La organización de las enseñanzas del citado Máster es común en las distintas universidades y especialidades (matemáticas, lengua, etc.), y se estructura, como mínimo, en tres módulos: a) El módulo *genérico*, común a todas las especialidades (entre 12 y 15 créditos según la Universidad); b) otro *específico*, que desarrolla las competencias relacionadas con las diferentes disciplinas o especialidades docentes (20 - 30 créditos); y c) el módulo de *Prácticum* (16 -20 créditos), que incluye un periodo de prácticas en instituciones educativas, bajo la supervisión de un tutor, y un Trabajo fin de Máster, como compendio de la formación adquirida.

Los contenidos de los módulos varían de una a otra universidad, aunque en casi todas se imparten algunas sesiones sobre estadística y su didáctica, variando la extensión, ya que algunas universidades dedican 3 créditos a la didáctica de la estadística y probabilidad, mientras que en otras, el tema sólo se incluye a través de la posibilidad de elegirlo para el Trabajo fin de Máster.

En la Universidad de Granada, el módulo genérico está compuesto por tres asignaturas de 4 créditos cada una: *Proceso y contextos educativos,* donde se presenta, por una parte, la evolución de la estructura de las instituciones educativas, la profesión docente, y la función tutorial, su situación actual y perspectivas de futuro; y por otra, criterios de desarrollo curricular en conjunción con el proyecto educativo de centro; *Aprendizaje y desarrollo de la personalidad*, que estudia los factores implicados en el proceso de enseñanza y aprendizaje, y su relación con el desarrollo de la personalidad del adolescente; y *Sociedad, familia y educación*, donde se describe la función social del sistema educativo, la familia y la comunidad, así como procedimientos para conseguir una participación conjunta y efectiva de éstos en la formación de los estudiantes.

El módulo específico se compone igualmente de tres materias: *Complementos de formación disciplinar* (6 créditos), que refuerza al estudiante en temas sobre historia y filosofía de las matemáticas; *Aprendizaje y enseñanza de las materias* (12 créditos), centrada en bloques de contenido tales como competencia matemática y análisis curricular, conocimiento profesional del profesor de matemáticas, aprendizaje matemático y características del razonamiento matemático de los alumnos/as de

Secundaria y Bachillerato; e *Innovación docente e investigación educativa* (6 créditos), donde se tratan estos temas, parte de ellos (4 créditos) centrados en educación matemática. La correlación y regresión podrían contemplarse en las actividades prácticas en cualquiera de las dos últimas materias.

El módulo de Prácticum está compuesto de dos materias: *Prácticas docentes* (10 créditos) con la que los estudiantes adquieren una experiencia práctica en enseñanza de la matemática en un centro educativo, y el *Trabajo fin de máster* (6 créditos), que el estudiante elaborará como síntesis del conjunto de competencias desarrolladas por las materias cursadas y la práctica docente, y que deberá exponer públicamente ante una comisión evaluadora.

Uno de los tipos más habituales de Trabajo fin de máster es la elaboración de una unidad didáctica, fundamentada, para la enseñanza de un tema de matemáticas en Educación Secundaria o Bachillerato; como ejemplo, citamos el hecho que en el curso 2012-2013 uno de los estudiantes realizó una unidad didáctica para la enseñanza de la correlación y regresión, dirigida a estudiantes del Bachillerato de *Humanidades y Ciencias Sociales*, y otro una unidad didáctica para la enseñanza de la correlación y regresión dirigida a estudiantes del Bachillerato en *Ciencia y Tecnología*. Es frecuente también la elaboración de unidades didácticas de otros temas de estadística.

Por último, se incluye un módulo de libre disposición, en el que los estudiantes elegirán dos asignaturas de 4 créditos entre las siguientes: *Atención a la diversidad y multiculturalidad, Atención a los estudiantes con necesidades especiales, Hacia una cultura de paz, Educación para la igualdad,* y *Organización y gestión de centros educativos*, aunque se ofrece la posibilidad de cursar estos 8 créditos con materias de otros másteres oficiales de la Universidad de Granada, y algunos estudiantes tomaron cursos de didáctica de la estadística, dentro del Máster en Didáctica de la Matemática.

5.3.2. LA ASIGNATURA

La asignatura *Innovación Didáctica e Investigación en Matemáticas* se desarrolla presencialmente en 15 sesiones de dos horas y media de duración, y se divide en dos períodos de docencia. El primer período (5 sesiones) es común a todas las especialidades, mientras el segundo (10 sesiones) se emplea para la formación específica a la especialidad de matemáticas. Los contenidos especificados en la guía docente de la asignatura son los siguientes:

1. El proceso de enseñanza y aprendizaje de las matemáticas. Criterios de idoneidad.
2. Investigación en educación matemática. Problemas y líneas. Criterios de calidad. Análisis de ejemplos.
3. Innovación en educación matemática. Fundamentación y finalidad. Buenas prácticas en la enseñanza. Análisis de ejemplos.
4. Recursos para la innovación e investigación: documentos, materiales, tecnología, metodología, interdisciplinariedad.
5. La difusión y aplicación en el aula de los resultados de la investigación e innovación. Fuentes de información.
6. El profesor como investigador e innovador. Análisis crítico de la práctica docente.

Igualmente hay un temario práctico, que incluye actividades de análisis didáctico, en las cuáles se pretende contribuir al desarrollo de las siguientes competencias específicas del título:

- CE39. Conocer y aplicar propuestas docentes innovadoras en el ámbito de las matemáticas
- CE40. Identificar problemas relativos a la enseñanza y aprendizaje de las matemáticas y plantear alternativas y soluciones.
- CE41. Analizar críticamente el desempeño de la docencia, de las buenas prácticas y de la orientación, utilizando indicadores de calidad.
- CE42. Conocer y aplicar metodologías y técnicas básicas de investigación y evaluación educativas y ser capaz de diseñar y desarrollar proyectos de investigación, innovación y evaluación.

En total se realizaron diez actividades de análisis didáctico, tres de las cuáles se desarrollaron con actividades prácticas cuyos datos se analizan en los estudios 2 y 3.

5.4. METODOLOGÍA

Los estudios que presentamos en este capítulo y el siguiente, son esencialmente cualitativos y exploratorios. Se pueden englobar en la denominada *investigación basada en diseño*, en la cual, se utiliza el diseño y el análisis sistemático de estrategias y herramientas instruccionales para la mejora de la enseñanza (Godino et al., 2013).

Una característica de esta metodología (DBRC, 2003) es que combina el diseño de situaciones o ambientes de enseñanza y aprendizaje (en nuestro caso dirigidas a la formación de profesores), y el desarrollo de teorías (sobre el conocimiento didáctico-matemático sobre la correlación y regresión). La investigación y el desarrollo configuran un ciclo de diseño de intervención - experimentación - análisis - rediseño; que en nuestro caso se ha experimentado en dos cursos sucesivos, con ajustes en el diseño de la situación de aprendizaje tras el análisis de la primera intervención. Asimismo, se trata de explicar, y no solo describir, cómo funciona el diseño en contextos reales. De acuerdo a la Orden (2007, p. 13):

> Una dimensión esencial del enfoque es la naturaleza altamente intervencionista de su metodología. Estos estudios son típicamente bancos de prueba para la innovación. El propósito es investigar las posibilidades de mejora educativa, incluyendo en el diseño nuevas formas de aprendizaje.

Más concretamente, el Estudio 2 incluye el diseño de una situación didáctica dirigida a evaluar y desarrollar el componente matemático del conocimiento didáctico-matemático (Godino, 2009; 2011) sobre la correlación y regresión en futuros profesores de Educación Secundaria y Bachillerato. Esta situación se diseña siguiendo el ciclo formativo propuesto por Godino, Batanero, Roa y Wilhelmi (2008), que comienza con el trabajo de los futuros profesores en un proyecto estadístico, que pueden utilizar con sus futuros estudiantes, y continua con una fase de análisis por parte de los futuros profesores de la idoneidad didáctica del proceso de enseñanza-aprendizaje basado en su experiencia personal con dicho proyecto (cuyos resultados se recogen en el Estudio 3).

El diseño incluye un entorno de aprendizaje completo, con tareas, materiales, herramientas y medios para secuenciar y apoyar el aprendizaje (Reimann, 2011). Los materiales están formados, en primer lugar, por datos estadísticos reales tomados del servidor de las Naciones Unidas sobre desarrollo humano (http://hdr.undp.org/es). Otras herramientas son el material impreso con tablas, gráficos, resúmenes de los datos que los futuros profesores han de interpretar; la hoja Excel, para facilitar los cálculos; y recursos interactivos en Internet.

Finalizado el diseño de la situación didáctica y los instrumentos de recogida de datos, se realizaron dos evaluaciones sucesivas, con la finalidad de mejorar el diseño al finalizar la primera experimentación. Tanto el diseño como la experimentación están apoyados en teorías instruccionales específicas (Cobb y Gravejemeir, 2008), en este caso, el enfoque ontosemiótico, y los antecedentes de investigación sobre la correlación y regresión, y sobre formación de profesores.

Una parte de las tareas, que se analizarán a lo largo del capítulo, se han propuesto a los futuros profesores en forma de cuestionario, que permiten tomar datos de su trabajo a lo largo de la situación didáctica. Otras se han realizado conjuntamente en clase, en el primer curso, o se han propuesto con carácter opcional, como ampliación a los participantes en el segundo.

Los cuestionarios han sido construidos siguiendo la metodología usual recomendada por APA, AERA y NCME (1999), para reforzar su validez de contenido y su fiabilidad. La definición semántica de las variables objeto de medición se apoya en fundamentos adecuados para definir "el conocimiento didáctico- matemático sobre la correlación y regresión", que es un constructo inobservable (León y Montero, 2002). Dichos fundamentos son el análisis conceptual de las investigaciones previas y de los libros de texto, expuestos en capítulos anteriores.

Los resultados obtenidos mediante las tareas del cuestionario, permiten evaluar la experimentación de la situación didáctica, proponiendo acciones de mejora de la misma (DBRC, 2003). Tratamos que de la investigación se obtengan, además de la evaluación de la propia situación didáctica y del conocimiento de los futuros profesores, algunos desarrollos teóricos, que puedan ser compartidos por profesores y sus formadores, y así apoyarles en su trabajo. Por otro lado, la situación didáctica diseñada puede utilizarse en otros talleres similares con profesores en formación y en ejercicio, para reforzar su conocimiento didáctico-matemático sobre la correlación y regresión.

Asimismo, se trata de compensar este tamaño limitado de muestra con una mayor profundidad de análisis de los datos, basado en el análisis de contenido (Weber, 1985), que permite efectuar inferencias mediante la identificación sistemática y objetiva de las características específicas de un texto (Ghiglione y Matalón, 1989). En base a la revisión bibliográfica presentada en el Capítulo 3, y al estudio de libros de texto (Capítulo 4), hemos analizado las respuestas escritas de los participantes para definir las categorías de análisis en cada tarea, y elaborar tablas que recogen la frecuencia de cada categoría. También se incluyen ejemplos comentados a lo largo del informe para aclarar las categorías presentadas. Finalmente, se realiza una síntesis cuantitativa, en los apartados principales.

En algunos casos, y para sintetizar los resultados, se presentan conjuntamente las respuestas de los dos cursos en que se evalúa la situación didáctica, comentando, cuando sea necesario, los cambios introducidos en el segundo año de experimentación.

5.4.1. DESCRIPCIÓN DE LAS MUESTRAS PARTICIPANTES

El proyecto se desarrolló en el aula en los cursos 2012-2013 y 2013-2014. En el primero participaron 23 futuros profesores y en el segundo 42, todos ellos alumnos del Máster de Formación del Profesorado de Educación Secundaria y Bachillerato. Como en toda investigación cualitativa, puesto que la muestra no es aleatoria, se exige explicitar los criterios de selección, describiéndola con detalle (Martínez, 2006). Siguiendo a este autor, se procura que la muestra represente lo mejor posible los subgrupos naturales; en nuestro caso, participaron todos los estudiantes que seguían la asignatura citada en el Máster de Secundaria. Asimismo, se ha elegido dos promociones sucesivas para alcanzar mayor representatividad, además de permitir mejorar el diseño y la experimentación de la situación didáctica.

Fueron informados de la finalidad del estudio, y cedieron sus datos, que son tratados de forma anónima. En las Tablas 5.4.1 y 5.4.2 se resume información de su titulación de procedencia, formación didáctica y experiencia docente. Se mostraron interesados por colaborar. Agradecemos su esfuerzo y el cuidado con que completaron las diferentes actividades.

Tabla 5.4.1. Titulación de acceso al Máster y año de finalización de los estudios de los participantes.

	Titulación		Año de finalización			Total (%)
			[1998, 2005]	[2006, 2010]	[2011, 2013]	
Muestra 1	Matemáticas				10	10(43,5)
	Estadística			2		2(8,7)
		Caminos[1]	2	3		5(21,7)
	Ingeniería	Telecomunicaciones			1	1(4,3)
		Topografía[2]	1			1(4,3)
	Arquitectura				1	1(4,3)
	Administración de empresas[3]			1		1(4,3)
	Ciencias Ambientales			1		1(4,3)
	No responden					1 (4,3)
	Total muestra 1 (%)		3(13,6)	7(31,8)	12(54,6)	23(100)
Muestra 2	Matemáticas				24	24 (57,1)
		Caminos[1]	2	4	4	10 (23,8)
	Ingeniería	Telecomunicaciones	1			1 (2,4)
		Química[4]			1	1(2,4)
	Arquitectura		1		3	4 (9,5)
	No responden					2 (4,8)
	Total muestra 2 (%)		4 (10)	4 (10)	32 (80)	42 (100)
Total (%)			7 (11,3)	11 (17,7)	44 (71)	65 (100)

[1]Ingeniería Técnica Superior de Caminos, Canales y Puertos; [2]Ingeniero Topógrafo e Hidráulico; [3]Administración y Dirección de empresas, [4]Ingeniería Química

Tabla 5.4.2. Experiencia didáctica de los participantes (% respecto a la muestra)

Titulación	Experiencia docente			Asignaturas en didáctica			Cursos, congresos, etc.		
	Si	No	Blanco	Si	No	Blanco	Si	No	Blanco
Matemáticas	22	12		11	23		4	30	
Estadística	2			1	1		1	1	
Ingeniería	11	8		1	18			19	
Arquitectura	2	3			5		1	4	
Administración de empresas		1			1			1	
CC. Ambientales		1			1			1	
No indican			3			3			3
Total	37 (56,9)	25(38,5)	3(4,6)	13 (20)	49(75,4)	3(4,6)	6 (9,2)	56(86,2)	3(4,6)

Los participantes habían cursado, en su mayoría, el Bachillerato de *Ciencias y Tecnología*, y observamos (Tabla 5.4.1) que su formación matemática es buena, ya que más de la mitad proceden de la licenciatura de matemáticas, y el resto de estadística, ingenierías, arquitectura, administración de empresas o ciencias ambientales. La mayoría finalizó el curso anterior al inicio del máster, sobre todo en la segunda muestra, aunque encontramos participantes que habían finalizado anteriormente; en algunos casos, debido a la reciente crisis económica sufrida en España, los estudiantes que provienen de arquitectura o ingeniería tenían la necesidad de reorientar su actividad profesional hacia la docencia.

En cuanto a su experiencia o conocimiento didáctico, más de la mitad de participantes habían impartido clases particulares (Tabla 5.4.2), sobre todo los licenciados en matemáticas e ingeniería. Otros habían impartido también clases particulares de otras materias de ciencias como física o química, nuevas tecnologías como programas informáticos SPSS o Fortran, dibujo técnico, idiomas o lenguaje. Los niveles de enseñanza en que han impartido docencia son principalmente Educación Secundaria y Bachillerato, y en algunos casos también primaria y/o universidad.

Observamos en los participantes una escasa formación en didáctica de la matemática, con pocas asignaturas regladas cursadas (sólo el 20% de la muestra). Esta formación se reduce a una o dos asignatura cuatrimestrales, que en algunos casos ha sido anual. También su participación en reuniones en el área de didáctica de la matemática (jornadas, congresos, cursos, etc.) es muy escasa. Sólo un futuro profesor indica haber realizado 280h de cursos de formación, pero no específicos en didáctica de la matemática (curso de formador de formadores y enseñanza a distancia).

5.4.2. MATERIAL Y TRABAJO DE LOS FUTUROS PROFESORES EN EL PROYECTO

La situación didáctica que analizamos en este capítulo, sirvió para proporcionar a los futuros profesores, por un lado, un modelo de actividad de innovación docente que emplease la tecnología, datos reales y el trabajo autónomo de los estudiantes de Bachillerato en torno a proyectos estadísticos. Por otro, para introducirlos en el concepto de idoneidad didáctica, que fue un contenido de este curso; punto que analizaremos en el Capítulo 6.

Se trabajó con los participantes a lo largo de dos sesiones prácticas (dos horas y media cada sesión). En la primera, y como parte de la realización del proyecto estadístico planteado, se recogieron las respuestas de los futuros profesores a las tareas del cuestionario, que sirvieron para evaluar y desarrollar el conocimiento matemático común y ampliado de los futuros profesores. Las actividades finales del proyecto, realizadas en el primer curso en una segunda sesión, colectivamente, se orientaron únicamente al desarrollo del conocimiento matemático de los profesores; en el segundo curso, algunos participantes las desarrollaron en forma más extensa a lo tratado en el aula y personalmente, como actividades opcionales del curso y fuera de su horario lectivo. El resto de la segunda sesión, así como una tercera completa, se dedicó al análisis de la idoneidad didáctica del proyecto, que se discute en el Capítulo 6.

La forma de trabajo en el aula fue proponer sucesivamente una o dos actividades para resolver, dar un tiempo de trabajo para que los futuros profesores aportasen sus soluciones por escrito. Una vez recogido el cuestionario correspondiente por el formador, se discuten colectivamente las soluciones propuestas a las actividades trabajadas antes de seguir con la siguiente. El formador de profesores moderaba la discusión, y realizaba una síntesis de lo aprendido. De este modo se mantenía el interés de los participantes, y no se les cansaba, puesto que los periodos de trabajo personal (10-15 minutos) se alternaban con los de debate; estos últimos de longitud variable.

Justificación del uso de un proyecto en el trabajo con los futuros profesores

La importancia de los proyectos como recurso didáctico en la clase de estadística es resaltada en Batanero y Díaz (2004, 2011) y MacGillivray y Pereira Mendoza (2011), pues, en lugar de introducir los conceptos y técnicas descontextualizadas, en un proyecto se presenta al estudiante las diferentes fases de una investigación estadística: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado.

Por consiguiente, el trabajo con proyectos permite al estudiante ejercitar diferentes componentes del modelo de razonamiento estadístico propuesto por Wild y Pfannkuch (1999), en particular, el ciclo de razonamiento estadístico. Como sugieren Murray y Gal (2002), el trabajo con proyectos desarrolla nuevas competencias pues, la comprensión, interpretación y reacción frente a la información estadística no sólo requiere conocimiento estadístico o matemático, sino también habilidades lingüísticas, conocimiento del contexto, capacidad para plantear preguntas, y una postura crítica que se apoya en un conjunto de creencias y actitudes.

Por otra parte, el trabajo de los alumnos con los proyectos permite aplicar uno de los criterios de evaluación sugerido en el Bachillerato para las *Matemáticas I* y *Matemáticas Aplicadas a las Ciencias Sociales I* (MEC, 2007b), que consiste en que el alumno sea capaz de abordar problemas de la vida real, organizando y codificando datos, formulando conjeturas sobre los mismos, seleccionando estrategias, y utilizando herramientas y modos de argumentación propios de las matemáticas para resolver problemas y realizar investigaciones. Es necesario, entonces, preparar al futuro profesor en la metodología de enseñanza de la estadística basada en proyectos.

Todas estas consideraciones, y el hecho de que se tratase de una asignatura de Innovación Docente e Iniciación a la Investigación en Matemáticas, nos llevó a basar la situación didáctica en el trabajo con un proyecto que combina, junto con esta metodología de trabajo, el uso de la tecnología y de datos reales.

Proceso de diseño del proyecto y de la situación didáctica

El proyecto propuesto a los futuros profesores fue construido para esta investigación, siguiendo la pauta de otros propuestos en Díaz y Batanero (2011), y en particular, incorporando algunas de las tareas y cuestiones relacionadas con la correlación y regresión que se exponen en uno de dichos proyectos (Batanero, Díaz y Gea, 2011; 2014).

Al tratarse de una investigación basada en diseño, se parte de un análisis preliminar que se recoge en los capítulos previos, e incluye el componente matemático (Sección 1.4) y epistemológico - histórico (Sección 1.3), el análisis curricular (Sección 1.5), el estudio de las investigaciones previas sobre correlación y regresión (Capítulo 3), y un análisis detallado de la presentación del tema en los libros de Bachillerato (Capítulo 4). Además, se analizaron en el Capítulo 2 diversos modelos sobre los componentes del conocimiento del profesor, adoptando el utilizado en esta investigación.

Como fruto del anterior análisis, y fijado el significado institucional de referencia de la correlación y regresión en nuestro trabajo, se seleccionaron los objetos matemáticos que se deseaba incluir dentro del diseño de la situación didáctica. De acuerdo a las recomendaciones internacionales sobre la enseñanza de la estadística (por ejemplo, Franklin et al., 2007), se deseaba trabajar con datos reales. Por ello se comenzó un proceso de selección de los datos que se utilizarían, consultando diversos servidores en Internet. Se deseaba, por un lado, que los datos fuesen comprensibles para los estudiantes, que tuviesen interés para ellos, pudiesen trabajarse mediante la hoja Excel, y se dispusiera de diferentes tipos de correlación, como se explicará en las siguientes secciones.

Seleccionado un conjunto de datos adecuado para la elaboración del proyecto, se comenzó un proceso cíclico de diseño y revisión de tareas, consultando las propuestas en los libros de texto o en las investigaciones analizadas, adaptándolas a nuestra situación para responder a nuestros objetivos. Dichas tareas debían estar secuenciadas en el orden lógico seguido en un estudio estadístico, y tener una dificultad creciente. Así, se comienza por tareas que refuerzan conocimientos previos requeridos en el estudio del tema, y se llega a las propias de la correlación y regresión, para finalmente tratar con actividades de ampliación o profundización.

Las versiones sucesivas de las tareas (hasta llegar a la versión final), fueron revisadas varias veces por la investigadora y su directora, así como por los miembros de nuestro equipo de investigación y profesores visitantes que realizaron estancias de investigación en nuestro Departamento. Asimismo, se ensayaron algunas versiones previas dentro de un curso de máster de investigación en didáctica de la matemática, para asegurar su comprensión y legibilidad.

Los contenidos matemáticos trabajados se incluyen en el bloque de contenidos de estadística y probabilidad de primer curso de Bachillerato, en las dos modalidades: *Ciencias y Tecnología* y *Humanidades y Ciencias Sociales*, y tiene conexiones con otros bloques de las asignaturas *Matemática I* y *Matemáticas aplicadas a las Ciencias Sociales I*, así como con otras materias de primer curso de Bachillerato como *Economía* o *Biología*.

Las tareas propuestas en la situación didáctica se diseñaron teniendo en cuenta los campos de problemas identificados en el estudio histórico, y otros objetos matemáticos abordados. Se cuidó el formato de la misma, y la situación que da origen al proyecto, procurando que las cuestiones que formaran parte de la tarea tuviesen potencial para promover la discusión en el desarrollo de la clase. Con los criterios anteriores, se llegó al proyecto *Relación de la Esperanza de vida con diversos indicadores del desarrollo humano* (Ver Anexo I) cuyos objetivos se analizan a continuación.

Objetivos planteados en el proyecto

Al comenzar la primera práctica, el formador de profesores presentó el proyecto a los futuros profesores, indicándoles los siguientes objetivos de su trabajo con dicho proyecto:

1. *Analizar un proyecto de innovación docente para la enseñanza de la estadística en Primer Curso de Bachillerato.* Puesto que se trabajaba en un curso de Innovación docente, los contenidos del citado curso se contextualizan por medio de ejemplos que permitan a los futuros profesores comprender mejor las características de la innovación docente, y desarrollar competencias de análisis de la calidad de dichas innovaciones.

2. *Adquirir competencias en el trabajo con proyectos y la enseñanza con proyectos.* Como se ha indicado, hoy día se recomienda el trabajo con proyectos en la clase de estadística, método que también podría ser adecuado para otros contenidos. Pero la enseñanza con esta metodología requiere una formación específica del profesor.

3. *Reflexionar sobre los conocimientos y competencias desarrolladas por los estudiantes.* El trabajo con proyectos permite desarrollar conocimientos estadísticos, además de competencias, tales como la capacidad discursiva de los estudiantes, y sus habilidades de pensamiento crítico (Nolan y Speed, 2002).

4. *Conocer fuentes de datos y recursos en Internet para la enseñanza de la estadística, desde una perspectiva multidisciplinar.* Estas fuentes de datos les permitiría a ellos diseñar sus propios proyectos para utilizarlos con sus estudiantes.

5. *Concienciarse sobre el impacto de las nuevas tecnologías en las competencias y conocimientos matemáticos requeridos por los estudiantes, y en general, por los ciudadanos.* El trabajo con la tecnología dentro del proyecto, permite a los futuros profesores comprobar cómo desaparece el interés en la capacidad de cálculo o representación gráfica. Por el contrario, aumenta la necesidad de competencia de búsqueda de información en Internet, uso de la tecnología para realizar el análisis estadístico, y habilidades de interpretación y razonamiento matemático.

El formador indica que este método es actualmente recomendado para la enseñanza de la estadística, y habla de algunas actuaciones tales como los concursos de proyectos estadísticos organizados por la Sociedad Española de Estadística e Investigación Operativa (www.incubadoradesondeos.es) y otras instituciones.

Tipos de actividades realizadas

En lugar de centrarnos exclusivamente en la correlación y regresión, se presentó a los futuros profesores un proyecto estadístico completo, donde se comienza con una pregunta de investigación, cuya respuesta se obtiene a partir del análisis de los datos. Las actividades incluidas del proyecto se pueden clasificar en tres grupos:

1. Un primer grupo sirven para familiarizar al estudiante con la principal variable dependiente (Esperanza de vida). Sirvieron para evaluar los conocimientos previos de los futuros profesores sobre contenidos que son prerrequisitos para el estudio de la correlación y regresión.

2. A continuación, se presentan cuatro actividades diseñadas propiamente para introducir el tema de correlación y regresión en el Bachillerato, que se utilizan en esta investigación también para evaluar los conocimientos de los futuros profesores sobre el tema.

3. Un tercer grupo de actividades profundizan o amplían los conocimientos sobre correlación y regresión, y proporcionan información sobre recursos didácticos utilizables en su enseñanza. En el primer año no fueron objeto de evaluación; se utilizan para desarrollar su conocimiento didáctico del tema, y evaluar posteriormente dicho conocimiento mediante el análisis de la idoneidad didáctica de todo el proyecto. En el segundo año, algunos participantes completaron estas actividades opcionalmente, proporcionando informes sobre las mismas, que permite analizar, en algunos casos, los conocimientos mostrados en estas actividades.

A continuación se presentan estos tres grupos de actividades. Para todas incluimos una solución correcta posible (experta), identificando la configuración epistémica de objetos matemáticos prevista, y los resultados de la evaluación. Asimismo, informamos brevemente del desarrollo de la experiencia, a partir del diario de observación recogido por las investigadoras.

5.5. DATOS Y ACTIVIDADES INICIALES

Fuente de datos

Los datos del proyecto fueron tomados de las Naciones Unidas (http://hdr.undp.org/es/data), que los usa en sus informes sobre desarrollo humano. El formador de profesores presentó a los participantes este servidor, entrando en él mismo. Los participantes pudieron observar que, dentro del servidor, la organización explica la procedencia y significado de los datos empleados para calcular el Índice de Desarrollo Humano (IDH), y otros indicadores compuestos. El formador de profesores indicó que todos ellos se publican periódicamente en el *Informe sobre Desarrollo Humano*, son proporcionados por diversas fuentes públicas internacionales, y representan las mejores y más actualizadas estadísticas disponibles para esos indicadores, en el momento en el que se prepara el Informe.

Los futuros profesores pudieron comprobar que los datos originales utilizados para estos estudios están a disposición del ciudadano en la página web citada; y pueden ser descargados en formato Excel, por lo que podrían ser analizados directamente, que es lo que se hace en este proyecto. El formador de profesores proporcionó previamente a los participantes una hoja Excel con los datos concretos que se analizarían durante la práctica, utilizando para ello la plataforma de docencia de la asignatura, donde, asimismo, se colocaron todos los materiales a utilizar.

Las Naciones Unidas proporciona un conjunto muy grande de variables (indicadores) agrupadas en varias categorías como salud, género, igualdad, economía, educación, etc. Es posible seleccionar las variables que se deseen, bien para todos los países que proporcionan estos indicadores (194 fueron seleccionados para trabajar en el proyecto) para uno o varios años, o para países de una zona geográfica específica.

Este servidor proporciona también una serie de recursos de visualización, que producen, por ejemplo, gráficos de burbuja o mapas, comparación de países, etc. Por todo ello, nos pareció de gran interés que los futuros profesores conociesen este recurso, y hacerles participar en una innovación docente basada en estos datos. Por otro lado, la gran cantidad de variables incluidas en esta base de datos incluye diversas intensidades y signos de la correlación entre algunos pares de estas variables, así como relaciones lineales y no lineales.

Para el proyecto se seleccionaron un conjunto de 9 variables para el último año disponible (2009). Una de ellas (la esperanza de vida) haría el papel de variable dependiente para el estudio de la correlación y regresión en el proyecto. El resto de variables se han elegido para tener en cuenta diversas intensidades, signos y tipos de relación (que se analizan con detalle en la Sección 5.7).

Inicio del trabajo con el proyecto

El formador de profesores comenzó el trabajo en el proyecto analizando con los participantes los datos que se utilizarían. Se señaló a los futuros profesores que este es un punto importante en cualquier estudio estadístico, pues como indicó Moore (1991), la estadística es la ciencia de los datos. Mientras que en otras ramas de matemáticas, los datos y contextos suelen ser imaginarios, y el interés se centra en los conceptos, en este proyecto (y en estadística, en general), los datos son tan importantes como los conceptos. Para ello se plantearon las siguientes preguntas:

Indicadores del desarrollo humano (IDH): En el fichero Excel se proporciona una serie de indicadores recogidos por las Naciones Unidas. ¿Podrías explicar qué significa cada una de las variables del fichero? ¿Quién y cómo las calcula? ¿Cómo se recogen los datos?

Se indicó a los futuros profesores que, cuando los estudiantes trabajan con datos reales aprenden los diferentes tipos de variables y escalas de medida, el manejo de datos faltantes, los errores de medición, aspectos de fiabilidad de los datos, y la consecuencia de los valores atípicos. Además, aumenta su motivación, y se inician en el trabajo interdisciplinar, aprendiendo a valorar la utilidad de la estadística (Hall, 2011). También se indicó que la aleatoriedad de las situaciones hace que los resultados no sean únicos, presentándose mayor variabilidad en los datos que en otras áreas de las matemáticas (Sánchez y Batanero, 2012).

A continuación, el formador, junto con los futuros profesores, analizaron cada una de las variables del fichero. El formador comenzaba preguntando si alguno conocía la variable en particular; generalmente, algún participante daba una definición aproximada, que el formador completaba leyendo la descripción dada en el servidor de Naciones Unidas. Dichas variables son las que se indican a continuación:

V1. *Esperanza de vida al nacer:* Número de años que un recién nacido espera vivir, si se mantienen las tasas de mortalidad de ese país durante la vida del niño. El formador de profesores indicó que esta sería la variable dependiente en el proyecto. También explicó que se pensó que potencialmente podría tener interés para un estudiante conocer su esperanza de vida, la de su país, comparar con otros países, y analizar los factores que la determinan. Se informó que la fuente que suministra los datos es el Departamento de asuntos Económicos y Sociales de las Naciones Unidas (UNDESA).

V2. *Índice de Desarrollo Humano (IDH):* Esta variable era desconocida para los participantes, aunque algunos adivinaron su significado: "*algo que indica si un país está más o menos desarrollado*". Los participantes observan que la variable varía de 0 a 1; el formador informa que mide los logros promedios de un país en tres dimensiones básicas: una vida larga y saludable, conocimiento, y un nivel económico de vida razonable. Además, se explica a los participantes que esta variable se elige por ser la variable objetivo en el informe preparado por las Naciones Unidas. Se informa que estos datos son calculados por la Oficina encargada del Informe del Desarrollo Humano (HDRO), basándose en datos ofrecidos por el Departamento de asuntos Económicos y Sociales de las Naciones Unidas (UNDESA), el Instituto de Estadística de la UNESCO, el Banco Mundial, y el Fondo Monetario Internacional (IMF).

V3. *PIB per cápita (en US$):* Prácticamente todos los participantes han oído hablar de la renta per cápita o del Producto Interior Bruto. De ahí fue fácil deducir que esta variable mide los ingresos de un país al año (obtenidos por todos los productores residentes y de todos los impuestos a los productos menos los subsidios, dividido por el número de habitantes; la fuente que suministra los datos es el Banco Mundial. Su unidad de medida es miles de dólares por habitante. En el contexto actual de crisis, los participantes estaban motivados por las variables económicas.

V4. *Tasa de fecundidad entre adolescentes:* El formador explica que la variable indica el número de partos al año por cada 1.000 mujeres vivas entre 15 y 19 años en el país dado. Se discute el por qué las Naciones Unidadas recopilan datos de esta variable; se concluye que junto con la siguiente, son variables relacionadas con la salud y con las desigualdades sociales. Se informa que los datos son recopilados por el Departamento de asuntos Económicos y Sociales de las Naciones Unidas (UNDESA). Resultó de interés a los participantes por referirse a jóvenes.

V5. *Tasa de mortalidad de niños menores de cinco años:* Muchos participantes conocían el significado de la tasa de mortalidad; el formador de profesores aclara que en este caso mide el número de muertes de menores de 5 años por cada 1.000 nacidos vivos al año; la fuente que suministra los datos es UNICEF. Se introduce fundamentalmente por tener una correlación negativa con la dependiente.

V6. *Gasto público en salud*: % del PIB dedicado a financiar la salud. Puesto que en la situación actual de crisis se está disminuyendo el gasto público en salud en España, esta variable interesó a los participantes. Los datos los proporciona el Banco Mundial.

V7. *Índice de educación*: Índice que varía entre 0 y 1, y tiene en cuenta el número medio de años de escolarización de los adultos y el número esperado de años de escolarización de los niños. Se incluye por su interés potencial para futuros profesores. La fuente que suministra los datos es la Oficina encargada del Informe sobre Desarrollo Humano (HDRO).

Finalmente, se añadieron dos variables, para tener ejemplos de correlaciones muy bajas o inexistentes con la dependiente:

V8. *Población, total* considerando ambos sexos (en miles). Número de habitantes (en miles) de cada país. La fuente que suministra los datos es el Departamento de asuntos Económicos y Sociales de las Naciones Unidas (UNDESA).

V9. *Población urbana*: % de la población que vive en ciudades. Estos datos se calculan en el Departamento de asuntos Económicos y Sociales de las Naciones Unidas (UNDESA).

Se analizó con los futuros profesores la naturaleza de las variables. Como vemos, todas son cuantitativas, y salvo V5, continuas; el formador indicó que se debe llevar cuidado cuando se interpreten los valores de estas variables. También recordó que, aunque V1, V3, V4 y V8 pueden ser tratadas como continuas al poseer naturaleza decimal, en realidad miden unidades enteras (años, dólares, nacimientos y habitantes, respectivamente). Las variables V6 y V9 son porcentajes, por tanto toman un valor entre 0 y 100, y el resto representan promedios, concretamente las variables V2 y V7 representan números índices, acotados entre 0 y 1.

Una vez que se proporcionó a los participantes el fichero Excel con los datos, y para motivarlos, se les explicó que en la actualidad diversos organismos internacionales, como las Naciones Unidas, tratan de obtener indicadores de la calidad de vida de sus ciudadanos, y estudiar su evolución, así como los factores que los determinan (Ridgway, Nicholson y McCusker, 2008). También se reiteró que estos datos son de libre disposición en Internet, y pueden descargarse en formato Excel. Por ello, se requieren nuevas habilidades en la interpretación de la información, y surge la necesidad de que los ciudadanos sean estadísticamente cultos. Los futuros profesores tuvieron oportunidad de acceder a estos datos y otros complementarios en Internet.

5.6. ACTIVIDADES DE EVALUACIÓN DEL CONOCIMIENTO ESTADÍSTICO PREVIO

Las cuatro primeras actividades que se plantearon en el proyecto, permitieron evaluar y recordar conocimientos previos de los futuros profesores, relevantes para la correlación y regresión; en una clase de Bachillerato permitirían al profesor evaluar estos conocimientos en sus estudiantes. Comprenden la interpretación de una serie de gráficos estadísticos, la elección de un promedio representativo e interpretación de percentiles.

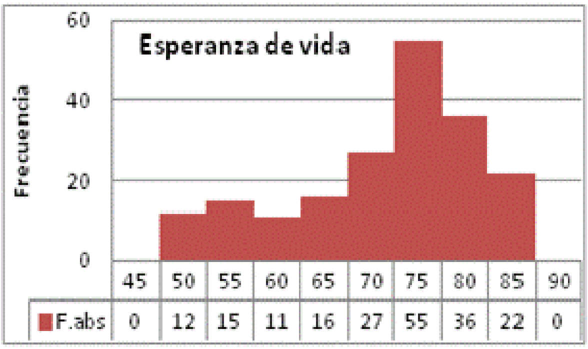
5.6.1. INTERPRETACIÓN DE GRÁFICOS ESTADÍSTICOS

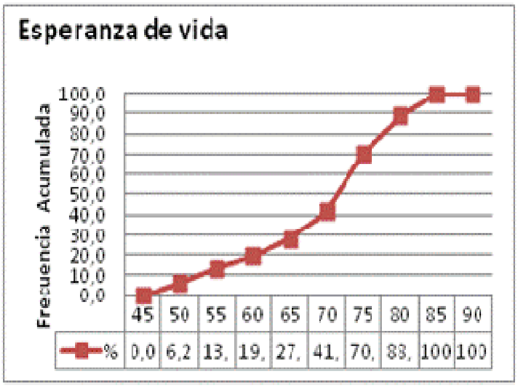
Para evaluar el conocimiento de los futuros profesores sobre la variable estadística, su distribución y representación gráfica, se les propuso las actividades que se analizan a continuación. El formador de profesores las presentó conjuntamente, y motivó a los participantes indicándoles que, puesto que el objetivo del proyecto es analizar cuáles son las variables que influyen en la esperanza de vida, en primer lugar se quiere conocer a fondo la distribución de dicha variable.

5.6.1.1. ANÁLISIS A PRIORI Y DESARROLLO

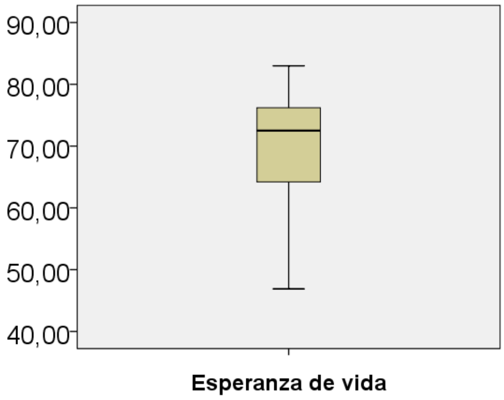
Se presentó a los participantes la Actividad 1, indicándoles que se pretendía que interpretasen diferentes resúmenes estadísticos y representaciones gráficas, explicando la información que aportan cada uno de ellos en el estudio de la distribución de la variable Esperanza de vida. Se les ofrecieron distintas tablas de frecuencias (absolutas y porcentajes), y resúmenes estadísticos de tendencia central (media, mediana, moda), posición (mínimo, máximo y cuartiles) y dispersión (desviación típica) de la esperanza de vida.

1. *Esperanza de vida.* Hemos preparado una tabla de frecuencias, gráficos y calculado algunos estadísticos de la esperanza de vida en el total de países de la muestra. Interpreta estos gráficos y los estadísticos y dinos qué nos indican sobre la distribución de la esperanza de vida.





Estadísticas de esperanza de vida			
Media	69,17	Máximo	83,00
Mediana	72,50	Primer cuartil	64,20
Moda	72,70	Tercer cuartil	76,10
Mínimo	46,90	D. Típica	9,88



Nuestra hipótesis, confirmada a lo largo del desarrollo de la actividad, es que el hecho de mezclar distintas representaciones (histograma, diagrama de frecuencias porcentuales acumuladas y gráfico de caja), propiciaría un espacio de traducción entre ellas. Como se verá en la evaluación, muchos participantes obtuvieron los estadísticos directamente de las gráficas y tabla presentadas Todo ello propició el proceso de transnumeración (Wild y Pfannkuch, 1999), obteniendo información no disponible en los datos brutos, mediante un cambio de representación.

Solución experta y objetos matemáticos implicados

En el debate final de la actividad hubo acuerdo en que el diagrama acumulativo es el más sencillo de interpretar de los tres gráficos. Se reconoce que informa del porcentaje de países por debajo de un cierto valor para la variable Esperanza de vida, en el conjunto de países de la muestra. Como ejemplo, un participante indicó que el 69,7% aproximadamente tiene una esperanza de vida de 70 años o menos. El formador recordó que el gráfico permitirá situar cualquier país en su percentil, en función de la esperanza de vida. Resaltó que la frecuencia acumulada es una función creciente, siendo más acusado el crecimiento en el intervalo de edad de 70 a 80 años, que en el resto de intervalos. No obstante, en el debate se observó que algunos participantes confundieron en este gráfico frecuencias y frecuencias acumuladas. Este punto se reforzó en la segunda implementación del proyecto, haciendo énfasis en la interpretación de este diagrama.

El formador recordó que el histograma nos proporciona información del número de países, según su esperanza de vida, y diferentes intervalos de valores. Como ejemplo, observa que hay 11 países con esperanza de vida en torno a 60 años (marca de clase; intervalo [57,5-62,5]). Los participantes reconocen que permite deducir fácilmente el intervalo de edad modal, que se sitúa alrededor de los 75 años, y la moda de la variable, que es 72,7 años.

Otros participantes observan que la distribución de la variable es asimétrica, con cola a la izquierda, y aunque se pudiera observar esta característica al comparar media y mediana de la tabla, se visualiza mucho mejor en el histograma. Algunos futuros profesores del primer año preguntan en qué intervalo se debe incluir un valor que coincide con el extremo del intervalo. El formador recuerda que puede haber diferentes criterios, aunque usualmente se colocaría en el intervalo siguiente (es decir, donde coincide con el extremo inferior).

El formador pasa a debatir el gráfico de caja que, por un lado, indica que no hay valores atípicos, y, por otro, divide la población en cuatro partes de igual número de efectivos. Un participante indica que esta información ya la proporcionaban la mediana y los cuartiles. El formador enfatiza el amplio rango de variación que presenta esta variable, donde el 75% de los países posee una esperanza de vida inferior a los 76,1 años de edad (Q_3), concentrándose entre los 64,2 y 76,1 el 50% del total de países de la muestra.

Llegados a este punto, algún participante en el primer año de la experimentación indicó que no estaba seguro de comprender el significado de la mediana. El formador de profesores recordó la definición, y para aclararla, añadió que, ordenados los datos de menor a mayor valor de la variable, se tomaría el dato central (o los dos centrales si el número de datos es par), siendo este dato central (en caso de número impar) el que correspondería a la mediana; si hubiese dos datos centrales con valores diferentes se tomaría como mediana la media de estos dos valores.

Para asegurar la comprensión de este concepto, el formador de profesores propuso como ejemplo calcular la mediana de la estatura de los participantes. Ellos se desplazaron a una parte del aula donde había suficiente espacio; se ordenaron por estaturas, y se sacó al que estaba en el centro; era una chica que indicó el valor de su estatura; el formador de profesores indicó que esta estatura era el valor mediano. Recordó entonces que la mediana es el centro de la distribución, pero una vez que ésta se ordena, siguiendo el orden numérico. La actividad sirvió para relajar la clase; gustó mucho a los participantes, quienes indicaron que la usarían con sus estudiantes.

En la Tabla 5.6.1, presentamos un resumen de la configuración de objetos matemáticos implicados en el desarrollo de la Tarea 1. Cobran protagonismo los conceptos de distribución, y estadísticos de centralización, posición y dispersión, y procedimientos como lectura de gráficos, tablas, y traducción entre diferentes representaciones. De ahí que el lenguaje sea un elemento de gran importancia para esta tarea, así como la justificación y argumentación de los estudiantes.

Tabla 5.6.1. Objetos matemáticos implicados en la Tarea 1

Situación problema	P0.Organización/Reducción de datos
Lenguaje	L1.Verbal L3.Tabular: tablas de frecuencias L4.Gráfico: histograma, gráfico de caja y diagrama de frecuencias acumuladas L5.Numérico
Conceptos	CI1.Variable estadística, valores, rango, intervalos, extremos, marcas de clase CI2.Distribución de una variable estadística CI3.Medidas de tendencia central: media, mediana y moda, percentiles, cuartiles CI4.Medidas de dispersión: desviación típica CI7.Números enteros y decimales, porcentajes
Propiedades	PPI2. Relación entre frecuencia absolutas, acumulada, porcentaje
Argumentos	A1. Análisis de ejemplos/ contraejemplos A2. Uso de Gráfico A5. Verbal – deductivo
Procedimientos	PC3.Traducción entre representaciones: gráfica, tabular y numérica PCI1.Lectura de gráficos PCI2.Lectura de tablas

5.6.1.2. RESULTADOS Y DISCUSIÓN

Las respuestas dadas por los futuros profesores a la Tarea 1 en la fase de resolución (antes del debate) se presentan a continuación, clasificadas según el gráfico o estadístico que se interpreta. Se ha valorado la corrección de la interpretación por parte de cada participante, y se han categorizado según los objetos matemáticos utilizados en dicha interpretación, aportando ejemplos para clarificar las categorías.

Interpretación del diagrama acumulativo

Las interpretaciones de los futuros profesores de este diagrama se han clasificado en las siguientes categorías:

Describe el contenido del gráfico. Cuando se describe el significado del gráfico acumulativo de modo general, por ejemplo, indicando que se representan las frecuencias acumuladas de la esperanza de vida en el conjunto de países de la muestra. En algunos casos, esta explicación se acompaña de la interpretación de datos particulares del gráfico, como en la siguiente respuesta correcta, en la que el futuro profesor muestra un nivel de lectura entre los datos (Curcio, 1989):

> En el gráfico se observa, que sólo el 6’2% de los países tienen una esperanza de vida menor de 50 años, que es el dato más drástico. Un dato destacado es que el 47% de los países tienen una esperanza de vida entre 70 y 80 años. (JP).

Encontramos interpretaciones incorrectas como en el primero de los siguientes ejemplos, donde el participante confunde frecuencias acumuladas (lo que crece con la esperanza de vida) y absolutas. Este conflicto semiótico podría limitarse a una interpretación incorrecta de un gráfico (representación) o a confusión de conceptos. El segundo ejemplo confunde la frecuencia con el número de muertes (FR):

> Como podemos observar en el gráfico 1 existen 0 países donde la esperanza de vida es de 45 años, el 6'2% tiene una esperanza de vida de 50 años, y la esperanza de vida va aumentando hasta que el 100% de los países tienen una esperanza de vida entre los 85 y 90 años. (ER).
> Hasta los 70 años ha muerto el 40% de la población, en los últimos 20 años muere el resto ≅ 60%, hay una mayor concentración de muertes en los últimos 20 años (FR).

Interpreta el mínimo. Cuando se interpreta el menor valor de esperanza de vida de los países del estudio, interpretación que también aparece en la lectura de gráficos en el trabajo de Arteaga (2011). Aunque debiera ser fácil para los futuros profesores, encontramos interpretaciones incorrectas como por ejemplo EGA, pues el mínimo de la gráfica se refiere a la Esperanza de vida, no a la edad de muerte. Esta explicación puede deberse a falta de capacidad de comunicación escrita, pues los participantes saben, por su experiencia, que muchas personas mueren con más de 45 años. En todo caso, el futuro profesor no llega al nivel de lectura de datos según Curcio (1989):

> Antes de los 45 años no muere nadie (EGA).

Interpreta el máximo. Esta categoría también aparece en el trabajo de Arteaga (2011), y la encontramos cuando se interpreta la mayor edad alcanzada en esperanza de vida por los países del estudio. Por lo general, estas interpretaciones son incorrectas porque confunden la frecuencia acumulada con la absoluta y/o el valor de la variable, conflicto semiótico ya señalado. En los siguientes ejemplos, ChC interpreta la igualdad de frecuencia acumulada como igualdad de esperanza de vida, es decir, confunde valor y frecuencia de la variable, error encontrado por Cobo (2003) y Arteaga (2011). En el segundo caso, MM piensa que al ser el máximo de la frecuencia acumulada igual a 100, todos los países tienen igual esperanza de vida.

> La esperanza de vida a los 85 años es igual a la esperanza de vida a los 90 años; (ChC).
> También se observa que existe el 100% de países que tienen la misma esperanza de vida. (MM).

Interpreta el rango. Cuando se aporta información del intervalo de valores que toma la variable, categoría también hallada por Arteaga (2011), y que en nuestro caso son siempre correctas. Supone un nivel de lectura entre los datos, ya que los alumnos obtienen previamente los extremos del gráfico para deducir el rango. Algunos participantes interpretan el rango en vez del máximo y mínimo de la variable, como el siguiente ejemplo:

> El gráfico 1 nos dice que las esperanzas de vida de todos los países están entre 45 y 90. (AJD).

Interpreta la moda. Cuando se interpreta la esperanza de vida más frecuente entre los países del estudio; también encontrado en Arteaga (2011). Son pocos los futuros profesores que interpretan este parámetro, todos de modo correcto, y en la mayoría de los casos no se da directamente la moda, sino el intervalo modal:

> La mayoría de países tienen su esperanza de vida comprendida entre los 70-75 años. (EGA).
> La esperanza de vida suele estar entre los 80 y 75 años. (ANL).

Interpreta los cuartiles y/o la mediana. Algunos participantes estiman estos estadísticos a partir del diagrama acumulativo, dando sus estimaciones. Para ello, observan la frecuencia a partir del gráfico, por lo que muestran un nivel de lectura entre los datos (Curcio,1989) al realizar comparaciones y cálculos con los datos del gráfico:

> El primer cuartil está aproximadamente en 63, la mediana en 72 y el tercer cuartil en 77 (AJD).

> Según la gráfica 1, la esperanza de vida de la mitad de los países aproximadamente (un 41%) es de entre 45 y 70 años. El resto, un 59% tendrá una esperanza de vida de más de 70 años. Me ϵ [70,75] (MIH).

Interpreta la pendiente. En esta categoría se utilizan conocimientos de representación gráfica de funciones pues, aunque el diagrama acumulativo es siempre creciente, la tasa de crecimiento (pendiente) no es homogénea. El alumno lo nota, y obtiene de ello una conclusión sobre la esperanza de vida, aunque en general, incorrecta. En general, se suele interpretar la tasa de crecimiento con el número de muertes en los países, mostrando un conflicto semiótico al confundir dos conceptos (valor de la variable, tasa de crecimiento y frecuencia).

En el primer ejemplo que mostramos, se indica el intervalo de mayor pendiente, donde mayor número de países tiene esa Esperanza de vida; el participante indica que hay más muertes a esa edad, lo cual es impreciso aunque corresponde a la misma idea. En el segundo ejemplo, se concluye erróneamente que en el intervalo de mayor pendiente se encuentra la moda de la Esperanza de vida. Esta categoría no aparece en Arteaga (2011).

> Según la pendiente del gráfico podemos ver que de 45 a 65 años no se muere mucha gente; pero de 70 a 75 años, al ser la pendiente más pronunciada, indica que se muere mucha gente en este intervalo. (EGO).
> Desde la primera gráfica observamos que la mayor diferencia se tiene entre los 70-75 años, luego nos indica que la esperanza de vida está entre estos valores. (MJG).

Indica cómo calcular algunos estadísticos a partir del gráfico. Algunos participantes explican cómo obtener algunos estadísticos de la Esperanza de vida a través del diagrama. Esta categoría tampoco aparece en Arteaga (2011). Por ejemplo, JMV explica cómo obtener el número de países que poseen una Esperanza de vida concreta al restar las frecuencias acumuladas que se incluyen en el listado horizontal:

> […] y como tenemos la función acumulada deberíamos de restar los valores para obtener las frecuencias; si hacemos esto podemos ver que la mediana está en torno a los 75 años, esto lo podemos ver aproximadamente en el diagrama de caja (4) gráficamente. Para calcular la media hay que hacer la proporción [fórmula de la media] Para calcular la probabilidad vivir menos que [expresión probabilidad] sería el ver de dicha función partido por 100. (JMV).

En la Tabla 5.6.2 podemos observar las diferentes interpretaciones de los futuros profesores al diagrama acumulativo. Para interpretar esta tabla, y las siguientes en esta sección, hay que tener en cuenta que, en general, se hace más de una interpretación por participante; al final del análisis, a modo de síntesis, se presentará el número medio de interpretaciones correctas e incorrectas por participante, según cada muestra.

Destacan las interpretaciones que describen el contenido del gráfico y las que interpretan los valores máximos y mínimos, seguidos de la del rango y la pendiente. En general, el porcentaje de interpretaciones correctas es alto en los dos grupos, lo que muestra el buen conocimiento del gráfico de los futuros profesores que hacen estas interpretaciones. Son excepciones la interpretación de la pendiente y del máximo y mínimo, que es incorrecta porque varios alumnos confunden la frecuencia, o el valor de la variable al hacer esta interpretación. Al compara los cursos en que se realizó la experiencia, vemos una mejora en la interpretación, globalmente (pasando del 66,7%

global de correctas 72,2 %). Asimismo, se observa mejora en la interpretación del mínimo y de la pendiente, que fueron los puntos más conflictivos en la primera muestra.

Tabla 5.6.2. Interpretaciones del diagrama acumulativo y % respecto a categoría y muestra

	Muestra 1			Muestra 2	
Interpretación	Correcta	Incorrecta	Blanco	Correcta	Incorrecta
Contenido	8(72,7)	3(27,3)		18(66,7)	9(33,3)
Máximo	6(85,7)	1(14,3)		5(83,3)	1(16,7)
Mínimo	1(16,7)	5(83,3)		5(71,4)	2(28,6)
Rango	7(100)			4(100)	
Moda	1(100)			2(100)	
Cuartiles o mediana	1(100)			3(100)	
Pendiente	2(35,3)	4(66,7)		15(65,2)	8(34,8)
Indica estadísticos	2(100)				
No interpreta			1(100)		
Total interpretaciones	28(66,7)	13(31)	1(2,4)	52(72,2)	20(27,8)

Puesto que no todos los alumnos interpretan un mismo aspecto, y las categorías no son excluyentes, hemos considerado de interés analizar la distribución del número de interpretaciones correctas en los dos grupos (Figura 5.6.1). Observamos unas distribuciones muy similares; lo más frecuente fueron una o dos interpretaciones correctas; en pocas ocasiones aparecen tres o cuatro (segundo grupo). Alrededor del 78% en la primera muestra y el 74% en la segunda hace al menos una interpretación correcta de este gráfico.

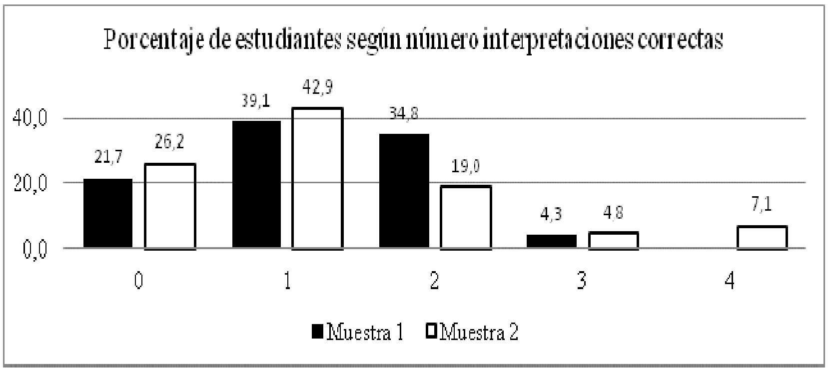


Figura 5.6.1. Número de interpretaciones correctas del diagrama acumulativo en las dos muestras

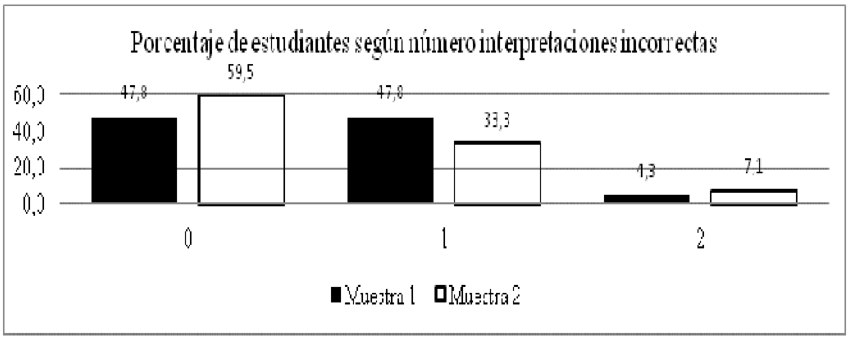


Figura 5.6.2. Número de interpretaciones incorrectas del diagrama acumulativo en las dos muestras

Igualmente, algunos participantes hicieron más de una interpretación incorrecta (Figura 5.6.2). Al comparar la distribución de interpretaciones incorrectas en las dos muestras, se observa mayor número de participantes que no cometen errores o sólo uno en la segunda muestra. Deducimos que la presentación más cuidadosa que se hizo del diagrama acumulativo el segundo año favoreció su interpretación correcta por los futuros profesores, aunque de todos modos, hay poca diferencia, debido a que los resultados fueron buenos los dos años.

Interpretación del histograma

Encontramos menor número de interpretaciones del histograma que del diagrama acumulativo, donde se repiten las mismas categorías, y encontramos algunas nuevas:

Describe el contenido del gráfico. Cuando explica el significado del histograma, indicando que cada rectángulo se refiere al número de países que poseen una determinada Esperanza de vida, y en algunos casos, la interpretación de casos particulares del histograma, como el primero de los ejemplos siguientes. En el segundo ejemplo se describe la frecuencia en cada intervalo del histograma:

> En este caso la frecuencia representa el nº de países cuya esperanza de vida se encuentra en el intervalo del eje horizontal. Por ejemplo, en el intervalo marcado se interpreta que hay 16 países cuya esperanza de vida está entre 60 y 65 años. (ME).
> Hay 55 países que tienen una esp. de vida de 75, 36 países con esp. de vida de 80 y 22 países que tienen una esperanza de 85, son pocos los países que tienen una esperanza de vida menor a 65 años. (DG).

Ejemplos de interpretaciones incorrectas son las de CM, quien confunde la media con el valor de la variable, error descrito por Cobo (2003). En menor medida, encontramos futuros profesores que interpretan la frecuencia como personas y no países, o confunden la esperanza de vida con el número de muertes en los países. Otro ejemplo es el de MIC, quien confunde la frecuencia absoluta con frecuencia porcentual:

> Existen unos 55 países que tienen una esperanza de vida de 75 años, unos 36 países que tienen una media de 80 años, unos 27 países que tienen una media de esperanza de vida de 70 años. Los demás países que son menos, tienen una esperanza de vida de menos de 65 años. (CM).
> Este gráfico nos da el porcentaje de países que tienen una edad determinada como esperanza de vida. Por ejemplo 12 de los países tienen una esperanza de vida de $\cong$ 50 años. El dato más interesante es que el 55 (más de la mitad de los países) tienen una esperanza de 75 años. (MIC).

Interpreta la moda. Cuando se interpreta el intervalo modal, o se indica cuál es la moda de la distribución a partir del histograma, así como lo hacen los dos casos siguientes; el primero dando la moda y el segundo el intervalo modal:

> Podemos interpretar que en la mayoría de los países la esperanza de vida es de 75 años. (ARM).
> […] que la moda está en el intervalo entre 70 y 75 y los siguientes intervalos más frecuentes son de 75 a 80 y de 65 a 70, (AJD).

Encontramos algunos errores en los futuros profesores al interpretar la frecuencia, de países como número de personas. Además, aunque la esperanza de vida modal es 75, puede no coincidir con la edad a la que muere la mayoría de personas, como indica

JFM, pudiendo deberse este error a la confusión entre la frecuencia absoluta y el valor de la variable. En menor medida encontramos confusión entre la moda y la media, como en JC, mostrando falta de comprensión del efecto de valores atípicos sobre la media:

> Lo que destaca del segundo gráfico es que la mayor frecuencia se da en los 75 años, en este baremo se encuentra la media de la esperanza de vida de esos países se encuentra de 72’5 y 82’5 años (JC).
> También indica que a los 75 años aproximadamente, muere mucha gente (JFM).

