

OBJETOS MATEMÁTICOS LIGADOS A LA REGRESIÓN EN LOS TEXTOS ESPAÑOLES DE BACHILLERATO

Carmen Batanero, M. Magdalena Gea, Danilo Díaz-Levicoy y Gustavo R. Cañadas

Educación Matemática, 27(2), 9-35, 2015

Resumen

El objetivo de este trabajo es caracterizar la presentación de la regresión en los libros de texto españoles de Bachillerato. Para ello se analizan y clasifican los campos de problema, procedimientos, conceptos y propiedades asociados a la regresión en dieciséis libros de texto de Bachillerato utilizados en España, ocho de la modalidad de Ciencia y Tecnología y ocho de la modalidad de Humanidades y Ciencias Sociales. En el caso de los conceptos, se estudia si su definición es operacional, estructural o mediante ejemplo. Los resultados indican que no hay grandes diferencias en la presentación de estos objetos matemáticos en los textos dirigidos a las dos modalidades de Bachillerato. Encontramos amplia variedad del número y tipo de propiedades presentadas y no se suele incluir la valoración de la bondad de ajuste o la construcción de modelos no lineales. Estos resultados proporcionan criterios para mejorar la presentación de la regresión en los textos de Bachillerato.

Palabras clave: *Regresión, Libros de texto, Bachillerato, Conceptos y Propiedades.*

Abstract

The aim of this paper is to characterize the presentation of regression in Spanish high-school textbooks. To achieve this aim we analyse the problem-fields, procedures, concepts and properties linked to regression in sixteen high Spanish textbooks, eight in the modalities Humanities and Social Sciences, and another eight directed to Science and Technology. We classify the concepts identified according to their definition (operational, structural or using examples). Results show no much difference between the textbooks directed to the two high school modalities. We found a wide variety of number and type of properties presented. The assessment of the goodness of fit and the building of non linear models is hardly included. These results provide criteria to improve the presentation of regression in high school textbooks.

Keywords: *Regression, Textbooks, High school, Concepts and Properties.*

Fecha de recepción: 9 de enero de 2015

Fecha de aprobación: 15 de abril de 2015

1. INTRODUCCIÓN

La regresión es un concepto fundamental por su utilidad en la modelización y predicción de fenómenos diversos (Engel y Sedlmeier, 2011). Junto a la correlación, extiende la dependencia funcional a variables aleatorias. En España se estudia en el primer curso de Bachillerato, siendo el primer tema que se incluye en el bloque de contenidos sobre Estadística y Probabilidad. Dicho bloque es uno de los cuatro que configuran la asignatura de matemáticas en la modalidad de *Ciencias y Tecnología* y de los tres que la conforman en la modalidad de *Humanidades y Ciencias Sociales*. Los alumnos de este curso tienen una edad entre 16 y 17 años, siendo obligatoria la enseñanza de las matemáticas en ambas modalidades.

Este contenido aparece en las orientaciones curriculares de España, según el Ministerio de Educación y Ciencia (MEC), de forma similar en ambas modalidades en la forma siguiente: “*Distribuciones bidimensionales. Relaciones entre dos variables estadísticas. Regresión lineal*” (MEC, 2007: 45449). También se especifica; “*Se pretende comprobar la capacidad para estimar y asociar los parámetros relacionados con la correlación y la regresión con las situaciones y relaciones que miden*” (MEC, 2007: 45450).

La investigación didáctica sobre este tema se centra preferentemente en la correlación, describiéndose dificultades como no apreciar la correlación inversa, tener un sentido determinista o local de la correlación o identificar correlación con causalidad (Estepa y Batanero 1995; Estepa, 2008; Zieffler y Garfield, 2009). No ha habido tanto interés por la regresión o por la forma en que se enseña o se presenta en los libros de texto, que son un paso intermedio en la enseñanza, pues desarrollan las directrices curriculares (Herbel, 2007).

El libro de texto es un soporte curricular, a través del cual se vehicula el conocimiento que las instituciones educativas han de transmitir (Escolano, 2009). La presentación del tema y los problemas propuestos se basan con frecuencia en los presentados en ellos (Shield y Dole, 2013). Ello explica el interés que recibe la investigación sobre textos de matemáticas por la comunidad internacional (e.g., Fan y Zhu, 2007). En España, los libros de texto son escritos por equipos de profesores con prestigio entre sus compañeros, deben ajustarse a los contenidos fijados por el currículo y son revisados por el Ministerio de Educación. Tanto alumnos como profesores se apoyan en el libro de texto que, con frecuencia, determina el contenido de la materia. Los profesores de matemáticas son en su mayoría licenciados en matemáticas; en algunos casos licenciados en ciencias o ingenieros. En los últimos años, para ser profesor de matemáticas se requiere también poseer un Máster específico orientado a la formación de estos profesores.

El objetivo de este trabajo es analizar los objetos matemáticos asociados a la regresión en los libros de texto del Bachillerato español. Forma parte de un proyecto de investigación más amplio, en el que se analiza tanto la regresión como la correlación en los mismos textos utilizados en el estudio; posteriormente, se utilizan los resultados obtenidos para diseñar talleres formativos dirigidos al profesorado (Gea, 2014). Algunos resultados del mismo proyecto se han publicado en Gea, Batanero, Cañadas y Contreras (2013), Gea, Batanero, Arteaga, Cañadas y Contreras (2014) y Gea, Batanero, Fernandes y Gómez (2014).

2. MARCO TEÓRICO

Nos basamos en el enfoque Onto-semiótico (EOS) (Godino y Batanero, 1994; Godino, 2002; 2012), donde, las matemáticas se conciben como respuesta a situaciones problemáticas extra o intra matemáticas. En este marco teórico el término “objeto matemático” tiene un significado muy amplio; se le asigna una entidad cultural y puede verse desde el punto de vista institucional (por ejemplo, una institución de enseñanza) o personal (de cada persona). Los objetos matemáticos surgen de las prácticas (acciones u operaciones) para resolver campos de problemas específicos. Según los autores, un objeto matemático es cualquier entidad o cosa a la cual nos referimos, o que interviene de algún modo en la actividad matemática. El EOS clasifica los objetos matemáticos en *situación-problema, lenguaje, conceptos, proposiciones, procedimientos, y argumentos*.

En el marco teórico, el significado de un objeto, que sería el conjunto de prácticas asociadas al campo de problemas de donde surge dicho objeto; puede variar en distintas instituciones y por tanto, tiene un carácter relativo. De ello se deduce la importancia de analizar la enseñanza y los recursos utilizados en ella, como los libros de texto, ya que la progresiva construcción del significado personal del estudiante depende directamente de los objetos que se describan y apliquen en su quehacer matemático, y el libro de texto se utiliza en este sentido.

En este trabajo nos centramos en las situaciones-problemas, procedimientos, conceptos y propiedades que intervienen en el trabajo con la regresión. Los dos primeros constituyen el conocimiento procedimental y los dos últimos el conceptual. Ambos tipos de conocimiento son polos de un continuo, siendo el conocimiento conceptual más flexible y generalizable, pues incluye la comprensión implícita o explícita de los principios de un dominio dado y sus interrelaciones (Rittle-Johnson, Siegler y Alibali, 2001).

