UNIVERSIDAD DE GRANADA

Facultad de Ciencias de la Educación

Departamento de Didáctica de la Matemática

FUNDAMENTOS PARA UN ESTUDIO SOBRE LA DIDÁCTICA DE LA CORRELACIÓN Y REGRESIÓN

TRABAJO FIN DE MÁSTER

María Magdalena Gea Serrano

Dirigido por:

Dra. Carmen Batanero Bernabéu

GRANADA, 2012

FUNDAMENTOS PARA UN ESTUDIO SOBRE LA DIDÁCTICA DE LA CORRELACIÓN Y REGRESIÓN

Carmen Batanero Bernabeu

María Magdalena Gea Serrano

Departamento de Didáctica de la Matemática
Facultad de Ciencias de la Educación
Universidad de Granada
18071 Granada

ISBN:

Depósito legal:

Esta obra forma parte del proyecto: EDU2010-14947 (MICINN-FEDER) con la colaboración de la beca BES-2011-044684 (MICINN-FEDER), dentro del grupo de investigación *Teoría de la Educación Matemática y Educación Estadística* de la Universidad de Granada, perteneciente al Plan Andaluz de Investigación, Desarrollo e Innovación de la Junta de Andalucía (FQM126).

AGRADECIMIENTOS

Agradezco de manera muy especial a mi profesora Dr. Carmen Batanero Bernabéu la enseñanza que me ha brindado en cada momento que la he necesitado, su tiempo y dedicación en este trabajo así como las continuas ideas que han posibilitado que este trabajo se haya presentado en tiempo y forma.

A los profesores de la Universidad de Granada del Departamento de Didáctica de la Matemática por la enseñanza recibida en cada uno de las asignaturas cursadas.

INDICE

INDICE

INTRODUCCIÓN

El Trabajo Fin de Máster que se desarrolla en esta Memoria se centra en las nociones de correlación y regresión y se plantea con la finalidad de servir de fundamentación teórica para mi futura tesis doctoral. Dicha tesis se centrará en investigar el conocimiento profesional del profesor de matemáticas en torno a estas nociones, en el sentido propuesto por Ball, Thames y Phelps (2008), en futuros profesores de Educación Secundaria y Bachillerato.

La aportación de conocimiento de este trabajo se refiere, principalmente, al análisis de investigaciones específicas desarrolladas en torno a la correlación y regresión, y por otro lado al análisis de dos libros de texto en que se presentan estas nociones, basándonos en el Enfoque Ontosemiótico de la cognición e instrucción matemática, herramienta desde la cual pretendo desarrollar mi futura tesis. Estas y otras aportaciones se distribuyen en los siguientes capítulos:

El primer capítulo muestra la importancia de la Educación estadística, se analizan la correlación y regresión desde diversos puntos de vista (origen histórico, importancia en estadística, los objetos matemáticos relacionados con el tema y su enseñanza). Se finaliza con los objetivos del trabajo.

El segundo capítulo desarrolla los antecedentes de investigación, organizada en los siguientes apartados: el razonamiento covariacional y las tareas covariacionales; las tareas correlacionales, pasos y estrategia, estimación de la correlación a partir de diferentes representaciones, sesgos y concepciones incorrectas, desarrollo del razonamiento correlacional en la enseñanza, sugerencias para la enseñanza del tema y estudios sobre libros de texto.

Tras este recorrido teórico, que proporciona el soporte de nuestra investigación, se presenta en el tercer capítulo un análisis de dos libros de texto posteriores a la publicación del Real Decreto de Enseñanzas Mínimas en Bachillerato (MEC, 2007b) para acercarnos un poco más a la forma en que estos conceptos se introducen en la enseñanza en este nivel educativo. Utilizando nuestro marco teórico, que se presenta en el Anexo 1, se describen los diferentes objetos matemáticos y la forma en que se presentan, así como algunos conflictos semióticos encontrados en los libros.

La Memoria se completa con las conclusiones, referencias, el marco teórico (Anexo 1), el significado de la regresión y correlación en nuestro estudio e importancia en estadística (Anexo 2), una lista de términos utilizados en cada libro de texto (Anexo 3), una tarea covariacional resuelta por uno de los libros de texto analizados (Anexo 4) y el conjunto de trabajos que he ido realizando en este período de Máster (Anexo 5).

CAPÍTULO 1.
EL PROBLEMA DE INVESTIGACIÓN

1.1. INTRODUCCIÓN

En este capítulo se presenta la problemática de la investigación, y se justifica su importancia. En primer lugar se describe la emergencia reciente de la educación estadística como campo propio de investigación, y se resalta la importancia de la alfabetización estadística de los ciudadanos. En las siguientes secciones, hacemos una somera descripción de la regresión y correlación, describiendo su origen histórico y resaltando su actual importancia, en particular, en el análisis exploratorio de datos. Seguidamente se analiza la presencia de la Estadística en la enseñanza de las Matemáticas, realizando un estudio más detallado de las nociones de regresión y correlación. Se finaliza con la presentación de nuestros objetivos para este trabajo.

1.2. DE LA ESTADÍSTICA A LA EDUCACIÓN ESTADÍSTICA

La Estadística es una disciplina que ha acompañado al hombre desde su existencia. Como señala Gutiérrez (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*". El autor aconseja distinguir entre *estadísticas* en plural y *estadística* en singular. Las estadísticas, son las colecciones de datos obtenidos mediante el proceso estadístico, mientras que la estadística, como ciencia, "*estudia el comportamiento de los fenómenos llamados de colectivo.*" (Gutiérrez, 1994, p.22). El material de estudio sería esta información recogida, tratada mediante un modo propio de razonamiento "*el método estadístico*" que genera *previsiones* en caso de incertidumbre.

Hace unos años, pocos investigadores se interesaban por la enseñanza y aprendizaje de la Estadística, pero en la actualidad, asistimos a gran interés en este tema (Estrada, 2002). La preocupación por la educación estadística se muestra en los *ICOTS* (*International Conference on Statistical Education*) que se iniciaron en 1982 o las actividades de sociedades de estadística o de educación que han dedicado temas específicos a la educación estadística como por ejemplo: los Congresos Internacionales de Educación Matemática (*ICME*), *ASA* (*American Statistical Association,*), *AERA* (*American Educational Research Association* y *Sociedad Española de Investigación en Educación Matemática* (Batanero, 2002).

Otro indicador es la fundación en 1991 de la IASE (International Association for Statistical Education) y las revistas orientadas a los profesores como *Teaching Statistics, Induzioni* o *Journal of Statistics Education* (Ottaviani, 1998). También aparecen revistas de investigación como *Statistics Education Research Journal* (SERJ), *Statistique et Enseignement* y *Technology Innovations in Statistics Education*. 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. La principal razón radica en la presencia que ésta adquiere en nuestro entorno, como muestra el libro de Tanur et al. (1972) que subraya la naturaleza humana de la estadística y su presencia en la sociedad donde se requiere un pensamiento estadístico, que se resalta en la siguiente sección.

Necesidad de alfabetización estadística

Es también evidente la presencia de la estadística en la vida diaria del ciudadano (prensa, radio, TV, libros y revistas especializadas, etc.). La estadística puede ser utilizada como método de investigación y como vehículo de transmisión de información y así el ciudadano precisa una terminología y un tipo de razonamiento que requieren de una enseñanza y aprendizaje específicos. Esta necesidad ha sido observada por los organismos productores de estadística. Las estadísticas elaboradas por el gobierno de un país influyen en las vidas profesionales y personales de sus conciudadanos. Es por ello necesario fortalecer el pensamiento estadístico en todos los sectores de nuestra población (Wallman, 1993). Ello lleva a diversas perspectivas en torno a la alfabetización estadística. Por ejemplo, Wallman (1993) nos indica 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, pág. 1)

La cooperación de los ciudadanos para facilitar información en cualquier investigación, o la desconfianza ante el trato de la información suministrada, se relaciona con la "*alfabetización estadística*". Por ello el consumidor**,** no sólo debe entender mejor las estadísticas sino potenciar su pensamiento estadístico (Moore, 1990).

1.3. CORRELACIÓN Y REGRESIÓN

La regresión y correlación subyacen en multitud de métodos estadísticos, pues generalizan la idea de dependencia funcional (Batanero, 2001). Disciplinas como la Psicología, Sociología y la Didáctica Estadística aportan conocimiento sobre estos tópicos, corroborando la dificultad intrínseca del ser humano en la emisión de juicios de asociación. La importancia de estas investigaciones se debe a que 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). La exigencia de un diseño de enseñanza idóneo para estas nociones, no es banal, pues observamos su limitada presencia en la normativa que regula el sistema educativo español[1], aunque autores, como King (2000) han tratado de inducir el razonamiento covariacional 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 y su importancia actual en estadística.

1.3.1. ORIGEN HISTÓRICO

El día a día está repleto de ejemplos en los que tomamos decisiones atendiendo al grado de asociación existente entre las variables que consideramos influyentes. Los juicios sobre la posible asociación de variables, acompañan al ser humano desde el comienzo de su existencia. A modo de 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, motivaron cuestiones que dieron un avance considerable al conocimiento científico como es la construcción de calendarios en cuanto a la medición del tiempo.

Los primeros trabajos científicos que encontramos sobre el tema proceden de 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 donde, la labor investigadora desarrollada por investigadores como Lambert-Adolphe-Jacques Quetelet (1796-1874 fueron de gran apoyo en cuanto al desarrollo de las nociones de correlación y regresión (Hald, 1998).

[1] Los contenidos de regresión y correlación se contemplan en el primer curso de Bachillerato en las asignaturas de Matemáticas I y Matemáticas aplicadas a las Ciencias Sociales I de las modalidades de *Ciencias y Tecnología* y *Humanidades y Ciencias Sociales* respectivamente.

Quetelet, (nacido en Ghent, Bélgica, obtuvo el doctorado en Matemáticas con una tesis sobre secciones cónicas) fue director del observatorio astronómico de Bruselas. En relación con la correlación sus aportaciones se circunscriben a sus estudios sobre el hombre medio, estudiando medias y desviaciones de medidas antropométricas, analizando la influencia de variables independientes como el sexo, edad, profesión o nivel de educación. Relaciona dos o más variables y llega a obtener una ecuación de una hipérbola que relaciona la edad y la altura de las personas entre cero y 30 años. Introduce la distribución normal en Biometría (Hald, 1998). Su originalidad es haber considerado la dispersión y extender el uso de la ley normal (bien conocida en Astronomía) al ajuste de medidas antropométricas, (Seal, 1967).

Otro avance consistió en el estudio conjunto de la variación de dos medias realizado por Francis Galton, quien 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 1.1 que llamó “*quincunx*” (1889), también llamado aparato de Galton.

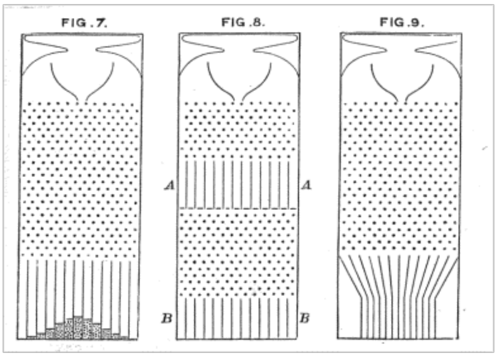


Figura 1.1. Versiones del quincux en “*Natural Inheritance”* (Hald, 1998, p. 605)

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 influían 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 7 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 7 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 1.2) 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 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 de entre hijos del mismo grupo de padres, variación que se observa en el quincunx. La identidad estadística de las dos generaciones significa que $\sigma_x^2 = \sigma_y^2$, en consecuencia, $\sigma_x^2 = r^2 \cdot \sigma_y^2 + \sigma_y^2$, 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).

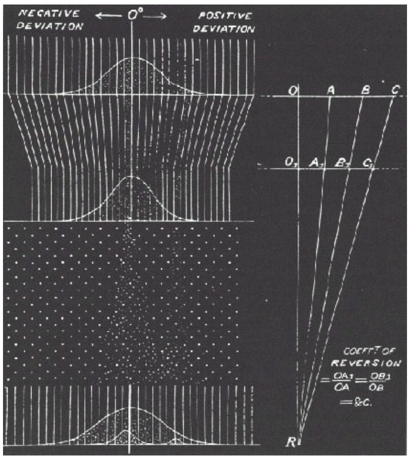


Figura 1.2. Ilustración de Galton de la identidad estadística de dos generaciones (Hald, 1998, p. 606)

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* (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. En su obra, *"Natural Inheritance"* (1889), propone un formidable programa de investigación biométrica: estudiar estadísticamente la variabilidad y la plasticidad de las formas vivientes, a fin de confirmar matemáticamente el mecanismo de la evolución descrito por Darwin (Benzecri, 1982). Este programa es continuado por prestigiosos autores como Edgeworth, Pearson, Yule, Seppard (Hald, 1998) que en tiempos posteriores desarrollaron estas ideas en profundidad.

1.4. ENSEÑANZA DE LA CORRELACIÓN Y REGRESIÓN

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

1.4.1. CONSIDERACIONES GENERALES SOBRE LA ENSEÑANZA DE LA ESTADÍSTICA

La estadística comienza ya en la Educación Infantil en Andalucía (Consejería de Educación, 2008a) 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.*".

Para ello, y mediante "*situaciones siempre vinculadas a su entorno y vivencias cotidianas*" se plantean la necesidad de propuestas educativas que impliquen "*la recogida de datos y la organización de los mismos.*", donde con la ayuda del maestro se 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).

Una noción que deben tratarse desde una temprana edad, como señala el currículo correspondiente a la Educación Infantil en Andalucía (Consejería de Educación, 2008a), es la de situación aleatoria, pues los alumnos perciben el mundo que les rodea desde una concepción determinista. La interdisciplinariedad de la estadística debiese hacerse notar, proporcionando a los niños ricas experiencias que favorezcan el reconocimiento de los aspectos aleatorios y deterministas de nuestro entorno.

El carácter continuo de los procesos de enseñanza en la educación básica (Educación Primaria y Educación Secundaria Obligatoria), mantienen una formulación sensiblemente paralela, considerando los aspectos estadísticos 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. 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, 2007b): *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).

La introducción de nuevos temas de contenido estadístico en los currícula supone un reto para los profesores al tener que preparar, adaptar e impartir los "nuevos contenidos". A todas estas consideraciones, hemos de añadir el aspecto interdisciplinar que podemos relacionar con las Ciencias Sociales, Biología, Geografía e Historia. Las orientaciones de la Ley de Educación de Andalucía en cuanto a la visión interdisciplinar de los contenidos en su enseñanza, hace la tarea del profesor aún más complicada. Se añaden las dificultades de tipo filosófico, epistemológico y didáctico, ligadas a la interpretación y aplicación de los conceptos estadísticos en situaciones prácticas. Una aproximación didáctica a esta problemática es la evidenciada por Batanero (2001) cuando señala la importancia del aprendizaje de la estadística, con una doble funcionalidad:

- Ayudan al alumno a *entender su entorno*, mucho antes de que tome conciencia de la complejidad matemática subyacente.
- *Preparan el conocimiento analítico posterior*. Debido al escaso tiempo de que se dispone, algunos profesores justifican la necesidad de ceñir la enseñanza de los contenidos estadísticos a la transmisión de una serie de fórmulas. Pero, según nuestro marco teórico, los objetos matemáticos emergen del conjunto de prácticas (situaciones-problema) que el alumno realiza. De este modo, cuando el alumno interacciona con los objetos matemáticos involucrados en las diferentes situaciones-problema, adquieren significado personal en el alumno. Si las nociones estadísticas se presentan a los alumnos de un modo limitado, les impide adquirir un razonamiento superior, que permita ser generalizado y no se limite a la práctica particular. Pueden propiciar la impresión de que las técnicas estadísticas son más importantes que los resultados obtenidos. En cualquier caso, el alumno se limita a aplicar la estadística como un conjunto de fórmulas (que no entiende), que ofrecen conclusiones que, tristemente, no saben interpretar.

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 totalmente contraria a los principios que rigen la actual Ley Orgánica de Educación, donde se pretende proporcionar una educación de *calidad* a *todos* los ciudadanos y ciudadanas. Pues ello implicaría que una parte importante de la población quedaría privada de la adquisición de conceptos y técnicas de gran importancia para su vida. Por otra parte, 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 la probabilidad desde edades tempranas.

No vamos a negar la dificultad de ciertas nociones estadísticas, pero como decíamos anteriormente, existen diferentes niveles de comprensión susceptibles de ser adaptados a cada etapa educativa. Para ello, trabajar las nociones estadísticas básicas desde la Educación Infantil, adecuando paulatinamente los niveles de profundidad a la madurez de los alumnos, permitirá construir una idea más clara de los diferentes contextos de los datos, del fundamento de las técnicas estadísticas aplicadas, y en definitiva, una visión más real de la ciencia Estadística. Además la introducción de la estadística permite ampliar la cultura estrictamente determinista común en la clase de Matemáticas.

Dadas estas consideraciones sobre la enseñanza de los contenidos estadísticos, nos centramos a continuación en las nociones de regresión y correlación.

1.4.2. LA CORRELACIÓN Y REGRESIÓN EN LAS ORIENTACIONES CURRICULARES

Emitir juicios de asociación efectivos en la toma de decisiones, implica en particular, el dominio de las nociones de correlación y asociación. Estas nociones no son tratadas como contenidos de enseñanza en la educación básica, aunque por ello no debemos desligar su presencia implícita en el estudio de las tablas de contingencia. Una breve mención a ellas aparece ya en la Educación Primaria (MEC, 2006, p.43099) donde se hace mención a *"Lectura e interpretación de tablas de doble entrada de uso habitual en la vida cotidiana."*.

También implícitamente se presenta en la asignatura de Matemáticas en cuarto de ESO (MEC, 2007a), en sus dos opciones: A y B, donde como criterio de evaluación se indica que "*Se pretende, además, que los resultados obtenidos se utilicen para la toma de decisiones razonables en el contexto de los problemas planteados*" (pp. 758 y 760), en cuanto 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, donde se indica que: "*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 postobligatoria de Bachillerato, las nociones de 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:

> 6. 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.

En la asignatura de Matemáticas aplicadas a las Ciencias Sociales I, se concretan estos contenidos en el bloque 3 denominado “*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:

> 6. 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.

Al desarrollar los núcleos de contenidos propuestos en dicho Real Decreto (MEC, 2007b) se remite a estos contenidos, que deben desarrollarse teniendo en cuenta cuatro núcleos transversales: La resolución de problemas; Aprender de y con la Historia de las Matemáticas; Introducción a los métodos y fundamentos matemáticos y Modelización matemática (Consejería de Educación, 2008b).

Como hemos indicado, la relevancia y el sentido educativo de estas nociones es manifestado ya, en el currículo correspondiente a la Educación Secundaria Obligatoria, aun no contemplándose explícitamente la enseñanza de estas nociones en el desarrollo curricular de esta etapa educativa. La adquisición de estas nociones es un útil para todo ciudadano en su vida en sociedad, y ésta es la motivación principal de nuestro estudio.

1.4.3. PERSPECTIVA INTERNACIONAL

La asociación, correlación y regresión también se incluye en los documentos curriculares de otros países en la etapa de secundaria. Como ejemplo, analizamos el currículo americano, donde se incluye en los grados 6-8 y 9-12, es decir, en edades correspondiente a nuestra secundaria y Bachillerato.

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 que es necesario que los ciudadanos tengan una base suficiente sobre el análisis de datos y algunas nociones elementales de probabilidad para poder razonar de manera estadística; destreza necesaria 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 el modelo de 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 a aprender 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: estudiando y analizando los promedios y estadísticos de dispersión de dichas variables y por medio de 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 mayor 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 y discusión y debates sobre las actividades y resultados de las mismas. Unos de los objetivos sería que los estudiantes tuviesen una participación más activa.

A raíz de estudiar dichos documentos, se concluye que el currículo norteamericano es muy avanzado en lo que se refiere a estadística y se supone que al ingresar en la universidad los estudiantes comprenden los rudimentos de la estadística descriptiva y análisis de datos. Este currículo ha tenido mucha influencia en el desarrollo del actual currículo de Educación Primaria en España, por lo que sus recomendaciones debieran tenerse en cuenta en la formación de profesores. 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.

