

## Definiciones asociadas a la distribución de datos bidimensionales en textos españoles de bachillerato

M. Magdalena Gea<sup>1</sup>, Carmen Batanero<sup>1</sup>, J. António Fernandes<sup>2</sup> y Emilse Gómez<sup>3</sup>

<sup>1</sup>Universidad de Granada, mmgea@ugr.es, batanero@ugr.es

<sup>2</sup>Universidade do Minho, jfernandes@ie.uminho.pt

<sup>3</sup>Universidad Nacional de Colombia, egomezt@unal.edu.co

**Resumen.** *Se estudia la presentación de la distribución de datos bidimensionales en ocho libros de texto españoles de bachillerato. Se analizan las definiciones de variable estadística y distribución bidimensional, frecuencia y distribución marginal y condicional. Para cada una se analiza el modo (formal o mediante ejemplo) en que se definen, el lugar del tema en que se incluye la definición, y su uso a lo largo del estudio de la estadística bidimensional. La presentación es, a veces, incompleta o parcialmente correcta y otras se equiparan conceptos que no son equivalentes. Todo ello puede repercutir en el aprendizaje de los estudiantes a los que van dirigidos estos textos.*

**Palabras clave:** distribución estadística bidimensional; libros de texto; bachillerato; definición.

### 1. Introducción

La delimitación de la distribución en un estudio estadístico es primordial para el análisis posterior, en particular, para el estudio de la relación entre dos variables, o la predicción de una a partir de la otra, es decir, el estudio de la correlación y regresión (Crocker, 1981; Estepa et al., 2012; Gea, 2012; Moritz, 2004). Como señala Estepa (2007, p. 126):

En el estudio de la relación entre dos variables es de sumo interés distinguir si las dos variables constituyen una distribución bidimensional o no. Si las variables constituyen una distribución bidimensional se pueden realizar estudios de correlación y regresión, en caso contrario la correlación y regresión no tendrían sentido.

Este tema adquiere gran relevancia en estadística, y se incluyen en España en el primer curso de bachillerato (estudiantes de 16 años) de las modalidades de *Ciencias y Tecnología*, y *Humanidades y Ciencias Sociales* (MEC, 2007), con contenidos similares. Así, para el estudio de de datos bidimensionales en la modalidad de *Humanidades y Ciencias Sociales*, se concretan los siguientes contenidos: “*Distribuciones bidimensionales. Interpretación de fenómenos sociales y económicos en los que intervienen dos variables a partir de la representación gráfica de una nube de puntos. Grado de relación entre dos variables estadísticas. Regresión lineal. Extrapolación de resultados.*” (MEC, 2007, p. 45475). Sin embargo, su enseñanza no es simple, pues es necesario comprender los diferentes tipos de frecuencias y

distribuciones unidimensionales asociadas a la distribución bidimensional, que algunos alumnos confunden (Estepa y Batanero, 1995; Estepa, 2008). Más concretamente, en el estudio de Estepa (2007) sólo el 52% de los alumnos, elige en un ítem de opción múltiple la definición correcta de distribución bidimensional. Un porcentaje elevado (33%) indican que una distribución bidimensional consiste en dos conjuntos diferentes de datos relacionados, sin percatarse de que los datos pueden provenir de individuos de la misma muestra.

Algunos autores han analizado la presentación de la correlación y regresión en libros de texto (Lavallo, Micheli y Rubio, 2006; Sánchez Cobo, 1999; Sánchez Cobo, Estepa y Batanero, 2000), pero no se centran específicamente en las definiciones de las variables y distribuciones bidimensionales y los tipos de frecuencias asociadas.

El objetivo de este trabajo es completar los anteriores, analizando estas definiciones en una muestra de libros de texto de bachillerato. En lo que sigue analizamos los fundamentos, métodos y resultados del estudio, finalizando con algunas conclusiones.

## **2. Fundamentos**

### *2.1. Marco teórico*

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

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

Godino (2003) indica que un objeto matemático se caracteriza por su definición y el enunciado de sus propiedades; pero estas pueden variar según las distintas instituciones en que se trate, y por tanto, hemos de concederles un carácter relativo. Puesto que las *definiciones de conceptos* son evocadas por el estudiante cuando se enfrenta a una situación problema, es importante analizar el tratamiento de éstas en la enseñanza, ya

que la progresiva construcción del significado de estas nociones depende directamente de los conceptos que se definan y utilicen.

## 2.2. Antecedentes

Aunque hay una amplia investigación sobre los libros de texto de matemáticas, esta tradición es mucho menor en el caso de la estadística y probabilidad, donde encontramos algunos ejemplos como los de Ortiz, Batanero y Serrano (2001), Azcárate y Serradó (2006) y Cobo y Batanero (2004).