También tendremos en cuenta las ideas de Sfard (1991), quien describe un concepto como constructo correspondiente al universo matemático formal y diferencia entre su definición estructural (describiendo sus condiciones o propiedades) y la operacional (cuando se define mediante una expresión o fórmula). Skemp (1993) indica que los nuevos conceptos no pueden ser comunicados simplemente a través de una definición, sino que sería necesario proporcionar una colección apropiada de ejemplos, apoyados en los conceptos que ya se conocen. Joyce y Weil (1996) consideran varios aspectos a tener en cuenta en la enseñanza de un concepto: (a) su nombre o etiqueta, que puede ser una palabra o símbolo; (b) sus atributos esenciales, o propiedades específicas, que lo diferencian de otros conceptos; (c) los ejemplos y contraejemplos del concepto; y (d) la regla de definición, que es una afirmación sintética y precisa elaborada a partir de los atributos esenciales que lo caracterizan. Vergnaud (1990), por su parte, considera el concepto matemático como una tripleta formada por el conjunto de situaciones que lo hacen significativo; el conjunto de invariantes que lo constituyen; y las representaciones simbólicas asociadas. En este trabajo analizaremos los invariantes asociados al concepto, dados por su definición, y las propiedades que se asocian a éstos; las representaciones fueron analizadas en un trabajo previo (Gea, Batanero, Fernandes y Gómez, 2014).

3. ANTECEDENTES

Una investigación destacada entre las relacionadas con la comprensión de la regresión es la de Estepa (1994), quien señala como puntos importantes la comprensión de la diferencia entre variable dependiente e independiente. Sánchez Cobo (1999) identifica la confusión de muchos estudiantes entre las dos rectas de regresión y sus dificultades en la interpretación de los coeficientes de regresión, y su relación con la pendiente de la recta y tipo de correlación. Eisenhauer (2003) encuentra estudiantes que diferencian los casos en que debe aplicarse o no una recta de regresión que cruce el origen de coordenadas. Otro error frecuente al interpretar la regresión, en una situación en que se realizan dos pruebas sucesivas sobre los mismos sujetos, es interpretar como efecto de un tratamiento el hecho de que cualquier valor atípico en la primera medida suele tender hacia el centro de la distribución en la segunda (efecto de regresión hacia la media; ver Engel y Sedlmeier, 2011). Por otra parte, Agnelli, Konic, Peparrelli y Flores (2009) indican que algunas de estas dificultades pueden estar ligadas al estudio previo de la función lineal que se generaliza excesivamente.

La investigación sobre libros de texto de matemáticas es amplia, siendo menor el análisis de contenidos de estadística, aunque encontramos algunos ejemplos como los de Cobo y Batanero (2004) y Díaz-Levicoy (2014). Respecto a la regresión, destacamos las investigaciones de Sánchez Cobo (1999), quien estudia las definiciones de conceptos relacionados con la regresión en once libros de texto de Bachillerato, publicados entre 1977 y 1990, clasificándolas según se definan explícitamente o bien mediante una fórmula, o bien una mezcla de las anteriores. En su estudio a veces no diferencia entre conceptos y propiedades por lo que éste será un punto original de nuestro trabajo, además de presentar una actualización de su estudio a libros publicados recientemente.

Lavalle, Micheli y Rubio (2006) analizan la enseñanza de la regresión en siete textos argentinos de Bachillerato, considerando los conceptos y procedimientos asociados, así como sus relaciones. Encuentran distintos niveles de profundidad en el tratamiento del tema. Solo un texto define el concepto de recta de regresión; cuatro incluyen la ecuación de la recta, siendo únicamente dos los que además utilizan funciones de ajuste distintas a la lineal. En la mayoría de los textos analizados por las autoras se trata también la estimación de la variable dependiente utilizando la función de ajuste.

Por su parte, Gea y cols. (2013) realizan un estudio empírico de algunas variables, como tipo e intensidad de la correlación en los problemas de correlación y regresión propuestos en una muestra de libros textos de Bachillerato de Ciencias Sociales. En Gea, Batanero, Fernandes y Gómez (2014) analizan, en estos mismos libros de texto, las representaciones gráfica y tabular de la distribución bidimensional; tema que el estudiante aprende antes de comenzar el estudio de la regresión. En Gea, Batanero, Arteaga, Cañadas y Contreras (2014) se estudia el lenguaje simbólico, expresiones algebraicas y verbales que utilizan los libros de texto en los temas de correlación y regresión. En este artículo se complementan los anteriores, estudiando las situaciones-problemas, procedimientos, conceptos y propiedades asociados a la regresión en los textos mencionados y otros ocho textos de Bachillerato de Ciencia y Tecnología.

4. METODOLOGÍA

Se analizaron 16 libros de texto de primer curso de Bachillerato en la modalidad de Humanidades y Ciencias Sociales y de Ciencia y Tecnología, posteriores al Decreto de Bachillerato (MEC, 2007). Los libros se eligieron por ser los más utilizados en la enseñanza pública en la Comunidad Autónoma de Andalucía (después de una consulta a la Consejería de educación sobre el tema) y estar publicados en editoriales de gran tradición y prestigio. En el Anexo 1 se presentan los libros analizados junto con un código utilizado a lo largo del trabajo; dicho código será

H1, H2, etc. para los textos correspondiente a Humanidades y Ciencias Sociales y T1, T2, etc. para los textos dirigidos a Ciencia y Tecnología.

Resaltamos el hecho de que todos han estado vigentes hasta la fecha. De cada editorial se tomaron los libros dirigidos a las dos modalidades de Bachillerato citadas. Se realizó un análisis de contenido, que difiere de otras técnicas de estudio documental (por ejemplo, del método histórico), porque sustituye en lo posible las interpretaciones y subjetividad del estudio de documentos por procedimientos estandarizados, con el fin de convertir en datos los contenidos analizados en los documentos (León y Montero, 2002). Se siguieron los siguientes pasos:

- En primer lugar se categorizan todos los problemas presentados, identificando dos tipos principales de situaciones problemas.
- Se identifican los procedimientos presentados, bien porque se introduzcan explícitamente, explicando el procedimiento paso a paso, o bien porque sea necesario para resolver los problemas propuestos.
- Se identifican todos los conceptos implicados en el tratamiento de la regresión. Para cada uno de ellos se analizó la forma en que se describe o presenta el concepto, que puede ser mediante ejemplos (E) o mediante una definición explícita formal (F); que de acuerdo a Sfard (1991) puede ser operacional (O) o bien estructural (S), o bien mezclando varios de estos tipos de definiciones. Se ha tenido también en cuenta si los ejemplos se proponen antes de la definición, como es sugerido por Skemp (1993).
- Seguidamente se consideran las propiedades asociadas a cada uno de los anteriores conceptos, que no fueron consideradas en trabajos previos.

5. RESULTADOS

Con la metodología descrita se analizaron cada uno de los textos de la muestra, identificando la presencia o ausencia de los objetos matemáticos identificados. A continuación se presentan estos resultados, describiendo las categorías consideradas e introduciendo un ejemplo, cuando sea necesario, para la mejor comprensión. Se finaliza cada una de las secciones con tablas en las que se resumen los resultados y con las conclusiones respecto a la forma en que los textos presentan los campos de problemas, procedimientos, conceptos y propiedades ligados a la regresión.