1.5. OBJETIVOS DEL TRABAJO

Nuestro trabajo se inscribe dentro del Proyecto de Investigación EDU2010-14947 “Evaluación y desarrollo de conocimientos matemáticos y didácticos de profesores. Aplicación a los contenidos relacionados con la estadística y probabilidad”. Más concretamente, en una futura tesis doctoral se pretende construir un instrumento válido y fiable de evaluación de los *conocimientos matemáticos para la enseñanza de la correlación y regresión* según el modelo de Ball, Thames y Phelps (2008), apoyado en el Enfoque Onto-semiótico para la didáctica de la matemática (Ver Anexo 1) utilizando la metodología propuesta por Godino (2009).

Un primer paso para la construcción de este cuestionario es construir los fundamentos adecuados para definir de una forma válida "el conocimiento matemático para la enseñanza de la correlación y regresión", que es un constructo inobservable (León y Montero, 2003). Dichos fundamentos están constituidos por el marco teórico del estudio (Anexo 1), la revisión de las investigaciones sobre correlación y regresión (Capítulo 2) y por el estudio curricular y matemático de estos objetos que se ha llevado a cabo en este capítulo. Asimismo se complementa con el estudio del tema en dos libros de texto (Capítulo 3). Consecuentemente, los objetivos de esta Memoria de Máster son los siguientes:

O1. Analizar los contenidos sobre correlación y regresión desde el punto de vista elemental y descriptivo para identificar los objetos matemáticos específicos en los que se enfocará la evaluación del conocimiento matemático para la enseñanza de los futuros profesores. Este objetivo se aborda en este Capítulo 1.

O2. Analizar asimismo los contenidos curriculares del tema en el Bachillerato, que también se desarrolla en este capítulo, donde además se extiende el análisis curricular a los estándares del NCTM y el proyecto GAISE.

O3. Profundizar en el marco teórico que será fundamento del trabajo de tesis futuro. En el Anexo 1 se realiza un resumen extenso de los principales documentos producidos en el Enfoque Onto-semiótico.

O4. Realizar una revisión bibliográfica y clasificación sobre la investigación específica relacionada con el tema. Este estudio se recoge en el Capítulo 2 y permitirá familiarizarse con el campo de investigación, y recoger ítems que podrían ser útiles en la construcción de los que se incorporen al cuestionario de profesores. También permitirá recoger ejemplos de respuestas correctas e incorrectas dadas por los estudiantes a ítems de evaluación, que pudieran usarse en la construcción de ítems dirigidos a la evaluación de los profesores.

O5. Realizar un análisis del contenido que sobre correlación y regresión se incluye en dos libros de texto del Bachillerato de Ciencias Sociales. Este estudio se desarrolla en el Capítulo 3 en que, utilizando elementos de nuestro marco teórico se analizan los campos de problemas, lenguaje, propiedades, conceptos, propiedades, procedimientos y argumentos presentados. Todo ello con la finalidad de completar el significado de referencia de la correlación y regresión para nuestra futura tesis.

CAPÍTULO 2.
ANTECEDENTES

2.1. INTRODUCCIÓN

El presente capítulo, constituye una síntesis de las investigaciones más relevantes sobre la regresión y correlación. Aunque nos restringimos al estudio de la asociación entre variables cuantitativas, para enmarcar el estudio, se comienza con una introducción al *razonamiento covariacional* y los pasos en tipo de tareas. Seguidamente, se describen las estrategias intuitivas de los estudiantes en la detección de la correlación, a partir de diagramas de dispersión, los sesgos en estas tareas, las concepciones, precisión en la estimación del coeficiente de correlación desde diversas representaciones, y desarrollo del razonamiento correlacional con la enseñanza. Un segundo bloque es el análisis de la libros de texto, donde comenzamos con unas consideraciones sobre la transposición didáctica y se describen algunas investigaciones sobre la presentación de la estadística en los textos de secundaria.

2.2. INVESTIGACIONES SOBRE LA CORRELACIÓN Y REGRESIÓN

2.2.1. RAZONAMIENTO COVARIACIONAL

Explicar, controlar y predecir los sucesos que se presentan en nuestro día a día, depende de habilidades para detectar covariaciones (Alloy y Tabachnik, 1984). El razonamiento covariacional como actividad cognitiva fundamental no está exento de dificultades (Moritz, 2004; Zieffler, 2006; McKenzie y Mikkelsen, 2007). Alloy y Tabachnik (1984) plantean la cuestión: "*¿Bajo qué condiciones son los organismos precisos en la detección de la covariación entre sucesos?*" (p. 113), reduciendo a dos las fuentes de información relevantes para percibir la covariación entre dos eventos:

- Las *expectativas previas* o *creencias*. El "individuo" posee ideas o juicios sobre la asociación de los fenómenos tratados, debidas a la experiencia previa en situaciones similares o bien a otras fuentes por ejemplo, la transmisión cultural o biológica.
- La *información objetiva* de la situación aleatoria o información acerca de las relaciones de los fenómenos en el entorno, que puede ser débil debida a la poca experiencia con los fenómenos de la situación, o una información ambigua.

Atendiendo a esta caracterización, McKenzie y Mikkelsen (2007) aportan un punto de vista bayesiano al razonamiento covariacional, indicando que nuestro sistema cognitivo se aproxima a la tarea de covariación desde un marco inferencial. Sugieren que el ser humano utiliza el procedimiento bayesiano para evaluar la covariación, dado que las personas son sensibles a las creencias previas y a la rareza de los datos. También indican que las condiciones en el laboratorio influyen en el razonamiento de los sujetos.

2.2.2. PASOS EN UNA TAREA COVARIACIONAL

Diversos autores como Crocker (1981) o Moritz (2004) distinguen subtareas en que se puede descomponer un estudio covariacional, desde recoger datos para estudiar la asociación estadística hasta 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 fundamente para hacer predicciones o emitir juicios.

Moritz (2004) describe una secuenciación resumida en cuatro pasos, donde el inicial es la generación de hipótesis sobre la asociación de las variables de estudio, precisando, además, la representación gráfica de los datos recolectados. Una implicación didáctica es la importancia de considerar la construcción de gráficas estadísticas en el proceso de enseñanza. Por ejemplo, las gráficas de series temporales sirven para introducir la asociación de un modo natural.

Zieffler (2006) indica que la mayoría de los estudios se centran en el paso (6) anterior, es decir, determinar si existe relación entre dos variables y justificar esta decisión. En las tablas de contingencia abarca de cuatro categorías de respuestas:

- *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, inadecuadamente.
- *Covariación avanzada.* Los juicios usan adecuadamente las cuatro celdas.

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

- Las creencias o teorías previas de los individuos sobre la relación entre las variables de estudio, tienen una 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;

- Hay gran dificultad para el razonamiento covariacional negativo, esto es, para identificar la asociación inversa (Estepa, 1994);
- Los juicios de asociación tienden a evidenciar una asociación inferior a la que realmente presentan las variables de estudio; y
- Existe una tendencia a establecer relaciones de causalidad en el estudio de la asociación de entre variables (también descrito por Estepa, 1994),

2.2.3. LAS TAREAS DE CORRELACIÓN

En el desarrollo del razonamiento covariacional, se requieren objetos matemáticos relacionados con el análisis de datos bivariados. Los más importantes son los referidos a la correlación y regresión entre variables numéricas, donde nos centramos para fundamentar el diseño del proceso de enseñanza y la evaluación de la capacidad de nuestro alumnado en el desempeño de las nociones tratadas ya que:

> *"Junto con el desarrollo del currículum para promover el razonamiento del estudiante en estadística se debe proceder a la evaluación de este razonamiento."* (Zieffler, 2006, p. 3).

La investigación en Psicología llevada a cabo en un amplio rango de edad (desde primaria hasta estudiantes universitarios), ha producido resultados sólidos (Zieffler, 2006). Y aunque las investigaciones didácticas se interesen más por cómo los estudiantes razonan en un contexto educativo las tareas dadas 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 la pobreza de razonamiento del ser humano en tareas relativas a su resolución (Zieffler, 2006; McKenzie y Mikkelsen, 2007). En las siguientes secciones, se recogen los resultados más relevantes.

2.2.4. ESTRATEGIAS EN LA DETECCIÓN DE LA CORRELACIÓN A PARTIR DE DIAGRAMAS DE DISPERSIÓN

Diversas investigaciones (Sánchez Cobo, 1998; Sánchez Cobo, Estepa y Batanero, 2000; Cañadas, 2010; Cañadas, Batanero, Contreras y Arteaga, 2011) llevan a cabo un estudio pormenorizado de la resolución de las tareas de covariación propuestas a los alumnos, clasificándose los tipos de problemas en tres tipos (Sánchez Cobo, 1998):

a. *Juicios de asociación en tablas de contingencia.* Se trata de 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 asociación entre dos variables numéricas.

c. *Juicios de asociación en la comparación de muestras.* Se estudia 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, 1998; Cañadas, 2010). Nosotros nos restringimos al estudio de las estrategias empleadas en la detección de la covariación entre dos variables numéricas, es decir, a partir de diagrama de dispersión.

Estrategias empleadas en diagramas de dispersión

Estepa y Batanero (1996) estudian las estrategias en los juicios de asociación en una muestra de 213 estudiantes del curso preuniversitario, entre ellas cuando los datos se dan en un diagrama de dispersión, donde puede determinarse visualmente la asociación entre las dos variables. Clasifican las estrategias encontradas en tres tipos:

- Estrategias correctas:
    1. *Comparación global.* Cuando el alumno da un juicio global correcto 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.* Se compara la forma de la nube de punto con una función conocida, por ejemplo, lineal o cuadrática.
    4. *Interpretación* correcta de puntos aislados que cumplen el tipo de relación existente entre las variables.
- Estrategias incorrectas:
    5. *Interpretación incorrecta de puntos aislados*. El alumno emplea pares aislados de valores para interpretar, de forma incorrecta, la relación entre las variables.
    6. *Teorías previas.* Se usan las teorías previas sobre el contexto para argumentar la asociación, es decir, se manifiesta explícitamente 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.
    9. *Causalidad.* A pesar de la asociación empírica, se argumenta que no existe asociación entre las variables, ya que no existe relación causa-efecto.

A modo de réplica, Sánchez Cobo (1998), corrobora la existencia de tres ejes que explican las relaciones existentes entre las tareas planteadas y las estrategias empleadas:

- En primer lugar, encuentra una oposición entre el razonamiento numérico y 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, 1998, p. 252).
- También se muestra un empleo de argumentaciones conjuntas (numéricas, gráficas, y teorías previas) donde tan solo se requiere un argumento gráfico, principalmente en situaciones cercanas al alumno donde "*le es menos necesario el recurso a estrategias complementarias*." (Sánchez Cobo, 1998, p. 253).
- En tercer lugar, existe 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. En este caso el alumno tiende a responder según sus teorías previas o con argumentos incorrectos.

2.2.5. ESTIMACIÓN DE LA CORRELACIÓN A PARTIR DE DIFERENTES REPRESENTACIONES

Sánchez Cobo (1998) hizo notar que los alumnos 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. La estimación es más precisa al estimar el coeficiente de correlación a partir de un diagrama de dispersión, así como la tarea inversa y cuando la 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, 1998, p.210).

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.
- La dificultad en diferenciar la variable explicativa de la explicada en el cálculo de la recta de regresión. 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.

El razonamiento sobre la covariación implica procesos de traducción entre los datos, sus representaciones gráficas y descripciones verbales, así como procesos tales como el cálculo estadístico, uso de modelos matemáticos para el ajuste de datos y traducción desde y hacia las expresiones algebraicas y gráficas utilizadas, así como juicio sobre posible existencia de relaciones causales. Moritz (2004), presenta un resumen (Figura 2.2.5.1) donde se reflejan estas destrezas o habilidades requeridas.

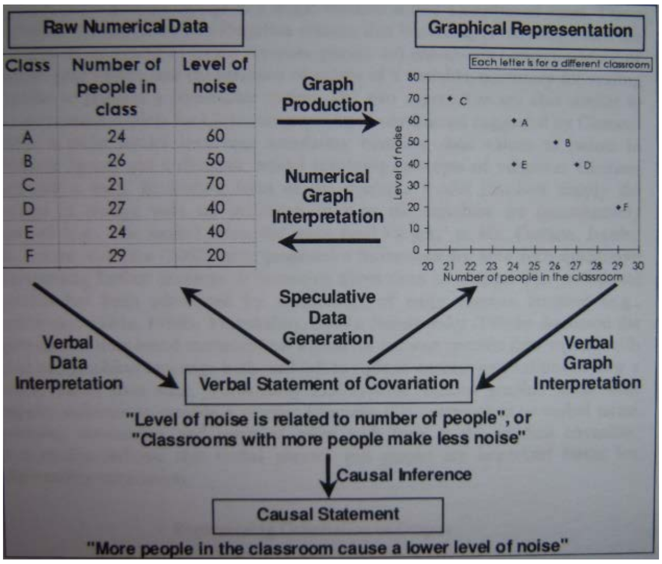


Figura 2.2.5.1. Representaciones y destrezas de la covariación estadística (Moritz, 2004, p. 230).

El estudio de Sánchez Cobo y cols. (2000), nos acerca un poco más a comprender el modo en que los alumnos universitarios adquieren la noción de correlación, todo ello cimentado en estudios previos (Batanero, Estepa y Godino, 1997; Batanero y Godino, 1998; Batanero, Godino y Estepa, 1998; Estepa 1994; Estepa y Batanero, 1996; Estepa y Sánchez Cobo, 1998). Sánchez Cobo y cols. (2000), abordan una investigación que considera cuatro formas de representar la correlación entre dos variables cuantitativas:

b. *Descripción verbal,* cuando describimos una distribución bivariada mediante el lenguaje natural
c. *Tabla de valores,* o presentación de un conjunto de pares de valores numéricos de una distribución bivariada,
d. *Diagrama de dispersión,* cuando el conjunto de pares de valores de una distribución bivariada se presentan mediante un diagrama cartesiano, y
e. *Coeficiente de correlación,* cuando se da el coeficiente de correlación existente entre dos modalidades de una distribución bivariada.

Con las seis tareas, con cinco apartados (subtareas) cada una, 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, (subtarea t2d).
- *Interdependencia*: Cuando la ocurrencia de una variable X influye en la ocurrencia de una variable Y, y viceversa (como por ejemplo en la subtarea t1a).
- *Dependencia indirecta*: Dos variables pueden mostrar cierta covariación debido a la variación de una tercera variable que está correlacionada con ambas, produciendo una asociación aparente (como en la subtarea t2e).
- *Concordancia*: Correlación producida por la ordenación de un conjunto de datos por dos personas de forma independiente (por ejemplo, la subtarea t4b).
- *Covariación casual*: Cuando parece que en la covariación de dos variables hay cierta sincronía, lo cual podría interpretarse como la existencia de asociación entre ambas; sin embargo, ésta es casual o accidental (como en la subtarea t1e).

Se atiende igualmente a la intensidad y dirección de la correlación, a la posible linealidad, las teorías previas de los alumnos, la precisión de las estimaciones, las estrategias empleadas, y la identificación de situaciones reales en que se presente un valor dado del coeficiente de correlación. Los alumnos muestran una buena capacidad de estimación de la correlación, mejorando cuando mayor es la intensidad de ésta.

Moritz (2004) se estudia tres destrezas importantes: (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, (c) e interpretar tablas de frecuencias de datos covariados. Destaca las siguientes dificultades:

- Considerar tan sólo la correspondencia de datos bidimensionales aislados, en lugar de considerar la tendencia global de éstos;

- 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;
- Las 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.

McKenzie y Mikkelsen, (2007 indican que, aunque las creencias y teorías previas son imprecisas e inexactas, son de interés, 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.

2.2.6. 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. 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 la *correlación ilusoria* (Chapman y Chapman, 1967). Se define del siguiente modo:

> *”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 o siente frente a lo que cualquier informe le pueda proporcionar (Chapman y Chapman, 1967). Por ello muchos autores proponen advertir a nuestros estudiantes de la dificultad que conlleva un razonamiento covariacional correcto, con el fin de sensibilizar ante el peligro que acarrean sus sesgos (Chapman y Chapman, 1967,1969; Moritz, 2004). En concreto, Chapman y Chapman defienden la necesidad de incluir, en el periodo de formación de los futuros profesionales de la medicina, la evidencia manifiesta de estos sesgos y en la toma de decisiones dada su responsabilidad (Chapman y Chapman, 1967, 1969). Engel y Sdelmeir (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.

2.2.7. CONCEPCIONES SOBRE LA CORRELACIÓN

En su estudio Estepa (1994) observar 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 asociación, superpuestas con otras correctas. A lo largo de sus trabajos describe las siguientes (Estepa, 1994; Estepa y Batanero, 1996; Batanero, Estepa y Godino, 1997):

- *La concepción determinista de la asociación*. Los alumnos tienden a asignar un único valor de la variable independiente a cada uno de los valores considerados de la variable dependiente. Esto es, la relación de las variables sólo es considerada desde un punto de vista funcional. En casos extremos, exigen la existencia de una fórmula que ligue las dos variables (*concepción algebraica* de la asociación).
- *La concepción local de la asociación.* Los alumnos se limitan a confirmar la asociación según un subconjunto de éstos que de algún modo justifique algún tipo de patrón, obviando la tendencia global de los datos.
- *Concepción unidireccional:* En este caso del estudiante que no admite la asociación inversa, considerándose la intensidad de la asociación, pero no su signo y considerando la dependencia inversa como independencia.
- *Concepción causal:* Cuando sólo considera la dependencia entre variables si puede adjudicarse a la presencia de una relación causal entre las mismas.
- *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 estarlos A y C.

Esta propiedad no se comprende por parte de los estudiantes (Castro-Sotos, Van Hoof, Van den Noortgate, & Onghena, 2009).

2.2.8. DESARROLLO DEL RAZONAMIENTO CORRELACIONAL EN LA ENSEÑANZA

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

Estudio de Estepa

Estepa (1994) realiza también un estudio del efecto de la enseñanza basada en un uso intensivo de ordenadores en un curso de análisis exploratorio de datos con 22 futuros maestros. Batanero, Estepa y Godino (1997) estudian los cambios después de la instrucción, mediante un cuestionario, observándose que 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, pues en un diagrama en que la correlación se debía a concordancia (y no a relación causa-efecto) no mejoraron las estrategias. Los autores indican que gran parte de los estudiantes de su muestra superaron la concepción determinista y la concepción local de la asociación, pero, en general, no se observa mejora respecto a la concepción causal de la correlación. Por consiguiente, creen 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,se deduce que el ordenador por si sólo no es suficiente para que los estudiantes adquieran competencia en el análisis de datos.

Actos de comprensión de la asociación

Los autores analizaron, mediante grabaciones, entrevistas y análisis de las tareas escritas y resueltas por ordenador, el proceso de enseñanza para dos estudiantes, y describieron los siguientes actos de comprensión (Sierpinska, 1994) de 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,* pues el papel de la condición y lo que se condiciona no es intercambiable. Falk (1986) y otros autores han señalado que los estudiantes no discriminan entre la probabilidades P(A/B) y P(B/A). Muchos estudiantes en el estudio 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 sesión 5, los estudiantes no descubrieron esta propiedad.
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 esa correlación ignora la distinción entre explicativas y variables de respuesta, mientras en la regresión esta diferencia es esencial y causa mucha confusión para los estudiantes.
5. *Una correlación positiva señala a una asociación directa entre las variables*. Aunque, en la sesión 6, los 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.
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. No usaron el término "asociación inversa" explícitamente, ni diferenciaron entre los dos tipos de la relación 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.

Investigación de Sánchez Cobo

El estudio descriptivo del razonamiento covariacional, es objeto de interés de varios investigadores, como la desarrollada por Sánchez Cobo (1998) quien intenta caracterizar el significado personal que los alumnos universitarios dan a la correlación y regresión al finalizar un curso de introducción de Estadística en la Universidad, mediante la descripción de los errores conceptuales y procedimentales de los alumnos, la estimación que hacen del coeficiente de correlación a través de diferentes representaciones de la correlación (verbal, tabla y gráfico) y la capacidad de traducción entre dichas representaciones.

Su muestra estuvo compuesta de 193 estudiantes de las carreras de la Diplomatura de Empresariales (104) y de la 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 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 (1998) 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, 1998, p.210).

El autor sugiere que esto puede ser debido, a que los diagramas de dispersión son una de las representaciones que los alumnos han usado con frecuencia en la enseñanza. 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.
- La dificultad de diferenciar la variable explicativa de la explicada en cuanto al cálculo de la recta de regresión. 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. Casi la mitad de los sujetos del estudio consideran que si existe correlación positiva, ésta se deduce en una dependencia lineal.