Interpreta el intervalo de menor frecuencia. Generalmente se hace en forma correcta como PP, aunque también encontramos interpretaciones incorrectas como por ejemplo IM, que confunde la frecuencia con el valor de la variable, conflicto señalado anteriormente:

> El menor número de países (11 países) tienen una esperanza de vida de 60 años (PP).
> Mientras que aproximadamente a los 60 años, la esperanza de vida decrece (IM).

Interpreta el rango. Algunos futuros profesores calculan en rango de la variable a partir de la gráfica, por tanto, estarían en nivel de lectura entre los datos (Curcio, 1989). Al igual que en el diagrama acumulativo, la interpretación suele ser correcta:

> No hay ningún país con la esperanza de vida inferior a 50, ni superior a 85. (EGA).

Interpreta el máximo. Algunos participantes interpretan adecuadamente el máximo de la variable a partir de la gráfica, como por ejemplo AJD, por lo que estarían en nivel de lectura entre los datos (Curcio, 1989). Encontramos también algunas interpretaciones en las que se suele manifestar el conflicto señalado anteriormente, en el que se confunde la esperanza de vida con la edad de muerte, como por ejemplo MJG.

> La esperanza de vida máxima es de 82’5 a 87’5 años, y esta se da solamente en 22 países. (AJD).
> También con todos estos gráficos observamos que en los diferentes países la población no supera los 85 años (MJG).

Interpreta la media. Algunos participantes interpretan la media de la variable a partir del histograma, llegando a conclusiones erróneas ya que, por lo general, se confunde este concepto con la mediana o la moda. No tienen en cuenta que al ser la distribución asimétrica negativa, la media debe ser menor que la moda; un ejemplo es el siguiente:

> Además podemos intuir que la media se encontrará entre 65 y 85 porque es donde se concentra más los datos (MC).

Interpreta los cuartiles y/o la mediana. Algunos participantes interpretan los cuartiles y/o la mediana en el histograma, generalmente apoyándose en los estadísticos de la tabla de frecuencias, pero señalando su presencia en el histograma, por ejemplo:

> La mayoría de los países tiene esperanza de vida entre 64’2 - 76’10 (1^{er}, 3^{er} cuartil) (entre los 67’5 - 82’5 años) (FR).

Interpreta la simetría de la distribución. Encontramos interpretaciones de la asimetría de la distribución en el histograma al observar que, el intervalo por debajo de la moda es más amplio que por encima de ella, como el siguiente ejemplo, donde además se relaciona con el diagrama de caja, para fundamentar esta interpretación.

> Es importante tener en cuenta que en el 2º [gráfico: histograma], podemos ver que la gran mayoría de los países se encuentra por encima de los 65 años, luego la mayoría se encuentra en un nivel alto y no está centrada, cosa que también podemos comprobar […]. (JG).

En la Tabla 5.6.3 se muestran los resultados, donde podemos observar que los futuros profesores interpretan este gráfico, y con un alto índice de respuestas correctas, que incluso se incrementa algo (72,2% al 74%) el segundo año. Generalmente interpretan el contenido o describen la moda, aunque también se observa otra variedad de interpretaciones, sobre todo en la segunda muestra. Asimismo, observamos errores al describir el contenido (mayores en la segunda muestra), al interpretar la moda (menores en la segunda) y al interpretar la media en la primera muestra.

Tabla 5.6.3. Interpretación del histograma y % respecto a categoría y muestra

	Muestra 1		Muestra 2		
Interpretaciones	Correcta	Incorrecta	Correcta	Incorrecta	Blanco
Contenido	12(80,0)	3(20,0)	17(65,4)	9(34,6)	
Moda	5(71,4)	2(28,6)	22(78,6)	6(21,4)	
Intervalo menor frecuencia	3(100)		1(25,0)	3(75,0)	
Rango	5(100)		6(100)		
Media		5(100)			
Simetría de la distribución	1(100)		2(100)		
Máximo			5(83,3)	1(16,7)	
Mínimo			3(100)		
Cuartiles o mediana			1(100)		
No interpreta					1(100)
Total interpretaciones	26(72,2)	10(27,8)	57(74)	19(24,7)	

En la Figura 5.6.3 se compara la distribución del número de interpretaciones correctas del histograma en las dos muestras, siendo en este caso ligeramente mayor el porcentaje de participantes que cometen algún error en la segunda. No obstante, vemos que el 83% en la primera muestra y el 79% en la segunda dan al menos una interpretación correcta del histograma, siendo lo más frecuente una en la primera muestra y dos en la segunda.

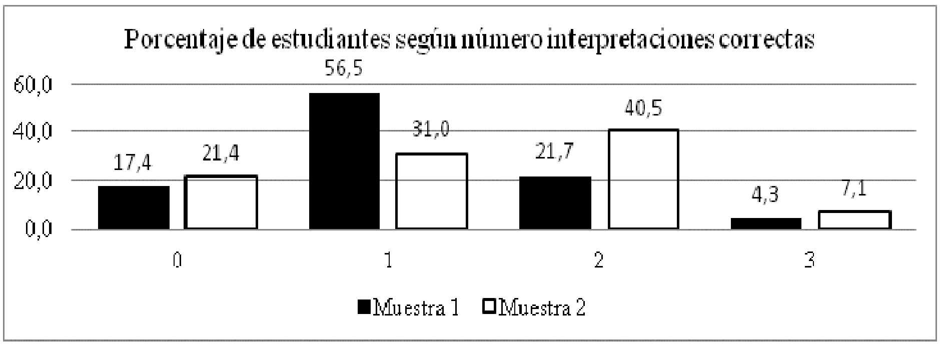


Figura 5.6.3. Interpretaciones correctas del histograma en las dos muestras

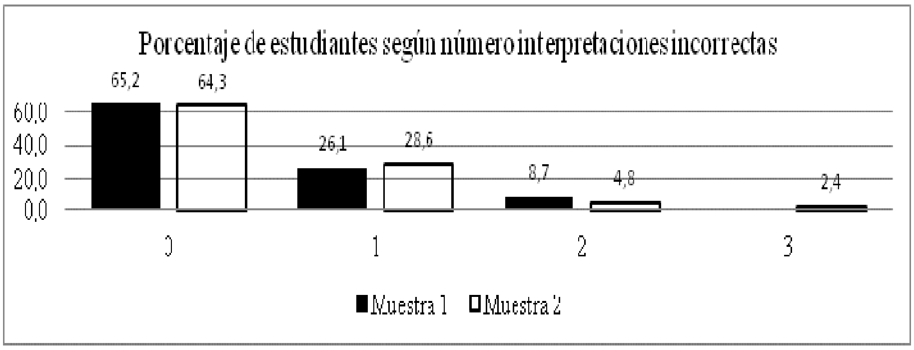


Figura 5.6.4. Interpretaciones incorrectas del histograma en las dos muestras

Al comparar las distribuciones del número de respuestas incorrectas en las dos muestras (Figura 5.6.4), vemos que son escasas. Alrededor del 65% en las dos muestras no comenten errores, y pocos comenten más de uno. Estos casos son debidos, principalmente, a cierta dificultad en la interpretación de la media o la moda a partir del histograma, ya que los confunden; además, en el caso de la moda, interpretan el intervalo de mayor frecuencia con la edad a la que muere más gente, algo similar ocurre al interpretar el intervalo de menor frecuencia. Los resultados son similares en las dos muestras, lo que unido al mayor número de interpretaciones correctas en la segunda, reflejan en ella resultados algo mejores

Interpretación del diagrama de caja

El diagrama de caja es el gráfico que menor número de futuros profesores han interpretado. A continuación describimos las categorías, similares a las descritas para los gráficos anteriores.

Describe el contenido del gráfico. Cuando se explica el significado del gráfico a través de sus componentes (cuartiles, máximo y mínimo, y puntos extremos). En la mayoría de los casos, los participantes indican cómo se organiza la información, completando a veces algunos datos particulares, como hace AMA. Destacamos que algunos futuros profesores, como JC, indican además, la ausencia de datos atípicos.

> En este gráfico se observan de forma aproximada (pues no se da en papel milimetrado) los siguientes datos sobre la esperanza de vida: Mínimo en: 47 años; Máximo en: 82 años; Mediana próxima a: 73 años; cuartiles respectivos a 64 años y 76 años. (AMA).
> No hay datos atípicos porque si no saldrían puntos sueltos (JC).

En esta categoría encontramos interpretaciones incorrectas, como las que se muestran a continuación; en la primera se considera que el diagrama representa tan sólo la información de la esperanza de vida de una cuarta parte de los países del estudio, mientras que en la segunda, el participante confunde media y mediana.

> Se refiere a la cuarta parte de todos los países que intervienen en la muestra (LT).
> […] Existe una dispersión en torno a los datos de la media. (PP).

Interpreta el máximo o el mínimo. Encontramos estas dos categorías similares a las ya descritas para los otros gráficos, como se ve en los siguientes ejemplos:

> […] la máxima esperanza de vida 83, (AJD).

El gráfico 3 nos dice que la mínima esperanza de vida es aproximadamente 47, (AJD).

Interpreta los cuartiles y/o la mediana. Encontramos algunas interpretaciones que muestran un adecuado nivel de lectura entre los datos (Curcio, 1989) al interpretar conjuntamente el primer y tercer cuartil, como se muestra en el siguiente ejemplo:

El 50% central de los países tienen una E.V. entre 69'2 y 76'10 años. (MAG).

Otras interpretaciones incorrectas de los cuartiles y la mediana se deben a que no se considera que los porcentajes entre cada cuartil son iguales. Este conflicto se suele manifestar porque los participantes quieren justificar la asimetría de la distribución, como por ejemplo AMC. En otros casos, se muestra un conflicto semiótico al confundir los conceptos media, moda y mediana, como por ejemplo MI.

En el gráfico 4, en el diagrama de caja, vemos como la mayoría de los países se encuentran por debajo de la mediana 72’5 años. (AMC).
En el 3º gráfico, la línea de la mediana nos indica dónde está la mayor cantidad de individuos del estudio, en este caso: esperanza de vida 72’5. (MI).

Interpreta la simetría de la distribución. Cuando se interpretan los valores extremos del diagrama, y su situación respecto a la caja. La mayoría de estas interpretaciones son correctas:

También en el 4º ya que la caja (entre el primer y tercer cuartil) se encuentra más cerca del máximo que del mínimo. También podemos ver esto con la mediana con respecto a los cuartiles, encontrándose más próximo al tercero que al primero. (JG).

Interpreta el rango. Al igual que en los gráficos anteriores, encontramos interpretaciones del rango, manifestando así un nivel avanzado de lectura entre los datos (Curcio, 1989). Mostramos dos ejemplos correctos; en el segundo, se calcula el rango intercuartílico:

El diagrama de caja esta acotado por arriba y alargado por abajo. Esto significa que la E.V. en el 50% de los países que más viven varía en un rango de valores pequeños. La E.V. en el 50% de los países que menos E.V. tienen varía en un rango alto de valores. (MAG).
Además, podemos calcular el rango intercuartílico: Q_3 - Q_1 = 11,9 (el 50% de la población)

Encontramos interpretaciones incorrectas como la siguiente, en la que se confunde el rango con el rango intercuartílico:

La gráfica 4 habla de una esperanza de vida de entre 63 y 77 años. (VC).

Interpreta la moda. Cuando se interpreta la esperanza de vida más frecuente entre los países del estudio, generalmente en comparación con la mediana. Encontramos un futuro profesor que interpreta este parámetro, de modo incorrecto, pues confunde el máximo con el hecho de que en la moda la frecuencia sea la mayor.

[Mientras que la mediana se sitúa en torno a 72’5 años] que coincide más o menos con el valor en el que los países obtienen mayor esperanza de vida. (IM).

Indica estadísticos en el diagrama sin interpretarlos. Estos participantes marcan los cuartiles, y en algunos casos los valores máximo y mínimo de la variable sobre el diagrama, pero no los interpretan. Los participantes reconocen los estadísticos que lo generan, pero no son capaces de interpretarlos de manera conjunta. Mostramos a continuación uno de estos casos, en los que el futuro profesor indica los estadísticos en el diagrama de caja, pero dice no conocer el gráfico.

El gráfico de la caja no lo reconozco (ATL).

Tabla 5.6.4. Interpretación del diagrama de caja y % respecto a categoría y muestra

	Muestra 1			Muestra 2		
Interpretación	Correcta	Incorrecta	Blanco	Correcta	Incorrecta	Blanco
Contenido	4(66,6)	2(33,3)		13(100)		
Máximo	3(100)			4(80)	1(20)	
Mínimo	3(100)			4(100)		
Cuartiles o mediana	4(100)			14(77,8)	5(22,2)	
Simetría	5(100)			1(50)	1(50)	
Rango	1(100)			7(87,5)	1(12,5)	
Indica estadísticos	8(100)			8(100)		
Moda					1(100)	
No interpreta			2(100)			9(100)
Total interpretaciones	28(87,5)	2(6,3)	2(6,3)	51(73,9)	9(13)	9(13)

En la Tabla 5.6.4 observamos la variedad de interpretaciones, con mayor porcentaje correctas que en los gráficos anteriores, aunque también es mayor el porcentaje de futuros profesores que no lo interpretan (sobre todo en la segunda muestra). Como se comentó anteriormente, los futuros profesores tienden a indicar los cuartiles en el diagrama de caja junto a los valores máximo y mínimo, sin interpretar los mismos a través de la variable que se estudia.

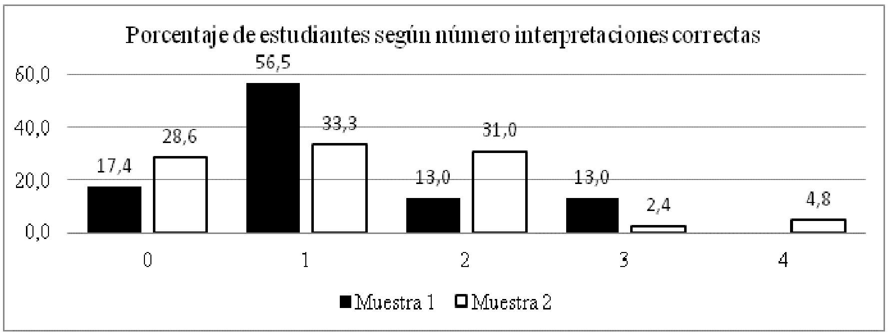


Figura 5.6.5. Interpretaciones correctas del diagrama de caja en las dos muestras

En la Figura 5.6.5 comparamos las distribuciones de interpretaciones correctas en este gráfico, donde mayor el porcentaje de incorrectas, y que no lo interpretan en la segunda muestra. En la primera, lo más frecuente es una interpretación correcta por estudiante, y en la segunda, se reparte entre una y dos interpretaciones. No obstante, apenas hay errores (Figura 5.6.6).

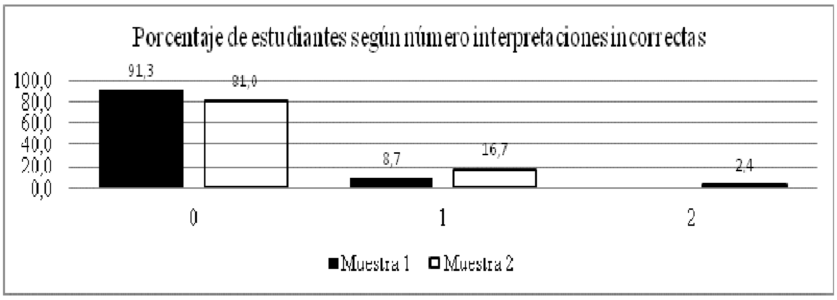


Figura 5.6.6. Interpretaciones correctas del diagrama de caja en las dos muestras

Tabla de resúmenes estadísticos

Todos los participantes interpretan la tabla de estadísticos, con gran diversidad de respuestas, atendiendo al estadístico que interpretan; aunque en algunos casos, los participantes se limitan a repetir el valor de los mismos, por lo que se incluye una nueva categoría: *Indica y/o describe los estadísticos pero no los interpreta*, para recoger este tipo de interpretaciones. La mayoría de los participantes valoran positivamente disponer del valor exacto de los estadísticos, ya que los gráficos tan sólo permiten obtener un valor estimado. A continuación describimos cada una de estas categorías.

Interpreta la moda. Cuando señalan el valor de la moda y realizan alguna interpretación; por ejemplo, indicando que es el valor más usual:

> Lo más repetido en cuanto a la esperanza de vida, es decir, la edad con mayor predominio es de 72,7 años. (MJG).

Encontramos una interpretación incorrecta en la que se confunde esperanza de vida con edad de la muerte, como ocurría en la interpretación de otros gráficos:

> Podemos saber también gracias a la moda que a los 75 años es cuando más gente se muere. (MC)

Interpretan la media. Cuando los participantes realizan alguna interpretación del valor medio de la variable; en la mayoría de los casos se limitan a indicar el valor medio. En otros, como el siguiente ejemplo, se alcanza una mayor complejidad, al conectar media, mediana y moda:

> Efectivamente la tabla indica que la media es sobre los 70 años, que casi coincide con el valor de la mitad de la distribución y con el valor más frecuente. (EGO).

Una interpretación incorrecta es la de MC, pues 69 no es la esperanza de vida sino la media de la esperanza de vida.

> Aportan exactitud, viendo así que la esperanza de vida es 69 años. (MC).

Interpreta la desviación típica. Son los participantes que interpretan el valor de la desviación típica, indicando que es una medida de dispersión, en relación con la distribución de valores alrededor de la media:

> La variación a la media es en torno a 9,88 años. (CM).
> […] la desviación típica es bastante alta, casi 10, y esto quiere decir que los datos no están nada centralizados. Es una pena, quiere decir que hay mucha diferencia entre unos países y otros. (AU).

Encontramos algunas interpretaciones incorrectas de la desviación típica, como mostramos en el siguiente ejemplo, donde se considera el valor 9,88 como alto, y se asume por ello que los datos estarían poco concentrados respecto a la media. Esto no es cierto, y el uso de la palabra centrado puede suponer que indica que la distribución no es simétrica debido al valor de la desviación típica, por tanto, un conflicto en la interpretación de la simetría.

> Que la desviación típica sea 9,88 quiere decir que los datos no están muy centrados en torno a la media que es 69,17 (ChC).

Interpreta los cuartiles o la mediana. Un ejemplo es E.V., que interpreta el significado del cuartil en su doble enfoque, es decir, tanto a los países que tienen una esperanza de vida menor que dicho cuartil como mayor:

> La mitad de los países tiene una esperanza de vida de 72'5 años o menos. La otra mitad tiene una E.V. de 72'5 años o más. El 25% de los países tiene una E.V. de 64'20 años o menos, el 75% de los países tiene una E.V. de 64'2 años o más. El 75% de los países tiene una E.V. de 76'10 años o menos, el 25% de los países tiene una E.V. de 76'10 años o más. (MAG).

Las dos siguientes interpretaciones son incorrectas, por la poca precisión en el uso del lenguaje, ya que, en la primera de ellas se considera que todos los países dentro del rango del percentil poseen la misma esperanza de vida, en lugar de indicar que su esperanza de vida sería menor a dicho valor. En la segunda, la moda sí se halla en el intervalo de mayor frecuencia pero no tiene por qué encontrarse la mediana en dicho intervalo, por lo que hay un conflicto al aplicar a la mediana una propiedad de la moda:

> El 25% de los países posee una esperanza de vida de 64'20 años y el 75% posee 76'10 años. (MM).
> Tiene sentido que la mediana y la moda se hallen en el intervalo que más frecuencias tiene (MRA).

Interpreta el mínimo. La mayoría de interpretaciones se reducen a explicar que en los países estudiados, la menor esperanza de vida es de 46,90 años. Aunque parezca trivial, encontramos un participante que muestra un razonamiento incorrecto:

> La persona que antes se muere es a los 46,90 años. (CM).

Interpreta el máximo. Al igual que ocurría con las interpretaciones del mínimo, los participantes, por lo general, se limitan a interpretar el máximo contextualizando el dato de la tabla, con la variable de estudio, encontrando una interpretación errónea al confundir el valor de la variable con la edad de la muerte:

> […] y la última que se muere es a los 83 años. (CM).

Interpreta el rango. Tan sólo un participante interpreta este estadístico, ofreciendo el intervalo a partir de los valores máximo y mínimo de la tabla:

> […] y que se mueven en el intervalo de [46'90,83'00]. (MRA).

Indica y/o describe los estadísticos pero no los interpreta. Cuando el participante se limita a repetir la información de la tabla sin contextualizar los datos. Se incluyen en esta categoría aquellas descripciones del significado de algunos estadísticos que aparecen en la tabla, que aunque correctas, no integran el significado de la variable que se analiza, y por tanto, están descontextualizadas.

> Media → Punto de equilibrio; Mediana → Percentil 50; Moda → Valor que más se repite; Mínimo → Valor más pequeño; Máximo → Valor más grande; Primer cuartil → Percentil 25; Tercer cuartil → Percentil 75; Desv. Típica → Dispersión con respecto a la media. (ME).

Un participante interpreta incorrectamente la información de la tabla, ya que los datos que se exponen no son discretos, lo que manifiesta un conflicto semiótico entre los conceptos discreto y continuo.

> La tabla aporta parámetros discretos calculados a partir de los datos de la serie. (ATL).

En la Tabla 5.6.5 se resumen los resultados de las interpretaciones de los estadísticos de la tabla, donde encontramos que la mayoría responden a la tarea, con alta proporción de respuesta correcta en cada categoría, y en las dos muestras. La proporción de respuestas correctas en este caso es mayor en la primera muestra.

Tabla 5.6.5. Interpretación de estadísticos y % respecto a categoría y muestra

	Muestra 1		Muestra 2		
	Correcta	Incorrecta	Correcta	Incorrecta	Blanco
Interpreta moda	5(83,3)	1(16,7)	7(87,5)	1(12,5)	
Interpreta media	8(88,9)	1(11,1)	8(80)	2(20)	
Interpreta desviación típica	3(60)	2(40)	3(60)	2(40)	
Interpreta cuartiles y/o mediana	3(60)	2(40)	14(93,3)	1(6,7)	
Interpreta mínimo	4(80)	1(20)	9(100)		
Interpreta máximo	4(80)	1(20)	9(90)	1(10)	
Interpreta rango	1(100)		4(100)		
Indica estadísticos sin interpretar	13(92,9)	1 (7,1)	6(100)		
No interpreta					18(100)
Total	41(82)	9(18)	60(70,6)	7(8,2)	18(21,2)

Lo más frecuente en la primera muestra fue dar el valor de los estadísticos de la tabla sin interpretarlos desde el contexto que se estudia, seguido de la interpretación de la media y moda. Por el contrario, en la segunda muestra hay mayor interpretación de los cuartiles y mediana, el mínimo o el máximo.

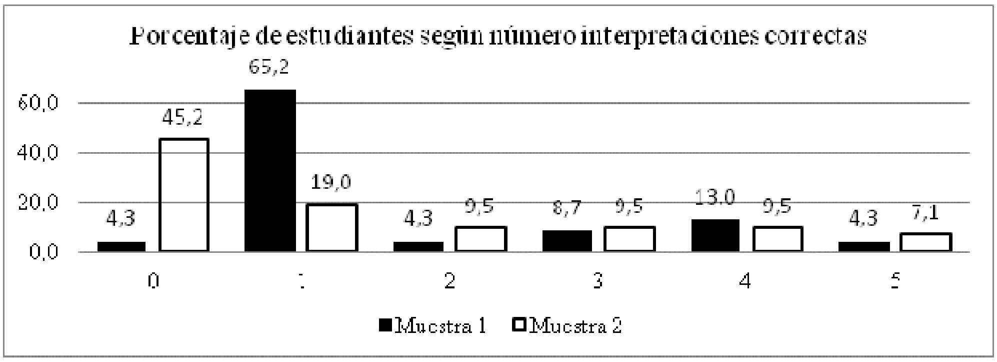


Figura 5.6.7. Interpretaciones de estadísticos en las dos muestras

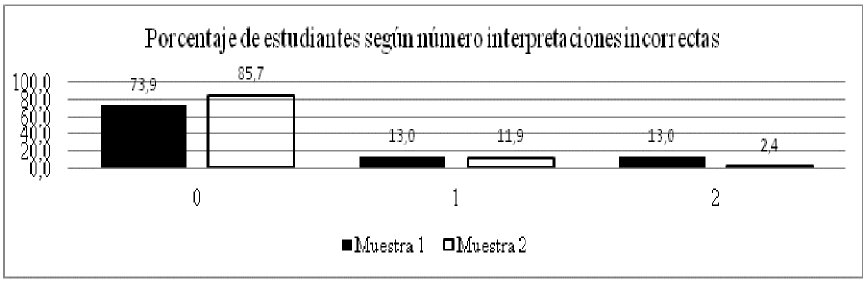


Figura 5.6.8. Interpretaciones incorrectas de estadísticos en las dos muestras

En la Figura 5.6.7 comparamos las distribuciones del número de respuestas correctas, que es mayor en la primera muestra pues en la segunda un gran porcentaje no añade nuevas interpretaciones de los estadísticos, ya utilizados en las gráficas; lo que se explica, pues ya los habían comentado. Hay que hacer notar, que en la primera muestra muchos se limitan a dar el valor de los estadísticos sin añadir interpretación de los mismos. El limitarse a repetir el valor del estadístico sin interpretar, da indicios de una posible dificultad de los futuros profesores en cuanto a interpretar los datos en un contexto. Este resultado es preocupante, ya que es deseable que un futuro profesor sea capaz de contextualizar la información resumida que le ofrece un dato particular, mediante el uso de diferentes representaciones de la información.

En todo caso, y al igual que en el gráfico de la caja, en los que interpretan, apenas hay errores (Figura 5.6.8).

Síntesis de resultados en la interpretación de tablas y gráficos

Para comparar los resultados en los cuatro apartados de la Tarea 1, en la Tabla 5.6.6 se presentan estadísticos resumen del número de respuestas correctas por alumno en cada una de las tareas relacionadas con conocimientos previos. Observamos que en promedio, tenemos más de una respuesta correcta por alumno, con resultados muy similares, algo mejores en la segunda muestra en la interpretación de gráficos y en la primera en la interpretación de la tabla; es en esta última parte donde hay más dispersión de resultados, como comentamos anteriormente.

Tabla 5.6.6. Estadístico del número de interpretaciones correctas por alumno en cada muestra

Interpretaciones correctas	Muestra	Media	D. típica.
Polígono acumulativo	1	1,22	0,85
	2	1,24	1,12
Histograma	1	1,13	0,76
	2	1,33	0,90
Diagrama de Caja	1	1,22	0,90
	2	1,21	1,05
Tabla	1	1,78	1,48
	2	1,40	1,68

Respecto a los errores de interpretación (Tabla 5.6.7) los resultados son más variados. En primer lugar se observa que es menor el número de errores que de interpretaciones correctas en cualquiera de las tareas y muestra. Además, apenas se observan errores en la interpretación del diagrama de caja y tabla, encontrando diferencias en las muestras; el diagrama de caja posee mayores interpretaciones incorrectas en la segunda muestra, y la tabla mayor error en la interpretación en la primera muestra, pero esto es de esperar, ya que más participantes dejaron sin comentar la tabla en la segunda muestra. La mayor proporción aparecen en el diagrama acumulativo e histograma, a pesar de que estos gráficos son bien conocidos por los participantes.

Tabla 5.6.7. Estadístico del número de interpretaciones incorrectas por muestra

Interpretaciones incorrectas	Muestra	Media	Desviación típ.
Polígono acumulativo	1	0,57	0,59
	2	0,48	0,63
Histograma	1	0,43	0,66
	2	0,45	0,71
Diagrama de Caja	1	0,09	0,29
	2	0,21	0,47
Tabla	1	0,39	0,72
	2	0,17	0,44

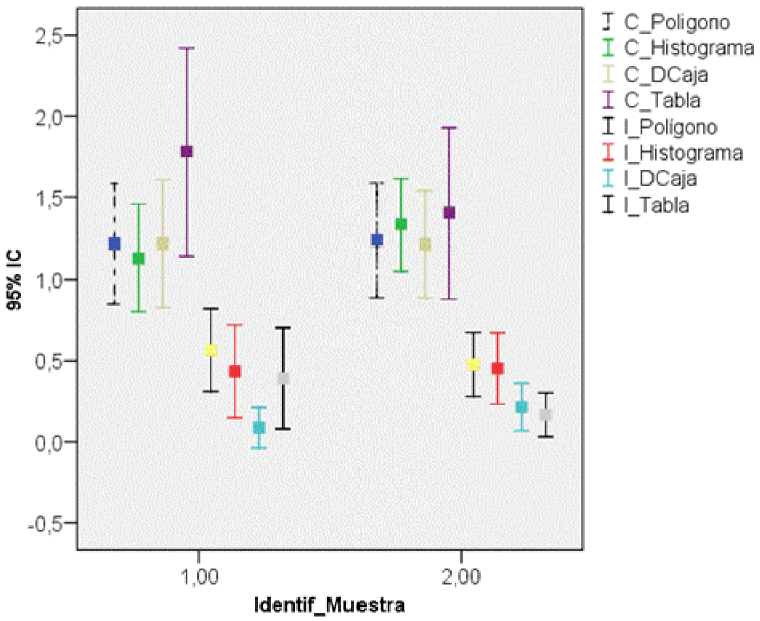


Figura 5.6.9. Intervalos de confianza de las respuestas correctas e incorrectas en las dos muestras

En la Figura 5.6.9 presentamos los intervalos de confianza del número de interpretaciones correctas e incorrectas en cada muestra. Observamos que en las dos hay una diferencia estadísticamente significativa entre el número de interpretaciones correctas e incorrectas. Por otro lado, no se observan diferencias significativas entre las dos muestras ni para el número medio de correctas o de incorrectas en la interpretación de ninguno de los gráficos o la tabla.

No obstante, al comparar el número de interpretaciones correctas, se observa una mayor media y variabilidad (aunque no estadísticamente significativa) en la interpretación de la tabla, en la primera muestra, y también en el número de errores de interpretación.

Podemos decir, en todo caso, que nuestros resultados son mejores que los obtenidos por Arteaga (2011) en cuanto a la lectura de gráficos, aun cuando nosotros hemos considerado como incorrectas aquellas producciones en las que los participantes utilizan un lenguaje impreciso. En dicha investigación, fue muy pequeño el porcentaje de participantes que llega a interpretar el gráfico, mientras que, como hemos visto en nuestro caso, la mayoría los interpreta en contexto y de un modo correcto.

5.6.2. ELECCIÓN DE UN PROMEDIO REPRESENTATIVO

Para evaluar el conocimiento de los futuros profesores sobre algunos estadísticos, se plantearon dos tareas adicionales. La primera propone la elección de un estadístico de centralización que represente, lo mejor posible, a la variable Esperanza de vida.

5.6.2.1. ANÁLISIS A PRIORI Y DESARROLLO

2. Teniendo en cuenta la tabla anterior y el gráfico de caja ¿Cuál de los promedios (media, mediana y moda) representaría mejor esta variable? ¿Por qué?

El formador de profesores presentó esta actividad, indicando que su fin era promover la reflexión sobre las diferentes medidas estadísticas que pueden usarse en un estudio, y sobre el significado del término “representativo”. Indicó que a menudo se necesita trabajar con valores aproximados, como medio de combinación de análisis parciales, para obtenerse con ello índices globales. También aclaró que los valores de la Esperanza de vida son datos agregados (pues son medias ponderadas, no valores de individuos), lo cual tiene repercusión en la interpretación de estadísticos, como puede ser los promedios de dicha variable.

Finaliza la presentación sugiriendo que esta actividad podría utilizarse para dar significado a la noción de media ponderada, concepto difícil para los estudiantes (Batanero, 2000). Igualmente recuerda que algunos estudiantes podrían tener dificultad en comprender el significado de la mediana, respecto a la cual se han descrito numerosos conflictos semióticos (Mayén, Díaz y Batanero, 2009). Para ayudar a visualizar su significado, se incluye el gráfico de caja, que permite visualizar la dispersión, además de valorar la situación de los cuartiles respecto de la mediana.

Solución experta y configuración de objetos matemáticos

En el debate se llega al consenso de que, en este ejemplo, media, mediana y moda tienen un valor muy similar, pero la media es inferior a la mediana o moda, por lo que sería preferible usar la mediana o moda para representar los datos. Esta elección también se puede justificar por la asimetría de la variable, pues la media es un estadístico no resistente a valores extremos o falta de simetría (Batanero, 2000).

De los objetos matemáticos implicados en la resolución de la tarea (Tabla 5.6.8), podemos advertir el protagonismo de los conceptos de estadístico de centralización y posición, media, mediana y moda. Se pretende hacer emerger las propiedades de dichas medidas de centralización, como que la media es sensible a los valores atípicos, mientras que la mediana es un estadístico robusto (PPI3). Se destaca la argumentación gráfica y deductiva, y el uso de ejemplos.

Tabla 5.6.8. Objetos matemáticos implicados en la Tarea 2

Situación problema	P0.Organización/Reducción de datos
Lenguaje	L1.Verbal (escrito y oral)
	L2.Simbólico: notación de estadísticos de centralización y dispersión
	L4.Gráfico: gráfico de caja
	L5.Numérico
Conceptos	CI2.Distribución de una variable estadística
	CI3.Medidas de tendencia central y posición
Propiedades	PPI3.La media es un estadístico sensible a valores extremos y asimetría.
	PPI3. La mediana es un estadístico “robusto”.
	PPI3. Al calcular la media intervienen todos los valores y en la mediana no
	PPI3. En el cálculo de la mediana es importante el orden de los datos.
Argumentos	A1. Análisis de ejemplos/ contraejemplos
	A2. Uso de Gráfico
	A5. Verbal – deductivo
Procedimientos	PC3.Traducción entre representaciones
	PCI1.Lectura de gráficos
	PCI1.Localización de estadísticos en gráficos

5.6.2.2.RESULTADOS Y DISCUSIÓN

Los resultados en la Tarea 2 se presentan a continuación, y se clasifican atendiendo a la elección del promedio más representativo, y su justificación. Para ambos apartados se ha valorado la corrección de los resultados, y la justificación aportada.

Elección de la media. Esta elección es incorrecta, dada la asimetría de la distribución. Entre las justificaciones de los participantes, sólo la de CM es correcta. En las incorrectas encontramos confusión de media y mediana, o no comprender el rango de la variable, como la siguiente:

> El dato más representativo es la media porque es el que engloba a todos los países en general (CM).
> La media es el valor más representativo porque dista del valor máximo y mínimo 9'88 y la dispersión de los valores máximo y mínimo no es muy grande. (JMV).

Elección de la moda. Algunos futuros profesores eligen la moda en esta tarea. La mayoría de las justificaciones son correctas; por ejemplo, la de DG. Encontramos dos incorrectas, como PG, quien confunde la edad de muerte con el valor de la variable, conflicto descrito en la sección anterior.

> Creo que la moda porque es más representativo de la muestra → 55 países con 75 años de esperanza de vida. (DG).
> El promedio que mejor representa la esperanza de vida es la moda que es el valor más frecuente de edad en el que se producen más fallecimientos. (PG)

Elección de la mediana. Todos los futuros profesores que optan por esta elección la justifican correctamente. Suelen indicar que la media se ve afectada por la asimetría de la distribución (EM), o que es útil pues separa los países en dos mitades iguales. Las pocas justificaciones incorrectas se deben a confusión entre media y mediana, como AV.

> La mediana porque la media está demasiado afectada por los valores extremos. (EM).
> La mediana porque hace una idea general de todos los países, aunque existan excepciones (AV)

Elección de la mediana y la moda como promedio más representativo. Cuando el sujeto indica que ambos estadísticos son válidos como representantes de la esperanza de vida. Muchos optan por estos dos estadísticos, sin justificar su elección, otros aportan justificaciones correctas, como las siguientes:

> La media es un mal representante debido a la asimetría de la distribución. Yo tomaría la mediana. Aprovechando que la distribución es unimodal y que hay una gran concentración de datos en torno a la moda, también sería un buen representante de la distribución. (MAG).
> O la mediana o la moda ya que la media es peor porque hace que se compensen los datos (en este caso la media nos hace ver que se vive menos de lo que realmente se vive) la mediana o moda son más representativas. (MVC).

Elección de la media y la moda. Cuando se eligen estos dos estadísticos; se indica que ambos estadísticos se parecen; y que la mediana no tiene en cuenta el valor de los datos, que es incorrecto pues el valor influye en el orden de los mismos (GS). Sólo la justificación de AU es correcta, aunque la elección no lo sea

> Pienso que serían más acertados tanto la media como la moda, no así la mediana, ya que esta se limita a separar observaciones sin tener en cuenta los valores. Porque son los valores que nos marcan los porcentajes que más se presentan. (GS).
> Tengo razones para decir que la que mejor lo representa es la media o la moda. Con la moda sabré cual es la esperanza de vida más común y así podré saber cuál es la esperanza más normal. No obstante no sabré qué pasa con "todo" el mundo. Con la media tendré un poco más de información, si la esperanza media es 69, me da más información que la moda, aunque con esta no sabré cual esperanza será más común. Me quedo con las dos. (AU).

No se elige ningún promedio como más representativo. Algunos participantes no eligen ningún promedio, aunque sí responden a la tarea con justificaciones incorrectas. Por ejemplo AJD indica que un único dato es insuficiente para representar a la variable:

> Yo creo que ninguno es representativo por sí solo, habría que darlos junto más medidas. (AJD).

Tabla 5.6.9. Frecuencia y % de respuestas y justificación en cada muestra

		Muestra 1			Muestra 2		
		Tipo de justificación		No	Tipo de justificación		No
Tipo de elección		Correcta	Incorrecta	justifica	Correcta	Incorrecta	justifica
Correcta	Moda	1(100)			5(71,4)	2(28,6)	
	Mediana	4(100)			6(66,7)	2(22,2)	1(11,1)
	Mediana y moda	3(33,3)	2(22,2)	4(44,4)	10(52,6)	7(36,8)	2(10,5)
Incorrecta	Media	1(33,3)	2(66,7)		1(50)	1(50)	
	Media y moda				1(33,3)	2(66,7)	
	Responde y no elige		4(100)			2(100)	
No responde				2(100)			
Total		9(39,1)	8(34,8)	6(26,1)	23(54,8)	16(38,1)	3(7,1)

En la Tabla 5.6.9 observamos un gran porcentaje de respuestas correctas (70% en la primera muestra que sube al 83% en la segunda). Respecto a las justificaciones correctas, el porcentaje es mucho menor (39,1% en la primera muestra y 54,8 en la segunda). Valoramos positivamente los casos en que se da una respuesta incorrecta pero se justifica correctamente, ya que da indicios de que los futuros profesores conocen las propiedades de estos conceptos, aunque muestren dificultad en la elección de uno u otro, como muestra el siguiente ejemplo:

> El dato más representativo es la media porque es el que engloba a todos los países en general. Aunque la moda también es importante (CM)

Por tanto, a pesar del alto índice de respuestas correctas, un grupo de participantes no comprenden significativamente los estadísticos de centralización, que deben enseñar en su futura profesión docente. Encontramos también dos futuros profesores que no resuelven la tarea, uno de ellos manifiesta que no sabe leer el diagrama de caja, y el otro que no recuerda el significado de la mediana y la moda, ambos en la primera muestra.

5.6.3. INTERPRETACIÓN DE PERCENTILES

La tercera tarea que planteamos para analizar el conocimiento previo de los futuros profesores, versa sobre los estadísticos de orden y su localización en gráficos.

5.6.3.1. ANÁLISIS A PRIORI Y DESARROLLO

El formador presentó la Tarea 3, indicando que se trata de evaluar la habilidad en la lectura de percentiles desde el diagrama acumulativo.

> 3. España tuvo en 2009 una esperanza de vida al nacer de 80,9. ¿En qué percentil se sitúa*?*

Solución experta y objetos matemáticos

En el debate posterior a las soluciones escritas, el formador recuerda que el diagrama acumulativo de frecuencias, que ya se había recordado en la primera actividad, permite visualizar el percentil en que se sitúa España en cuanto al valor 80,9 años de esperanza de vida. Leyendo en la gráfica el eje *Y*, el valor 80, muy aproximado al dado, se ve que corresponde a un porcentaje acumulado inferior a 90, y aproximadamente igual a 88%; por tanto, el percentil correspondiente es el de 88%, aproximadamente. Algunos participantes indican que también se podría aproximar dicho percentil mediante la tabla de frecuencias porcentuales acumuladas adjunta al pie de dicho diagrama. Si localizamos en la tabla el valor del intervalo (77,5-82,5) cuyo centro es 80 lleva asociado un porcentaje acumulado de 88.

El formador compara los dos métodos, indicado que cada uno tiene su dificultad; y en todos ellos interviene la proporcionalidad. En el caso de la lectura del gráfico, cabe decir que la información no está directamente disponible, y su búsqueda implica un nivel de lectura de “interpolación” o “leer más allá de los datos”, según Curcio (1989).

En la Tabla 5.6.10 presentamos la configuración de objetos matemáticos implicados en el desarrollo de la actividad. Como se puede observar, se introduce la notación de percentil, así como este concepto, el de rango de percentil y sus propiedades; destacan los procedimientos, en torno al objeto matemático de estadísticos de posición, que cobran gran protagonismo en la tarea, así como la interpolación y lectura más allá de los datos. También se advierte una riqueza del lenguaje en la tarea.

Tabla 5.6.10. Objetos matemáticos implicados en la Tarea 3

La situación problema	P0. Organización/Reducción de datos
Lenguaje	L1.Verbal (escrito y oral) L2.Simbólico: porcentaje, y notación de percentil L3.Tabular L4.Gráfico L5.Numérico
Conceptos	CI1.Variable estadística, valores, rango, intervalos, extremos, marcas de clase CI2.Distribución de una variable estadística CI3.Medidas de tendencia central y posición: percentiles, orden, porcentajes
Propiedades	PPI3.El percentil del r% deja por debajo el r% ciento de los datos.
Argumentos	A2. Uso de gráfico A5. Verbal – deductivo
Procedimientos	PC3.Traducción entre representaciones: gráfica, tabular y numérica PC7.Uso de tecnología: hoja Excel PCI1.Lectura de gráficos: localización de percentil en el diagrama de frecuencias acumuladas, interpolación de datos PCI2.Lectura de tablas: tabla de porcentajes acumuladas

5.6.3.2. RESULTADOS Y DISCUSIÓN

Las respuestas a la tarea fueron variadas, pudiendo establecerse varios grupos. Algunos futuros profesores establecen un intervalo de valores como respuesta, mientras que otros ofrecen un valor determinado como percentil a la edad que se pedía.

Para evaluar la corrección de las mismas, partimos del hecho de que no se disponía de otros de medios para determinar exactamente el percentil solicitado que la gráfica, por lo que no se esperaba un valor concreto exacto. En este sentido, y al igual que en el trabajo de Arteaga (2011), el hecho de confundir los elementos del gráfico o no comprender correctamente algunos conceptos hace que los estudiantes cometan errores o imprecisiones.

Consideramos correctas las respuestas que dan un valor del percentil entre P_{80} y P_{95} e incorrectas el resto; en estos casos el futuro profesor no establece una relación precisa entre el eje Y y el valor 80,9 de la esperanza de vida desde el eje X o una aproximación al mismo.

Dos participantes dieron la respuesta (P_{88}, P_{100}), que se han considerado correctas por las justificaciones que aportan. Sus explicaciones evidencian que han establecido una relación precisa del eje Y con el valor de la esperanza de vida de España, aunque han dejado un margen muy amplio en su respuesta:

Entre el percentil 88 y el percentil 100, más cercano al percentil 88. (MRA).

En la Tabla 5.6.11 se muestran las respuestas de los futuros profesores a la tarea, donde podemos observar el alto índice de respuestas correctas (83%). El resto de las respuestas pueden considerarse parcialmente correctas, principalmente las que incluyen valores mayores que P_{95}, pues la pregunta no exigía un valor concreto para el percentil solicitado. De este modo, los resultados de esta tarea podrían considerarse satisfactorios.

Tabla 5.6.11. Cálculo de un percentil a partir del diagrama acumulativo

Tipo de respuesta	Muestra 1	Muestra 2	Total
Correcta(P_{80} , P_{95})	16(24,6)	38(58,5)	54(83,1)
Incluyen valores mayores que P_{95}	6(9,2)	1(1,5)	7(10,8)
Incluyen valores menores que P_{80}		3(4,6)	3(4,6)
No responde	1(1,5)		1(1,5)
Total	23(35,4)	42(64,6)	65(100)

5.7. ACTIVIDADES DE EVALUACIÓN DEL CONOCIMIENTO SOBRE CORRELACIÓN Y REGRESIÓN

En esta sección se analizan las cuatro tareas específicamente diseñadas para el estudio de la correlación y regresión, que permitirán recoger datos sobre el conocimiento matemático común de los futuros profesores (Ver capítulo 2). Al presentar estas actividades, el formador de profesores les indica que su secuenciación, dentro de la clase de Bachillerato, favorece el aprendizaje significativo de estas nociones. También les informa, que en el trabajo con futuros profesores servirá para recordar el tema, a la vez que se propone una metodología alternativa (de innovación docente) para el trabajo en el aula.

En primer lugar, por razón de formato, se presentan a los participantes simplemente los diagramas de dispersión, sin plantear ninguna pregunta, aunque al presentar la actividad, el formador de profesores pide a los participantes que analicen las gráficas que se presentan. Se elige el diagrama de dispersión, que permite visualizar características de dicha dependencia como es la intensidad, el sentido, y establecer a simple vista un modelo funcional de mejor ajuste a la distribución que se representa. Todos los participantes indicaron estar familiarizados con este gráfico.

Se presentan una amplia tipología de diagramas de dispersión, y se utiliza una escala del eje vertical común, de 40 a 90 años en la variable dependiente (Esperanza de vida) en todos los gráficos, para visualizar mejor los datos. Además, se ha modificado el eje horizontal para tener en cuenta, en cada variable, su rango de variación. Así, por ejemplo, el diagrama de la Población total no permitía visualizar fácilmente los datos, dado el valor máximo de esta variable (China con 1334908,8 habitantes), y aparecen aquellos países cuya población es inferior a los 50 millones de habitantes, con lo que el diagrama de dispersión visualiza a 171 de los 194 países que se analizan (90% de los datos aproximadamente).

Variables que afectan a la Esperanza de vida. En las gráficas siguientes se representan los diagramas de dispersión de la *Esperanza de vida al nacer* en función de algunas variables incluidas en el fichero.

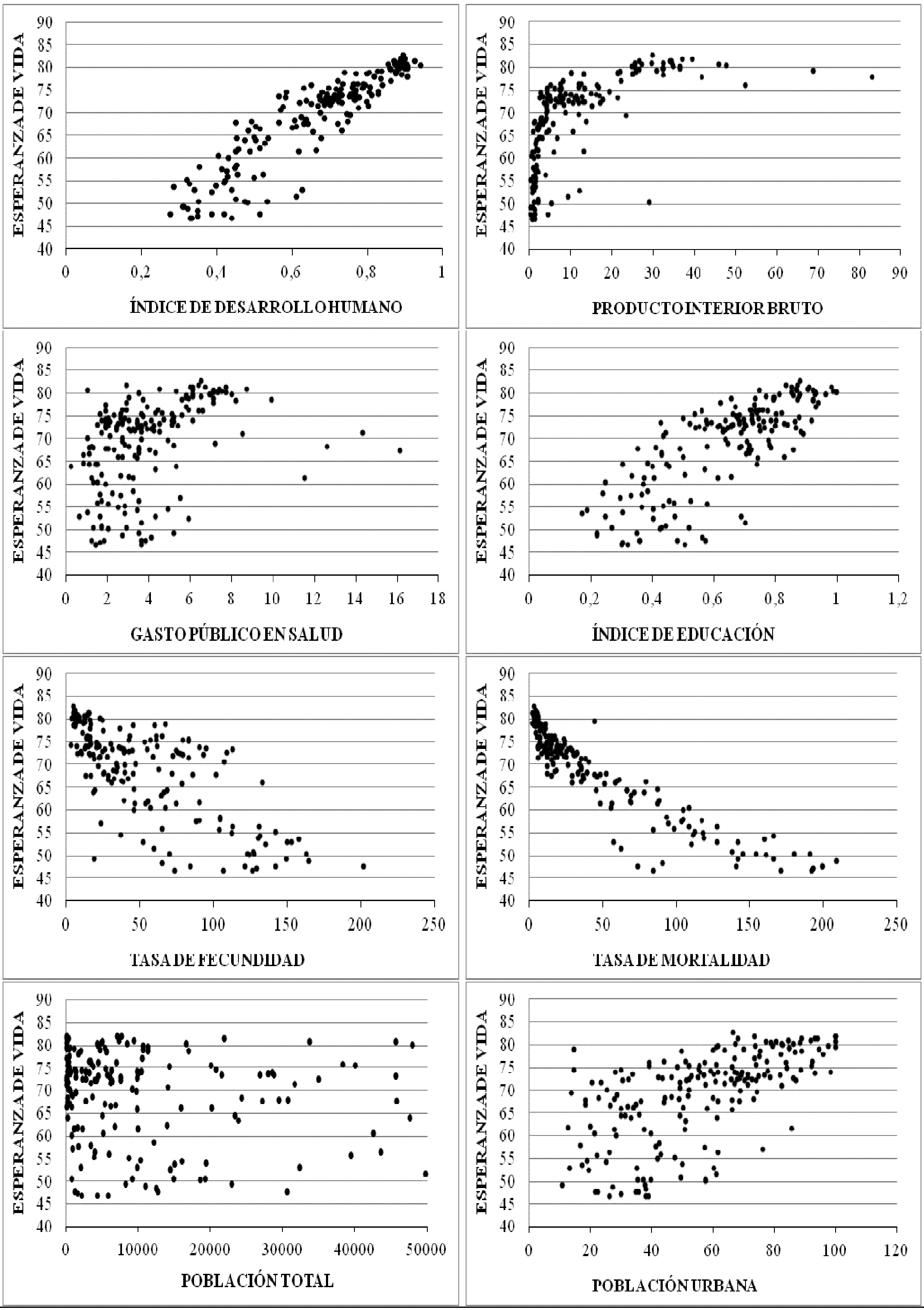


Tabla 5.7.1. Coeficiente de correlación, intensidad y signo en los diagramas de dispersión presentados

Variable	Coeficiente de correlación	Intensidad	Signo
Índice de Desarrollo Humano	0,91	Alta	Positivo
PIB per cápita	0,61	Moderada	Positivo
Tasa de fecundidad adolescentes	-0,73	Moderada-alta	Negativo
Tasa de mortalidad de niños	-0,92	Alta	Negativo
Gasto público en salud	0,38	Baja	Positivo
Índice de educación	0,78	Moderada-alta	Positivo
Población, total	0,00	Independencia	No tiene
Población urbana	0,62	Moderada	Positivo

Tabla 5.7.2. Modelo de ajuste, explicación de la relación y teorías previas en los diagramas de dispersión

Variable	Modelo de ajuste	Explicación de la relación	Teorías previas
Índice de Desarrollo Humano	Lineal	Interdependencia	Coincide
PIB per cápita	Logarítmica	Dependencia Indirecta	Coincide
Tasa de fecundidad adolescentes	Lineal-Exponencial	Dependencia Indirecta	No hay teoría
Tasa de mortalidad de niños	Exponencial	Causal unilateral	Coincide
Gasto público en salud	Polinómica	Causal unilateral	Coincide*
Índice de educación	Lineal	Dependencia Indirecta	Coincide
Población, total	Independencia	Independencia	No hay teoría
Población urbana	Lineal	Dependencia Indirecta	En contra

*Coincide, pero es menor a la correlación esperada

Se ha elegido variedad de relaciones de dependencia de la Esperanza de vida en los ocho diagramas que se presentan, teniendo en cuenta las variables de tarea descritas por Estepa (1994), Sánchez Cobo (1999) y Gea, et al, (2013) (Tablas 5.7.1 y 5.7.2):

- *V1*: *Signo de la correlación*: Se incluyen correlaciones negativas (Tasa de mortalidad y Tasa de fecundidad), un caso de independencia, y el resto correlaciones positivas. Se toma esta variable porque, tanto Estepa (1994) como Sánchez Cobo (1999), indican que los estudiantes perciben mejor las correlaciones positivas.

- *V2*: *Intensidad de la correlación,* que podemos deducir del valor absoluto del coeficiente de correlación. Destacamos dos diagramas de dispersión de alta intensidad: el Índice de desarrollo humano y la Tasa de mortalidad. Se incluye un caso de independencia (Población total), una correlación baja (Gasto público en salud) y el resto son moderada o moderadas altas. Los valores de los coeficiente de correlación han sido calculados con SPSS, y salvo para la variable Población total (con coeficiente de correlación 0,004) que se sustituyó por el valor 0, todos ellos son significativos al nivel 0,01 (bilateral).

- *V3*: El *modelo de ajuste* a los datos del diagrama de dispersión. En unos casos el modelo lineal proporciona un buen ajuste (Índice de desarrollo humano, Índice de educación, Tasa de fecundidad y Población urbana) mientras que en otros no tanto, como el PIB (modelo logarítmico) y la Tasa de mortalidad (modelo exponencial).

- *V4. Explicación de la posible correlación en los datos.* Se han tenido en cuenta algunas de las categorías de posible relación consideradas por Barbancho (1973): dependencia causal unilateral, interdependencia, tercera variable o dependencia indirecta, y covariación casual.

- *V5: Relación entre teorías previas y datos:* Si la relación entre las variables es la esperada por los participantes, si es contraria, o no hay teorías al respecto. En el primer caso, se percibe más correctamente la correlación; en caso de teorías en contra de la relación, los participantes se ven influenciados por las mismas.

Sobre estos diagramas de dispersión, se plantearon una serie de preguntas que se describen a continuación.

5.7.1. ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DEL DIAGRAMA DE DISPERSIÓN

Con la Tarea 5 se pretende evaluar la capacidad para estimar una correlación a partir del diagrama de dispersión, deducir su signo, y ver la consistencia entre estimación y asignación de un coeficiente de correlación, entre varios dados. Se proporcionó a los futuros profesores 12 coeficientes de correlación diferentes, que incluye los observados entre las variables, los valores 0 y 1 (independencia y correlación lineal perfecta) y dos más, uno de ellos más bajo que los observados (0,2) y otro intermedio (0,5). Se ofrecen más valores para el coeficiente de correlación que diagramas de dispersión.

5. *Relación de cada variable con la Esperanza de vida (continuación).*

a. Asigna una puntuación entre 0 (si no hay relación) y 1 (máxima intensidad de la relación) según la intensidad de esta relación. Asigna también un signo (+ o -) según la relación sea directa o inversa.

Variable	Intensidad de la relación (entre 0 y 1)	Relación directa o inversa (signo)	Coeficiente de correlación
Índice de Desarrollo Humano			
PIB per cápita			
Tasa de fecundidad adolescentes			
Tasa de mortalidad de niños			
Gasto público en salud			
Índice de educación			
Población, total			
Población urbana			

b. A continuación se presentan algunos coeficientes de correlación. Trata de encontrar cuáles corresponden a la relación de la *Esperanza de vida* con cada una de las variables de la tabla.

Coeficiente de correlación lineal r =	-0,40	1	0,78	0,91	-0,92	0,5	0,62	0,2	-0,73	0	0,38	0,61

5.7.1.1. ANÁLISIS A PRIORI Y DESARROLLO

El formador presentó la actividad, indicando a los futuros profesores que en un aula de Bachillerato, permitiría introducir algunos conocimientos sobre la dependencia aleatoria entre dos variables cuantitativas. También sirve para dar sentido a la utilidad de la correlación y regresión, enriqueciendo la idea de dependencia funcional. A continuación, se pidió a los participantes que completasen por escrito la actividad, en base al análisis de los diagramas de dispersión:

1. Estimasen la intensidad de la relación que se presenta en los diferentes diagramas, determinando un valor entre 0 y 1 que cuantifique su propia percepción, y señalando

en la columna anexa si la relación se considera de tipo directa o inversa mediante el signo + o –, respectivamente;

2. A continuación, asignasen a cada una de las variables que aparecen, uno de los coeficientes de correlación lineal que se ofrecen en la tabla adjunta.

Esta es la misma forma en que Estepa (1994) y Sánchez Cobo (1999) plantean la pregunta en una actividad semejante, quienes encontraron buena estimación de los coeficientes de correlación si las expectativas previas de los estudiantes no contradicen los datos, y mayor precisión cuando la relación es intensa y directa. Los autores también señalan la dificultad de algunos estudiantes al ordenar valores del coeficiente de correlación. La tarea se incluye para comprobar si algunos futuros profesores comparten estas dificultades.

Solución experta y objetos matemáticos implicados

En el debate posterior de las soluciones aportadas por los futuros profesores, el formador explicó que la estimación de la intensidad de las diferentes relaciones de dependencia (primera columna del apartado a), está sujeta a la percepción subjetiva de cada participante, luego cabe esperar diferencias en sus estimaciones. En el segundo año de la experiencia, el formador de profesores también presentó un resumen estadístico de las estimaciones aportadas por los participantes en el primer año de la experiencia, comparando con los valores reales, y estudiando el efecto de las diferentes variables de tarea. El formador de profesores también añadió que esta es una tarea utilizada en los estudios sobre percepción de la correlación (por ejemplo en Sánchez Cobo, 1999; y Cañadas, 2012, con tablas de contingencia).

Se llegó a un consenso de que las relaciones de mayor intensidad son las del Índice de desarrollo humano, y la Tasa de mortalidad, que tienen diferente signo. Le siguen el Índice de educación, la Tasa de fecundidad y la Población urbana, encontrando una intensidad baja en el Gasto público en salud y nula en la Población total. En las estrategias que expusieron los participantes para la parte final de la tarea, algunos primeramente identifican los dos diagramas con correlación negativa, observando la tendencia, y asignan el mayor valor del coeficiente al que presenta menor dispersión. Una vez asignado el valor cero al caso de independencia, quedaría asignar el resto de coeficientes (cinco a elegir entre los nueve restantes).

En la Tabla 5.7.3 se muestran los objetos matemáticos más relevantes de la tarea, que se refieren principalmente a los conceptos de variable estadística bidimensional y dependencia, así como sus tipos, correlación y coeficiente de correlación lineal; y diversas propiedades que relacionan el sentido e intensidad de la correlación con el valor absoluto y signo del coeficiente de Pearson, o con la forma y dispersión del diagrama de dispersión. Implícitamente aparece el concepto de función (CI6), aunque explícitamente no se pida que el participante ajuste un modelo funcional a los datos de los diferentes diagramas. Este concepto (CI6) debiese emerger de manera natural, aunque pueda ocurrir que no.

Destacamos la importancia del lenguaje gráfico y los argumentos basados en visualizaciones y análisis de ejemplos, aunque también se recurre al verbal deductivo.

En el plano procedimental, aparece la interpretación de gráficos, traducción de representaciones y estimación de la correlación.

Tabla 5.7.3. Objetos matemáticos implicados en la Tarea 5

La situación problema	P0.Organización/Reducción de datos P1.Analizar la existencia de relación entre variables
Lenguaje	L1.Verbal: etiquetas de los ejes L2.Simbólico: sentido de la dependencia L4.Gráfico: diagrama de dispersión L5.Numérico: ejes
Conceptos	C1.Variable estadística bidimensional; distribución C3.Dependencia funcional/estadística/independencia C4.Covarianza y/o correlación CI1.Variable estadística, valores, rango CI6.Funciones
Proposiciones Propiedades	PP26a. El coeficiente de correlación se acerca a 1 o -1 si la relación se aproxima a una línea recta. La intensidad de la dependencia lineal será más fuerte según proximidad a 1 o -1 PP26b. Interpretación del signo: positivo (negativo) si las variables tienen dependencia directa (inversa). PP18. Dispersión de la nube, correlación y regresión: A menor dispersión de los datos más próximo será el coeficiente de correlación a 1 o -1. Si los puntos se sitúan sobre una curva hay dependencia funcional.
Argumentos	A1. Análisis de ejemplos/ contraejemplos A2. Uso de Gráfico A5. Verbal – deductivo
Procedimientos	PC2.Representación gráfica: diagrama de dispersión PC3.Traducción entre representaciones: gráfica y numérica PC4. Interpretación correlación: Estimación de la correlación

5.7.1.2. RESULTADOS Y DISCUSIÓN

Para analizar los datos aportados por los participantes en sus respuestas escritas, se han tenido en cuenta el valor estimado y asignado (entre los dados), y el signo.

Valor estimado para el coeficiente

En la Tabla 5.7.4 se presentan los valores mínimo, máximo, media, y desviación típica del valor estimado para el coeficiente de correlación en las diferentes variables. Se acompaña de los valores reales del coeficiente de correlación (última columna) para una mejor interpretación de los resultados. Observamos que los valores medios se acercan, en general, al valor absoluto del coeficiente de correlación, aunque la precisión de la estimación es mayor para las correlaciones altas, y dentro de ellas, es mayor para las positivas que para las negativas.