5.1. CAMPOS DE PROBLEMAS

En nuestro marco teórico, los problemas son el origen o razón de ser de la actividad matemática; se postula que el resto de objetos matemático emergen de las prácticas, personales o institucionales, realizadas para resolverlos (ver, por ejemplo, Godino, 2002; Godino y Batanero, 1994). Al analizar la actividad matemática interesan, más que ejemplos particulares de problemas, clases o campos de problemas, relacionados con cada objeto matemático. El concepto de “problema” es muy amplio, pues incluye cualquier situación o tarea que requiera actividad matemática. Por ello, en el análisis de los libros de texto, hemos englobado como “problemas”, tanto los problemas en sí, como los ejercicios, tareas o actividades. El análisis de todos ellos sirvió para identificar dos clases o campos de problemas principales referidos a la regresión, que se ejemplifican en la Figura 1 y que se describen a continuación, asignando a cada uno de ellos un código: CP1 y CP2.

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

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

- ¿Puede establecerse, de forma fiable, una fórmula que relacione el año con el número de nacimientos?
- ¿Cuántos nacimientos pueden estimarse en 2008?
¿Y en 2010? ¿Qué puede estimarse para 2050?
- ¿Es fiable esta última estimación?
Razona la respuesta.

Figura 1. Ajuste de un modelo a los datos y estimación mediante el modelo de regresión. ([H7]: 262)

CP1. Ajuste de un modelo a los datos bivariantes. Los problemas que hemos incluido en esta categoría piden a los estudiantes ajustar una función a los datos bivariantes, obteniendo su expresión algebraica (pregunta (a) en la Figura 1). Otras veces se sugiere trazar en forma aproximada la recta de regresión sobre un diagrama de dispersión.

En menor medida, encontramos algunas situaciones donde se pide comprobar propiedades del modelo de regresión lineal, asignar una recta entre varias dadas a un diagrama de dispersión, o determinar el ángulo que forman las dos rectas de regresión. En los textos analizados, estos problemas usualmente sólo proponen el modelo de regresión lineal, que es el más sencillo y asequible para los estudiantes.

CP2. Realizar estimaciones de un valor de la variable dependiente, a partir de un modelo de regresión. Una vez calculada la función de ajuste a la distribución bidimensional, los textos proponen problemas estimación; por ejemplo, el mostrado en el apartado (b) de la Figura 1. Son muchos los problemas donde, una vez calculada la función de ajuste, se pide realizar una estimación en valores de la variable independiente incluidos (interpolaciones) y excluidos (extrapolaciones) en el rango de los datos. Sánchez Cobo (1999) encontró en su análisis este mismo tipo de problemas. Al igual que Lavalle, Micheli y Rubio (2006), algunos textos son imprecisos ([H6], [T6], [H4] y [T4]), porque no señalan que el valor estimado a partir de la ecuación de regresión es un valor esperado o promedio. Hemos incluido en esta categoría las tareas que piden al estudiante analizar la bondad del ajuste del modelo de regresión (pregunta (c) en la Figura 1), que el alumno ha de resolver mediante el cálculo del coeficiente de determinación. Otro ejemplo de este campo de problemas es la siguiente tarea: “¿Cuál sería la fiabilidad de un ajuste bidimensional con $r = 0,7$?” ([T6]: 187).

En la Tabla 1 se resume el análisis realizado de los campos de problemas. En todos los libros se presentan los dos campos de problemas descritos, que aunque con frecuencia similar, es algo más alta para el campo P1, que supone, aproximadamente el 52,6% del total de problemas presentados en los textos de Humanidades y Ciencias Sociales y el 52,7% en los de Ciencias y Tecnología. Estos porcentajes varían según los textos, pues unas editoriales proponen más problemas tipo CP1 (ajustar una recta a los datos): [H1], [H4], [H5], [H7], [H8], [T1], [T4], [T5], [T7], [T8], y el resto al contrario. No hay mucha diferencia entre los dos textos de la misma editorial.

Tabla 1. Frecuencias (y porcentaje) de campos de problemas por texto

Campo de problemas	H1	H2	H3	H4	H5	H6	H7	H8	Total
CP1	44(57,9)	26(39,4)	26(42,6)	26(59,1)	50(51,0)	28(45,2)	59(54,6)	60(65,9)	319(52,6)
CP2	32(42,1)	40(60,6)	35(57,4)	18(40,9)	48(49,0)	34(54,8)	49(45,4)	31(34,1)	287(47,4)
Total	76	66	61	44	98	62	108	91	606
	T1	T2	T3	T4	T5	T6	T7	T8	Total
CP1	45(58,4)	26(39,4)	22(47,8)	26(59,1)	53(54,1)	28(45,2)	56(54,9)	47(58,8)	303(52,7)
CP2	32(41,6)	40(60,6)	24(52,2)	18(40,9)	45(45,9)	34(54,8)	46(45,1)	33(41,3)	272(47,3)
Total	77	66	46	44	98	62	102	80	575

5.2. PROCEDIMIENTOS

Los procedimientos que describen los textos para resolver los anteriores problemas de ajuste y predicción se describen a continuación, identificados por un código P1 a P6.

P1. Ajuste de la recta de mínimos cuadrados. El modelo de regresión usualmente presentado es el ajuste lineal, utilizando el criterio de mínimos cuadrados (es decir, la recta que hace mínima la suma de los cuadrados de las diferencias entre los valores observados de la variable bidimensional y los estimados por la recta). Se presentan dos procedimientos de cálculo, que requieren el cálculo previo de las medias, varianzas, desviaciones típicas y covarianza. El primer procedimiento es la ecuación punto-pendiente, o ecuación de la recta que pasa por el centro de gravedad y tiene como pendiente el coeficiente de regresión de Y sobre X (análogamente para la recta de regresión de X sobre Y):

$$\text{Coeficiente de regresión de } Y \text{ sobre } X: b_{yx} = \frac{\sigma_{yx}}{\sigma_x^2} = r \frac{\sigma_y}{\sigma_x}$$

$$\text{Recta de regresión de } Y \text{ sobre } X: y - \bar{y} = b_{yx} \cdot (x - \bar{x})$$

Otros textos, plantean la ecuación de la recta a partir de la ordenada en el origen (análogamente para la recta de regresión de X sobre Y):

Ordenada en el origen $a = \bar{y} - b_{yx} \cdot \bar{x}$; Recta de regresión de Y sobre X: $\hat{y} = a + b_{yx} \cdot x$

P2. Ajuste de otros modelos de regresión. Tan sólo [H8] y [T8] incluyen procedimientos para ajustar otros modelos de regresión. En primer lugar, mediante un ejercicio resuelto, explican el procedimiento desarrollado por Tukey para ajustar la recta de regresión respecto a la mediana, que se utiliza cuando los datos muestran una tendencia lineal fuerte, y aparecen datos atípicos. También, mediante un ejercicio resuelto, introducen el método de cambio de variable para ajustar una función no lineal a los datos (en [H8] los datos presentan una dependencia exponencial, y en [T8] una parabólica). También se pide comparar estos modelos de ajuste con el lineal y justificar la elección del mejor modelo, tarea que ayudará al estudiante a comprender mejor la regresión.