Investigación de Zieffler

Algunos factores que podrían explicar el razonamiento covariacional de nuestros estudiantes son la secuenciación de los temas en un curso de estadística, o el desarrollo del razonamiento de los estudiantes sobre temas anteriores. La investigación de Zieffler (2006) pretende, caracterizar el desarrollo del razonamiento covariacional de los estudiantes en cuanto a datos bivariados cuantitativos y en un segundo nivel, estudiar los factores explicativos del razonamiento covariacional. El autor realiza una investigación en un curso universitario de introducción con 113 estudiantes y parte de evaluaciones del razonamiento de los estudiantes sobre las nociones referidas a la distribución de una variable estadística, previo análisis de diseño de muestras (muestreo) y AED. Dos profesores ordenaron al azar los temas referentes al análisis bidimensional. Uno comenzó 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 se enseñó al final de curso.

Zieffler observó un desarrollo del razonamiento covariacional marcado por la idiosincrasia de cada individuo. 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. Un hallazgo interesante es que la mayor parte del cambio ocurre al inicio de la instrucción, previa 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, aunque es posible que la brevedad de la unidad en el curso de introducción a la estadística, afecte al desarrollo posterior de este razonamiento.

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, previa a otros temas, como por ejemplo la inferencia, dadas las anécdotas a que se enfrentó con los estudiantes. Por último, en cuanto a la tercera cuestión planteada, se encuentra una marcada influencia del estudio de la variable estadística unidimensional en el desarrollo del razonamiento covariacional. Respecto a estas dos últimas cuestiones, Sánchez Cobo advierte que:

> "*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, 1998, p. 291).

luego se defiende la secuenciación de trabajar en primer lugar la correlación y luego abordar la regresión. Además, este autor aboga por

> "*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, como ha puesto de manifiesto los resultados de esta investigación"* (Sánchez Cobo, 1998, p. 291).

2.3. ANÁLISIS DE LIBROS DE TEXTO

2.3.1. INTRODUCCIÓN

El estudio del proceso de enseñanza y aprendizaje del saber matemático debe ser concebido desde un enfoque sistémico, cuya base o esquema de investigación se fundamenta en 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 requiere de ciertas modificaciones para poder ser enseñado (constituyéndose así en saber a enseñar), y la noción de transposición didáctica supone una poderosa herramienta que da respuesta a las restricciones (anteriormente expuestas) que pesan sobre el sistema de enseñanza. Según Chevallard:

> "*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 (unas veces son de la vida real y otras veces surgen del propio saber matemático y sus relaciones). Sin embargo, el productor del saber, el investigador, cuando lo comunica, lo despersonaliza y lo descontextualiza, quedando oculto todo este proceso de generación. Éste se comunica de manera limpia, secuenciada.

Cuando el saber se encuentra constituido, las transformaciones, reformulaciones, generalizaciones y aplicaciones que sufre, algunas veces pueden hasta hacer que pierda su identidad. Algunas veces, los saberes son destruidos sea porque se identifican con otros saberes ya conocidos, o bien, porque se incluyen en saberes más fuertes, o, porque simplemente se olvidan. El trabajo del profesor es, en cierta medida, inverso al del productor del saber, debe producir una recontextualización y una repersonalización de los conocimientos. Al respecto, cabe señalar la siguiente reflexión de Chevallard:

> "*Es muy necesario que el proceso de aprendizaje sea secuencial: pero el orden de aprendizaje no es isomorfo en relación con el orden de exposición del saber; el aprendizaje del saber no es el calco del texto del saber*" (Chevallard, 1991, p.74).

Tengamos en cuenta que estos conocimientos van a ser los conocimientos del alumno, es decir, una respuesta natural a las condiciones particulares y personales que son indispensables para que los nuevos conocimientos tengan sentido para el estudiante. Cuando el estudiante se ha apropiado de los nuevos conocimientos con sentido y ha adquirido un cierto dominio de ellos, debe re-descontextualizar y re-despersonalizar sus nuevos conocimientos e identificarlos con los existentes en la comunidad matemática científica. Es por ello necesaria la transformación/adaptación de los contenidos de saber para constituirse en objetos de enseñanza, evitando, de algún modo, la ilusión de transparencia en la enseñanza de los objetos matemáticos, implicando en ciertos casos la creación de diversos objetos a exigencia de este proceso, como fue el caso de los diagramas de Venn en cuanto a la enseñanza de la teoría de conjuntos en la escuela primaria (Chevallard, 1991).

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.

Son varios los autores que, en nuestro grupo de investigación han analizado los libros de texto, utilizando elementos del enfoque ontosemiótico, con la finalidad de establecer un significado de referencia en su trabajo, bien para la construcción de cuestionarios de evaluación, o para el diseño de procesos de estudio. A nivel universitario, destacamos los trabajos de Alvarado (2007) sobre el teorema central del límite, Cobo (2003) y Mayén (2009) sobre las medidas de tendencia central, Ortiz (1999) sobre conceptos probabilísticos y Olivo (2008) sobre intervalos de confianza. Debido a la restricción de longitud de la Memoria, no nos es posible describir con detalle los anteriores trabajos.

2.3.2. ESTUDIO SOBRE LA PRESENTACIÓN DE LA CORRELACIÓN Y REGRESIÓN EN LOS LIBROS DE TEXTO

Una primera investigación sobre el tema es la de Sánchez Cobo (1998), quien analiza 11 libros de texto de tercer curso de Bachillerato, nivel en que en aquella fecha 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 taxonomía 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, que pudieran transmitir la visión de las matemáticas como conjunto de reglas y hechos a ser recordados. 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.

Otra conclusión es que "*apenas se presentan definiciones de tipo instrumental-relacional*" (Sánchez Cobo, 1998, p. 287) en los libros de texto, siendo en su mayoría de tipo instrumental, pudiéndose con ello "*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, 1998, 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 el tema de la covariación. La mayoría de los textos incluye 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 la vertiente 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 (1998) 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 presentar a los alumnos contextos y datos reales.

Como complemento a la investigación desarrollada por Sánchez Cobo y cols. (2000) encontramos la desarrollada por Lavalle y cols. (2006). Los autores muestran la

> *"necesidad de investigar acerca de cuáles son los procesos de enseñanza que favorecen un aprendizaje significativo para los alumnos."* [Su estudio supone] *"un punto de partida para el análisis de cómo funcionan las actividades propuestas en la práctica áulica y sus consecuencias didácticas"* (Lavalle y cols., 2006, p. 404).

Las autoras analizan del 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), así como los procedimientos con los que se vincula, y sus relaciones entre ellos. Estos son:

- *Variable estadística*, tanto unidimensional como bidimensional: variable explicativa (X) y variable explicada/respuesta (Y); Identificar las variables en estudio; Distinguir entre una variable respuesta y una variable explicativa; Identificar las unidades de medida; Calcular la media y la desviación.
- *Datos* bidimensionales que se tratan para ser comparados, analizados e interpretados, es decir, pares ordenados que pueden ser representados mediante puntos del plano cartesiano; 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.); Observando el gráfico

de dispersión de una relación lineal, indicar aproximadamente qué tipo y grado de relación lineal existe (directa, inversa)

- *Gráfico 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. Interpretar valores del coeficiente de correlación.
- *Recta de regresión*, para obtener una relación funcional entre las variables que contemple 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;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 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 de sus correspondientes valores asociados en la variable explicativa (X). Si el valor de X que será sustituido en la recta de ajuste mínimo cuadrática se encuentra en el rango de valores observados de dicha variable, ocurre una interpolación. Si el valor de X se encuentra fuera del rango de valores, ocurre una extrapolación; Calcular valores estimados de y. Indicar gráficamente el error de estimación Interpretar el error como desviación

Las autoras analizan siete libros de texto, centrándose en: el enfoque con que se presentan las nociones tratadas, el nivel de profundidad, si se deducen las fórmulas referentes a las nociones del apartado anterior, 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.

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.

Las autoras proponen realizar un nivel de enseñanza de tipo descriptivo, donde no se pretende tratar las nociones de modelo ni de inferencia estadística. Con un marcado cuidado por disponer de información previa a la secuencia de enseñanza relativa a los conocimientos matemáticos y estadísticos de los alumnos (función, ecuación lineal, variación, etc.), y haciendo presente en su enseñanza los resultados de la literatura en torno a los errores y dificultades del aprendizaje de los alumnos en cuanto a las nociones de regresión y correlación, sugieren una serie de actividades referidas al tratamiento de datos bivariados, la relación entre las variables y la noción de correlación y la noción de regresión basado en tres pasos fundamentales: (1) tratamiento intuitivo, (2) aproximación gráfica y (3) análisis numérico.

2.4. CONCLUSIONES

En este capítulo se ha presentado un resumen de las principales investigaciones relacionadas con las ideas previas y el aprendizaje de la correlación y regresión y un breve resumen de algunas investigaciones sobre libros de texto.

Aunque el estudio de la covariación no ha tenido mucha relevancia en didáctica de la matemática, encontramos una amplia literatura en psicología, que describe, en general, el razonamiento covariacional, como una competencia fundamental para la toma de decisiones. El análisis de las estrategias intuitivas 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. Podemos destacar, entre otras, las aportaciones que varios autores han sugerido en cuanto a la enseñanza de los conceptos de regresión y correlación, como son la negociación de significados (Ballman, 1997; McClain y Cobb, 2001), el uso de proyectos de investigación (Groth y Powell, 2004), la inclusión en la enseñanza del AED (Batanero, Estepa y Godino, 1991; Estepa, 1994; Batanero, Godino y Estepa, 1998), la simulación (Ballman, 1997) y la utilización del desarrollo histórico de estas nociones (Estepa y Sánchez Cobo, 1994).

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 onto-semiótico. 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 de nuestra investigación.

CAPITULO 3.

LA CORRELACIÓN Y REGRESIÓN EN LOS LIBROS DE TEXTO

3.1. INTRODUCCIÓN

En este capítulo se analiza la presentación de las nociones de regresión y correlación en dos libros de texto de Matemáticas aplicadas a las Ciencias Sociales I, de primer curso de Bachillerato. Se pretende que sus resultados, junto con el estudio curricular expuesto en el Capítulo 1, y los antecedentes (Capítulo 2) aporten al profesor información para el diseño del significado institucional pretendido. La importancia del análisis de los libros de texto se debe a que nos permiten observar algunos resultados de la transposición didáctica (Chevallard, 1991), es decir de los cambios que experimenta el conocimiento matemático, cuando es adaptado para pasar a ser objeto 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).

Cuando se construye el saber matemático, este se encuentra personalizado y contextualizado por las situaciones y problemas que se han estudiado, pero el productor del saber, al comunicarlo, lo despersonaliza y lo descontextualiza, quedando oculto este proceso de generación. Éste se comunica de manera limpia, secuenciada y a veces, las transformaciones que sufre, pueden hacer que pierda su identidad (Chevallard, 1991).

3.2. METODOLOGÍA DEL ANÁLISIS

Se trata de una investigación cualitativa, pues según Kirk y Miller (1986), en ella se incluyen *"la inducción analítica, el análisis de contenido, la semiótica y ciertas manipulaciones de archivos, informáticas y estadísticas"* (pg. 10). Se mantiene una concepción global fenomenológica, que trata de identificar la naturaleza profunda de las realidades (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.

La muestra utilizada (Tabla 3.2.1) es intencional, por tanto no se aspira a generalizar sino que se busca la comparabilidad y traducibilidad (Goetz y Lecompte, 1998), con lo que la responsabilidad de la generalización 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 han de estar definidas 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.

Tabla 3.2.1. Libros de texto utilizados en el análisis

Código	Referencia
T1	Colera, J., Oliveira, M.J., García, R. y Santaella, E. (2008). Matemáticas aplicadas a las Ciencias Sociales I. Madrid: Grupo Anaya.
T2	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.

Se ha realizado un 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). 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. El "enfoque ontosemiótico" nos proporciona además una categorización de entidades matemáticas que aportan orientación y concreción al análisis de contenido. Seguimos el mismo método utilizado en la investigación de Cobo (2003), el cual consiste en los siguientes pasos:

1. Seleccionados los libros, y el capítulo 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 elementos de significado 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.
4. Elaboración de tablas que resuman los resultados y permitan 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 análisis matemático llevado a cabo en el Anexo 2. Con estos elementos podemos determinar el significado institucional que servirá posteriormente, junto con el análisis de las investigaciones previas (Capítulo 2), para construir un cuestionario de evaluación del conocimiento matemático para la enseñanza de la correlación y regresión en futuros profesores.

3.3. SITUACIONES-PROBLEMAS

Como se expone en Godino (2002), los objetos matemáticos emergen para dar solución a problemas compartidos en el seno de instituciones específicas. Por tanto, el primer elemento analizado son las situaciones que dotan de sentido a estos objetos matemáticos y que se pueden presentar, bien como ejercicios o como ejemplos a lo largo del texto. En su estudio, Sánchez Cobo (1998) estudia el contenido matemático, centrándose en los aspectos procedimentales (construcción de la tabla de frecuencias, cálculo distribuciones marginales o condicionales, cálculo de momentos, cálculo del coeficiente de correlación, representación gráfica y cálculo de los coeficientes de regresión). No clasifica específicamente los campos de problemas, por lo que este será un punto original de nuestro trabajo.

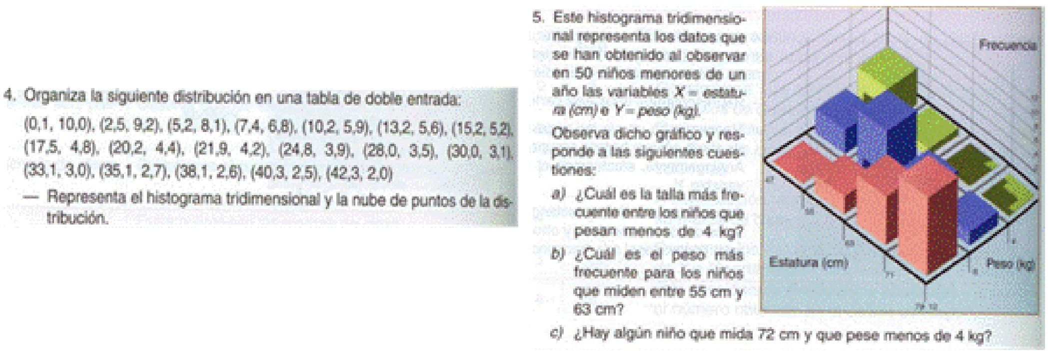
Al analizar el significado de la correlación y regresión en nuestro estudio (Anexo 2), se indicó que, al estudiar la posible relación entre dos variables estadísticas, las principales preguntas del investigador son las siguientes: ¿Hay alguna relación entre las variables? ¿Es intensa o moderada? ¿Directa o inversa? ¿Puedo usar una variable para predecir la otra? (Batanero, Díaz y Gea, 2011).

De las anteriores preguntas se derivan los campos de problemas que se proponen en los libros de texto, aunque se suele facilitar la solución o guiarla, de modo, que, generalmente, estos problemas se convierten en ejercicios. Nosotros no hemos diferenciado entre problemas y ejercicios, pues al ser nueva la situación para el alumno, puede para él, constituir un verdadero problema. Se añade el que denominamos *P0*, relacionado con la organización de los datos. A continuación se describen, haciendo finalmente alusión a la diversidad de contextos expuestos en ellos.

Situaciones-problemas

P0. Organización de datos bidimensionales y representación en el registro gráfico y tabular. Este campo de problemas incluye ejercicios o ejemplos que conducen al alumno a representar los datos, así como a la lectura de dichas representaciones (Figura 3.3.1). La organización y representación de datos es un primer requisito para el estudio de la correlación y constituye en sí mismo una situación problemática, ya que hay más de una solución posible. Esta categoría se puede relacionar con los contenidos "construir la tabla de frecuencias de la distribución bidimensional" y "representación gráfica del diagrama de dispersión" de Sánchez Cobo (1998), quien no considera otros tipos de representaciones tenidas en cuenta por nosotros.

En los textos analizados hemos advertido una intención de ejercitación del alumno en la representación de los datos, desde el punto de vista gráfico y tabular. Aunque el texto [T1] describe a grandes rasgos cómo es una tabla de doble entrada, así como su representación en un diagrama de dispersión, el texto [T2] se detiene mucho más en este aspecto, destacamos, a modo de ejemplo, en la Figura 3.3.1 esta intencionalidad.


4. Organiza la siguiente distribución en una tabla de doble entrada:
(0,1, 10,0), (2,5, 9,2), (5,2, 8,1), (7,4, 6,8), (10,2, 5,9), (13,2, 5,6), (15,2, 5,2), (17,5, 4,8), (20,2, 4,4), (21,9, 4,2), (24,8, 3,9), (28,0, 3,5), (30,0, 3,1), (33,1, 3,0), (35,1, 2,7), (38,1, 2,6), (40,3, 2,5), (42,3, 2,0)
— Representa el histograma tridimensional y la nube de puntos de la distribución.

5. Este histograma tridimensional representa los datos que se han obtenido al observar en 50 niños menores de un año las variables *X* = *estatura (cm)* e *Y* = *peso (kg)*.
Observa dicho gráfico y responde a las siguientes cuestiones:
a) ¿Cuál es la talla más frecuente entre los niños que pesan menos de 4 kg?
b) ¿Cuál es el peso más frecuente para los niños que miden entre 55 cm y 63 cm?
c) ¿Hay algún niño que mida 72 cm y que pese menos de 4 kg?

Figura 3.3.1. Ejercicios de representación tabular y gráfica de datos bidimensionales ([T2], p. 220)

P1. Analizar la existencia de relación entre variables. Como se expuso en el estudio histórico, la correlación y regresión surgen del trabajo de Galton para tratar de medir la relación entre características físicas de sucesivas generaciones de la misma especie, problema que se extiende poco a poco al análisis de relación entre diversas variables estadísticas (Hald, 1998). Este campo de problemas aparece con frecuencia en los textos analizados, fue también considerado por Lavalle y cols. (2006) y podemos asimilarlo al tipo de tarea “interpretación” en Sánchez Cobo (1998). Las tareas que principalmente se plantean siguen un propósito común y es que una vez organizados los datos y representados (gráfica o tabularmente), en algunos casos se acompañe de una reflexión sobre la dependencia funcional o estadística de dichos datos, enfocado al estudio de la dependencia lineal entre las variables, así como sobre la intensidad de ésta. Podemos subdividir este problema en los que se exponen a continuación.

P1.1. Definir las variables que aparecen en un estudio estadístico bidimensional. Este problema fue considerado en el estudio de Lavalle y cols. (2006) pero no por Sánchez Cobo (1998). En el texto [T2] se propone una colección de ejercicios en que se describe un supuesto estudio y se pide a los estudiantes indicar cuáles son las variables estadísticas consideradas. Por ejemplo, “*Duración e importe de las llamadas telefónicas urbanas efectuadas en una ciudad durante 24 horas*” ([T2], p. 216).

P1.2. Deducir la existencia de una dependencia funcional o estadística. Otro problema es estudiar la existencia de una dependencia funcional o aleatoria, es decir, diferenciar si existe o no relación entre las variables (o bien hay independencia) y en caso de relación, decidir si a cada valor de la variable independientes corresponde un único valor de la variable dependiente (relación determinista o funcional) o varios (relación estadística o aleatoria). Por ejemplo, en el texto [T1] encontramos un problema donde se pide, además de distinguir las variables de estudio, su posible dependencia, y que se reflexione sobre si conforman una distribución bidimensional ([T1], p.234).

P1.3. Determinar la intensidad de la relación entre variables. Un segundo paso, una vez detectada la relación, es estudiar su intensidad, que variará desde el caso de independencia hasta la dependencia funcional. El objeto matemático que permitirá deducir esta intensidad puede ser el diagrama de dispersión, y de forma más precisa, la covarianza y el coeficiente de correlación lineal (Batanero y Díaz, 2008).

Las tareas planteadas sobre este punto se orientan preferentemente en determinar la intensidad de la correlación a partir del coeficiente de correlación y se expone también un procedimiento detallado para calcular el coeficiente de correlación mediante una calculadora. En ocasiones se trata de estimar dicho coeficiente a través de varios registros como el gráfico o tabular (Figura 3.3.2).

10. Considera la distribución de la siguiente tabla:

a) Dibuja el diagrama de dispersión de la distribución y describe el grado, el sentido y el tipo de la correlación que se observa.

b) Calcula el coeficiente de Pearson.