Al contrario que lo observado en otras investigaciones con estudiantes universitarios (por ejemplo, por Cañadas, 2012, con tablas de contingencia), se ha estimado el caso de independencia (Población total) con un valor medio cercano a cero. En el trabajo de este autor, en una tabla de contingencia con independencia perfecta, se estimó un valor medio del coeficiente de correlación igual a 0,47. Por su parte, Sánchez Cobo, Estepa y Batanero (2000) informan de una actividad de estimación de la correlación a partir de varias representaciones con estudiantes de enfermería y empresariales. En una de ellas, donde se estima la correlación a partir de un diagrama de dispersión con independencia perfecta, se estima un valor medio 0,26. En consecuencia, nuestros futuros profesores muestran una mayor competencia en la identificación de la independencia estadística en diagramas de dispersión, y la estimación de su correlación, que en estudios previos.

Tabla 5.7.4. Estadísticos del valor absoluto estimado para el coeficiente de correlación

	Mínimo	Máximo	Media	Desv. típ.	Valor r
Índice de Desarrollo Humano	,70	1,00	,88	,08	0,91
PIB per cápita	,20	1,00	,71	,18	0,61
Tasa de fecundidad	,00	1,00	,52	,22	-0,73
Tasa de mortalidad de niños	,00	1,00	,82	,23	-0,92
Gasto público en salud	,00	1,00	,39	,23	0,38
Índice de educación	,00	1,00	,68	,18	0,78
Población, total	,00	,80	,05	,14	0,004
Población urbana	,00	1,00	,39	,21	0,62

Observamos una influencia de las teorías previas en la estimación de la correlación, principalmente en los casos en que los datos se muestran en contra de las mismas (Población urbana). Así mismo, encontramos una sobre estimación en el caso del PIB, donde los participantes esperan una correlación fuerte. Todo ello podría explicarse por el fenómeno de correlación ilusoria (Chapman y Chapman, 1967). Es decir, algunos futuros profesores podrían basar su estimación de la correlación atendiendo, de algún modo, a expectativas o esquemas propios referidos a las variables que estudian. Por otra parte, destacamos el caso de las correlaciones negativas (Tasa de fecundidad, Tasa de mortalidad), ya que encontramos una estimación a la baja, lo que podría indicar, en parte de los futuros profesores, una concepción unidireccional de la correlación, descrita por Estepa (1994).

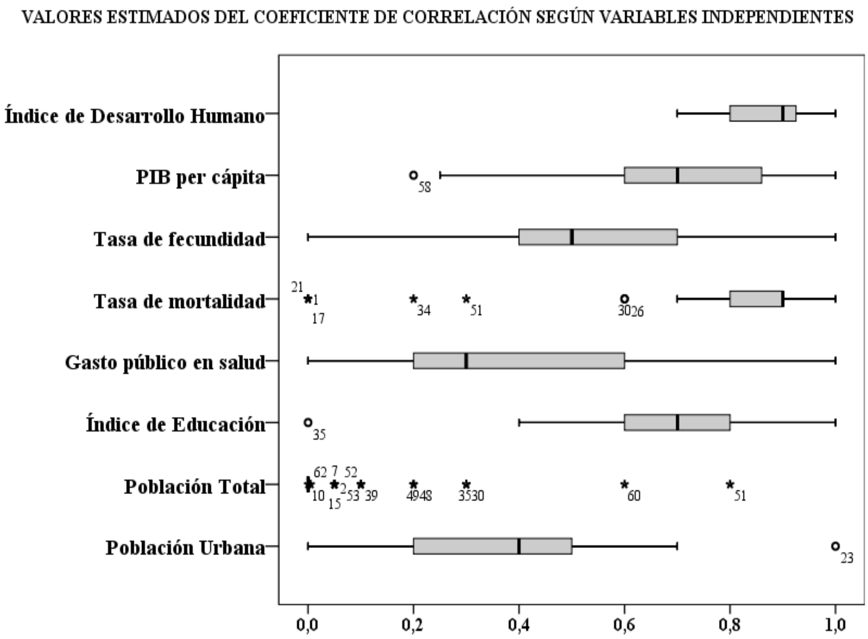


Figura 5.7.1. Gráficos de caja del valor estimado para el coeficiente de correlación

En cuanto a la dispersión de las estimaciones, que se observa en las desviaciones típicas y en la anchura de los gráficos de caja (Figura 5.7.1), nos da idea de la homogeneidad de la respuesta de los participantes a la tarea. La respuesta es muy homogénea en la Población total, lo que nos confirma la buena competencia de los participantes para detectar la independencia. Igualmente son extremadamente consistentes al estimar la correlación para el Índice de Desarrollo Humano, al igual que para la Tasa de mortalidad, donde la mayor desviación típica se debe a los casos atípicos. Observamos que las variables que muestran mayor dispersión en la estimación son el Gasto público y la Tasa de fecundidad, cuya función de ajuste no es lineal.

En general, hay pocos casos atípicos en las gráficas. En la tasa de mortalidad, algunos futuros profesores dan correlaciones muy bajas, incluso cero, al asignar una correlación nula cuando los datos presentan correlación negativa (casos 1, 17 y 21); por tanto, muestran claramente la concepción unidireccional (Estepa, 1994). En cuanto a la Población total, la mayoría de los participantes dan valor cero o muy próximo, De hecho, tanto la mediana como los dos cuartiles toman valor cero, por lo cual, no se aprecia la caja en el gráfico. Los casos atípicos son los pocos participantes que dan valores mayores que cero.

Valor asignado a partir de una lista de correlaciones

En la Tabla 5.7.5 incluimos los resúmenes estadísticos de los valores asignados al coeficiente de correlación, entre los dados en una lista. La tarea fue resuelta en casi la totalidad de la muestra. Observamos que los valores medios son ahora más próximos al verdadero valor de la correlación que en la actividad de estimación; además, han bajado las desviaciones típicas (Figura 5.7.2), sobre todo en la Tasa de mortalidad, el Gasto público, el Índice de educación y la Población urbana; por lo que dar la lista de valores facilitó mucho la tarea, y se puede deducir un mayor acuerdo en la respuesta.

Tabla 5.7.5. Estadísticos del valor absoluto estimado para el coeficiente de correlación

	Mínimo	Máximo	Media	Desv. típ.	Valor r
Índice de Desarrollo Humano	,41	1,00	,89	,09	0,91
PIB per cápita	-,42	1,00	,64	,23	0,61
Tasa de fecundidad	-,75	,50	-,54	,22	-0,73
Tasa de mortalidad de niños	-,92	,20	-,87	,17	-0,92
Gasto público en salud	,20	,91	,39	,17	0,38
Índice de educación	,38	,91	,69	,12	0,78
Población, total	-,40	,50	,02	,13	0,004
Población urbana	,20	,62	,45	,15	0,62

Para la Población total, aunque la mayoría de asignaciones de los futuros profesores son correctas, encontramos cinco que cambian su asignación previa de 0, o próxima a cero. Tres de ellos asignan ahora el valor 0,2; otro el valor -0,4; y otro 0,5. A pesar de estos cambios, el valor medio asignado en este ítem es muy cercano a cero, por lo que podemos considerar que los futuros profesores tienen mejor percepción de la independencia que los participantes en las investigaciones previas (Capítulo 3).

Para completar este análisis, mostramos en la Tabla 5.7.6 las respuestas por grado de corrección. Se ha considerado correcta si la asignación es exacta, y parcialmente correcta si se asigna un valor del coeficiente que difiera del exacto, en valor absoluto, en menos de una décima. Sólo para dos variables (Tasa de mortalidad y Población total) los resultados son correctos en más del 75% de la muestra. Sin embargo, la mayoría de los futuros profesores asignan correcta o parcialmente correcta la correlación. La mayor dificultad fue asignarla en el PIB, el Gasto en salud, y la Población urbana, las dos primeras con ajuste no lineal, y la última con correlación que contradice las teorías previas.

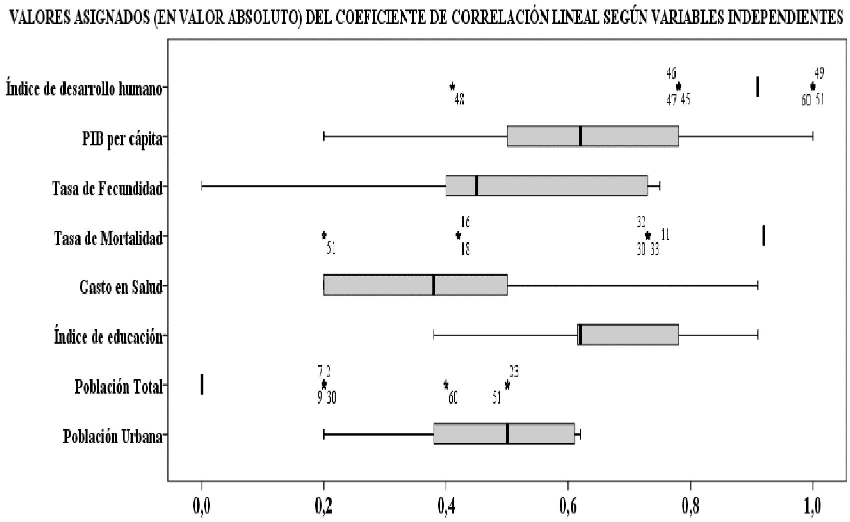


Figura 5.7.2. Gráficos de caja del valor asignado (valor absoluto) para el coeficiente de correlación

Tabla 5.7.6. Frecuencias de las asignaciones del coeficiente de correlación a las variables del proyecto

Variable	Correctas (%)	Parcialmente correctas (%)	Incorrectas (%)	No contestan (%)
Índice de Desarrollo Humano	39(60)	10(15,4)	13(20)	3(4,6)
PIB per cápita	9(13,8)	7(10,8)	46(70,8)	3(4,6)
Tasa de fecundidad	29(44,6)	2(3,1)	31(47,7)	3(4,6)
Tasa de mortalidad	54(83,1)		8(12,3)	3(4,6)
Gasto en salud	24(36,9)		38(58,5)	3(4,6)
Índice de educación	25(38,5)		37(56,9)	3(4,6)
Población, total	54(83,1)		8(12,3)	3(4,6)
Población urbana	9(13,8)	12(18,5)	42(64,6)	2(3,1)

Tabla 5.7.7. Correlación entre valores estimados y asignados del coeficiente de correlación

Variable	Correlación	Valor p	Respuestas
Índice de Desarrollo Humano	0,491	0,000*	61
PIB per cápita	0,509	0,000*	61
Tasa de fecundidad adolescentes	-0,241	0,062	61
Tasa de mortalidad de niños	-0,234	0,070	61
Gasto público en salud	0,789	0,000*	60
Índice de educación	0,545	0,000*	61
Población, total	0,194	0,137	60
Población urbana	0,628	0,000*	61

* Correlación significativa

En la Tabla 5.7.7 se muestran los coeficientes de correlación entre los valores estimados de la correlación (en la primera pregunta de la actividad) y los valores asignados (en la segunda), con el fin de estudiar la consistencia de la respuesta.

Observamos consistencia de respuesta en todas las variables, salvo para las variables con correlación negativa. Los valores asignados y estimados correlacionan muy poco, y negativamente, para las otras variables que parecen peor percibidas por los participantes. La falta de correlación para el caso de independencia se explica por ser los valores muy próximos a cero.

Signo asignado a la correlación

El signo asignado a la correlación (Tabla 5.7.8) se ha identificado fácilmente. En el caso de independencia (Población total), la mayoría no lo indica o asigna un valor cero, al concebir correctamente que la independencia supone ausencia de signo. Hay mayor diversidad en el Gasto público en salud, quizá porque se trate de la única variable con intensidad baja, a pesar de que los futuros profesores son consistentes en la estimación y asignación de dicha intensidad, según los análisis anteriores.

Tabla 5.7.8. Asignaciones del signo de la correlación a cada variable

Variable	Positivas (%)	Negativas (%)	No asignan (%)	Signo
Índice Desarrollo Humano	64(98,5)		1(1,5)	+
PIB per cápita	62(95,4)	2(3,1)	1(1,5)	+
Tasa de fecundidad	1(1,5)	61(93,8)	3(4,6)	-
Tasa de mortalidad	1(1,5)	63(96,9)	1(1,5)	-
Gasto público en salud	54(83,1)	1(1,5)	10(15,4)	+
Índice de educación	63(96,9)		2(3,1)	+
Población, total	12(18,5)	7(10,8)	46(70,8)	Independencia
Población urbana	60(92,3)		5(7,7)	+

5.7.2. RELACIÓN ENTRE CORRELACIÓN Y CAUSALIDAD

El proyecto continúa cuando el formador pide a los participantes que valoren la posible causalidad en las relaciones que se establecen en los diferentes diagramas, así como diferenciar los tipos de dependencia (lineal o no). También explica que la existencia de correlación no siempre se debe a una relación causa - efecto, sino que se presentan otras categorías como puede ser la existencia de terceras variables que producen la correlación (relación indirecta), interdependencia, concordancia, correlación espuria o casual, consideradas por Barbancho (1973).

5.7.2.1. ANÁLISIS A PRIORI Y DESARROLLO

6. *Explicación de la relación.* ¿Cuál de las variables analizadas tiene una relación de causa-efecto con la esperanza de vida? En la siguiente tabla indica las variables que consideres tienen una relación causal con la esperanza de vida y explica por qué.

Variable	¿Por qué?

El formador indica a los futuros profesores que la finalidad de la tarea es ayudar a los estudiantes a diferenciar correlación y causalidad, puesto que, de acuerdo a Estepa (1994) en muchos casos los estudiantes no diferencian entre estos dos conceptos, y manifiestan una concepción causal de la asociación estadística, que incluso es resistente a la instrucción. Se pidieron ejemplos de situaciones no causales en que aparece correlación; entre otros se indicaron: "la talla o peso de un padre, sabiendo la talla o peso del hijo adulto"; que están claramente correlacionados, pero la posible relación causal iría de padre a hijo y no al contrario.

Se explicó a los participantes los tipos de covariación sugeridos por Barbancho (1973), utilizando ejemplos que clarificasen las diferentes categorías. Se explicó la importancia de diferenciar correlación y causalidad en contextos como investigación o diagnóstico médico.

Con esta tarea se pretende analizar si algunos participantes manifiestan la concepción causal, y analizar las posibles concepciones que puedan poseer sobre la correlación, así como sus creencias previas sobre la relación de la esperanza de vida con otras variables.

Solución experta y objetos matemáticos implicados

En la Tabla 5.7.9 se analiza la explicación de la correlación de la variable esperanza de vida con cada una de las variables del estudio, utilizando las categorías de covariación identificadas por Barbancho (1973). Como se mostró en la Tabla 5.7.2, se han tenido en cuenta algunas de estas categorías de posible relación; además de la causalidad, se consideran relaciones de dependencia indirecta e interdependencia (además de la independencia).

Tabla 5.7.9. Explicaciones de la correlación entre la Esperanza de vida y otra variable

Variable	Relación	¿Por qué?
I. Desarrollo Humano	Interdependencia	En el cálculo del IDH se computa la esperanza de vida; un cambio en el IDH proporciona un cambio en la esperanza de vida.
PIB per cápita	Dependencia indirecta	Un país con PIB elevado tendrá mayor gasto en salud y educación, que hacen variar la esperanza de vida, es una relación indirecta.
Tasa de fecundidad adolescentes	Dependencia indirecta	La correlación es debida al efecto de otras variables; por ejemplo, la menor educación hace que haya mayor número de embarazos entre adolescentes y que el nivel de vida sea bajo. Esto repercute en los cuidados de salud y la esperanza de vida es menor.
Tasa de mortalidad de niños	Causal unilateral	Si hay muchos niños que mueren pequeños la duración de su vida computa en el cálculo de la esperanza de vida del país y la hace bajar, luego la relación es inversa.
Gasto público en salud	Causal unilateral	A mayor/menor gasto en sanidad pública mejor/peor cuidado de la salud, luego más aumenta o disminuye la esperanza de vida de un país, siendo la relación directa.
Índice de educación	Dependencia indirecta	Se trata de un índice compuesto que refleja la educación de un país pero su correlación se debe a variables como el PIB, que repercuten en una mayor esperanza de vida
Población total	No hay relación	No presenta correlación
% Población urbana	Dependencia indirecta	Hay variables que indirectamente producen la correlación; por ejemplo, el índice de educación o el PIB aumentan con el % de la población urbana, y hace subir la esperanza de vida

En la Tabla 5.7.10 presentamos los objetos matemáticos implicados en el desarrollo de la Actividad 6. Se puede observar la importancia de las proposiciones relacionadas con la tipología de covariación entre dos variables estadísticas, ya que esta actividad se diseña específicamente para hacer emerger la propiedad PP25a, y por ello, el futuro profesor deberá enfrentarse a algunos tipos de covariación considerados por Barbancho (1973). Será también muy importante la argumentación. Observamos también, que en esta actividad no sólo se requieren conocimientos estadísticos, sino también un razonamiento avanzado, que relacione del contexto con la correlación.

Tabla 5.7.10. Objetos matemáticos implicados en la Tarea 6

La situación problema	P1.Analizar la existencia de relación entre variables
Lenguaje	L1.Verbal (escrito y oral) L4.Gráfico L5.Numérico
Conceptos	C1.Variable estadística bidimensional; distribución: v. dependiente e independiente C3.Dependencia funcional/estadística/independencia C4.Covarianza y/o correlación
Proposiciones Propiedades	PP25a. Correlación no implica una causalidad: Cuando dos variables covarían puede ser por dependencia indirecta, correlación casual o interdependencia.
Argumentos	A1. Análisis de ejemplos/ contraejemplos A2. Uso de Gráfico A5. Verbal – deductivo
Procedimientos	PC4.Cálculo e interpretación covarianza y/o correlación: tipos de covariación

5.7.2.2. RESULTADOS Y DICUSIÓN

Los argumentos utilizados por los futuros profesores para justificar la causalidad, se han clasificado atendiendo a las siguientes categorías:

Crecimiento/decrecimiento. Se justifica la relación de tipo causal porque el crecimiento (decrecimiento) en la variable independiente implica un crecimiento (decrecimiento) en la variable dependiente. Sería correcta en algunos casos, pero no en general, pues una causa puede actuar sobre un efecto en forma no lineal:

> Cuanto más avances tiene una población o mayor desarrollo humano posee, mayor será la esperanza de vida. (PP).

También se aplica este tipo de argumentación, erróneamente, al confundir la variable independiente con la variable dependiente, como por ejemplo:

> Nos indica que cuanto menor sea la esperanza de vida mayor será la mortalidad (ChC)

Existencia de terceras variables. Cuando se justifica la causalidad por el efecto de otras variables. De este modo, no se diferencia causalidad de correlación indirecta:

> Mientras más alto sea el PIB per cápita por persona más bienestar se tendrá y por tanto la esperanza de vida mejorará. (LT).

En algunos casos se utiliza erróneamente este tipo de argumentación, como por ejemplo, al justificar una Esperanza de vida baja por una superpoblación (CM), o confundir la variable dependiente e independiente, como por ejemplo PG:

> A mayor fecundidad de la población, más población habrá. Menor será la esperanza de vida por superpoblación. (CM).
> Va directamente ligado con el índice de desarrollo humano, pues cuanto más desarrollo tiene un país, más escolaridad, más educación. (PG).

Modo en que se calcula la Esperanza de vida: En algunos casos se justifica la causalidad por puntos que afectan al cálculo de la Esperanza de vida:

> A menor mortalidad infantil, mayor será el índice de esperanza de vida puesto que crecerá la media. (CL).

Influencia en la variable dependiente. Cuando se argumenta la causalidad mediante la descripción de la influencia de la variable independiente en la dependiente:

> Porque si la información sobre cómo evitar enfermedades, y cómo mantener un estilo de vida saludable está a disposición de los ciudadanos gracias a la educación que se les proporciona, estos pueden poner remedio para prolongar y mantener su salud (PJ).
> Si gastas mas en salud e investigación se encuentran curas a nuevas enfermedades y por tanto la esperanza de vida aumenta. (GM).

Correlación. Cuando se identifica la correlación y la causalidad explícitamente, a veces con referencias a la probabilidad, y en otras. con una explicación del efecto del crecimiento o decrecimiento de la variable independiente en la Esperanza de vida:

> Están correlacionadas y además podríamos decir que si una sociedad está más desarrollada es más probable que vivas más años (EGO).
> El índice de correlación está cercano a 1 así que las variables son causa y efecto (MIH).

En la Tabla 5.7.11 se presenta el conjunto de variables consideradas por los futuros profesores en la tarea, junto al tipo de justificación que aportan a la relación de tipo causal. Observamos que las variables que se citan con mayor frecuencia fueron el Índice de desarrollo humano y la Tasa de mortalidad, seguida del PIB per cápita. Por otra parte, y como ocurrió con la estimación y asignación del coeficiente de correlación, la Población total se considera por todos los participantes como independiente de la esperanza de vida.

Tabla 5.7.11. Frecuencia y porcentaje (%) de participantes según argumentación de relaciones causales

Variable	Crecimiento	Terceras variables	Cálculo	Influencia	Correlación	Seleccionan la variable
Índice Desarrollo Humano	18(27,7)	30(46,2)		5(7,7)	5(7,7)	58(89,2)
PIB per cápita	5(7,7)	26(40)		3(4,6)	3(4,6)	37(56,9)
Tasa de fecundidad	2(3,1)	5(7,7)	1(1,5)	1(1,5)		9 (13,8)
Tasa de mortalidad	23(35,4)	3(4,6)	17(26,2)	7(10,8)	2(3,1)	52(80)
Gasto público en salud	2(3,1)	10(15,4)		8(12,3)	1(1,5)	21(32,3)
Índice de educación	4(6,2)	7(10,8)		10(15,4)	1(1,5)	22(33,8)
Población, total						
Población urbana	1(1,5)			1(1,5)		2(3,1)

El principal argumento de la supuesta relación causal por los futuros profesores es la existencia de terceras variables que, afectando o siendo afectadas por la variable independiente, a su vez influyen en la esperanza de vida. En estos casos, algunos futuros profesores podrían no diferenciar la relación causal de la dependencia indirecta. De hecho, la descripción de la influencia directa solo es citada 35 veces, y la relación indirecta 81.

Es igualmente frecuente el argumento del crecimiento/decrecimiento de la variable independiente en la dependiente (lo argumentan hasta el 35,4% de participantes en la tasa de mortalidad). Como hemos indicado, esta explicación podría ser aceptable; aunque no siempre. Por último, algunos pocos casos, explícitamente, identifican correlación y relación causal, todas en consonancia a sus teorías previas, sobre todo en el Índice de desarrollo humano y el PIB.

5.7.3. DETERMINACIÓN INTUITIVA DE UNA FUNCIÓN DE AJUSTE

Aceptada la existencia de una relación entre algunas de las variables estudiadas y la esperanza de vida, el siguiente paso es deducir una posible función de ajuste, lo que se estudia en la siguiente tarea.

5.7.3.1. ANÁLISIS A PRIORI Y DESARROLLO

7. *Cuál de las variables sirve mejor para predecir la Esperanza de vida?*
   a. En la siguiente tabla, ordena las variables, según pienses que sirve para predecir mejor la *Esperanza de vida*.
   b. ¿Podríamos para alguna de estas variables hallar una función matemática para predecir, aproximadamente, *Esperanza de vida* a partir de la otra variable? ¿Qué tipo de función?

Variable	Número de orden, según poder de predicción	Tipo de función
Índice de Desarrollo Humano		
PIB per cápita		
Tasa de fecundidad adolescentes		
Tasa de mortalidad de niños		
Gasto público en salud		
Índice de educación		
Población, total		
Población urbana		

El formador de profesores presentó la tarea, justificando la utilidad de ordenar las variables en cuanto a su poder de predicción (apartado a). Además, se les pidió en segundo lugar, que determinaran intuitivamente el modelo de regresión.

Solución experta y objetos matemáticos implicados

Los gráficos muestran una dependencia aproximadamente lineal (Índice de desarrollo humano, Población urbana o Índice de educación), o no lineal (en las otras variables), excepto para el caso de la Población total, donde se muestra independencia. El formador de profesores pidió a los participantes utilizar funciones conocidas (lineales, polinómicas, exponenciales, potenciales o logarítmicas), y valorar cuál de ellas se aproximaría mejor a los datos, en cada caso. En la Tabla 5.7.12 se presenta una posible solución a esta tarea, y aunque las resoluciones de los participantes difieran, el orden establecido será similar.

En la Tabla 5.7.13 presentamos un resumen de los objetos matemáticos implicados en el desarrollo de la actividad. Son tratados conjuntamente los conceptos de correlación y regresión, además, con una gran riqueza en cuanto al lenguaje. Se recuerdan conceptos sobre funciones, que el futuro profesor enseñará en otros temas, por ejemplo, la forma aproximada de diversas familias de funciones, También se observa que el coeficiente de correlación lineal mide la dependencia de tipo lineal, pero puede haber dependencia funcional alta (por ejemplo exponencial o logarítmica) sin que el valor del coeficiente sea elevado.

Tabla 5.7.12. Orden de las variables, según poder de predicción

Variable	Número de orden, según poder de predicción	Tipo de función	Valor r	Coef. det.
Índice de Desarrollo Humano	2	Lineal	0,91	0,82
PIB per cápita	3	Logarítmica	0,61	0,63[1]
Tasa de fecundidad adolescentes	5	Lineal-Exponencial	-0,73	0,54[2]
Tasa de mortalidad de niños	1	Exponencial	-0,92	0,86[3]
Gasto público en salud	7	Polinómica	0,38	0,23[4]
Índice de educación	4	Lineal	0,78	0,61
Población, total	8	Independencia	0,004	0,00
Población urbana	6	Lineal	0,62	0,38

[1]Modelo logarítmico; [2]Modelo lineal o exponencial; [3]Modelo exponencial; [4]Polinomio de grado cuatro

Tabla 5.7.13.Objetos matemáticos implicados en la Tarea 7

Situación problema	P2.Predecir una variable en función de otra
Lenguaje	L1.Verbal (escrito y oral) L2.Simbólico: expresiones algebraicas (funciones) L3.Tabular: tablas de valores para las funciones propuestas L4.Gráfico: diagrama de dispersión L5.Numérico: ordinales
Conceptos	C1.Variable estadística bidimensional; distribución C3.Dependencia funcional/estadística/independencia C4.Covarianza y/o correlación C5.Regresión CI6.Funciones
Proposiciones Propiedades	PP26a. El coeficiente de correlación mide el grado de relación de los datos. Se aproxima a 1 o -1 cuando la relación entre las variables se aproxime a una línea recta. Cuando los puntos se sitúan sobre una curva, hay dependencia funcional PP33. El coeficiente de determinación informa de la proporción de varianza de la variable dependiente explicada por el modelo de regresión y la bondad de ajuste PP45 Relación covarianza, correlación, regresión: Signo de correlación, signo covarianza y pendiente rectas de regresión.
Procedimientos	PC1.Representación tabular: uso de tablas para el ajuste del modelo PC2.Representación gráfica: diagrama de dispersión PC4.Cálculo e interpretación covarianza y/o correlación PC5.Cálculo e interpretación de modelos de regresión PC6.Predicción a través de modelos de regresión
Argumentos	A1. Análisis de ejemplos/ contraejemplos A2. Uso de Gráfico A5. Verbal – deductivo

Destacamos el lenguaje gráfico, así como numérico (números ordinales). El futuro profesor sólo debe indicar el tipo de función y no su ecuación, aunque ha de reconocer la forma aproximada de la nube de puntos, comparando con funciones que conoce (lineal, parábola, exponencial, etc.); así como apreciar el grado de dispersión para asignar el orden en cuanto al poder de predicción. La tarea permite ejercitar procesos de generalización y modelización, y sirve de introducción al cálculo de la función de ajuste con Excel.

5.7.3.2.RESULTADOS Y DISCUSIÓN

Orden de fiabilidad en la predicción

Para analizar los resultados de la tarea, se han caracterizado las respuestas como parcialmente correctas cuando se asigna un orden que difiere en una unidad del orden correcto. Se incluye la media y desviación típica del orden asignado por los futuros profesores, y se añade una columna con los valores reales del orden de las variables según su poder de predicción para facilitar la lectura de los resultados (Tabla 5.7.14).

Tabla 5.7.14. Frecuencia y porcentajes del grado de corrección del orden asignado

Variable	Orden predicción	Orden asignado		Grado de corrección			
		Media	D. Típica	Correctas (%)	Parcialmente correctas (%)	Incorrectas (%)	Blanco (%)
Índice Desarrollo Humano	2	1,56	0,71	20(30,8)	44(67,7)		1(1,5)
PIB per cápita	3	3,30	1,44	15(23,1)	30(46,1)	19(29,2)	1(1,5)
Tasa de fecundidad	5	4,80	1,09	33(50,8)	19(29,2)	12(18,5)	1(1,5)
Tasa de mortalidad	1	1,89	1,07	27(41,5)	24(36,9)	13(20)	1(1,5)
Gasto público en salud	7	5,41	1,50	17(26,2)	21(32,3)	25(38,5)	2(3,1)
Índice de educación	4	3,97	1,21	25(38,5)	23(35,4)	16(24,6)	1(1,5)
Población, total	8	7,54	0,95	46(70,8)	11(16,9)	6(9,2)	2(3,1)
Población urbana	6	6,50	1,04	16(24,6)	36(55,4)	12(18,5)	1(1,5)

Observamos que el orden de predicción medio asignado es muy cercano al orden real en todas las variables, algo menos en el Gasto en salud o la Tasa de mortalidad, lo que indica una buena competencia en la tarea.

Los futuros profesores, en su mayoría, ordenan de modo correcto o parcialmente correcto las variables según su poder para predecir la esperanza de vida, y las respuestas se distribuyen en torno al valor correcto dada la proximidad del valor medio al valor real, y la baja desviación típica. Los tres primeros órdenes se suelen asignar de modo correcto a las variables Tasa de mortalidad, Índice de desarrollo humano y PIB; y aunque para esta última aparezcan mayor variabilidad de órdenes asignados, todos los casos se agrupan en torno al valor central. Igualmente encontramos una buena asignación del orden de predicción en la Tasa de fecundidad, siendo la que mayor cantidad de respuestas correctas de los participantes posee.

Se observa mayor dificultad en las asignaciones de orden para correlaciones menos intensas, como el Gasto público en salud, pero en general, son bastante precisas. Además, se confirma la buena capacidad de los futuros profesores de apreciar la independencia, ya que un alto porcentaje de ellos asigna este caso como el peor para predecir la Esperanza de vida (un 70,8% de participantes).

Reconocimiento de la función de ajuste

En la Tabla 5.7.15 presentamos la función de ajuste, sugerida por los futuros profesores para cada una de las variables independientes trabajadas en el proyecto. La mayoría las identifican correctamente, sobre todo en las variables Índice de desarrollo humano, Índice de educación, PIB, y Tasa de fecundidad, tres de las cuales, (salvo el PIB), con función de ajuste lineal.

En general, los participantes han reconocido bien la forma de la función lineal, exponencial y logarítmica, mostrando su alta preparación matemática; incluso indicando con frecuencia su signo o tipo de crecimiento. En algunos casos, principalmente en la Población total, los participantes perciben la alta dispersión de los datos indicándolo en sus respuestas, aunque no indican que no es pertinente una función de ajuste a los mismos. El alto índice de no respuesta para dicha variable, da a entender que algunos futuros profesores no asignan una función porque no consideran que la haya.

Tabla 5.7.15. Frecuencia (y porcentaje) de función de ajuste considerada

Variable	Tipo de función prevista	Correcta	Lineal	Otra	No responde
Índice Desarrollo Humano	Lineal	63(96,9)		1[1](1,5)	1(1,5)
PIB per cápita	Logarítmica	56(86,2)		7[2](10,8)	2(3,1)
Tasa de fecundidad	Lineal-Exponencial	54(83,1)			11(16,9)
Tasa de mortalidad	Exponencial	34(52,3)	14(21,5)	16[3](24,6)	1(1,5)
Gasto público en salud	Polinómica	29(44,6)	4(6,2)	6[4](9,2)	26(40)
Índice de educación	Lineal	63(96,9)			2(3,1)
Población, total	Independencia	16(24,6)	1(1,5)		48[5](73,8)
Población urbana	Lineal	35(53,8)		6[6](9,2)	24(36,9)

[1]Parabólica; [2]Hiperbólica (1), Polinómica (5), Exponencial (1); [3]Hiperbólical (5), Polinómica (11); [4]Exponencial (2), Logarítmica (3), Independencia (1); [5]Indican alta dispersión (4); [6]Polinómica (5), Independencia (1);

5.8. ACTIVIDADE FORMATIVAS FINALES Y TRABAJOS OPCIONALES DE ALGUNOS PARTICIPANTES

Finalizada la recogida de datos de evaluación, se continuó el trabajo en el aula, en forma colectiva, con algunas actividades adicionales. En la primera edición de la experiencia, las actividades simplemente se realizaron colectivamente en el aula. Para ello, el formador de profesores, utilizando el cañón de proyección, y con la intervención de algunos participantes, iba iniciando la actividad, pidiendo a los futuros profesores su respuesta, cuando fuese necesario.

El segundo año se dio también la posibilidad de realizar un trabajo complementario individual, para mejorar la calificación en la asignatura. Dicha actividad podría ser alguna de las que se describen a continuación, y que se realizaron colectivamente en el aula el primer año. Las actividades se presentaron también en el aula, pero en vez de realizarlas por completo, únicamente se llevó a cabo un ejemplo, dejando a los participantes finalizarlas individualmente fuera del aula, y dejando quince días para completarlas.

En lo que sigue, se describen las actividades formativas finales, tal como se llevaron a cabo en el aula el primer año, y las soluciones aportadas voluntariamente por algunos futuros profesores el segundo año en que se realiza la experiencia. Aunque fueron pocos los que las completan, por tener carácter opcional y debido a la carga de trabajo en el resto de las materias, consideramos de interés analizar sus soluciones, que nos permiten evaluar (aunque con carácter muy exploratorio) el conocimiento matemático avanzado, y algunos aspectos de su conocimiento sobre la enseñanza de la correlación y regresión, en una pequeña muestra de participantes.

5.8.1. DETERMINACIÓN DE LA FUNCIÓN DE AJUSTE CON EXCEL

En la actividad 8, el objetivo es completar la actividad anterior con el cálculo formal de diversas funciones de ajuste, utilizando Excel. Con ello, se pretendía dar a conocer a los futuros profesores la herramienta que proporciona Excel, muy sencilla de utilizar para el estudio de la correlación y regresión, no sólo lineal, pues incluye diversas funciones de ajuste.

Como se vio en el apartado anterior, la mayoría resolvió satisfactoriamente la tarea 7, en la que implícitamente se trataba la noción de coeficiente de determinación, pues se pedía ordenar las variables independientes según su poder de predicción sobre la esperanza de vida. También se les pidió indicar la función que serviría para tal ajuste. Finalizada dicha tarea, el formador de profesores introduce la siguiente actividad.

8. Utilizando Excel encuentra la función que mejor describa la esperanza de vida a partir de cada una de las variables (entre las disponibles en el programa) y escribe la expresión algebraica

Variable independiente	Expresión algebraica de la función que usaríamos para predecir la esperanza de vida $Y=f(X)$
Índice de Desarrollo Humano	
PIB per cápita	
Tasa de fecundidad adolescentes	
Tasa de mortalidad de niños	
Gasto público en salud	
Índice de educación	
Población, total	
Población urbana	

Asimismo, se propone a los participantes cambiar la variable dependiente y repetir el análisis con otras nuevas variables. Puesto que la actividad matemática es la misma en las dos tareas, las analizamos conjuntamente.

9. Elige otra variable del fichero (Índice de Desarrollo Humano; PIB per cápita; Tasa de fecundidad adolescentes; Tasa de mortalidad de niños; Gasto público en salud; Índice de educación; Población, total o Población urbana) como variable dependiente. Utilizando los datos dados en la hoja Excel, estudia su posible relación con el resto de variables, incluida la esperanza de vida. Utilizando Excel encuentra la función que mejor describa la esperanza de vida a partir de cada una de las variables (entre las disponibles en el fichero) y escribe la expresión algebraica

5.8.1.1. ANÁLISIS A PRIORI Y DESARROLLO

El formador de profesores introdujo, en forma dinámica, las herramientas de la hoja de cálculo Excel que permiten completar estas actividades. Dichas herramientas permiten visualizar, junto con el diagrama de dispersión, diferentes funciones de ajuste junto con su expresión algebraica y el valor del coeficiente de determinación. Se consideran los modelos exponencial, lineal, logarítmico, polinómico (grado 2 a 6) o potencial.

El formador de profesores mostró estas opciones a los futuros profesores, justificando su utilidad para el estudio de la correlación y regresión en Bachillerato, pues el tiempo que el estudiante debiera ocupar realizando esos laboriosos cálculos, lo puede emplear en analizar las diferencias, ventajas e inconvenientes de un modelo u otro.

Solución experta y objetos matemáticos implicados

Los futuros profesores conocen Excel, cuyo uso se recomienda en la estadística en la Educación Secundaria Obligatoria. (MEC. 2007a). El formador de profesores, colectivamente, comienza a trabajar con los participantes con la primera de las variables. En gran grupo, y junto al formador, realizan ensayos con varios modelos, resaltando para cada uno el significado de los parámetros de su expresión algebraica, y el porcentaje de varianza explicada (R^2, coeficiente de determinación). Por ejemplo, se pregunta a los futuros profesores por el significado de la pendiente y la ordenada en el origen para la función lineal. Antes de pasar a un nuevo modelo, se piden conjeturas sobre si el ajuste mejorará o no. Pide propuestas sobre cuál de estos modelos utilizar, teniendo en cuenta tanto la mejora del coeficiente de determinación, como obtener una función sencilla.

El gráfico producido por defecto no contiene la línea de tendencia ni su ecuación (Figura 5.8.1). El formador utiliza la opción de diseño para elegir un diagrama de dispersión que sí proporciona esta información. En este ejemplo, la recta de regresión, que es la opción por defecto, tiene por ecuación $Y = 52{,}05X + 34{,}78$.

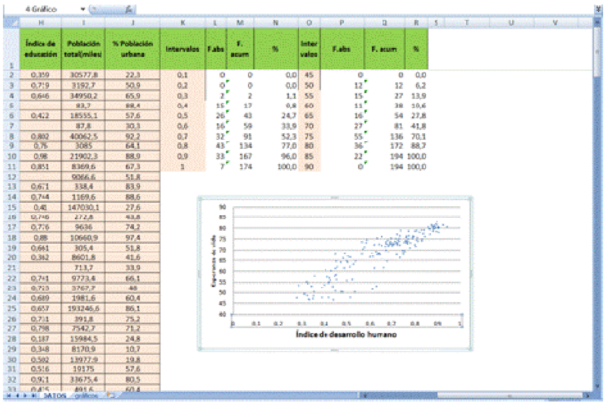


Figura 5.8.1. Elaboración inicial de un diagrama de dispersión con ayuda de Excel

Se interpreta que la esperanza de vida mínima es 34,7 años (ordenada en el origen) y por cada punto que sube el índice de desarrollo humano se incrementa en 52 años. Ello permite hacer predicciones, conociendo el Índice de desarrollo humano de un país. El cuadrado del coeficiente de correlación (0,8211), indica que el 82% de la varianza de la variable *Y* (esperanza de vida) queda explicada por el modelo lineal. El formador de profesores, junto con los participantes, prueban otros modelos de ajuste para ver si mejora este porcentaje (mejora la predicción). En la Figura 5.8.2 se presenta el modelo lineal, junto a un modelo polinómico, pero apenas hay mejora, por lo cual, la mejor decisión es utilizar el modelo lineal en este caso, que es simple y proporciona un buen ajuste.

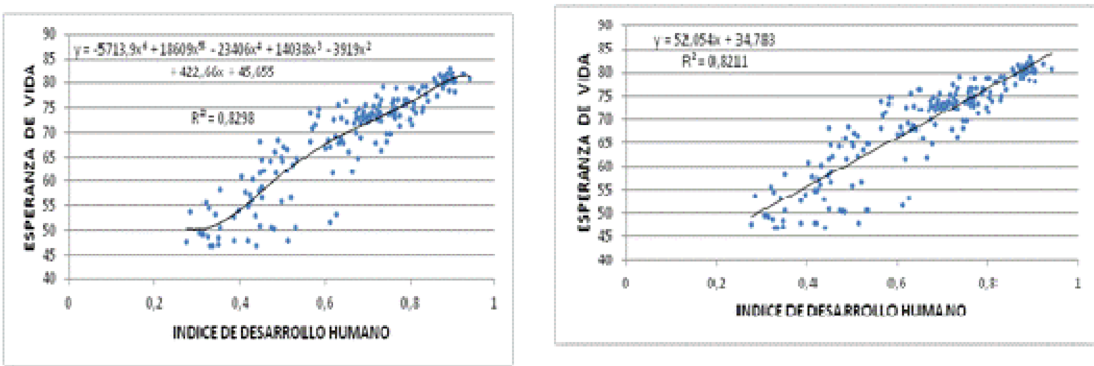


Figura 5.8.2. Ajuste funcional a la relación de dependencia con la variable IDH

Tabla 5.8.1. Expresiones algebraicas y coeficiente de determinación de funciones de ajuste

Variable independiente	Expresión algebraica de la función que usaríamos para predecir la Esperanza de vida	Coeficiente determinación	Tipo de función
Índice desarrollo humano	y = 52,054x + 34,783	0,82	Lineal
Producto Interior Bruto	y = 6,2216ln(x) + 57,807	0,63	Logarítmica
Tasa de fecundidad	y = -0,1692x + 77,682	0,54	Lineal
Tasa mortalidad niños	$y = 77{,}533e^{-0{,}003x}$	0,86	Exponencial
Gasto Público en salud	$y = 0{,}005x^4 - 0{,}1599x^3 + 1{,}4069x^2 - 1{,}8107x + 63{,}813$	0,23	Polinómica
Índice de educación	y = 38,777x + 44,166	0,61	Lineal
Población total	No hay ajuste	0	
% Población urbana	y = 0,2614x + 54,548	0,38	Lineal

Del mismo modo se trabaja con el resto de variables independientes. La Tabla 5.8.1 resume el análisis realizado de la esperanza de vida con el resto de variables. Al decidir el mejor ajuste, se toma en cuenta la simplicidad de la expresión y la mejora sustancial de la precisión. Además, en la discusión con los estudiantes resultó útil comparar ambos ajustes, en correspondencia con la aproximación intuitiva que se realizó en el ejercicio anterior. Esta tarea, junto a la anterior, son las diseñadas específicamente para el tratamiento de la regresión y el análisis del coeficiente de determinación, como medida de la bondad del ajuste, y proporción de la varianza explicada por el modelo.

En la Tabla 5.8.2 presentamos los objetos matemáticos implicados en el desarrollo de esta actividad, y la actividad 9. Su importancia radica en el tratamiento del coeficiente de determinación, y de su relación con la correlación y regresión, cuyas propiedades serán el objetivo principal de la tarea. Se recuerdan diversas funciones elementales, mostrando su utilidad como modelos de ajuste a los datos. La argumentación se realiza, tanto a través de análisis de ejemplos, como gráficamente y de forma verbal deductiva.

Tabla 5.8.2. Objetos matemáticos implicados en las Tareas 8 y 9

Situación problema	P2.Predecir una variable en función de otra
Lenguaje	L1.Verbal (escrito y oral) L2.Simbólico: coeficiente de determinación y expresiones algebraicas (funciones) L4.Gráfico: diagrama de dispersión L5.Numérico
Conceptos	C1.Variable estadística bidimensional; distribución C3.Dependencia funcional/estadística/independencia C4.Covarianza y/o correlación C5.Regresión C6.Coeficiente de determinación CI6. Funciones: lineal, polinómica, logarítmica, potencial, exponencial; significado de los parámetros; representación gráfica
Proposiciones Propiedades	PP25a. Una correlación intensa entre variables no implica causalidad PP26a. El coeficiente de correlación mide la correlación lineal entre las variables PP32a. La recta de regresión hace mínima la suma de los cuadrados PP33. El coeficiente de determinación informa de la varianza de la variable dependiente explicada por el modelo de regresión PP45 Relación covarianza, correlación, regresión. PP18. Dispersión de la nube, correlación y regresión PP7. El coeficiente de determinación informa del grado de ajuste del modelo de regresión a los datos, que varía entre 0 y 1
Argumentos	A1. Análisis de ejemplos/ contraejemplos A2. Uso de Gráfico A5. Verbal – deductivo
Procedimientos	PC2.Representación gráfica: diagrama de dispersión PC4.Cálculo e interpretación covarianza y/o correlación PC5.Modelos de regresión atendiendo al coeficiente de determinación PC6.Predicción a través de modelos de regresión. PC7.Uso de la tecnología: uso de Excel.

5.8.1.2. SOLUCIONES APORTADAS POR ALGUNOS PARTICIPANTES

Un total de 16 participantes en el segundo año del estudio (AA, AG, BS, CM, GO, HD, IE, JP, JPA, MA, ME, MG, MJG, MMG, RS, VC) ajustaron, con ayuda de Excel, modelos de ajuste a la esperanza de vida de cada una de las variables independientes. En la Tabla 5.8.3 presentamos un resumen de los modelos de ajuste presentados. Otro participante más, (JC), indica correctamente el tipo de función pero no ajusta una particular.

Por ejemplo, la alumna CM determina todas las funciones de ajuste, y para decidir el mejor modelo utiliza elementos numéricos (coeficiente de determinación) y gráficos (Figura 5.8.3), como podemos ver en su argumento, que reproducimos a continuación. No obstante, ha confundido coeficiente de correlación y de determinación (pues el programa ofrece el valor R^2 y no el valor r. La alumna trata de ajustar un modelo a todas las variables, incluso a la Población total, que no correlaciona con la esperanza de vida. Siempre elige un modelo no lineal, incluso en los casos en que el ajuste lineal es muy fuerte; posiblemente trata de mejorar el ajuste, y no percibe que el ajuste ya es bueno pues ha tomado el coeficiente de determinación como coeficiente de correlación.

Como vemos la variable que mejor describe la esperanza de vida es la tasa de mortalidad de niños, cuya función tiene como expresión algebraica y = 0,0008x2 - 0,3096x + 79,411, puesto que si nos fijamos en la columna añadida en la tabla (coeficiente de correlación) es la que tiene mayor coeficiente de correlación. Además, lo podemos ver en la gráfica siguiente ya que se aprecia perfectamente como la línea de tendencia ajusta los datos con una precisión bastante buena (CM).

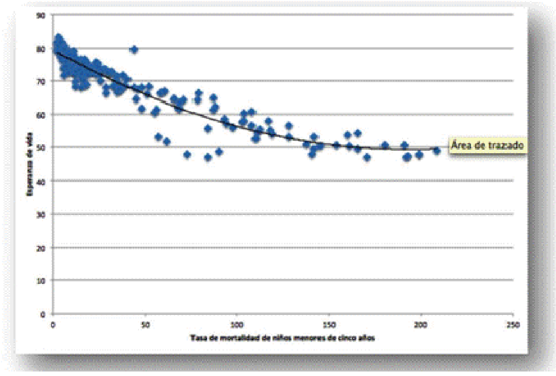


Figura 5.8.3. Gráfica proporcionada por CM en su informe sobre funciones de ajuste

En la Tabla 5.8.3 presentamos los modelos ajustados, que no se reducen al lineal, sino que los futuros profesores incluyen funciones logarítmicas, exponenciales y polinómicas de diferente grado. Los participantes, en este segundo caso, han comparado el coeficiente de determinación para elegir el modelo que lo haga máximo.

Tabla 5.8.3. Tipos de modelos ajustados por algunos participantes

Variable independiente	Expresión algebraica de la función de ajuste a la Esperanza de vida		Lineal	Logarítmica/ Exponencial	Polinómica
Índice desarrollo humano	y = 52,054x + 34,783	Lineal	AA, BS, HD, IE, JP, JPA, MG, MJG, MMG, RS, VC		AG, CM, GO, MA,ME, PG
Producto Interior Bruto	y = 6,2216ln(x) + 57,807	Logarít-mica	MJG,	AA, BS, CM, GO, HD, IE, JP, JPA, MA, ME, MG, MMG, PG RS, VC	AG,
Tasa de fecundidad	y = -0,1692x + 77,682	Lineal	AA, JP, JPA, VC, MG, MJG, RS	BS, CM, GO, HD, IE, MA,ME, PG	AG,
Tasa mortalidad niños	$y = 77,533e^{-0,003x}$	Exponen-cial	JPA, MG, MJG, VC,	BS, HD, IE	AA, AG, CM, GO, JP, MA, ME, PG
Gasto Público en salud	$y = 0,005x^4 - 0,1599x^3 + 1,4069x^2 - 1,8107x + 63,813$	Polinómi-ca	MJG,	JPA, VC, HD, IE MG	AA, AG, BS, CM, GO, JP, MA,ME, PG
Índice de educación	y = 38,777x + 44,166	Lineal	AA, BS, HD, IE, JP, JPA, MG,MJG, VC		AG, CM, GO, MA,ME, PG
Población total	No hay ajuste		BS, MG. MJG,	CM, GO, JP, MA	
% Población urbana	y = 0,2614x + 54,548	Lineal	AA, HD, IE, JP, JPA,MJG, VC, MG	BS,	AG, CM, GO,MA, PG

El modelo no siempre coincide con el previsto en nuestro análisis, pues en nuestro caso, hemos simplificado a veces, tomando el lineal o exponencial /logarítmico, en vez de utilizar un modelo polinómico, como hacen con frecuencia los futuros profesores (a veces de grado tres o cuatro). No obstante, puesto que esta solución proporciona mayor varianza explicada, es correcta. Observamos en estos casos, cómo los futuros profesores se dividen en dos grupos: los que usan el criterio de máxima varianza explicada y los que optan por un modelo más simple.

Todos los participantes realizan un ajuste correcto, representando gráficamente los datos, y cambiando la escala del eje X para visualizar los datos con facilidad. Únicamente MJG muestra dificultades en la tarea, porque no cambia la escala para las diferentes variables, con lo cual, visualiza con dificultad la relación con la Esperanza de vida. Este futuro profesor opta por aplicar el modelo lineal en todas las variables.

Varios participantes tratan de encontrar un modelo de ajuste para la relación con la Población total, a pesar de que se vio la independencia y ellos mismos la interpretaron correctamente. Otros participantes indican expresamente "*No hay expresión algebraica representativa*" (AG, HD) o "*No hay función*" (AA). Y el resto, simplemente no lo ajusta; todo ello confirma la buena percepción de la independencia por estos futuros profesores, que se advirtió en análisis precedentes.

Análisis de otras variables dependientes

Los 17 futuros profesores anteriores (AA, AG, BS, CC, CM, GO, HD, IE, JC, JP, JPA, MA, ME, MG, MMG, PG, VC), incluyendo el que en la primera actividad sólo indicó el tipo de función sin expresarla, eligieron una variable dependiente diferente, y ajustaron modelos considerando su relación con alguna de las independientes, generalmente en forma correcta. Enviaron los ficheros Excel o los gráficos con los resultados del análisis. En general, elegida una variable dependiente, la cruzan con el resto, y determinan una función de ajuste.

Estos participantes determinan correctamente un modelo de ajuste a sus variables dependientes, y generalmente, obtienen una conclusión sobre cuál de ellas proporciona mejor predicción o, como en el siguiente caso, si ninguna sirve para predecir la dependiente:

> Como podemos observar, no podemos predecir el gasto público de un país en función de ninguna de las variables, ya que, como podemos ver, el coeficiente de correlación es muy pequeño en todos los casos (MA).

En la Tabla 5.8.4 se presenta una síntesis de los modelos ajustados por cada uno de los futuros profesores, que hicieron opcionalmente esta actividad. Como vemos, cada uno eligió libremente la variable dependiente, y de hecho, todas las variables del fichero han sido tomadas como dependientes por algún alumno, excepto la Población total, donde, como se ha dicho, se consideró por todos independiente. Observamos que, en general, y con pocas excepciones, una vez que el participante elige una variable dependiente, la cruza con las del resto del fichero.

También observamos que al ajustar sus modelos, sólo seis futuros profesores ajustan la variable elegida con la Población total; aunque hay algunas excepciones. Así, por ejemplo, MA la cruza con el Gasto público en salud, obteniendo una función exponencial $y = 6{,}7927 \cdot x^{0{,}086}$, aunque con poco coeficiente de determinación (0,081).

Tabla 5.8.4. Análisis de otras variables dependientes con los mismos datos por algunos participantes

	Variable independiente							
Variable Dependiente	Esperanza de vida	PIB	TF	TM	GPS	IE	PT	PU
Índice desarrollo humano	GO	GO MMG	GO	GO	GO	GO		GO
Producto Interior Bruto	AG, IE, JP, MG		AG, IE, JP, MG	AG, IE, JP, MG	AG, IE, JP, MG	AG, IE, JP, MG	JP, MG	AG, CC, IE, JC, JP, MG
Tasa de fecundidad	CM	CM		CM	CM	CC, CM, JC	CM	CM
Tasa mortalidad niños	HD,	HD,	HD,		CC, HD, JC	HD		HD,
Gasto Público en salud	MA, PG	MA, PG	MA, PG	MA, PG		CC, JC MA, PG	MA,	CC, JC MA, PG
Índice de educación	AA	AA	AA	AA	AA			AA, CC, JC
Población total								
% Población urbana	CC, JPA, VC	JPA, VC	JPA, VC	JPA, VC	JPA, VC	JPA, VC	JPA, VC	

5.8.2. ANÁLISIS DE NUEVOS DATOS

En esta tarea se ofrece al futuro profesor espacios de investigación y reflexión, para poder elegir sus propios datos, analizarlos y establecer conclusiones sobre otras variables disponibles en la web de las Naciones Unidas.

5.8.2.1. ANÁLISIS A PRIORI Y DESARROLLO

10. *Otros datos disponibles en Internet.* Visita el servidor de las Naciones Unidas en (http://hdr.undp.org/es/estadisticas/datos/). Elige algunas otras variables que te interesen; descárgalas en formato Excel y estudia su relación. Elige una de ellas como variable dependiente y encuentra la función de ajuste con cada una de las otras variables elegidas.

El formador de profesores mostró nuevamente a los participantes el servidor, y explicó la forma en que podrían elegirse variables para introducirlas en la hoja Excel, al igual que se había hecho con la utilizadas en el proyecto. Como ejemplo, pidió a los futuros profesores propuestas de nuevas variables a analizar, y se descargaron algunas al ordenador del profesor.

El formador resaltó la cantidad de variables disponibles para analizar su posible dependencia con la Esperanza de vida, o plantear nuevos estudios. También explicó que este es sólo uno de los muchos servidores que proporcionan datos; citó los del Instituto Nacional de Estadística, como otro ejemplo. Se explicó, que en el trabajo con estudiantes de Bachillerato, el objetivo es que el estudiante formule nuevas preguntas que requieran disponer de datos para resolverlas, desarrollando su razonamiento estadístico. También se espera que conozca mejor la forma en que se recogen estas variables, y aprecie la complejidad del proceso de elaboración de las estadísticas demográficas o económicas, así como su utilidad.

Solución experta

La actividad presenta la dirección del servidor de Naciones Unidas, desde donde se pueden elegir otros datos, clasificados en los grupos siguientes: *Desarrollo Humano*, *Movilidad Humana*, *Cambio climático* (Figura 5.8.4). Cada uno de ellos, a su vez, se abre en varias bases de datos, de las cuales se pueden seleccionar variables para países y años.

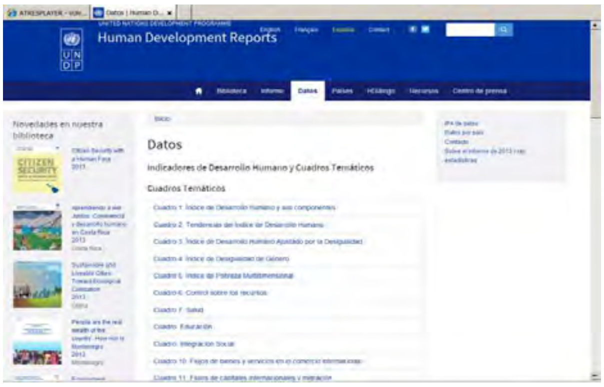


Figura 5.8.4. Bases de datos de las Naciones Unidas

Seleccionando algunas de estas variables, podríamos analizar la dependencia de la Esperanza de vida con otras variables diferentes a las del proyecto; por ejemplo, la Tasa de participación en la fuerza de trabajo (relación de mujeres-hombres de entre 15-64 años en el mercado laboral) o la Tasa de alfabetización de adultos (porcentaje de la población de 15 años alfabetizada). El futuro profesor ha de ser capaz de seleccionar datos de dicho servidor, en formato Excel, y repetir el análisis efectuado en la actividad anterior, obteniendo sus conclusiones del mismo.

Los objetos matemáticos implicados en esta actividad se presentan junto a los de siguiente tarea 11, pues en su desempleo se emplean los mismos (Ver Tabla 5.8.6).

5.8.2.2.SOLUCIONES APORTADAS POR ALGUNOS PARTICIPANTES

Un total de trece participantes (AA, AG, GO, HD, IE, JC, JP, JPA, MC, ME, MG, PG, VC) seleccionaron nuevos datos del servidor de las Naciones Unidas, repitiendo el análisis ya realizado anteriormente. Fueron unos pocos menos que los que realizaron la anterior actividad pues, al parecer, el exportar los datos resultó más complicado.

Se eligió alguna variable nueva como independiente; generalmente el futuro profesor elige una sola variable dependiente y la cruza con varias independientes, como vemos en la Tabla 5.8.5, donde, en la primera columna indicamos la variable dependiente, y en la segunda, todas las independientes analizadas por el mismo participante. Algunos futuros profesores también toman nuevas variables dependientes, proponiendo un nuevo estudio (y no el de la Esperanza de vida o el desarrollo humano)

Así, IE (primera fila) selecciona siete variables independientes nuevas sobre igualdad de género, que cruza con el Índice de desarrollo humano. Obtiene todos los modelos de regresión correctamente, pero no comenta sus conclusiones. Son otros los alumnos que se interesan por este tema, por ejemplo JP, quien toma como variable dependiente la proporción de mujeres en el parlamento y HD, quien relaciona la mortalidad infantil con varios indicadores de igualdad de género.

Por su parte, GO elige como campo la Investigación y desarrollo tecnológico, tomando como variable dependiente el gasto en I+D+I, y una serie de variables independientes relacionadas con la tecnología e investigación, aportando una gama de coeficientes de determinación desde 0,189 hasta 0,749.

Tabla 5.8.5. Modelos de regresión ajustados por los futuros profesores en la Tarea 10

Variable Dependiente	Variables independientes
Índice desarrollo humano (IE)	Tasa de escolaridad, Mujeres con al menos Educación Secundaria, Hombres con al menos Educación Secundaria, Índice de desigualdad de género, Tasa de participación laboral (Mujeres), Tasa de participación laboral (Hombres), Mujeres en el parlamento nacional
Índice desarrollo humano (AG)	Tasa de escolaridad, Mujeres con al menos Educación Secundaria, Hombres con al menos Educación Secundaria, Tasa de participación laboral (Mujeres), Tasa de participación laboral (Hombres), Desempleo juvenil, Satisfacción con el trabajo
Índice desarrollo humano (PG)	Años de escolarización Adultos, Tasa bruta de matriculación, gasto en educación
Esperanza de vida (MG)	Esperanza de vida en 1980; Esperanza de vida en 2012
Esperanza de vida (ME)	Tasa bruta de matriculación, Años de educación promedio, Ingreso nacional bruto (INB) per cápita, Población bajo la línea de pobreza de ingresos, Tasas de alfabetización de adultos, Superficie forestal , tasa de mortalidad maternal
Tasa mortalidad infantil (HD)	Índice de desigualdad de género, Tasa de participación laboral (Mujeres), Población con al menos Educación Secundaria (Mujeres), Mujeres en el parlamento nacional, Población con al menos Educación Secundaria.
Gasto Público en Investigación, Desarrollo e Innovación (I+D+i) (GO)	Número de investigadores, Tasa de graduados en Ciencias e Ingeniería, Patentes registradas, Grado de electrificación, Número de Ordenadores Personales, Número de usuarios de Internet, Número de abonados a servicios de telefonía móvil y fija
Índice de ingresos (JPA, VC)	Años educación promedio, Indice de salud, Población mujeres, Tasa participación trabajo mujeres/hombres
Tasa de empleo (AA)	Desempleo juvenil; Trabajo infantil; Pérdida total en el IDH debido a la desigualdad; Satisfacción general en la vida; Satisfacción con la libertad de elección; Satisfacción con el trabajo; Confianza en la gente; Satisfacción con la comunidad; Confianza en el Gobierno Nacional; Percepción de la seguridad; Tasa de homicidios.
% Población con estudios secundarios (CC)	GPD per cápita, % de población vacunada, tasa de desempleo juvenil, número de usuarios de Internet
% Mujeres en parlamento (JP)	Tasa de mortalidad maternal; N. abonados a telefonía móvil; índice de desigualdad
Puntuación PISA en Matemáticas (BS)	% de la población con al menos estudios secundarios
Puntuación PISA en Ciencias (BS)	% de la población con al menos estudios secundarios
Puntuación PISA en Comprensión lectora (BS)	% de la población con al menos estudios secundarios

Observamos un interés por el empleo y la educación en estos participantes. Así, JPA y VC relacionan el Índice de ingresos con varias variables educativas; y AA analiza la Tasa de empleo en función de varios indicadores de desigualdad de género, educación y Confianza en diversas instituciones. Incluso un participante se interesa por explicar los resultados de los estudios PISA, analizando las puntuaciones en matemáticas, comprensión lectora y ciencias respecto al abandono en la Educación Secundaria, como expone en sus comentarios:

> Dentro de la gran masa de datos que podemos obtener de esta página, me resultan de especial interés aquellos referentes a educación. En concreto, vamos a estudiar cómo afecta un temprano abandono o fracaso escolar, reflejado con la variable referente al porcentaje de población que tiene, al menos, la Educación Secundaria Obligatoria, o equivalente, a los diferentes resultados que los diferentes países obtienen en el ámbito de matemáticas, lectura o ciencia en el informe PISA, que quizás sea el más conocido en el ámbito informativo en los países que lo realizan (BS).