P3. Cálculo del coeficiente de determinación R^2 . Pocos textos incluyen el procedimiento de cálculo del coeficiente de determinación, que en el caso de la regresión lineal coincide con el cuadrado del coeficiente de correlación. Este coeficiente mide el porcentaje de la varianza de la variable dependiente que es explicada por el modelo de regresión. En caso de tomar su valor máximo (1) los datos de la variable dependiente se ajustarían perfectamente al modelo, siendo la relación entre las variables de tipo funcional. A pesar de su importancia, sólo en [H5] y [T5] se define este coeficiente y el método de cálculo, que es un complemento al estudio de la correlación, al permitir valorar la bondad del ajuste. Como vemos, en la presentación, se explican las propiedades de este coeficiente:

Si multiplicamos r^2 por 100, se obtiene el porcentaje de cambio de Y explicado por X. Así, si $r = 0$, los cambios en la variable X explican el 0% de los cambios en la Y, o sea, las variables X e Y son linealmente independientes. Y si $r = 1$ (o $r = -1$), la variación en la Y se explica al 100%, por la variación de la X; en este caso, las variables X e Y son linealmente dependientes. Fuera de estos casos límite, el porcentaje explicado es $100 \cdot r^2$ ([T5]: 363).

P4. Valorar la bondad del ajuste del modelo de regresión. El coeficiente de determinación permite valorar la bondad del ajuste del modelo de regresión, es decir el grado en que los puntos se aproximan al modelo (función) de regresión. Habrá un buen ajuste cuando los puntos estén próximos a la función elegida de ajuste. En algunos textos se valora el ajuste interpretando únicamente el coeficiente de determinación como proporción de la varianza, como hemos visto en el ejemplo anterior, pero no se indica si los puntos se aproximan o no a la función de ajuste.

P5. Determinar una predicción. En todos los textos se plantean procedimientos de estimación, utilizando la recta de regresión, que se reducen a la sustitución de valores en su ecuación. En su mayoría no se explicita de modo formal, sino mediante un ejercicio resuelto. En algunos textos, como [H1], [H2], [T1] y [T2], se indica que, si el coeficiente de correlación es alto, podría utilizarse una sola recta para establecer predicciones (las pendientes de las rectas de regresión son prácticamente las mismas, y poseen un punto en común: el centro de gravedad) pero, aunque esta propiedad sea cierta, se debe tener cuidado en que el estudiante no generalice indebidamente.

P6. Valorar la bondad de la predicción. Cuando se calculan estimaciones utilizando el modelo de regresión es importante que se razone sobre los resultados obtenidos, ya que para algunos valores de la variable independiente la predicción de la dependiente tendría un gran error. El procedimiento que se suele utilizar para analizar la fiabilidad de la estimación de valores es comprobar si los valores de la variable independiente que se utilizan para realizar la estimación son cercanos a la media, y si el valor del coeficiente de correlación está próximo a 1, en valor absoluto, como el ejemplo siguiente:

- Las predicciones realizadas a partir de una recta de regresión no son fiables si entre X e Y no hay un alto grado de correlación lineal, es decir, si r no es, en valor absoluto, cercano a 1.
- La fiabilidad de una recta de regresión es mayor cuanto mayor sea el número de datos considerados para calcularla. Las predicciones obtenidas para valores próximos al punto medio de la distribución son más fiables que las obtenidas para valores muy alejados. ([H3]: 227).

En la Tabla 2 observamos que en todos los textos se incluyen los procedimientos de ajuste de la recta de regresión por el método de mínimos cuadrados y los relativos al cálculo de predicciones y la valoración de fiabilidad de predicción. En cuanto a otros procedimientos, encontramos diferencias, principalmente en los textos de cada editorial; por ejemplo, el texto [T8] presenta procedimientos relativos al cálculo del coeficiente de determinación y la bondad del ajuste del modelo de regresión, mientras que en [H8] no se incluyen.

La valoración de la bondad del ajuste del modelo, así como el cálculo del coeficiente de determinación no se suele incluir en los textos analizados. Tan sólo [H5], [T5] y [T8] tratan ambos procedimientos, y en [H6] y [T6] la bondad del ajuste del modelo, sin el cálculo del coeficiente de determinación. En la investigación de Lavalle, Micheli y Rubio (2006), de siete libros analizados por estas autoras la ecuación de la recta de regresión aparece

únicamente en cuatro, la estimación de valores en cinco y la valoración de la bondad de las mismas en tres. En cuanto a procedimientos de ajuste de otros modelos de regresión no lineales, los resultados son similares a los nuestros.

Tabla 2. Procedimientos de análisis de regresión

Procedimientos	H1	H2	H3	H4	H5	H6	H7	H8
P1. Ajuste de la recta de regresión	x	x	x	x	x	x	x	x
P2. Ajuste de otros modelos de regresión								x
P3. Cálculo del coeficiente de determinación					x			
P4. Valorar la bondad del ajuste del modelo de regresión					x	x		
P5. Determinar una predicción	x	x	x	x	x	x	x	x
P6. Valorar la fiabilidad de la predicción	x	x	x	x	x	x	x	x

	T1	T2	T3	T4	T5	T6	T7	T8
P1. Ajuste de la recta de regresión	x	x	x	x	x	x	x	x
P2. Ajuste de otros modelos de regresión								x
P3. Cálculo del coeficiente de determinación					x			x
P4. Valorar la bondad del ajuste del modelo de regresión					x	x		x
P5. Determinar una predicción	x	x	x	x	x	x	x	x
P6. Valorar la fiabilidad de la predicción	x	x	x	x	x	x	x	x

5.3. CONCEPTOS

En primer lugar describiremos todos los conceptos implicados en el tratamiento de la regresión hallados en el tema, en cada uno de los textos analizados. Seguidamente analizamos la *forma* en que se describe o presenta el concepto, que puede ser mediante ejemplos (E), usualmente utilizando implícitamente una definición operacional, o mediante una definición explícita formal; que de acuerdo a Sfard (1991) puede ser operacional (O) o bien estructural (S), o bien mezclando varios de estos tipos de definiciones. Se ha tenido también en cuenta si los ejemplos se proponen antes de la definición, como es sugerido por Skemp (1993). Cada concepto tendrá un código C1 a C6.

C1. Variable dependiente e independiente. En la expresión $Y = f(x)$, los valores de la variable independiente x serán fijados; la variable dependiente Y es la que tratamos de predecir mediante un modelo, por ejemplo, la recta de regresión. Su diferencia ocupa un lugar central en la regresión (Estepa, 1994), con objeto de expresar una variable en función de otra. Una definición explícita se encuentra en [H8]: “*La variable dependiente es aquella que se quiere estimar, y la variable que se utiliza para ello se denomina variable independiente*” (p. 254). Implícitamente se describe este concepto al resaltar que existen dos rectas de regresión.