El primer antecedente relacionado con nuestro trabajo es el de Sánchez Cobo (1999), quien analiza la correlación y regresión en once libros de texto españoles de bachillerato publicados entre 1987 y 1990. Como consecuencia, ofrece una taxonomía de definiciones y un análisis de las demostraciones, desde el punto de vista de la función que realizan y las componentes que la integran. Muestra una tendencia formal en la presentación del tema, y el uso mayoritario de ejemplos basados en representaciones gráficas, así como un fuerte sesgo en los ejemplos presentados hacia la correlación positiva. Más recientemente, Lavalle et al. (2006) analizan la correlación y regresión en siete libros de texto argentinos de bachillerato, observando un enfoque mayoritariamente socio-constructivista, con un nivel de profundidad adecuado, donde también se plantean más actividades bajo una asociación directa que inversa.

## 3. Metodología

En este estudio se analizaron ocho libros de texto, todos ellos correspondientes al currículo actual de Bachillerato en la modalidad de Humanidades y Ciencias Sociales (MEC, 2007). Estos textos (ver Tabla A1 en Anexo) actualmente se utilizan en el aula, y corresponden a las editoriales de mayor tradición y prestigio, siendo los más utilizados en la enseñanza pública en Andalucía.

En ellos se analizaron las definiciones de los conceptos asociados a la distribución de datos bidimensionales, del que parte el trabajo con la correlación y regresión, considerando las siguientes variables:

- V1: *Concepto definido*. En primer lugar se han clasificado todos los objetos matemáticos relacionados con la distribución de datos bidimensionales que aparecen en los textos, asignando a cada uno una categoría que se usará en las tablas de resultados del análisis. Sobre cada una de estas categorías de la variable primaria se analizan tres variables adicionales en las definiciones:
- V2: *Momento* en que se presenta la definición en el libro de texto, que puede ser al

inicio del tema (I), al final del mismo, es decir, después de haberlo utilizado en ejemplos o problemas (F) o se define varias veces a lo largo del tema (VM).

- V3: *Forma* de presentación del concepto, que puede ser mediante ejemplos (E), mediante una definición formal (F) o bien una mezcla de ambos: proponiendo ejemplos y luego definiendo formalmente el concepto (EF), o al contrario (FE).
- V4: *Uso* en el tema, esto es, si el uso del concepto es continuado en el tema (S), si se usa poco (P), o se define pero no se usa (N).

En lo que sigue presentamos y discutimos los resultados. Para la variable principal (V1) se realiza un análisis cualitativo, mostrando cuando sea preciso ejemplos de la forma en que se define en los textos. Para el resto de variables, debido a la limitación de espacio, nos restringimos a la presentación y comentarios de tablas de resultados.

## **4. Resultados y discusión**

### *4.1. Conceptos considerados (VI)*

En primer lugar presentamos las categorías de conceptos identificados en nuestro análisis, algunas de las cuáles se han subdividido en categorías secundarias.

*C1. Variable estadística y distribución bidimensional.* Cuando realizamos un estudio estadístico, en cada elemento que constituye la muestra se toman datos de una o varias *variables*. Cada una de estas variables es una característica que se pretende investigar, y está determinada por los valores que ha tomado en los distintos individuos. Si para cada individuo se consideran dos variables, tendremos una variable estadística bidimensional; el conjunto de todos sus valores y frecuencias forma la distribución bidimensional. En esta categoría hemos analizado las definiciones de variable estadística bidimensional, frecuencia doble y distribución bidimensional.

*C11. Variable estadística bidimensional.* Sólo cinco de los ocho textos analizados ([H3], [H4], [H6], [H7] y [H8]) presentan esta definición, mientras que en el resto encontramos la definición de distribución bidimensional. Todos los textos introducen uno de estos dos conceptos al inicio del tema, aunque el modo de definirla varía de un texto a otro, siendo frecuente introducirla con ejemplos. Consideramos correctas estas definiciones cuando se indica que la variable bidimensional se conforma de dos variables estadísticas sujetas a un mismo estudio, incluyendo también la notación  $(X, Y)$ , como el ejemplo mostrado en la Figura 1.

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

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

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

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

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

Figura 1. Definición de variables estadística bidimensional con ejemplo ([H4], p. 218).

En este sentido, todas las definiciones encontradas en los textos son correctas, salvo la siguiente, que limita el concepto de variable estadística bidimensional a variables cuantitativas ([H8], p.248):

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

*C12. Frecuencias dobles.* Introducida la variable estadística bidimensional, se definen los diferentes tipos de frecuencia (doble, marginal y/o condicionada), absolutas o relativas. Todos los textos incluyen la definición de frecuencia doble, salvo [H6], generalmente explicando la forma de cálculo de la misma, es decir en forma operacional (Sfard, 1991). En el texto [H3] (p. 217) se puede leer:

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.