X	–2	–2	2	2	2	–2
Y	2	0	2	–2	0	–2
X	–1	1	0	1	0	–1
Y	1	–1	–2	1	2	–1

c) Relaciona los resultados obtenidos en los apartados anteriores.

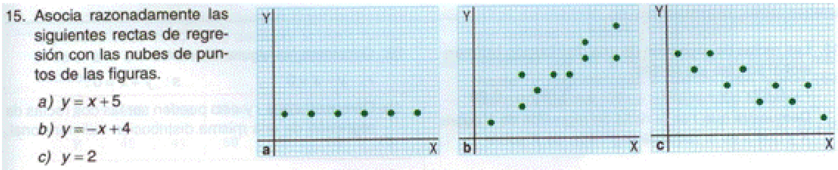
Figura 3.3.2. Ejercicio de estimación de la intensidad de la dependencia estadística ([T2], p. 225)

P1.4. Determinar la dirección de la relación entre variables. Un tercer problema, para el caso de relación lineal consistiría en estudiar el signo, determinando si la relación es directa (al crecer una variable crece la otra) o inversa (al crecer una la otra disminuye). Hacemos notar que, en caso de relación curvilínea (exponencial o cuadrática, por ejemplo), este problema no tendría sentido.

P2. Predecir una variable en función de otra. Una vez detectada la existencia de una relación entre dos variables (problema de correlación), interesa encontrar alguna ecuación que nos permita obtener una de las variables en función de la otra (problema de regresión), problema que también aparece en los libros de texto y se puede asimilar a la categoría "predicción" del tipo de tarea de Sánchez Cobo (1998). En los textos analizados se plantean estos ejercicios. Se expresa la ventaja de disponer de las rectas de regresión para obtener estimaciones, mediante el ajuste lineal a la distribución bidimensional, se introduce la nomenclatura y propiedades de las rectas de regresión y los coeficientes de regresión. También este problema se suele descomponer en varios subproblemas en los libros de texto en la forma que sigue.

P2.1. Analizar el ajuste lineal entre variables. En primer lugar es necesario decidir si una recta sería un buen modelo matemático para describir los datos, punto que sólo se considera el texto [T2], donde además se proponen ejercicios al respecto. En ambos textos, las tareas que inicialmente se plantean piden trazar a ojo las rectas, o representarlas una vez calculadas. Se persigue generar una idea intuitiva del trazo de la recta como "mejor" ajuste a los datos del diagrama de dispersión, sin detenerse demasiado en la deducción matemática de dicha recta, dado que aplicar el criterio de mínimos cuadrados requeriría del uso de derivadas parciales, que los estudiantes de este nivel educativo no conocen. Sánchez Cobo (1998) no menciona esta actividad.

Los libros analizados presentan diferentes casos con correlación positiva y negativa, más o menos intensa, así como correlación nula. Aún así, destacamos algunas diferencias: el texto [T2] presenta ejercicios de asignación de una recta de regresión a un diagrama de dispersión (Figura 3.3.3), que no se hace en el texto [T1]; y existe una mayor cantidad de ejercicios puramente algebraicos (Figura 3.3.4) en el texto [T2] (2 ejercicios resueltos y 8 propuestos) que en el texto [T1] (2 ejercicios).

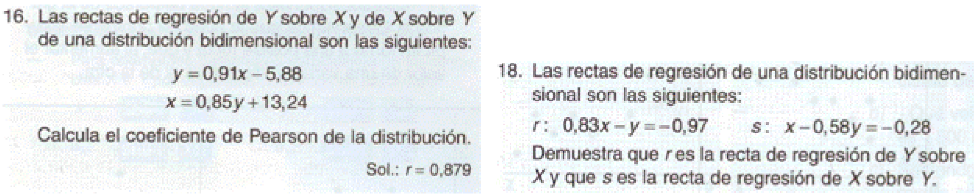

15. Asocia razonadamente las siguientes rectas de regresión con las nubes de puntos de las figuras.

a) $y = x + 5$

b) $y = -x + 4$

c) $y = 2$

Figura 3.3.3. Ejercicio de análisis de regresión lineal, ([T2], p. 229)


16. Las rectas de regresión de Y sobre X y de X sobre Y de una distribución bidimensional son las siguientes:

$$y = 0{,}91x - 5{,}88$$
$$x = 0{,}85y + 13{,}24$$

Calcula el coeficiente de Pearson de la distribución.

Sol.: $r = 0{,}879$

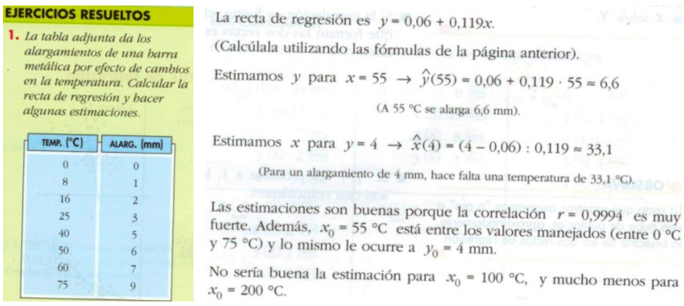
18. Las rectas de regresión de una distribución bidimensional son las siguientes:

r: $0{,}83x - y = -0{,}97$ s: $x - 0{,}58y = -0{,}28$

Demuestra que r es la recta de regresión de Y sobre X y que s es la recta de regresión de X sobre Y.

Figura 3.3.4. Ejercicios de análisis de regresión lineal, ([T2], p. 230)

P2.2. Hacer estimaciones mediante el ajuste lineal entre variables. Una vez calculada la recta de regresión, se plantean ejercicios de estimación (Figura 3.3.5). Al igual que es sugerido en Lavalle y cols. (2006), el cálculo de la estimación lleva asociado una reflexión en torno al valor esperado o promedio y el valor real, ya que puede haber diferentes valores observados ([T1], pp. 231 y 236).


EJERCICIOS RESUELTOS

1. *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.*

TEMP. (°C)	ALARG. (mm)
0	0
8	1
16	2
25	3
40	5
50	6
60	7
75	9

La recta de regresión es $y = 0{,}06 + 0{,}119x$.

(Calcúlala utilizando las fórmulas de la página anterior).

Estimamos y para $x = 55 \rightarrow \hat{y}(55) = 0{,}06 + 0{,}119 \cdot 55 \approx 6{,}6$

(A 55 °C se alarga 6,6 mm).

Estimamos x para $y = 4 \rightarrow \hat{x}(4) = (4 - 0{,}06) : 0{,}119 \approx 33{,}1$

(Para un alargamiento de 4 mm, hace falta una temperatura de 33,1 °C).

Las estimaciones son buenas porque la correlación $r = 0{,}9994$ es muy fuerte. Además, $x_0 = 55$ °C está entre los valores manejados (entre 0 °C y 75 °C) y lo mismo le ocurre a $y_0 = 4$ mm.

No sería buena la estimación para $x_0 = 100$ °C, y mucho menos para $x_0 = 200$ °C.

Figura 3.3.5. Estimación mediante el ajuste lineal ([T1], p. 231)

En la Tabla 3.3.1 se clasifica el modo en que las situaciones problemáticas se presentan al alumno. El porcentaje de ejercicios para resolver es similar; y aunque el primer texto ofrece más ejercicios resueltos que ejemplos, las diferencias son pequeñas. Sánchez Cobo (1998) no realiza esta clasificación en su estudio. En la Tabla 3.3.2 se muestran los resultados del análisis de los campos de problemas en los libros analizados,. Los dos textos (aunque con diferente frecuencia) tratan todos los campos descritos. Cabe advertir una diferencia ya que en el texto [T2] no se incluyen ejemplos o ejercicios resueltos relacionados con la calculadora.

Tabla 3.3.1. Frecuencias y (porcentajes) de categorías de problemas en los libros de texto

	T1		T2	
Ejemplos	21	(17,5)	31	(26,3)
Ejercicios	82	(68,3)	81	(68,6)
Ejercicios resueltos	12	(10)	6	(5)
Ejercicios/Ejemplos resueltos con calculadora	5	(4,2)	0	(0)
	120		118	

Tabla 3.3.2. Campos de problemas en los libros de texto

		T1	T2
P0. Organización de datos bidimensionales en el registro gráfico y tabular.			
	Ejemplos	p. 226, 233	p. 216, 217, 218, 219(3), 220(2),
	Ejemplo con calculadora	p. 233	
	Ejercicios[1]	p.225(2), 227, 239(4), 240	p. 220(3), 225, 228, 232(2), 233(6), 234(4)
P1. Analizar la existencia de relación entre variables.			
P1.1. Definir las variables estadísticas que aparecen en un estudio bidimensional.	Ejemplos		p. 216(2)
	Ejercicios	p. 238	p. 216(2)
P1.2. Deducir la existencia de una dependencia funcional o estadística.	Ejemplos	p.225(3), 226(3),	p. 221(4), 222(3)
	Ejercicios	p. 225(2), 238(2)	p. 221, 232(2), 233(2), 235
	Ejercicios resueltos	p. 234	
P1.3. Determinar la intensidad de la relación entre variables.	Ejemplos	p.226(2), 227	p. 222(2), 225
	Ejercicios	p.227, 229, 238 (5), 239(6), 240(7), 241(5)	p. 223(2), 225(2), 228, 229, 230(2), 232(2), 233(6), 234(3), 235 (3)
	Ejercicios resueltos	p.229, 234, 235, 237	p. 229, 230
	Ejercicios resueltos con calculadora	p.229, 235	
P1.4. Determinar la dirección de la relación entre variables.	Ejemplo	p. 225, 226, 227	p. 223(2), 225,
	Ejercicios	p. 225, 227, 238 (4), 239, 240(3), 241(2)	p. 223(2), 225(2), 229, 233
	Ejercicios resueltos	p. 234	p. 229
P2. Predecir una variable en función de otra.			
P2.1. Analizar el ajuste lineal entre variables	Ejemplos	p.226(3), 227, 232(3)	p. 223(2), 224(4), 227
	Ejercicios	p. 225, 227, 238(2), 239(6),240(6), 241(4)	p. 223(2), 228(3), 229, 230(2), 232(6), 234, 235
	Ejercicios resueltos	p.230, 231, 235, 236	p. 229, 230
	Ejercicios resueltos con calculadora	p.230, 235	
P2.2. Hacer estimaciones mediante el ajuste lineal entre variables.	Ejemplos		p. 227
	Ejercicios	p. 239(4), 240(4), 241(4)	p. 228(3), 231, 233(4), 234(3), 235
	Ejercicios resueltos	p.231, 236	p. 231

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 es permitir finalmente la recontextualización del saber.

Con objeto de analizar este punto, también tenido en cuenta por Sánchez Cobo (1998), se han estudiado los contextos de aplicación, encontrando una gran variedad, tanto para el desarrollo del tema como para el planteamiento de ejercicios. Los hemos clasificado en seis categorías: fenómenos biológicos (como la estatura de hijos – estatura de padres, precisamente el problema que históricamente dio origen a la idea de regresión), educativos (notas de exámenes: física-matemáticas; matemáticas-filosofía),

[1] Ejercicio = ejercicio, problema o cuestión con fines de ejercitación.

deportivos (distancia del jugador – número de encestes al jugar al baloncesto), experimentación en ciencias (aumento de peso de un animal – mg diarios de un fármaco; alargamiento de una barra metálica – temperatura a la que se expone; altura que alcanza una piedra y la fuerza con que se lanza), demografía-sociología (renta per cápita – índice de natalidad) y economía (consumo de energía per cápita y renta per cápita; kg de capturas de pescado y precio de subasta en la lonja). Hemos encontrado también una séptima categoría "descontextualizados".

Tabla 3.3.3. Contextos encontrados en el texto T1

Contextos	Ejemplos	Ejercicios en el desarrollo del tema	Ejercicios resueltos	Ejercicios propuestos al finalizar el tema
Biológico	p.225(1)	p.225(2)		p.238(2), 239(1), 240(1)
Ciencias	p. 225(1), 227(1)	p.225(3)	p. 231(1), 234(1)	p.238(2), 239(2),
Demografía-sociología		p.225(1), 227(1)		p.239(1), 241 (1)
Deportes	p.226(1)	p.225(2), 229(1)	p.237(1)	
Economía		p.225(3)	p.235(2), p.236(1)	p.238(3), 240(3), 241 (2)
Educativos	p.226(2)	p.229(1)	p.229(1), 230(1), p.234(1)	
Descontextualizados			p.233(1)	p.238(4), 239(4), 240(6), 241 (3)

Tabla 3.3.4. Contextos encontrados en el texto T2

Contextos	Ejemplos	Ejercicios en el desarrollo del tema	Ejercicios resueltos	Ejercicios propuestos al finalizar el tema
Biológico	p.216(2), 221	p.216, 220, 221(3), 228		p.232
Ciencias		p.228	p.231	p.232(5), 233(3), 234(2)
Demografía-sociología		p.221		p.233
Deportes	p.217, 220	p.216		p.232, 235(2)
Economía	p.219, 221, 225, 227	p.216(3),		p.231, 233(2), 234
Educativos	p.218, 219, 220	p.216(2), 221(2)		
Descontextualizados	p.219, 222(5), 223(4), 224(4),	p.220(2), 223(2), 225(2), 228	p.229(2), 230(2)	p.229(2), 230(4), 232(7), 234, 235(2)

En las tablas 3.3.3 y 3.3.4 presentamos los resultados. La principal diferencia es que el número de ejercicios descontextualizado es mayor en el [T2] (44,7%), mientras en el [T1] es el 27,6%. En el estudio de Sánchez Cobo (1998), el 31,7% de los ejercicios estaba descontextualizado.

3.4. DEFINICIONES (CONCEPTOS)

De acuerdo a Godino (2003), las *definiciones de conceptos* son evocadas por el alumno cuando realiza cualquier acción para resolver las cuestiones planteadas. Es por ello importante tener en cuenta los conceptos que permiten caracterizar el objeto de estudio en el sistema de prácticas planteadas. Sánchez Cobo (1998) no diferencia entre conceptos y propiedades, sino que engloba las dos categorías como "contenidos". Antes de exponer nuestros resultados, indicamos que el texto [T2] introduce la unidad recordando definiciones básicas como: estadística, población, individuo, variable estadística (cualitativa, cuantitativa, continua, discreta) y sumatorio.

Variable estadística bidimensional

El trabajo con la correlación y regresión parte de la definición de variable estadística bidimensional, pieza angular de todo estudio bivariado. Sánchez Cobo indica que algunos textos presentan dos variables estadísticas unidimensionales separadamente, mientras otros incluyen la variable estadística bidimensional pero no lo estudia específicamente como contenido. En nuestro estudio, el texto [T1] prescinde de esta noción, y tan sólo incluye el uso de la distribución bidimensional desde un enfoque funcional, con fin utilitario. Por el contrario, el texto [T2] expone la finalidad de un estudio bidimensional de modo claro y conciso al comienzo del tema y ofrece una definición más formal (Figura 3.4.1):

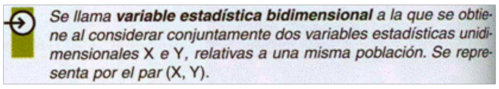

*Se llama **variable estadística bidimensional** a la que se obtiene al considerar conjuntamente dos variables estadísticas unidimensionales X e Y, relativas a una misma población. Se representa por el par (X, Y).*

Figura 3.4.1. Definición de variables estadística bidimensional, ([T2], p. 216)

Tabla de doble entrada

Sánchez Cobo indica que estas tablas se presentan asociadas a la definición de la variable estadística bidimensional y su tratamiento es diferente en los libros analizados. El texto [T1] da una definición de la tabla de doble entrada y da sentido a las frecuencias de ésta al finalizar el tema. Se presentan sólo cuatro (1 ejemplo a modo de definición, 1 ejercicio resuelto y 2 ejercicios propuestos) de un total de 24 tablas expuestas en el tema (4 ejemplos, 6 ejercicios resueltos y 14 ejercicios propuestos). Mayoritariamente se presentan tablas donde cada una de las variables que conforman la variable estadística bidimensional aparecen en una columna o fila y sólo el 17% de las tablas en el tema son propiamente de doble entrada.

En el texto [T2] se presentan 5 tablas de doble entrada (2 ejemplos de construcción y 3 ejercicios propuestos) de un total de 20 tablas en el tema (1 a modo de ejemplo, 1 como ejemplo/ejercicio resuelto, 1 ejercicio resuelto y 12 ejercicios propuestos) lo que supone 25% de las tablas. El texto [T2] ofrece un tratamiento más apropiado de la tabla de doble entrada como se puede observar en el planteamiento de ejercicios, en la descripción de los procedimientos de las diferentes representaciones gráficas (véase la sección 3.7), o en el tratamiento de otros conceptos.

Distribución bidimensional

El texto [T1] define la distribución bidimensional a través de ejemplos, señalando que es el "*conjunto de pares de valores* $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$" (véase el ejemplo presentado en la sección 3.9). Por el contrario, el texto [T2] realiza una definición más completa, indicando que se refiere al "*conjunto de todos los datos procedentes de la observación de una variable estadística bidimensional*" ([T2], p.216).

Frecuencia marginal

Sánchez Cobo (1998) indica que este concepto se asocia a la distribución bidimensional, aunque en nuestro análisis sólo es tratado en el texto [T2] en la construcción de la tabla de doble entrada, indicando que corresponde a la frecuencia absoluta de las celdas de la última fila ó columna de dicha tabla ([T2], p.217).

Distribución marginal

El texto [T1] usa esta noción sin una definición previa, a modo de etiquetado en una tabla de frecuencias, potenciando el conflicto semiótico que se destacará en la sección 3.9. Por su parte, el texto [T2] la define al margen y como complemento a la definición de frecuencia marginal del siguiente modo: "*Cuando se estudian por separado las variables unidimensionales X e Y que forman la variable bidimensional (X,Y), se habla de distribuciones marginales.*" ([T2], p. 217). Sánchez Cobo indica que este concepto se asocia a la distribución bidimensional para mostrar la forma en que se puede deducir aquella, de variables estadísticas unidimensionales.

Dependencia funcional / aleatoria o estadística/ independencia

De acuerdo a Sánchez Cobo (1998) este es un concepto fundamental, que permite integrar la correlación y regresión. Al contrario que en su estudio, los textos analizados describen la relación funcional, diferenciándola de la aleatoria o estadística. En el texto [T1], y a partir de un ejemplo, se asemeja la relación funcional a la existencia de una fórmula de cálculo exacto de la variable dependiente y el caso contrario a la estadística:

> "Si lanzamos una piedra hacia arriba, llegará tanto más alto cuanto más fuerte la lancemos. Y hay una fórmula que nos permite calcular, exactamente, la *altura* alcanzada en función de la *velocidad* con que se lanza. Es una *relación funcional.*
> Las personas, en general, pesan más cuanto más altas son. Pero no se podría dar una fórmula que permitiera obtener el peso de cualquier persona conociendo su estatura. La relación entre las variables *estatura-peso* es estadística. Se dice que hay una correlación entre ellas." ([T1], p.225).

En el texto [T2] se da una definición más formal (Figuras 3.4.2 y 3.4.3), haciéndose mención explícita a la noción de independencia de variables estadísticas (véase la Sección 3.9, donde se hace también referencia a la noción *incorreladas*).

*Entre dos variables estadísticas existe **dependencia funcional** si están relacionadas de forma que sea posible determinar con exactitud los valores que toma una de ellas a partir de los que toma la otra.*

*Entre dos variables estadísticas existe **dependencia estadística** o **correlación** cuando los valores que toma una de ellas están relacionados con los valores que toma la otra, pero no de manera exacta.*

Figura 3.4.2. Definición de dependencia funcional y estadística ([T2], p. 221)

*Dos variables estadísticas son **independientes** si no puede establecerse ninguna relación entre los valores que toma una de ellas y los que toma la otra.*

Figura 3.4.3. Definición de independencia ([T2], p. 221)

Coeficiente de correlación de Pearson

El texto [T1] emplea esta noción sin definirla (generando el conflicto que se destacará en la sección 3.9), no especifica su origen y utiliza la denominación "coeficiente de correlación" (Figura 3.6.1). En la Figura 3.4.4 se presenta el modo en que el texto [T1] introduce el coeficiente de correlación a través de su fórmula, a partir la cual será sencillo posteriormente razonar que el signo de la correlación y la covarianza son iguales, puesto que las desviaciones típicas siempre tienen signo positivo. El texto [T2] se detiene más en precisar la diferencia de medir cuantitativamente la correlación entre dos variables estadísticas y el caso en que únicamente se considere la correlación lineal.