C2. Regresión. Se conoce como tal el método que permite elegir, dentro de una familia de funciones, la que mejor se ajusta a un conjunto de datos, siguiendo ciertos criterios; por ejemplo, la recta de regresión es, entre todas las funciones lineales, la que mejor se ajusta a los datos, cuando el criterio es minimizar la suma de los cuadrados de las desviaciones de los datos a la recta. Este concepto suele tratarse de modo implícito. Una definición explícita es: “*Al análisis que pretende determinar la curva que mejor aproxima un diagrama de dispersión se le llama regresión*” ([H3]: 226).

C3. Modelos de regresión. En regresión consideraremos un modelo a cualquier función de ajuste, pues, como todo modelo matemático permite representar en forma simplificada (mediante una ecuación) el fenómeno en estudio. El concepto de modelo es fundamental en el tratamiento de la regresión, pues permite diferenciar la función de ajuste (modelo) de los datos observados. Sin embargo, raramente aparece explicitado en los textos. En algunos casos, no se define pero queda implícito, por ejemplo:

Si existe una correlación fuerte entre las variables X e Y , el análisis de la regresión permite encontrar la ecuación de la función matemática que mejor se ajusta a la nube de puntos. Esta puede ser una recta, una parábola, una exponencial, una cúbica, etc. ([H4]: 226).

La situación es parecida a la descrita por Lavalley, Micheli y Rubio (2006), donde sólo dos de los siete textos incluyen esta idea. Otros textos plantean tareas donde el modelo de ajuste más indicado es no lineal, sin definir la idea de modelo, al igual que en cinco de los libros analizados por Sánchez Cobo (1999).

C4. Coeficientes de regresión lineal. Una vez decidido utilizar el modelo lineal de regresión, es necesario determinar los parámetros de la recta de regresión, es decir, la ordenada en el origen y la pendiente. Las pendientes de las rectas de regresión se conocen como coeficientes de regresión lineal. Existen dos coeficientes de regresión que corresponden a las pendientes de las dos rectas de regresión asociadas a cada conjunto de datos. Sánchez Cobo (1999) no los analiza. En nuestro estudio, algunos libros los definen en forma parcialmente correcta pues sólo definen el coeficiente de regresión de Y sobre X :

La recta que hace mínima la suma $\sum d_i^2$ tiene por ecuación: $y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$

Se llama recta de regresión de Y sobre X . A la pendiente, $\frac{\sigma_{xy}}{\sigma_x^2}$, se la llama coeficiente de regresión ([H1]: 230).

C5. Rectas de regresión. Todos los textos analizados la definen y describen el método de mínimos cuadrados. Se encuentran también enunciados imprecisos como el siguiente donde no se indica cuál es la variable dependiente y cuál la independiente.

La tabla adjunta da los alargamientos de una barra metálica por efecto de cambios en la temperatura. Calcular la recta de regresión y hacer algunas estimaciones ([H1], p.231).

Tan sólo [H8] alude a la influencia de los datos atípicos sobre la recta de regresión y sugiere en estos casos usar la metodología de Tukey para obtener la recta de regresión respecto a la mediana. En cuanto a la utilidad predictiva de la recta de regresión, sólo la mitad de los textos analizados la resaltan. Estos resultados coinciden con los de Sánchez Cobo (1999), quien señala que se suele incluir el procedimiento de ajuste a los datos (siete de los once textos) sin discutir su utilidad predictiva (tres de los once textos). Aparentemente, la finalidad de la recta de regresión es calcularla, sin utilizarla posteriormente.

C6. Bondad de ajuste. Coeficiente de determinación. Estas nociones, ya descritas, ayudan al estudiante a dotar de significado al análisis de regresión. A pesar de ello, los textos analizados no suelen definirlos; tan sólo encontramos estas definiciones en [H5] y [H6], siendo parcial en [H6], que define la bondad de ajuste, pero no el coeficiente de determinación. Por otro lado, se presenta en forma confusa la diferencia entre fiabilidad del modelo y fiabilidad, que no son nociones equivalentes. La primera se refiere a la fiabilidad global de la predicción y la segunda se reduce a la de un punto particular.

En la Tabla 3 se resume la forma de definición de estos conceptos, observando un predominio de definiciones acompañadas de ejemplos, siguiendo entonces las recomendaciones de Skemp (1993) y Joyce y Weil (1996). Las excepciones son los conceptos de coeficiente de regresión y variable dependiente e independiente, cuya definición no se suele acompañar de ejemplos.

La recta de regresión se define en todos los libros, acompañada de ejemplos; en la mitad de los textos estos ejemplos se presentan antes de la definición y en el resto después de la misma. Generalmente se introduce de modo estructural, describiendo sus propiedades y características (Sfard, 1991). Los ejemplos presentados se apoyan gráficamente en diagramas de dispersión, a los que se añade la recta de mínimos cuadrados y con algunas representaciones que resaltan la diferencia entre las coordenadas de los puntos observados y las estimadas por medio de la recta (Ver ejemplo en Figura 2). Estos gráficos permiten también mostrar la tendencia de la variación conjunta. Finalmente, se ofrece la definición operacional de la misma, introduciendo las ecuaciones de las rectas. Esta misma tendencia se observa en la investigación de Sánchez Cobo (1999), donde por lo general, la recta de regresión se introduce mediante ejemplos en que el diagrama de dispersión muestra la recta de ajuste a los datos. Además, se muestran los dos tipos de definición, estructural y operacional, mientras en otros conceptos se limitan a uno de estos tipos de definiciones.

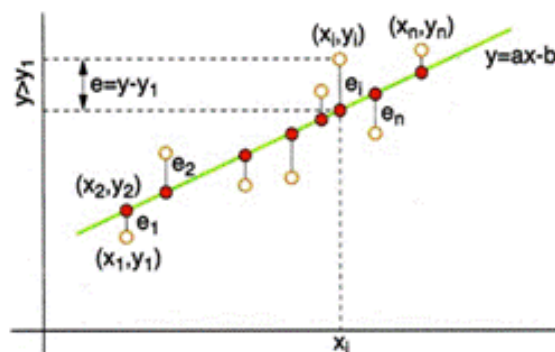


Figura 2. Coordenadas observadas de los puntos en el diagrama de dispersión y valores estimados mediante la recta de regresión ([T5]: 364)

Tabla 3. Conceptos asociados a la regresión y forma de introducción

Conceptos	H1	H2	H3	H4	H5	H6	H7	H8
C1.Variable dependiente e independiente					O			O
C2.Regresión			ES	SE				
C3.Modelos no lineales de regresión					SOE	SOE		
C4.Rectas de regresión	ESO	EOS	SO	SOE	SOE	SOE	SOE	SO
C5.Coefficientes de regresión	O	SO		O				
C6.Bondad de ajuste. Coef. de determinación					SOE	OE		

	T1	T2	T3	T4	T5	T6	T7	T8
C1.Variable dependiente e independiente					O			
C2.Regresión			ES	SE				
C3.Modelos no lineales de regresión					SOE	SOE		
C4.Rectas de regresión	ESO	EOS	SO	SOE	SOE	SOE	SOE	SO
C5.Coefficientes de regresión	O	SO		O				
C6.Bondad de ajuste. Coef. de determinación					SOE	OE		SOE

E= Definición mediante ejemplo; O = Definición operacional, S =Definición estructural

El concepto general de regresión o la variable dependiente e independiente apenas se definen, en unos pocos textos. Igual ocurre con la bondad del ajuste y los coeficientes de regresión; estos últimos con un tratamiento meramente operacional de la regresión, pudiendo así olvidar la importancia de sus aspectos estructurales y con ello, un aprendizaje más significativo.