En resumen, este grupo de futuros profesores ha demostrado, no sólo competencia para seleccionar diferentes datos del servidor de las Naciones Unidas, sino también conocimiento del contexto para seleccionar variables relevantes, que les sirvan para estudiar fenómenos de su interés, planteando y resolviendo nuevos problemas. Con todo ello, muestran un alto grado de conocimiento avanzado en estadística, y en particular, de la correlación y regresión.

5.8.3. VISUALIZACIÓN DE DATOS

11. *Visualización de los datos:* Visita los recursos de exploración de datos de las Naciones Unidas en http://hdr.undp.org/en/data-explorer. Representa en un diagrama de burbujas la esperanza de vida (en eje Y) eligiendo otras variables para el eje X, el tamaño y color de la burbuja. Observa el cambio general y el que experimenta España a lo largo del tiempo. Prepara un informe con tus conclusiones.

En esta Actividad se presenta el diagrama de burbujas dinámico, que aporta puntos de vista complementarios a los hasta ahora ofrecidos por el diagrama de dispersión. Su utilización ofrece la ventaja de visualizar otras variables a la vez que se representan los datos bidimensionales.

5.8.3.1. ANÁLISIS A PRIORI Y DESARROLLO

Solución experta y objetos matemáticos implicados

El formador presentó la tarea, mostrando a los participantes el enlace que nos dirige al conjunto de herramientas de tratamiento y visualización de datos que ofrece las Naciones Unidas. En particular, se les indica que el objetivo es trabajar con el *Explorador de datos públicos (Public Data Explorer)*. El formador pregunta si ya lo conocían, pero el servidor era nuevo para todos los futuros profesores.

El formador entra en la aplicación, y explica a los futuros profesores que este diagrama de burbujas permite representar hasta cinco variables simultáneamente, usando los dos ejes, el tamaño y color con que se grafique cada dato, y además, se puede variar la gráfica en el tiempo pulsando un marcador. Como ejemplo, representa la Esperanza de vida (eje *Y*), en función de la variable Población con al menos Educación Secundaria completa (eje *X*), tomando a la vez la Población total (tamaño de los círculos, que representan cada país de la muestra), y el Índice de Desarrollo Humano (color de cada uno de los datos bidimensionales que se representan). Es posible destacar algunos países si se desea, y así, el formador marcó España y algún otro país para comparar (China el primer año, y Alemania el segundo). En gran grupo y con el formador de profesores, se interpreta el gráfico y las principales diferencias, así como el tipo de relación entre las dos variables principales.

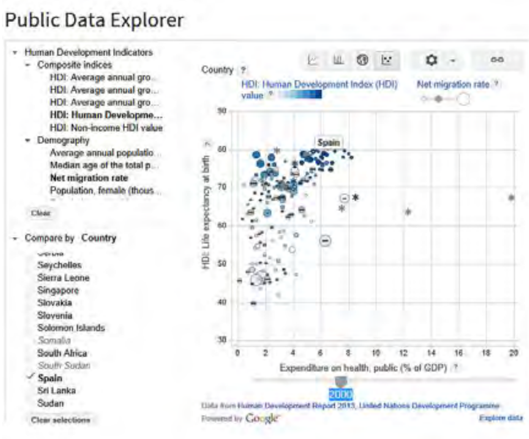


Figura 5.8.5. Gráfico de burbujas disponible en el *Explorador de datos* de la ONU

Tabla 5.8.6. Objetos matemáticos implicados en la Tarea 11

Situación problema	P0.Organización/Reducción de datos P1.Analizar la existencia de relación entre variables
Lenguaje	L1.Verbal (escrito y oral) L3.Tabular: Tablas de datos de las Naciones unidas L4.Gráfico L5.Numérico
Conceptos	C1.Variable estadística bidimensional; distribución: v. dependiente e independiente C3.Dependencia funcional/estadística/independencia C4.Correlación CI1.Variable estadística, valores, rango de la variable, tipos (discreta, continua) CI2.Distribución de una variable estadística CI5.Población (censo) - muestra CI6 Números enteros y decimales, porcentajes
Propiedades	PP25a. Una correlación intensa entre variables no implica una causalidad entre ellas: PP18. Dispersión de la nube, correlación y regresión: si los puntos se sitúan sobre una curva nos encontramos ante una dependencia funcional.
Argumentos	A1. Análisis de ejemplos/ contraejemplos A2. Uso de Gráfico A5. Verbal - deductivo
Procedimientos	PC2.Representación gráfica PC3.Traducción entre representaciones: gráfica, tabular y numérica PC4.Cálculo e interpretación correlación: tipos de covariación PC4.Estimación de la correlación PC7.Uso de tecnología: recursos orientados a bases de datos de la ONU PCI1.Lectura de gráficos

El formador muestra cómo analizar en forma dinámica la evolución de todo ello en el tiempo, a lo largo de 100 años (Figura 5.8.5), y los futuros profesores observan la variación de la Esperanza de vida, y sobre todo, de la Población con estudios secundarios en los dos países comparados. Para el segundo año, se propone como actividad de ampliación representar algunas otras variables, y obtener sus conclusiones. Explicó para ello, cómo se pueden cambiar las variables en el gráfico.

En la Tabla 5.8.6 se presentan los objetos matemáticos implicados en el desarrollo esta Actividad (y la actividad 10). Se presenta gran riqueza de objetos matemáticos, principalmente los procedimientos de interpretación y traducción de gráficos, conceptos y propiedades que utilizarán los futuros profesores en su desempeño.

5.8.3.2.SOLUCIONES APORTADAS POR ALGUNOS PARTICIPANTES

Once participantes (AG, CM, GO, JPA, IE, MA, MG, MMG, PG, RS, VC,), en el segundo año, utilizaron este recurso de visualización para obtener conclusiones sobre los datos, mostrando un razonamiento multivariante, como se aprecia en el siguiente ejemplo de CM (que completa con los gráficos mostrados en la Figura 5.8.6, y otros complementarios, así como algunas tablas de datos).

Este futuro profesor es capaz de obtener conclusiones sobre la evolución de la Esperanza de vida, y otras variables globalmente; y también para algunos países, pues posteriormente señala, tanto las excepciones a la regla como los casos de países que más han progresado.

> He realizado un diagrama de burbujas con las esperanza de vida en el eje Y y el Índice de Desarrollo Humano en el eje X; luego, en el color de las burbujas la tasa de mortalidad en niños menores de cinco años y en el tamaño, la población total, ya que como hemos visto en el ejercicio 1 son las variables que mejor la describen. Por lo tanto, nos aportarán más información. Gracias a la visualización de los datos que nos permite hacer esta página se observa como la esperanza de vida en España ha ido aumentando a medida que pasan los años, al igual que el Índice de Desarrollo Humano. En particular, podemos observar en los diagramas siguientes las diferencias entre 1990 y 2010 La esperanza de vida en 1990 era de 77 años con una tasa de mortalidad de niños menores de cinco años de un 1.1% y un IDH de 0.76 mientras que en 2010 la esperanza de vida era de aproximadamente 81 años (81.2 para ser exactos) con tasa de mortalidad de niños de un 0.5% y un IDH de 0.88. (CM).

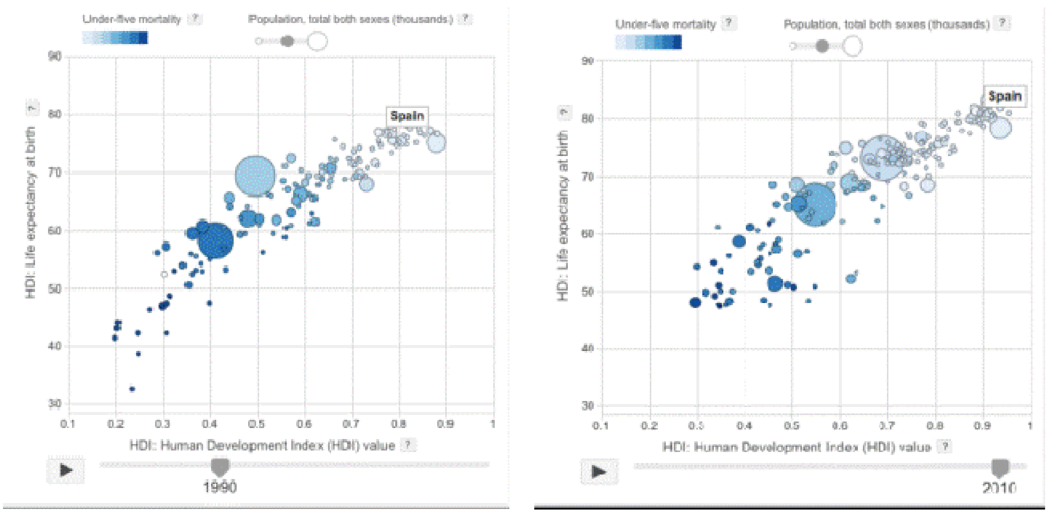


Figura 5.8.6. Gráficos aportados por CM

Igualmente ocurre en otros casos, como por ejemplo CO, quien analiza la relación entre la Esperanza de vida y el Gasto público en salud, utilizando como variables explicativas la Población total y la Tasa de mortalidad infantil, y variando a lo largo del tiempo. Este futuro profesor es capaz de analizar los datos, concluyendo la variación de cada una de las variables estudiadas con el tiempo, no solamente globalmente, sino también identificando los periodos de tiempo en que hay mayor variación. Es capaz también de realizar un análisis crítico, identificando otras posibles variables explicativas no consideradas en el gráfico:

> A la vista de estos datos, se observa el incremento experimentado en la esperanza de vida, el gasto en sanidad y el total de la población mientras que la tasa de mortalidad infantil descendía.
> Si se observa la evolución de los datos, se evidencia un progresivo aumento de la esperanza de vida, que discurre a la par con el aumento del gasto público en Sanidad, el cual experimenta un gran crecimiento en los años 2.006, 2.007 y 2.008; decelerándose e incluso experimentando un leve descenso en los años 2.009 y 2.010 achacable al inicio de la crisis económica.
> Esta dinámica influye en la esperanza de vida y en el descenso de la tasa de mortalidad infantil, que en los últimos años han seguido con su tendencia creciente y decreciente respectivamente, si bien se hace patente una desaceleración en dicha tendencia en los últimos años.
> En lo referente a la población total, se observa que su crecimiento ha sido sostenido a lo largo de toda la serie cronológica, sin que le afectara en un grado apreciable el rápido incremento del gasto público en Sanidad, lo cual es explicable si se tienen, además, otros factores no estudiados en el diagrama, tales como los movimientos migratorios de la población (CO).

En la Tabla 5.8.7 se resumen las visualizaciones de los participantes. En general, presentaron un informe incluyendo varias copias del gráfico de burbujas, estudiado en diferentes momentos del tiempo para apoyar su argumentación. Al trabajar con el gráfico verían, en forma continua, la variación en los 100 años; para poder comunicar sus impresiones, hicieron varias grabaciones de la pantalla para copiarlas en el informe.

Tabla 5.8.7. Variables representadas en el gráfico de burbujas

V. Dependiente	V. independiente	Tiempo	Color	Tamaño
Esperanza de vida (CM)	Índice desarrollo humano	si	TM	PT
Esperanza de vida (MG, MMG)	Índice de educación	si	ID	PT
Esperanza de vida (IE, RS,)	Años escolarización	si	IDM	PT
Esperanza de vida (AG)	Años escolarización	si	PT	ID
Esperanza de vida (GO)	Gasto Público en salud	si	TM	PT
Esperanza de vida (AA, MA)	Gasto Público en salud	si	ID	PT
Esperanza de vida (AA)	Mortalidad mujeres adultas	si	ID	PT
Esperanza de vida (JPA, VC)	Mortalidad infantil	si	ID	PT
Esperanza de vida (AA)	Usuarios de internet	si	ID	PT
Esperanza de vida (PG)	Mortalidad adulta masculina	si	Ratio suicidio	Muerte por cólera
Esperanza de vida (CC)	Producto interior bruto	si	TM	PT
Esperanza de vida (CC)	% población empleada	si	TM	PT
Esperanza de vida (CC)	% de superficie cultivada	si	TM	PT
Esperanza de vida (JP)	Población graduada en ciencias e ingeniería	si	ID	Tasa de emigración
Índice desarrollo humano(IE)	Esperanza de vida	si	PIB	PT
Años esperados escolarización (HD)	Esperanza de vida	si	ID	PT

Hubo casos aislados en que se eligieron variables difíciles de relacionar en el mismo gráfico, por ejemplo, PG relaciona la Esperanza de vida con la mortalidad masculina, y usa en el color y tamaño la ratio de suicidios y muerte por cólera, variando el tiempo. La única conclusión que saca este futuro profesor es que la Esperanza de vida aumenta a lo largo del siglo, pero no comenta la fuerte relación decreciente con la tasa de mortalidad masculina; ni con la ratio de suicidios, que se observan fácilmente en la gráfica. No se aprecia bien en la misma el efecto de las muertes por cólera.

Encontramos también futuros profesores que expresan en su informe su opinión sobre las posibilidades educativas del trabajo con datos reales, como es el caso de ME, quien no usa directamente el gráfico de burbujas, pero si otros gráficos proporcionados por el servidor de las Naciones Unidas, y representa en ellos múltiples variables, proporcionando un informe detallado de conclusiones. Al final de su informe hace la siguiente valoración global:

> Como conclusión, la web de las Naciones Unidas con la que hemos trabajado permite obtener multitud de datos interesantes para impartir clases de estadística en Secundaria (últimos cursos) y Bachillerato. No sólo por la gran cantidad de datos que proporcionan, sino también por la posibilidad de trabajar con estos datos matemáticamente y posteriormente razonar los resultados en un contexto público. Además facilitan la introducción de herramientas tecnológicas como la hoja de cálculo y la gestión del aprendizaje propio por parte del alumno en las tareas de investigación o proyectos que se les puedan encomendar.
> Por último me gustaría destacar la transversalidad de este tipo de tareas que pueden gestionarse en común con profesorado de otras áreas como las ciencias sociales (ME).

5.8.4. TRABAJOS CON RECURSOS EN INTERNET

5.8.4.1. ANÁLISIS A PRIORI Y DESARROLLO

Se plantea una tarea en la que los futuros profesores pueden explorar distintos applets, y deberán indicar sus ventajas y limitaciones, así como su potencial, en el estudio de la correlación y regresión.

12. Recursos en Internet. Explora algunos de los applets (estos u otros) que pueden usarse en el estudio de la correlación y regresión y prepara un informe explicando cómo los usarías en clase y qué se puede aprender con ellos.
    - Educación Navarra: http://docentes.educacion.navarra.es/msadaall/geogebra/figuras/e3regresion.htm
    - Comparación y predicción del coeficiente de correlación: http://www.math.usu.edu/~schneit/CTIS/scorrelation/
    - Trazado de la línea de regresión: http://bcs.whfreeman.com/ips4e/cat_010/applets/CorrelationRegression.html

Esta actividad pide al futuro profesor explorar la posibilidad que algunos applets tienen en el estudio de la correlación y regresión, tomados de una selección más amplia que se llevó a cabo mediante la búsqueda en Internet.

Solución experta y objetos matemáticos implicados

Con el uso de los diferentes applets, se permite al futuro profesor disponer de variedad de situaciones-problema, con la posibilidad de validar procedimientos, o evidenciar propiedades relativas a las nociones de correlación y regresión. Esto es así, ya que de manera casi instantánea, mediante la manipulación interactiva de sus diferentes objetos matemáticos (básicamente puntos/datos bidimensionales y trazado de rectas), se dispone de gran variedad de diagramas de dispersión, que visualizan y permiten estos tratamientos.

Para describir cada uno de estos applets, y dado que su mayoría se diseñan para el tratamiento conjunto de la correlación y regresión, utilizamos la Tabla 5.8.8, donde se incluyen los objetos matemáticos implicados, clasificados según los applets, cuya lectura nos permite realizar comparaciones entre ellos (filas) a la vez que describirlos (columnas).

El primero de ellos (Figura 5.8.7) representa gráficamente, mediante cuadrados sombreados, los cuadrados de las diferencias entre las ordenadas de la recta de regresión y la de los puntos observados. Permite ajustar la recta de regresión, “a ojo”, o automáticamente mediante el sistema; representa también el centro de gravedad y el coeficiente de correlación. Es posible trabajar con datos generados por el sistema o introducir los propios datos.

En el segundo applet (Figura 5.8.8) tiene distintas opciones. La figura mencionada corresponde a la opción de elegir un diagrama de dispersión entre cuatro posibles, para un valor del coeficiente de correlación. Otras opciones son las siguientes:

- Variando el valor del coeficiente de correlación, se obtienen nubes de puntos con dicho valor del coeficiente, bien en forma de animación (varía aleatoriamente el valor del r) o a petición del usuario. Se puede también fijar un valor de r, cambiando la desviación típica de las variables X e Y, observando cómo afectan éstas a la tendencia de la nube de puntos, para r fijo.

- Se pueden comparar sobre la misma pantalla hasta cuatro diagramas de dispersión simultáneos.

- Otras opciones permiten ordenar nubes de puntos, asignando un coeficiente de correlación de entre los propuestos para cada gráfico, con tres niveles de dificultad.

El tercero de los applets propuestos (Figura 5.8.9), proporciona una pantalla en la que se pueden añadir puntos, pinchando en la posición deseada con el ratón. A cada punto que se añade, varía el valor del coeficiente de correlación, con lo cual, puede observarse el efecto de cada punto sobre dicho valor.

Se puede añadir la recta de mínimos cuadrados, y dos rectas perpendiculares que pasan por el centro de gravedad; al seguir añadiendo o moviendo puntos de lugar, se observa también el efecto de estos cambios sobre estas tres rectas.

Como podemos observar en la Tabla 5.8.8, el desarrollo de esta Actividad 11 moviliza casi la totalidad de objetos matemáticos que conforman el significado de referencia de nuestro estudio, descrito en el Estudio 1 (Capítulo 4). Destacamos la abundancia de propiedades que el alumno moviliza, y la relevancia del procedimiento de traducción entre representaciones, presente en todos los applets, aunque difiera de unos a otros. Además, la argumentación: verbal y de ductiva; y basada en casos/ejemplos, se comparte en todos

Tabla 5.8.8. Objetos matemáticos implicados en las actividades con Applets (Tarea 12)

	Objetos matemáticos implicados en cada applet	AP1	AP2	AP3
Lenguaje	L1.Verbal (escrito y oral)	x	x	x
	L2.Simbólico	x	x	x
	L3.Tabular	x	x	x
	L4.Gráfico	x	x	x
	L5.Numérico	x	x	x
Conceptos	C1.Variable estadística bidimensional; distribución	x	x	x
	C3.Dependencia funcional/estadística/independencia	x	x	x
	C4.Covarianza y/o correlación			x
	C5.Regresión	x		x
	C6.Coeficiente de determinación			
	CI1.Variable estadística, valores, rango tipos intervalos	x		x
	CI3.Medidas de tendencia central y posición	x		x
	CI4.Medidas de dispersión: desviación típica			
	CI6.Funciones			x
Propiedades	PP2. Centro de gravedad	x		x
	PP5. Covarianza: signo; puntos en cuadrantes; datos atípicos			x
	PP6. Correlación: signo y rango de variación	x		x
	PP7.Intensidad de la dependencia	x		x
	PP10.Recta de mínimos cuadrados	x		x
	PP11.Regresión lineal y no lineal	x		
	PP12.Propiedades de las estimaciones			x
	PP13.Dos rectas diferentes; diferencia v. dependiente/independiente			
	PP14.Centro de gravedad y recta de regresión	x		x
	PP15.Significado de los coeficientes	x		
	PP16.Signo de correlación, signo covarianza y pendiente rectas de regresión	x		x
	PP17.Correlación y regresión	x		x
	PP18. Dispersión de la nube, correlación y regresión	x		x
	PP20.Estimaciones y coeficientes			x
Procedimientos	A2.Análisis de ejemplos	x	x	x
	A3.Visualizaciones	x		
	A4.Gráfico	x		
	PC1.Representación tabular	x		
	PC2.Representación gráfica	x	x	x
	PC3.Traducción entre representaciones	x	x	x
	PC4.Cálculo e interpretación covarianza y/o correlación	x		x
	PC5.Cálculo e interpretación de modelos de regresión	x		x
	PC6.Predicción a través de modelos de regresión	x		x
	PC7.Uso de tecnología	x	x	x
Prob.	P0.Organización/Reducción de datos	x	x	x
	P1.Analizar la existencia de relación entre variables	x	x	x
	P2.Predecir una variable en función de otra	x		x

.

5.8.4.2.SOLUCIONES APORTADAS POR ALGUNOS PARTICIPANTES

Fueron once los participantes (AA, AG, CC, CM, GO, HD, IE, JPA, ME, RS, VC) que realizaron esta actividad de ampliación, presentando todos ello el análisis de los tres applets (Tabla 5.8.9). Así, por ejemplo, IE analiza el primero de los applets propuestos (Figura 5.8.7), donde ha elegido las opciones de representar la recta de regresión y el coeficiente de correlación. Sugiere las siguientes actividades para llevar a cabo en el aula, que muestra su conocimiento de la faceta mediacional e interaccional en este tema:

> En clase podrían plantearse distintas actividades como:
> - Introducir los datos de una tabla y observar el coeficiente de correlación que se obtiene.
> - Modificar los datos de forma que el coeficiente de correlación sea mayor de 0.95.

- Introducir una recta de regresión que se ajuste lo máximo posible a la nube de puntos, teniendo en cuenta la suma del área de los cuadrados, y observar posteriormente si es la correcta o no.
- Realizar una previsión para nuevos datos a partir de la recta de regresión (IE).

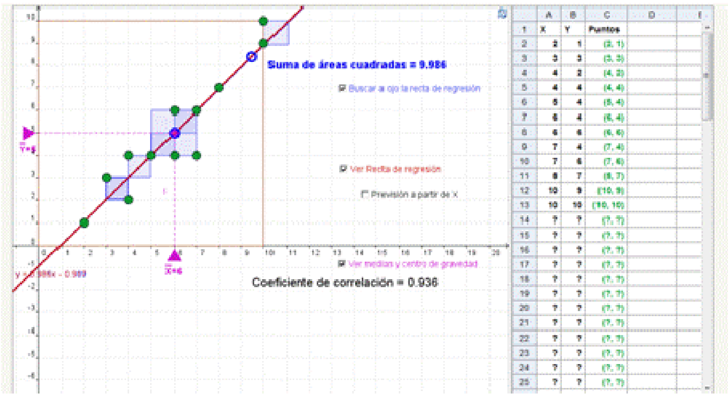


Figura 5.8.7. Pantalla aportada por IE en su análisis del primer applet

Respecto al segundo applet, IE presenta varias pantallas, una de las cuales reproducimos en la Figura 5.8.8; sus comentarios se exponen a continuación. Observamos la madurez de este futuro profesor pues, por un lado, es capaz de sugerir una amplia variedad de actividades que podrían trabajarse en la clase con ayuda del applet; y por otro lado, es capaz de percibir su limitación, pues se reduce al estudio de las propiedades del coeficiente de correlación, pero no se relaciona con la regresión, luego no ayuda a percibir la utilidad del estudio de la correlación.

> Este Applet ayuda a los alumnos a comprender el significado del coeficiente de correlación y le ayuda en la predicción de dicho coeficiente. En clase podrían plantearse distintas actividades tales como: a) Obtener gráficas con distintos coeficientes de correlación; b) - Descubrir el coeficiente de correlación de distintas gráficas, dadas o no distintas opciones de respuesta, c)- Descubrir gráficas con el mismo coeficiente de correlación.
> Este Applet es interesante para que los alumnos comprendan qué significa el coeficiente de correlación. Sin embargo, no lo considero demasiado útil, pues sólo se limita a eso. Es decir, puede inducir al alumnado a descubrir un valor en función de que los puntos estén más juntos o no. Pero no a que tomen conciencia de para qué puede usarse, ni comprender que los puntos son datos estadísticos y que el coeficiente de correlación te ayuda a comprender qué relación hay entre ellos. Considero, que aunque tiene bastante variables y eso me parece positivo, es demasiado abstracto y se reduce sólo al uso de un concepto.

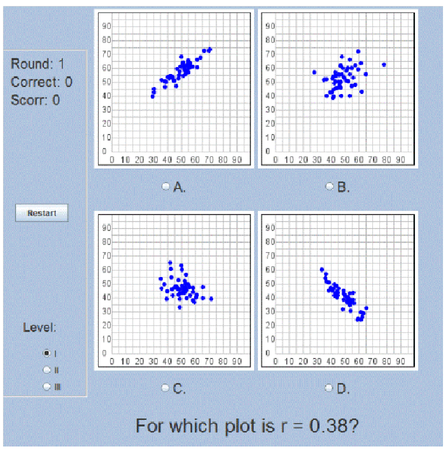


Figura 5.8.8. Pantalla aportada por IE en su análisis del segundo applet

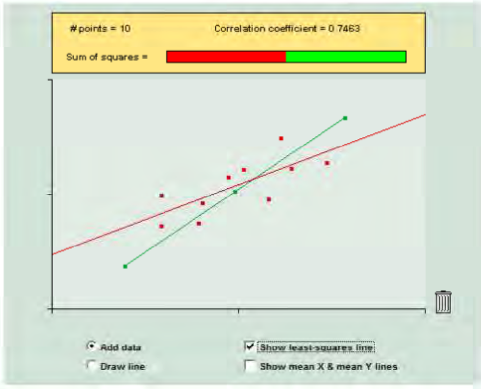


Figura 5.8.9. Pantalla aportada por IE en su análisis del tercer applet

La alumna IE también analiza el tercer applet (Figura 5.8.9), mostrando varias pantallas, con diferentes conjuntos de puntos, y el dibujo de las rectas: la de regresión y una trazada "a ojo", con diferentes conjuntos de datos. Su comentario va en la misma línea del anterior. Considera que el applet es útil, sobre todo para el estudio formal del tema, pero no para mostrar a los alumnos la utilidad de la regresión en el trabajo con datos reales.

> Este Applet ayuda a los alumnos a comprender el significado de la recta de regresión. En clase podrían plantearse distintas actividades tales como: Realizar una nube de puntos y descubrir la recta de regresión o Realizar una recta y situar un nube de puntos cuya recta de regresión sea la recta dibujada. Pero añado lo mismo que en el Applet anterior, y es que es demasiado abstracto. Se pierde el verdadero significado de esto: la lectura de gráficas, la predicción de datos mediante la recta de regresión, etc.

Tabla 5.8.9. Análisis de applets realizados por algunos futuros profesores

Aplet comentado	Describe sus posibilidades	Sugiere actividades para el aula	Critica sus limitaciones
Primer aplet	AA, AG, CC, CM, GO, HD, IE, JP, JPA, ME, RS, VC	AA, AG, CM, HD, IE, ME, JP, JPA, RS, VC	IE
Segundo applet	AA, AG, CC, CM, GO, HD, IE, JP, JPA, ME,RS, VC	AA, AG, CM, HD,IE, ME, JP, JPA, RS, VC	IE
Tercer applet	AA, AG, CC, CM, GO, IE, JP, JPA, ME, RS, VC	AA, AG, CM, IE, ME, JP, JPA, RS, VC	IE

Hay casos aislados (que no aparece en la Tabla 5.8.9) como PG, que en vez de analizar estos applets buscan otros alternativos. Así, PG explora un applet de la página de Bioestadística (http://e-stadistica.bio.ucm.es/mod_regresion/regresion_applet.html) describiendo su contenido en la forma siguiente:

> Se trata de un applet en Java en el que se trata de encontrar la recta de regresión, junto con su gráfica, que mejor se ajuste o mejor aproxime a una serie de datos. Una vez introducidos los datos, pulsando el botón Calcular situado en la izquierda de la ventana, la aplicación representará gráficamente la colección de puntos de los datos, establecerá la recta de regresión maestral y otros datos de interés estimación puntual de la varianza, coeficientes de correlación y de determinación múltiple, análisis de la varianza de la regresión. Esta aplicación me resulta interesante por su facilidad de implantar en el aula, pues los conocimientos previos que debe tener el alumnado son mínimos y no les va a presentar demasiadas dificultades.

5.9. CONCLUSIONES SOBRE LA EVALUACIÓN Y DESARROLLO DEL CONOCIMIENTO MATEMÁTICO DE LOS FUTUROS PROFESORES

A lo largo del capítulo, se han descrito las actividades de evaluación y desarrollo llevadas a cabo durante dos cursos sucesivos, con muestras de futuros profesores. Para finalizar este estudio, se aportan a continuación las conclusiones obtenidas en relación a sus objetivos e hipótesis.

Conclusiones respecto a los objetivos

Nos planteamos inicialmente dos objetivos, el primero de los cuáles se enunció en la forma siguiente:

O2.1. Elaborar una situación didáctica e instrumentos de evaluación dirigidos a futuros profesores, que permita contextualizar la correlación y regresión, nos proporcione información sobre sus conocimientos matemáticos (comunes y avanzados), y contribuya a desarrollarlos.

Para cumplir este objetivo, se diseñó el proyecto trabajado con los futuros profesores que, además, permite en su primera parte obtener datos de evaluación. En las primeras secciones del capítulo se ha descrito su forma de construcción, y a lo largo del mismo se han analizado las actividades que lo componen. Siguiendo la técnica de investigación basada en diseño, hemos producido un material didáctico dirigido a la evaluación y formación de futuros profesores, que puede utilizarse en otras experiencias de formación. Será importante, siguiendo los criterios de este tipo de investigación, valorar la ingeniería producida. Usaremos los criterios de idoneidad didáctica (Wilhelmi, Font y Godino, 2005; Godino, Contreras y Font, 2006) para ello en el siguiente capítulo.

Una primera conclusión respecto a este objetivo, es la alta *idoneidad epistémica* de este proyecto, debido a la representatividad del significado institucional de referencia que se definió en el Estudio 1, respecto al implementado en el proyecto (ver Tabla 5.8.10). Observamos que se recogen los tres campos principales de problemas, identificados en el significado de referencia, lenguaje verbal (simbólico, gráfico y numérico), los principales conceptos y proposiciones, así como argumentos variados. Asimismo, los procedimientos incluyen métodos tradicionales (como la lectura e interpretación de gráficos) y otros apoyados en la tecnología.

El proyecto tuvo una *idoneidad cognitiva adecuada*, como se mostró en los buenos resultados de la evaluación en una amplia parte de este significado, tanto mediante las respuestas escritas de los participantes en el estudio, como para, en algunos casos, las actividades de ampliación llevadas a cabo voluntariamente el segundo año de la experiencia. Como se verá en el Estudio 3, la *idoneidad afectiva* fue alta, como así lo manifiestan en sus respuestas de análisis didáctico los futuros profesores, quienes valoran la motivación del trabajo con datos reales y con tecnología, así como por el contexto utilizado.

El proyecto se pudo completar en el tiempo previsto, y los recursos utilizados (hoja Excel, datos y herramientas de visualizaciones de las Naciones Unidas y applets) estuvieron disponibles para los alumnos, y su uso fue sencillo, lo que contribuye a la idoneidad mediacional.

Tabla 5.8.10. Significado institucional implementado en el proyecto

Problemas	P0. Organización/Representación de datos	a1, a2, a3, a5, a9
	P1.Analizar la relación entre variables	a5, a6, a9
	P2.Predecir una variable en función de otra	a7,a8
Lenguaje	L1.Verbal (escrito y oral)	todas
	L2.Simbólico: coeficiente de determinación	a2, a3, a5, a7, a8, a9
	L3.Tabular	a1, a3, a7,a9
	L4.Gráfico; L5.Numérico	todas
Conceptos	C1.Variable estadística bidimensional; distribución	a5, a7,a6, a8, a9
	C3.Dependencia funcional/estadística/independencia	a5, a6, a7, a8, a8
	C4.Covarianza y/o correlación	a5, a6, a7, a8, a9
	C5.Regresión	a7, a8,a9
	C6.Coeficiente de determinación. Bondad de ajuste	a8,a9
	CI1.Variable estadística, valores, rango, tipos; intervalos, extremos, marcas de clase	a1,a3,a5,a9
	CI2.Distribución de una variable estadística	a1, a2, a3, a9
	CI3.Medidas de tendencia central y posición	a1,a2,a3
	CI4.Medidas de dispersión: desviación típica	a1
	CI5.Población (censo) - muestra	a9
	CI6. Funciones: lineal, polinómica, logarítmica, potencial, exponencial; significado parámetros; representación gráfica	a5, a7.a8,a9
	CI7.Números enteros y decimales, porcentajes	a1,a9
Proposiciones	PPI2. Relación entre frecuencia absolutas, acumulada, porcentaje	a1
	PPI3. Propiedades de media y mediana	a2. a3
	PP25a. Una correlación intensa no implica causalidad	a6, a9
	PP26a. El coeficiente de correlación mide la correlación lineal, interpretación del valor absoluto	a5, a7, a8
	PP26b. Interpretación del signo del coef. de correlación lineal.	a5
	PP32a. La recta de regresión hace mínima la suma de cuadrados	a8,a9
	PP33. El coeficiente de determinación informa de la varianza de la v. dependiente explicada por el modelo y la bondad del ajuste	a5, a7, a8, a9
	PP45 Relación covarianza, correlación, regresión.	a7, a8
	PP18. Dispersión de la nube, correlación y regresión	a5, a7, a8, a9
	PP7. El coeficiente de determinación informa del grado de ajuste del modelo y de la varianza explicada: varía entre 0 y 1	a8, a9
Procedimientos	PC1.Representación tabular	a7
	PC2.Representación gráfica	a5, a7, a8, a9
	PC3.Traducción entre representaciones	a1, a2, a3, a5, a9
	PC4.Cálculo e interpretación correlación o covarianza	a5, a6, a7, a8, a9
	PC5.Cálculo e interpretación de modelos de regresión	a7, a8,a9
	PC6.Predicción a través de modelos de regresión	a7, a8
	PC7.Uso de la tecnología: uso de Excel.	a3, a8, a9
	PCI1.Lectura de gráficos	a1, a2, a3, a9
	PCI2.Lectura de tablas	a1, a3, a9
Argumentos	A1. Ejemplos-contraejemplos	a1,a2, a5,a6,a7, a8,a9
	A2. Uso de representaciones gráficas	Todas
	A3. Verbal – deductivo	a1,a2,a3 a5, a6, a8,a9
	A5.Deductivo-algebraico	a7

La *idoneidad interaccional* se podría mejorar si se dispusiese de mayor tiempo de discusión de las soluciones. No obstante, el análisis de las respuestas escritas ha permitido detectar conflictos. Los visibles en la primera muestra se trataron de solucionar en la segunda, con éxito en algunos puntos, como la interpretación de algunos de los gráficos. Finalmente, pensamos que el proyecto tiene idoneidad ecológica pues responde a las directrices curriculares, tiene un componente de innovación docente, y conecta con problemas sociales y otras materias de interés del alumno. Todos estos puntos se analizan con detalle en el Estudio 3, en el que son los propios participantes quienes realizan un análisis de la idoneidad de este proyecto.

O2.2.Realizar un estudio exploratorio de evaluación de los conocimientos matemáticos (comunes y avanzados) en futuros profesores, utilizando la situación didáctica e instrumento de evaluación construidos para tal efecto.

A lo largo del capítulo se han proporcionado análisis detallados de las respuestas escritas de los futuros profesores en los dos primeros grupos de actividades del proyecto. Dichas respuestas han permitido describir el conocimiento común (y parte del avanzado) de los participantes sobre los siguientes puntos:

- Gráficos estadísticos elementales, lectura e interpretación de los mismos: gráfico de caja, histograma y diagrama acumulativo. (Conocimiento común).
- Elección de un promedio representativo y determinación de un percentil a partir de una representación gráfica. (Conocimiento común).
- Estimación de la correlación a partir de diagramas de dispersión, (Conocimiento común) y efecto de diversas variables de tarea sobre la estimación (Conocimiento común y avanzado). Ordenación y asignación de coeficientes de correlación. (Conocimiento común).
- Determinación intuitiva de la mejor función de ajuste a los datos; valoración intuitiva del orden de predicción de diversas variables. (Conocimiento común).

Asimismo, algunas actividades escritas, así como las actividades de ampliación opcionales realizadas por algunos futuros profesores el segundo año de la experiencia han permitido evaluar los siguientes conocimientos matemáticos avanzados, el primero de ellos en toda la muestra, y el resto en un pequeño número de participantes:

- Explicación de la correlación en términos de relaciones causales o de otro tipo de covariación.
- Ajuste de funciones a los datos utilizando Excel. Interpretación de modelos de ajuste.
- Selección de nuevos datos desde Internet y propuesta de otros análisis de correlación y regresión.
- Visualización de datos multidimensionales en el gráfico de burbujas dinámico de las Naciones Unidas y obtención de conclusiones sobre los mismos.

Adicionalmente, el análisis de applets en Internet y propuesta de posible uso de los mismos en la clase de Bachillerato, ha permitido evaluar y desarrollar algunos conocimientos didácticos de un grupo reducido de futuros profesores.

Conclusiones respecto a las hipótesis

Aunque de carácter exploratorio, quisimos establecer de manera informal algunas hipótesis sobre los resultados de la experiencia, que discutimos a continuación.

H1. Al evaluar los conocimiento de los futuros profesores *se prevé encontrar algunas dificultades en conceptos estadísticos elementales requeridos para el trabajo con la correlación y regresión.*

En general, los futuros profesores han mostrado un buen conocimiento de los gráficos estadísticos utilizados en el proyecto: histograma, polígono acumulativo de frecuencias, gráfico de la caja y diagrama de dispersión, que supieron interpretar y contextualizar respecto a la investigación presentada en el proyecto. En este sentido, los futuros profesores mostraron niveles adecuados de lectura de gráficos (Curcio, 1989), fueron capaces de integrar la estadística con el contexto, una de las habilidades de razonamiento estadístico descritas por Wild y Pfannkuch (1999).

Igualmente es fuerte el conocimiento de las ideas de dependencia funcional y aleatoria, diversos tipos de funciones y sus parámetros, correlación y su interpretación, regresión y su utilidad, así como significados de los parámetros de diversos modelos de regresión, incluida la lineal.

Paradójicamente, y confirmando en parte esta hipótesis, donde más dificultades hemos encontrado ha sido en los conceptos estadísticos pre-requisitos, es decir, en los gráficos y resúmenes estadísticos, aunque las dificultades fueron menores que en otras realizadas con profesores de educación primaria, como la de Jacobbe (2008). Algunas de las dificultades observadas han sido las siguientes:

- Confusión entre frecuencia ordinaria y acumulada en la lectura del diagrama de frecuencias acumuladas.
- Confusión entre promedio y valor de la variable, ya mostrado en la investigación de Cobo (2007).
- Dificultad en contextualizar la variable en el histograma y diagrama de frecuencias acumuladas, mostrando confusión entre frecuencia y valor de la variable. Sobre todo en las interpretaciones de la pendiente, el máximo y el mínimo.
- Falta de criterio para elegir un promedio representativo, llegando a elegir automáticamente la media, incluso en distribuciones asimétricas.
- Interpretar la desviación típica como un estadístico que mide la asimetría o valorar inadecuadamente el valor de una desviación típica.
- Interpretación incorrecta del diagrama de la caja, que también aparece en Pfannkuch (2006).
- Calcular incorrectamente o interpretar incorrectamente un percentil.
- Confundir variable independiente y dependiente en el estudio de la regresión, error señalado por Ruiz (2006) en su estudio de la variable aleatoria.

Igualmente se han encontrado algunas creencias erróneas sobre la causalidad, como identificarla con función lineal. No obstante, todos estos problemas sólo se han encontrado en una proporción relativamente pequeña, inferior en todos los casos al 25% de los futuros profesores.

H2. Se espera que los futuros profesores se interesen por el trabajo con datos reales y tecnología en la actividad planteada, y ésta permita desarrollar su conocimiento matemático (común y avanzado) sobre la correlación y regresión.

Esta hipótesis se ha confirmado, tanto por el estudio de evaluación, como por sus resultados, pues la mayoría de los participantes han dado estimaciones razonables del coeficiente de correlación, asignando igualmente uno adecuado a un diagrama de dispersión, y han sido capaces de diferenciar relaciones causales y no causales.

Asimismo, las actividades de ampliación llevadas a cabo por un grupo reducido de participantes en el segundo año han mostrado el aprendizaje del uso de las herramientas de Excel para el análisis de regresión. Estos futuros profesores fueron capaces de plantear nuevos problemas de correlación con variables de su interés, analizarlas, y encontrar un modelo de ajuste con Excel; además de obtener conclusiones sobre la bondad del ajuste de diferentes variables independientes para una misma variable dependiente.

Otros obtuvieron y analizaron nuevos datos, a partir del servidor de las Naciones Unidas, llegando a la obtención de conclusiones a partir del análisis estadístico, y concluyendo por tanto un ciclo completo de modelización. Igualmente se ha mostrado la capacidad de interpretación de visualizaciones dinámicas, y la obtención de conclusiones a partir de datos multivariantes, manejando simultáneamente cinco variables, y detectando las tendencias de cada una. Incluso en algunos casos, se ha llegado a detectar mediante el análisis, excepciones a esas tendencias globales, y a exponer hipótesis explicativas de las mismas. Finalmente, el análisis de los applets propuestos por algunos de los participantes ha mostrado sus conocimientos didácticos, al sugerir modos de uso en el aula, o incluso criticar sus limitaciones.

CAPÍTULO 6.

CONOCIMIENTO DIDÁCTICO DE FUTUROS PROFESORES DE EDUCACIÓN SECUNDARIA Y BACHILLERATO

6.1. INTRODUCCIÓN

En este capítulo se describe el Estudio 3, orientado a evaluar y desarrollar las distintas facetas del componente didáctico del conocimiento didáctico-matemático sobre la correlación y regresión (Godino, 2009; 2011), en una muestra de futuros profesores de Educación Secundaria y Bachillerato. Como se ha especificado en la última sección del Capítulo 2, algunas de estas facetas pueden asimilarse a categorías del modelo MKT (Hill, Ball y Schilling, 2008).

Más concretamente, nos centramos en el análisis de las respuestas de estos futuros profesores a una actividad de análisis de la idoneidad didáctica (Godino, Bencomo, Font y Wilhelmi, 2006) del proyecto trabajado por ellos mismos en el Estudio 2. Esta nueva actividad, fue realizada en una sesión posterior del mismo curso (tercera sesión) y, una vez discutidas colectivamente, mediante un debate con el formador, las soluciones al proyecto (la descripción detallada de estas sesiones se presenta en el Capítulo 5).

En lo que sigue se describe, en primer lugar, los objetivos e hipótesis del Estudio 3, la muestra participante, el método, y el contexto de la actividad. Seguidamente, se analizan las preguntas formuladas a los futuros profesores, y las respuestas obtenidas, clasificadas en bloques que atienden a los diferentes componentes de la idoneidad didáctica. Se finaliza con las conclusiones obtenidas sobre los objetivos e hipótesis del estudio.

6.2. OBJETIVOS E HIPÓTESIS DEL ESTUDIO 3

El objetivo general al que se responde con este tercer estudio es *realizar un estudio exploratorio del conocimiento didáctico-matemático sobre la correlación y regresión, en estudiantes que se preparan para ser profesores de Matemáticas de Educación Secundaria y Bachillerato*, que se describió en el Capítulo 1. De modo más específico, como se ha indicado anteriormente, podemos distinguir dos objetivos específicos en este Estudio 3:

O3.1.Realizar un estudio exploratorio de evaluación de distintas facetas del componente didáctico del conocimiento didáctico-matemático sobre la correlación y regresión (Godino, 2009; 2011), en una muestra de futuros profesores de Educación Secundaria y Bachillerato, puestos en juego al realizar el análisis de la idoneidad didáctica del proyecto descrito en el Estudio 2. Al mismo tiempo, se trataba de desarrollar dicho conocimiento en los futuros profesores mediante la actividad práctica.

Godino, Font y Wilhelmi (2008) sugieren que la didáctica de la matemática debe aspirar a la mejora de la enseñanza, aportando una racionalidad que permita el análisis y crítica de medios, fines y cambios en los procesos educativos; por ello, introducen sus criterios de “idoneidad”. Los autores indican que la idoneidad didáctica es una herramienta de análisis, que puede ser útil para la formación de profesores, ayudándoles a que adquieran las herramientas que necesitarán para aprender a enseñar. La idoneidad didáctica y las herramientas para su análisis y valoración, establecen, según Godino, Font y Wilhelmi (2008), un puente entre una didáctica descriptiva - explicativa y su aplicación para el diseño, implementación y evaluación de intervenciones educativas específicas. Con la intención de aportar información sobre los conocimientos didácticos de los futuros profesores de Educación Secundaria y Bachillerato, se utilizó en nuestro trabajo la Guía de Análisis de la Idoneidad Didáctica (Godino, 2011), y en lo que sigue, valoramos la aplicación de la misma por los futuros profesores en el estudio.

Los componentes y descriptores de la idoneidad didáctica fueron introducidos en el curso que seguían los estudiantes en forma teórica, que se conjugaba con la discusión de ejemplos en otros bloques de contenidos del área de matemáticas. Con ello, se trataba de proporcionar a los futuros profesores una herramienta de análisis de sus proyectos de innovación didáctica, sus unidades didácticas, o su práctica docente. Concretamente se esperaba que aplicasen esta herramienta en la valoración de las unidades didácticas o experiencias docentes incluidas en sus Memorias de Prácticas, y/o sus Memorias Fin de Máster.

O3.2.Identificar y clasificar algunos ejemplos de conocimientos relacionados con cada una de las facetas del conocimiento didáctico en el modelo de Godino (2009, 2011) de los futuros profesores, e interpretarlas en las diferentes categorías del modelo MKT (Hill, Ball y Schilling, 2008).

La importancia de este objetivo se deduce de la escasez de modelos sobre el conocimiento del profesor para enseñar estadística, y todavía más, del análisis de ejemplos de estos conocimientos en el tema particular de correlación y regresión, que se puso de manifiesto en el Capítulo 2. Se trata de aportar información sobre los conocimientos de los futuros profesores de Educación Secundaria y Bachillerato, describiendo algunos elementos necesarios en el conocimiento didáctico para la enseñanza de la correlación y regresión.

En relación a estos objetivos específicos, se han planteado también algunas hipótesis previas, que a continuación exponemos.

Hipótesis del Estudio 3

Aunque el Estudio 3 se realiza sobre una muestra pequeña de futuros profesores, debido a su novedad, tiene carácter exploratorio. Al igual que se hizo en el Estudio 2, queremos aventurar algunas hipótesis sobre lo que se espera encontrar, entendidas, como en aquél caso, como expectativas más que como hipótesis formales. Estas hipótesis son las siguientes:

H1. Se espera que los futuros profesores de Educación Secundaria y Bachillerato sean capaces de analizar la idoneidad didáctica del proceso de estudio de la correlación y regresión experimentado en el Estudio 2, en sus diferentes componentes.

Más concretamente, esperamos que los futuros profesores sean capaces de proporcionar respuestas adecuadas a la mayor parte de los descriptores de los componentes de la idoneidad didáctica en el cuestionario de evaluación que completarán en el Estudio 3. Dicho cuestionario, como se verá, es una adaptación de la guía de análisis de la idoneidad didáctica propuesta por Godino (2009, 2011). Al completarla, los participantes han de poner en juego diferentes componentes de su conocimiento didáctico sobre el tema.

Puesto que dicho conocimiento didáctico depende fuertemente del conocimiento del contenido, según Hill, Ball y Schilling (2008), y dado que los resultados del análisis del conocimiento matemático sobre correlación y regresión en los profesores participantes fueron buenos, se justifica esta expectativa.

H2. Se prevé que el análisis cualitativo de las respuestas en la tarea de análisis de la idoneidad didáctica permita identificar ejemplos de conocimientos didácticos *específicos sobre la correlación y regresión, y sobre estadística, que puedan utilizarse para enriquecer los modelos disponibles de conocimiento didáctico - matemático para la enseñanza del tema.*

Nos basamos, al establecer esta hipótesis, en la capacidad argumentativa y de análisis crítico de los futuros profesores, que ya se puso de manifiesto en el Estudio 2, especialmente al analizar ejemplos de actividades opcionales de ampliación, llevadas a cabo por los participantes en dicho estudio. Esperamos que en el Estudio 3, los participantes apliquen dichas capacidades en sus respuestas, y ello nos permita identificar y clasificar los ejemplos citados.

6.3. MÉTODO

Muestra participante y contexto

En el Estudio 3 participaron los 23 futuros profesores que formaron la muestra 1 (curso 2012-2013) del Estudio 2, y cuyas características específicas se describieron con detalle en el Capítulo 5. Dos de ellos solo pudieron completar la valoración de la idoneidad epistémica, quedando el resto de componentes valorados por 21 futuros profesores.

Esta muestra de futuros profesores realizó una segunda práctica, que continúa a la descrita en el Capítulo 5, que fue realizada dentro de la misma asignatura. Siguiendo el modelo de ciclo formativo propuesto por Godino y Batanero (2008), una vez que los participantes realizaron personalmente el proyecto estadístico, diseñado para sus alumnos de Bachillerato, se siguió su formación realizando una práctica de análisis didáctico. En ella se implementa el modelo de análisis de la idoneidad didáctica de procesos de estudio de las matemáticas (Godino, 2009; 2011), que los profesores en formación pueden adaptar de manera crítica para su futuro trabajo.

Los participantes conocían que el proyecto está pensado para trabajar la correlación y regresión en primer curso de Bachillerato (en las dos modalidades que incluyen el tema). El proceso de estudio seguido fue similar al que podrían seguir los estudiantes de Bachillerato cuando trabajasen el proyecto. En esta segunda práctica, se pidió a los futuros profesores valorar la idoneidad didáctica del proceso de estudio vivido en el aula (incluyendo en el mismo el trabajo extraescolar realizado con las actividades de ampliación). Para completarla, se suministró a los futuros profesores una adaptación de la Guía de Análisis de la Idoneidad Didáctica propuesta por Godino (2011), que contenía diferentes apartados de valoración y descriptores de los mismos (Anexo 2).

Para motivar la práctica, el concepto de idoneidad didáctica se fue trabajando componente a componente en dos sesiones sucesivas (dos horas y media cada una) de la asignatura, mostrando ejemplos de su aplicación para el análisis de otros procesos de estudio. Se pidió a los futuros profesores que fuesen completando, con el mayor detalle posible, los apartados de la pauta, proporcionando también una valoración de los distintos apartados de la idoneidad. Finalmente, se propuso una pregunta en que los futuros profesores habrían de sugerir mejoras globales al proyecto.

Análisis de datos

Recogidos los informes, se procedió al análisis de su contenido, utilizando técnicas cualitativas, y definiendo algunas variables que luego se utilizarán en el análisis estadístico de los datos, con la intención de dar una valoración cuantitativa del conocimiento didáctico-matemático que mostraron los futuros profesores al desarrollar la práctica.

El análisis de contenido es un tipo de análisis cualitativo, fundamentado en la idea de que las unidades del texto pueden clasificarse en un número reducido de categorías (Weber, 1985). Sirve para efectuar inferencias mediante la identificación, sistemática y objetiva, de las características específicas de un texto (Ghiglione y Matalón, 1989). Su objetivo final es la búsqueda del significado implícito en el texto, a partir de un estudio sistemático del mismo. Para realizar el análisis, hemos recurrido a nuestro conocimiento previo sobre el proyecto, adquirido a través del análisis a priori del mismo, y nuestros estudios previos de estadística y didáctica, adquirido a través de la titulación y el Máster en investigación en didáctica de la matemática. También nos hemos apoyado en la revisión de antecedentes, que se recogen en el Capítulo 3.

A partir de los resultados de este análisis previo, hemos resumido el contenido del texto y definido categorías, estudiando su presencia o ausencia en los protocolos escritos de los participantes. Es un análisis directo, puesto que seguimos estrictamente el contenido de la unidad de análisis, sin ir más allá de lo que esta contiene. Para cada pregunta se asignó a cada participante un nivel, atendiendo al conocimiento mostrado en la aplicación del descriptor. Estos niveles desarrollan los propuestos por Arteaga (2011), quien aplicó únicamente los tres primeros, recogiendo en el último los niveles 4 y 5, que nosotros proponemos en nuestro estudio; ya que en su trabajo, al tratarse de futuros profesores de educación primaria, el análisis fue más incompleto. Los niveles de análisis que utilizaremos en todos los descriptores son los siguientes:

0. Cuando el futuro profesor no hace referencia al descriptor. Algunos participantes dejan sus respuestas en blanco en este apartado, no habiendo comprendido la pregunta, o no siendo capaz de aplicarla en este contexto.

1. Se hace referencia, pero se limita a copiar casi literalmente el descriptor, sin aplicarlo al proyecto analizado.

2. Se hace referencia y se aplica el descriptor, pero no se centra específicamente en el proyecto desarrollado, sino en aspectos anecdóticos o no estrictamente matemáticos. Por ejemplo, al valorar los contenidos matemáticos se centra en el contexto utilizado (Esperanza de vida) y no en el propio contenido. También consideramos en esta categoría el caso en que el futuro profesor aplica una parte del descriptor correctamente y otra incorrectamente, por no haberlo interpretado correctamente.

3. El futuro profesor hace referencia y aplica el descriptor a contenidos matemáticos del proyecto y a la forma en que se trabajó con el proyecto en el aula, pero la aplicación es incompleta, porque se centra en un aspecto no recogido por el descriptor. Por ejemplo, si se le pide valorar las situaciones-problema propuestas en el proyecto, y en su lugar, valora una propiedad.

4. Se hace una aplicación correcta del descriptor, utilizando contenidos matemáticos u otros componentes del trabajo en el proyecto (recursos, metodología, etc.) en forma consistente con la pregunta. Razona mediante un único ejemplo. Un caso

sería el alumno que es capaz de identificar correctamente un tipo de problema propuesto en el proyecto cuando se le pide valorar las situaciones-problema.

5. El futuro profesor hace una aplicación correcta y consistente del descriptor, utilizando contenidos matemáticos u otros componentes del trabajo en el proyecto (recursos, metodología, etc.) en forma consistente con la pregunta. Razona mediante dos o más ejemplos. Sería idéntico, pero más completo al caso anterior.

En lo que sigue se presenta la pauta utilizada, que es una adaptación de la Guía de Análisis de la Idoneidad Didáctica propuesta por Godino (2009; 2011), que es la utilizada en la investigación de Arteaga (2011). También se analizan las respuestas, y se informa de los resultados del análisis de las producciones de los participantes. Solo se incluyen ejemplos de los niveles 2 a 5 en la aplicación de los descriptores, pues los dos primeros niveles, o bien son respuestas en blanco, o reproducen el descriptor casi al pie de la letra.

6.4. VALORACIÓN DE LA IDONEIDAD EPISTÉMICA

El primer componente analizado fue la idoneidad epistémica del proyecto, es decir, la representatividad del contenido matemático implementado en el mismo, con respecto al significado institucional pretendido para el tema. Se proporcionó a los futuros profesores la pauta, que reproducimos en la Tabla 6.4.1.

Tabla 6.4.1. Pauta de análisis de la valoración de la idoneidad epistémica (pertinencia del contenido)

		Justificación
Situaciones-problemas	1.1.	Los problemas que se presentan, ¿te parece que son útiles para contextualizar, aplicar y ejercitar los contenidos de correlación y regresión? ¿Por qué?
	1.2.	¿Se proponen situaciones para que el alumno de Bachillerato invente nuevos problemas? ¿Cuáles?
Lenguaje matemático	1.3.	¿Se usan diferentes modos de expresión matemática (verbal, gráfica, simbólica...), traducciones y conversiones entre los mismas? ¿Cuáles?
	1.4.	¿Hay actividades de representación e interpretación del lenguaje matemático? ¿Cuáles?
Conceptos, propiedades, procedimientos	1.5.	¿Qué conceptos, propiedades y procedimientos de la correlación y regresión habría que explicar previamente para trabajar este proyecto?
	1.6.	¿Se proponen tareas donde los alumnos tengan que reconocer definiciones propiedades o procedimientos? ¿Cuáles?
Argumentos	1.7.	¿Son las explicaciones, comprobaciones y demostraciones adecuadas para Bachillerato?
	1.8.	¿Se incluyen situaciones donde el alumno tenga que argumentar? ¿Cuáles?
Relaciones	1.9.	Los contenidos matemáticos (problemas, definiciones, propiedades, etc.) se relacionan y conectan entre sí en el proyecto?

Su aplicación requiere, por parte del futuro profesor, de un conocimiento matemático del contenido común y especializado del tema, pues, además de conocer los objetos matemáticos, debe ser capaz de reconocer su presencia o ausencia en el proceso de estudio propuesto, valorando su pertinencia y adecuación. El objetivo de esta actividad es que los futuros profesores pongan en práctica y desarrollen su conocimiento especializado (Ball, Lubienski y Mewborn, 2001), a la vez que profundicen en su conocimiento común sobre el tema (Godino, 2011).

Para estudiar la idoneidad epistémica, los participantes han de ser capaces de reconocer los conceptos y procedimientos, los distintos lenguajes usados, los tipos de justificaciones, propiedades, así como las diferentes situaciones-problema implicadas en el tema, y ser capaces de reconocerlas en el proceso de estudio. Es decir, considerar las configuraciones epistémicas de objetos y procesos en el proceso de estudio llevado a cabo, y compararlo con las configuraciones pretendidas en la enseñanza; en nuestro caso, con las orientaciones curriculares (Godino, Bencomo, Font y Wilhelmi, 2006).

6.4.1. IDENTIFICACIÓN INICIAL DE CONTENIDOS

6.4.1.1. ANÁLISIS A PRIORI Y DESARROLLO

Puesto que era posible que los futuros profesores no recordasen los contenidos estadísticos fijados para primer curso de Bachillerato, el formador de profesores propuso una actividad inicial, consistente en examinar los contenidos propuestos por el Ministerio (MEC, 2007b; Batanero, Arteaga, y Gea, 2011) para las asignaturas *Matemáticas I* y *Matemáticas aplicadas a las Ciencias Sociales I* (Tabla 6.4.2). Se incluyeron otros bloques, además de la estadística, porque en el proyecto también se ejercitan contenidos de los mismos, y se les presentó con la siguiente consigna:

> La siguiente tabla resume los contenidos de las asignaturas de Matemáticas en dos especialidades de primer curso de Bachillerato según el Decreto de Enseñanzas Mínimas del MEC. Subraya todos los contenidos trabajados en el proyecto.

Se esperaba que los futuros profesores reconocieran en el proyecto planteado los contenidos: "*Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal*", del Bachillerato en *Ciencias y Tecnología*, que se amplían a "*Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación grafica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados*" en el Bachillerato de *Humanidades y Ciencias Sociales*.

En el debate posterior a la actividad, el formador indicó a los futuros profesores que todos los diagramas de dispersión presentados en el proyecto corresponden a distribuciones bidimensionales de la Esperanza de vida con alguna otra variable y, dado el contexto del que se tomaron los datos, la actividad permite interpretar fenómenos sociales (tales como la influencia del Índice de desarrollo humano o del Índice de educación sobre la Esperanza de vida). Del mismo modo, al introducir el Producto interior bruto, se interpretan fenómenos económicos, y ello se hace de forma gráfica.

La actividad de estimación de la correlación, o asignar un coeficiente a cada gráfico, permite trabajar el grado de relación entre dos variables, y la búsqueda de una función que se ajuste a los datos (regresión lineal). El proyecto no pide directamente determinar nuevos valores de una variable en función de otra, pero se trabaja la regresión al estimar la capacidad de predicción de las variables (ordenar las variables según poder de predicción), y determinar intuitiva o formalmente la función de ajuste.