RECUERDA

La desviación típica de una variable x se calcula así:

$$\sigma_x = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2}$$

Correlación

El valor de la **correlación** entre las dos variables de una distribución bidimensional viene dado por la expresión:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

σ_{xy} es la covarianza

σ_x, σ_y son las desviaciones típicas de cada variable

Figura 3.4.4. Fórmula del índice de correlación, ([T1], p. 228)

Correlación

Esta noción se incluye en todos los textos del estudio de Sánchez Cobo (1998), así como en los textos que hemos analizado. En nuestro análisis observamos que la correlación es utilizada como sinónimo de dependencia estadística, sin diferenciar que las variables sean o no numéricas. Sin embargo, desde el punto de vista matemático, se precisa que el término "correlación" se aplica sólo a variables cuantitativas. Sólo el texto [T2] hace mención a la *correlación espúrea,* a modo de anotación al margen del texto:

> "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.” ([T2], p.221)

Covarianza

La covarianza proporciona un primer coeficiente de asociación para variables cuantitativas y es presentado en los dos textos analizados de modo formal, al igual que en la mayoría de textos del estudio de Sánchez Cobo (1998):

“La covarianza es la media aritmética de los productos de las desviaciones de cada par de datos respecto de sus medias” ([T2], p.235)

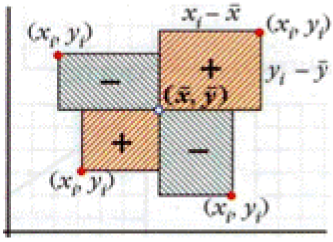


Figura 3.4.5. Posición relativa de los datos bidimensionales a las medias de cada variable unidimensional que dividen el eje cartesiano en cuatro cuadrantes ([T1], p. 228)

Desde nuestro punto de vista, es necesario ofrecer al alumnado una fórmula clara, además de exponer todas las expresiones equivalentes, al menos aquellas que se pretendan utilizar a lo largo del texto (en la sección 3.9 se señalará un conflicto semiótico relacionado con este aspecto). Destacamos el tratamiento que el texto [T1] ofrece a la construcción de esta noción, mediante la división en cuatro cuadrantes de la nube de puntos por las rectas correspondientes a las medias de cada variable (Figura 3.4.5), ya que permite al alumnado desarrollar una comprensión más significativa. 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 Anexo 2).

Regresión

Además de la definición se introducen la recta y sus coeficientes. A diferencia de los textos de Sánchez Cobo (1998), en el texto [T2] se define la noción de regresión y se precisa con ello la finalidad del estudio de la regresión lineal:

“Al análisis que pretende determinar la curva que mejor aproxima un diagrama de dispersión se le llama *regresión*. En este libro estudiaremos el caso de la regresión lineal, es decir, la determinación de la *recta* que mejor aproxima una nube de puntos.” ([T2], p. 226)

Recta de regresión. Sánchez Cobo (1998) indica que en su estudio, el apartado sobre la regresión se conforma principalmente alrededor de la determinación de las dos rectas. En nuestro caso, los libros analizados describen la recta de regresión de una manera práctica, utilizando gráficos en los que se representan conjuntos de datos bivariantes, con ejemplos que muestran la distancia de los puntos de la nube a la recta (Figura 3.4.6) así como la tendencia de la variación conjunta. Se destaca la diferencia entre el valor observado o dato real (ordenadas del punto) y el predicho o dato teórico (punto sobre la recta). Se concluye mostrando la definición de la recta de regresión de modo formal, y su justificación a través del método de los mínimos cuadrados.

Coeficiente de regresión / pendiente de la recta de regresión. En los textos analizados se presenta la definición de la pendiente de la recta de regresión mediante su fórmula $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}$. En el caso del texto [T2] pasa muy desapercibida, como una anotación al margen con el fin de remarcar la propiedad de los productos de las pendientes, pero en el texto [T1] sí ocupa un lugar relevante. Hacemos notar la imprecisión en dicha definición, dado que si disponemos de dos rectas de regresión (*Y* sobre *X* y *X* sobre *Y*) el texto define a la pendiente de la recta de regresión de *X* sobre *Y* como coeficiente de regresión de *X* sobre *Y* mientras que no guarda el mismo cuidado a la hora de definir la pendiente de la recta de regresión de *Y* sobre *X* denominándolo tan sólo como coeficiente de regresión. En el estudio de Sánchez Cobo (1998) el coeficiente de regresión sólo se incluye en la cuarta parte de los libros analizados.

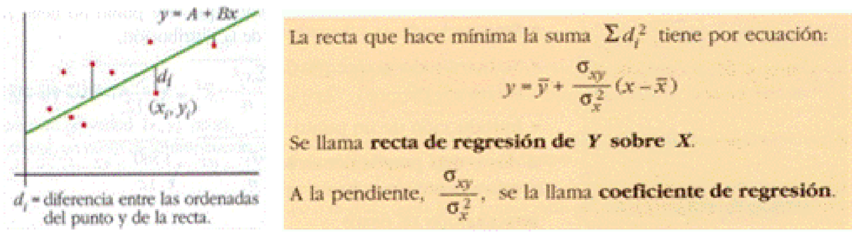


Figura 3.4.6. Fórmula de la recta de regresión de *Y* sobre *X* ([T1], p. 230)

Centro de gravedad de la distribución bidimensional

La recta de regresión tiene la propiedad de pasar por el punto cuyas coordenadas son las medias de las dos variables y que se es el centro de gravedad de la distribución. En los textos analizados esta noción aparece como anotación al margen, y el texto [T2] lo denomina *punto medio de la distribución*. Sánchez Cobo no analiza este concepto.

Diagrama de barras tridimensional

El texto [T1] realiza una aproximación a este gráfico desde un tratamiento impreciso que se relatará en la sección 3.9. En el texto [T2] se define propiamente un diagrama de barras tridimensional como un gráfico tridimensional utilizado para representar datos bidimensionales no agrupados en intervalos donde para cada dato se levanta una barra de altura proporcional a su frecuencia absoluta (Figura 3.7.5). Sánchez Cobo (1998) no hace referencia a este diagrama.

Pictograma tridimensional

Sólo el texto [T2] define este gráfico, que no aparece en el estudio de Sánchez Cobo (1998) y es tratado como una variante del diagrama de barras donde cada barra es sustituida por dibujos que representen las variables de estudio (Figura 3.7.6).

Histograma tridimensional

Con esta noción ocurre algo parecido que con el diagrama de barras tridimensional, tampoco descrito por Sánchez Cobo (1998). El texto [T1] realiza una aproximación a esta noción desde un tratamiento impreciso que se relatará en la sección 3.9. El texto [T2] define propiamente un histograma tridimensional como un gráfico tridimensional utilizado para representar datos bidimensionales agrupados en intervalos donde para cada par de intervalos de clase se levanta un prisma de volumen proporcional a su frecuencia absoluta (Figura 3.7.7).

Nube de puntos / Diagrama de dispersión

La nube de puntos o diagrama de dispersión es la representación gráfica más utilizada en el desarrollo de los temas de los textos analizados, siendo la representación en coordenadas cartesianas de una variable estadística bidimensional (Sánchez Cobo, 1998). El texto [T1] (p. 226) la introduce con ejemplos en que se distingue la intensidad y la direccionalidad de la dependencia entre las variables y se define como el conjunto de datos de una distribución bidimensional representados en un eje cartesiano. Esta definición se completa al final del tema cuando se introduce la tabla de doble entrada, donde se advierte que la frecuencia absoluta de cada dato no tiene que ser necesariamente uno. Así, la representación de cada dato debe “hincharse” proporcionalmente a su frecuencia absoluta (Figura 3.9.3). La definición que ofrece el texto [T2] es más concisa dado que su construcción parte directamente de una tabla de doble entrada en que las frecuencias absolutas de los datos no necesariamente son la unidad. Señalamos que en los textos analizados no se hace mención a los diagramas de burbuja por considerarse una extensión de los diagramas de dispersión ([T1]) o bien como diagrama de dispersión propiamente dicho ([T2]) (Figura 3.7.8).

En la Tabla 3.4.1 se presentan los conceptos incluidos en los libros analizados, observándose unas definiciones más precisas y completas en el texto [T2], mientras que el [T1] suele usar los conceptos a partir de ejemplos Se observa alguna confusión, que se detallará en el análisis de conflictos semióticos (sección 3.9).

Tabla 3.4.1. Conceptos de los libros analizados

Conceptos	T1	T2
Variable estadística bidimensional	Usa sin definición previa	Define desde el comienzo
Tabla de doble entrada	Presencia anecdótica; uso preferente tabla dos columnas /filas	Mayoría de tablas de doble entrada
Distribución bidimensional	A través de ejemplos; posteriormente como conjunto de puntos	Definición más precisa
Frecuencia marginal	A partir de un ejemplo	No define
Distribución marginal	Usa sin definición previa	Complemento de la definición de frecuencia marginal
Dep. funcional / estadística / independencia	Mediante ejemplo	Más formal Define independencia
Correlación	Sinónimo de dependencia y mediante fórmula	Define incorreladas Correlación espúrea
Coeficiente de correlación	Usa sin definir No especifica su origen	Diferencia la correlación lineal Referencia al origen histórico
Covarianza	Formalmente Mediante división del plano en cuadrantes	Formalmente
Regresión	Recta de regresión mediante ejemplos y fórmula. Coeficiente de regresión mediante fórmula y de modo impreciso	Recta de regresión mediante fórmula Se precisa la finalidad Coeficiente de regresión como anotación al margen.
Centro gravedad	Anotación al margen	Anotación al margen
Diagrama barras tridimensional	Definición imprecisa	Define correctamente
Pictograma tridimensional		Variante del diagrama de barras
Histograma tridimensional	Definición imprecisa	Define correctamente
Diagrama de dispersión	A través de ejemplos; diferencia intensidad y signo Incluye el diagrama de burbujas	Más concisa Incluye el diagrama de burbujas

3.5. LENGUAJE

Otro elemento de significado que analizamos es el lenguaje. Los estudiantes se deben familiarizar con él dado que sirve para enunciar las definiciones y propiedades de dicho objeto, así como para representar los problemas y datos.

Términos

Encontramos una variedad de términos asociados a la correlación y regresión, que hemos dividido en dos grupos; por un lado los *términos básicos* que debe conocer el alumno al iniciar el tema y por otro lado los *términos específicos* del tema. Además, hemos considerado una clasificación de términos coloquiales utilizados en los textos, que aluden a términos específicos y que se utilizan como sinónimos de aquellos. Presentamos a continuación los términos que hemos encontrado.

Términos básicos: amplitud de intervalo, ángulo entre dos rectas, área de un rectángulo, área de un punto, bisectriz de dos rectas, coordenadas de un punto, dato, datos no agrupados, datos agrupados, desviación típica, distancia, distribución, ecuación, ejes cartesianos, estimación, fiabilidad, frecuencia, frecuencia absoluta, individuo, intervalo de clase, marca de clase, máximo, media aritmética, método de reducción al absurdo, mínimo, muestra, ordenada de un punto, ordenada en el origen, parámetro, pendiente de una recta, población, prisma, probabilidad, proporcionalidad, recta, rectas coincidentes, rectas perpendiculares, subíndice, sumatorio, tabla de datos, tabla de frecuencias, tendencia, valor absoluto, valor de la variable, variable estadística, variable cualitativa, variable cuantitativa, variable cuantitativa continua, variable cuantitativa discreta, variable unidimensional, varianza, volumen.

Términos específicos: centro de gravedad/punto medio de la distribución bidimensional; coeficiente de correlación de Pearson; coeficiente de regresión; correlación; correlación espuria; correlación curvilínea; correlación fuerte, débil o nula; correlación lineal; correlación perfecta; correlación positiva o negativa; covarianza; diagrama de dispersión; diagrama de barras tridimensional; distribución bidimensional; distribución marginal; frecuencia marginal; histograma tridimensional; incorrelada; independencia; método de mínimos cuadrados; nube de puntos; pictograma tridimensional; recta de regresión; recta de regresión de Y sobre X; recta de regresión de X sobre Y; regresión; regresión lineal; relación/dependencia estadística; relación/dependencia funcional; tabla de doble entrada; valor esperado/predicción; variación conjunta; variable bidimensional.

Términos coloquiales con significado matemático. En la Tabla 3.5.1 se presentan los utilizados en los textos, junto al término matemático al que hacen alusión.

Notación simbólica

Las notaciones simbólicas se utilizan para referirse a conceptos o propiedades. Tal es su relevancia en la matemática, que el simbolismo permite una comunicación comprimida entre individuos pudiendo trabajar a un alto nivel de complejidad dando significado a los símbolos. En la Tabla 3 (Anexo 3) exponemos la simbología utilizada en los libros analizados. Al igual que Ortiz (1999) hemos encontrado notación funcional y el uso de subíndices y superíndices, que con frecuencia son variables.

Tabla 3.5.1. Términos coloquiales con significado matemático en los textos

T1	Término coloquial	Término matemático al que alude
	Estatura *normalita* ([T1], p.225)	Estatura *media*
	Grosso modo ([T1], p. 226 y p. 231)	Relación estadística y *correlación*
	Según lo *apretados* que estén los puntos ([T1], p. 227) Grado de "apretura" ([T1], p. 228)	*Dispersión*
	A ojo ([T1], p. 230, p. 232, p. 238)	Por aproximación o *ajuste*
	Rectas que "se *acoplan* bien" a la nube de punto ([T1], p. 230)	*Ajuste* lineal a la nube de puntos casi perfecto
	La recta de regresión se *amolda* a la nube de puntos ([T1], p. 231)	*Ajuste* lineal a la nube de puntos
	Hinchar los puntos proporcionalmente a su frecuencia ([T1], p. 233)	Representar circunferencias con *diámetro* proporcional a la frecuencia
	"*Buenas*" estimaciones	Estimaciones *fiables*
T2	Los puntos de la nube (…), están completamente en *desorden*. ([T2], p.222)	Están muy *dispersos*, (o bien) se aprecia gran *dispersión* en los puntos.

Expresiones algebraicas

En los textos analizados se incluyen numerosas expresiones algebraicas, es decir, combinaciones de letras, números y signos de operaciones que permiten traducir al lenguaje matemático expresiones del lenguaje habitual. Las letras suelen representar cantidades desconocidas, que pueden ser variables o incógnitas. Son frecuentes para dar la fórmula de la covarianza, los coeficientes de correlación y regresión o la expresión de la recta de regresión. En la Tabla 4 (Anexo 3) presentamos las más destacadas.

Representación tabular

De acuerdo a Ortiz (1999), las tablas estadísticas se usan con frecuencia para presentar a los alumnos datos empíricos. Ofrecen una estructuración particular del espacio, una estructura soporte de relaciones, presentando una parrilla no de números en sí mismos, sino de las relaciones entre las diferentes entradas en la tabla. En los libros analizados la representación tabular más utilizada es aquella en la que cada fila/columna de la tabla representa los datos relativos a cada uno de los individuos de la muestra. Además es común, avanzado el tema, que se vayan añadiendo columnas según la necesidad de los cálculos (cálculo de la covarianza, correlación lineal, etc.).

En cuanto a las tablas de doble entrada, el [T1] hace una breve descripción de qué son al finalizar el tema (ver Figura 3.7.1), pero no hace una descripción o definición precisa y son pocas las actividades donde el alumno trate con tablas de doble entrada. En concreto, dedica un ejercicio (resuelto) a estas tablas y propone dos ejercicios al final del tema con la etiqueta “PARA PROFUNDIZAR”.

Representaciones gráficas

La representación gráfica alcanza un estatus privilegiado en el estudio de las nociones de correlación y regresión. Tal es su importancia que el modo habitual de definir la correlación es mostrarla mediante un diagrama de dispersión y a partir de él exhibir sus propiedades (intensidad y signo) y el modo habitual de tratar la regresión es mostrarla 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:

Diagramas de dispersión o nube de puntos. Este objeto matemático se definió en la sección 3.4 y es la representación más común, al permitir observar la intensidad de la relación (mayor o menor dispersión de la nube de puntos), si la relación es o no lineal (por medio de su tendencia) así como el signo de la correlación, en caso de relación lineal (ver Figuras 3.6.3 y 3.8.3).

Nube de puntos con recta de regresión añadida: En general, los diagramas de dispersión que se presentan vienen acompañados de la recta de regresión, reflejando la tendencia de variación conjunta de las variables que conforman la distribución bidimensional (ver Figura 3.8.5).

Otros: Con menor frecuencia se presentan diagramas de barras o histogramas tridimensionales, o gráficos de burbuja (ver Figura 3.7.4, 3.7.5 y 3.7.6).

3.6. PROPOSICIONES

Los libros suelen presentar *proposiciones*, que se refieren a *atributos o propiedades;* son los enunciados que se realizan sobre un objeto matemático que marcan las acciones que se realizan sobre él. Regulan las relaciones de este objeto con otros y con ello se produce un enriquecimiento en su significado (Godino, Font, Contreras y Wilhelmi, 2006). A continuación describimos las más importantes:

Organización de datos en representación tabular

La suma de las frecuencias absolutas de una columna es la frecuencia absoluta del valor de X correspondiente a esa columna. Esta proposición junto con su análoga en la variable *Y* permiten al estudiante comprender los conceptos de frecuencia marginal y distribución marginal. Esta proposición se presenta únicamente en el texto [T2] aunque se utilice en el texto [T1] (ver sección 3.9).

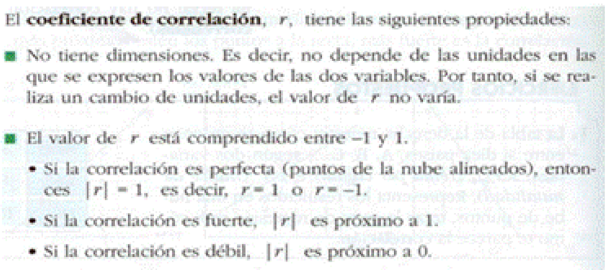
Los extremos de cada intervalo de clase sólo pueden pertenecer a uno de ellos. Esta proposición se presenta únicamente en el texto [T2] en el procedimiento de agrupación de datos en intervalos de clase (Figura 3.7.4).

Distribución bidimensional

El centro de gravedad no tiene por qué ser un punto de la distribución. Esta propiedad de la media aritmética (que no es una operación interna en el conjunto de datos), supone una dificultad para algunos estudiantes (Cobo, 2003) y se encuentra en los libros analizados.

Correlación

Signo y rango de variación de la correlación. Para poder usar correctamente el coeficiente de correlación, se requieren una serie de propiedades señaladas en la mayoría de manuales del estudio de Sánchez Cobo (1998). En nuestro estudio, por medio de ejemplos, los dos textos exponen el hecho de que la relación entre las variables puede ser funcional o estadística ("*Entre los casos extremos de dependencia funcional e independencia existe una amplia gama de situaciones en que se da dependencia estadística o correlación*" ([T2], p.222)); positiva o negativa, atendiendo a si un aumento de una variable implica un menor crecimiento de la otra variable (correlación negativa), y viceversa; así como los casos de independencia y correlación perfecta (véase además la sección 3.9). En ambos textos se ofrece el rango de valores de dicho coeficiente para una intensidad alta o baja de la correlación, pero sólo el texto [T1] ofrece una formalización (Figura 3.6.1).



El **coeficiente de correlación**, r, tiene las siguientes propiedades:

- No tiene dimensiones. Es decir, no depende de las unidades en las que se expresen los valores de las dos variables. Por tanto, si se realiza un cambio de unidades, el valor de r no varía.
- El valor de r está comprendido entre -1 y 1.
  - Si la correlación es perfecta (puntos de la nube alineados), entonces $|r| = 1$, es decir, $r = 1$ o $r = -1$.
  - Si la correlación es fuerte, $|r|$ es próximo a 1.
  - Si la correlación es débil, $|r|$ es próximo a 0.

Figura 3.6.1. Propiedades del coeficiente de correlación ([T1], p. 228)

Intensidad de la correlación. Se suele poner ejemplos donde se enfatiza el hecho de que diversas variables pueden tener correlación más o menos fuerte, siendo anecdótico el texto [T1] cuando indica: "(…) *Puede ser más o menos fuerte según lo apretados que estén los puntos de la nube en torno a una recta*" ([T1], p.227). Por el contrario el tratamiento de esta propiedad es más apropiado en el texto [T2] indicándose:

> "Si tenemos en cuenta únicamente el caso de la correlación lineal, se suele considerar el llamado coeficiente de Pearson.