Las definiciones suelen ser parcialmente correctas ([H1], [H2], [H3], [H5]), al no incluir la notación, aunque se describan propiedades asociadas (definición estructural) o su forma de cálculo (definición operacional). Así, por ejemplo, en [H3] no se indica la notación de frecuencia doble, aunque sí se relaciona este concepto y el de frecuencia condicional. Por otro lado, un ejemplo que combina ambos tipos de definición se presenta en la Figura 2, siendo este texto junto con [H7], los únicos que además de la frecuencia absoluta doble, definen también la frecuencia relativa doble.

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

Figura 2. Definición de frecuencia conjunta absoluta y relativa ([H4], p. 218).

*C13. Distribución bidimensional.* Todos los textos analizados hacen uso explícito de esta noción, a pesar de que la mitad no la definan, posiblemente por considerarse una noción equivalente a la de variable estadística bidimensional ([H4], [H6], [H7] y [H8]). Esta hipótesis fue ya sugerida por Sánchez Cobo (1999), en cuyo análisis sólo tres de

los once textos estudiados incluyen este concepto, y además alguno no diferencia la variable y distribución bidimensional. Puesto que no son conceptos totalmente equivalentes, pensamos que sería útil diferenciarlos, ya que dar la variable solo implica conocer su rango de variación y el significado de cada variable individual, mientras la distribución requiere también conocer la frecuencia de cada par de valores.

En los textos analizados, las definiciones son parcialmente correctas, porque sólo consideran el caso en que el número de categorías diferentes de las dos variables unidimensionales que la forman sea el mismo. Es decir, no se considera el caso de un conjunto de modalidades  $m \times n$  categorías ( $m$  en la variable  $X$  y  $n$  en la variable  $Y$ ), siendo  $m$  y  $n$  diferentes. Posiblemente se deba a que, al tratar de facilitar la enseñanza, no hay mucho uso de tablas de doble entrada; sino, para facilitar los cálculos posteriores de la covarianza, coeficiente de correlación lineal y parámetros asociados a la recta de regresión, los datos se presentan simplemente mediante un listado.

Otras definiciones vienen acompañadas de ejemplos, aprovechando para introducir una tabla o un diagrama de dispersión, como es el caso del texto [H1] (ver Figura 3). Hacemos notar la complejidad semiótica de este ejemplo, que combina tres representaciones diferentes del concepto (tabular, gráfica y verbal). En este texto, posteriormente se ofrece una definición estructural formal (Sfard, 1991) como conjunto de valores de dos variables estadísticas unidimensionales con notación simbólica.

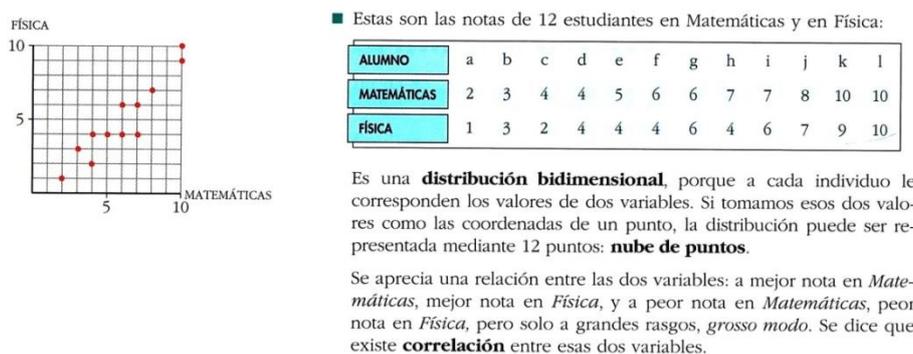


Figura 3. Definición de distribución bidimensional ([H1], p. 218).

*C2. Distribución marginal y condicional.* De la distribución doble pueden deducirse diferentes distribuciones unidimensionales: la distribución marginal para cada una de las variables  $X$  e  $Y$ , y además, fijado un valor de una de las variables (por ejemplo para  $X = x_i$ ) se puede deducir la distribución condicional de la otra variable dado este valor. En esta categoría se incluye la definición de la frecuencia y distribución marginal, y de la frecuencia y distribución condicional en los textos analizados.

C21. *Frecuencia marginal*. La frecuencia marginal de un valor  $X = x_i$  se obtiene sumando la frecuencia doble de todos los pares de valores en la distribución bidimensional que tengan este valor de  $X$ . Puede tratarse de una frecuencia absoluta o relativa. Sánchez Cobo (1999) encuentra este concepto tan sólo en cuatro de los once textos que analizó. En nuestro análisis se presenta en cinco de los ocho textos analizados ([H3], [H4], [H5], [H7] y [H8]), y es definida, en su mayoría, estructuralmente (formalmente) al inicio de su uso. Aunque ciertamente se utiliza poco, y en algunos casos, como en [H3], prácticamente nada.