5.4. PROPIEDADES

Los libros de texto suelen presentar atributos o propiedades de los conceptos definidos, cuya naturaleza es epistémica, y por tanto institucional. Se han encontrado las siguientes (designadas como P1 a P11):

P1. La recta de regresión hace mínima la suma de los cuadrados de las desviaciones de los puntos de la distribución bidimensional a la recta (diferencia de ordenadas o diferencia de abscisa, según la recta de regresión que se trate). Todos los textos la incluyen, generalizando así una propiedad de la media que los alumnos ya conocen (la media es el valor que minimiza la suma de desviaciones de los datos).

P2. Existen dos rectas de regresión diferentes. En todos los textos se indica implícitamente que hay dos rectas de regresión diferentes, aunque algunos lo hacen explícito. Como Sánchez Cobo (1999), creemos que si se presenta esta propiedad se podría prevenir que los estudiantes obtuvieran una de las rectas despejando la variable de la otra recta calculada previamente, lo cual es un error. Se indica generalmente que la estimación de cada variable se realiza con su correspondiente recta de regresión. Algunos textos señalan que si el coeficiente de correlación es próximo a 1 o -1, se puede utilizar una única recta de regresión para predecir cualquiera de las variables del estudio. Un ejemplo de esta propiedad es:

Las rectas de regresión de Y sobre X y de X sobre Y son distintas, por eso hay que saber qué variable es la dependiente, pues X e Y no son intercambiables. (Es posible que tenga sentido deducir la variable Y a partir de la X , mientras que deducir la X a partir de la Y carezca de significado) ([H5]:.258).

P3. Propiedades del coeficiente de determinación y varianza explicada. La principal propiedad que se introduce es que el coeficiente de determinación informa de la fiabilidad del ajuste lineal. Otros textos analizan la relación del coeficiente de determinación con la proporción de varianza de la variable dependiente explicada por la recta:

En ocasiones, con el fin de calcular la calidad o bondad del ajuste realizado mediante la recta de regresión y, por tanto, la fiabilidad de las predicciones que con ella se puedan realizar, se utiliza la expresión $(r^2 \cdot 100) \%$ que nos da el porcentaje en el que la variable Y se justifica por el valor de la variable X ([H6]: 185).

P4. La recta de regresión permite estimar valores de la variable dependiente a partir de la independiente. Las estimaciones obtenidas con la recta de regresión son aproximaciones, en términos de probabilidad, al valor real. No se suele resaltar el carácter aproximado de la estimación, aunque en la mayoría de los libros se usa de modo implícito, pues se indica que, al contrario que en la dependencia funcional, para cada valor de la variable independiente, corresponden varios valores de la dependiente, siendo el valor proporcionado por la recta de regresión el promedio de todos ellos. Así:

Las estimaciones siempre se realizan aproximadamente y en términos de probabilidad: es probable que si $x = x_0$ entonces y valga, aproximadamente, $\hat{y}(x_0)$ ([H1], p.230).

P5. La recta de regresión pasa por el centro de gravedad de la distribución. Es decir, el punto que tiene como coordenadas las medias de las dos variables (\bar{x}, \bar{y}) es un punto incluido en la recta y por tanto satisface su ecuación. En el estudio de Sánchez Cobo (1999), siete de los once textos citan esta propiedad, aunque dos lo hacen de modo implícito. En nuestro estudio aparece en la mayoría de los textos.

P6. Las estimaciones con la recta de regresión son mejores en valores cercanos a la media de la variable independiente. Un ejemplo es: “Las predicciones obtenidas para valores próximos al punto medio de la distribución son más fiables que las obtenidas para valores muy alejados” ([H3]: 227). No suele aparecer en muchos textos ya que, por lo general, la fiabilidad de la predicción se juzga por la proximidad del coeficiente de correlación a 1 o -1, aunque esta condición es insuficiente.

P7. La fiabilidad de la predicción con la recta de regresión aumenta con el tamaño de muestra. Se presenta en pocos de los textos analizados a pesar de su utilidad. Un ejemplo es: “La fiabilidad aumenta al aumentar los datos. Una recta obtenida a partir de pocos datos genera grandes riesgos, aunque r sea muy alto”. ([H5]: 260).

P8. Intensidad de correlación y ángulo de las rectas de regresión. Dependiendo de si las variables presentan dependencia funcional o independencia, las rectas de regresión son coincidentes o perpendiculares, respectivamente. No es muy habitual encontrar esta propiedad geométrica sobre el ángulo que forman las rectas en función del valor del coeficiente de correlación.

P9. Recta de regresión, covarianza y correlación. Se interpreta la covarianza según la proximidad de los puntos a la recta de regresión, y se relaciona el signo y los coeficientes de regresión, o pendiente de las rectas de regresión. En el estudio de Sánchez Cobo (1999) ninguno de los textos los incluye:

Según donde esté situado (x_i, y_i) respecto a (\bar{x}, \bar{y}) , el área $(x_i - \bar{x}) \cdot (y_i - \bar{y})$ será positiva o negativa. Si los puntos están próximos a una recta de pendiente positiva, los sumandos son casi todos positivos y la covarianza es grande ([H1]: 228).

De la pendiente de la recta de regresión se puede deducir igualmente el signo de la correlación:

Si la pendiente de la recta de regresión es positiva o negativa, la correlación se llama positiva o negativa, respectivamente ([H1]: 227).

P10. El producto de los dos coeficientes de regresión es r^2 . Esta proposición se encuentra en la mayoría de los textos, y también en los de Sánchez Cobo (1999).

P11. Relación entre las estimaciones con la recta de regresión y el coeficiente de correlación. La mayoría de los textos relacionan ambos conceptos como por ejemplo:

La fiabilidad de los cálculos obtenidos mediante las rectas de regresión será tanto mayor cuanto mayor sea el valor absoluto del coeficiente de correlación lineal r .

Si r está muy próximo a cero, no tiene sentido realizar ninguna estimación o previsión.

Si r está próximo a 1 o a -1, los valores reales serán, probablemente, próximos a nuestras estimaciones.

Si $r = 1$ o $r = -1$, los valores reales coincidirán con las estimaciones efectuadas ([H4]: 226)

En la Tabla 4 presentamos un resumen de las propiedades que se incluyen en el tema. Todos los textos menos [H4] y [T4] incluyen la propiedad de que la recta de regresión es aquella que minimiza la suma de cuadrados, que igualmente se presenta en la investigación de Sánchez Cobo (1999) en diez de los once textos que analiza. Igualmente, la mayoría de los textos indican la posibilidad de realizar predicciones a partir de ella (sólo la omiten [H8] y [T8]), y la existencia de dos rectas diferentes de regresión se presenta en la mayoría. Observamos diferencia en los textos, desde algunos que incluyen casi todas las propiedades citadas (como [H3] y [T3]) hasta otros como [H4] y [T4] que presentan pocas. Son también estos dos textos los únicos que omiten que el centro de gravedad es un punto de la recta de regresión y la relación entre la fiabilidad y proximidad del punto a dicho centro de gravedad.