Para llegar a la expresión de este coeficiente debemos definir antes un nuevo parámetro estadístico, llamado covarianza," ([T2], p.224).

Adimensionalidad del coeficiente de correlación. Sánchez Cobo (1998) 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 (Figura 3.6.1) algo que señala sólo el texto [T1].

Regresión

La recta de regresión es aquella que hace mínima la suma de los cuadrados de las distancias de los puntos de la distribución bidimensional. Una vez avanzado el tema, y tras definir la covarianza y la correlación, solemos encontrar la propiedad de que la recta de regresión minimiza las distancias de dicha recta a la nube de puntos. En cierto modo esta propiedad generaliza la equivalente de la media (la media es el valor que minimiza la suma de distancias de los datos), puesto que, en realidad, para cada valor de la variable independiente la recta de regresión proporciona el valor medio teórico de los valores de la variable dependiente. En ambos textos, y por medio del soporte gráfico, se relaciona la distancia de cada punto de la distribución bidimensional a una recta cualquiera (Figura 3.6.3), evidenciándose el interés de hacer mínima esta distancia, señalando el criterio de mínimos cuadrados.

Las estimaciones obtenidas con la recta de regresión son aproximaciones. Ello es debido a que, al contrario del caso de la dependencia funcional, para cada valor de la variable independiente, en la dependencia aleatoria corresponden varios valores de la dependiente, y el valor proporcionado por la recta de regresión es el promedio de todos ellos. Señalamos la advertencia que al respecto realiza el texto [T1] donde se indica que:

"Las estimaciones siempre se realizan aproximadamente y en términos de probabilidad: es probable que si $x = x_0$ entonces *y* valga, aproximadamente, $\hat{y}(x_0)$ " ([T1], P.230).

Las estimaciones llevadas a cabo con la recta de regresión solo deben hacerse dentro del intervalo de valores utilizados o muy cerca de ellos. La aplicación más directa de las nociones de correlación y regresión es la predicción de una variable respecto de otra. En ambos textos se advierte la importancia de esta propiedad, pues los estudiantes podrían tratar de usar la recta para extrapolar, lo cual no es válido matemáticamente. El texto [T2] señala al respecto: "*Las predicciones obtenidas para valores próximos al punto medio de la distribución son más fiables que las obtenidas para valores muy alejados*" ([T2], p.227).

La estimación de cada una de las variables se realiza con su correspondiente recta de regresión, es decir, las rectas no se obtienen despejando una de otra; aunque son muchos los estudiantes que cometen este error. En ambos textos se incluye esta advertencia (Figura 3.6.2).

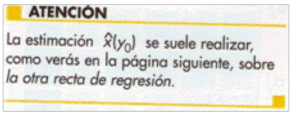

ATENCIÓN
La estimación $\hat{x}(y_0)$ se suele realizar, como verás en la página siguiente, sobre *la otra recta de regresión.*

Figura 3.6.2. Advertencia sobre la existencia de dos rectas de regresión ([T1], p. 231)

Relaciones entre conceptos

Relación entre el coeficiente de Pearson y la correlación. Como se señaló en la sección 3.4, el texto [T2] distingue las nociones de coeficiente de correlación de Pearson y correlación. Añade a este tratamiento los enunciados de la Figura 3.6.3, donde se ejemplifica tal relación.

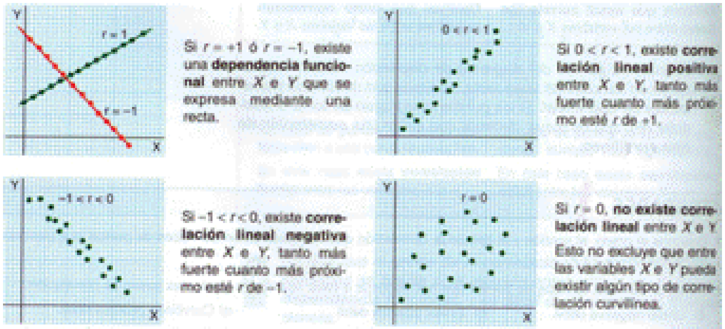


Figura 3.6.3. Relación entre el coeficiente de Pearson y la correlación ([T2], p. 224)

Relación entre signo de la correlación y pendiente de la recta. Debido a su expresión matemática, el coeficiente de correlación tiene el mismo signo que los coeficientes de regresión de las dos rectas. Además, intuitivamente es sencillo razonar que si la relación es directa (inversa), como se deduce del signo de la correlación, la recta ha de ser creciente (decreciente), por lo que su pendiente ha de ser positiva (negativa).

El ángulo formado por las rectas de regresión X sobre Y e Y sobre X informa sobre la intensidad de la correlación. Esta es una propiedad de interés porque relaciona propiedades geométricas con la regresión y correlación, ya que el coseno del ángulo de las dos rectas es el coeficiente de correlación. Sánchez Cobo (1998) indicó que los estudiantes tienen dificultad para comprender esta propiedad. Es por medio de ejemplos como el texto [T1] relaciona la posición de las rectas de regresión en cuanto a la amplitud del ángulo que éstas forman y la intensidad de la correlación (véase Figura 3.8.3). El texto [T2] no explicita tal relación.

Relación entre centro de gravedad y recta de regresión. En ambos textos se señala que el centro de gravedad de la distribución es un punto de cada recta de regresión. Se remarca la representación de este punto en el diagrama de dispersión, y se evidencia que el centro de gravedad es la intersección de las dos rectas de regresión gráficamente, aunque no se llega a demostrar.

Relación entre dispersión de la nube de puntos e intensidad de la correlación. Cuanto mayor es la dispersión menor es el valor del coeficiente de correlación y viceversa. Esta propiedad se incluye en el texto ([T1], p.226) mediante ejemplos.

Relación entre dispersión en la nube de puntos, recta de regresión y covarianza. Los libros analizados relacionan la dispersión de la nube de puntos y la covarianza mediante el uso de un gráfico auxiliar en que se representa el área de cada uno de los rectángulos que se pueden formar para cualquier dato bidimensional (ver Figura 3.8.2) haciéndose indicar que atendiendo a la posición de cada dato bidimensional, el área resultará positiva o negativa siendo cada uno de éstos valores un sumando de la fórmula de la covarianza. Continuando con la lectura del gráfico, se indica que si los datos bidimensionales se ajustan a una recta de pendiente positiva, los sumandos en la fórmula de la covarianza corresponderán a cálculos de áreas positivas pertenecientes a datos en el primer y tercer cuadrante (cuadrantes formados a partir de las rectas perpendiculares $x = \bar{x}$ e $y = \bar{y}$) y por tanto la covarianza será grande.

El producto de los dos coeficientes de regresión es r^2. Este enunciado lo encontramos en ambos textos y es demostrado mediante una igualdad algebraica.

Relación entre las estimaciones realizadas mediante la recta de regresión y el coeficiente de correlación. En ambos textos se señala que las estimaciones que se realicen con la recta de regresión serán mejores cuanto mayor sea el valor absoluto del coeficiente de correlación. A esta propiedad, el texto [T2] añade la importancia del tamaño de la muestra dado que: "*La fiabilidad de una recta de regresión es mayor cuanto mayor sea el número de datos considerados para calcularla.*" ([T2], p.227).

En la Tabla 3.6.1 presentamos las proposiciones encontradas en los textos analizados. Observamos en los dos textos gran cantidad de proposiciones y relaciones, aunque el primer texto no incluye las relacionadas con el resumen de los datos. Las propiedades presentadas en los dos textos son muy similares.

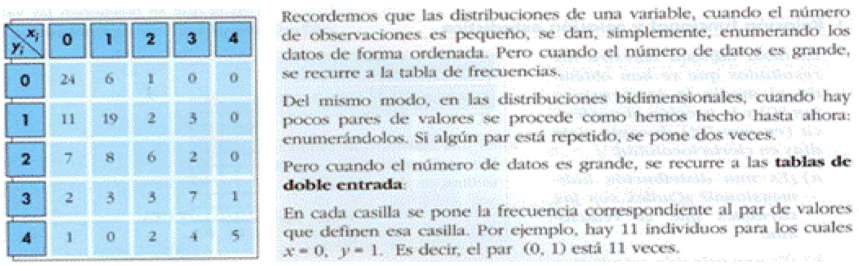
Tabla 3.6.1. Proposiciones en los libros analizados

Conceptos	Propiedades/relaciones	T1	T2
Organización de datos (tabular)	Suma de frecuencias absolutas fila o columna y frecuencia marginal.		X
	Extremos de intervalo de clase.		X
Distribución bidimensional	Centro de gravedad puede no ser punto de la distribución	X	X
Correlación	Signo y rango de variación.	X	X
	Intensidad.	X	X
	Adimensionalidad.	X	
Regresión	Recta mínimos cuadrados.	X	X
	Las estimaciones son aproximaciones.	X	X
	Las estimaciones deben hacerse en el rango de la variable.	X	X
	Las estimaciones deben hacerse próximas al punto medio.		X
	Dos rectas diferentes.	X	X
Relaciones	Coeficiente de Pearson y correlación.		X
	Signo de correlación y pendiente rectas.	X	X
	Angulo de rectas e intensidad de correlación.	X	Sin precisar
	Centro de gravedad y recta de regresión.	X	X
	Dispersión de la nube e intensidad de la correlación.	X	X
	Dispersión, rectas y covarianza.	X	X
	Producto de coeficientes de regresión	X	X
	Estimaciones y coeficiente correlación	X	X
	Importancia del tamaño de la muestra.		X

3.7. PROCEDIMIENTOS

También hemos analizado los *procedimientos*, es decir, el conjunto de algoritmos, técnicas de cálculo, planificaciones de ejecución, etc., que se presentan en los textos y facilitan la resolución de los campos de problemas analizados (sección 3.3) ya que es habitual proporcionar al alumnado modos de actuar ante las situaciones-problemas que se pueden plantear. En los textos analizados encontramos los siguientes.

Representación tabular de datos. Un primer paso en el estudio bidimensional sería construir una representación en la que se organicen los datos recogidos. El texto [T1] plantea la enseñanza de la correlación y regresión con el tratamiento de tablas en que las variables se apilan en filas o columnas y se limita a describir la tabla de doble entrada, propiamente dicha, al final del tema (Figura 3.7.1), Por el contrario, el texto [T2] describe los pasos para construir una tabla de doble entrada, tanto para datos discretos como continuos (Figura 3.7.2 y 3.7.3) y desarrolla el tema bajo su tratamiento.



$y_j \backslash x_i$	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

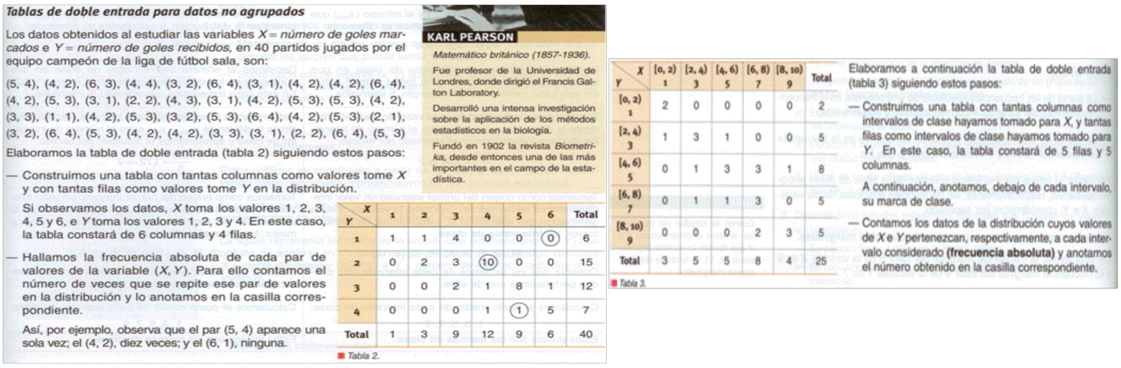
Recordemos que las distribuciones de una variable, cuando el número de observaciones es pequeño, se dan, simplemente, enumerando los datos de forma ordenada. Pero cuando el número de datos es grande, se recurre a la tabla de frecuencias.

Del mismo modo, en las distribuciones bidimensionales, cuando hay pocos pares de valores se procede como hemos hecho hasta ahora: enumerándolos. Si algún par está repetido, se pone dos veces.

Pero cuando el número de datos es grande, se recurre a las **tablas de doble entrada**.

En cada casilla se pone la frecuencia correspondiente al par de valores que definen esa casilla. Por ejemplo, hay 11 individuos para los cuales $x = 0$, $y = 1$. Es decir, el par (0, 1) está 11 veces.

Figura 3.7.1. Representación tabular de distribución bidimensional ([T1], p. 233)



Tablas de doble entrada para datos no agrupados

Los datos obtenidos al estudiar las variables X = *número de goles marcados* e Y = *número de goles recibidos*, en 40 partidos jugados por el equipo campeón de la liga de fútbol sala, son:

(5, 4), (4, 2), (6, 3), (4, 4), (3, 2), (6, 4), (3, 1), (4, 2), (4, 2), (6, 4), (4, 2), (5, 3), (3, 1), (2, 2), (4, 3), (3, 1), (4, 2), (5, 3), (5, 3), (4, 2), (3, 3), (1, 1), (4, 2), (5, 3), (3, 2), (5, 3), (6, 4), (4, 2), (5, 3), (2, 1), (3, 2), (6, 4), (5, 3), (4, 2), (4, 2), (3, 3), (3, 1), (2, 2), (6, 4), (5, 3)

Elaboramos la tabla de doble entrada (tabla 2) siguiendo estos pasos:

— Construimos una tabla con tantas columnas como valores tome X y con tantas filas como valores tome Y en la distribución.

Si observamos los datos, X toma los valores 1, 2, 3, 4, 5 y 6, e Y toma los valores 1, 2, 3 y 4. En este caso, la tabla constará de 6 columnas y 4 filas.

— 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.

Así, por ejemplo, observa que el par (5, 4) aparece una sola vez; el (4, 2), diez veces; y el (6, 1), ninguna.

Y \ X	1	2	3	4	5	6	Total
1	1	1	4	0	0	0	6
2	0	2	3	10	0	0	15
3	0	0	2	1	8	1	12
4	0	0	0	1	1	5	7
Total	1	3	9	12	9	6	40

Tabla 2.

KARL PEARSON

Matemático británico (1857-1936).

Fue profesor de la Universidad de Londres, donde dirigió el Francis Galton Laboratory.

Desarrolló una intensa investigación sobre la aplicación de los métodos estadísticos en la biología.

Fundó en 1902 la revista *Biometrika*, desde entonces una de las más importantes en el campo de la estadística.

Y \ X	[0, 2) 1	[2, 4) 3	[4, 6) 5	[6, 8) 7	[8, 10) 9	Total
[0, 2) 1	2	0	0	0	0	2
[2, 4) 3	1	3	1	0	0	5
[4, 6) 5	0	1	3	3	1	8
[6, 8) 7	0	1	1	3	0	5
[8, 10) 9	0	0	0	2	3	5
Total	3	5	5	8	4	25

Tabla 3.

Elaboramos a continuación la tabla de doble entrada (tabla 3) siguiendo estos pasos:

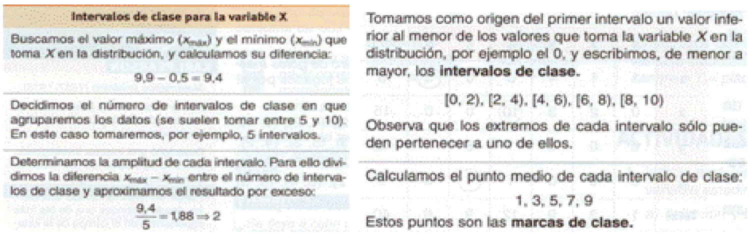
— Construimos una tabla con tantas columnas como intervalos de clase hayamos tomado para X, y tantas filas como intervalos de clase hayamos tomado para Y. En este caso, la tabla constará de 5 filas y 5 columnas.

A continuación, anotamos, debajo de cada intervalo, su marca de clase.

— Contamos los datos de la distribución cuyos valores de X e Y pertenezcan, respectivamente, a cada intervalo considerado (**frecuencia absoluta**) y anotamos el número obtenido en la casilla correspondiente.

Figura 3.7.2. Construcción de tabla de doble entrada (datos agrupados/no agrupados) ([T2], pp. 217-218)

Como anexo a la representación tabular, aunque tan sólo aparece en el texto [T2], encontramos el procedimiento para agrupar los datos de la distribución en intervalos de clase (Figura 3.7.3).



Intervalos de clase para la variable X

Buscamos el valor máximo ($x_{máx}$) y el mínimo ($x_{mín}$) que toma X en la distribución, y calculamos su diferencia:

$$9{,}9 - 0{,}5 = 9{,}4$$

Decidimos el número de intervalos de clase en que agruparemos los datos (se suelen tomar entre 5 y 10). En este caso tomaremos, por ejemplo, 5 intervalos.

Determinamos la amplitud de cada intervalo. Para ello dividimos la diferencia $x_{máx} - x_{mín}$ entre el número de intervalos de clase y aproximamos el resultado por exceso:

$$\frac{9{,}4}{5} = 1{,}88 \Rightarrow 2$$

Tomamos como origen del primer intervalo un valor inferior al menor de los valores que toma la variable X en la distribución, por ejemplo el 0, y escribimos, de menor a mayor, los **intervalos de clase.**

[0, 2), [2, 4), [4, 6), [6, 8), [8, 10)

Observa que los extremos de cada intervalo sólo pueden pertenecer a uno de ellos.

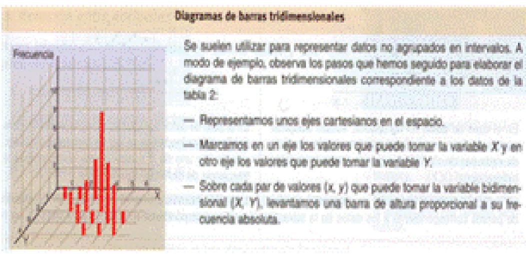
Calculamos el punto medio de cada intervalo de clase:

1, 3, 5, 7, 9

Estos puntos son las **marcas de clase.**

Figura 3.7.3. Proceso de construcción de intervalos de clase para la variable *X* ([T2], p.218)

Representación gráfica de datos. Los gráficos adquieren gran relevancia en la enseñanza y aprendizaje de las nociones de correlación y regresión. Se utilizan como nociones en sí mismas (sección 3.4) y como medio de expresión para informar sobre el grado, sentido y tipo de la correlación así como para el ajuste lineal entre las variables de estudio. En cuanto a su construcción, el texto [T1] describe brevemente cómo se construye el diagrama de dispersión para datos de frecuencia uno, indicando que cada dato se corresponde con las coordenadas de un punto en el eje cartesiano, y alude, al final del tema al modo de representar un diagrama de dispersión para datos de frecuencia distinta de uno (Figura 3.9.3). Presentamos a continuación (Figuras 3.7.4, 3.7.5, 3.7.6) los procedimientos que presenta el texto [T2].



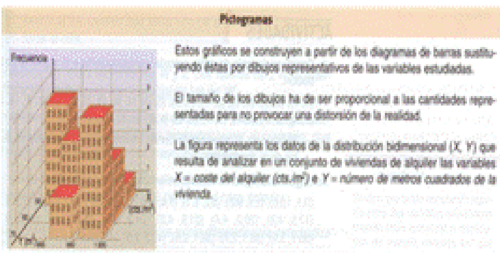
Diagramas de barras tridimensionales

Se suelen utilizar para representar datos no agrupados en intervalos. A modo de ejemplo, observa los pasos que hemos seguido para elaborar el diagrama de barras tridimensionales correspondiente a los datos de la tabla 2:

— Representamos unos ejes cartesianos en el espacio.

— Marcamos en un eje los valores que puede tomar la variable X y en otro eje los valores que puede tomar la variable Y.

— Sobre cada par de valores (x, y) que puede tomar la variable bidimensional (X, Y), levantamos una barra de altura proporcional a su frecuencia absoluta.