Tabla 6.4.2. Contenidos de matemáticas en Primer Curso de Bachillerato

MATEMATICAS I.	MATEMATICAS APLICADAS A LAS CIENCIAS SOCIALES I.
1. Aritmética y algebra: Números reales. Valor absoluto. Desigualdades. Distancias entre la recta real. Intervalos y entornos. Resolución e interpretación grafica de ecuaciones e inecuaciones. Utilización de las herramientas algebraicas en la Resolución de problemas.	*1. Aritmética y algebra:* Aproximación decimal de un numero real. Estimación, redondeo y errores. Resolución de problemas de matemática financiera en los que intervienen el interés simple y compuesto, y se utilizan tasas, amortizaciones, capitalizaciones y números índice. Parámetros económicos y sociales. Resolución de problemas del ámbito de las ciencias sociales mediante la utilización de ecuaciones o sistemas de ecuaciones lineales. Método de Gauss.
2. Geometría: Medida de un Angulo en radianes. Razones trigonométricas de un Angulo. Uso de formulas y transformaciones trigonométricas en la resolución de triángulos y problemas geométricos diversos. Vectores libres en el plano. Operaciones. Producto escalar. Modulo de un vector. Ecuaciones de la recta. Posiciones relativas de rectas. Distancias y ángulos. Resolución de problemas. Idea de lugar geométrico en el plano. Cónicas.	
3. Análisis: Funciones reales de variable real: clasificación y características básicas de las funciones polinómicas, racionales sencillas, valor absoluto, parte entera, trigonométricas, exponenciales y logarítmicas. Dominio, recorrido y extremos de una función. Operaciones y composición de funciones. Aproximación al concepto de límite de una función, tendencia y continuidad. Aproximación al concepto de derivada. Extremos relativos en un intervalo. Interpretación y análisis de funciones sencillas, expresadas de manera analítica o grafica, que describan situaciones reales.	*2. Análisis:* Expresión de una función en forma algebraica, por medio de tablas o de graficas. Aspectos globales de una función. Utilización de las funciones como herramienta para la resolución de problemas y la interpretación de fenómenos sociales y económicos. Interpolación y extrapolación lineal. Aplicación a problemas reales. Identificación de la expresión analítica y grafica de las funciones: polinómica, exponencial y logarítmica, valor absoluto, parte entera y racionales sencillas a partir de sus características. Las funciones definidas a trozos. Tasa de variación. Tendencias.
4. Estadística y Probabilidad: Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal. Estudio de la probabilidad compuesta, condicionada, total y a posteriori. Distribuciones binomial y normal como herramienta para asignar probabilidades a sucesos.	*3. Probabilidad y estadística:* Estadística descriptiva unidimensional. Tipos de variables. Métodos estadísticos. Tablas y gráficos. Parámetros estadísticos de localización, de dispersión y de posición. Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación grafica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados. Asignación de probabilidades a sucesos. Distribuciones de probabilidad binomial y normal.

Las actividades iniciales también permiten ejercitar los contenidos de estadística descriptiva unidimensional de la asignatura *Matemáticas aplicadas a las Ciencias Sociales I*: *"Tipos de variables. Métodos estadísticos. Tablas y gráficos. Parámetros estadísticos de localización, de dispersión y de posición"*. El proyecto permite igualmente trabajar con la aritmética y algebra, pues se opera con números decimales, distancias (entre los puntos del diagrama y la recta), intervalos (en el diagrama acumulativo e histograma), variables; y se usan herramientas algebraicas en la resolución del proyecto.

Respecto a geometría, se trabaja con coordenadas cartesianas, ecuaciones de la recta, y posición relativa de las dos rectas de regresión, así como ángulo que forman. La recta o línea de regresión también se puede considerar un lugar geométrico (de las medias condicionales). También se trabaja la idea de proyección de un punto sobre la recta. El componente de análisis viene dado por las funciones en el estudio de la regresión. En el proyecto aparecen la lineal, polinómica, exponencial y logarítmica, así como su expresión algebraica y representación gráfica. Se trabaja el dominio, recorrido y extremos. Se identifican e interpretan propiedades de funciones sencillas de forma gráfica y analítica. Se utilizan funciones en la resolución de problemas, e interpretación de fenómenos sociales y económicos.

Con la realización del proyecto, el futuro profesor ejercitaría, por tanto, estos contenidos, adquiriendo un conocimiento matemático común del mismo. Asimismo, se puede ejercitar un conocimiento más avanzado, puesto que se analizan objetos matemáticos no incluidos en dichos decretos, principalmente en las actividades de ampliación, que los futuros profesores realizaron y discutieron en clase con el formador. El análisis de dicho contenido, y su contextualización en un proyecto, desarrolla el conocimiento especializado del contenido del futuro docente (Godino, 2009).

6.4.1.2. RESULTADOS DE LA EVALUACIÓN

Las respuestas de los futuros profesores a esta actividad inicial se resumen en las Tablas 6.4.3 y 6.4.4, en las que se diferencia si los participantes seleccionan todos los contenidos de un apartado dentro de cada bloque, o parte de los mismos.

Así, por ejemplo, en la asignatura de *Matemáticas aplicadas a las Ciencias Sociales I*, el futuro profesor EGA selecciona todos los contenidos del bloque de Probabilidad y estadística ("*Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación grafica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados.*"), mientras que en la asignatura *Matemáticas I* selecciona sólo los siguientes: "*Relaciones entre dos variables estadísticas*", omitiendo la regresión lineal, que se desarrolla en el proyecto, y así la identifica para la modalidad de Bachillerato de *Humanidades y Ciencias Sociales*.

Observamos que todos los participantes identifican en el proyecto la correlación y regresión, junto con un buen número de contenidos de aritmética y análisis; siendo mejores los resultados que los obtenidos en actividades semejantes con profesores de educación primaria (por ejemplo, Mohamed, 2012 o Gómez, 2014).

De los resultados obtenidos podríamos pensar, que los futuros profesores otorgan mayor importancia a los contenidos estadísticos del Bachillerato de Ciencias Sociales, o son capaces de reconocerlos mejor. Ello es debido a que dicha especialidad incluye un repaso de la estadística descriptiva univariante, que fue reconocida en el proyecto por los futuros profesores. El tema más seleccionado en ambos casos fue la correlación y regresión.

Tabla 6.4.3. Contenidos reconocidos del currículo de Matemáticas I (Ciencia y Tecnología)

Contenidos	Completa	Parcial
1. Aritmética y algebra:		
Números reales. Valor absoluto. Desigualdades.	2	1
Distancias entre la recta real. Intervalos y entornos.	2	
Resolución e interpretación grafica de ecuaciones e inecuaciones.	1	
Utilización de las herramientas algebraicas en la resolución de problemas.	1	
2. Geometría:		
Ecuaciones de la recta. Posiciones relativas de rectas. Distancias y ángulos. Resolución de problemas.		2
3. Análisis:		
Funciones reales de variable real: clasificación y características básicas de las funciones poli nómicas, racionales sencillas, valor absoluto, parte entera, trigonométricas, exponenciales y logarítmicas.	3	6
Dominio, recorrido y extremos de una función.		1
Interpretación y análisis de funciones sencillas, expresadas de manera analítica o grafica, que describan situaciones reales.	12	1
4. Estadística y Probabilidad:		
Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal.	18	4
Estudio de la probabilidad compuesta, condicionada, total y a posteriori.		1

Tabla 6.4.4. Contenidos reconocidos del currículo de Matemáticas aplicadas a las Ciencias Sociales I

Contenidos	Completa	Parcial
1. Aritmética y algebra:		
Aproximación decimal de un numero real. Estimación, redondeo y errores.	2	1
Resolución de problemas de matemática financiera en los que intervienen el interés simple y compuesto, y se utilizan tasas, amortizaciones, capitalizaciones y números índice. Parámetros económicos y sociales.		1
2. Análisis:		
Expresión de una función en forma algebraica, por medio de tablas o de graficas.	1	2
Interpolación y extrapolación lineal. Aplicación a problemas reales.		1
Identificación de la expresión analítica y grafica de las funciones: polinómica, exponencial y logarítmica, valor absoluto, parte entera y racionales sencillas a partir de sus características. Las funciones definidas a trozos. Tasa de variación. Tendencias.	4	8
3. Probabilidad y estadística:		
Estadística descriptiva unidimensional. Tipos de variables. Métodos estadísticos. Tablas y gráficos. Parámetros estadísticos de localización, de dispersión y de posición.	12	8
Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación grafica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados.	19	4
Asignación de probabilidades a sucesos. Distribuciones de probabilidad binomial y normal.	1	1

Por otra parte, es en *Matemáticas I* donde se seleccionan más contenidos de otros bloques temáticos, como aritmética y álgebra o análisis (Figura 6.4.1). Lo que mejor se reconoce es el trabajo con algunos tipos de funciones. Además, se encuentran mayor cantidad de selecciones de apartados completos de contenidos referidos a la asignatura de *Matemáticas Aplicadas a las Ciencias Sociales I* de estadística, que si se compara con el resto de bloques temáticos (Figura 6.4.1).

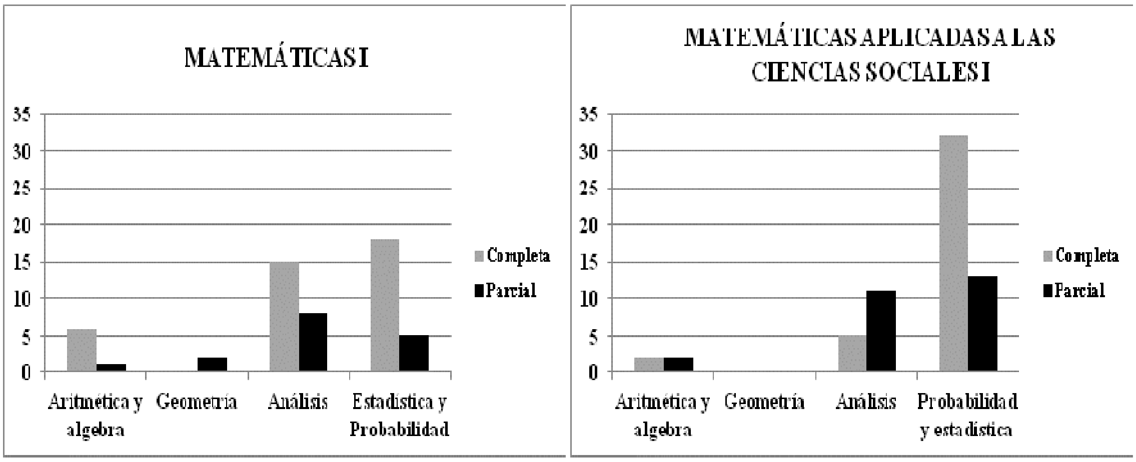


Figura 6.4.1. Contenidos reconocidos por los futuros profesores por bloques temáticos y asignatura.

6.4.2. SITUACIONES-PROBLEMAS

6.4.2.1. ANÁLISIS A PRIORI Y DESARROLLO

Una vez presentado a los futuros profesores el concepto de idoneidad epistémica y sus componentes, se analizó con ellos la importancia de las situaciones problemas, recordándoles su papel en el aprendizaje constructivista. Se propuso a los futuros profesores dos preguntas que se analizan a continuación:

1.1. Los problemas que se presentan, ¿te parece que son útiles para contextualizar, aplicar y ejercitar los contenidos de correlación y regresión? ¿Por qué?

En el proyecto aparecen los tres principales campos de problemas determinados en el análisis de los libros de texto (Estudio 1). En primer lugar, la interpretación de los diagramas de dispersión involucra el campo de problemas *P0. Organización de datos bidimensionales y representación en el registro gráfico*, pues los datos, originariamente vienen dados en forma de listado en la hoja Excel entregada como parte del proyecto. Aunque el futuro profesor no ha elaborado dichos diagramas, al final del proyecto se le propone que explore otras posibles relaciones entre las variables del fichero, lo que le requerirá tener que trabajar este campo de problemas. Asimismo, aparece cuando los participantes trabajan con las actividades de ampliación del proyecto, tanto en la exploración de otras variables en el servidor de Naciones Unidas, como en el estudio de recursos en Internet.

También aparecen en el proyecto los campos de problemas *P1. Analizar la existencia de relación entre variables* y *P2. Predecir una variable en función de otra*, tanto en el análisis de la relación de la Esperanza de vida con otras variables, como en las actividades finales del proyecto.

En el análisis de las situaciones problema se espera que los futuros profesores perciban que el proyecto planteado permite generar una muestra representativa de problemas de índole estadística. Aunque en todo proyecto estadístico se requiere también la recogida, recuento y resumen de datos (Connor, Davies y Payne, 2002), estas primeras fases se han dado hechas y sólo se pide la interpretación de resultados. Es posible también que algún futuro profesor aluda a que el proyecto permite contextualizar otros contenidos estadísticos, como los de variable y distribución, medidas de posición central y dispersión, y diferentes tipos de gráficos.

1.2. ¿Se proponen situaciones para que el alumno de Bachillerato invente nuevos problemas? ¿Cuáles?

Como en todo proyecto, se trata de estudiar la realidad (que no se ajusta completamente a la matemática); y para ello se plantean preguntas concretas que pueden resolverse con la estadística (Starkings, 1997). Aunque el proyecto contiene problemas específicos (como el buscar la función que mejor se ajuste a los datos) también se pueden identificar momentos en que el alumno podría inventar nuevos problemas.

Por ejemplo, un estudiante que trabajase con este proyecto podría preguntarse cuáles variables afectan a la Tasa de fecundidad entre adolescentes, o a cualquier nueva variable dependiente. Asimismo, al trabajar con la hoja Excel, que produce los gráficos y la línea de regresión usando las herramientas que incluye por defecto, se plantean nuevos problemas, como por ejemplo: ¿Cómo elegir las escalas del gráfico de dispersión para visualizar mejor los datos? ¿Sería mejor usar una función polinómica en vez de la recta? ¿Cómo se puede mejorar la estimación?, entre otras.

6.4.2.2. RESULTADOS DE LA EVALUACIÓN

Las respuestas de los futuros profesores se codificaron según los niveles descritos en la Sección 6.3, y que a continuación mostramos junto a algunos ejemplos.

Nivel 2. Se hace referencia a situaciones problema, pero no son específicas del proyecto desarrollado, no son estrictamente matemáticas, o se refieren a aspectos anecdóticos del proyecto.

En el descriptor 1.1 (problemas útiles para contextualizar los contenidos), varios futuros profesores justifican la pertinencia de las situaciones problema que se proponen en el proyecto, pero las que describen no son estrictamente matemáticas, como por ejemplo:

> Creo que sí, porque al tomar los datos de la página web de la ONU, trabajamos en el contexto público que los alumnos están acostumbrados a ver en los medios de comunicación. (ME, descriptor 1.1).
> Sí, porque son datos extraídos de la realidad y con gran relevancia social, que además asocian al alumno y a su entorno más inmediato (MAG, descriptor 1.1).

En el descriptor 1.2 (situaciones en las que el alumno inventa nuevos problemas) encontramos un ejemplo en ChC, quien ha identificado actividades de ampliación en la parte final del proyecto. Sin embargo, las describe de una forma muy genérica, sin citar los tipos particulares de problemas propuestos. Otro caso es DG, quien describe variables del proyecto como la población y su crecimiento, y no indica ninguna situación:

> En el proyecto de ayer no; se ofrece en el trabajo de ampliación (ChC, descriptor 1.2).
> Dentro de este proyecto sí hay situaciones diferentes que se pueden usar para inventar nuevos problemas aunque todos relativos a temas relacionados con la población y su crecimiento. Podría haber más variedad (DG, descriptor 1.2).

Nivel 3. El futuro profesor aplica el descriptor a contenidos matemáticos contenidos en el proyecto y a la forma en que se trabajó con el proyecto en el aula, pero la aplicación es incompleta, pues no se centra en situaciones problema.

Por ejemplo, ChC, en lugar de valorar las situaciones-problema propuestas en el proyecto, aplica el descriptor 1.1 describiendo una propiedad (grado de correlación), e identifica que dicha propiedad se visualiza mediante el diagrama de dispersión. Por su parte, EGO aplica este descriptor valorando aspectos del lenguaje (gráfico), o JMV analiza aspectos procedimentales que se utilizan en el proyecto. No se encuentran respuestas de futuros profesores al descriptor 1.2 en este nivel.

> Sí, porque la nube de puntos representa el grado de correlación entre dos variables (ChC, descriptor 1.1).
> Sí, ya que en el gráfico aparecen representados conceptos como el coef. de correlación a través de la recta y de la nube de puntos, por tanto se puede apreciar a través de dos tipos de representación (EGO, descriptor 1.1).
> Sí, son útiles porque el estudio de fenómenos reales que afectan al alumno, se hace de forma visual e interactiva en el ordenador (JMV, descriptor 1.1).

Nivel 4. Se hace una aplicación correcta y consistente del descriptor, utilizando situaciones problemas del proyecto en forma consistente con la pregunta, aunque razona mediante un único ejemplo.

Un caso en el descriptor 1.1 sería el futuro profesor que es capaz de identificar correctamente un tipo de problema propuesto en el proyecto, como realiza MC, quien identifica la existencia de problemas de correlación, aunque lo expresa en forma algo imprecisa:

> Sí, porque tienes que aplicar todos los conceptos por diferentes caminos. Tienes unas nubes de puntos para aplicar las cosas que sabes de correlación. Además está muy contextualizado y te permite ver su utilidad (MC, descriptor 1.1).

En el descriptor 1.2 sería el participante que es capaz de identificar situaciones-problema en las que un estudiante de Bachillerato pudiera inventar nuevos problemas, como por ejemplo EGA, quien menciona la posibilidad de analizar otras variables:

> Sí, porque hay distintas variables que el alumno puede analizar y relacionar como actividades complementarias (EGA, descriptor 1.2).

Nivel 5. El futuro profesor hace una aplicación correcta y consistente del descriptor, utilizando situaciones problemas del proyecto en forma consistente con la pregunta; y razona mediante dos o más ejemplos.

Encontramos varios futuros profesores con este nivel en las respuestas al descriptor 1.1, como por ejemplo PJ, quien valora positivamente la contextualización de las situaciones-problema propuestas en el proyecto, e indica que permiten adquirir gradualmente los contenidos pretendidos.

> Sí, porque comienzan hablándonos de una situación de la vida real y después se matematiza el contexto, aplicando la mayoría de los contenidos de correlación y regresión tales como nubes de puntos, coeficiente de correlación lineal, regresión, rectas de regresión, predicción. Y es un buen ejemplo de que con una situación determinada hacer un estudio completo de variables bidimensionales y su relación entre ellas (PJ, descriptor 1.1)

Otro ejemplo en el descriptor 1.2 es AJD, quien cita por un lado el problema de analizar diferentes variables (ya que en el fichero y el servidor de la ONU se podrían elegir muchas otras para analizar). Al mismo tiempo hace referencia específica al problema de elegir una función de ajuste. Por otro lado, ME indica aquellas nuevas situaciones que permiten contextualizar tanto la correlación como la regresión:

> Sí, porque hay muchas más variables que se pueden analizar; además queda abierto ajustar los datos a una función o a otra (AJD, descriptor 1.2)
> Si. Por ejemplo en la actividad en la que se habla de qué variables sirven para predecir mejor la esperanza de vida. O en las actividades de ampliación cuando se pide estudiar “otras variables que te interesen e influyan en el desarrollo humano” (ME, descriptor 1.2)

En la Tabla 6.4.5 se resumen los resultados obtenidos en cada una de las categorías descritas, que se representan gráficamente en la Figura 6.4.2. Observamos que el nivel más frecuente fue el 2, especialmente en el descriptor 1.2, y aproximadamente el 40% de los futuros profesores queda en este nivel o por debajo en el primer descriptor, y un 60% en el segundo.

Tabla 6.4.5. Frecuencia (y porcentaje) de niveles en la valoración de las situaciones problema

Nivel	Descriptor 1.1	Descriptor 1.2
0	1 (4,3)	2 (8,7)
1		2 (8,7)
2	8 (34,8)	9 (39,1)
3	4 (17,4)	
4	5 (21,7)	7 (30,4)
5	5 (21,7)	3 (13)
Total	23(100)	23(100)

No obstante, otro 40% de los participantes aplica los dos descriptores en uno de los dos niveles superiores. Estos resultados son mucho mejores que los obtenidos por Arteaga (2011), quien plantea el análisis de la idoneidad didáctica de un proyecto estadístico, experimentado por ellos mismos, a futuros profesores de Educación Primaria. En su caso, sólo alrededor del 20% de los participantes llegan a aplicar los descriptores correspondientes a las situaciones problemas a nivel 3, y el resto ni siquiera lo alcanza. Nuestro resultado evidencia la mayor preparación y el conocimiento matemático más completo de los futuros profesores en nuestra muestra.

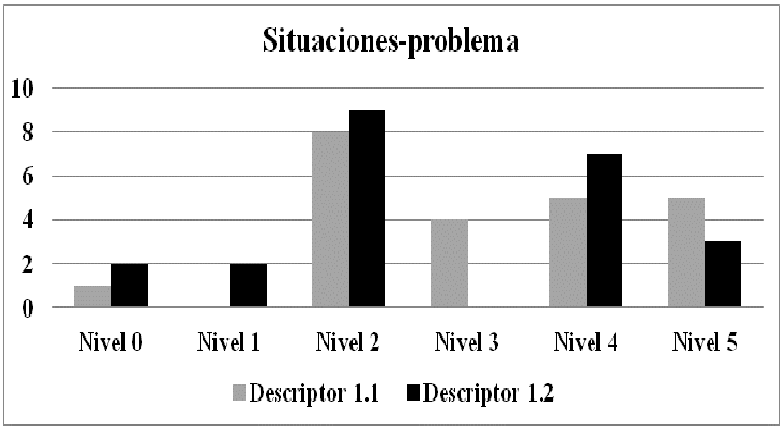


Figura 6.4.2. Frecuencia (y porcentaje) de niveles en la valoración de las situaciones problema

6.4.3. LENGUAJE

6.4.3.1. ANÁLISIS A PRIORI Y DESARROLLO

Sobre este componente de la idoneidad epistémica se plantearon dos preguntas:

1.3. ¿Se usan diferentes modos de expresión matemática (verbal, gráfica, simbólica...), traducciones y conversiones entre los mismas? ¿Cuáles?

Como se ha indicado en Gea, Contreras, Cañadas, y Arteaga (2012), la variedad de lenguaje matemático es imprescindible para una correcta comprensión de la correlación y regresión. En sus respuestas a las preguntas planteadas en el proyecto, los futuros profesores utilizaron términos y expresiones verbales relacionadas con la correlación o regresión, y en general, con la estadística. Por ejemplo: dato, variable, valor, distribución, frecuencia, población, muestra, correlación, signo, intensidad, coeficiente de correlación, recta de regresión, pendiente, centro de gravedad. Se utiliza también lenguaje simbólico para expresar las líneas de regresión o los diferentes estadísticos que intervienen en el proyecto. Aparecen tablas de frecuencias absolutas y porcentajes, y diversos gráficos: histogramas, gráficos acumulativos, diagramas de dispersión, gráficos de caja, diagramas de burbuja.

1.4. ¿Hay actividades de representación e interpretación del lenguaje matemático? ¿Cuáles?

Nolan y Speed (1999) resaltan la importancia que el trabajo con proyectos tiene para desarrollar la capacidad discursiva de los estudiantes, que también se desarrolla en el proyecto que implementamos. Los futuros profesores tuvieron que interpretar diferentes gráficos estadísticos, interpretar símbolos y términos estadísticos, así como utilizarlos al exponer sus resultados. Al trabajar con la hoja Excel se ha de interpretar el lenguaje específico, especialmente si se trabaja con algunas funciones; también la disposición tabular de la hoja donde cada fila corresponde a un país, y las columnas corresponden a las variables. En las actividades finales del proyecto se plantean nuevas situaciones de representación gráfica.

Todo lo anterior dota al proyecto de una razonable idoneidad en cuanto al lenguaje matemático utilizado, como fue resaltado por el formador en el debate posterior.

6.4.3.2. RESULTADOS DE LA EVALUACIÓN

Mostramos a continuación ejemplos de respuestas en que los futuros profesores aplican estos descriptores en los niveles 2 a 5.

Nivel 2. El futuro profesor aplica el descriptor, haciendo referencia al lenguaje en el proyecto, aunque de modo anecdótico o no estrictamente matemático. En otros casos, se aplica en forma parcial, pues no ha comprendido el sentido en que utilizamos la expresión “lenguaje matemático”.

Encontramos este nivel únicamente en el descriptor 1.4, principalmente porque los participantes centran su respuesta bien en actividades de interpretación, o en las de representación, como ATL.

> Se representan las “nubes de puntos”. De esta forma se está representando gráficamente el lenguaje matemático. También son una representación del lenguaje matemático las rectas de regresión porque indican visualmente el concepto (ATL; descriptor 1.4).

En otros casos, como ChC, se identifica correctamente las actividades de interpretación y representación en los diagramas de dispersión y otros gráficos usados en el proyecto; no obstante, se responde (inconsistentemente con la primera parte) que no hay actividades de interpretación del lenguaje, o es escasa su presencia en el proyecto; probablemente se entiende lenguaje matemático únicamente en sentido coloquial.

> Si. Todas las representaciones gráficas son actividades de representación e interpretación de los resultados estadísticos obtenido. De interpretación del lenguaje matemático no hay (ChC; descriptor 1.4)

Nivel 3. Se aplica el descriptor a contenidos matemáticos del proyecto pero no se centra en el lenguaje.

Encontramos este nivel en un futuro profesor que, al aplicar el descriptor 1.4, se refiere a aspectos procedimentales o propiedades, y no tanto al lenguaje, como se pedía:

> Si nos dan la ecuación de la recta de regresión sí que habría que hacer una interpretación matemática. Además, cuando nos dan la nube de puntos tenemos que interpretar si hay correlación o no (JFM; descriptor 1.4).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, indicando aspectos del lenguaje utilizado en el proyecto de forma consistente a la pregunta, aunque razona mediante un único ejemplo.

Las respuestas que generalmente dan los futuros profesores se refieren a la actividad 2, en la que se pedía interpretar distintos gráficos y estadísticos calculados de la variable dependiente. Para su resolución, el estudiante maneja distintas representaciones, las traducciones entre las mismas, y necesita interpretarlas. A pesar de ello, tan sólo se refiere a esta actividad, sin precisar otras que aparecen en el proyecto.

> Sí. En la actividad 2, cuando se pide interpretar los gráficos y los estadísticos (ME, descriptor 1.4).
> Sí, a partir de la representación gráfica de la esperanza de vida con respecto a otra variable de la nube de puntos se realizó una interpretación sobre ellas, indicando los posibles valores del coeficiente lineal, por ejemplo (PP; descriptor 1.4.)

Nivel 5. Se hace una aplicación correcta y consistente del descriptor y del lenguaje matemático trabajado en el proyecto mediante dos o más ejemplos.

Los futuros profesores se encuentran principalmente en este nivel al aplicar el descriptor 1.3, como por ejemplo ChC, aunque no siempre al aplicar el descriptor 1.4. Así, ChC se encuentra en el nivel 2 al aplicar 1.4, mientras que PJ aplica ambos en este nivel.

> Sí, para los mismos datos se ofrecen gráficos como histogramas, caja, tabla y gráfico de frecuencias acumuladas (ChC; descriptor 1.3)
> Sí se usan diferentes modos de expresión, por ejemplo los gráficos de histograma, de cajas, curva de frecuencias acumuladas… Esto es muy útil a la hora de apreciar distintas representaciones de una misma información para apreciar cómo varían visualmente y observar varias percepciones diferentes. Además los datos concretos nos aportan rigor, con lo que podemos discernir qué modo de representación es más objetivo (PJ; descriptor 1.3)
> Sí, por ejemplo cuando nos dan las nubes de puntos y los coeficientes de correlación lineal, tenemos que decidir qué tipo de correlación existe entre las variables a partir de la inspección visual y la interpretación del valor del coeficiente. Otro caso se da en la recta de regresión y la predicción de un valor a partir de otro, pues si no entendemos lo que representa dicha recta no podemos proceder al cálculo de predicciones (PJ; descriptor 1.4)

En la Tabla 6.4.6 y Figura 6.4.3 podemos observar el alto porcentaje de respuesta en los niveles 4 y 5, sobre todo en el descriptor 1.3 (diferentes modos de expresión matemática). Aun así, los futuros profesores muestran cierta dificultad al identificar tareas de representación e interpretación del lenguaje en el proyecto. De ahí el alto índice de respuesta al descriptor 1.4 en el nivel 2, donde aplican parte del descriptor correctamente (identifican actividades de interpretación o bien de representación, pero no ambas.

Tabla 6.4.6. Frecuencia (y porcentaje) de niveles en la valoración del lenguaje

Nivel	Descriptor 1.3	Descriptor 1.4
0		
1	3(13)	
2		10(43,5)
3		1(4,3)
4	3(13)	8(34,8)
5	17(73,9)	4(17,4)
Total	23(100)	23(100)

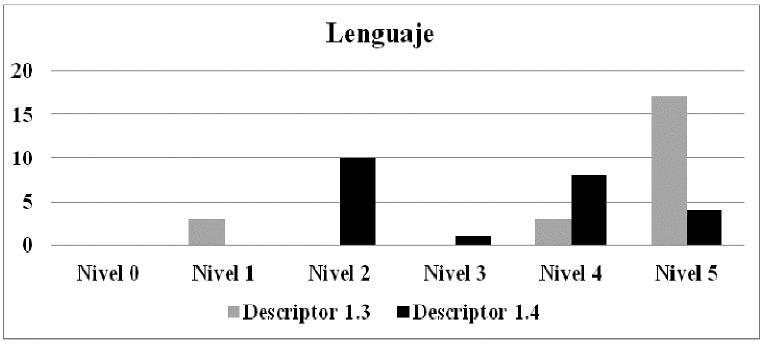


Figura 6.4.3. Frecuencia (y porcentaje) de niveles en la valoración del lenguaje

De nuevo, los resultados son mucho mejores que los de Arteaga (2011) con futuros maestros de primaria, quienes sólo llegaron al nivel 3 en la aplicación de estos descriptores. En su caso, como en el nuestro, fue más sencillo el descriptor referido a la variedad del lenguaje (1.3), que el referido a situaciones de interpretación y representación (1.4).

6.4.4. CONCEPTOS, PROPIEDADES Y PROCEDIMIENTOS

6.4.4.1. ANÁLISIS A PRIORI Y DESARROLLO

El formador de profesores indica a los participantes que el proyecto tiene como finalidad el aprendizaje de conceptos, propiedades y procedimientos relacionados con la correlación y la regresión, que a ellos le sirvió de refuerzo de los mismos. Les motiva, a continuación, sobre la importancia que para ellos tiene el saber reconocerlos en la situación que se trabaja, planteándose sobre este componente de la idoneidad epistémica las dos preguntas que se analizan a continuación.

1.5. ¿Qué conceptos, propiedades y procedimientos de la correlación y regresión habría que explicar previamente para trabajar este proyecto?

Podemos suponer que el estudiante comienza en ese momento el estudio de la correlación y regresión, como se indicó anteriormente. Así, el futuro profesor ha de reconocer aquellos objetos matemáticos utilizados en su enseñanza y aprendizaje como:

- *Conceptos y propiedades.* El alumno, al comenzar el trabajo, debe tener familiaridad con la variable estadística unidimensional, sus posibles distribuciones y representaciones, así como con las medidas de centralización (media, mediana, moda), posición (percentiles) y medidas de dispersión (desviación típica y varianza). Debe conocer las coordenadas cartesianas para interpretar los diagramas de dispersión, la representación de funciones sencillas (polinómica, exponencial, logarítmica), la distancia y proyección de un punto a una recta. Asimismo, debe conocer las propiedades más importantes de estos conceptos.
- *Procedimientos*: Los procedimientos requeridos son la lectura e interpretación de gráficos sencillos, determinación de percentiles, elección de un valor representativo para una distribución, interpretación de la ecuación de las funciones citadas. Así

como manejo de la hoja Excel para el cálculo de funciones sencillas y la producción de gráficos.

1.6. ¿Se proponen tareas donde los alumnos tengan que reconocer o aplicar definiciones propiedades o procedimientos? ¿Cuáles?

Hay varias situaciones en que los estudiantes pueden reconocer o aplicar definiciones, propiedades y/o procedimientos. Por ejemplo, al elegir la medida de centralización más adecuada para la Esperanza de vida; asignar el signo de la correlación, según la tendencia de un gráfico; graduar (primero mediante un valor numérico y luego con el coeficiente de correlación) la intensidad de dependencia según la dispersión de la nube; o bien, ordenar las variables independientes según su poder de predicción, proponiendo una función que se ajuste a los datos.

En este análisis se espera que los futuros profesores perciban que el proyecto comienza con el tratamiento de estos conceptos, y que para resolver las distintas tareas se hace uso de procedimientos en los que se aplican diferentes propiedades, que les permiten argumentar y justificar sus acciones.

6.4.4.2. RESULTADOS DE LA EVALUACIÓN

Mostramos a continuación ejemplos de respuestas en que los futuros profesores aplican estos descriptores en los niveles 2 a 5, omitiendo el nivel 3 al no encontrarse respuestas en este.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra parte se omite, o se cometen errores al no haberse comprendido totalmente su significado. También se puede aplicar correctamente el descriptor, pero haciendo uso de aspectos no estrictamente matemáticos.

Los futuros profesores suelen encontrarse en este nivel al aplicar el descriptor 1.5, porque a veces no diferencian conceptos, propiedades y/o procedimientos previos (que se requieren para la enseñanza de la correlación y regresión), como se pide. Por ello indican objetos referidos a la correlación y regresión propiamente dicha (Tabla 6.4.8). En el descriptor 1.6, LT presenta este nivel, pues muestra inconsistencia en su respuesta, y además, se refiere a la regresión, y no a contenidos previos:

> No se proponen tareas pero sí que tienen que reconocer definiciones y procedimientos de cálculo por ejemplo las propiedades de la recta de regresión y otros parámetros (LT; descriptor 1.6).

Nivel 4. Hace una aplicación correcta y consistente del descriptor, utilizando conceptos, propiedades y/o procedimientos utilizados en el proyecto, en forma consistente con la pregunta, aunque razona mediante un único concepto.

En la aplicación del descriptor 1.5 encontramos que la mayoría se refieren a propiedades o procedimientos relativos a la interpretación de la nube de puntos, como EGO; o a las medidas en una variable (centralización, dispersión y posición) como MM.

> La nube de puntos, el coeficiente de correlación, recta de regresión, dependencia (conceptos).

> Propiedades: $-1 \leq r \leq 1$; $r = +1, -1$ correlación fuerte. Procedimientos: interpretación de la nube de puntos, tipos de dependencia (EGO; descriptor 1.5).
> La media, la moda, la mediana, los percentiles, la varianza y la covarianza. Como procedimientos se tendría que explicar sus correspondientes cálculos. (MM; descriptor 1.5).

Al aplicar el descriptor 1.6, generalmente se hace referencia a las tareas iniciales del proyecto, como se muestra a continuación. CM utiliza el término significado, en lugar de propiedad, no asimilando los elementos primarios (lenguaje, conceptos, propiedades, procedimientos, argumentos y situación problema) que conforman el significado de la correlación o la regresión.

> Sí, por ejemplo, el cuadro de datos dado en la clase anterior de la media, moda, mediana… Para saber las relaciones entre ellos deben saber su significado. (CM; descriptor 1.6).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor y de los conceptos, propiedades y procedimientos que se trabajan en el proyecto mediante dos o más ejemplos.

En la aplicación del descriptor 1.5 sólo tres futuros profesores muestran este nivel en sus respuestas, ya que combinan riqueza de conceptos previos requeridos e indican al menos un procedimiento o propiedad asociado a los mismos. Así por ejemplo, DG se refiere a la construcción de la nube de puntos, o PJ, quien además indica un procedimiento:

> Por ejemplo, qué es una nube de puntos y cómo se construye, distribuciones marginales, tabla de frecuencias, media, varianza, … (DG; descriptor 1.5).
> Nubes de puntos, coeficiente de correlación lineal de Pearson. Recta de regresión, concepto de correlación, tipos de dependencia, cálculo de la recta de regresión, los distintos parámetros estadísticos. Interpretación de la nube de puntos. Predicción de una variable a partir de la otra. Propiedades: $r \in (-1,1)$, $\sigma > 0$ (PJ; descriptor 1.5).

Para el descriptor 1.6, encontramos gran variedad de respuesta. Algunos, como JG se limitan a justificar las tareas mediante la riqueza de conceptos que se trabajan; otros se refieren a procedimientos y propiedades como ATL:

> Sí, reconocer las definiciones de las medidas de centralización, dispersión, coeficiente de correlación (Pearson), regresión (JG; descriptor 1.6).
> Sí, al tener que asignar una relación positiva-negativa, fuerte-débil e incluso el coeficiente de correlación, se le está pidiendo que reconozca los conceptos relacionados con esto. También lo es la tarea de ajustar nubes de puntos con un tipo de función (lineal, exponencial, logarítmica). (ATL; descriptor 1.6).

Tabla 6.4.7. Frecuencia (y porcentaje) de niveles en la valoración de reglas

Nivel	Descriptor 1.5	Descriptor 1.6
0		1(4,3)
1		
2	15(65,2)	1(4,3)
3		
4	5(21,7)	10(43,5)
5	3(13)	11(47,8)
Total	23(100)	23(100)

Observamos (Tabla 6.4.7 y Figura 6.4.4) el alto índice de respuesta en los niveles 4 y 5 para el descriptor 1.6, no tanto para el descriptor 1.5. Creemos que se debe a la dificultad de los participantes de diferenciar lo que es la definición de un concepto de las propiedades que determinan los procedimientos que se realicen, y este hecho se agrava cuando deben considerar si se requieren previamente para la enseñanza de la correlación y regresión. Así es que se presenta un alto porcentaje de participantes en el nivel 2 para el descriptor 1.5, como se indicó anteriormente.

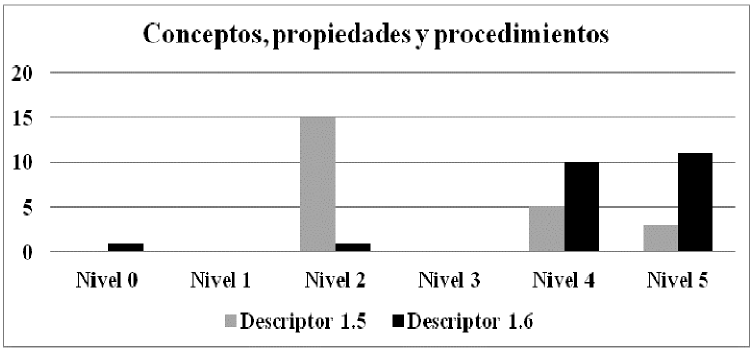


Figura 6.4.4. Frecuencia (y porcentaje) de niveles en la valoración de reglas

Tabla 6.4.8. Frecuencia y porcentaje de conceptos indicados en la valoración del descriptor 1.5.

Conceptos	Frecuencia(%)
Variable estadística unidimensional	6(26,1)
Medidas de centralización y dispersión	16(69,6)
Coordenadas cartesianas, Representación de funciones	11(47,8)
Correlación	19(82,6)
Regresión	9(39,1)

En la Tabla 6.4.8 se presentan las frecuencias y porcentajes del los conceptos que se esperaba que los futuros profesores indicaran, para profundizar en los que indican al valorar el descriptor 1.5. Algunos participantes señalaron todos ellos en sus respuestas. Se añaden en la tabla correlación y regresión, por encontrarlas citadas en la mayoría de las respuestas de los participantes, aunque eran respuestas incorrectas. Encontramos que la correlación es el concepto más indicado, seguido de las medidas de centralización y dispersión, coordenadas cartesianas y representación de funciones. Todos ellos son conceptos presentes en la actividad, que han sido bien reconocidos por los participantes. Observamos que, al contrario que en otras investigaciones en que se pide identificar objetos matemáticos en la tarea (e.j., Arteaga, 2011, Mohamed, 2012 o Gómez, 2014), no se incluyen objetos no matemáticos u objetos no presentes en la misma, evidenciando los buenos conocimientos didácticos de los futuros profesores en la identificación de los mismos.

6.4.5. ARGUMENTOS Y RELACIONES

6.4.5.1. ANÁLISIS A PRIORI Y DESARROLLO

La situación planteada es muy rica en argumentación, pues los futuros profesores deben proponer una conclusión sobre la relación de diferentes variables con la Esperanza de vida. Por otro lado, como se ha indicado, algunas correlaciones van en contra a las teorías previas de los participantes; por lo que se podría plantear en clase una confrontación. Por ello, el formador de profesores indica a los participantes que será necesario el debate y argumentación del profesor y los estudiantes en el desarrollo del proyecto. Sobre este componente de la idoneidad epistémica se planteó dos preguntas:

1.7. ¿Son las explicaciones, comprobaciones y demostraciones adecuadas para Bachillerato?

El proyecto en sí mismo no contiene argumentaciones. Los futuros profesores han de valorar la claridad de las explicaciones que se dieron en las sesiones, así como en la corrección de las posibles soluciones al proyecto. Asimismo, puesto que se entrega una solución de la actividad por escrito, se puede valorar la claridad de este documento.

1.8. ¿Se incluyen situaciones donde el alumno tenga que argumentar? ¿Cuáles?

Se espera que los futuros profesores valoren las posibilidades de argumentación que ofrece el proyecto, pues en el informe escrito, los estudiantes han de sacar conclusiones, apoyadas por los resultados del análisis de datos. Siguiendo a Murray y Gal (2002), hemos tratado de desarrollar en los estudiantes la comprensión e interpretación de la información estadística, que no sólo requiere conocimiento estadístico o matemático, sino también habilidades lingüísticas, capacidad para plantear preguntas, y una postura crítica ante la información. También Nolan y Speed (1999) resaltan la importancia de desarrollar la capacidad discursiva de los estudiantes, para ampliar sus habilidades de pensamiento crítico.

1.9. Los contenidos matemáticos (problemas, definiciones, propiedades, etc.) se relacionan y conectan entre sí en el proyecto?

De acuerdo con Godino, Batanero y Font (2007), los diversos objetos matemáticos aparecen ligados entre sí en las prácticas de resolución de problemas. Así, una situación-problema es el origen de la actividad matemática; el lenguaje sirve para representar los problemas, procedimientos, conceptos y proposiciones; los argumentos justifican los procedimientos y las soluciones de los problemas; y las proposiciones permiten relacionar los conceptos entre sí. Por otro lado, la estadística se conecta con otros bloques temáticos y los objetos matemáticos de los mismos, como se resalta en las orientaciones curriculares (MEC, 2007b, Consejería de Educación, 2008b). Por ello, se incluye una última pregunta, que se refiere a las relaciones establecidas en el proyecto entre los diversos tipos de objetos matemáticos:

En el proyecto se conecta el campo de problemas *P1. Analizar la existencia de relación entre variables* con el lenguaje gráfico (diagramas de dispersión), numérico y simbólico (coeficientes de correlación), tabular (tabla de estadísticos y/o frecuencias) y verbal (diferentes términos relacionados con la correlación), así como con los conceptos y propiedades asociados.

El campo de problemas *P2. Predecir una variable en función de otra,* se conecta con las expresiones verbales y simbólicas relacionadas con el estudio de la regresión (funciones de ajuste), tabular (uso de tablas de valores para el ajuste de una función) y verbal (diferentes términos relacionados con la regresión), así como con los conceptos y propiedades asociados. Por otro lado, estos dos campos de problemas se conectan entre sí, y con el campo *P0. Organización de datos bidimensionales.* Así mismo, todos ellos se relacionan con procedimientos como la representación gráfica de datos, interpretación de gráficos, trabajo con la hoja Excel, o cálculo de estadísticos.

Además, como sugiere Gattuso (2006), como en cualquier trabajo estadístico, es posible relacionar la aritmética, geometría y análisis, la resolución de problemas, o el razonamiento matemático.

6.4.5.2. RESULTADOS DE LA EVALUACIÓN

Mostramos a continuación ejemplos de respuestas en que los futuros profesores aplican estos descriptores en los niveles 2 a 5.

Nivel 2. Se hace referencia a las explicaciones, comprobaciones o demostraciones no específicas en el proyecto o no estrictamente matemáticos. También se aplica correctamente parte del descriptor mientras que otra parte se aplica incorrectamente o no se aplica.

Por ejemplo, EGO se refiere a las argumentaciones, pero en forma general. Así mismo, otros participantes indican aspectos no matemáticos como por ejemplo EGA, quien valora positivamente el contexto y la transversalidad del proyecto, o PJ, quien se refiere a aspectos cognitivos:

> Sí, ya que se pueden hacer más deducciones y los conceptos tienen más sentido y se entienden mejor (EGO; descriptor 1.7)
> En mi opinión sí. Además son temas bastante interesantes que pueden relacionar con otras asignaturas como economía, historia, biología, etc. (EGA; descriptor 1.7).
> Sí, puesto que a esta edad los adolescentes están en el último nivel de desarrollo cognitivo según Piaget, y han adquirido las competencias para que se les puedan mostrar estas comprobaciones y demostraciones y puedan asimilarlas. (PJ; descriptor 1.7).

Nivel 3. El futuro profesor aplica el descriptor pero de modo incompleto, ya que se centra en otros aspectos y no en la argumentación o las relaciones entre objetos.

Encontramos este nivel únicamente al aplicar el descriptor 1.7, donde los futuros profesores valoran generalmente conceptos en vez de argumentaciones:

> Sí, son conceptos básicos que deben saber de años anteriores no sólo en bachillerato (media, varianza y desviación típica por ejemplo) (CL; descriptor 1.7)

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, utilizando explicaciones, comprobaciones y/o demostraciones del proyecto o relaciones entre objetos matemáticos en forma consistente con la pregunta, aunque razona mediante un único ejemplo.

Este nivel de respuesta lo encontramos generalmente en la aplicación del descriptor 1.8, y en menor medida en el resto. Es habitual que los futuros profesores indiquen tareas iniciales del proyecto, como PP, aunque también se señalan aquellas específicas de correlación y regresión, como ME:

> Sí, por ejemplo en la actividad 6 donde se pide explicar qué variables consideras que tengan relación causa-efecto la esperanza de vida. (ME; descriptor 1.8).
> Sí, por ejemplo, tenemos que tener en cuenta la moda y la mediana para justificar cuál es un valor representativo de nuestro estudio. (PP; descriptor 1.9)

Nivel 5. Se aplica correcta y consistente el descriptor a la argumentación o relaciones entre objetos matemáticos trabajados en el proyecto mediante dos o más ejemplos.

Las respuestas de los futuros profesores generalmente se encuentran en este nivel, sobre todo al aplicar el descriptor 1.9. Esto puede dar indicios de la dificultad de los participantes de reconocer argumentaciones en el proyecto, pero no tanto de identificar conexiones y relaciones que se establecen entre objetos matemáticos en el mismo.

> Sí, aunque el coeficiente de determinación (R^2) a veces no se explica en bachillerato (solo el coeficiente de correlación lineal). (MAG; descriptor 1.7).
> Sí, ya que no son ejercicios de cálculo sino de relacionar y deducir, por tanto hay que argumentar. Serían los de la nube de puntos o la interpretación de gráficos. (EGO; descriptor 1.8)
> Sí porque todo se puede relacionar, la nube de puntos con la recta de regresión y el coeficiente de correlación así como la dependencia entre variables. (MC; descriptor 1.9)

Tabla 6.4.9. Frecuencia (y porcentaje) de niveles en la valoración de argumentos y relaciones

Nivel	Descriptor 1.7	Descriptor 1.8	Descriptor 1.9
0		1(4,3)	
1			1(4,3)
2	9(39,1)	1(4,3)	2(8,7)
3	3(13)		
4	2(8,7)	11(47,8)	4(17,4)
5	9(39,1)	10(43,5)	16(69,6)
Total	23(100)	23(100)	23(100)

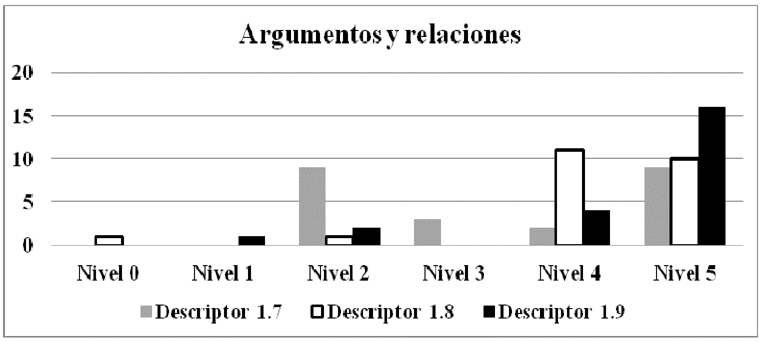


Figura 6.4.5. Frecuencia (y porcentaje) de niveles en la valoración de argumentos y relaciones

En la Tabla 6.4.9 y Figura 6.4.5 se presenta un resumen de los resultados obtenidos, donde podemos observar el alto porcentaje de respuesta en los niveles 4 y 5. A pesar de ello, pareciera que los participantes no hubiesen entendido correctamente el descriptor 1.7, ya que cerca de la mitad de las valoraciones al mismo se encuentran en los niveles 2 y 3, generalmente por indicar conceptos y no argumentos o demostraciones.

6.4.6. SÍNTESIS DE CONOCIMIENTOS DIDÁCTICOS EN LA FACETA EPISTÉMICA

Para resumir los conocimientos didácticos de los futuros profesores de la muestra en la faceta epistémica, se han calculado las puntuaciones medias obtenidas en cada uno de los descriptores propuestos, descritos anteriormente (Tabla 6.4.10 y Figura 6.4.6).

Tabla 6.4.10. Medias y desviaciones típicas en los descriptores de la idoneidad epistémica

Descriptor	Contenido	Media	D, Típica
D1.1	Problemas (contextualización	3,2	1,3
D1.2	Nuevos problemas	2,7	1,5
D1.3	Variedad de lenguaje	4,3*	1,3
D1.4	Interpretación /traducción	3,3	1,2
D1.5	Reglas previas requeridas	2,8	1,2
D1.6	Tareas de reconocimiento de reglas	4,2*	1,1
D1.7	Nivel adecuado argumentación	3,5	1,3
D1.8	Argumentación por el alumno	4,2*	1,1
D1.9	Relaciones entre objetos	4,4*	1,1
Total		3,6	1,4

* Media (en la aplicación del descriptor) superior a 4

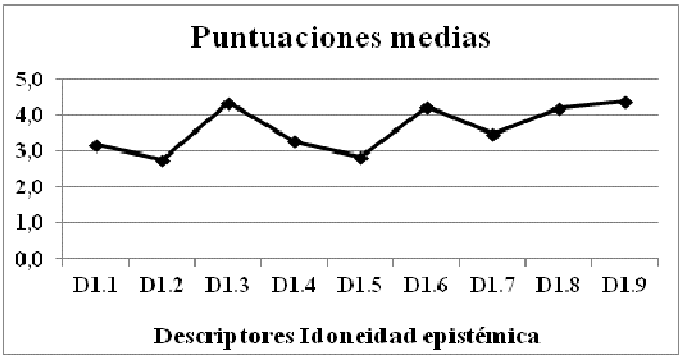


Figura 6.4.6. Puntuación media en la valoración de los descriptores de la idoneidad epistémica

Los resultados muestran un conocimiento adecuado en su faceta epistémica, ya que la media global (3.6) supera en un punto el nivel medio teórico (2.5) y es mucho mejor que el mostrado por los futuros profesores de educación primaria en el trabajo de Arteaga (2011) (puntuación media 1,11). Este conocimiento sería parte del conocimiento especializado del contenido, según Hill, Ball, y Schilling (2008), que debe incluir la capacidad para identificar los objetos matemáticos implícitos, o que se han hecho explícitos en una cierta situación de enseñanza. Por tanto, los resultados no sorprenden, debido a la relación entre el conocimiento especializado del contenido y el conocimiento común, y al hecho de que el conocimiento matemático sobre estadística es mucho mayor en los futuros profesores de secundaria que en los de primaria.

Otra aportación del análisis es la identificación, a partir de las respuestas de los estudiantes en los niveles 4 y 5, de conocimientos requeridos por parte del profesor en la faceta epistémica, que mostramos en la Tabla 6.4.11. Esta tabla se ha construido con ejemplos específicos de conocimientos manifestados por los participantes en las respuestas del análisis de la idoneidad epistémica (niveles 4 y 5), algunas de las cuáles hemos presentado a lo largo de la sección.

Tabla 6.4.11. Ejemplos de conocimientos didácticos de los futuros profesores (Faceta epistémica)

		Ejemplos de conocimientos
Situaciones-problemas	1.1.	Identifica ejemplos de problemas de contextualización de la correlación y regresión u organización de datos bidimensionales en el proyecto, como la búsqueda de una función de ajuste que describa la relación de la Esperanza de vida con otras variables.
	1.2.	Identifica situaciones en que los estudiantes tengan que proponer o modificar problemas relacionados con la correlación y regresión u organización de datos bidimensionales; en particular, el análisis de otras variables dependientes en el fichero de datos y en las actividades de ampliación.
Lenguaje matemático	1.3.	Reconoce que los diagramas de dispersión, las ecuaciones de las rectas y funciones de ajuste, y los términos y símbolos matemáticos en el proyecto, son distintos modos de expresión matemáticas ligadas a la correlación y regresión.
	1.4.	Reconoce las representaciones gráficas que permiten visualizar mejor la correlación y regresión; por ejemplo, viendo la necesidad de un cambio de escala en los datos.
	1.5.	Identifica situaciones de interpretación y traducción de representaciones de la correlación y regresión, como la búsqueda de un coeficiente de correlación o la estimación del mismo, a partir del diagrama de dispersión, o la interpretación de los parámetros en la expresión algebraica de la recta de regresión.
Conceptos, propiedades, procedimientos	1.6.	Reconoce los conceptos, propiedades y procedimientos asociados a la correlación y regresión que el alumno ha de aplicar al resolver cada problema relacionado, entre otros, la dependencia estadística y funcional, intensidad y sentido, correlación, ajuste, modelo, parámetros del modelo y su cálculo, predicción de una variable a partir de otra.
	1.7.	Reconoce cuáles de ellos han de ser explicados previamente y cuáles pueden ser construidos por el alumno con la ayuda del profesor para el trabajo en el tema.
	1.8.	Identifica y propone nuevas tareas en las que el alumno haya de reconocer o recordar conceptos, propiedades y procedimientos asociados a la correlación y regresión.
Argumentos	1.9.	Identifica argumentaciones adecuadas e inadecuadas para justificar una solución o una propiedad para el nivel del estudiante.
	1.10.	Identifica situaciones que requieran una argumentación por parte del estudiante, como por ejemplo las referidas a la justificación de situaciones causales y no causales en el proyecto, y la toma de decisión respecto al mejor modelo de ajuste.
Relaciones	1.11.	Identifica relaciones entre objetos matemáticos en el contexto del proyecto, como la relación entre la dirección de la dependencia y el signo del coeficiente de correlación o la dispersión de la nube de puntos y la intensidad de la correlación.
	1.12.	Es capaz de proponer nuevas tareas en que los estudiantes hayan de relacionar diferentes objetos matemáticos.

Por ejemplo, respecto a las situaciones problema, algunos futuros profesores han identificado la tarea 5 del proyecto (asignar un valor entre 0 y 1 según intensidad de relación entre variables, asignar un signo según sentido, asignar un coeficiente de correlación entre una lista) como un problema que permite contextualizar la idea de correlación, dependencia funcional y aleatoria y sus propiedades. Igualmente, hemos encontrado ejemplos en el análisis de futuros profesores que muestran estos conocimientos.

Pensamos que esta tabla permite describir en forma sintética y partiendo de los descriptores de la idoneidad epistémica de un proceso de estudio, algunos de los conocimientos didácticos, en su faceta epistémica, que serían requeridos en la enseñanza de la correlación y regresión. En este sentido, esta tabla, junto con otras complementarias que presentaremos, permite describir un modelo del conocimiento didáctico del profesor, desde una faceta epistémica del tema, que puede utilizarse para la evaluación o el desarrollo del mismo en cursos de formación de profesores.

6.5. VALORACIÓN DE LA IDONEIDAD COGNITIVA

Una segunda tarea en el análisis didáctico llevado a cabo al finalizar el proyecto fue valorar su idoneidad cognitiva, es decir, el grado en que los contenidos de correlación y regresión implementados (o pretendidos) son adecuados para los alumnos de Bachillerato. Godino, Wilhelmi y Bencomo (2005) indican que su análisis debe comparar los significados personales construidos por los alumnos a lo largo del proceso de enseñanza y aprendizaje con los significados institucionales.

De acuerdo a Godino (2009), el análisis de la idoneidad cognitiva de un proceso de instrucción permite desarrollar el conocimiento didáctico del contenido en relación con los estudiantes, en la terminología de Hill, Ball y Schilling (2008). Puesto que los futuros profesores no han implementado el proyecto, no han observado el aprendizaje de alumnos reales. Así es que, este análisis lo han de hacer teniendo en cuenta su propia experiencia en el trabajo con el mismo, y basándose en las características del alumnado de primer curso de Bachillerato.

En la Tabla 6.5.1 se presenta la pauta de análisis utilizada, y a continuación se describen las respuestas esperadas a las diferentes preguntas planteadas.

Tabla 6.5.1. Pauta de análisis de idoneidad cognitiva (conocimientos previos y diversidad)

		Justificación
Conocimientos previos	2.1.	¿Qué conocimientos iniciales son necesarios para el estudio del tema? ¿Se estudian en los cursos anteriores (educación secundaria)?
	2.2.	¿Se pueden conseguir el aprendizaje de los contenidos incluidos en el proyecto en alumnos de Bachillerato?
Atención a la diversidad	2.3.	¿Se incluyen actividades de ampliación y de refuerzo? ¿Cuáles?
Aprendizaje	2.4.	¿Las respuestas de los alumnos a las tareas del proyecto, permitirían evaluar el aprendizaje del alumno?
	2.5.	¿Sería necesario ampliar la evaluación con otras tareas? ¿De qué tipo?

6.5.1. CONOCIMIENTOS PREVIOS Y ATENCIÓN A LA DIVERSIDAD

6.5.1.1. ANALISIS A PRIORI Y DESARROLLO

En primer lugar, el formador de profesores motiva a los participantes sobre la importancia de reconocer qué conocimientos previos deben poseer los alumnos para poder trabajar en el proyecto, y como atender a los diferentes ritmos de aprendizaje. Para ello se plantean tres preguntas:

2.1. ¿Qué conocimientos iniciales son necesarios para el estudio del tema? ¿Se estudian en los cursos anteriores (educación secundaria)?

Se pide al futuro profesor indicar los conocimientos previos requeridos para iniciarse en el estudio de la correlación y regresión. Para comenzar a trabajar en el proyecto, se necesitarían unos conocimientos básicos de estadística descriptiva, comenzando por el concepto de población, muestra y unidad muestral; variable estadística, valores, frecuencias (absolutas, relativas, acumuladas), porcentajes y distribución. Debería diferenciar las variables cualitativas y cuantitativas, discretas y continuas, y conocer la agrupación en intervalos de clase, y conceptos asociados (extremos, marca de clase), así como su representación en la recta real.

Debería conocer los gráficos estadísticos utilizados en el proyecto: histograma, diagrama acumulativo y gráfico de caja, diferenciar la información que proporciona cada uno, saberlos leer e interpretar, principalmente a nivel de extracción de tendencias y extrapolación - interpolación, es decir, “leer entre los datos” y “leer más allá de los datos” de acuerdo a Curcio (1989). Debiera estar familiarizado con las medidas de posición central y dispersión, y saber usarlas para interpretar distribuciones de datos, reconociendo también la más adecuada en una situación problemática. Todos estos contenidos, y otros relacionados (propiedades, relaciones, representaciones verbales y simbólicas, procedimientos), se han estudiado a lo largo de la educación secundaria, y en el Bachillerato de *Ciencia y Tecnología,* se recuerdan en el primer tema del Bloque temático *Estadística y Probabilidad*.

Asimismo, serían necesarios conocimientos de otros bloques de contenido de las matemáticas, por ejemplo, el concepto de función, variable dependiente e independiente, valor, rango y dominio; características de la función lineal, ecuación, pendiente y ordenada en el origen; relación de la función lineal con la proporcionalidad; funciones polinómicas, exponenciales, potenciales y logarítmicas, ecuaciones, parámetros e interpretación de su representación gráfica.

2.2. ¿Se pueden conseguir el aprendizaje de los contenidos incluidos en el proyecto en alumnos de Bachillerato?

Para responder a esta pregunta, los futuros profesores debieran primero identificar cuáles son estos contenidos, que describimos a continuación, clasificados según los campos principales de problemas que se trabajan, identificados en el Estudio 1.