La relación que menos se establecen se refieren a la discusión de la importancia del tamaño de la muestra en el ajuste del modelo predictivo. El resto de propiedades también aparece con baja frecuencia.

Tabla 4. Propiedades de la regresión

	H1	H2	H3	H4	H5	H6	H7	H8
PR1. Propiedad de mínimos cuadrados	x	x	x		x	x	x	x
PR2. Dos rectas de regresión diferentes	x		x	x	x	x	x	
PR3. Porcentaje de varianza explicada (r^2)					x	x		
PR4. Estimación usando la recta de regresión	x	x	x	x	x	x	x	
PR5. Centro de gravedad y recta de regresión	x	x	x		x	x	x	x
PR6. Fiabilidad de la estimación y centro de gravedad	x	x	x		x	x	x	x
PR7. Fiabilidad de la estimación y tamaño muestral			x		x			
PR8. Intensidad de correlación y ángulo de las rectas	x		x				x	x
PR9. Recta de regresión, correlación y covarianza	x		x		x	x	x	
PR10. Producto de los coeficientes de regresión	x		x					
PR11. Coeficiente de correlación y fiabilidad de predicción	x	x	x	x	x	x	x	

	T1	T2	T3	T4	T5	T6	T7	T8
PR1. Propiedad de mínimos cuadrados	x	x	x		x	x	x	x
PR2. Dos rectas de regresión diferentes	x		x	x	x	x	x	
PR3. Porcentaje de varianza explicada (r^2)					x	x		
PR4. Estimación usando la recta de regresión	x	x	x	x	x	x	x	
PR5. Centro de gravedad y recta de regresión	x	x	x		x	x	x	x
PR6. Fiabilidad de la estimación y centro de gravedad	x	x	x		x	x	x	x
PR7. Fiabilidad de la estimación y tamaño muestral			x		x			
PR8. Intensidad de correlación y ángulo de las rectas	x		x				x	x
PR9. Recta de regresión, correlación y covarianza	x		x		x	x	x	
PR10. Producto de los coeficientes de regresión	x		x					x
PR11. Coeficiente de correlación y fiabilidad de predicción	x	x	x	x	x	x	x	

No encontramos diferencias en los textos dirigidos a cada modalidad, salvo en [H8] y [T8], siendo éste último mucho más completo en cuanto a las propiedades que incluye. La razón principal es que este texto define el coeficiente de determinación, por lo que incluye más propiedades referidas a la regresión.

6. CONCLUSIONES

El análisis realizado permite cumplir el objetivo de este trabajo, que fue caracterizar la presentación de la regresión en los libros de texto españoles de Bachillerato. Para ello se han analizado los campos de problemas, procedimientos, conceptos y propiedades, comparando entre textos de la misma modalidad de Bachillerato y entre las dos modalidades.

Todos estos análisis completan los estudios previos. Además de actualizar el estudio de Sánchez Cobo (1999), nuestro trabajo estudia en forma diferenciada los conceptos y propiedades, encontrando dos que no aparecen en los textos analizadas por este autor: la relación entre signo de la covarianza, correlación y la pendiente de las rectas de

regresión y el hecho de que el producto de los dos coeficientes de regresión es igual al coeficiente de correlación al cuadrado. Lavalle, Micheli y Rubio (2006) no analizan las propiedades incluidas en los textos y además su estudio se ha realizado sobre textos argentinos. Otro punto original en nuestro trabajo es la clasificación de las definiciones los conceptos siguiendo la propuesta de Sfard (1991) (estructural u operacional).

Completamos nuestros trabajos previos sobre las variables incluidas en las situaciones problema de correlación y regresión (Gea, Batanero, Cañadas y Contreras, 2013), la distribución de datos bidimensional (Gea, Batanero, Fernandes y Gómez, 2014) y el lenguaje de la correlación y regresión (Gea, Batanero, Arteaga, Cañadas y Contreras, 2014). En ninguno de ellos se analizan los conceptos, propiedades y procedimientos ligados a la regresión.

Los resultados de nuestro estudio muestran que los textos contemplan los dos campos de problemas considerados en el análisis; siendo aproximadamente la mitad de los problemas, ejercicios y actividades propuestas dirigidas a ajustar un modelo de regresión y el resto a realizar estimaciones con dicho modelo. Consecuentemente, todos los textos incluyen los procedimientos de ajuste de la recta de regresión por el método de mínimos cuadrados y los necesarios para realizar cálculo de predicciones y valorar la fiabilidad de la predicción de valores aislados de los datos.

También hemos encontrado algunas carencias que resumimos a continuación y que nos permiten dar una serie de recomendaciones para mejorar la presentación de la regresión en los libros de texto.

- Una primera recomendación sería incluir los procedimientos de cálculo del coeficiente de determinación, así como su interpretación, tanto para valorar la bondad global del ajuste del modelo como para medir el porcentaje de varianza explicado por la regresión. Son muy pocos los libros que incluyen estos procedimientos, que consideramos fundamentales para comprender la utilidad de la regresión.
- Sugerimos también ampliar el número de conceptos que se definen explícitamente en los textos, donde hemos encontrado muy pocas definiciones, exceptuando la de la recta de regresión. Una explicación de esta escasez de definiciones podría ser la amplitud que se dedica en los textos al estudio de la correlación. Pensamos que es importante que los textos den la relevancia debida a la regresión, pues una vez aceptada una correlación entre las variables, la verdadera utilidad de la misma está en tratar de ajustar un modelo de regresión a los datos. Recomendamos consecuentemente incluir las definiciones de regresión, modelo de regresión, modelos no lineales y coeficientes de regresión, que sólo son incluidos en algunos libros.
- Recomendamos, cuando sea posible, dar relevancia a la definición estructural de los conceptos pues encontramos, en los conceptos que se definen, un énfasis en los aspectos operacionales, sin poner atención al significado y aspectos interpretativos.
- Sería útil que las propiedades de la regresión se expliciten en los textos. En general, los textos analizados cubren todas las propiedades importantes relativas a la regresión; la variedad y cantidad de estas propiedades y relaciones muestra la gran complejidad semiótica del tema. Sin embargo, la mayoría de las veces las propiedades no se discuten explícitamente, sino que únicamente se presentan implícitamente y pueden quedar ocultas para el estudiante. Así, por ejemplo, se presentan los cálculos para determinar cada una de las rectas de regresión, pero no se analizan a fondo las propiedades que relacionan la regresión con la correlación o el significado de los coeficientes de regresión.
- Sería importante, en particular, que los libros resalten la relación entre coeficiente de correlación y regresión lineal para que los estudiantes lleguen a comprender que es posible un coeficiente de correlación bajo o moderado en un diagrama de dispersión de dependencia alta pero no lineal (por ejemplo, parabólica).