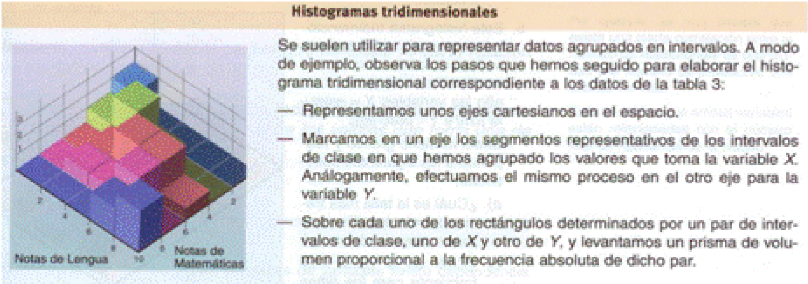
Pictogramas

Estos gráficos se construyen a partir de los diagramas de barras sustituyendo éstas por dibujos representativos de las variables estudiadas.

El tamaño de los dibujos ha de ser proporcional a las cantidades representadas para no provocar una distorsión de la realidad.

La figura representa los datos de la distribución bidimensional (X, Y) que resulta de analizar en un conjunto de viviendas de alquiler las variables X = *coste del alquiler (cts./m²)* e Y = *número de metros cuadrados de la vivienda.*

Figura 3.7.4. Construcción de un diagrama de barras tridimensional y pictograma ([T2], p.219)



Histogramas tridimensionales

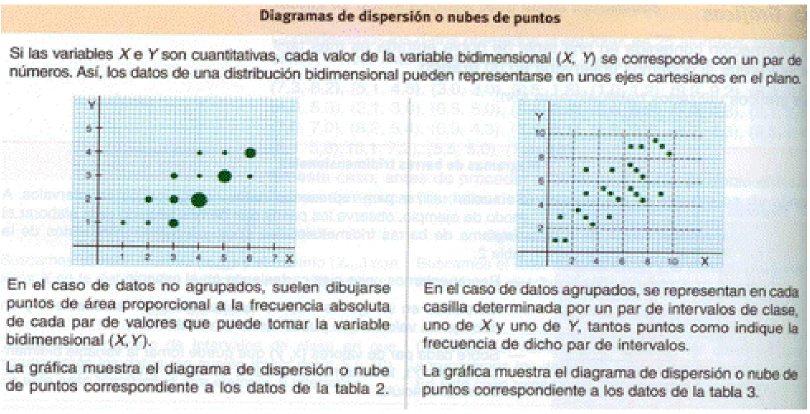
Se suelen utilizar para representar datos agrupados en intervalos. A modo de ejemplo, observa los pasos que hemos seguido para elaborar el histograma tridimensional correspondiente a los datos de la tabla 3:

— Representamos unos ejes cartesianos en el espacio.

— Marcamos en un eje los segmentos representativos de los intervalos de clase en que hemos agrupado los valores que toma la variable *X*. Análogamente, efectuamos el mismo proceso en el otro eje para la variable *Y*.

— Sobre cada uno de los rectángulos determinados por un par de intervalos de clase, uno de *X* y otro de *Y*, y levantamos un prisma de volumen proporcional a la frecuencia absoluta de dicho par.

Figura 3.7.5. Construcción de un histograma tridimensional ([T2], p.219)



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*).

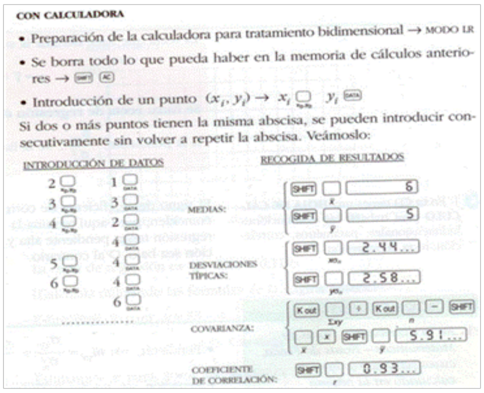
La gráfica muestra el diagrama de dispersión o nube de puntos correspondiente a los datos de la tabla 2.

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 3.7.6. Construcción de una nube de puntos/diagrama de dispersión ([T2], p.220)

Uso de herramientas tecnológicas para el tratamiento de la correlación y regresión. En los libros analizados se presentan los pasos necesarios para calcular parámetros estadísticos, el coeficiente de correlación lineal y las rectas de regresión mediante el uso de la calculadora. En la clasificación de los campos de problemas de las Tablas 3.3.1 y 3.3.2 se informa sobre la presencia de las actividades referidas específicamente con el uso de la calculadora. En la Figura 3.7.7 se puede observar este tratamiento en el cálculo del coeficiente de correlación lineal.



CON CALCULADORA

• Preparación de la calculadora para tratamiento bidimensional → MODO LR

• Se borra todo lo que pueda haber en la memoria de cálculos anteriores → SHIFT AC

• Introducción de un punto $(x_i, y_i) \rightarrow x_i$ y_i DATA

Si dos o más puntos tienen la misma abscisa, se pueden introducir consecutivamente sin volver a repetir la abscisa. Veámoslo:

Figura 3.7.7. Cálculo del coeficiente de correlación con calculadora ([T1], p. 229)

Ambos textos hacen alusión a la hoja de cálculo, como herramienta para la representación gráfica, y el tratamiento de la distribución bidimensional y las tablas de doble entrada. Además, en el texto [T2] se ofrece un enlace a la aplicación Descartes:

http://descartes.cnice.mec.es/materiales_didacticos/Variables_estadisticas_bidimensionales_regresion_correlacion/Indice.htm, (desarrollado por *Leoncio Santos Cuervo* para el Ministerio de Educación en el año 2001), como unidad didáctica con actividades y aplicaciones interactivas.

Cálculo de covarianza y /o correlación. De modo informal se indica cómo puede apreciarse la correlación entre dos variables: "… *la correlación entre dos variables (más o menos fuerte, positiva o negativa) se aprecia mediante el grado de "apretura" de los puntos de la nube*" ([T1], p.228). Aunque se utilizan mayoritariamente cálculos mediante fórmulas. En ambos textos se presentan ejercicios resuelto como modelo de resolución, contemplando además el texto [T1] el uso de la calculadora.

Cálculo analítico de la recta de regresión. El procedimiento de cálculo analítico varía de un texto a otro. En el texto [T2] se utiliza mayoritariamente la sustitución de la pendiente y la ordenada en el origen en la expresión: $y = A + B \cdot x$ (Figura 3.7.8).

En el caso del texto [T1], se sustituyen los valores en la expresión de la recta punto-pendiente: $y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$.

En el ejemplo 2 ya calculamos los parámetros $\bar{x} = 6{,}833$; $\bar{y} = 4{,}833$; $\sigma_{xy} = 0{,}559$; $\sigma_x = 1{,}069$ y $\sigma_y = 0{,}626$. Así pues, sustituyendo en las expresiones de los coeficientes *A* y *B* se obtiene la recta de regresión de *Y* sobre *X*:

$$A = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x} = 4{,}833 - \frac{0{,}559}{(1{,}069)^2} \cdot 6{,}833 = 1{,}491 \quad B = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0{,}559}{(1{,}069)^2} = 0{,}489 \longrightarrow y = 1{,}491 + 0{,}489x$$

Figura 3.7.8. Cálculo de la recta de regresión de *Y* sobre *X* ([T1], p. 227)

Cabe hacer mención al procedimiento que sugiere el texto [T1] para el cálculo de la recta de regresión de *X* sobre *Y* (Figura 3.7.9).

La ecuación de la **recta de regresión de *X* sobre *Y*** es:

$$x = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

O bien, despejando *y*:

$$y = \bar{y} + \frac{\sigma_y^2}{\sigma_{xy}}(x - \bar{x})$$

El número $\frac{\sigma_{xy}}{\sigma_y^2}$ se llama **coeficiente de regresión de *X* sobre *Y*.**

Observa en la segunda ecuación que este coeficiente no es la pendiente de la recta, sino su inversa.

Figura 3.7.9. Recta de regresión de *X* sobre *Y* ([T1], p. 232)

Resolución de una tarea. Aunque no se trata propiamente de un procedimiento, en el texto [T2] se resuelve una tarea en forma algorítmica, que hemos considerado de interés incluirla como Anexo 4 por el modo en que se asemeja a la resolución de un proyecto que incluye el tratamiento de la correlación y regresión.

Estimación/Predicción a través de la recta de regresión. En ambos textos se indica el procedimiento para realizar estimaciones (Figuras 3.7.10 y 3.7.11).

a) Sustituyendo $x = 9$ en la recta de regresión de Y sobre X, obtenemos $\hat{y} = 1{,}491 + 0{,}489 \cdot 9 = 5{,}892 \Rightarrow \hat{y} \approx 6$

b) Sustituyendo $y = 3{,}5$ en la recta de regresión de X sobre Y, obtenemos $\hat{x} = -0{,}061 + 1{,}427 \cdot 3{,}5 = 4{,}934 \Rightarrow \hat{x} \approx 5$

Figura 3.7.10. Predicción de valores a partir de las correspondientes rectas de regresión ([T2], p. 227)

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 3.7.11. Predicción de valores a partir de las correspondientes rectas de regresión ([T1], p. 231)

Tabla 3.7.1. Procedimientos en los libros analizados

Procedimientos	T1	T2
Representación tabular	Ejemplo	Ejemplos, Datos agrupados y no agrupados
Representación gráfica	Diagrama de dispersión Histograma tridimensional	Diagrama de barras tridimensional Pictograma tridimensional Histograma tridimensional Diagrama de dispersión bi y tridimensional
Herramientas tecnológicas	Uso de calculadora en ejemplos o ejercicios resueltos. Hoja de cálculo en un CD	Procedimientos de cálculo con calculadora. Alusión a la hoja de cálculo y unidad didáctica en Internet.
Cálculo covarianza y/o correlación	A mano y con calculadora	A mano y con calculadora
Cálculo recta de regresión	Fórmula punto- pendiente	Sustituyendo la pendiente y ordenada en el origen
Predicción a través de la recta de regresión	Sustituyendo en la correspondiente recta	Sustituyendo en la correspondiente recta
Resolución de un “mini” proyecto		Planteamiento y resolución de los pasos a seguir en la resolución de un proyecto sobre correlación y regresión.

En la Tabla 3.7.1 se presentan los procedimientos encontrados. Observamos que se incluyen procedimientos relacionados con los diferentes conceptos y también el uso de la tecnología, e incluso en el texto [T2] un mini-proyecto.

3.8. ARGUMENTOS

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, convicción, explicación, sistematización, descubrimiento o comunicación (Sánchez Cobo, 1998). Los 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 (1998) se recogen en todos los textos al igual que en nuestra revisión, y a continuación los presentamos.

A1. Justificación con ejemplos o contraejemplos. Es frecuente encontrar ejemplos acompañados de argumentos que justifique su resultado y/o su aplicación. Destacamos a continuación algunos de éstos:

A1.1. Justificación de la representación del diagrama de dispersión. Tan sólo el texto [T2] se refiere a la tipología de la variable en cuanto al diagrama de dispersión e indica que: "*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*" ([T2], p.220) añadiendo posteriormente un ejemplo (ver Figura 3.7.7).

A1.2. Justificación de la relación entre la pendiente de la recta de regresión y signo del coeficiente de correlación. Por lo general, este tipo de argumentaciones aparecen en ejemplos o ejercicios resueltos (Figura 3.3.4).

A1.3. Justificación de la relación entre *dispersión de los puntos e intensidad de la correlación*. Por medio de ejemplos de soporte gráfico se justifica que cuanto más juntos están los puntos en la nube de puntos, más intensa es la correlación en el ajuste lineal ([T1], p.226). Señalamos al respecto un ejercicio resuelto del texto [T2] (Figura 3.8.1) donde se pide asignar los coeficientes de correlación a las nubes de puntos representadas. En dicha resolución, cada una de las asignaciones implica el uso de estos argumentos.

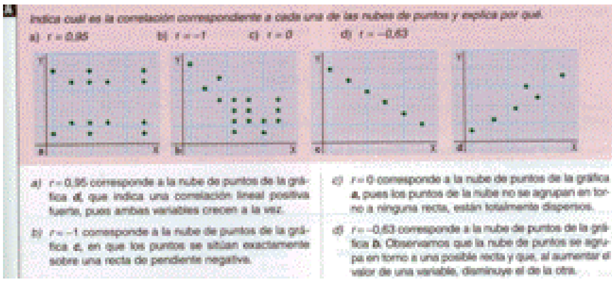


Figura 3.8.1. Relación entre la nube de puntos y el coeficiente de correlación lineal ([T2], p. 229)

A2. Uso de propiedades de gráficos auxiliares para apoyar una argumentación verbal o simbólica. También en el estudio de Sánchez Cobo (1998) se indica que los gráficos se utilizan con frecuencia como ejemplos en la argumentación, más que como ejercicios de representación de datos. Suelen emplearse para el desarrollo de explicaciones y "demostraciones" relativas a la correlación y regresión, como son:

A2.1. Uso de la división del plano en cuatro cuadrantes mostrando gráficamente la *relación entre el signo de la covarianza, la forma de la nube, así como la dispersión de los puntos respecto a la recta de regresión.* El texto [T1] relacionan estos tres conceptos por medio del soporte gráfico de forma similar a la realizada en Holmes (2001):

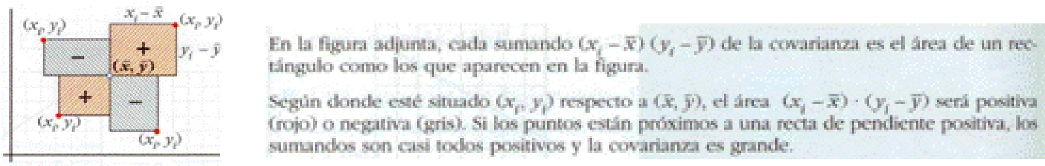


Figura 3.8.2. Relación covarianza, nube de puntos y recta de regresión ([T1], p. 228)

A2.2. Justificación gráfica de dependencia funcional o estadística. En ambos textos se utilizan representaciones gráficas que facilitan la interpretación de la dependencia estadística. El texto [T2] considera además la relación curvilínea (Figura 3.8.3).

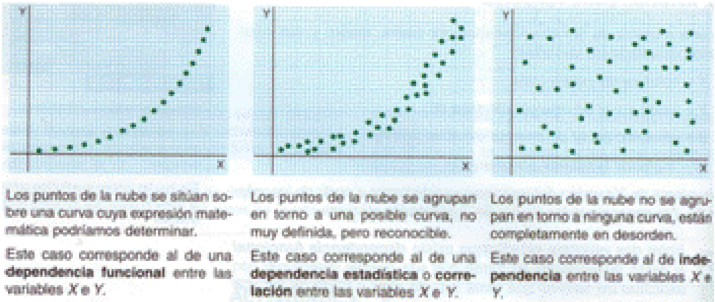


Figura 3.8.3. Interpretación gráfica de la relación entre variables ([T2], p. 222)

A2.3. Justificación gráfica de la intensidad, sentido y tipo de correlación. En ambos textos se utiliza una representación gráfica que facilite la interpretación de la intensidad y sentido de la correlación entre variables destacando el texto [T2] ya que además presenta la correlación curvilínea (Figura 3.8.4).

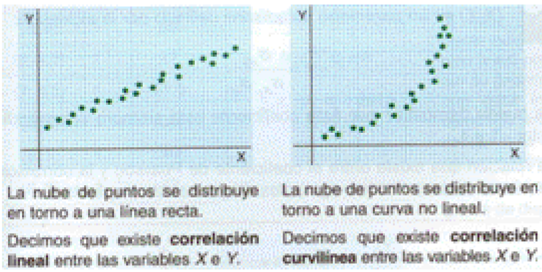


Figura 3.8.4. Tipo de correlación entre variables ([T2], p. 223)

A2.4. Justificación gráfica del significado del ángulo que forman las rectas de regresión X sobre Y e Y sobre X. Destacamos el texto [T1] donde por medio del soporte gráfico se relaciona el significado de dicho ángulo y su relación con la intensidad del coeficiente de correlación (Figura 3.8.5).

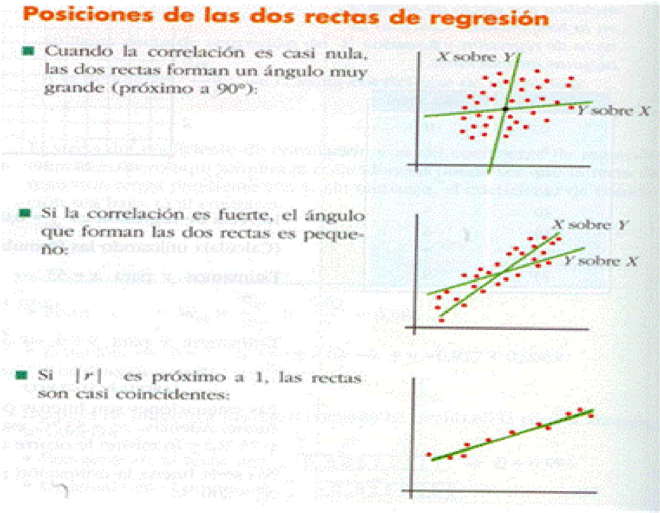


Figura 3.8.5. Recta de regresión de X sobre Y ([T1], p. 232)

A3. Razonamientos verbales deductivos. Son la mayoría, como en el trabajo de Sánchez Cobo (1998). Se trata, generalmente, de razonamientos no excesivamente formalizados, que de forma deductiva justifican una propiedad. Se destacan:

A3.1. Justificación de la intensidad de la correlación. Se encuentran en el texto [T2] y se realizan desde el cálculo del coeficiente de correlación y en menor medida con soporte gráfico. En la Figura 3.8.6 mostramos un ejemplo, entre otros.

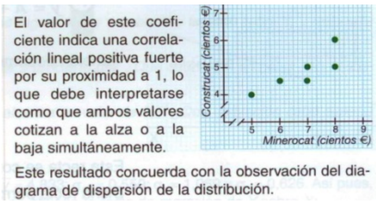


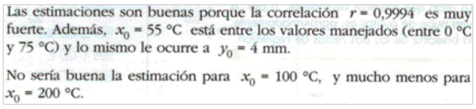
Figura 3.8.6. Justificación e interpretación del coeficiente de correlación ([T2], p. 225)

A3.2. Justificación de la distribución marginal. Sólo el texto [T2], y tras explicar la construcción de una tabla de doble entrada (sección 3.7), concluye que en la última fila (o columna) se encuentran las frecuencias absolutas de la variable *Y* (o *X*).

A3.3. Justificación de la predicción mediante la recta de regresión. En el texto [T2], y como motivación al estudio de la regresión lineal, encontramos argumentos como: "(…) *es lógico pensar que, si hay una curva en torno a la cual se agrupan los puntos de un diagrama de dispersión, ésta ha de dar una aproximación de los valores reales* (…)" ([T2], p. 226) (véase además Anexo 4).

A3.4. Justificación del uso del método de mínimos cuadrados. En el texto [T2] se menciona la necesidad de utilizar un método alejado del subjetivismo de un trazado por ajuste manual. Se justifica así la ventaja de su uso ([T2], p. 226).

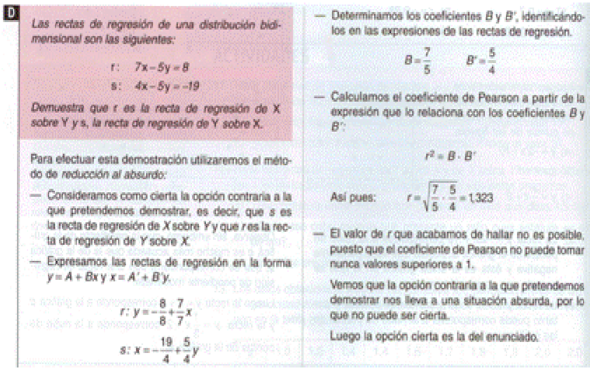
A3.5. Justificación de la relación entre la intensidad de la correlación y bondad de *las estimaciones realizadas con la recta de regresión.* En los textos analizados es habitual justificar la bondad de las estimaciones recurriendo a la intensidad de la correlación entre las variables de estudio (Figura 3.8.7).


Las estimaciones son buenas porque la correlación $r = 0{,}9994$ es muy fuerte. Además, $x_0 = 55$ °C está entre los valores manejados (entre 0 °C y 75 °C) y lo mismo le ocurre a $y_0 = 4$ mm.

No sería buena la estimación para $x_0 = 100$ °C, y mucho menos para $x_0 = 200$ °C.