P0. Organización/representación de datos bidimensionales. Un primer contenido, en relación a este campo de problemas, son los diagramas de dispersión y su interpretación. Los estudiantes estarán ya familiarizados con la representación en coordenadas cartesianas y la representación de funciones sencillas.

P1.Analizar la existencia de relación entre variables. En las actividades propuestas se persigue que los estudiantes diferencien entre dependencia funcional, independencia y correlación. Mediante la interpretación de diagramas de dispersión se pretende que distingan la dependencia directa de la inversa (a partir de la tendencia de la nube de puntos); que asocien la mayor o menor dispersión de la nube de puntos con la intensidad de la correlación, y den una estimación correcta del coeficiente de correlación (en primer lugar de forma intuitiva) y después, identifiquen el coeficiente más adecuado a cada diagrama a partir de una lista dada de valores. Se espera también que den una explicación satisfactoria de la correlación, bien una dependencia causal u otra de los tipos de posibles de relaciones descritas por Barbancho (1973).

P2. Predecir una variable en función de otra. Para justificar la búsqueda de un modelo matemático de ajuste a los datos, una primera actividad propone ordenar las variables independientes en función del poder de predicción sobre la dependiente. De este modo aparece una propiedad que relaciona correlación y regresión: que el cuadrado del coeficiente de correlación (coeficiente de determinación) mide la proporción de varianza explicada por el modelo de regresión. Esta propiedad solo se institucionaliza una vez realizada la actividad, y aparece posteriormente cuando, con ayuda de Excel, los futuros profesores proponen el modelo que mejor se ajuste a los datos.

En el caso del Índice de desarrollo humano, el Índice de educación, e incluso de la Tasa de fecundidad y Población urbana, se podría a priori deducir que una recta iría bien como modelo. El estudiante podría dar la ecuación general de la recta $y = a + bx$ indicando que b sería la pendiente (indicando si es positiva o negativa), y a la ordenada en el origen. Para el Producto Interior Bruto se podría pensar en una parábola o una función logarítmica, y dar su expresión aproximada, o indicar el tipo de función. Posteriormente, el estudiante debe utilizar Excel para completar la tarea, comprender el significado del cuadrado del coeficiente de correlación como medida de bondad de ajuste, e interpretar la expresión algebraica de diferentes familias de funciones: lineal, exponencial, logarítmica y parabólica.

En resumen, con estas actividades se trabajan los contenidos que marca la normativa curricular (MEC 2007b). Además, se responde a uno de los criterios de evaluación sugeridos por el MEC (2007b), que se refleja en las Matemáticas de ambas modalidades (MEC 2007b, pp. 45450 y 45476):

> Comprobar la capacidad de apreciar el grado y tipo de relación existente entre dos variables, a partir de la información gráfica aportada por una nube de puntos; así como la competencia para extraer conclusiones apropiadas, asociando los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden" (MEC 2007b, p. 45476).

Los futuros profesores deberán valorar el aprendizaje de los estudiantes con el desarrollo del proyecto mediante el análisis de su propia experiencia en el mismo, y basándose en las características del alumnado de primer curso de Bachillerato.

2.3. ¿Se incluyen actividades de ampliación y de refuerzo? ¿Cuáles?

Se pueden considerar como actividad de refuerzo aquella en que se pide al estudiante que elija otra variable del fichero como variable dependiente, y que estudie su relación con el resto de variables del fichero, haciendo uso de Excel en el ajuste de los datos. Como actividades de ampliación, por ejemplo, la actividad en que se pide al estudiante visitar el servidor de las Naciones Unidas, y analizar otras que puedan explicar la Esperanza de vida; tarea que es muy abierta, y permite al estudiante plantear sus propios problemas de correlación y regresión. Precisamente la fase de planteamiento de preguntas es una de las más difíciles, porque los alumnos rara vez comienzan con un problema claramente formulado. El papel del profesor es ayudarles a pasar de un tema general a una pregunta que pueda contestarse (Batanero y Díaz, 2004).

También la actividad donde se pide al estudiante obtener nuevas conclusiones sobre la Esperanza de vida a partir del diagrama de burbujas, disponible en las Naciones Unidas (http://hdr.undp.org/es/estadisticas/). El estudiante puede elegir otras variables explicativas y representar simultáneamente hasta cinco diferentes (los dos ejes, el color, tamaño de la burbuja y el eje de tiempo, que además, puede variar dinámicamente). Esta actividad es, por consiguiente, una introducción elemental al análisis multivariante. De acuerdo a Ridgway, McCusker y Nicholson (2007), los datos y representaciones multivariantes tienen interacciones complejas, que en muchas ocasiones, no están relacionadas linealmente, por lo que el currículo de la escuela debiera preparar a los estudiantes para interpretar este tipo de datos.

6.5.1.2. RESULTADOS DE LA EVALUACIÓN

Mostramos a continuación ejemplos de respuestas a cada una de las preguntas descritas en los niveles 2 a 5.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra se omite, se cometen errores, o se hace uso de aspectos no estrictamente matemáticos.

En la aplicación al descriptor 2.3 (actividades de ampliación/refuerzo) no hay respuestas en este nivel, y sólo una en el descriptor 2.1 (conocimientos previos) pues no se valora si se tratan en cursos anteriores los conceptos indicados. Las respuestas de los futuros profesores al descriptor 2.2 (se puede conseguir el aprendizaje) se encuentran generalmente en este nivel. Por ejemplo, ChC señala aspectos no relacionados con el proyecto, y GM puntos anecdóticos del mismo.

> Sí, en Bachillerato se profundiza más, también se estudia contraste de hipótesis y temas más avanzados. La capacidad de aprendizaje de alumnos de Bachiller abarca este tema. (ChC; descriptor 2.2).
> Sí, puesto que de esta manera más lúdica, los contenidos mejor. (GM; descriptor 2.2).

Nivel 3. Se aplica en forma correcta el descriptor pero de modo incompleto ya que, se centra en otros puntos y no en la pregunta específica formulada en el descriptor.

Encontramos un único participante en este nivel de aplicación al descriptor 2.3, que se centra en conceptos y no en actividades de ampliación o refuerzo:

> Sí, ya que tú estás impartiendo la correlación lineal y hay nube de puntos que se ajustan mejor a otro tipo de funciones. (JFM; descriptor 2.3).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

En la aplicación del descriptor 2.1 no consideramos como contenidos previos los específicos de la correlación y regresión, citados en este descriptor por algunos participantes, entre ellos, el coeficiente de correlación o la recta de regresión. Así es que algunos futuros profesores señalan tan sólo un contenido previo, entre todos los que se señalan:

> Son necesarios conocimientos de regresión, nube de puntos, funciones. En teoría deberían darse estos conceptos al final de la E.S.O, pero creo que en la mayoría de los casos no da tiempo de explicarlo (ATL; descriptor 2.1).

Para el descriptor 2.3 es habitual encontrar respuestas en que el futuro profesor indica únicamente una actividad de las denominadas “de ampliación”; y en el descriptor 2.2, los profesores en este nivel únicamente valoran la potencialidad del proyecto en el aprendizaje de la correlación y regresión por trabajar con datos reales.

> Sí, porque, por ejemplo, hay una actividad en la que te dan las nubes de puntos y tiene que identificar el tipo de función a la que se aproxima (que puede ser no lineal) (MC; descriptor 2.3).
> Sí, tomando otro conjunto de datos rehacer la actividad, estudiando los resultados (DG; descriptor 2.3).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Son frecuentes estas respuestas en el descriptor 2.1. Consideramos en este nivel las respuestas al descriptor 2.3 que se refieren a las actividades de ampliación de modo global.

> La estadística descriptiva, funciones, regresión y correlación… Sí se suelen estudiar en 3º y 4º E.S.O (AJD; descriptor 2.1).
> Sí. La tarea de repetir el ejercicio con otras variables, o bien, intentar ajustes de funciones diferentes en función del coeficiente de correlación óptimo (ATL; descriptor 2.3).

En la Tabla 6.5.2 y Figura 6.5.1 observamos el alto índice de respuesta en los niveles 4 y 5, sobre todo en los descriptores 2.1 (reconocer conocimientos previos para el estudio del tema y considerar si se trabajan en Secundaria) y 2.3 (actividades que atiendan a la diversidad de alumnado: ampliación y refuerzo). Para el descriptor 2.2, como indicamos anteriormente, se muestra una tendencia de los futuros profesores a matizar aspectos relativos a la madurez del estudiante, y no tanto a precisar contenidos propios del proyecto. De nuevo, los resultados superan los de Arteaga (2011), que suelen estar a nivel 1 o 2.

Además, en la Tabla 6.5.3 se recogen los contenidos que los futuros profesores consideran como más pertinentes que sepan los estudiantes previo al estudio del tema (descriptor 2.1). Encontramos que las diferentes medidas de una variable (centralización, dispersión y posición) junto a las representaciones gráficas y su interpretación son los conceptos que mayormente identifican los futuros profesores, como necesarios que los estudiantes conozcan antes del estudio del tema. Encontramos pocos casos en los que se selecciona como conocimientos previos el análisis de funciones o las características de una variable unidimensional y su distribución, siendo por lo general seleccionadas conjuntamente las representaciones gráficas y las medidas de la variable. Todos ellos son necesarios en la actividad, que han sido bien reconocidos por los participantes.

Tabla 6.5.2. Frecuencia (y porcentaje) de niveles en la valoración de conocimientos previos y atención a la diversidad

Nivel	Descriptor 2.1	Descriptor 2.2	Descriptor 2.3
0			
1		3(14,3)	2(9,5)
2	1(4,8)	8(38,1)	
3			1(4,8)
4	3(14,3)	2(9,5)	9(42,9)
5	17(81)	8(38,1)	9(42,9)
Total	21(100)	21(100)	21(100)

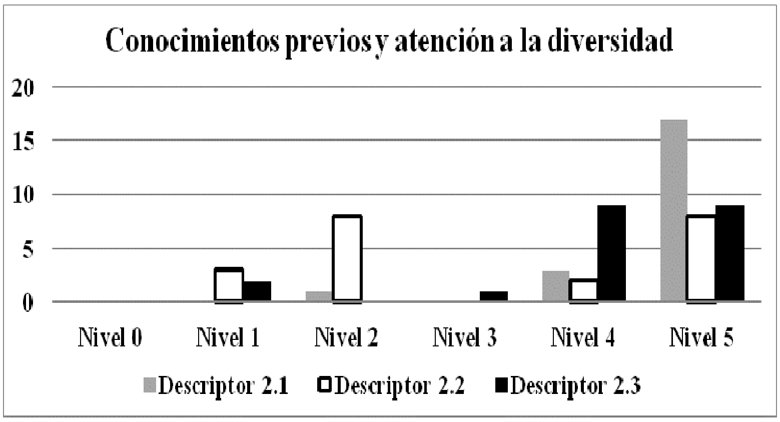


Figura 6.5.1. Frecuencia (y porcentaje) de niveles en la valoración de conocimientos previos y atención a la diversidad

Tabla 6.5.3. Frecuencia (y porcentaje) de conceptos indicados en la valoración de conocimientos previos

Conceptos	Muestra 1
Estadística descriptiva	4(19)
Variable estadística (valor, tipo, frecuencia, distribución)	6(28,6)
Representaciones gráficas	16(76,2)
Medidas de centralización, dispersión y posición	18(85,7)
Funciones	4(19)
Correlación	4(19)
Regresión	9(42,9)

Observamos que, al contrario que en otras investigaciones en que se pide identificar objetos matemáticos en la tarea (por ejemplo, Arteaga, 2011, Mohamed, 2012, o Gómez, 2014), no se incluyen objetos no matemáticos u objetos no presentes en el proyecto.

6.5.2. APRENDIZAJE

6.5.2.1. ANALISIS A PRIORI Y DESARROLLO

Tras el análisis de los futuros profesores de los conocimientos previos para el estudio del tema, y la adecuación del proyecto para atender a la diversidad del alumnado, un aspecto importante de la idoneidad cognitiva del mismo es valorar el aprendizaje del estudiante. Para ello, el futuro profesor deberá considerar aspectos relativos a la evaluación de los contenidos que se implementan, y la adecuación de las tareas de evaluación. A continuación, se describen las preguntas que se plantearon al respecto:

2.4. ¿Las respuestas de los alumnos a las tareas del proyecto, permitirían evaluar el aprendizaje del alumno?

Se pide al futuro profesor que valore si las respuestas de los estudiantes en el desarrollo del proyecto permiten una adecuada evaluación de su aprendizaje. Ello implica suponer que el proyecto se desarrolla con estudiantes de Bachillerato, o reflejarse ellos mismos tras la implementación recibida, donde se llevó a cabo una discusión de las respuestas a las preguntas planteadas en el proyecto.

El análisis de dichas respuestas pretendería conocer el conocimiento final de los estudiantes. No se ha realizado una evaluación inicial, por tanto, formalmente no se puede evaluar el aprendizaje; no obstante, en el supuesto de que el proyecto se utiliza para contextualizar e iniciar el estudio de la correlación y regresión, los datos recogidos permitirían evaluar dicho aprendizaje en los estudiantes, principalmente, por la riqueza en argumentación que poseen las situaciones problema que se plantean.

2.5. ¿Sería necesario ampliar la evaluación con otras tareas? ¿De qué tipo?

En el proyecto no se trabajan todos los objetos matemáticos relacionados con la correlación y regresión. Por este motivo, la evaluación nos informa únicamente del aprendizaje sobre los contenidos tratados. Muchas propiedades utilizadas implícitamente a lo largo del proyecto tampoco se evalúan formalmente. Por ejemplo, no se evalúa formalmente que los alumnos asocien la intensidad de la correlación con la pendiente de la recta de regresión (creciente, decreciente) o con el signo del coeficiente de correlación, a pesar de que en el debate sobre las soluciones estas propiedades emerjan.

En este sentido, esperamos que los futuros profesores indiquen que sería necesario completar la evaluación con algunas preguntas de tipo teórico, como por ejemplo, que interpreten el significado de los coeficientes a y b en la ecuación de la recta de regresión $y = a + b \cdot x$, o bien pedirles, una vez construida esta recta de regresión, que formulen la recta de regresión de X en función de Y para evaluar si los estudiantes discriminan las dos líneas de regresión, y por ende, los conceptos de variable dependiente e independiente. También podrían proponer ejercicios de cálculo de los diferentes coeficientes, o de representación de datos bivariantes en un diagrama de dispersión, así como otras actividades de traducción entre representaciones.

6.5.2.2. RESULTADOS DE LA EVALUACIÓN

Presentamos a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2 a 5.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra parte se omite, se cometen errores, o se aplica correctamente pero haciendo uso de aspectos anecdóticos del proyecto.

La valoración del descriptor 2.5 (nuevas tareas de evaluación) se encuentra generalmente en este nivel, ya que no se suelen argumentar las respuestas con contenidos específicos del proyecto. Por ejemplo, ChC indica que se deberían proponer tareas relacionadas con la vida real, aspecto contemplado el proyecto. Otros futuros profesores contestan que sí se permite evaluar el aprendizaje del alumno pero no señalan cómo:

> Sí, tareas relacionadas con la vida real, como estaturas, edades… Realizar un pequeño examen. (ChC; descriptor 2.5).
> Sí, por ejemplo tareas de ampliación, o en un contexto más personal para que los alumnos se

> motiven e impliquen más en la materia. (CM; descriptor 2.5).

En el descriptor 2.4 (el proyecto permite evaluar el aprendizaje) encontramos también este nivel de respuesta, generalmente porque mezclan su respuesta con la de la siguiente pregunta. Así por ejemplo, ME señala aspectos no estrictamente matemáticos:

> Sí, pero al ser un trabajo diseñado para hacer en grupo, habría que ampliar la evaluación con tareas individuales para comprobar que todos los miembros del grupo han alcanzado los objetivos de aprendizaje. (ME; descriptor 2.4).

Nivel 3. Se hace una aplicación correcta pero incompleta del descriptor, ya que se centra en puntos no recogidos por el descriptor.

Encontramos una sola respuesta en este nivel, que se refiere a los conceptos que se pueden evaluar pero no se indica la adecuación de las preguntas que se plantean para valorar dicho aprendizaje:

> Sí, ya que estás evaluando su conocimiento de parámetros estadísticos, las rectas de regresión, etc. (JFM; descriptor 2.4).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

Los futuros profesores no presentan este nivel en la aplicación al descriptor 2.4. En el descriptor 2.5, por lo general, los futuros profesores proponen un único tipo de actividad, o un único aspecto para añadir a la evaluación. Por ejemplo, ATL considera la necesidad de preguntar a los alumnos por definiciones formales de los conceptos.

> Sí, alguna más de ampliación en las que se trabaje con la ecuación de regresión más claramente, como ver el cálculo ya que en el proyecto no se refleja (DG; descriptor 2.5).
> Creo que sería necesario poner alguna tarea más conceptual donde se pueda deducir el grado de conocimientos del alumno sobre los puntos de los que trata el tema, aunque e proyecto que se ha realizado es bastante completo. (ATL; descriptor 2.5).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos. Encontramos este nivel mayoritariamente en el descriptor 2.4 y también el 2.5:

> Sí, ya que tienen que interpretar, no calcular, y por tanto se puede saber si han entendido correctamente el concepto, ya que si fuese sólo calcularlo quizás lo hacen de forma automática sin entender para qué sirve (EGO; descriptor 2.4).
> Sí, ampliaría con tareas que incluya un poco de más cálculo. No sólo que sepan interpretar los conceptos; sino cómo llegar por ejemplo a obtener el coeficiente de correlación o la recta de regresión para interpretarla posteriormente (MC; descriptor 2.5).

En la Tabla 6.5.4 y Figura 6.5.2 observamos el alto índice de respuesta en los niveles 4 y 5, sobre todo el descriptor 2.4 en el nivel 5. También encontramos un índice de respuesta considerable en el nivel 2 en ambos descriptores, de modo algunos futuros profesores valoran de modo parcial estos descriptores; en general, porque responden a la metodología de evaluación y no tanto a objetos matemáticos, o aportan criterios como otros contextos aludiendo a los datos reales, algo que sí se trata en el proyecto. Los resultados de nuevo mejoran los de Arteaga (2011). Por consecuencia, las respuestas muestran un grado razonable de desarrollo de la competencia del profesor para proponer o valorar pruebas de evaluación del conocimiento de los estudiantes.

Tabla 6.5.4. Frecuencia (y porcentaje) de niveles en la valoración del aprendizaje

Nivel	Descriptor 2.4	Descriptor 2.5
0		
1	2(9,5)	
2	5(23,8)	8(38,1)
3	1(4,8)	
4		7(33,3)
5	13(61,9)	6(28,6)
Total	21(100)	21(100)

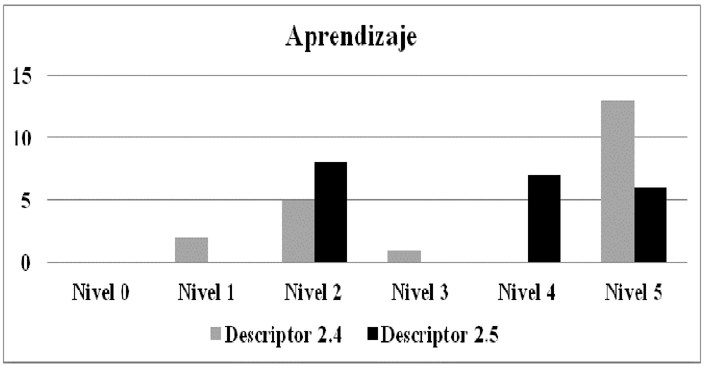


Figura 6.5.2. Frecuencia (y porcentaje) de niveles en la valoración del aprendizaje

6.5.3. SÍNTESIS DE CONOCIMIENTOS DIDÁCTICOS EN LA FACETA COGNITIVA

Para resumir los conocimientos didácticos de los futuros profesores de la muestra en la faceta cognitiva, se han calculado las puntuaciones medias obtenidas en cada uno de los descriptores descritos anteriormente (Tabla 6.5.5 y Figura 6.5.3). Este conocimiento sería parte del conocimiento del contenido y el estudiante según Hill, Ball, y Schilling (2008), que debe incluir la capacidad para reconocer los conocimientos, estrategias, razonamientos, y dificultades o posibles errores de los estudiantes.

Tabla 6.5.5. Medias y desviaciones típicas en los descriptores de la idoneidad epistémica

Descriptor	Contenido	Media	D, Típica
D2.1	Conocimientos iniciales	4,7*	0,7
D2.2	Aprendizaje factible	3,2	1,6
D2.3	Atención a la diversidad	4,1*	1,2
D2.4	Evaluación factible	3,8	1,6
D2.5	Ampliación tareas evaluación	3,5	1,3
Total		3,9	1,4

* Media (en la aplicación del descriptor) superior a 4

Los resultados también son razonablemente buenos, pues la puntuación media global (3,9) supera a la media teórica (2.5), y mejor que en el trabajo de Arteaga (2011), donde sólo llegaron a una puntuación media de 1,55. Lo más sencillo fue identificar qué conocimientos previos se requieren por parte de los estudiantes para trabajar la enseñanza de la correlación y regresión, así como que el proyecto permite tener en cuenta la diversidad de estudiantes (puntuaciones sobre 4). En este componente de la idoneidad ningún descriptor tiene puntuaciones inferiores a 3, dando muestra del adecuado conocimiento didáctico de los futuros profesores en el componente cognitivo del tema.

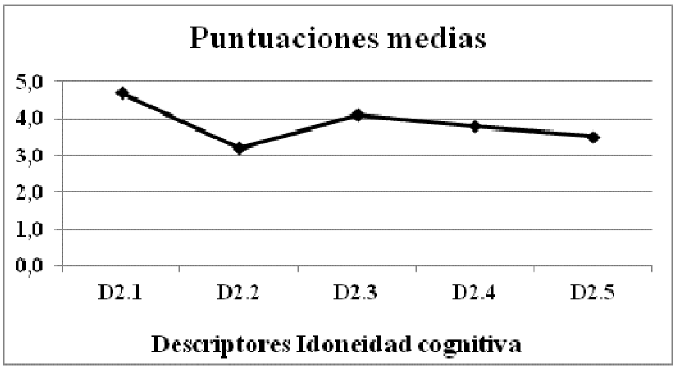


Figura 6.5.3. Puntuación media en la valoración de los descriptores de la idoneidad cognitiva.

Tabla 6.5.6. Ejemplos de conocimientos didácticos de los futuros profesores (Faceta cognitiva)

	Ejemplos de conocimientos
Conocimientos previos	2.1. Reconoce que para iniciar el tema se requiere un conocimiento de la estadística descriptiva unidimensional (tablas, gráficos y resúmenes de estadísticos), la representación en coordenadas cartesianas y las funciones elementales. 2.2. Recuerda que todos estos conceptos han sido aprendidos en la Educación Secundaria, y están contenidos en el currículo de ese nivel educativo.
Atención a la diversidad	2.3. Reconoce en el proyecto actividades de ampliación y refuerzo como análisis de nuevas variables, trabajo con applets, etc. 2.4. Es capaz de proponer variantes de las tareas del proyecto que permitan la atención a la diversidad y el refuerzo del aprendizaje; por ejemplo, utilizar diferentes funciones de ajuste a los mismos datos, o proponer otros proyectos smiliares, entre otras.
Aprendizaje	2.5. Evalua la dificultad prevista de los contenidos incluidos en el proyecto. Reconoce si los contenidos son asequibles a los alumnos de Bachillerato 2.6. Reconoce que las respuestas escritas a las tareas propuestas permiten evaluar el aprendizaje del estudiante; por ejemplo, su comprensión e interpretación de gráficos y medidas de posición central, diagramas de dispersión, sentido e intensidad de la correlación, etc. 2.7. Reconoce posibles conflictos de aprendizaje que podrían observarse por medio de las tareas propuestas; por ejemplo, confusión entre correlación y causalidad, confusión de la variable dependiente e independiente, interpretación incorrecta del coeficiente de correlación, ordenación incorrecta de coeficientes de correlación. 2.8. Propone tareas complementarias de evaluación del conocimiento sobre la correlación y la regresión; por ejemplo, tareas de cálculo de los distintos coeficientes, de definición de conceptos, o de justificación de propiedades.

Finalizamos el análisis mostrando en la Tabla 6.5.6 ejemplos de conocimientos didácticos específicos de la correlación y regresión en su faceta cognitiva, mostrados en las respuestas de los futuros profesores en los niveles 4 y 5; y que pueden utilizarse para construir un modelo del conocimiento didáctico del profesor sobre este tema.

Así, por ejemplo, algunos futuros profesores han reconocido como conocimientos previos requeridos la estadística descriptiva unidimensional; han recordado cómo estos conceptos se incluye en el currículo de la Educación Secundaria Obligatoria; han identificado las actividades de ampliación y refuerzo como actividades que permiten la atención a la diversidad; evalúan la dificultad prevista de las tareas, en relación a los estudiantes a que van dirigidas; y reconocen modos de evaluación en el proyecto, proponiendo otros complementarios, tanto conceptuales, como de cálculo.

6.6. VALORACIÓN DE LA IDONEIDAD AFECTIVA

Se pidió también a los futuros profesores valorar la idoneidad afectiva del proyecto, que informa del grado de implicación, interés y motivación de los estudiantes respecto al mismo (Godino, Contreras y Font, 2006). Al igual que el caso de la idoneidad cognitiva, este análisis tiene relación con el conocimiento del contenido y los estudiantes (Ball, Lubienski y Mewborn, 2001).

Sobre este punto se propuso la pauta de evaluación mostrada en la Tabla 6.6.1, con seis preguntas relacionadas con los intereses, actitudes y emociones de los estudiantes. Todos estos aspectos los incluye Gómez Chacón (2000) en la matemática emocional. Mientras las emociones son respuestas o sentimientos inmediatos, positivos o negativos, que se producen mientras se estudia matemáticas o estadística, las actitudes son más estables y se manifiestan como respuestas emocionales por repetición, que por lo general, inducen una predisposición a la acción. Estrada (2002) propone los siguientes ejemplos de emociones relacionados con la enseñanza y aprendizaje de la estadística:

- Respecto a la materia, en este caso la estadística, puede parecer fácil y asequible para cualquiera, o por el contrario, difícil si se consideran las habilidades que se requieren, principalmente en los procesos de interpretación y argumentación, que no todos poseen.
- En el caso de la estadística, se puede percibir como más o menos conectada con las matemáticas, y trasladar a ella las actitudes o sentimientos hacia las matemáticas. Por ejemplo, se puede pensar que es demasiado cálculo o demasiado abstracta.
- Sobre la enseñanza que se recibe; por ejemplo, que los ejemplos usados son o no aplicables a la realidad.
- Sobre uno mismo, sobre cómo se aprende estadística o matemáticas: es torpe; se le da bien, comprende/no comprende, etc.
- Sobre la utilidad o valor de la matemática y su importancia en su futuro profesional.

Tabla 6.6.1. Pauta de análisis de la valoración de la idoneidad afectiva (interés de la actividad para los estudiantes)

	JUSTIFICACIÓN
Intereses y necesidades	4.1. ¿Piensas que las tareas tienen interés para los alumnos? 4.2. Las tareas propuestas ¿permiten valorar la utilidad de las matemáticas en la vida cotidiana y profesional? ¿por qué?
Actitudes	4.3. ¿Se promueve la participación de los estudiantes en las actividades, la responsabilidad, etc.? 4.4. ¿Se favorece la argumentación en situaciones de igualdad; el argumento se valora en sí mismo y no por quién lo dice?
Emociones	4.5. ¿Promueve el proyecto la autoestima, ayudando a evitar el rechazo, fobia o miedo a las matemáticas? 4.6. ¿Se resaltan las cualidades estéticas de las matemáticas? ¿Qué otras actitudes o emociones positivas hacia las matemáticas permitiría desarrollar?

6.6.1. INTERESES Y NECESIDADES

Una vez presentado a los futuros profesores el concepto de idoneidad afectiva, se pidió que valoraran diferentes aspectos de la misma, bien desde las emociones y sensaciones que ellos mismos sintieron, como propios alumnos a los que se dirigía el proyecto, bien desde su experiencia didáctica, que es considerable (Ver el Capítulo 5).

6.6.1.1.ANÁLISIS A PRIORI Y DESARROLLO

A continuación se analizan las diferentes preguntas planteadas y se describen las respuestas esperadas.

4.1. ¿Piensas que las tareas tienen interés para los alumnos?

Se espera que los futuros profesores encuentren interesantes las tareas planteadas, así como trabajar el tema con un proyecto, que es una metodología innovadora respecto a la utilizada habitualmente en clase de matemáticas. En este sentido, el trabajo con proyectos permite al estudiante trabajar las diferentes fases de una investigación estadística, y como sugiere Holmes (1997), su desempeño aumenta la motivación de los estudiantes, ya que los datos tienen para él un significado y tienen que ser interpretados.

Por otra parte, las tareas se diseñan en un contexto internacional, con datos reales, y abordando una cuestión de interés en la vida real. No sólo resulta interesante el análisis de la Esperanza de vida en los países de la muestra, sino que la comparación de la situación de España con otros países es motivadora, especialmente en la situación actual de crisis económica.

La exploración de los visualizadores de las Naciones Unidas, o el que ellos mismos puedan elegir otras variables de su interés, pueden ser otra de las fuentes de motivación. Además, el uso de applets suele atraer mucho al estudiante, ya que se suelen tratar como elementos de exploración, por lo que se sienten atraídos.

4.2. Las tareas propuestas ¿permiten valorar la utilidad de las matemáticas en la vida cotidiana y profesional? ¿por qué?

Al trabajar en el proyecto, los futuros profesores han completado el ciclo de investigación propuesto por Wild y Pfannkuch (1999) ya que, comenzado con una pregunta, buscan unos datos (en este caso, los proporcionó el formador), los analizan e interpretan, y finalmente, responden a la pregunta planteada. Este mismo método podría seguirse con otras investigaciones, por lo que el proyecto muestra la utilidad de la estadística a la hora de realizar investigaciones (McGilliwray y Pereira-Mendoza, en preparación). Es también importante en una asignatura de innovación docente e investigación, pues les proporciona un método a seguir en futuros trabajos.

Por otro lado, el formador de profesores ha insistido en la necesidad actual de cultura estadística, ingrediente esencial en una sociedad democrática; así como para el trabajo profesional. Dicha cultura puede incrementarse mediante la interpretación de los diferentes gráficos (Ridgway, Nicholson y McCusker, 2008).

6.6.1.2. RESULTADOS DE LA EVALUACIÓN

Presentamos a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2 a 5.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra parte se omite, se cometen errores, o se usan aspectos no estrictamente matemáticos

Encontramos pocas respuestas en este nivel. Por ejemplo, ChC valora la utilidad del proyecto por el uso de la probabilidad, aspecto que no se trata en el proyecto. En otros casos, como PP, la respuesta no se centra específicamente en el proyecto.

> Sí, nos expresan datos cotidianos, probabilidades (ChC; descriptor 4.2).
> Sí, con las matemáticas se pueden recoger datos y hacer estudios para ver que hay que hacer cambios, cómo hacerlos y dónde para obtener los resultados que esperamos (PP; descriptor 4.2).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

Los futuros profesores señalan principalmente el contexto de las tareas como elemento principal de interés del alumno, y con el que se puede ver la utilidad de la matemática en la vida cotidiana. Otros señalan el trabajo con proyectos como elemento motivador en sí mismo:

> Sí, porque están enmarcados en una situación real tal como es el análisis de la esperanza de vida. (AJD; descriptor 4.2).
> Sí, porque permite constatar o rechazar objetivamente relaciones fuertes entre variables sociales que, intuitivamente, el alumno sospechara que están relacionadas o no (MAG; descriptor 4.2).
> Pienso que debe ser una tarea interesante para ellos porque se trabaja como un proyecto y eso les da cierta libertad (ATL; descriptor 4.1)

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Encontramos respuestas de futuros profesores en este nivel, sobre todo el descriptor 4.2. De nuevo se refieren al interés del contexto, que se puedan analizar desde una perspectiva comparativa (AJD), o la importancia de la matematización (PJ):

> Sí, porque llama la atención y es motivador trabajar con datos reales, además de que pueden situar su país de procedencia con respecto a los demás y eso es interesante. (AJD; descriptor 4.1)
> Sí, porque encontrar una respuesta al problema que plantea el proyecto sólo se puede hacer si matematizamos con los datos que se nos proporciona y razonamos de forma justificada mediante las leyes lógicas de la estadística para dar una respuesta objetiva pues, de otro modo sólo tendremos una idea intuitiva y no contrastada de la solución correcta. (PJ; descriptor 4.2).

En la Tabla 6.6.2 y Figura 6.6.1 se resumen los resultados obtenidos en cada uno de los niveles descritos, donde podemos observar el alto porcentaje de respuestas en los niveles 4 y 5, en los que los futuros profesores suelen justificar el interés de las tareas por el contexto que tratan (66.7% de respuestas en el nivel 4, descriptor 4.1), y la valoración de la utilidad de la matemática con diferentes argumentos (descriptor 4.2).

Tabla 6.6.2. Frecuencia (y porcentaje) de niveles en la valoración de intereses y necesidades

Nivel	Descriptor 4.1	Descriptor 4.2
0		
1	1(4,8)	
2	2(9,5)	3(14,3)
3		
4	14(66,7)	6(28,6)
5	4(19)	12(57,1)
Total	21(100)	21(100)

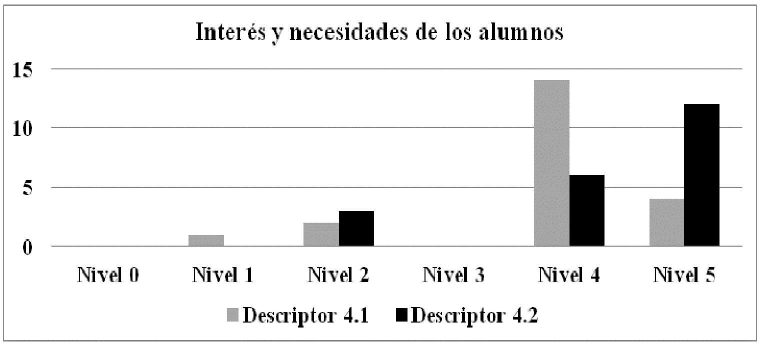


Figura 6.6.1. Frecuencia (y porcentaje) de niveles en la valoración de intereses y necesidades

6.6.2. ACTITUDES Y EMOCIONES

6.6.2.1. ANÁLISIS A PRIORI Y DESARROLLO

Para completar el análisis de la idoneidad afectiva del proyecto, el formador de profesores explicó la necesidad de establecer en el aula un buen clima de trabajo. Así se concreta en la normativa curricular, que indica que el profesor debe fomentar la igualdad de derechos y oportunidades en sus estudiantes, el interés por el trabajo cooperativo, y generar una actitud crítica hacia los resultados que se obtienen (MEC, 2007b).

Se pidió a los futuros profesores valorar si las tareas generan actitudes y emociones positivas, para lo que se plantearon dos preguntas asociadas a cada uno de estos aspectos, que describimos a continuación.

4.3. ¿Se promueve la participación de los estudiantes en las actividades, la responsabilidad, etc.?

Los estudiantes han de resolver personalmente las tareas, y se les deja bastante libertad para elegir el método. Además, se deben responsabilizar de su aprendizaje pues necesitan integrar y asimilar las diferentes respuestas de sus compañeros con la suya, sobre todo en las tareas finales que son muy abiertas. Se espera que el futuro profesor valore el trabajo activo del estudiante en el desarrollo del proyecto.

Por otro lado, en las actividades de ampliación y refuerzo se puede incluso elegir nuevas variables para analizar y plantear nuevos problemas. También se pide analizar algunos recursos en Internet, con lo que el estudiante podría buscar otros diferentes a los sugeridos por el profesor. El futuro profesor debiera reconocer que las tareas motivan la participación e implicación del estudiante en el estudio del tema.

4.4. ¿Se favorece la argumentación en situaciones de igualdad; el argumento se valora en sí mismo y no por quién lo dice?

Esta pregunta no tiene relación con el proyecto en sí mismo, sino con la forma en que se lleva a cabo en la clase. Puesto que una parte se ha desarrollado por escrito, será necesario que el formador de profesores organice una fase de corrección y discusión colectiva de las actividades, en la que se fomente la igualdad de argumentación entre los participantes, y se genere, de modo paralelo, un espacio para la evaluación del alumno. El futuro profesor podría valorar este punto basándose en su propia experiencia, como si se tratase de un estudiante de Bachillerato que ha realizado el proyecto.

4.5. ¿Promueve el proyecto la autoestima, ayudando a evitar el rechazo, fobia o miedo a las matemáticas?

El rechazo y las actitudes negativas hacia las matemáticas son frecuentes, especialmente cuando el estudiante ha experimentado una enseñanza demasiado formal, donde no se presentan suficientes aplicaciones que doten de sentido a los conocimientos presentados (Estrada, 2002). También algunos participantes podrían recurrir a su experiencia docente, que se centra principalmente en clases de apoyo, donde habrán podido percibir la actitud de rechazo que algunos estudiantes sienten hacia las matemáticas. Un motivo pudiera ser la inseguridad del estudiante en su respuesta, que desemboca en una negativa o temor a la participación en clase.

Cuando los conceptos, propiedades y procedimientos se requieren para resolver situaciones problemáticas, el estudiante puede comprender la necesidad de dichos conocimientos, y valorar el aprendizaje da las matemáticas. Cuando las tareas son asequibles al estudiante, y puede llegar a resolverlas en su totalidad, o en una parte importante, el estudiante se siente motivado, aumentando con ello su autoestima, y comprendiendo que es capaz de hacer matemáticas. Esperamos que el futuro profesor reconozca estas características en el proyecto y, por tanto, de una valoración positiva y justificada a esta pregunta.

4.6. ¿Se resaltan las cualidades estéticas de las matemáticas? ¿Qué otras actitudes o emociones positivas hacia las matemáticas permitiría desarrollar?

Aunque las cualidades estéticas de las matemáticas pudieran pasar desapercibidas al estudiante, muchos matemáticos las han destacado (Guzmán, 2003). En este proyecto en particular, la matemática permite descubrir el patrón oculto en los datos brutos mediante un proceso de transnumeración (Wild y Pfannkuch, 1999). Así, de la lista desordenada de datos, se pasa al diagrama de dispersión (o gráfico de burbujas o mapa), de donde emerge el modelo matemático de regresión, que expresa a la vez el orden de las partes (distribución) y la interrelación entre las variables (correlación). La posibilidad de extraer una tendencia desde el desorden de los datos puede ser vista como una cualidad estética.

Respecto a otras emociones positivas, podríamos citar algunas de las disposiciones que Wild y Pfannkuch (1999) incluyen en el pensamiento estadístico, puesto que el trabajo con el proyecto desarrolla este tipo de pensamiento. Se trata de la curiosidad, imaginación, mentalidad abierta a nuevas ideas, capacidad lógica, y compromiso con el trabajo. Se espera que el futuro profesor aprecie las distintas cualidades estéticas y emociones que se producen con el proyecto.

6.6.2.2. RESULTADOS DE LA EVALUACIÓN

Mostramos a continuación ejemplos de respuestas en que los futuros profesores aplican estos descriptores en los niveles 2 a 5.

Nivel 2. Se hace referencia a las actitudes o emociones, no centrándose específicamente en el proyecto, o bien se aplica parcialmente.

Los futuros profesores valoran las actitudes en este nivel, principalmente, porque no se centran específicamente en el proyecto. Por ejemplo, se valora la participación del estudiante (descriptor 4.3), pero condicionada a la gestión del profesor. Para valorar las emociones del alumno (descriptor 4.6), se refieren a la utilidad de las matemáticas, más que a sus cualidades estéticas; se considera que el proyecto ayuda a tomar conciencia de la aplicación de la matemática (descriptor 4.5), pero no tanto a disminuir la fobia hacia las mismas:

> Depende de cómo el profesor enfoque la actividad (DG; descriptor 4.3).
> Dependerá en gran medida de la capacidad que tenga el profesor para transmitir el conocimiento a sus alumnos para que estos sean capaces de superar las tareas e incrementar la autoestima. Es adecuado combinar tareas sencillas con tareas algo más complejas (MRA; descriptor 4.5).
> Sí, el ajuste de una nube de puntos dispersa y desordenada mediante una función muestra cómo las matemáticas pueden condensar la información y simplificarla mediante un modelo sencillo. También desarrolla la capacidad para matematizar. (MAG; descriptor 4.6).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, utilizando emociones o actitudes promovidas en el proyecto, en forma consistente con la pregunta, aunque razona mediante un único ejemplo.

Las respuestas de los futuros profesores que se encuentran en este nivel, se refieren a la responsabilidad y participación de los estudiantes en el proyecto (descriptor 4.3). Hacen referencia a la necesidad de superar las tareas de modo individual, a los debates, etc.:

> Sí, porque cada alumno independientemente tiene que enfrentarse a las actividades (MC; descriptor 4.3).
> Sí, realizando tareas que han sido propuestas para subir nota (responsabilidad). La realización de actividades complementarias. (ChC; descriptor 4.3).

También encontramos que los futuros profesores valoran que el proyecto favorece la igualdad por usar un lenguaje no sexista (descriptor 4.4). También se contribuye a ver las matemáticas divertidas, y que el uso de la hoja Excel motiva al alumnado por disminuir su formalidad y la realización de cálculos (descriptor 4.5).

> Sí, porque habla en lenguaje general, no diferencia entre hombres y mujeres. (MC; descriptor 4.4)
> Se ven las matemáticas como algo más ameno y entretenido y eso puede ayudar a mejorar la situación ante las matemáticas disminuyendo la fobia hacia ellas (CM; descriptor 4.5).
> Sí, porque al utilizar la hoja Excel mejora muchísimo la autoestima puesto que no es necesario hacer cálculos complejos, el mismo programa nos facilita el cuanto a cálculos (LT; descriptor 4.5).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Las respuestas de los futuros profesores al valorar el descriptor 4.6 se encuentran, generalmente, en este nivel. Se señalan aspectos como la representación gráfica, la satisfacción del uso de la correlación para contrastar sus intuiciones, entre otras.

> Sí, se resaltan las cualidades estéticas de las matemáticas. Permite establecer la media de un conjunto de datos, interpretarlos gráficamente y establecer un grado de correlación entre los datos. Gráficos con colores. (ChC; descriptor 4.6)
> Con las representaciones gráficas se podría favorecer la estética de las matemáticas. Creo que el asumir retos o el afán de superación pueden ser actitudes que se pueden ver desarrolladas con esta tarea, ya que serían consecuencia de la mejora de la autoestima (ATL; descriptor 4.6).

Igualmente encontramos valoraciones del resto de descriptores en este nivel, citando la responsabilidad y la participación individual del estudiante, el tratamiento de la igualdad en el proyecto, y la ayuda a reducir la fobia a las matemáticas.

> Sí que se promueve la participación, pero habría que controlar que trabajaran todos los miembros del grupo. La responsabilidad se puede probar con la realización de las actividades complementarias (ME; descriptor 4.3)
> Este proyecto no hace referencia discriminatoria alguna. Se tratan los datos de forma objetiva y además al analizar distintos países se puede promover la interculturalidad en clase como tema transversal (PJ; descriptor 4.4).
> Como es un proyecto que va avanzando el alumno desde conocimientos más básicos hasta razonamientos propios. En este sentido se favorece su autoestima y puede ayudar a que se reduzca la aversión contra las matemáticas (ATL; descriptor 4.5).

En la Tabla 6.6.3 y Figura 6.6.2 se presenta un resumen de los resultados obtenidos. Podemos observar el alto índice de no respuesta, principalmente en la valoración de la argumentación en situaciones de igualdad (descriptor 4.4, con un

66.7%), seguido de la contribución del proyecto a la autoestima, erradicando la fobia o el miedo a las matemáticas (descriptor 4.5, con un 42.9%). Asimismo, encontramos un tratamiento similar en los descriptores 4.3 y 4.6, donde los futuros profesores responden de acuerdo a los niveles 4 y 5 respectivamente.

Tabla 6.6.3. Frecuencia (y porcentaje) de niveles en la valoración de actitudes y emociones

Nivel	Descriptor 4.3	Descriptor 4.4	Descriptor 4.5	Descriptor 4.6
0	1(4,8)	14(66,7)	9(42,9)	4(19)
1				
2	2(9,5)	3(14,3)	5(23,8)	7(33,3)
3				
4	12(57,1)	3(14,3)	2(9,5)	1(4,8)
5	6(28,6)	1(4,8)	5(23,8)	9(42,9)
Total	21(100)	21(100)	21(100)	21(100)

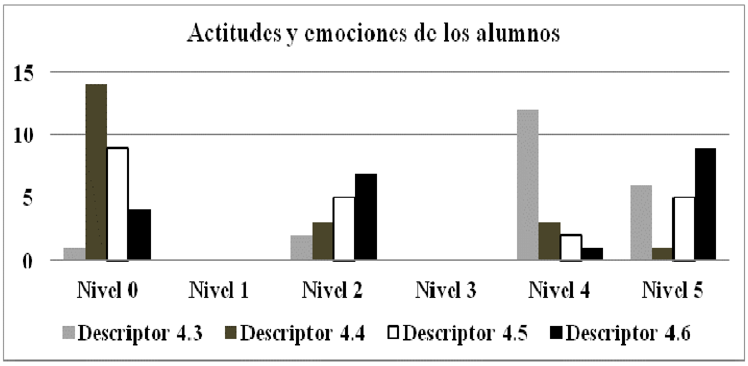


Figura 6.6.2. Frecuencia (y porcentaje) de niveles en la valoración de actitudes y emociones

6.6.3. SÍNTESIS DE CONOCIMIENTOS DIDÁCTICOS EN LA FACETA AFECTIVA

Para resumir los conocimientos didácticos de los futuros profesores de la muestra en la componente afectiva del proyecto, se han calculado las puntuaciones medias obtenidas en cada uno de los descriptores, que sirven para evaluar el conocimiento del contenido y el estudiante según Hill, Ball, y Schilling (2008).

En este componente se han obtenido peores resultados en los descriptores relacionados con favorecer la igualdad y autoestima de los estudiantes (Tabla 6.6.4 y Figura 6.6.3); posiblemente estos puntos no se lograron ver con claridad al trabajar con el proyecto, por lo que, muchos participantes los dejaron en blanco. No obstante, la puntuación media total es ligeramente mayor que la teórica, pues los tres primeros apartados obtuvieron altas puntuaciones, y bastante mejor que la obtenida en Arteaga (2011), salvo en los descriptores 4.4 y 4.5.

Como en los componentes anteriores, en la Tabla 6.6.5 se muestran ejemplos de conocimientos sobre la faceta afectiva, mostrados en las respuestas de los futuros profesores a estos descriptores, en los niveles 4 y 5. Como se sugiere en las orientaciones curriculares (MEC, 2007b), ellos reconocen que es importante presentar la matemática como una ciencia viva, que forma parte de nuestra cultura, y que resulta útil para interpretar la realidad. También indican que se debe fomentar en el estudiante una disposición abierta y positiva hacia las matemáticas, que se reflejará en la dinámica de resolución de problemas que traten contextos de relevancia social, como el propuesto en el proyecto.

Tabla 6.6.4. Medias y desviaciones típicas en los descriptores de la idoneidad afectiva

Descriptor	Contenido	Media	D, Típica
D4.1	Interés para el alumno	3,9	1,0
D4.2	Utilidad de la matemática	4,3*	1,0
D4.3	Participación de los estudiantes	3,9	1,2
D4.4	Igualdad entre estudiantes	1,1	1,7
D4.5	Autoestima; rechazo a la fobia	2,0	2,1
D4.6	Cualidades estéticas y emociones positivas	3,0	2,0
Total		3	1,9

* Media (en la aplicación del descriptor) superior a 4

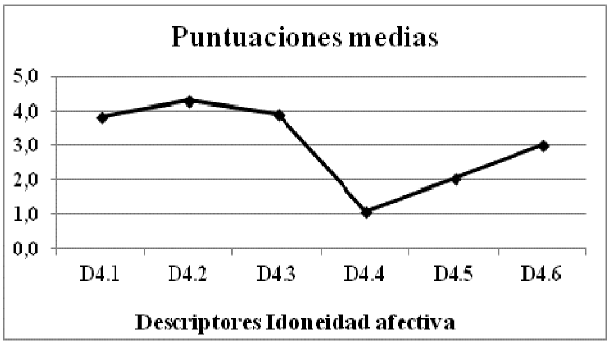


Figura 6.6.3. Puntuación media en la valoración de los descriptores de la idoneidad afectiva

Tabla 6.6.5. Ejemplos de conocimientos didácticos de los futuros profesores (Faceta afectiva)

	Justificación
Intereses y necesidades	4.1. Valora el interés de las tareas, debido a que el proyecto da al alumno cierta libertad. 4.2. Valora el interés que tiene para el alumno relacionar datos de variables sociales; igualmente el hecho de poder comparar el propio país con otros. 4.3. Reconoce que al ser reales los datos de la *Esperanza de vida*, el proyecto permite ver la utilidad de las matemáticas. Valora la utilidad de las matemáticas para dar respuestas objetivas a preguntas de interés.
Actitudes	4.4. Reconoce que el trabajo independiente en el proyecto promueve la participación de los estudiantes en las actividades y su responsabilidad e iniciativa. 4.5. Observa que se favorece la argumentación en situaciones de igualdad; por ejemplo, se evita el sexismo.
Emociones	4.6. Reconoce que al hacer las matemáticas más amenas se evita el rechazo; igualmente al trabajar con Excel se evitan errores en cálculo, que aumenta la actitud positiva. 4.7. Se reconoce que los gráficos resaltan las cualidades estéticas de las matemáticas. 4.8. Indica que, como es un proyecto que avanza al alumno desde conocimientos más básicos hasta razonamientos propios, se favorece su autoestima, y puede ayudar a que se reduzca la aversión contra las matemáticas.

Hemos encontrado ejemplos de futuros profesores que han sabido valorar el interés de las tareas, del uso de datos reales y la tecnología, y del contexto; así como las ventajas de dar autonomía al alumno, y promover su responsabilidad, para fomentar actitudes y emociones positivas hacia las matemáticas.

6.7. VALORACIÓN DE LA IDONEIDAD MEDIACIONAL E INTERACCIONAL

Se incluyen en este apartado el análisis y los resultados de la valoración de la idoneidad mediacional e interaccional del proyecto por los futuros profesores. Dichas dimensiones están relacionadas en el proyecto, ya que se desarrolla fundamentalmente mediante el uso de la tecnología, que condiciona la interacción en el aula.

6.7.1. RECURSOS Y MATERIALES

De acuerdo a Godino (2011), el análisis de la idoneidad mediacional trata de valorar la disponibilidad y el uso adecuado de los recursos didácticos en un proceso de enseñanza y aprendizaje. Se considera “recurso”, no sólo los recursos manipulativos, apuntes, textos, etc., sino la tecnología, el tiempo disponible u organización de la clase.

Su análisis permite profundizar en el conocimiento del contenido matemático y la enseñanza, en la terminología de Ball, Lubienski y Mewborn (2001), y requiere que el futuro profesor valore, no sólo los recursos, sino también cómo usarlos y para qué propósito (Llinares, 2009). Aunque la guía primitiva contenía preguntas sobre el número de alumnos, la organización y el horario de la clase, no se incluyeron en la práctica realizada con los futuros profesores, ya que estos aspectos se habían fijado de antemano, siendo muy diferentes de los que cabría esperar para su enseñanza en un centro de Bachillerato.

En la Tabla 6.7.1 se presenta la pauta de análisis utilizada para valorar la idoneidad mediacional del proyecto.

Tabla 6.7.1. Pauta de análisis de la valoración de la idoneidad mediacional

	Justificación
Recursos y Materiales	3.1. Los recursos informáticos usados, ¿permiten introducir buenas situaciones, lenguajes, procedimientos, argumentaciones sobre la correlación y regresión? 3.2. ¿Las definiciones y propiedades son contextualizadas y motivadas usando situaciones de la vida real y visualizaciones?

6.7.1.1. ANALISIS A PRIORI Y DESARROLLO

El formador de profesores motiva a los participantes sobre la importancia de utilizar los recursos tecnológicos en la enseñanza del tema, resaltada en las recomendaciones curriculares para la enseñanza de las matemáticas en Bachillerato (MEC, 2007b). Su uso no solo ayuda a la mejor comprensión de conceptos, sino también a la resolución de problemas, ya que se prestaría más atención a la interpretación de los resultados que a los cálculos. Con todo ello, se espera que los futuros profesores valoren la adecuación de los recursos utilizados en el proyecto, para lo que se plantean las siguientes preguntas:

3.1. Los recursos informáticos usados, ¿permiten introducir buenas situaciones, lenguajes, procedimientos, argumentaciones sobre la correlación y regresión?

Los recursos utilizados son principalmente informáticos; de ahí que se modificara la pregunta original de Godino (2009), ya que no se usan materiales manipulativos. Estos recursos se han descrito anteriormente con detalle y fueron de varios tipos:

- Las bases de datos de las Naciones Unidas, que permite obtener datos reales (eligiendo variables, número de valores, año, etc.), directamente en formato Excel. Permite introducir muchos problemas de estudio de la correlación y regresión.
- La hoja de cálculo Excel, utilizada como almacenamiento de los datos, organización tabular y gráfica, y cálculo. Proporciona una serie de herramientas muy útiles, sobre todo de funciones de ajuste conocidas por el estudiante, así como el coeficiente de determinación. Permite al estudiante trabajar el lenguaje gráfico y simbólico.
- Los recursos de visualización que ofrece las Naciones Unidas para representar los datos mediante diferentes gráficos, en especial el gráfico de burbujas, utilizado en las actividades de ampliación, descritas en el Capítulo 5.
- Un conjunto de applets sobre correlación y regresión, cuya exploración se propone como actividad de ampliación, que permiten al estudiante descubrir relaciones y propiedades entre los objetos matemáticos implicados.

3.2. ¿Las definiciones y propiedades son contextualizadas y motivadas usando situaciones de la vida real y visualizaciones?

Como se ha indicado, en el proyecto se trabajan los campos principales de problemas que dan sentido a la correlación y regresión, identificados en el Estudio 1. Se espera que el futuro profesor valore la utilidad de disponer de datos reales para el trabajo del tema, que se refieren a una problemática actual como es la Esperanza de vida de un país.

Mientras que, en otras ramas de las matemáticas el interés se centra más en los conceptos, y se trabaja con frecuencia con datos y contextos imaginarios, en el desarrollo del proyecto los datos son tan importantes como los conceptos. Es tan importante en sí, conocer las características de la distribución de la Esperanza de vida o de qué variables específicas depende, que comprender el significado del coeficiente de correlación. Hay también un amplio uso de visualizaciones bivariantes y multivariantes, e incluso dinámicas, como ya se ha descrito anteriormente.

6.7.1.2. RESULTADOS DE LA EVALUACIÓN

Se presentan a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2 a 5.

Nivel 2. Se aplica parte del descriptor correctamente, o se usan aspectos no estrictamente matemáticos o anecdóticos del proyecto.

Encontramos respuestas en este nivel sólo en el descriptor 3.1, que se aplica correctamente, aunque no centrándose específicamente en el proyecto:

> Sí, porque pueden observar de forma rápida y sencilla la correlación o no entre variables, ya que gráficamente se puede observar y las gráficas se pueden generar con tan sólo un par de clicks utilizando algún programa. (LT; descriptor 3.1).
> Sí, ya que de una manera más visual los alumnos entienden mejor la materia y están más atentos y motivados por ser la clase en los ordenadores, a la vez que aprenden. (CM; descriptor 3.1).

Nivel 3. Se hace una aplicación correcta del descriptor, pero de modo incompleto, ya que se centra en otros objetos matemáticos, no tanto en los recursos utilizados.

Encontramos un futuro profesor que, al valorar el descriptor 3.1, responde según aspectos procedimentales del propio recurso informático, más que de la pertinencia de su uso para la enseñanza y aprendizaje del tema:

> Sí, porque con los programas de ordenador podemos hacer cálculos mucho más rápido que a mano y representaciones de gráficos al instante. Y ello nos ahorrará tiempo para ir directamente a analizar las relaciones entre las variables, pero, si previamente los alumnos no saben representar y analizar por ellos mismos las variables, dichos programas informáticos a mi parecer no sirven de gran ayuda, pues los alumnos tendrán “huecos” en su aprendizaje. (PJ; descriptor 3.1)

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

La mayoría de participantes valoran la idoneidad mediacional en este nivel de respuesta, sobre todo en el descriptor 3.2, donde los futuros profesores valoran positivamente la problemática analizada en el proyecto. Para el descriptor 3.1, los futuros profesores se refieren, principalmente, a los elementos procedimentales o el lenguaje gráfico, como por ejemplo ChC, quien indica la utilidad de Excel para ajustar modelos de regresión:

> Sí, nos permite relacionar cada gráfica con el tipo de función que está relacionada (exponencial, logarítmica, lineal) mediante Excel (ChC; descriptor 3.1).
> Sí, porque hemos relacionado variables de la vida real como la esperanza de vida y la tasa de mortalidad, el PIB, … (MC; descriptor 3.2).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Encontramos respuestas de los futuros profesores en este nivel para ambos descriptores, pero sobre todo en el descriptor 3.1. Algunos ejemplos son:

> Sí, ya que se puede ver en el momento la representación gráfica de los datos y modificarla instantáneamente al cambiarlos, y así se observa claramente cómo repercute en el gráfico, y el significado de la regresión y la correlación. (EGO; descriptor 3.1).
> Sí. El Excel es una herramienta bastante potente que además de expresar algebraicamente y gráficamente la recta de regresión así como la nube de puntos. Los applets también son idóneos. (MAG; descriptor 3.1).
> Sin duda. El ejemplo de los países es muy ilustrativo. Y el tema es bastante visual (nube de puntos, tablas de frecuencias, rectas de regresión, …) (MRA; descriptor 3.2).

Tabla 6.7.2. Frecuencia (y porcentaje) de niveles en la valoración de recursos y materiales

Nivel	Descriptor 3.1	Descriptor 3.2
0	1(4,8)	
1		1(4,8)
2	5(23,8)	
3	1(4,8)	
4	7(33,3)	18(85,7)
5	7(33,3)	2(9,5)
Total	21(100)	21(100)

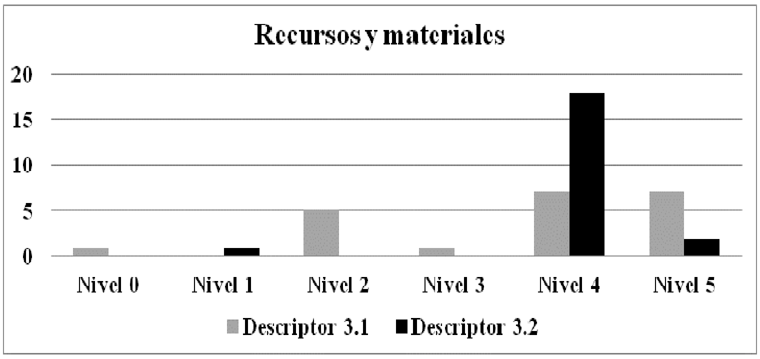


Figura 6.7.1. Frecuencia (y porcentaje) de niveles en la valoración de recursos y materiales.

En la Tabla 6.7.2 y la Figura 6.7.1 podemos observar la distribución de respuestas de los participantes según los diferentes niveles, encontrando un alto porcentaje de respuesta de los futuros profesores en los niveles 4 y 5, que en el descriptor 3.2 se concentra en torno al nivel 4 (85.7% de participantes) y para el descriptor 3.1 se distribuye de manera equivalente. Destacamos, para este descriptor, el alto porcentaje de futuros profesores cuya valoración se encuentra en un nivel 2, debido principalmente a que sus respuestas no se centran en aspectos específicos del proyecto, como se comentó anteriormente.

6.7.2. INTERACCIÓN DOCENTE-DISCENTE

La idoneidad interaccional valora el grado en que la interacción en el aula permite identificar y resolver conflictos de significado, y favorecer la autonomía en el aprendizaje. De acuerdo a Godino (2009), la reflexión sistemática sobre la misma, permite incrementar el conocimiento del contenido y la enseñanza, en términos de Hill, Ball, y Schilling (2008).

Las primeras actividades del proyecto permiten evaluar si los estudiantes tienen los conocimientos previos requeridos, o qué parte han olvidado. Se pide a continuación tratar de identificar los coeficientes de correlación adecuados a cada gráfico. En esta sesión, aflorarán algunos conflictos e ideas erróneas de los estudiantes, pues, como se ha visto, los coeficientes no siempre corresponden con los valores esperados.

En lugar de rebatir directamente estas ideas, el profesor organizará un debate en el que los estudiantes confronten ellos mismos estas ideas con los resultados empíricos del análisis de datos, una vez realizados los cálculos y el ajuste de las líneas de regresión con Excel. La intención es provocar en los estudiantes un conflicto cognitivo que ellos mismos traten de resolver, tal y como el formador de profesores llevó a cabo en la implementación del proyecto con los futuros profesores.