Finalizamos indicando que el libro de texto es sólo una de las herramientas que el profesor usa en su labor docente; el impacto sobre el aprendizaje del estudiante depende no sólo del mismo libro, sino de los conocimientos previos del estudiante, así como de la forma que el profesor usa el texto en el aula (Lowe y Pimm, 1996). En este sentido, es importante que las propiedades y relaciones entre conceptos no se presenten de forma apresurada, sino que el profesor en la clase organice un debate y reflexión en torno a las mismas. Este será el modo de conseguir un aprendizaje significativo, para evitar posibles errores como no diferenciar las dos rectas de regresión (Sánchez Cobo, 1999) y las concepciones incorrectas identificadas por Estepa (1994).

Agradecimientos

Proyecto Proyecto EDU2013-41141-P (MEC) y grupo FQM126 (Junta de Andalucía).

Referencias

Agnelli, H., Konic, P., Peparelli, N. Z. y Flores, P. (2009). La función lineal obstáculo didáctico para la enseñanza de la regresión lineal. *Revista Iberoamericana de Educación Matemática*, 17, pp. 52-61.

- Cobo, B. y Batanero, C. (2004). Significados de la media en los libros de texto de secundaria. *Enseñanza de las Ciencias*, 22 (1), pp. 5-18.
- Díaz- Levicoy, D. (2014). Un estudio empírico de los gráficos estadísticos en libros de texto de educación primaria española. Tesis de Máster. Universidad de Granada.
- Engel, J. y Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En: C. Batanero, G. Burrill y C. Reading (editores). *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp. 247-258). Springer. New York.
- Escolano, A. (2009). El manual escolar y la cultura profesional de los docentes. *Tendencias Pedagógicas*, 14, pp. 169-180.
- Eisenhauer, J. G. (2003). Regression through the origin. *Teaching Statistics*, 25 (3), pp. 76-80.
- 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), pp. 257-270.
- Estepa, A. y Batanero, C. (1995). Concepciones iniciales sobre la asociación estadística. *Enseñanza de las Ciencias*, 13 (2), pp.155-170.
- Fan, L. y Zhu, Y. (2007). Representation of problem-solving procedures: A comparative look at China, Singapore, and US mathematics textbooks. *Educational Studies in Mathematics*, 66 (1), pp. 61-75.
- Gea, M. (2014). *La correlación y regresión en Bachillerato: Análisis de libros de texto y del conocimiento de los futuros profesores*. Tesis doctoral. Universidad de Granada.
- 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 (editores.). *Investigación en Educación Matemática XVII* (pp. 293-300). Sociedad Española de Investigación en Educación Matemática. Bilbao.
- Gea, M. M., Batanero, C., Arteaga, P., Cañadas, G. R., Contreras, J. M. (2014). Análisis del lenguaje sobre la correlación y regresión en libros de texto de bachillerato. *SUMA*, 76, pp. 37-45
- Gea, M. M., Batanero, C., Fernandes, J. A. y Gómez, E. (2014). La distribución de datos bidimensionales en los libros de texto de matemáticas de Bachillerato. *Cuadrante*. XXIII (2), pp. 147-172.
- Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 22 (2-3), pp. 237-284.
- Godino, J. D. (2012). Origen y aportaciones de la perspectiva ontosemiótica de investigación en Didáctica de la Matemática. En: A. Estepa, A. Contreras, J. Deulofeu, M. C. Penalva, F. J. García y L. Ordóñez (editores). *Investigación en Educación Matemática XVI* (pp. 49 - 68). Sociedad Española de Investigación en Educación Matemática. Jaén.
- Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, 14 (3), pp. 325-355

Herbel, B. A. (2007). From intended curriculum to written curriculum: Examining the "voice" of a mathematics textbook. *Journal for Research in Mathematics Education*, 38 (4), pp. 344-369.

Joyce, B. y Weil, M. (1996). *Models of teaching*. Boston: Allyn & Bacon.

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), pp. 383-406.

León, O. G. y Montero, I. (2002). *Métodos de Investigación*. McGraw Hill. Madrid.

Lowe, E. y Pimm, D. (1996). 'This is so': a text on texts. En: A. Bishop, K. Clements, C. Keitel, J. Kilpatrick y C. Laborde (editores), *International Handbook of Mathematics Education*, (pp. 371-410). Dordrecht: Kluwer.

MEC (2007). *Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura de Bachillerato y se fijan sus enseñanzas mínimas*. Ministerio de Educación y Ciencia. Madrid.

Rittle-Johnson, B., Siegler, R. S., y Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, 93 (2), pp. 346-362.

Sánchez Cobo, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral no publicada. Universidad de Granada.

Skemp, R. R. (1993). *The psychology of learning mathematics*. London: Penguin Books.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22 (1), pp. 1-36.

Shield, M. y Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics*, 82 (2), pp. 183-199.

Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques*, 10 (2-3), 133-170.

Zieffler, A. y Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8 (1), pp. 7-31.

Anexo 1: Textos utilizados en el análisis

[H1] Colera, J., Oliveira, M.J., García, R. y Santaella, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Grupo Anaya.

[H2] 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.

[H3] Vizmanos, J. R., Hernández, J. y Alcaide, F. Moreno, M. y Serrano, E. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Grupo SM.

[H4] Arias, J. M. y Maza, I. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Grupo Editorial Bruño.

[H5] Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santon, D. y de Vicente, M. (2008). *Matemáticas aplicadas a las Ciencias Sociales I*. Madrid: Santillana Educación.

[H6] Martínez, J. M., Cuadra, R., Heras, A. (2008). *Matemáticas aplicadas a las Ciencias Sociales. 1º Bachillerato*. Madrid: McGraw-Hill.

[H7] Bescós, E. y Pena, Z. (2008). *Matemáticas aplicadas a las Ciencias Sociales. 1º Bachillerato*. Vizcaya: Oxford University Press.

[H8] Monteagudo, M. F. y Paz, J. (2008). *1º Bachillerato. Matemáticas aplicadas a las Ciencias Sociales*. Zaragoza: Luis Vives.

[T1] Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). *Matemáticas I*. Madrid: Grupo Anaya.

[T2] Arias, J. M. y Maza, I. (2011). *Matemáticas I*. Madrid: Grupo Editorial Bruño.

[T3] Biosca, A., Doménech, M., Espinet, M. J., Fandos, M. J. y Jimeno, M. (2008). *Matemáticas I*. Barcelona: Guadiel - Grupo Edebé.

[T4] Monteagudo, M. F. y Paz, J. (2008). *1º Bachillerato. Matemáticas. Ciencias y Tecnología*. Zaragoza: Edelvives (Editorial Luis Vives).

[T5] Martínez, J. M., Cuadra, R., Barrado, F. J. (2007). *Matemáticas 1º Bachillerato*. Madrid: McGraw-Hill.

[T6] Bescós, E. y Pena, Z. (2009). *Matemáticas. 1º Bachillerato*. Navarra: Oxford University Press España.

[T7] Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santos, D. y de Vicente, M. (2008). *Matemáticas 1º Bachillerato*. Madrid: Santillana Educación.

[T8] Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008). *Matemáticas I*. Madrid: Ediciones SM.