Figura 3.8.7. Justificación de la estimación ([T1], p.231)

A4. Demostración por reducción al absurdo. En el texto [T2] se plantean y resuelven tareas aplicando el método de reducción al absurdo (Figura 3.8.8).


D

Las rectas de regresión de una distribución bidimensional son las siguientes:

$$r: \quad 7x - 5y = 8$$
$$s: \quad 4x - 5y = -19$$

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 que r es la recta de regresión de Y sobre X.

— Expresamos las rectas de regresión en la forma $y = A + Bx$ y $x = A' + B'y$.

$$r: y = -\frac{8}{5} + \frac{7}{5}x$$
$$s: x = -\frac{19}{4} + \frac{5}{4}y$$

— Determinamos los coeficientes B y B', identificándolos en las expresiones de las rectas de regresión.

$$B = \frac{7}{5} \qquad B' = \frac{5}{4}$$

— 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 3.8.8. Resolución de una tarea por el método de reducción al absurdo. ([T2], p. 230)

En la Tabla 3.8.1 se presentan los argumentos encontrados, la mayoría de los cuáles se tratan de ejemplos y contraejemplos o bien gráficos como soporte de argumentación, resultado también encontrado en otros trabajos sobre libros de texto, como los de Alvarado (2007) sobre el teorema central del límite y Olivo (2008), sobre intervalos de confianza. El texto [T2] incluye también argumentos deductivos e incluso presenta el método de reducción al absurdo a modo de procedimiento de un modo general. Es por tanto más formalizado en sus argumentos.

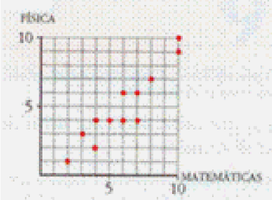
Tabla 3.8.1. Argumentos en los libros analizados

Argumentos		T1	T2
Ejemplos/ contraejemplos	Relación dispersión-intensidad.	X	X
	Relación pendiente-signo correlación.	X	X
	Representación diagrama de dispersión.		X
Gráficos auxiliares	Covarianza-diagrama dispersión-recta.	X	
	Dependencia funcional o estadística.	X	X
	Intensidad y sentido de correlación.	X	X
	Ángulo entre rectas de regresión.	X	
	Tipo de correlación		X
Verbales deductivos	Intensidad de correlación y estimación	X	X
	Distribución marginal.		X
	Recta de regresión para la predicción.		X
	Uso del método de mínimos cuadrados.		X
	Intensidad de la correlación.		X
Reducción al absurdo	Como método general		X

3.9. CONFLICTOS SEMIÓTICOS

En el análisis realizado hemos encontrado asignaciones imprecisas de significado, susceptibles de provocar en el alumno un conflicto semiótico si el profesor no está atento al uso que de ellas pueda llevarse a cabo. A continuación presentamos algunos.

Confusión de un concepto con su representación tabular y/o gráfica. El texto [T1] plantea la definición de una *distribución bidimensional* en dos momentos diferentes. En primer lugar se le atribuye el significado a través de una representación gráfica y tabular (Figura 3.9.1), incitando al alumno a confundir su representación (gráfica o tabla) con el propio objeto (distribución). Más adelante, se da una definición verbal (Figura 3.9.2). Desde nuestro punto de vista hubiese sido preferible completar la definición inicial.



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**.

Figura 3.9.1. (Texto T1, p. 226)

El conjunto de pares de valores (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) se llama **distribución bidimensional**. 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**.

Figura 3.9.2. (Texto T1, p. 226)

Algo parecido ocurre con el uso de la noción de *distribución marginal* de *X* e *Y*, donde se utiliza dicha noción como simple etiquetado de una tabla de frecuencias y a diferencia del caso anterior ni se define dicha noción ([T1], p.237).

Aplicar inapropiadamente las diferentes representaciones gráficas. Histogramas y diagramas de barras tridimensionales. El texto [T1] describe cómo representar gráficamente los datos de una tabla de doble entrada, pero no asigna un nombre que permita identificar cada una de estas representaciones (Figura 3.9.3). Esto puede confundir al alumno en el uso del histograma y del diagrama de barras tridimensional. Suponemos que se está graficando un histograma tridimensional, por la distancia que deja entre las barras, pero la variable que se representa es cuantitativa discreta, y se debiese haber dejado espacio entre éstas tal como corresponde a un diagrama de barras tridimensional.

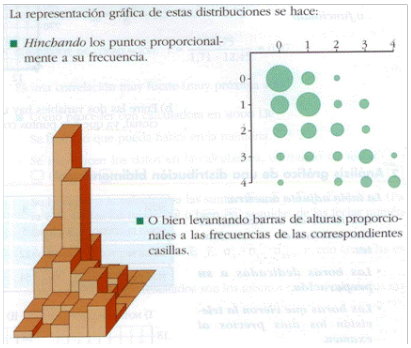


Figura 3.9.3. Representación gráfica de una distribución bidimensional ([T1], p. 233)

Interpretación demasiado generalizada del concepto de correlación. En los textos analizados se suele utilizar la noción de correlación como sinónimo de relación estadística entre variables (véase la Sección 4.4). En realidad, este término sólo se aplica a variables numéricas y el problema radica con el tratamiento de la correlación lineal. El texto [T1] no precisa la distinción entre correlación y correlación lineal y cuando las variables no se encuentran relacionadas indica: "…*la nube de puntos es amorfa y no sugiere ninguna recta: no hay correlación entre las variables (se dice que son incorreladas)*". Se deja a cargo del alumno, como algo transparente, la diferencia entre dependencia lineal y correlación significativa. Todo esto nos hace pensar que la confusión no proviene de un tratamiento generalizado de la noción de correlación sino más bien de una falta de significación de esta noción.

Imprecisión en la definición de los términos correlación y regresión. En la introducción al tema en el texto [T1] se presenta brevemente el estudio que motivó la creación de las nociones de correlación y regresión debido a Galton. La cuestión es que no se hace alusión alguna al término correlación, generando pues una imprecisión en el tratamiento de estas nociones como exponemos a continuación:

> "Según Galton, la estatura de los hijos regresa hacia la media de la población. De ahí el término *regresión*, que, desde entonces, se utiliza para designar una relación estadística cualquiera." ([T1], p.224)

Confusión en el tratamiento de la fórmula de la covarianza. Cuando se introduce la covarianza de modo formal sería recomendable utilizar una formulación clara, que si es posible venga acompañada de todas las expresiones equivalentes que se pretendan utilizar en el texto. El texto [T2] utiliza la siguiente formulación:

$$\sigma_{xy} = \frac{\sum_{i,j}\left(x_i - \bar{x}\right)\cdot\left(y_j - \bar{y}\right)\cdot n_{ij}}{N} = \frac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x}\cdot\bar{y}$$

acompañada, tanto al margen como en el texto y al inicio de la unidad, de unas anotaciones para hacerla más entendible. El problema radica en el texto [T1] porque al rehusar del formalismo del sumatorio, se potencia un conflicto en el alumno en cuanto al tratamiento de la fórmula de la covarianza. Se presenta así un problema resuelto (Figura 3.9.4) donde la notación no clarifica el procedimiento.

$y_i \backslash x_i$	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

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1{,}08}{1{,}22\cdot 1{,}5} = 0{,}68$$

Figura 3.9.4. Cálculo del coeficiente de correlación lineal ([T2], p. 224)

No se precisa la diferencia entre las dos rectas de regresión en cuanto a su uso en la estimación. Es de destacar el riesgo que conlleva realizar estimaciones de las variables de estudio mediante una misma recta de regresión. Destacamos el planteamiento que se lleva a cabo en un ejercicio en el texto [T1] 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*” ([T1], p. 231), pero pensamos que la ingenuidad del alumno traspasa las fronteras de estas advertencias incurriendo en los errores detectados en las investigaciones detalladas en el Capítulo 2.

3.10. CONCLUSIONES DEL ESTUDIO

En este capítulo se ha llevado a cabo un estudio detallado de la presentación de la correlación y la regresión en dos libros de texto publicados el año siguiente a la publicación del *Real Decreto de Enseñanzas Mínimas* (MEC, 2007b), los dos dirigidos a la modalidad de *Humanidades y Ciencias Sociales*. Se 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 (Anexo 1). A continuación exponemos nuestras principales conclusiones, todas ellas provisionales, y podrían interpretarse como hipótesis para un futuro estudio ampliado ya que la muestra de textos analizados es muy pequeña.

- 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.
- 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 (1998), que 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.
- 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. Respecto al cambio de representación, generalmente se pide al estudiante pasar de un listado a un gráfico, haciendo pocas actividades de traducción de otros tipos de representaciones consideradas por Sánchez Cobo (1998) o Lavalle y otros (2006).
- Hay 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 los contextos y no presentar la estadística como una colección de recetas o fórmulas de cálculo.
- 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 barra 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.
- 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.

CAPITULO 4.

CONCLUSIONES DEL ESTUDIO

4.1. INTRODUCCIÓN

Como punto final de este trabajo de Máster, en este capítulo exponemos las conclusiones sobre cada uno de los objetivos fijados y algunas ideas para continuar nuestra investigación y realizar la tesis doctoral.

4.2. CONCLUSIONES RESPECTO A LOS OBJETIVOS

En el primer capítulo de la Memoria se fijaron sus objetivos, indicando también los capítulos en los que se trataba de lograrlos. A continuación incluimos de nuevo estos objetivos, para exponer las principales conclusiones alcanzadas sobre cada uno de ellos.

- *O1. Analizar los contenidos sobre correlación y regresión desde el punto de vista elemental y descriptivo* para identificar los objetos matemáticos específicos en los que se enfocará la evaluación del conocimiento matemático para la enseñanza de los futuros profesores.

Puesto que nuestro marco teórico reconoce la diferencia y mutua relación entre los significados personales e institucionales de los objetos matemáticos, y la relatividad de dichos significados a las instituciones de enseñanza (Godino y Batanero, 1994; Godino, 2003) una tarea inicial en la investigación es fijar el significado institucional de referencia. Por esto motivo se realizó un análisis detallado de los diferentes tipos de objetos matemáticos que se consideran en este enfoque, ligados a la correlación y regresión.

La principal conclusión obtenida es que, incluso desde el punto de vista elemental y descriptivo, el estudio de la correlación y regresión es complejo, debido a la diversidad de objetos matemáticos involucrados, que, además están interrelacionados entre ellos y con muchos otros objetos estadísticos y matemáticos. Por un lado, la correlación y regresión se apoyan en el conocimiento previo de muchos objetos, como los de variable estadística, frecuencias y sus tipos, medidas de posición central y dispersión y gráficos estadísticos. Por otro servirá de base al estudio futuro de una amplia variedad de objetos estadísticos de amplia utilidad en la investigación y la vida profesional, entre los que destacan la correlación múltiple y no lineal, el análisis de varianza y el análisis multivariante. Estas conclusiones confirman el interés de un estudio didáctico del tema.

- *O2. Analizar asimismo los contenidos curriculares del tema en el Bachillerato,* que también se desarrolla en el capítulo 1, donde además se extiende el análisis curricular a los estándares del NCTM y el proyecto GAISE.

Respecto a este objetivo, hemos podido observar la similitud de contenidos en los currículos españoles, respecto a los americanos, concretizados en los estándares del NCTM(2000) y el proyecto GAISE (Franklin et al., 2007). Sin embargo, en estos últimos el tema se inicia en la educación secundaria, aunque en España se restringe al Bachillerato. Además el análisis permitió delimitar la extensión y nivel de profundidad con que los objetos se presentan en el Bachillerato.

- *O3. Profundizar en el marco teórico que será fundamento del trabajo de tesis futuro.* En el Anexo 1 se realiza un resumen extenso de los principales documentos producidos en el Enfoque Onto-semiótico.

El marco teórico es potencialmente útil para nuestro futuro trabajo, como ya se ha empezado a mostrar en esta Memoria. Por ejemplo, la tipología de objetos se ha usado para analizar los libros de texto y podrá ampliarse, considerando los procesos matemáticos, así como sus facetas duales. Ello nos permitirá profundizar en el análisis de los libros de texto presentado en esta Memoria, así como en el diseño de instrumentos de evaluación y en el análisis de respuestas con los mismos. Los constructos “conflicto semiótico” e “idoneidad didáctica” se han utilizado, aunque muy someramente en la valoración de los libros de texto. En el futuro el primero se utilizará en la tesis para analizar las respuestas de los futuros profesores y el segundo, con mucho más detalle para analizar la idoneidad del cuestionario de evaluación.

- *O4.Realizar una revisión bibliográfica y clasificar la investigación específica relacionada con el tema.*

La principal conclusión obtenida de este estudio, que se recoge en el Capítulo 2, es que apenas existen investigaciones centradas en el conocimiento de futuros profesores sobre la correlación y regresión (en ninguno de los componentes de dicho conocimiento). Por ello pensamos enfocar la investigación futura en este punto; en particular analizando los componentes y facetas del conocimiento del profesor para enseñar el tema. Otras conclusiones se refieren a la diversidad de conflictos semióticos ligados al tema y las variables didácticas de las posibles tareas en las situaciones-problemas relacionadas que nos ayudarán a construir ítems para el cuestionario, teniendo en cuenta estos resultados.

- *O4. Realizar un análisis del contenido que sobre correlación y regresión se incluye en dos libros de texto del Bachillerato de Ciencias Sociales.*

Aunque muy limitado, debido al número de textos analizados, este estudio nos permite aventurar algunas conjeturas, que deberán estudiarse en una muestra más amplia de libros. En primer lugar, al menos en los libros analizados, la presentación del tema no es homogénea, por lo que, para el mismo currículo, los estudiantes podrían recibir una formación diferente en función del texto utilizado.

Por otro lado, se han observados conflictos semióticos en los libros, en relación a algunas propiedades y representaciones gráficas, que podrían extenderse a otros textos. Es también demasiado alto el porcentaje de ejercicios y problemas contextualizado, lo que no favorece la adquisición de un razonamiento estadístico completo en los estudiantes, al suprimir la parte interpretativa, que es la más importante en modelos de razonamiento estadístico, como por ejemplo, los de Wild y Pfannkuch (1999). Todo ello nos da criterio para la mejora de los libros de texto y de la enseñanza del tema.

4.3. LÍNEAS DE INVESTIGACIÓN FUTURAS

Puesto que la principal carencia observada en los antecedentes del tema es la relacionada con los conocimientos y formación de profesores, nuestra intención es continuar la línea de investigación emprendida con un estudio sobre los conocimientos del profesor en relación con la correlación y regresión.

Más concretamente, como este tema se incluye actualmente en el currículo de Bachillerato nos interesaremos por los diversos componentes del conocimiento matemático de los futuros profesores de este nivel educativo para la enseñanza de la correlación y regresión. La población potencial serán los alumnos del Máster de Educación Secundaria y Bachillerato, especialidad de Matemáticas. Utilizaremos para analizar y evaluar los componentes de dicho conocimiento, tanto el enfoque onto-semiótico, y en particular, las categorías de componentes y facetas en el conocimiento del profesor descrita por Godino (2009), como el modelo propuesto por Ball, Thames y Phelps (2008).

El estudio realizado en esta Memoria de Máster proporciona las bases de dicho trabajo, en el que realizaremos, en primer lugar un estudio teórico sobre las componentes que deseamos evaluar. Seguidamente construiremos un cuestionario válido y fiable que nos permite obtener la información necesaria. Dicho cuestionario partirá del significado de referencia pretendido, delimitado a partir del estudio matemático, el estudio curricular y de libros de texto y las investigaciones previas.

Las investigaciones analizadas en esta Memoria de Máster proporcionan, asimismo, ítems que podrían adaptarse para el futuro cuestionario, que estará basado en el conocimiento matemático especializado y ampliado del contenido, así como en el conocimiento del contenido matemático y los estudiantes.

REFERENCIAS

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.

Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores.* Tesis Doctoral. Universidad de Granada.

Arteaga, P., Batanero, C., Cañadas, G. y Contreras, J. M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números 76*, 55-67.

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.

Azcárate, P. y Serradó, A. (2006). Tendencias didácticas en los libros de texto de matemáticas para la Eso. *Revista de Educación*, *340*, 341-378.

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.

Barbancho, A. G. (1973). *Estadística elemental moderna*. Barcelona: Ariel.

Barret, G. B. (2000).The coefficient of determination: understanding r^2 y R^2. *Mathematics Teacher*, *93* (3), 230-234.

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. (2004). Statistics education as a field for research and practice. *Proceedings of ICME-10.* CD-ROM. ISBN-978-87-7349-733-3. Copenhague: International Commission for Mathematical Instruction. Regular Lecture.

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. 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* (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., 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., 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. 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.

Benzecri, J. P. (1982). *Histoire et prehistoire de l'analyse des données*. París: Bordás.

Bisquerra, R. (1989). *Métodos de investigación educativa.* Barcelona: P.P.U.

Brousseau, G. (1997). *Theory of didactical situations in mathematics. Didactique des mathématiques, 1970-1990.* London: Kluwer.

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.

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: www.stat.auckland.ac.nz/~iase/serj/.

Chapman, L. J. (1967) Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior 6*(1), 151-155.

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.

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.

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.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin 90* (2), (pp. 272-292).

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.

Engel, J. y Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill, and 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. (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.. Batanero, C. y Sánchez, F. T. (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 Matemática Andaluza de Educación Matemática Thales

Estepa, A. y Gea, M. M. (2011). La enseñanza-aprendizaje de la asociación estadística. *Actas 15 Jornadas sobre el Aprendizaje y la Enseñanza de las Matemáticas*. Gijón: Servicio de Publicaciones de la Federación Española de Sociedades de Profesores de Matemáticas (FESPM). ISBN

Estepa, A. y Gea, M. M. (2012). Conocimiento para la enseñanza de la asociación estadística. En J. J. Ortiz (Ed.), *Investigaciones actuales en educación stadística y formación de profesores* (pp. 23-40). Melilla: Departamento de Didáctica de la Matemática. Facultad de Educación y Humanidades. 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.

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/.

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. 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. (2013). 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. (2013). 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., 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. *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

Ghiglione, R. y Matalón, B. (1989). *Las encuestas sociológicas*. México: Trillas.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. J*ournal for Research in Mathematics Education*, 38 (5), 427-437.

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. 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. 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., Rivas, M., Castro, W. F y Konic, P. (2008). Desarrollo de competencias para el análisis didáctico del profesor de matemáticas. *Actas de las VI Jornadas de Educación Matemática Región de Murcia.* Centro de Profesores y Recursos de Murcia. Disponible en: www.ugr.es/~jgodino/.

Gómez Chacón, I. M. (2000). *Matemática emocional. Los afectos en el aprendizaje matemático.* Madrid: Narcea.

Goetz, J. P. y Lecompte, M. D. (1998). *Etnografía y diseño cualitativo en educación*. Morata: Madrid.

Groth, R. E y Powell, N. N. (2004). Uso de los proyectos de investigación para ayudar al desarrollo del pensamiento estadístico en los estudiantes. *Mathematics Teacher*, *97* (2), 106-109.

Gutiérrez, C. S. (1994). *Filosofía de la estadística*. Valencia; Servicio de publicaciones de la Universidad de Valencia.

Hald, A. (1998). *A history of mathematical statistics. From 1750 to 1930*. New York: John Wiley.

Holmes, P. (2001). Correlation: From Picture to formula. *Teaching Statistics*, *23* (3), 67-71.

Johnson, R. y Kuby, P. (2004). *Estadística elemental*. México: Thompson.

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.

León, O. G. y Montero, I. (2003). *Métodos de Investigación*. Madrid: Mc Graw Hill.

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

Mcclain, K. y Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, *45* (1-3), 103-129.

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.

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.

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.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: VA.

Nolan, D., y Speed, T.P. (1999). Teaching statistics theory through applications. *American Statistician*, *5*3, 370-375.

Newman, J. R. (1956). *Sigma. El mundo de las matemáticas*. Barcelona: Grijalbo.

Olivo, E. (2008). *Significados de los intervalos de confianza para los estudiantes de ingeniería en México*. Tesis doctoral. Universidad de Granada.

Ortiz, J. J. (1999). *Significado de los conceptos probabilísticos elementales en los textos de Bachillerato*. Tesis Doctoral. Universidad de Granada.

Sánchez Cobo, F. T. (1998). *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).

Sierpinska, A. (1994). *Understanding in mathematics*. London: Falmer Press.

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.

Wild, C. J., y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67* (3), 223-265.

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.