En cuanto al desempeño de las tareas y la gestión de la clase, el futuro profesor podrá valorar su propia experiencia en el proyecto. Donde se organizó un trabajo en parejas, para favorecer la interacción entre estudiantes y el contraste de diferentes puntos de vista, como analizaremos más adelante (interacción entre alumnos y autonomía). Por otro lado, en diversas fases se produce un informe escrito, con lo que se favorece la capacidad de argumentación del estudiante, al tiempo que se contribuye al desarrollo de su autonomía en el aprendizaje.

Sobre este componente de la idoneidad, y para valorar la interacción docente - discente, se plantearon las preguntas que se incluyen en la Tabla 6.7.3, y que describimos a continuación.

Tabla 6.7.3. Pauta de análisis de la idoneidad interaccional (Interacción docente-discente)

	Justificación
Interacción docente-discente	5.1. ¿Sirven las soluciones al proyecto para que el profesor reconozca y resuelva las dificultades y errores de los alumnos? 5.2. ¿Se usan en el proyecto diversos recursos retóricos y argumentativos para implicar y captar la atención de los alumnos? 5.3. ¿Facilita el proyecto la inclusión de los alumnos en la dinámica de la clase?

6.7.2.1. ANÁLISIS A PRIORI Y DESARROLLO

Al analizar este componente de la idoneidad interaccional, se espera que los futuros profesores perciban que la forma en que se ha trabajado el proyecto planteado permite diversas formas de discurso en el aula. Al comienzo, el profesor introducirá el proyecto, tratando de interesar a los estudiantes tanto por su temática, como por la metodología de enseñanza de la estadística, con la intención de que algunos estudiantes expongan o debatan sus propias ideas al respecto.

Se describe a continuación las preguntas que se plantean para valorar la interacción docente-discente en el desarrollo del proyecto.

5.1. ¿Sirven las soluciones al proyecto para que el profesor reconozca y resuelva las dificultades y errores de los alumnos?

Si el estudiante completa todos los apartados del proyecto, y participa activamente en los debates del aula, el profesor obtendría información de sus posibles dificultades, tanto en los contenidos requeridos previamente, como en los propios de correlación y regresión. Así, por ejemplo, podría detectar la asignación incorrecta de un signo al coeficiente de correlación; dificultades de ordenación de los coeficientes de correlación, descritas por Sánchez Cobo (1999); o atribución indebida de una relación causal a alguna de las variables, lo cual sería indicativo de la existencia de la concepción causal de la correlación (Estepa, 1994), entre otras.

Sería más difícil de identificar aquellas dificultades y conflictos de manejo de Excel, a menos que el estudiante preguntase específicamente al profesor. Aún así, se podría observar los posibles errores en las producciones finales del estudiante; por ejemplo, la confusión entre la variable dependiente e independiente (Ruiz, 2006) o el uso acrítico de las opciones del software. El grado en que es posible resolver los conflictos identificados dependerá del tiempo disponible y número de alumnos que lo presentan, pues si un conflicto aparece frecuentemente, el profesor insistiría en este punto, pero si afecta a un solo estudiante, podría pasar inadvertido.

5.2. ¿Se usan en el proyecto diversos recursos retóricos y argumentativos para implicar y captar la atención de los alumnos?

El proyecto parte de una pregunta de investigación que los estudiantes han de resolver, por lo cual, con una filosofía mixta (constructivista con apoyo del profesor, que proporciona los medios y conduce a los estudiantes a cada paso) se debe gestionar el aprendizaje. Podemos identificar en el proyecto y el trabajo con el mismo varios recursos retóricos y argumentativos: la exposición por parte del profesor; el uso de ejemplos y contraejemplos; el planteamiento de preguntas abiertas que se van cerrando progresivamente; el uso de la transnumeración para facilitar la comprensión; la exploración de datos mediante gráficos dinámicos del servidor de Naciones Unidas; la exploración de conceptos mediante simuladores, entre otros.

En el trabajo con el proyecto por los futuros profesores, el mismo análisis de datos y sus resultados se usaron como argumentos para apoyar las conclusiones sobre las intuiciones de la clase. El debate entre futuros profesores, así como entre el formador de profesores y los participantes, sirvió para resolver los conflictos que se presentaron. Por todo ello, se espera que los futuros profesores valoren positivamente el uso de estos recursos para implicar y captar la atención de los estudiantes de Bachillerato, habiendo vivido ellos mismos esta experiencia.

5.3. ¿Facilita el proyecto la inclusión de los alumnos en la dinámica de la clase?

Los futuros profesores participaron en la clase de modo activo, resolviendo de modo individualizado y en parejas las tareas, y participando en el debate conjunto de las soluciones. Las frecuentes preguntas del profesor a los estudiantes, posibilitan que los estudiantes integren diferentes puntos de vista sobre las tareas que resuelven, así como que ellos también planteen nuevas preguntas. Por todo ello, se espera que reconozcan la potencialidad del proyecto para la inclusión de los estudiantes en la dinámica de la clase; al igual que en los casos anteriores, recordando su experiencia en el desempeño de la tareas.

6.7.2.2. RESULTADOS DE LA EVALUACIÓN

Se presentan a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2 a 5.

Nivel 2. Se aplica parte del descriptor correctamente, o se cometen errores. También se puede hacer uso de aspectos no estrictamente matemáticos, o anecdóticos.

Encontramos este nivel de respuesta, principalmente, en la valoración del descriptor 5.1, bien porque no centran la respuesta específicamente en el proyecto (MAG); se refieren a aspectos anecdóticos (CL), o porque aplican parte del mismo; así por ejemplo, se centran en cómo están formuladas las preguntas (DG), o si son adecuadas para evaluar conceptos (JG):

> Sí, siempre se puede preguntar dudas al profesor y según sus respuestas se pueden ver sus dificultades y errores. (DG; descriptor 5.1).
> Sí, en el proyecto se pueden encontrar los errores habituales ya que se puede evaluar y seleccionar las mayores dificultades. (JG; descriptor 5.1).
> Si el trabajo se realiza en grupo habría que estar atento a los errores que puedan cometer aunque entre ellos mismos los podrían subsanar (CL; descriptor 5.1).

También encontramos participantes que responden al descriptor 5.2 o 5.3 en este nivel, aplicándolo correctamente pero sin centrarse en aspectos propios del proyecto (AJD)

> Sí, porque tienen que prestar atención para posteriormente poder responder. (AJD; descriptor 5.2).
> Sí, porque es un proyecto. (JMV; descriptor 5.3).

Nivel 3. Se hace una aplicación correcta del descriptor, pero de modo incompleto ya que se centra en otros objetos matemáticos no recogidos por el mismo.

Los futuros profesores valoran únicamente el descriptor 5.2 en este nivel, además, con gran frecuencia. Los participantes se refieren al lenguaje; a aspectos de las situaciones problema; o incluso a aspectos procedimentales por el uso del ordenador:

> Sí, se usa imágenes de gráficos, pruebas escritas, y uso del ordenador para realizar cálculos y búsquedas de datos. (JMV; descriptor 5.2).
> Los programas informáticos y el power point. (JG; descriptor 5.2).
> Sí, pero creo que lo que más capta la atención de los alumnos es que sean actividades en un contexto distinto al que están acostumbrados. (ME; descriptor 5.2)

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

Lo futuros profesores suelen reconocer que el proyecto contribuye a la inclusión de los estudiantes en la dinámica de la clase (descriptor 5.3) en este nivel, señalando la importancia del debate, el interés de la problemática que trata, la oportunidad de trabajar en grupo o por parejas determinadas tareas, el interés por el uso de la tecnología, así como el formato de las tareas:

> Sí, porque participan en el proyecto, sobre todo si, posteriormente, se hace una puesta en común. (MC; descriptor 5.3).
> Depende tanto del profesor como del alumno, aunque está pensado para que sí se participe. (DG; descriptor 5.3).
> Sí, permite hacer trabajos en grupo o por parejas (análisis de otras variables). (MAG; descriptor 5.3).
> Sí, porque el formato de las actividades es más atractivo que la típica tarea: “calcula la media” o “calcula r para estas dos variables”… (ME; descriptor 5.3).

Para el resto de descriptores, sólo un futuro profesor se encuentra en este nivel de respuesta en cuanto al descriptor 5.1, ya que sólo se refiere a la interpretación de los gráficos para reconocer dificultades y errores de los estudiantes:

> Sí, a la hora de interpretar los gráficos. (AJD; descriptor 5.1).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Encontramos este nivel de respuesta en la valoración de todos los descriptores. Por ejemplo, en el descriptor 5.1 el futuro profesor no sólo valora la interpretación de los gráficos, sino que señala la relación de conceptos y su dificultad para el reconocimiento de otras destrezas como errores en el cálculo (MC), o valora además las soluciones a las tareas de regresión (ChC), entre otras argumentaciones.

> Sí, a la hora de interpretar los gráficos y asignar el nivel de ajuste el profesor puede observar si el alumno ha entendido o no los conceptos (ChC; descriptor 5.1).
> Sirven sólo para reconocer si el alumno sabe la interpretación y relación de los conceptos. Pero no para reconocer errores en el cálculo o en el desarrollo de algunas destrezas, (MC; descriptor 5.1).

Tabla 6.7.4. Frecuencia (y porcentaje) de niveles en la valoración de la idoneidad interaccional docente-discente

Nivel	Descriptor 5.1	Descriptor 5.2	Descriptor 5.3
0	1(4,8)	1(4,8)	1(4,8)
1		1(4,8)	
2	13(61,9)	7(33,3)	3(14,3)
3		9(42,9)	
4	1(4,8)		11(52,4)
5	6(28,6)	3(14,3)	6(28,6)
Total	21(100)	21(100)	21(100)

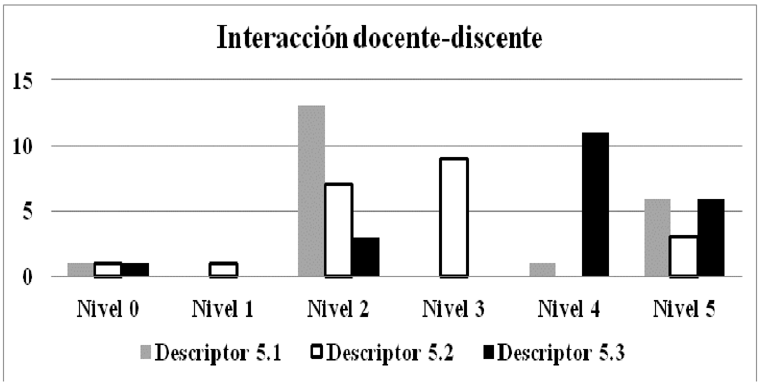


Figura 6.7.2. Frecuencia (y porcentaje) de niveles en la valoración de la interacción docente-discente.

Podemos observar (Tabla 6.7.4 y Figura 6.7.2) el alto porcentaje de respuestas en el nivel 2, principalmente en la valoración al descriptor 5.1, pudiendo dar indicios de cierta dificultad en los futuros profesores de relacionar las soluciones de las tareas con las dificultades y errores que los estudiantes muestren en el aprendizaje del tema. Del mismo modo encontramos las valoraciones al descriptor 5.2, donde los futuros profesores no se refieren a los recursos retóricos y argumentativos utilizados en el proyecto, a pesar de haber sido explicados previamente, destacando mayormente la influencia del uso de la tecnología en la atención del alumno. El descriptor 5.3 (potencial del proyecto para la inclusión de los alumnos en la dinámica de la clase) sería el mejor valorado, con un alto índice de respuesta en los niveles 4 (52,4%) y 5 (28,6).

6.7.3. INTERACCIÓN ENTRE ALUMNOS Y AUTONOMÍA

El proyecto se fundamenta en el aprendizaje constructivista. Se trata de provocar en el estudiante la necesidad del nuevo conocimiento matemático que se enseña, a partir de las situaciones problema que se plantean. Para ello, las tareas tienen un fuerte componente de trabajo en grupo, con objeto de que los estudiantes confronten sus ideas con el resto de compañeros, y se responsabilicen de integrarlas para aplicarlas en las siguientes tareas que se plantean.

El primer debate comienza con las actividades iniciales sobre la Esperanza de vida, diseñadas como evaluación previa del conocimiento del alumno. Una vez realizados los cálculos y el ajuste de las líneas de regresión con Excel, por parejas o en pequeños grupos (3 o 4 estudiantes), se confrontarán los resultados obtenidos. Por último, las preguntas finales (ítems de evaluación y problema abierto) contribuyen a la evaluación. Sobre este componente de la idoneidad interaccional se plantearon las preguntas que se incluyen en la Tabla 6.7.5, y que describimos a continuación.

Tabla 6.7.5. Pauta de análisis de la idoneidad interaccional entre alumnos y la autonomía

	Justificación
Interacción entre alumnos	5.4. ¿Se favorece el diálogo y comunicación entre los estudiantes? ¿Cómo?
Autonomía	5.5. ¿Se contemplan momentos en los que los estudiantes asumen la responsabilidad del estudio? ¿Por qué?

6.7.3.1.ANÁLISIS A PRIORI Y DESARROLLO

5.4. ¿Se favorece el diálogo y comunicación entre los estudiantes? ¿Cómo?

El diálogo entre los estudiantes, como se ha dicho, se vio favorecido desde el comienzo del curso con el trabajo en pequeños grupos o parejas. Además, en la práctica que se llevó a cabo en el aula de clase, el profesor, en determinados momentos, fomentó el debate, donde tuvieron que discutir y argumentar sobre sus soluciones. Se espera que el futuro profesor valore su experiencia con el proyecto y la traslade a un aula de Bachillerato.

5.5. Se contemplan momentos en los que los estudiantes asumen la responsabilidad del estudio? ¿Por qué?

Este proyecto, en particular, se caracterizó por el grado de autonomía que en el aprendizaje se otorgaba a los futuros profesores, quienes tuvieron libertad para elegir las funciones de ajuste (en el estudio de la regresión), ampliar el análisis con nuevas variables, e interpretar nuevos gráficos o recursos interactivos en Internet. Por ello, el futuro docente podría considerar el trabajo realizado en el proyecto, y trasladarlo al de un estudiante de Bachillerato, ya que el desempeño de las tareas depende del propio estudiante y no tanto del profesor.

Mostramos a continuación ejemplos de respuestas de los futuros profesores a las preguntas descritas.

6.7.3.2. RESULTADOS DE LA EVALUACIÓN

Se presentan a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2 a 5.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra parte se omite o se cometen errores.

Encontramos un futuro profesor que valora la autonomía del estudiante en el estudio (descriptor 5.5) en este nivel, pues pareciera que lo único que hace el estudiante en el proyecto es interpretar los resultados de las tareas (DG). También encontramos este nivel al aplicar el descriptor 5.4:

> Depende de cómo se trate. Si dejamos cuestiones abiertas que fomente el debate se favorece la comunicación entre ellos. (MRA; descriptor 5.4).
> Creo que lo único que aportan es su interpretación de los resultados, (DG; descriptor 5.5).

Nivel 3. Se hace una aplicación correcta del descriptor, pero de modo incompleto ya que se centra en otros objetos matemáticos no recogidos por el mismo.

Encontramos un futuro profesor que aplica el descriptor 5.5 en este nivel, reconoce la necesidad del estudiante de argumentar sus respuestas, en lugar de indicar momentos en los que el alumno se responsabiliza de su aprendizaje:

> Sí, porque tienen que justificar sus respuestas según ellos piensen, (AJD; descriptor 5.5)

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

Lo futuros profesores suelen reconocer que el proyecto contribuye al diálogo y comunicación entre los estudiantes (descriptor 5.1), aunque se refieren únicamente al debate como medio para ello:

> Sí, a la hora de poner en común los resultados, se establecen debates que favorecen el diálogo y comunicación entre los alumnos. (ChC; descriptor 5.4).
> En la resolución del proyecto, verificando los resultados, comparando y fomentando la participación del alumnado. (JG; descriptor 5.4).

En cuanto al descriptor 5.5, los futuros profesores suelen considerar como responsabilidad del estudiante, únicamente, la realización de las tareas de ampliación:

> Sí, el momento de realizar tareas que no son obligatorias es un momento de responsabilidad. (ChC; descriptor 5.5).
> Sí, por ejemplo, cuando los alumnos tienen que decidir qué recta se ajusta mejor a la nube de puntos y tienen la responsabilidad de decidir si es una curva o una recta. (JFM; descriptor 5.5).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Encontramos este nivel de respuesta en la valoración de ambos descriptores, como por ejemplo:

> En las preguntas iniciales y en la puesta común final se pueden hacer debates sobre la opinión de cada uno al respecto. No obstante, si hacemos que las actividades se hagan en grupos los alumnos pueden interactuar para decidir qué respuesta es más adecuada asociar a cada caso (PJ; descriptor

5.4).
Sí, porque por ejemplo, en el ejercicio de redactar, son ellos los que deben escribir lo que interpretan, y en el resto de ejercicios también, además hay actividades complementarias. (JG; descriptor 5.5)

A continuación se resumen (Tabla 6.7.6 y Figura 6.7.3) los resultados obtenidos en cada uno de estos niveles. Podemos observar la adecuación de las respuestas de los futuros profesores en la valoración de la interacción entre los alumnos y su autonomía en el aprendizaje en el desarrollo del proyecto, encontrándose principalmente en el nivel 4, ya que, como se comento anteriormente, los futuros profesores sólo indican el debate y las tareas de ampliación en las valoraciones a los descriptores 5.4 y 5.5, respectivamente.

Tabla 6.7.6. Frecuencia (y porcentaje) de niveles en la valoración de la interacción y autonomía

Nivel	Descriptor 5.4	Descriptor 5.5
0		1(4,8)
1		
2	3(14,3)	1(4,8)
3		1(4,8)
4	13(61,9)	10(47,6)
5	5(23,8)	8(38,1)
Total	21(100)	21(100)

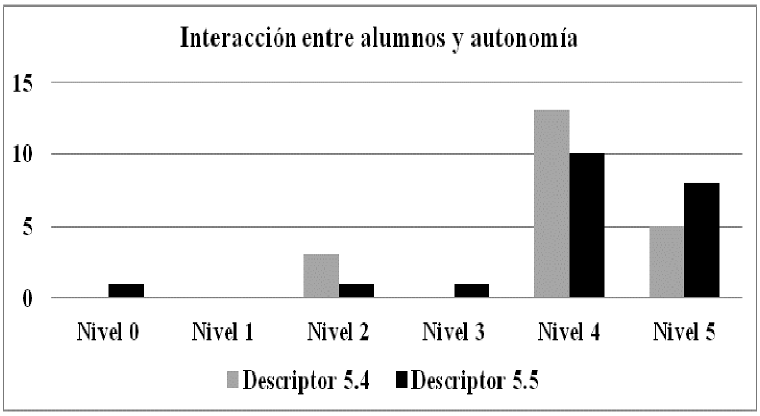


Figura 6.7.3. Frecuencia (y porcentaje) de niveles en la valoración de la interacción y autonomía

6.7.4. SÍNTESIS DE CONOCIMIENTOS DIDÁCTICOS EN LA FACETA MEDIACIONAL E INTERACCIONAL

Para resumir los conocimientos didácticos de los futuros profesores de la muestra en las facetas mediacional e interaccional, se han calculado las puntuaciones medias obtenidas en cada uno de los descriptores propuestos (Tabla 6.7.7). Este conocimiento sería parte del conocimiento del contenido y la enseñanza, según Hill, Ball, y Schilling (2008).

Los resultados muestran (Tabla 6.7.7 y Figura 6.7.4) puntuaciones por encima de la media teórica (2,5) en todos los apartados, y globalmente; y como en otros componentes de la idoneidad didáctica, mucho mejores que los correspondientes en la investigación de Arteaga (2011). En este caso, ningún descriptor supera al 4, no obstante, hay tres con esta puntuación: contextualización de situaciones, comunicación entre alumnos y autonomía y responsabilidad. Lo más difícil fue reconocer los recursos retóricos y argumentativos, quizás porque valoraron la mayor parte del proyecto, que se resolvió por escrito, sin considerar aquellos recursos utilizados de modo oral, en los debates.

Tabla 6.7.7. Medias y desviaciones típicas en los descriptores de la idoneidad mediacional e interaccional

Descriptor	Contenido	Media	D, Típica
D3.1	Valoración de recursos	3,6	1,4
D3.2	Contextualización de situaciones	4,0	0,7
D5.1	El profesor reconoce dificultades	2,9	1,5
D5.2	Recursos retóricos y argumentativos	2,7	1,2
D5.3	Inclusión de alumnos en la dinámica	3,8	1,3
D5.4	Comunicación entre alumnos	4,0	0,9
D5.5	Autonomía y responsabilidad	4,0	1,2
Total		3,6	1,3

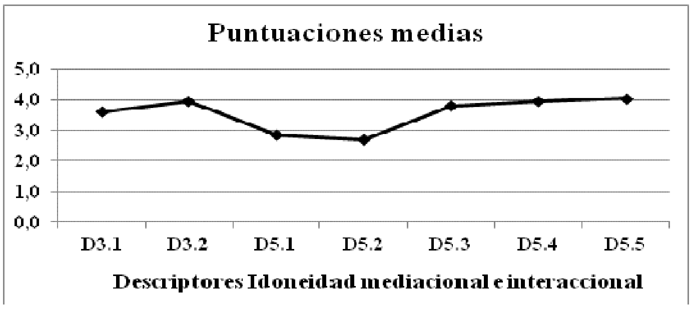


Figura 6.7.4. Puntuación media en la valoración de los descriptores de idoneidad mediacional e interaccional

Como en otros apartados, finalizamos la síntesis presentando ejemplos de conocimientos de las facetas interaccional y mediacional (Tabla 6.7.8), que algunos participantes mostraron en su análisis, especialmente los que razonan a nivel 4 y 5. Observamos, en primer lugar, la valoración crítica de la tecnología, no sólo como facilitadora de los cálculos, sino como recurso didáctico que permite la innovación y mejora de la motivación de los estudiantes.

Se valora también el formato de las tareas, que permite recoger datos de evaluación para reconocer y posteriormente solventar (en los periodos de debate) posibles conflictos de significado en el aprendizaje. Los futuros profesores también identifican la ayuda que supone el trabajo por parejas, y los momentos en que los estudiantes han de tomar responsabilidad, sobre todo al elegir variables explicativas, funciones de ajustes, o al plantear nuevos problemas.

Tabla 6.7.8. Ejemplos de conocimientos didácticos de los futuros profesores (Faceta interaccional y mediacional)

	Justificación
Recursos Materiales	1. Reconoce la conveniencia de los recursos informáticos usados en el proyecto para introducir los diferentes tipos de objetos matemáticos relacionados con la correlación y regresión. Por ejemplo, se reconocen las ventajas de visualización y cálculo proporcionadas por Excel y los applets interactivos. 2. Reconoce que la tecnología sirve como motivación a los estudiantes, y libera tiempo de cálculo, que puede dedicarse a actividades de interpretación. 3. Valora la contextualización de la matemática en situaciones de la vida real; se reconoce que el uso de datos reales (de las Naciones Unidas) aumenta la valoración que el estudiante puede dar al tema y, en general, a la estadística.
Interacción docente-discente	4. Reconoce que las soluciones al proyecto permiten al profesor identificar las dificultades y errores de los alumnos, y ayudar a superarlas, aunque también se indica que esto depende del profesor y el alumno, y no sólo del proyecto. Por ejemplo, a la hora de interpretar los gráficos y asignar la función de ajuste, el profesor puede observar si el alumno ha entendido o no los conceptos. 5. Indica que el formato de la tarea, más atractivo que el problema tipo, sirve para implicar y captar la atención de los alumnos. 6. Se valora que el proyecto permite la inclusión de los alumnos en la dinámica de la clase, en la resolución de problemas en grupo, en la exposición de las respuestas en clase, y al manejar conjuntamente un programa informático.
Interacción entre alumnos	7. Indica que el trabajo por parejas favorece el diálogo y comunicación entre los estudiantes, sobre todo a la hora de poner en común los resultados; se reconoce que los debates que se establecen contribuyen igualmente al diálogo y la comunicación.
Autonomía	8. Identifica los momentos en los que los estudiantes asumen la responsabilidad del estudio. Por ejemplo, al realizar tareas que no son obligatorias, o cuando tienen que decidir qué recta se ajusta mejor a la nube de puntos, ya que tienen la responsabilidad de decidir si es una curva o una recta.

6.8. VALORACIÓN DE LA IDONEIDAD ECOLÓGICA

Por último se analizan los informes en que los alumnos valoran la idoneidad ecológica, que, según Godino (2009), indica el grado en que el proceso de estudio se ajusta al proyecto educativo del centro y a la sociedad. Permitirá la reflexión sobre los aspectos que Shulman (2007) describe como conocimiento curricular, entendido como contexto educativo, fines, propósitos y valores de la educación.

Como se comentó en el análisis de la idoneidad epistémica (Sección 6.4), los futuros profesores conocen los contenidos que se concretan en la normativa vigente, no sólo los referidos al bloque de estadística y probabilidad; además, tienen una perspectiva intra/interdisciplinar de los mismos. Con todo ello, el formador de profesores explica a los participantes la importancia de conocer qué se debe enseñar sobre la correlación y regresión en primer curso de Bachillerato, según lo que indican las directrices curriculares vigentes, así como valorar la implementación y evaluación que se ha realizado con el desarrollo del tema respecto a las mismas.

Sobre esta componente de la idoneidad didáctica, se plantearon diversas preguntas, que atienden a diferentes aspectos de la idoneidad ecológica. Los agrupamos en dos ámbitos, uno referido la adaptación del proyecto al currículum y su apertura hacia la innovación didáctica, y otro a las conexiones intra e interdisciplinares, junto a su adaptación a la formación socio-profesional y cultural del estudiante.

6.8.1. ADAPTACIÓN AL CURRÍCULO Y APERTURA HACIA LA INNOVACIÓN DIDÁCTICA

Para valorar la adaptación a las directrices curriculares, y su apertura hacia la innovación didáctica, se plantearon las preguntas que se presentan en la Tabla 6.8.1, y que a continuación se describen.

Tabla 6.8.1. Pauta de análisis de la valoración de la idoneidad ecológica (adaptación curricular, formación socio-profesional, apertura a la innovación, conexiones)

	Justificación
Adaptación al currículo	6.1. Los contenidos, su implementación y evaluación ¿se corresponden con las directrices curriculares?
Apertura hacia la innovación didáctica	6.2. ¿Hay en el proyecto un componente de innovación? 6.3. ¿Se integran las nuevas tecnologías en el proyecto educativo?

6.8.1.1. ANÁLISIS A PRIORI Y DESARROLLO

El formador de profesores explicó a los participantes la importancia de considerar las directrices curriculares en el diseño, implementación y evaluación de la enseñanza del tema. Además, se explicó la relevancia que la Comunidad Autónoma de Andalucía otorga a la actividad investigadora en el aula, como fuente de conocimiento, con objeto de conectar la teoría y la práctica (Consejería de Educación, 2008b). A continuación, se describen las preguntas que se plantearon al respecto:

6.1. Los contenidos, su implementación y evaluación ¿se corresponden con las directrices curriculares?

Como se ha indicado, el proyecto trabaja los contenidos "*Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal*", del Bachillerato en *Ciencias y Tecnología*, que se amplían a "*Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación grafica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados*" en Ciencias Sociales. También trabaja los contenidos de estadística descriptiva unidimensional, que se concretan en este mismo bloque: "*Tipos de variables. Métodos estadísticos. Tablas y gráficos. Parámetros estadísticos de localización, de dispersión y de posición*". Se espera que los futuros profesores valoren la adaptación del proyecto a los mismos, y que además, consideren su vinculación con el tratamiento de otros contenidos de aritmética y algebra, geometría, y análisis, todos ellos incluidos en el Decreto de Enseñanzas Mínimas (MEC, 2007b).

Por otra parte, el Real Decreto indica que las matemáticas deben responder a la necesidad de resolver problemas prácticos reales, que ayuden a formalizar las intuiciones del estudiante, e incluyan el uso de herramientas tecnológicas como mejora de la comprensión de conceptos y en la resolución de problemas. Se espera que los futuros profesores valoren la pertinencia de la metodología del trabajo con proyectos en cuanto a estas orientaciones.

6.2. ¿Hay en el proyecto un componente de innovación?

Se trata de una innovación docente, y como tal, se plantea dentro de la asignatura Innovación docente e Investigación educativa en Matemáticas (descrita en el Capítulo 5). Una innovación es un cambio con mejora, respecto a un objetivo previamente determinado. La mejora puede ser una idea, un material, una práctica, un contenido o una metodología, pues hay una gran diversidad de experiencias que pueden considerarse una innovación educativa (Ortega et al., 2007). Estos autores introducen los siguientes criterios para caracterizar la investigación educativa, que se cumplen en el Proyecto:

- *Novedad:* La innovación introduce algo nuevo, que origina una mejora con respecto a una situación bien definida en un contexto o hábitos. En nuestro caso, la novedad se origina por el tipo de problema planteado (problema realista en vez de artificial), los datos utilizados (muchas más variables y casos que lo habitual) y el método (introducir los problemas antes que los conceptos), así como algunos de los gráficos usados en las actividades de ampliación.
- *Intencionalidad e interiorización:* Hay una intención explícita de cambiar (la enseñanza tradicional) y los actores del sistema lo aceptan (en este caso el formador y los futuros profesores).
- *Creatividad:* Hay un elemento original. Aunque el trabajo con proyectos estadísticos tiene ya una amplia tradición, las preguntas planteadas, las variables elegidas, y la forma en que se plantean son originales.
- *Profundidad:* La innovación genera cambios en las concepciones, las actitudes y las prácticas educativas; supone una verdadera transformación. En este caso, se trata sólo de una propuesta, pero de aceptarse en un centro educativo, supondría un cambio en las creencias de los docentes y en sus métodos didácticos.
- *Pertinencia*. La innovación propuesta sería pertinente, en cuanto recoge los principales contenidos educativos, a la vez que permite un aprendizaje más completo por parte de los estudiantes; no requiere muchos medios y es factible en un centro de Bachillerato.

Se espera que los futuros profesores se refieran a alguno de estos aspectos, desde su experiencia en el desarrollo del mismo, y esto se conjugue con la necesidad que indican las orientaciones curriculares de que la resolución de problemas ofrezca una visión amplia y científica de la realidad (MEC, 2007b).

6.3. ¿Se integran las nuevas tecnologías en el proyecto educativo?

Esta pregunta repite de algún modo alguna de las anteriores, pues ya se ha analizado con detalle el modo en que se integra la tecnología a lo largo del proyecto. Se espera que el futuro profesor se refiera a ellas, si no lo hubiese hecho anteriormente.

6.8.1.2. RESULTADOS DE LA EVALUACIÓN

Se presentan a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2, 4 y 5.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra parte se omite, se cometen errores, o se hace uso de aspectos no estrictamente matemáticos.

Encontramos muy pocas respuestas de los futuros profesores en este nivel, y en general, se deben a no especificar aspectos del proyecto, como por ejemplo:

> Sí, porque están basados en unos conocimientos previos que deben tener adquiridos los alumnos si se han aplicado los contenidos curriculares (ATL; descriptor 6.1).
> No lo sé, quizá la utilización de programas informáticos en clase, trabajar de manera online (CM; descriptor 6.2).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

Encontramos este nivel principalmente en la valoración de la adaptación del proyecto al currículo (descriptor 6.1), y en su mayoría, se refieren a la actividad inicial de identificación de contenidos. Consideramos la respuesta en este nivel, ya que no se incluye referencia a la metodología o la evaluación que se realiza en el proyecto:

> Sí, porque hicimos el ejercicio de la correspondencia entre los contenidos vistos aquí y los que se encuentran en el currículo. (MC; descriptor 6.1).
> Sí, se adaptan al contenido del currículo de matemáticas como vimos anteriormente, en el ejercicio que teníamos que subrayar (CM; descriptor 6.1).

También encontramos que los futuros profesores valoran la innovación (descriptor 6.2) en este nivel, haciendo referencia al contexto, uso de datos reales o nuevas tecnologías:

> Sí, porque trabajar con las nuevas tecnologías es otra forma de trabajar en el aula (MC; descriptor 6.2).
> Sí, el uso de datos reales y oficiales, que no suele ser común en el aula (MAG; descriptor 6.2).
> Sí, podemos usar hojas de cálculo para completar el estudio y representar los datos en tablas o gráficos (CL; descriptor 6.3).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Encontramos este nivel de respuesta principalmente al valorar la incorporación de las nuevas tecnologías en el proyecto (descriptor 6.3), donde los futuros profesores suelen indicar el uso de Internet y la hoja de cálculo Excel, como componentes de la innovación:

> Sí, se usa Internet, hoja de cálculo, applets… (AJD; descriptor 6.3).
> Sí, al usar applets y herramientas informáticas como el Excel, recoger datos de Internet (MAG; descriptor 6.3).

También encontramos respuestas en este nivel en los descriptores 6.1 y 6.2, como por ejemplo:

> Los contenidos se encuentran en el currículo y la presentación de las actividades y la forma de evaluar también. (ME; descriptor 6.1).
> Sí, entran dentro de la estadística bidimensional descriptiva y de la geometría analítica (uso de la recta) (MAG; descriptor 6.1).
> Sí, ya que se utilizan nuevas tecnologías como programas informáticos e internet. (EGO;

descriptor 6.2).

Se resumen, en la Tabla 6.8.2 y Figura 6.8.1, los resultados obtenidos en cada uno de estos niveles, donde vemos que, en genera,l estos descriptores se aplican en un nivel alto (4 y 5); como en los casos anteriores, los resultados son bastante mejores que en Arteaga (2011). En los descriptores 6.1 y 6.2 los participantes se limitan a mostrar un ejemplo, por lo que la mayoría está a nivel 4, mientras que en el tercero, al citar las nuevas tecnologías, suelen proponer varios ejemplos, razonando a nivel 5.

Tabla 6.8.2. Frecuencia (y porcentaje) de niveles en la valoración de la adaptación curricular y la innovación didáctica

Nivel	Descriptor 6.1	Descriptor 6.2	Descriptor 6.3
0	2(9,5)	1(4,8)	
1			1(4,8)
2	2(9,5)	1(4,8)	1(4,8)
3			
4	14(66,7)	12(57,1)	5(23,8)
5	3(14,3)	7(33,3)	14(66,7)
Total	21(100)	21(100)	21(100)

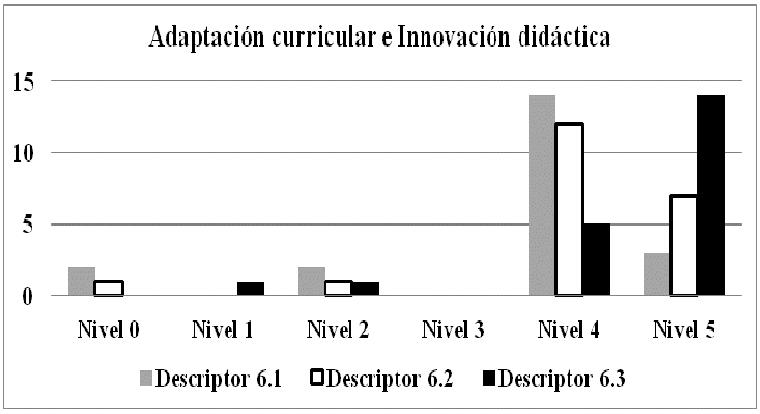


Figura 6.8.1. Frecuencia (y porcentaje) de niveles en la valoración de la adaptación curricular y la innovación didáctica

6.8.2. ADAPTACIÓN SOCIO-PROFESIONAL Y CULTURAL. CONEXIONES INTRA E INTERDISCIPLINARES

El formador de profesores explicó a los participantes la importancia de proyectar en el alumno la enseñanza hacia su futuro más inmediato, así como relacionar la matemática con otras materias. En las directrices curriculares encontramos indicaciones en este sentido, y en la Comunidad Autónoma de Andalucía se señalan explícitamente estos principios (Consejería de Educación, 2008b, p.98):

> b) La visión interdisciplinar del conocimiento, resaltando las conexiones entre diferentes materias y la aportación de cada una a la comprensión global de los fenómenos estudiados.
> c) La aplicación de lo aprendido a las situaciones de la vida cotidiana, favoreciendo las actividades que capaciten para el conocimiento y análisis del medio que nos circunda y de las variadas actividades humanas y modos de vida.
> d) El aprovechamiento de las diversas fuentes de información, cultura, ocio y estudio presentes en la sociedad del conocimiento.

e) La toma de conciencia sobre temas y problemas que afectan a todas las personas en un mundo globalizado, entre los que se considerarán la salud, la pobreza en el mundo, el agotamiento de los recursos naturales, la superpoblación, la contaminación, el calentamiento de la Tierra, la violencia, el racismo, la emigración y la desigualdad entre las personas, pueblos y naciones.

Para valorar la adaptación socio-profesional y cultural de los contenidos tratados en el proyecto, y las conexiones intra e interdisciplinares que se consiguen con su desarrollo, se plantearon las preguntas que se presentan en la Tabla 6.8.3, y que a continuación se describen, presentnado, además, el análisis de respuestas que los futuros profesores realizan al valorar dichos descriptores.

Tabla 6.8.3. Pauta de análisis de la valoración de la idoneidad ecológica

	Justificación
Adaptación socio-profesional y cultural	6.4. ¿Contribuyen los contenidos a la formación socio-profesional de los estudiantes?
Conexiones intra e interdisciplinares	6.5. ¿Se relacionan los contenidos con otros contenidos intra e interdisciplinares ¿con cuáles?

6.8.2.1. ANÁLISIS A PRIORI Y DESARROLLO

6.4. ¿Contribuyen los contenidos a la formación socio-profesional de los estudiantes?

Los contenidos estadísticos, en general, contribuyen en gran medida a la formación profesional, pues la estadística es una de las disciplinas más incluidas en todo tipo de formación posterior, tanto universitaria como en el trabajo profesional. Como indican Arteaga, Batanero, Contreras y Cañadas (2011), la constante presencia de la estadística en nuestra sociedad, contribuye a mostrar su utilidad, como una herramienta metodológica que permite analizar la variabilidad, determinar relaciones entre variables, diseñar estudios y experimentos, y tomar decisiones adecuadas en situaciones de incertidumbre.

Esta conexión entre la estadística y su aplicabilidad en la vida cotidiana, se podría llevar al aula, aprovechando la presencia de datos y de todo tipo de gráficos que encontramos en los medios de comunicación (Espinel, 2007). Además, el aprendizaje de la construcción e interpretación de gráficos bivariantes y multivariantes, presentes en el proyecto, permitirá mejorar el aprendizaje de otras asignaturas, ya que son muchas las ciencias que utilizan este tipo de representaciones.

Finalmente, destacamos que el proyecto contribuye al desarrollo del razonamiento correlacional, que se encuentra presente en la vida cotidiana del ser humano, como actividad cognitiva fundamental para su desarrollo (Moritz, 2004; Zieffler, 2006). Aunque, como se ha visto en los antecedentes son muchas las dificultades de los estudiantes con este tipo de razonamiento. Esperamos que los futuros profesores sean conscientes de todas estas contribuciones del proyecto a la formación de los estudiantes.

6.5. ¿Se relacionan los contenidos con otros contenidos intra e interdisciplinares ¿con cuáles?

Ya se analizaron anteriormente las muchas conexiones de la correlación y regresión con otras ramas de las matemáticas (intradisciplinares), y se espera que el futuro profesor indique aquellas que considere más relevantes. En cuanto a las conexiones interdisciplinares que se generan con el proyecto, se espera que consideren su relación con las ciencias sociales o la economía, siendo la principal de ellas la demografía, ya que la principal variable dependiente es la Esperanza de vida, siendo tomadas las variables explicativas,de campos muy diversos como ciencias de la salud, educación y economía. Luego su vinculación también es pertinente.

Sería posible también ampliar estas conexiones, tomando otras variables incluidas en el servidor de las Naciones Unidas.

6.8.2.2. RESULTADOS DE LA EVALUACIÓN

Se presentan a continuación ejemplos de respuestas de los futuros profesores al aplicar cada uno de los descriptores en los niveles 2, 4 y 5.

Nivel 2. Se aplica parte del descriptor correctamente, mientras que otra parte se omite, se cometen errores, o se introducen aspectos no matemáticos.

Encontramos pocas respuestas en este nivel, que se deben principalmente a que los futuros profesores no se centran específicamente en el proyecto. Por ejemplo, ChC se refiere al coeficiente de correlación, pero indica el de variación, y además contesta mal la segunda parte, pues no se trataron los contrastes de hipótesis ni intervalos de confianza:

> Sí, por ejemplo el coeficiente de variación con la media y desviación típica. La media y desviación típica con los intervalos de confianza y contraste de hipótesis (ChC; descriptor 6.5).

Nivel 4. El futuro profesor hace una aplicación correcta y consistente del descriptor, aunque razona mediante un único ejemplo.

Estas respuestas son frecuentes, principalmente para el descriptor 6.5. Los futuros profesores se suelen referir al análisis matemático como contenidos intramatemáticos, y a contenidos de ciencias sociales como interdisciplinares. En cuanto a la contribución del proyecto en la formación socio-profesional de los estudiantes (descriptor 6.4), generalmente se refieren a la utilidad de la estadística en diversos ámbitos profesionales, o a la alfabetización estadística:

> Sí, porque son contenidos que se usan en diversos ámbitos profesionales (MC; descriptor 6.4).
> Con contenidos de asignaturas de Ciencias Sociales (ME; descriptor 6.5).
> Sí, Por ejemplo con análisis, identificación de funciones a través de su representación gráfica, estimación, errores, parámetros sociales, etc. (AMA; descriptor 6.5).

Nivel 5. Se hace una aplicación correcta y consistente del descriptor mediante dos o más ejemplos.

Encontramos varias respuestas en ambos descriptores para este nivel. Algunos ejemplos son:

> Se relacionan los contenidos estadísticos con los de análisis (funciones). También se relacionan

con materias de la rama de sociales donde se estudian sociología-demografía-estado de bienestar (ATL; descriptor 6.5)
Se pueden relacionar con contenidos de asignaturas de Ciencias Sociales: economía, geografía e historia… (CL; descriptor 6.5)

Destacamos la respuesta de un futuro profesor, pues valora la contribución de los contenidos a la formación socio-profesional de estudiantes de Bachillerato en *Humanidades y Ciencias Sociales*, sin pensar en los beneficios que puede aportar a los estudiantes en la otra modalidad:

Sí, sobre todo si nos encontramos en el Bachillerato de Humanidades y Ciencias Sociales pues conectamos asignaturas y los introducimos en estudios de índole social (PJ; descriptor 6.4).

A continuación, se resumen (Tabla 6.8.4 y Figura 6.8.2) los resultados obtenidos en cada uno de estos niveles. Vemos que, en general, en estos dos descriptores la mayor parte de respuestas se distribuyen en los niveles 4 y 5, mientras que en la investigación de Arteaga (2011) se quedaron en los niveles 1 y 2; de nuevo, se observa la mejor competencia de los futuros profesores de Secundaria y Bachillerato.

Tabla 6.8.4. Frecuencia (y porcentaje) de niveles en la valoración de la adaptación a la formación socio-profesional y cultural y la conexión intra e interdisciplinar

Nivel	Descriptor 6.4	Descriptor 6.5
0	4(19)	1(4,8)
1		
2	1(4,8)	3(14,3)
3		
4	6(28,6)	9(42,9)
5	10(47,6)	8(38,1)
Total	21(100)	21(100)

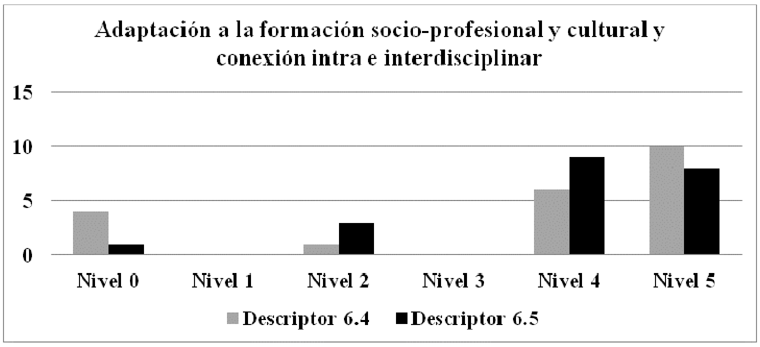


Figura 6.8.2. Frecuencia (y porcentaje) de niveles en la valoración de la adaptación a la formación socio-profesional y cultural y la conexión intra e interdisciplinar

6.8.3. SÍNTESIS DE CONOCIMIENTOS DIDÁCTICOS EN LA FACETA ECOLÓGICA

Para resumir los conocimientos didácticos de los futuros profesores de la muestra en la faceta epistémica, se han calculado las puntuaciones medias obtenidas en cada uno de los descriptores propuestos. Este conocimiento sería parte del conocimiento del contenido y el currículo según Hill, Ball, y Schilling (2008). Los resultados, de nuevo, son buenos (Tabla 6.8.5 y Figura 6.8.3), mostrando la madurez de los participantes; también la pequeña desviación típica es signo de homogeneidad entre ellos; y todos superan el nivel 3, acercándose mucho al 4, que referente a las nuevas tecnologías es superado, con un valor medio de 4.4.

Tabla 6.8.5. Medias y desviaciones típicas en los descriptores de la idoneidad ecológica

Descriptor	Contenido	Media	D, Típica
D6.1	Directrices curriculares	3,6	1,4
D6.2	Innovación didáctica	4,0	1,1
D6.3	Nuevas tecnologías	4,4*	1,0
D6.4	Formación socio-profesional y cultural	3,6	1,9
D6.5	Relación intra e interdisciplinar	3,9	1,3
Total		3,9	1,4

* Media (en la aplicación del descriptor) superior a 4

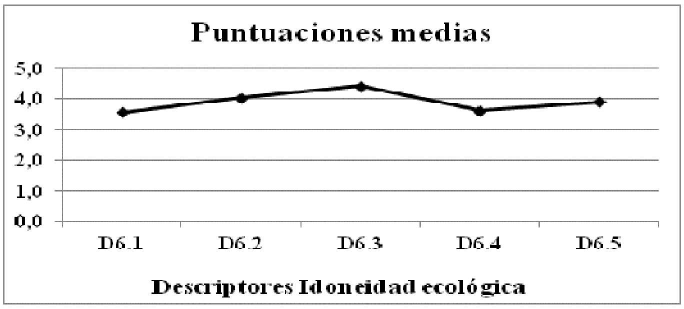


Figura 6.8.3. Puntuación media en los descriptores de la idoneidad ecológica

Tabla 6.8.6 Ejemplos de conocimientos didácticos de los futuros profesores (Faceta ecológica)

		Justificación
Adaptación al currículo	6.6.	Reconoce los contenidos curriculares para el tema, dentro de los implementados en el proyecto; en particular, el estudio de las distribuciones bidimensionales, correlación y regresión, y su aplicación a problemas sociales y económicos.
	6.7.	Identifica los principios metodológicos citados en las directrices curriculares, como uso de la tecnología o de contextos reales en el proyecto.
Apertura hacia la innovación didáctica	6.8.	Indica características de la innovación didáctica, como problemas no convencionales y abiertos, uso de visualizaciones dinámicas, uso de datos reales tomados de internet, y autonomía del estudiante para la realización del proyecto.
	6.9.	Valora positivamente la utilización de la tecnología para el cálculo, tratamiento de datos, visualización y motivación del alumnado.
Adaptación socio-profesional y cultural	6.10.	Reconoce la utilidad del aprendizaje para la futura vida profesional o futuros estudios de los estudiantes.
Conexiones intra e interdisciplinares	6.11.	Determina relaciones del tema con álgebra y cálculo, aritmética, ciencias sociales y economía, así como el mejor conocimiento del mundo actual por parte del alumno.

Como en los anteriores componentes, presentamos en la Tabla 6.8.6 ejemplos de conocimientos puestos de manifiesto en la muestra de futuros profesores en el análisis de la idoneidad ecológica. Observamos, como algunos de ellos reconocen, las características de una innovación docente en el proyecto; y son capaces de establecer conexiones intra e inter disciplinares con la futura formación y vida profesional del estudiante.

6.9. SÍNTESIS DE LA EVALUACIÓN

Para finalizar el Estudio 3, y valorar globalmente el conocimiento didáctico puesto de manifiesto en la actividad por los futuros profesores, se compara a continuación la dificultad relativa de aplicación de las diversas facetas en la pauta de análisis de la idoneidad didáctica, y se calcula y analiza una puntuación total.

6.9.1. DIFICULTAD COMPARADA DE LAS FACETAS DEL CONOCIMIENTO DIDÁCTICO

Cada uno de los componentes de la idoneidad didáctica ha tenido un número diferente de descriptores. Para comparar estas facetas entre sí, para cada alumno, se ha calculado el nivel medio alcanzado en cada una de las facetas; y en la Tabla 6.9.1 presentamos la media y desviación típica global del conjunto de participantes.

Tabla 6.9.1. Medias y desviaciones típicas en las componentes de la idoneidad didáctica

Descriptor	Media	D. Típica
Epistémica	3,6	1,4
Cognitiva	3,9	1,4
Mediacional	3,8	1,1
Afectiva	3	1,9
Interaccional	3,5	1,4
Ecológica	3,9	1,4
Total	3,6	1,5

Observamos (Tabla 6.9.1), que el mínimo de estas medias (3) es mayor que el nivel medio teórico (2.5), de modo que los futuros profesores, en su conjunto, son capaces de aplicar, por encima de un nivel medio teórico, el análisis de los componentes de la idoneidad didáctica. También destacamos dos componentes que resultaron particularmente sencillos, alcanzando casi el nivel medio de 4, que son la idoneidad cognitiva y ecológica. Interpretamos este resultado como que los participantes son capaces de valorar adecuadamente y razonadamente los aspectos cognitivos (aprendizaje, conocimientos previos, atención a la diversidad, evaluación) y ecológicos (conexiones con el currículo, conexiones con otras materias, apertura a la innovación docente en el proyecto) de la enseñanza de la correlación y regresión con el proyecto. Los resultados son mucho mejores que en el estudio de Arteaga (2011), con futuros profesores de educación primaria, en los que estas puntuaciones medias en todas las componentes se situaron por debajo de 2.

En la Figura 6.9.1 presentamos los intervalos de confianza del 95% de estas puntuaciones medias, observando solapamiento entre ellos, lo que confirma la uniformidad de respuesta en torno al nivel 4, excepto en el caso de la idoneidad afectiva, que aparece con puntuación significativamente menor que el resto.

Interpretamos este hecho como que los futuros profesores tuvieron mayor dificultad en valorar los aspectos afectivos ligados al proyecto y cómo se tienen en cuenta en el mismo; posiblemente, por la menor formación recibida en el campo de la psicología en comparación con matemáticas, didáctica de la matemática, didáctica general o pedagogía, muestra de ello es el alto índice de no respuesta en cuanto a las situaciones del proyecto que propician la igualdad; o la contribución del proyecto a erradicar la fobia o miedo a las matemáticas, como se describió en secciones anteriores.

En la Figura 6.9.2, los gráficos de caja muestran la variabilidad relativa de la puntuación en cada componente de la idoneidad, siendo mayor también en la idoneidad afectiva; no obstante, aparecen casos atípicos de estudiantes en todas las puntuaciones, generalmente con resultados bajos; han sido futuros profesores que dejaron en blanco algunas preguntas, o bien tuvieron una tendencia a responderlas a nivel más bajo que sus compañeros.

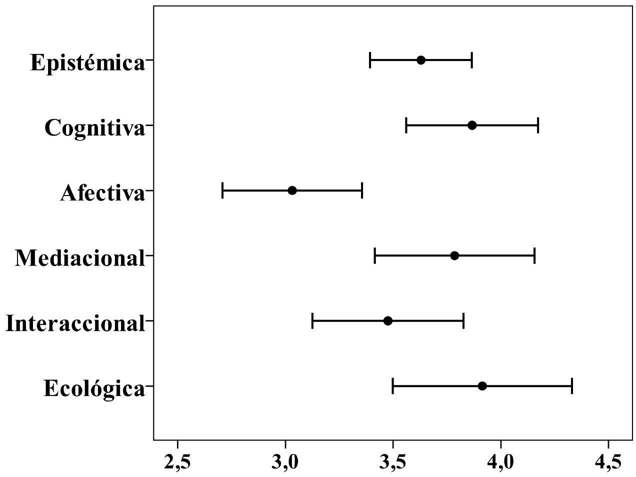


Figura 6.9.1. Puntuación media e intervalo de confianza del 95% en componentes de la idoneidad didáctica

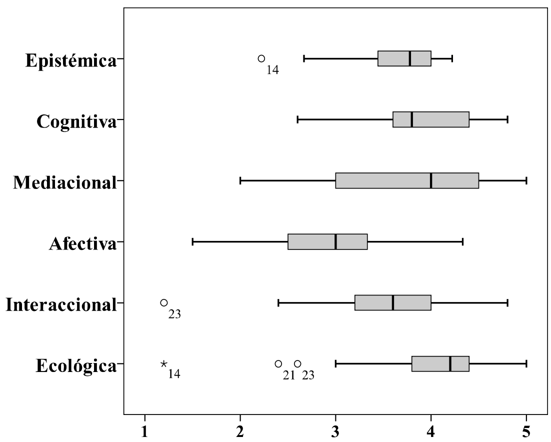


Figura 6.9.2. Gráficos de caja de nivel medio en componentes de la idoneidad didáctica

6.9.2. PUNTUACIÓN TOTAL EN LA PRUEBA DE EVALUACIÓN

También hemos calculado una puntuación total para cada estudiante, sumando las puntuaciones medias obtenidas de su valoración en los seis componentes de idoneidad didáctica (Figura 6.9.3). Esta puntuación, teóricamente, varía de 0 a 30 puntos, ya que un alumno tendría una puntuación cero si deja todos los descriptores en blanco y alcanzaría la puntuación 30 si fuese capaz de alcanzar un nivel medio 5 en todos los componentes de la idoneidad didáctica.

Observamos, en la Figura 6.9.3, una variación de la puntuación entre 16 y 27 puntos; es decir, no se llega a alcanzar el máximo 30 por ningún participante, pero si un valor muy cercano (27); el mínimo obtenido indica que prácticamente se aplican todos los descriptores a nivel medio cercano a 3. La mayor parte se sitúa entre 21 y 25 puntos (nivel medio entre 3,5 y 4).

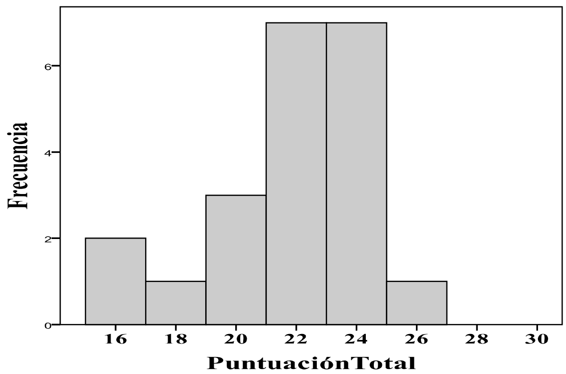


Figura 6.9.3. Puntuación total en la pauta de evaluación

Los resultados se pueden ver igualmente en el gráfico de caja (Figura 6.9.4), donde observamos que el 50% central de valores está en los límites señalados, y que, como en el caso de las puntuaciones por componentes, hay unos pocos casos atípicos que corresponden a futuros profesores que obtienen una baja puntuación global.

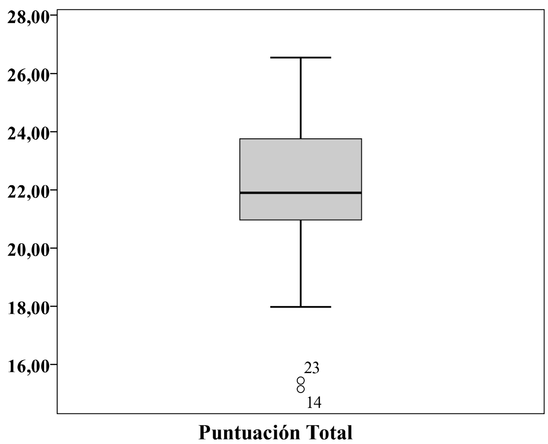


Figura 6.9.4. Puntuación total en la pauta de evaluación

Sugerencias de cambio

La valoración de la idoneidad didáctica del proyecto finalizó con una pregunta abierta en la que los futuros profesores podían aportar ideas para mejorar la idoneidad didáctica global. De los 21 profesores que completaron la valoración de la idoneidad didáctica del proyecto, sólo 12 hicieron estas recomendaciones. Es posible que el resto no tuvieran nuevas ideas para mejorar el proyecto, o bien, al ser largo el cuestionario, consideraron innecesario responder a esta pregunta.

La aportación que más se repite se refiere al contexto (8 futuros profesores, 38.1%), como por ejemplo CM, proponiendo contextos más cercanos al estudiante, como los personales.

> Quizá una adaptación más real a la vida diaria del alumno. Tratar temas más cotidianos o cercanos al alumno (CM).

También se propone la inclusión de actividades procedimentales (4 futuros profesores, 19%), como por ejemplo ATL; y encontramos dos estudiantes que sugieren ambos aspectos.

> Completaría el proyecto con actividades en las que sea necesario calcular los parámetros estadísticos. (ATL)

Destacamos las recomendaciones de tres futuros profesores, uno de ellos porque no se refiere tanto al proyecto, sino más bien a la metodología (PJ); otro pues considera que los contextos a trabajar en el Bachillerato de *Ciencias y Tecnología* debiesen ser más formales o no tan enfocados a la sociedad (ME), en contra de lo que se indica en las orientaciones curriculares; y otro futuro profesor (JMV), pues se refiere a la utilidad de incluir unas orientaciones al uso de la hoja de cálculo, previas al trabajo con el mismo:

> Proponer trabajar por grupos. Pedir un informe final con las conclusiones sobre el estudio que han realizado. Pedirles que diseñen después un proyecto ellos mismos sobre alguna cuestión real que les interese y den respuesta a la pregunta que se planteen de forma coherente utilizando técnicas similares a las usadas en este proyecto (PJ).
> Para la rama de Humanidades me parece genial, pero para la rama Tecnológica quizá se podrían usar otros datos relacionados con la Física, la Química, las Ingenierías… (ME).
> Tal vez sería bueno introducir algunas actividades de repaso de conocimientos y actividades introductorias al uso de la Hoja Excel en el ordenador para que no hubiera problema en su uso y en manejo (JMV).

6.10. CONCLUSIONES SOBRE EL CONOCIMIENTO DIDÁCTICO

A lo largo de este capítulo hemos descrito el Estudio 3, realizado con una de las muestras que tomó parte en el Estudio 2, y que estaba dirigido a evaluar y desarrollar algunos aspectos del conocimiento didáctico de los futuros profesores participantes en sus diferentes facetas.

Para ello, y una vez que los futuros profesores participantes resolvieron el proyecto sobre la Esperanza de vida descrito en el Capítulo 5, y se debatieron colectivamente las soluciones, llevaron a cabo un análisis de la idoneidad didáctica de la experiencia educativa vivida, utilizando una adaptación de la pauta propuesta por Godino (2009; 2011).

Para finalizar el Estudio 3, y al igual que se ha hecho con los anteriores, exponemos en este capítulo nuestras principales conclusiones, primeramente respecto a los objetivos, y a continuación respecto a las hipótesis planteadas para este estudio.

Conclusiones respecto a los objetivos

Como se expuso al comienzo del capítulo, el objetivo principal del Estudio 3 *fue realizar un estudio exploratorio del conocimiento didáctico matemático sobre la correlación y regresión, en estudiantes que se preparan para ser profesores de Matemáticas de Educación Secundaria y Bachillerato,* objetivo que se ha alcanzado en la medida que se cumplen los objetivos específicos planteados para tal fin, y que pasamos a explicar a continuación.

O3.1. Realizar un estudio exploratorio de evaluación de distintas facetas del componente didáctico del conocimiento didáctico matemático sobre correlación y regresión (Godino 2009; 2011), en una muestra de futuros profesores de Educación Secundaria y Bachillerato, puestos en juego al realizar el análisis de la idoneidad didáctica del proyecto descrito en el Estudio 2; al tiempo que se desarrollaba dicho conocimiento.

Pensamos que este objetivo se ha cumplido, a partir del análisis detallado de las respuestas escritas de los participantes al cuestionario proporcionado, en que se les plantean preguntas de valoración de las diferentes componentes y descriptores de la idoneidad didáctica. Puesto que previamente fue introducido a los participantes el concepto de idoneidad didáctica con otros ejemplos, la actividad contribuye, asimismo, al desarrollo de las diferentes facetas de su conocimiento didáctico.

Todo ello se muestra, en primer lugar, por el hecho de que los participantes han valorado a un nivel medio o alto (3 o superior) la mayoría de los descriptores propuestos, siendo capaces de analizar el proyecto que completaron en el Estudio 2 desde múltiples perspectivas, e identificando, para cada una de ellas, el grado en que el proyecto es o no idóneo para los fines previstos. Puesto que los participantes no tenían experiencias de enseñanza de estadística con proyectos, es claro que el conocimiento didáctico mostrado fue adquirido a partir de la experiencia realizada.

Los análisis realizados por los participantes fueron especialmente buenos (cercanos al nivel 4) en las facetas cognitivas, ecológica y mediacional, seguida por epistémica e interaccional (sobre 3,5) y más pobre la afectiva (3). Los participantes mostraron su competencia en identificar los conocimientos previos requeridos; la forma en que el proyecto contempla la evaluación y atención a la diversidad; las relaciones que se establecen con las directrices curriculares, con otras ramas de las matemáticas, en especial de las ciencias sociales; y el componente de innovación. En consecuencia, la guía de análisis de la idoneidad didáctica, adaptada de la de Godino (2009; 2011) se ha revelado un instrumento útil en la formación de las diferentes facetas del conocimiento didáctico del profesor.

Los participantes valoraron especialmente el uso de la tecnología en el trabajo con la estadística. Ello coincide con las orientaciones curriculares que recomiendan el uso de la tecnología en la enseñanza de las matemáticas, en las dos modalidades de Bachillerato en que se imparte esta materia, y concretan como objetivo de enseñanza el desarrollo de la capacidad de emplear los recursos tecnológicos para obtener y procesar información, y así facilitar la comprensión de fenómenos dinámicos. En la modalidad de *Humanidades y Ciencias Sociales*, por ejemplo, se indica:

> Hacer uso de variados recursos, incluidos los informáticos, en la búsqueda selectiva y el tratamiento de la información gráfica, estadística y algebraica en sus categorías financiera, humanística o de otra índole, interpretando con corrección y profundidad los resultados obtenidos de ese tratamiento. (MEC, 2007b; p.45475)

Un segundo objetivo fue *Identificar y clasificar algunos ejemplos de conocimientos relacionados con cada una de las facetas del conocimiento didáctico en el modelo de Godino (2009; 2011) de los futuros profesores, e interpretarlas en las diferentes categorías del modelo MKT (Hill, Ball y Schilling, 2008).*

También pensamos que este objetivo se ha alcanzado razonablemente, pues el análisis cualitativo de las respuestas de los futuros profesores a la Guía de análisis de la idoneidad didáctica nos ha permitido extraer y clasificar ejemplos de conocimientos didácticos en las diferentes facetas. Estos ejemplos se han tomado de las respuestas clasificadas a nivel 4 y 5, donde los estudiantes aplican correctamente y consistentemente con el proyecto los descriptores propuestos.

Los resultados se han presentado en forma de tablas al realizar una síntesis global de los resultados en cada uno de los tipos de idoneidad. El conjunto de estas tablas es un primer paso en la construcción de un modelo de conocimientos del profesor para enseñar estadística por medio de proyectos, y más concretamente, para enseñar el tema particular de correlación y regresión. Este modelo puede servir de guía para la organización de actividades formativas de los profesores, que tengan en cuenta la adquisición de todos estos conocimientos.

Hipótesis del Estudio 3

Aunque el Estudio 3 se realiza sobre una muestra pequeña de futuros profesores, y debido a su novedad, tiene carácter exploratorio, al igual que se hizo en el Estudio 2, se expusieron unas hipótesis iniciales que ahora discutimos.

H1. Se espera que los futuros profesores de Educación Secundaria y Bachillerato sean capaces de analizar la idoneidad didáctica del proceso de estudio de la correlación y regresión experimentado en el Estudio 2, en sus diferentes componentes.

Esta hipótesis queda confirmada por los resultados detallados del estudio. Como hemos visto, las respuestas de los futuros profesores se clasificaron en cinco niveles, estando la mayoría a nivel 3 o superior; muchas de ellas en niveles 4 y 5. Los resultados fueron mucho mejores que los obtenidos con futuros profesores de primaria en el análisis de otros proyectos más sencillos (que sólo involucran tablas, gráficos y medidas de posición y dispersión) en el estudio de Arteaga (2011).

Este resultado es razonable, puesto que dicho conocimiento didáctico depende fuertemente del conocimiento del contenido según Hill, Ball y Schilling (2008), y, dado que los resultados del análisis del conocimiento matemático sobre correlación y regresión en los profesores participantes fueron buenos, se justifican los mismos.

Una consecuencia de la confirmación de esta hipótesis es la utilidad de la actividad, así como de la Guía de análisis de la idoneidad didáctica para la formación de los profesores. Esta guía les resulta asequible, son capaces de argumentar su valoración de los diferentes descriptores y componentes, con lo cual, se introducen en un proceso de reflexión sobre la práctica docente, que al mismo tiempo desarrolla sus propios conocimientos.

H2. Se espera que el análisis cualitativo de las respuestas en la tarea de análisis de la idoneidad didáctica permita identificar ejemplos de conocimientos didácticos específicos sobre la correlación y regresión y sobre estadística que puedan utilizarse para enriquecer los modelos disponibles de conocimiento didáctico-matemático para la enseñanza del tema.

Igualmente se confirmó esta hipótesis, pues a lo largo del capítulo hemos indicado ejemplos de conocimientos de los futuros profesores que completan los descritos en trabajos previos, por ejemplo, los recogidos en Estepa y Gea (2012) y Gea (2012). Como hemos visto, los argumentos de los participantes en el estudio han sido muy variados, no sólo al comparar unos componentes con otros, sino incluso dentro de la misma pregunta.

Así, por ejemplo, al valorar la adecuación de los problemas planteados (idoneidad epistémica), se hace referencia al hecho de que con una sola situación se pueda introducir el estudio completo de la correlación y regresión, o que la intensidad de la correlación se pueda visualizar a partir de la dispersión de la nube, o el uso de aplicaciones interactivas en la resolución de los problemas. Al valorar el lenguaje, se cita no sólo los utilizados al plantear los problemas (gráficos o estadísticos en los enunciados), sino a los que obtendrán los estudiantes en la solución; por ejemplo, se citan las rectas o funciones de regresión.

Los futuros profesores son capaces de sugerir nuevas aplicaciones (biología, medio ambiente, economía) o nuevas preguntas de evaluación (preguntas dirigidas a valorar la comprensión conceptual, la competencia de cálculo). Identifican las situaciones de interpretación o representación, por ejemplo, de los diagramas de dispersión o de los parámetros de la función de ajuste. Reconocen posibles conflictos de aprendizaje y sugieren formas de ayudar a solventarlos. Igualmente, identifican las características de innovación docente del proyecto, de acuerdo con las directrices curriculares, que hemos resaltado en Batanero, Arteaga, y Gea, (2011).

Todos estos tipos de conocimiento, reconocidos en las respuestas de los participantes, se han clasificado en tablas para cada uno de los diferentes tipos de idoneidad, y constituyen una guía para la evaluación y/o desarrollo del conocimiento didáctico para la enseñanza de la correlación y regresión, apoyada en el uso de la tecnología.

CAPÍTULO 7.
CONCLUSIONES

7.1. INTRODUCCIÓN

Para finalizar la Memoria, se expone en este capítulo un resumen de las principales conclusiones obtenidas en nuestra investigación. Puesto que cada uno de los estudios que la componen tiene objetivos e hipótesis específicos, ya en los Capítulos 4, 5 y 6 se discutieron con detalle las conclusiones respecto a dichos objetivos e hipótesis, por lo que ahora no se repiten.

En lo que sigue analizamos, en primer lugar, las conclusiones obtenidas respecto a los dos objetivos generales planteados para la investigación, descritos en el Capítulo 1; seguidamente se resaltan las principales aportaciones a la investigación y la enseñanza del tema; y finalmente se proponen algunas líneas de investigación futuras, que podrían complementar nuestro trabajo.

7.2. CONCLUSIONES SOBRE LOS OBJETIVOS GENERALES

En el Capítulo 1 se describen los dos objetivos generales que dan origen a nuestra investigación, y a continuación aportamos, para cada uno de ellos, las conclusiones que se han obtenido.

O1. Realizar un análisis detallado de la presentación de la correlación y regresión en una muestra de libros de texto de Bachillerato, con la finalidad de caracterizar el significado institucional pretendido en las dos modalidades en que se contempla su enseñanza.

Creemos que este objetivo se ha cumplido razonablemente mediante el Estudio 1, en el cual se analizaron 16 libros de texto de Bachillerato, todos ellos vigentes en este momento, publicados por editoriales de amplia difusión, y de las dos modalidades de Bachillerato que incluyen el estudio de la correlación y regresión.

La base del análisis fue el análisis epistémico del significado de la correlación y regresión, desde el punto de vista descriptivo, presentado en el Capítulo 1, complementado por un estudio histórico, algunos de cuyos resultados se publicaron en Estepa y Gea (2010) y Estepa, Gea, Cañadas y Contreras (2012). Como consecuencia, se obtuvo una primera versión del significado institucional de referencia de la correlación y regresión para nuestro trabajo. Igualmente, se analizan en dicho capítulo las directrices curriculares españolas y andaluzas, así como las americanas, que tomamos de referencia.

El análisis realizado en el Estudio 1 permitió caracterizar el significado institucional pretendido para el tema en el actual Bachillerato, que luego se utilizó para asegurar una idoneidad epistémica razonable en las actividades del proyecto implementado, que sirvieron de instrumentos de evaluación y desarrollo del conocimiento didáctico-matemático de los futuros docentes en los Estudios 2 y 3. Ello fue posible mediante la categorización y análisis de los diferentes tipos de objetos primarios considerados en nuestro marco teórico, que en otras investigaciones previas no se distinguen (como por ejemplo, Sánchez Cobo (1999), quien no distingue entre concepto y propiedad), y la determinación de su distribución en los textos, en función de diferentes variables didácticas, algunas de las cuáles han sido aportaciones de la investigación realizada, (por ejemplo, analizar el momento en que se define un concepto y el uso del mismo en el tema).

El estudio se complementó con un análisis de los CDs que acompañan a algunos textos, que permite apreciar el énfasis desigual dado a diferentes tipos de tecnología en el estudio de la correlación y regresión en dichos textos. También se observa la falta de actividades con datos reales y proyectos, que permiten aprovechar estas herramientas tecnológicas, al tiempo que se desarrollan, no sólo las competencias de cálculo o la comprensión conceptual, sino el razonamiento estadístico de los estudiantes.

Finalmente, se identifican y categorizan, en relación a los objetos matemáticos primarios, los conflictos semióticos encontrados en los textos analizados. Estos conflictos incluyen representaciones inadecuadas al tipo de datos, o ambiguas en su notación; conceptos no equivalentes que se equiparan en los textos; propiedades que se generalizan en forma excesiva; o argumentos poco precisos e incluso confusos.

Todos estos puntos son comparados entre textos de la misma editorial, entre editoriales de la misma especialidad de Bachillerato y entre los dos Bachilleratos. Se observan y describen diferencias importantes, no tanto en cuanto a la modalidad de Bachillerato, sino entre diferentes editoriales, que indican la responsabilidad del profesor en la selección y el uso de los libros en el aula.

O2. Realizar un estudio exploratorio de evaluación del conocimiento didáctico-matemático sobre correlación y regresión, según el modelo propuesto en nuestro marco teórico (Godino, 2009; 2011), reinterpretando algunos de sus componentes y facetas en términos del modelo MKT (*Mathematical Knowledge for Teaching*) de Hill, Ball, y Schilling (2008).

Para cumplir este segundo objetivo se organizaron los Estudios 2 y 3, en los cuales, siguiendo la metodología de investigación basada en diseño (DBRC, 2003; Reiman, 2011), se construyeron, experimentaron y evaluaron actividades formativas, que también permitieron acercarse a los conocimientos de los estudiantes que se preparan como profesores de matemáticas de Secundaria y Bachillerato.

En el Estudio 2 (Capítulo 5), se diseña y experimenta la primera de estas actividades. Elegimos la forma de proyecto estadístico (Anexo 1), por ser una metodología actualmente sugerida para la enseñanza de la estadística (Batanero y Díaz, 2004; 2011; MacGillivray y Pereira Mendoza, 2011), y por realizarse la experimentación en el contexto de una asignatura de innovación docente. En él, los futuros profesores analizan diversos indicadores sociales relacionados con la Esperanza de vida, utilizando datos reales del servidor de las Naciones Unidas. Esta actividad se experimenta en dos cursos sucesivos, con muestras de 23 y 42 futuros profesores respectivamente.

Los resultados permiten evaluar los conocimientos matemáticos comunes y avanzados (Godino, 2009; 2011) sobre la interpretación de diagramas de dispersión, la estimación de la correlación, identificación del signo, asignación de un coeficiente de correlación entre varios dados, explicación de la correlación en relación con algunos de los tipos de covariación propuestos por Barbancho (1973), la identificación de un modelo de ajuste a los datos, y la ordenación de variables respecto a su poder de predicción sobre una variable dependiente. Igualmente, se evaluó el conocimiento sobre conceptos estadísticos previos en las dos muestras de estudiantes.

Los resultados revelan un buen conocimiento sobre la correlación y regresión de todos los puntos citados, y algo menor sobre la interpretación de gráficos estadísticos univariantes (diagrama acumulativo de frecuencias, histograma y gráfico de caja) y promedios, donde una pequeña proporción de participantes muestran errores descritos en las investigaciones previas, y por tanto, habría que reforzar su formación. Parte de estos resultados se presentarán en Batanero, Gea, Díaz y Cañadas (2014).

En el segundo año de la experimentación, un pequeño grupo de estudiantes llevan a cabo, de forma voluntaria, actividades de ampliación; que implican el trabajo con la hoja Excel, para ajustar un modelo de regresión; selección y análisis de nuevos datos, a partir del servidor de las Naciones Unidas; visualización de datos multivariantes dinámicos, con herramientas disponibles en dicho servidor; así como análisis de applets disponibles para la enseñanza. Este pequeño grupo de futuros profesores, unido a los pocos participantes de la primera muestra que también las realizaron, nos permite acercamos al conocimiento avanzado de la correlación y regresión de estos futuros profesores, así como a la faceta mediacional de su conocimiento didáctico, que es razonablemente buena, ya que muestran algún desarrollo de dichos componentes de su conocimiento mediante las actividades sugeridas, algunas de las cuales se muestran en el Capítulo 5.

En el Estudio 3, descrito en el Capítulo 6, los futuros profesores analizan la idoneidad didáctica del proyecto realizado por ellos mismos en sesiones anteriores, llevado a cabo con la primera de las muestras participantes en el Estudio 2. Del análisis de las respuestas escritas a la pauta de análisis (Anexo 2), se describen algunos componentes de las diferentes facetas del conocimiento didáctico de los profesores sobre la correlación y regresión en el modelo de Godino (2009; 2011).

El análisis detallado de las respuestas escritas de los futuros profesores en la actividad, indica unas competencias didácticas razonables de análisis de la enseñanza y aprendizaje de la correlación y regresión, en función de cada una de las facetas analizadas, puesto que en todos los descriptores de los diferentes componentes de la idoneidad didáctica, el nivel medio de la muestra de estudiantes se sitúa por encima del nivel medio teórico de la pauta de análisis utilizada. Además, este nivel es muy superior al obtenido en una actividad similar por Arteaga (2011), con futuros profesores de educación primaria. Resultados especialmente buenos se encuentran en la valoración de la idoneidad epistémica, mediacional y ecológica.

Por tanto, los futuros profesores son capaces de reconocer los objetos matemáticos implícitos en el trabajo con el proyecto planteado, y valorar su representatividad, respecto a los contenidos sobre correlación y regresión propuestos en las directrices curriculares. Igualmente, valoran y encuentran aplicaciones de los instrumentos tecnológicos utilizados, y relacionan el proyecto con los contenidos curriculares de estadística y otras áreas de matemáticas, así como con otras materias, y la sociedad en que vive el alumno.

7.3. PRINCIPALES APORTACIONES DEL TRABAJO

Pensamos que con nuestra investigación hemos realizado diferentes contribuciones, tanto a la investigación sobre educación estadística y formación de profesores, como para la enseñanza y innovación curricular. En esta sección sintetizamos estas aportaciones.

Contribución a la investigación en educación estadística

Nuestro trabajo aporta nuevo conocimiento en esta línea de investigación pues, como se ha argumentado a lo largo de la memoria, son muy escasas, incluso a nivel internacional, las investigaciones centradas específicamente sobre la correlación y regresión, recogidas principalmente en Estepa (2004), Estepa y Gea (2011; 2012) y Engel y Sedlmeier (2011). Más aún, la mayoría de las existentes se han enfocado desde el campo de la psicología, o bien, si se realizan desde la didáctica, se han preocupado por describir los errores o dificultades de los estudiantes.

Nuestro trabajo sobre los libros de texto (Estudio 1) es novedoso, pues el único realizado en España (Sánchez Cobo, 1999) analizó libros de planes anteriores, y llevó a cabo un estudio mucho menos extenso y profundo que el nuestro. Sobre este punto aportamos, en primer lugar, la descripción detallada del significado implementado en estos libros, con su caracterización de los diferentes tipos de objetos matemáticos tratados y las variables que influyen en los mismos. Además, el análisis epistémico del significado de la correlación y regresión, desde el punto de vista descriptivo, y de las directrices curriculares, desarrollada en el Capítulo 1, ha permitido realizar una síntesis del significado de referencia de la correlación y regresión en el nivel de Bachillerato.

Otro punto original es el análisis del uso de la tecnología (o sugerencias de uso) en los textos, que nos permite dar criterios para mejorarlos y para el trabajo con los mismos en el aula.

Contribución a la investigación sobre formación de profesores

Igualmente, abordamos un tema novedoso en el campo de la formación de profesores, donde, como se ha descrito en el Capítulo 2, apenas se investiga el tema de estadística, y menos aún la correlación y regresión. En nuestro trabajo, abordamos el conocimiento didáctico matemático sobre la correlación y regresión, de una forma completa en los Estudios 2 y 3, utilizando un modelo del conocimiento del profesor (Godino, 2009, 2011) igualmente poco tratado en este campo.

Una primera contribución en esta línea, es la descripción detallada de los conocimientos matemáticos (común y ampliado) sobre contenidos previos, y los relacionados con la correlación y regresión, en las muestras de profesores participantes en el Estudio 2. Abordamos en este estudio los principales objetos matemáticos considerados en el significado implementado en los libros de texto, así como conocimiento matemático ampliado sobre tecnología, y la relación entre correlación y causalidad.

Igualmente, con el Estudio 3 se proporciona descripción detallada de las diferentes facetas del conocimiento didáctico del profesor, en el modelo propuesto por Godino (2009, 2011), pocas de las cuáles han sido trabajadas en la investigación previa. Además del estudio cuantitativo, que proporciona un instrumento de evaluación de dichos conocimientos, las respuestas cualitativas han sido utilizadas para proponer ejemplos de conocimientos específicos de cada una de dichas facetas. Su conjunto es un paso previo para construir un modelo del conocimiento didáctico del profesor para la enseñanza de la correlación y regresión.

Implicaciones para la enseñanza

Los estudios anteriores proporcionan también materiales utilizables para la docencia.

En primer lugar, los ejemplos mostrados a lo largo del análisis de los textos pueden utilizarse en la construcción de pruebas de evaluación, o para alertar a los profesores sobre un uso crítico de los textos en el aula. Así como proporcionarles criterios pare elegir dichos libros de texto.

En los Estudios 2 y 3 mostramos también, junto a los resultados de la evaluación, el desarrollo de los conocimientos de los profesores. En consecuencia, las actividades diseñadas podrían ser útiles para la formación de otros futuros profesores sobre la correlación y regresión. Igualmente, el proyecto utilizado con los profesores en el Estudio 2 puede ser útil para la enseñanza de la correlación y regresión con estudiantes de Bachillerato, o de primeros cursos de Universidad.

Los recursos identificados en el Estudio 1, y su análisis, también pueden orientar la labor del profesor. Como se señala en las orientaciones curriculares, los recursos tecnológicos permiten evitar tediosos cálculos, que poco o nada aportan al tratamiento de la información. Por ello es que el profesor debiera estar atento a ello, y potenciar el uso de la Hoja de Cálculo Excel, ya que es una herramienta muy completa para el tratamiento de la correlación y regresión. Igualmente debiera apoyarse en datos reales y recursos interactivos (applets) y de visualización como los analizados en este trabajo.

Otras aportaciones

A lo largo del trabajo hemos citado algunas publicaciones relacionadas, que serían otras aportaciones realizadas, junto con otras no citadas en la Memoria.

Por otro lado, los estados de la cuestión desarrollados, tanto sobre la investigación previa en correlación y regresión, como sobre la formación de profesores para enseñar estadística, mejoran los realizados por los investigadores que nos precedieron y pueden ser una base para la investigación futura.

Añadimos a estas aportaciones el análisis matemático e histórico del tema, y el conjunto de referencias recopiladas, que pueden apoyar otras investigaciones sobre la correlación y la regresión.

7.4. LÍNEAS DE INVESTIGACIÓN FUTURAS

En esta Memoria se han trabajado con muestras limitadas de libros de texto y de futuros profesores, lo que, por un lado, indica algunas limitaciones del trabajo, y por otro, señala líneas de continuación de investigación futuras.

Respecto a los libros de texto, además de analizar los objetos matemáticos, como se ha llevado a cabo en nuestro trabajo, se podría completar el estudio con el análisis de los procesos matemáticos que se desarrollan en los libros de texto para la enseñanza de la correlación y regresión; o de las facetas duales del los objetos matemáticos incluidos en el tema. Este es un punto que pensamos completar del análisis realizado, que dará información de las diferentes configuraciones aportadas por los textos en la enseñanza del tema.

Igualmente, se podría completar el estudio con textos de nivel universitario. Por otra parte, el estudio de unidades didácticas o de recursos interactivos, disponibles en Internet, sería una nueva línea de investigación respecto al Estudio 1.

En la línea de formación de profesores, es claro que una continuación directa es ampliar los tamaños de muestra utilizados, y comprobar la estabilidad de los hallazgos con nuevas muestras de estudiantes de la Universidad de Granada, o incluso ampliar el contexto geográfico para ver hasta qué punto los resultados son locales. El diseño de otras actividades formativas, tanto sobre la parte matemática, como sobre la didáctica, serían nuevas líneas de investigación.

REFERENCIAS

Agnelli, H., Konic, P., Peparelli, N. Z. y Flores, P. (2009). La función lineal obstáculo didáctico para la enseñanza de la regresión lineal. *Revista Iberoamericana de Educación Matemática, 17*, 52 - 61.

Alloy, L. B. y Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91* (1), 112-149.

Alvarado, H. (2007). *Significados del teorema central del límite en la enseñanza de la estadística en ingeniería.* Tesis doctoral. Universidad de Granada.

American Psychological Association, American Educational Research Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Autor

Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores.* Tesis Doctoral. Universidad de Granada.

Azcárate, P. (1995). *El conocimiento profesional de los profesores sobre las nociones de aleatoriedad y probabilidad. Su estudio en el caso de la educación primaria.* Tesis doctoral inédita. Universidad de Cádiz, Cádiz.

Ball, D. L., Thames, M. H., y Phelps, G. (2005). Articulating domains of mathematical knowledge for teaching. Trabajo presentado en la American Education Research Association Conference. Disponible en: wwwpersonal. umich.edu/~dball/.

Ball, D. L., Thames, M. H. y Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education, 59* (5), 389-407.

Ball, D. L., Lubienski, S. T., y Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. En V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). Washington, DC: American Educational Research Association.

Barbancho, A. G. (1973). *Estadística elemental moderna*. Barcelona: Ariel.

Batanero, C. (2000). Significado y comprensión de las medidas de tendencia central. *UNO, 25*, 41-58.

Batanero, C. (2001). *Didáctica de la estadística*. Granada: Departamento de Didáctica de la Matemática.

Batanero, C. (2002). Los retos de la cultura estadística. Trabajo presentado en las *Jornadas Interamericanas de Enseñanza de la Estadística.* Buenos Aires, Octubre, 2002. Disponible en: www.ugr.es/~batanero/.

Batanero, C., Arteaga, P. y Gea, M. (2011). El currículo de estadística: Reflexiones desde una perspectiva internacional. *UNO, 59,* 9-17.

Batanero, C., Burrill, G. y Reading, C. (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study*. New York: Springer.

Batanero, C., Burrill, G., Reading, C. y Rossman, A. (Eds.) (2008). *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and IASE 2008 Round Table Conference.*

Monterrey, Mexico: ICMI e IASE. Disponible en http://iase-web.org/Publications.php.

Batanero, C. y Díaz, C. (2004). El papel de los proyectos en la enseñanza y aprendizaje de la estadística. En J. Patricio Royo (Ed.), *Aspectos didácticos de las matemáticas* (pp. 125-163). Zaragoza: ICE.

Batanero, C. y Díaz, C. (2008). *Análisis de datos con Statgraphics*. Granada: Departamento de Didáctica de la Matemática.

Batanero, C. y Díaz, C. (2011), *Estadística con proyectos*. Granada: Departamento de Didáctica de la Matemática.

Batanero, C. y Díaz, C. (2012). Training teachers to teach probability: Reflections and challenges. *Chilean Journal of Statistics 3*(1), 3-13.

Batanero, C., Díaz, C. y Gea, M. M. (2011). Estadísticas de la pobreza y desigualdad. En C. Batanero y C. Díaz (Eds.), *Estadística con proyectos*. (pp. 73-96). Granada: Departamento de Didáctica de la Matemática.

Batanero, C. Díaz, C. y Gea, M. (2013). Las estadísticas de pobreza y desigualdad. En E. Sánchez (Ed.), *Elementos de estadística y su didáctica a nivel bachillerato* (pp. 69-98). México: Secretaría de Educación Pública.

Batanero, C., Estepa, A. y Godino, J. (1991). Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria. *Suma*, *9*, 25-31.

Batanero, C., Estepa, A. y Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. En J. B. Garfield y G. Burrill, (Eds.), *Research on the role of technology in teaching and learning statistics. IASE Round Table Conference papers* (pp. 191-205). Voorburg, The Netherlands: Internacional Statistical Institute.

Batanero, C., Estepa, A., Godino, J. D. y Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, *27*, 151-169.

Batanero, C., Garfield, J. B., Ottaviani, M. G. y Truran, J. (2000). Research in statistics education. Some priority questions. *Statistics Education Research Newsletter, 2*(2). Disponible en: http://iase-web.org/Publications.php.

Batanero, C., Gea M. M., Díaz, C. y Cañadas, G. R. (2014). Building high school pre-service teachers knowledge to teach correlation and regression. *9th International Conference on Teaching Statistics "Sustainability in statistics education" (ICOTS 9)* (trabajo invitado). Flagstaff, AR: IASE., Julio, 2014. *Ponencia invitada*..

Batanero, C. y Godino, J. D. (1998). Understanding graphical and numerical representations of statistical association in a computer environment. En L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee y W. Wong, (Eds). *Proceedings of the Fifth International Conference on Teaching Statistics*, (Vol. *2*, pp. 1017-1024). Voorburg, The Netherlands: International Statistical Institute.

Batanero, C. y Godino, J. (2005). Perspectivas de la educación estadística como área de investigación. En R. Luengo (Ed.), *Líneas de investigación en didáctica de las matemáticas* (pp. 203-226). Badajoz: Universidad de Extremadura.

Batanero, C., Godino, J. D., y Estepa, A. (1998). Building the meaning of statistical association through data analysis activities (Research Forum). En A. Olivier y K. Newstead, (Eds.), *Proceedings of the 22nd Conference of the Internacional Group for the Psychology of Mathematics Education*, (pp. 221-236). Stellembosch, South Africa: Universidad de Stellenbosh.

Batanero, C., Godino, J. D., y Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education*, *12*. Disponible en: http://www.amstat.org/publications/jse/.

Batanero, C., Serrano, L. y Green, D. R. (1998). Randomness, its meanings and implications for teaching probability. *International Journal of Mathematics Education in Science and Technology*, 29 (1), 113-123.

Batanero, C., Godino, J. y Navas, F. (1997). Some misconceptions about averages in prospective primary teachers. En E. Pehknonen (Ed.), Proceedings of 21 PME Conference (v.1, pp. 276). University of Lahti.

Benzecri, J. P. (1982). *Histoire et prehistoire de l'analyse des données*. París: Bordás.

Beyth-Marrom, R. (1982). Perception of correlation reexamined. *Memory and cognition, 10 (6),* 511-519.

Bisquerra, R. (1989). *Métodos de investigación educativa.* Barcelona: P.P.U.

Brown, C. y Borko, H. (1992). Becoming a mathematics teacher. En D. A. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 209-239). New York: MacMillan.

Bruno, A. y Espinel, M. C. (2005). Recta numérica, escalas y gráficas estadísticas: un estudio con estudiantes para profesores. *Formación del Profesorado e Investigación en Educación Matemáticas 7,* 57-85.

Burgess, T. A. (2008). Teacher knowledge for teaching statistics through investigations. En C. Batanero, G. Burrill, C. Reading y A. Rossman (2008). *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.* Monterrey: ICMI e IASE. Disponible en: http://iase-web.org/Publications.php.

Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 259-270). Springer Netherlands.

Burrill, G. y Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in Training teachers. En C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education - A joint ICMI/IASE Study* (pp. 57-69). Dordrecht: Springer.

Cabriá, S. (1994). *Filosofía de la estadística.* Valencia: Servicio de Publicaciones de la Universidad.

Cai, J. y Gorowara, C. C. (2002). Teachers' conception and constructions of pedagogical representations in teaching arithmetic average. *Proceedings of the Sixth International Conference on Teaching of Statistics, Cape Town.* Voorburg, The Netherlands: International Statistical Institute. Disponible en: http://iase-web.org/Publications.php

Callingham, R., y Watson, J. (2011). Measuring levels of statistical pedagogical content knowledge. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 283-293). Springer Netherlands.

Canada, D. L. (2008). Conceptions of distribution held by middle school students and preservice teachers. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: ICMI e IASE. Disponible en: http://iase-web.org/Publications.php.

Cañadas, G. (2010). Las tablas de contingencia en la formación de profesionales de psicología. Tesis de Máster. Universidad de Granada.

Cañadas, G. (2012). *Comprensión intuitiva y aprendizaje formal de las tablas de contingencia en alumnos de psicología.* Tesis doctoral. Universidad de Granada.

Cañadas, G., Batanero, C., Contreras, J. M. y Arteaga, P. (2011). Estrategias en el estudio de la asociación en tablas de contingencia por estudiantes de psicología. *Educación Matemática, 23* (2), 5-32.

Cardeñoso, J. M. (1998). *Las creencias y conocimientos de los profesores de primaria andaluces sobre la matemática escolar. Modelización de las concepciones sobre aleatoriedad y probabilidad*. Tesis doctoral. Universidad de Cádiz.

Casey, S.A. (2008). *Subject matter knowledge for teaching statistical association*.Tesis doctoral. Universidad de Illinois.

Casey, S.A. (2010). Subject matter knowledge for teaching statistical association. *Statistics Education Research Journal 9*(2), 50-68. Disponible en http://iase-web.org/Publications.php.

Castro-Sotos, A. E., Vanhoof, S., Van Den Noortgate, W. y Onghena, P. (2009). The transitivity misconception of Pearson’s correlation coefficient. *Statistics Education Research Journal 8*(2), pp. 33-55. Disponible en: http://iase-web.org/Publications.php.

Chapman, L. J. y Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, *72* (3), 193-204.

Chapman, L. J. y Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74* (3), 271-280.

Chevallard, Y. (1991). *La transposición didáctica. Del saber sabio al saber enseñado*. Buenos Aires: Aique.

Chick, H. L. y Pierce, R. U. (2008). Teaching statistics at the primary school level: beliefs, affordances, and pedagogical content knowledge. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: ICMI e IASE. Disponible en: http://iase-web.org/Publications.php.

Cobb, P. y Gravemeijer, K. (2008). Experimenting to support and understand learning processes. Én A.E. Kelly, R.A. Lesh y J. Y. Baek (Eds.), *Handbook of design research methods in education. Innovations in science, technology,engineering and mathematics learning and teaching* (pp. 68-95). Mahwah, NJ: Erlbaum.

Cobo, B. (2003). *Significado de las medidas de posición central para los estudiantes de secundaria.* Tesis doctoral. Universidad de Granada.

Cobo, B. y Batanero, C. (2004). Significados de la media en los libros de texto de secundaria. *Enseñanza de las Ciencias, 22* (1), 5-18.

Connor, D., Davies, N. y Payne, B. (2002). Web-based project and key skill work. *Teaching Statistics*, *24*(2), 62-65.

Common Core Standards Initiative. (2010). *Standards for mathematical practice.* Online: http://www.corestandards.org/

Consejería de Educación. (2007). *Orden de 10 de agosto de 2007, por la que se desarrolla el currículo correspondiente a la Educación Secundaria Obligatoria en Andalucía.* Sevilla: Autor.

Consejería de Educación. (2008a). *Orden de 5 de agosto de 2008, por la que se desarrolla el Currículo correspondiente a la Educación Infantil en Andalucía.* Sevilla: Autor.

Consejería de Educación. (2008b). *Orden de 5 de agosto de 2008, por la que se desarrolla el currículo correspondiente al Bachillerato en Andalucía*. Sevilla: Autor.

Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional.* Tesis Doctoral. Universidad de Granada.

Cook, T. D. y Reichardt, C. S. (2000). *Métodos cualitativos y cuantitativos en investigación evaluativa.* Paideia.

Cooney, T. (1994). Research and teacher education: In search of common ground. *Journal for Research in Mathematics Education*, *25*(6), 608-636.

Cordero, F. y Flores, R. (2007). El uso de las gráficas en el discurso matemático escolar. Un estudio socioepistemológico en el nivel básico a través de los libros de texto. *Revista Latinoamericana de Matemática Educativa, 10* (1), 7-38.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin 90* (2), 272-292.

Curcio, F. R. (1989). *Developing graph comprehension.* Reston, VA: N.C.T.M.

D’Amore, B. y Godino, J. D. (2007). El enfoque ontosemiótico como un desarrollo de la teoría antropológica en didáctica de la matemática. *Revista Latinoamericana de Investigación en Matemática Educativa*, *10* (2) 191-218.

D'Amore, B., Font, V. y Godino, J. D. (2007). La dimensión metadidáctica en los procesos de enseñanza y aprendizaje de la matemática. *Paradigma, XXVIII* (2), 49-77.

DBRC (The Design Based Research Collective) (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Duval, R. (1993). *Semiosis et noesis. Lecturas en Didáctica de la Matemática: Escuela Francesa.* México: Sección de Matemática Educativa del CINVESTAV-IPN.

Engel, J. y Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp, 247-258). New York: Springer.

Estepa, A. (1994). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral. Universidad de Granada.

Estepa, A. (2004). Investigación en educación estadística. La asociación estadística. En R. Luengo (Ed.). *Líneas de investigación en Educación Matemática,* (pp. 227-255). Badajoz: Servicio de Publicaciones. Universidad de Extremadura.

Estepa, A. (2008). Interpretación de los diagramas de dispersión por estudiantes de Bachillerato. *Enseñanza de las Ciencias, 26* (2), 257-270.

Estepa, A. y Batanero, C. (1995). Concepciones iniciales sobre la asociación estadística. *Enseñanza de las Ciencias, 13* (2), 155-170.

Estepa, A. y Batanero, C. (1996). Judgments of correlation in scatter plots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education, 4,* 25-41.

Estepa, A. y Gea, M. M. (2010). El origen de la noción de correlación y la enseñanza. En J. Berral, M. de la Fuente y España, F. (Eds.) *Actas al XIII Congreso de Enseñanza y Aprendizaje de las Matemáticas,* (pp. 202-212). Córdoba: Sociedad Andaluza de Educación Matemática Thales

Estepa, A. y Gea, M. M. (2011). La enseñanza-aprendizaje de la asociación estadística. *Actas de las 15 Jornadas sobre el Aprendizaje y la Enseñanza de las Matemáticas*. Gijón: Federación Española de Sociedades de Profesores de Matemáticas.

Estepa, A. y Gea, M. M. (2012). Conocimiento para la enseñanza de la asociación estadística. En J. J. Ortiz (Ed.), *Investigaciones actuales en educación estadística y formación de profesores* (pp. 23-40). Melilla: Universidad de Granada

Estepa, A., Gea, M. M., Cañadas, G. R. y Contreras, J. M. (2012). Algunas notas históricas sobre la correlación y regresión y su uso en el aula. *Números, 81*, 5-14.

Estepa, A. y Sánchez Cobo, F. T. (1994). Desarrollo histórico de la idea de asociación estadística. *Épsilon, 30,* 61-74.

Estepa, A. y Sánchez Cobo, F. T. (1998). Correlation and regresión in secondary school text books. En Pereira-Mendoza, L., Seu, L., Wee, T. y Wong, W. (Eds), *Proceedings of the Fifth Internacional Conference on Teaching of Statistics* (pp. 671-676). Voorburg, The Netherlands: Internacional Statistical Institute.

Estrada, M. A. (2002). *Análisis de las actitudes y conocimientos estadísticos elementales en la formación del profesorado*. Tesis doctoral. Universidad Autónoma de Barcelona.

Estrada, A., Batanero, C. y Fortuny, J. M. (2004). Un estudio de evaluación de conocimientos estadísticos en profesores en formación e implicaciones didácticas. *Educación Matemática, 16,* 89-112

Espinel, C. (2007). Construcción y razonamiento de gráficos estadísticos en la formación de profesores. *Investigación en Educación Matemática XI,* 99-119.

Even, R. (1990). Subject matter knowledge for teaching and the case of functions. *Educational Studies in Mathematics, 21*(6), 521-544.

Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics*

Education, *24*(2), 94-116.

Even. R. y Ball, D. (2009). *The professional education and development of teachers of mathematics. The 15th ICMI Study*. New York: Springer.

Falk, R. (1986). Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics.* (pp. 292 – 297). Victoria, Canada: International Statistical Institute.

Fischbein, E. (1975). *The intuitive sources of probabilitics thinking in children*. Dordrecht, Holland: Reidel.

Font, V., Godino, J. D. y D'Amore, B. (2007). An onto-semiotic approach to representations in mathematics education*. For the Learning of Mathematics*, *27* (2), 2-7.

Franklin, C. Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework.* Alexandria, VA: American Statistical Association. Disponible en: www.amstat.org/education/gaise/.

Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review 70*(1), 1-25.

Galton, F. (1869). Hereditary genius. Reproducido parcialmente en J. R Newman. (1956). *Sigma. El mundo de las matemáticas* (pp. 241-248) Barcelona: Grijalbo.

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45*, 135-145.

Garfield, J. B. y Ben-Zvi, D. (2008). Preparing school teachers to develop students' statistical reasoning. En C. Batanero, G. Burrill, C. Reading y A. Rossman (2008). (Eds.), *Proceedings of the Joint ICMI /IASE StudyTeaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey, Mexico: ICMI e IASE. Disponible en: http://iase-web.org/Publications.php.

Gea, M. M. (2012). Correlation and regression in the training of teachers. En L. Gaiser y D. Curcic (Eds.), *Bridging gaps in the Mediterranean research space: 4th EMUNI Research Souk. The Euro-Mediterranean Student Research Multi-conference* (pp.153-159). El. Knjiga. - Portoroz: EMUNI University.

Gea, M.. Batanero, C., Arteaga, P., Cañadas, G. y Contreras, J. M. (En prensa). Análisis del lenguaje sobre la correlación y regresión en libros de texto de bachillerato. *SUMA*

Gea, M. M., Batanero, C., Arteaga, P. y Cañadas, G. R. (2013). Justificaciones en el tema de correlación y regresión en textos españoles de Bachillerato. *EM TEIA - Revista de Educaçao Matemática e Tecnológica Iberoamericana,* 4(2), 1 - 19

Gea, M. M., Batanero, C. y Cañadas, G. R. (2013). Un estudio empírico de los problema de correlación y regresión en libros de texto de Bachillerato. En J. A. Fernandes, F. Viseu, M. H. Martinho y P. F. Correia (Eds.), *Atas do III Encontro de Probabilidades e Estatística na Escola* (pp. 71-81). Braga: Centro de Investigação em Educação (CIEd). Universidade do Minho.

Gea, M., Batanero, C., Cañadas, G. y Arteaga, P. (2013a). La organización de datos bidimensionales en libros de texto de Bachillerato. *I Jornadas Virtuales de Didáctica de la Estadística, Probabilidad y Combinatoria*. Granada: SEIEM.

Gea, M. M., Batanero, C., Cañadas, G. R. y Contreras, J. M. (2013b). Un estudio empírico de las situaciones-problema de correlación y regresión en libros de texto de bachillerato. En A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 293-300). Bilbao: Sociedad Española de Investigación en Educación Matemática (SEIEM).

Gea, M. M., Batanero, C., Cañadas, G. R. y Contreras, J. M. (En prensa). Los problemas de correlación y regresión en los libros de texto españoles de bachillerato. *Perspectivas da Educaçao Matematica.*

Gea, M., Batanero, C., Cañadas, G., Contreras, J. M. y Arteaga, P. (2013). La estimación de la correlación: variables de tarea y sesgos de razonamiento. En A. Salcedo (Ed.), *Educación estadística en América Latina: Tendencias y* perspectivas (pp. 361-384). Caracas;

Universidad Central de Venezuela.

Gea, M., Contreras, J. M., Cañadas, G. y Arteaga, P. (2012). Comprendiendo la correlación a partir de sus representaciones. Trabajo presentado en el *XIV Congreso sobre Enseñanza y Aprendizaje de las Matemáticas: "Diversidad y Matemáticas"*. Thales. Málaga. 2012.

Gea, M. M., Contreras, J. M., Arteaga, P. y Cañadas, G. R. (2012). El lenguaje sobre la correlación y regresión: un estudio de dos libros de texto. En H. Pinto, H. Jacinto, A. Henriques, A. Silvestre y C. Nunes (Eds.). *Atas do XXIII Seminário de Investigaçao Matemática* (pp. 415-428). Lisboa: Associaçao de Professores de Matemática.

Gea, M. M., Batanero, C., Fernandes, J. A. y Gómez, E. (En prensa). La distribución de datos bidimensionales en los libros de texto de matemáticas de Bachillerato. *Quadrante.*

Ghiglione, R. y Matalón, B. (1989). *Las encuestas sociológicas*. México: Trillas.

Gitirana, V., Guimarães, G., Magina, S. y Carzola, I. (2008). Construção de gráficos de coluna: abilidade de alunos e professores. Presentado en el Encuentro Latinoamericano de Educación Estadística (ELEE), Monterrey, México.

Goetz, J. P. y Lecompte, M. D. (1998). *Etnografía y diseño cualitativo en educación.* Morata: Madrid.

Godino, J. D. (1996) Mathematical concepts, their meaning, and understanding. En L. Puig y A. Gutierrez (Eds.), *Proceedings of XX Conference of the International Group for the Psychology of Mathematics Education* (v.2, pp. 417-425). Universidad de Valencia.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, *22* (2-3), 237-284.

Godino, J. D. (2009). Categorías de análisis de los conocimientos del profesor de matemáticas. *UNION 20*, 13-31.

Godino, J. D. (2011). Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas. Conferencia presentada en el *XIII CIAEM-IACME*, Recife, Brasil.

Godino, J. D. (2014). *Síntesis del enfoque ontosemiótico del conocimiento y la instrucción matemática: motivación, supuestos y herramientas teóricas.* Universidad de Granada. Disponible en, http://www.ugr.es/local/jgodino/eos/sintesis_EOS_14abril14.pdf

Godino, J. D. y Batanero, C (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, *14* (3), 325-355.

Godino, J. D. y Batanero, C. (2008). Formación de profesores de matemáticas basada en la reflexión guiada sobre la práctica. Conferencia presentada en el VI Congreso Iberoamericano de Educación Matemática, Puerto Montt. Chile.

Godino, J. D. y Font, V. (2007). Algunos desarrollos de la Teoría de los significados sistémicos. Anexo al artículo : Significado institucional y personal de los objetos matemáticos. En Godino, J. D. y Batanero, C (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, *14* (3), 325-355.

Godino, J. D. y Batanero, C. (2009). Formación de profesores de matemáticas basada en la reflexión guiada sobre la práctica. Conferencia en el *VI CIBEM*, Puerto Montt, Chile, Enero, 2009. Disponible en: www.ugr.es/~jgodino/.

Godino, J. D. Batanero, C., Contreras, A., Estepa, A., Lacasta, E. y Wilhelmi, M. (2013). Didactic engineering as design-based research in mathematics education. Trabajo presentado en *CERME 7*, Antlya, Turquía.

Godino, J. D., Batanero**, C.** y Flores, P. (1999). El análisis didáctico del contenido matemático como recurso en la formación de profesores. En *Homenaje al profesor Oscar Sáenz Barrio* (pp. 165-185). Granada: Departamento de Didáctica y Organización Escolar.

Godino, J. D. Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. ZDM. *The International Journal on Mathematics Education*, *39* (1-2), 127-135.

Godino, J. D., Batanero, C., Roa, R. y Wilhelmi, M. R. (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Proceedings of the Joint ICMI /IASE StudyTeaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education*. Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education. Online: http://www.ugr.es/~icmi/iase_study/.

Godino, J. D., Bencomo, D., Font, V. y Wilhelmi, M. R. (2006) Análisis y valoración de la idoneidad didáctica de procesos de estudio de las matemáticas. *Paradigma, 27* (2), 221-252

Godino, J. D., Contreras, A. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathématiques*, *26* (1), 39-88.

Godino, J. D., Font, V., Contreras, A. y Wilhelmi, M. R. (2006). Una visión de la didáctica francesa desde el enfoque ontosemiótico de la cognición e instrucción matemática. *Revista Latinoamericana de Investigación en Matemática Educativa*, *9* (1), 117-150.

Godino, J. D., Font, V., Wilhelmi, M. R. y Castro, C. (2009). Aproximación a la dimensión normativa en Didáctica de las Matemáticas desde un enfoque ontosemiótico. *Enseñanza de las Ciencias*, *27* (1), 59-76.

Godino, J. D., Font, V. y Wilhelmi, M.R. (2008). Análisis didáctico de procesos de estudio matemático basado en el Enfoque Ontosemiótico. Publicaciones, 38, (pp. 25-49).

Gómez, E. (2014). *Evaluación y desarrollo del conocimiento matemático para enseñar la probabilidad en futuros profesores de educación primaria*. Tesis doctoral. Universidad de Granada

Gómez Chacón, I. M. (2000). *Matemática emocional. Los afectos en el aprendizaje matemático.* Madrid: Narcea.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. J*ournal for Research in Mathematics Education*, *38* (5), 427-437.

Groth, R. E. y Bergner, J. A. (2006). Preservice elementary teachers' conceptual an procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, *8*, 37-63.

Guzmán, M. (2003). Los goces estéticos del quehacer matemático. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales 97* (3), 351-457.

Hacking, I. (1990). *The taming of chance*. Cambridge, MA: Cambridge University Press.

Hald, A. (1998). *A history of mathematical statistics. From 1750 to 1930*. New York: John Wiley.

Herbel, B. A. (2007). From intended curriculum to written curriculum: Examining the" voice" of a mathematics textbook. *Journal for Research in Mathematics Education*, *38* (4), 344-369.

Hall, J. (2011). Engaging teachers and students with real data: Benefits and challenges. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 335-346). Springer Netherlands.

Hill, H. C., Ball, D. L. y Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39, 372-400.

Hill, H. C., Sleep, L., Lewis, J. M. y Ball, D. (2007). Assessing teachers' mathematical knowledge. En F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111-155). Greenwich, CT: Information Age Publishing, Inc. y NCTM.

Holmes, P. (2001). Correlation: From picture to formula. *Teaching Statistics*, *23* (3), 67-71.

Holmes, P. (1997). Assessing project work by external examiners. En I. Gal y J. B: Garfield (Eds.), The assesment challenge in statistics education (pp. 153-164). Voorburg: IOS Press.

Jacobbe, T. (2008). Elementary school teachers' understanding of the mean and median. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference. Monterrey

Mexico: ICMI e IASE. Disponible en: http://iase-web.org/Publications.php.

Jaworski, B. (1994). *Investigating mathematics teaching: A constructivist inquiry*. London: Falmer.

Johnson, R. y Kuby, P. (2004). *Estadística elemental*. México: Thompson.

Kieran, C., Krainer, K. y Shaughnessy, J. M. (2013). Linking research to practice: Teachers as key stakeholders in mathematics education research. En M.A. (Ken) Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, y F. K.S. Leung (Eds.). *Third international handbook of mathematics education* (pp. 361-392). Springer New York.

King, M. (2000). Scatter diagrams and the excel chart wizard. *Micromath, 16* (3), 31-34.

Kirk, J. y Miller, M. L. (1986). *Reliability and validity in cualitative research*. Newbury Park, CA: Sage University Paper.

Krippendorff, K. (1997). *Metodología de análisis de contenido. Teoría y práctica.* Barcelona: Paidos.

Lavalle, A. L., Micheli, E. B. y Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *Revista Latinoamericana de Investigación en Matemática Educativa*, *9* (3), 383-406.

Lee, H. S., & Hollebrands, K. F. (2011). Characterising and developing teachers' knowledge for teaching statistics with technology. . En C. Batanero, G. Burrill y C. Reading (Eds.), Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education (pp. 359-369). Springer Netherlands.

León, O. G. y Montero, I. (2002). *Métodos de Investigación*. Madrid: Mc Graw Hill.

Llinares, S. (1998). La investigación «sobre» el profesor de matemáticas: aprendizaje del profesor. *Aula*, *10,* 153-179.

Llinares, S. (2009). Competencias docentes del maestro en la docencia en matemáticas y el diseño de programas de formación. *UNO. Revista de Didáctica de las Matemáticas*, *51*, 92-101.

Llinares, S., Sánchez, V. y García, B. M. (1994). Conocimiento del contenido pedagógico del profesor. Tareas y modos de representación de las fracciones. *Revista de Educación, 304,* 199-225.

LLinares S. y Krainer K. (2006). Mathematics (student) teachers and teacher educators as learners. En A. Gutierrez y P. Boero (Eds) Handbook of Research on the Psychology of Mathematics Education. (pp. 429 – 459). Rotherdam / Taipei: Sense Publishers.

López Noguero, F. (2002). El análisis de contenido como método de investigación. *Revista de Educación*, *4,* 167-180.

Makar, K. y Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27-54. Disponible en http://iase-web.org/Publications.php.

Martínez, M. (2006). La investigación cualitativa (síntesis conceptual). *Revista de Investigación en Psicología* 9(1), 123-146.

Mayén, S. (2009). *Significados de las medidas de posición central para estudiantes mexicanos de Bachillerato.* Tesis doctoral. Universidad de Granada.

Mayén, S., Díaz, C. y Batanero, C. (2009). Conflictos semióticos de estudiantes con el concepto de mediana. *Statistics Education Research Journal 8(2),* 74-93**.** Disponible en http://iase-web.org/Publications.php.

MacGillivray, H. y Pereira Mendoza, L. (2011). Teaching statistical thinking through investigative projects. En C. Batanero, G. Burrill, y C. Reading, (Eds.) *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI and IASE study* (pp. 109-120). New York: Springer.

Martínez, M. (2006). La investigación cualitativa. *Revista de Investigación en Psicología, 9* (1). Disponible en: sisbib.unmsm.edu.pe/bvrevistas/investigacion_psicologia/

McKenzie, C. R. M., y Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment.

Cognitive Psychology, 54 (1), 33-61.

MEC (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas de la Educación primaria*. Madrid: Autor.

MEC (2007a). *Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria.* Madrid: Autor.

MEC (2007b). *Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas*. Madrid: Autor.

MEC (2007c), *ORDEN ECI/3858/2007, de 27 de diciembre, por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de las profesiones de Profesor de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanzas de Idiomas.* Madrid: Autor.

Mickelson, W. y Heaton, R. (2004). Primary teachers' statistical reasoning about data. In D. Ben-Zvi & J. Garfield (Eds.), The challenges of developing statistical literacy, reasoning, and thinking (pp. 327-352). Dordrecht, Netherlands: Kluwer.

Ministerio de Educación (2009). *PISA 2009. Programa para la evaluación internacional de alumnos de la OCDE. Informe español.* Madrid: Autor.

Mohamed, N. (2012). *Evaluación del conocimiento de los futuros profesores de educación primaria sobre probabilidad.* Tesis Doctoral. Universidad de Granada.

Monteiro, C. y Ainley, J. (2006). Student teachers interpreting media graphs. En A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil:* International Statistical Institute and International Association for Statistical Education. Online: http://www.stat.auckland.ac.nz/~iase.

Monteiro, C. y Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education 2* (3), 188-207. Disponible en: http://www.iejme/.

Moore, D. S. (1991). Teaching statistics as a respectable subject. En F. Gordon y S. Gordon (eds.), *Statistics for the Twenty-First Century* (pp. 14-25). Mathematical Association of America.

Moore, D. S. (1995). *Estadística aplicada básica.* Barcelona: Antoni Bosch

Moritz, J. (2004). Reasoning about covariation. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 221-255). Dordrecht: Kluwer.

Murray, S. y Gal, I. (2002). Preparing for diversity in statistics literacy: Institutional and educational implications. En B. Phillips (Ed.). *ICOTS-6 papers for school teachers*. [CD-ROM]. Cape Town: International Association for Statistics Education.

Naciones Unidas (2011). *Informe sobre desarrollo humano 2011. Sostenibilidad y equidad: Un mejor futuro para todos.* Ginebra: Autor. Disponible en: http://hdr.undp.org/es/informes/mundial/idh2011/

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: VA.

Newman, J. R (1956). *Sigma. El mundo de las matemáticas.* Barcelona: Grijalbo.

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education, 21,*509 -523.

Nolan, D., y Speed, T. P. (2002). Teaching statistics theory through applications. *American Statistician, 5*3, 370-375.

Olivo, E. (2008). *Significados de los intervalos de confianza para los estudiantes de ingeniería en México*. Tesis doctoral. Universidad de Granada.

Orden, la, A. (2007). El nuevo horizonte de la investigación pedagógica. *Revista Electrónica de Investigación Educativa 9* (1). Disponible en: medicina.iztacala.unam.mx/medicina/.

Ortega, P., Ramírez, M. E., Torres, J. L., López, A. E., Servín, Y., Suárez, L. y Ruiz, B. (2007).

Modelo de innovación educativa. Un marco para la formación y el desarrollo de una cultura de la innovación. *RIED: Revista Iberoamericana de Educación a Distancia* 10(1-2),145-173.

Ortiz, J. J. (1999). *Significado de los conceptos probabilísticos elementales en los textos de Bachillerato*. Tesis Doctoral. Universidad de Granada.

Ortiz, J. J., Batanero, C. y Serrano, L. (2001). El lenguaje probabilístico en los libros de texto. *Suma*, *38*, 5-14.

Orton, A. (1990). *Didáctica de las matemáticas.* Madrid: M.E.C. y Morata.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*. 13, 25-45.

Pearson, K. (1965). Somes incidents in the early history of biometry and statistics 1890-1894, *Biometrika*. 52, 3-18.

Peña, D. (2002). *Análisis de datos multivariables.* Madrid: McGraw-Hill.

Perez Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico.* Madrid: Ediciones de la Universidad Autónoma de Madrid.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, *5*(2), 27-45. Disponible en http://iase-web.org/Publications.php.

Pinto, J. (2010). *Conocimiento didáctico del contenido sobre la representación de datos estadísticos: Estudios de casos con profesores de estadística en carreras de psicología y educación*. Tesis Doctoral. Universidad de Salamanca.

Ponte, J. P. (2001). Investigating in mathematics and in learning to teach mathematics. En T. J. Cooney y F. L. Lin (Eds.), *Making sense of mathematics teacher education* (pp. 53-72). Dordrecht, Netherlands: Kluwer Academic Publishers.

Ponte, J. P. y Chapman, O. (2006). Mathematics teachers' knowledge and practices. In A. Gutierrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 461-494). Roterdham: Sense.

Quintas, S., Oliveira, H. y Freitas, R. (2013). O conhecimento didático de duas professoras no ensino da relação bivariada na estatística. Trabajo presentado en el *Seminário "a literacia e raciocínio estatísticos: desafios para o currículo e a aprendizagem"*. Instituto de Educação, Universidade de Lisboa

Recio, A. M. (1999). Una aproximación epistemológica a la enseñanza y aprendizaje de la demostración matemática. Tesis doctoral. Universidad de Granada,

Reimann, P. (2011). Design-based research. En L. Markauskaite et al. (Eds.), *Methodological choice and design* (pp. 37-56). Springer.

Reys, B. J., Reys, R. E. y Chavez, O. (2004). Why mathematics textbooks matter. *Educational Leadership, 61*(5), 61–66.

Ridgway, J., McCusker, S. y Nicholson, J. (2007, Agosto). Engaging citizens in reasoning with evidence. Trabajo presenado en la *56th Session of the Internacional Statistical Institute*, Lisboa, 2007.

Ridgway, J., Nicholson, J. y McCusker, S. (2008). Mapping new statistical Literacies and Iliteracies. Trabajo presentado en el *11th International Congress on Mathematics Educ*ation, Monterrey, Mexico.

Rittle-Johnson, B. y Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(1), 175-189

Rittle-Johnson, B., Siegler, S.. y Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process, *Journal of educational Psychology 93*(2), 343-362.

Rothery, A. (1980). *Children reading mathematics*. Worcester: College of Higher Education.

Ruiz, B. (2006). *Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria*. Tesis de Maestría. CICATA. México.

Sánchez, E. y Batanero, C. (2012). Manejo de la información. En E. Sánchez (Coord.), Enseñanza y aprendizaje de las matemáticas. Casos y perspectivas (pp. 64-92). México, D. F.: Secretaría de Educación Pública.

Sánchez Cobo, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral. Universidad de Granada.

Sánchez Cobo, F. T., Estepa, A. y Batanero, C. (2000). Un estudio experimental de la estimación de la correlación a partir de diferentes representaciones. *Enseñanza de las Ciencias*, *18* (2), 297-310.

Schleppegrell, M. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly*, *23*, 139-159.

Seal, H. L. (1967). The historical development of the Gauss linear model. *Biometrika*, 54, 1-24).

Shulman (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15* (2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: foundations of new reform. *Harvard Educational Review*, *57* (1), 1-22.

Sierpinska, A. (1994). *Understanding in mathematics*. London: Falmer Press.

Skemp, R. (1980). *Psicología del aprendizaje de las matemáticas.* Madrid: Morata

Silva y Coutinho, 2008

Sorto, M. A. (2004). *Prospective middle school teachers' knowledge about data analysis and its application to teaching*. Tesis doctoral. Universidad del Estado de Michigan.

Starkings, S. (1997). Assessing students' projects. En I. Gal y J. B. Garfield (Eds.), *The assesment challenge in statistics education* (pp. 139-152). Voorburg: IOS Press.

Stohl, H. (2005). Probability in teacher education and development. En G. Jones (Ed.). *Exploring probability in schools: Challenges for teaching and learning* (pp. 345-366). New York: Springer.

Tauber, L. (2001). *La construcción del significado de la distribución normal en un curso de análisis de datos*. Tesis doctoral. Universidad de Sevilla.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: a synthesis of the research. En D. A. Grouws (Ed.), *Handbook on Mathematics Teaching and Learning* (p. 127-146), Macmillan, New York.

Thompson, D. R. y Rubenstein, R. N. (2000). Learning mathematics vocabulary: Potential pitfalls and instructional strategies. *Mathematics Teacher*, *93*, 568-574.

Tukey, J. (1962). The future of data analysis. *Annals of Mathematical Statistics*, *33*, 1-67.

Tukey, J. W. (1977). *Exploratory data analysis*. New York: Addisson Wesley.

de Villiers, M. (1993). *El papel y la función de la demostración en matemáticas*. *Épsilon*, 26, 15-30.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88* (421), 1-8.

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of data and chance. *Journal of Mathematics Teacher Education*, 4, 305-337.

Watson, J.M. (2006). *Statistical literacy at school: Growth and goals.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wassong, T. y Biehler, R. (2010). A model for teacher knowledge as a basis for online courses for Professional development of statistics teachers En C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010)*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

Weber, R. P. (1985). *Basic content analysis.* Londres: Sage.

Wilhelmi, M. R., Font, V. y Godino, J. D. (2005). Bases empiriques de modèles théoriques en didactique des mathématiques: réflexions sur la théorie de situations didactiques et le point de vue ontologique et sémiotique. Trabajo presentado en el *Colloque International Didactiques: quelles references epistemologiques?*. Paris; Association Francophone Internationale de Recherche Scientifique en Education. Disponible en www.ugr.es/local/jgodino/.

White, A. L., Jaworski, B., Agudelo-Valderrama, C., y Gooya, Z. (2013). Teachers learning from teachers. En M.A. (Ken) Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, y F. K.S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 393-430). Springer New York.

Wild, C. J., y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67* (3), 223-265.

Wood, T. (Ed.) (2008). *The international handbook of mathematics teacher education*. Rotterdam: Sense Publishers.

Zieffler, A. S. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Tesis doctoral. Universidad de Minnesota.