Las definiciones son parcialmente correctas, principalmente porque no ofrecen la notación asociada a estas nociones, aunque los textos [H4], [H7] y [H8] se aproximan a ella, por ejemplo, con el uso de sumatorios ([H7]) o de letras ([H4]). En este caso (Figura 4), la notación de la frecuencia marginal ( $A_i$ ) es confusa pues no es habitual para las frecuencias; además no se explica el significado del subíndice, que se refiere a las modalidades de la variable, ni tampoco el rango de variación de la variable  $X$ .

- Frecuencia absoluta marginal de la variable  $X_i, A_i$ 

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

que representa el número de elementos en los que  $X$  toma el valor  $x_i$ .
- Frecuencia absoluta marginal de la variable  $Y_j, B_j$ 

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

Figura 4. Definición de frecuencia marginal ([H4], p. 218).

El caso más llamativo es el texto [H8], que incluye a definición la notación adecuada:  $f_{.m}$  y  $f_{n.}$  al definir el concepto pero posteriormente la cambia por otra ( $f_i$ ) en las tablas que aparecen a lo largo del tema. Además, define esta noción utilizando simplemente el lenguaje simbólico, sin emplear la terminología verbal “*frecuencia absoluta marginal*”.

C22. *Distribución marginal*. El conjunto de todos los valores de una de las variables, junto con la frecuencia marginal de los mismos, constituye la distribución marginal. Sánchez Cobo (1999) no la incluye entre los conceptos que analiza, aunque indica que la distribución marginal se asocia a la distribución bidimensional, para mostrar cómo deducir de aquella, algunas variables estadísticas unidimensionales. Además, la equipara a la noción de frecuencia marginal, lo que no es totalmente correcto, pues una distribución corresponde a un conjunto de valores de la variable, junto con sus frecuencias, mientras una frecuencia puede corresponder a un valor aislado.

Las definiciones que se presentan en los textos analizados son parcialmente completas, ya que no se ofrece la notación asociada, siendo tan sólo correcta en [H4] (ver Figura 5). En general, vienen acompañadas de ejemplos que clarifican su construcción. Concretamente, el texto [H1] usa esta noción sin una definición previa, como etiquetado en una tabla de frecuencias. Por su parte, el texto [H3] la define al margen y como complemento a la definición de frecuencia marginal, del siguiente modo ([H3], p. 217):

Cuando se estudian por separado las variables unidimensionales  $X$  e  $Y$  que forman la variable bidimensional  $(X, Y)$ , se habla de distribuciones marginales.

*C23. Frecuencia condicional.* La frecuencia absoluta de un valor  $X = x_i$  condicionado a que la variable  $Y$  tome un determinado valor, como por ejemplo  $Y = y_j$ , se corresponde con la frecuencia absoluta doble del dato  $(x_i, y_j)$ . La relativa sería igual al cociente entre la frecuencia absoluta doble y la marginal del valor que sirve de condición. Es importante que el estudiante se percate de que la variable que condiciona aporta información nueva a la cuestión que se plantee y por ello la frecuencia relativa condicional es, en general, diferente de la frecuencia doble condicional. Esta comprensión será de gran ayuda para dar sentido a la regresión, como modelo de ajuste a los datos, con objeto de predecir una variable a partir de la otra.

Sánchez Cobo (1999) tan sólo encuentra este concepto en uno de los once textos que analiza aunque no precisa el modo en que se presenta o el uso que se realiza de ésta. En nuestro caso, la definición tan sólo se incluye en [H4], que también presenta la definición operacional de las frecuencias relativas condicionales, apoyándose en la representación tabular de la distribución bidimensional (ver Figura 5). Este ejemplo se orienta principalmente a la definición de distribución condicional, que describimos a continuación.

*C24. Distribución condicional.* Además de las dos distribuciones marginales de una distribución bidimensional, se pueden deducir diferentes distribuciones condicionales. La distribución condicional sería el conjunto de valores de una de las variables junto con las frecuencias condicionadas respecto a un valor de la otra variable.

Tan sólo el texto [H4] (ver Figura 5) presenta una definición de esta noción, que de engloba a la definición de la frecuencia condicional, aunque, como hemos indicado, sería importante diferenciar estos conceptos. Su definición se apoya en la representación tabular, y aunque tiene algunas imprecisiones, se considera correcta. No se entiende, por

ejemplo, la exigencia de usar la primera columna (o fila), además de una intermedia, pues podría inducir a confusión, y el alumno pensar en tomar dos intermedias. No se menciona que el hecho de tomar la primera columna (fila) es porque contiene las categorías de la variable que se condiciona. Previendo su dificultad, después de su definición, el texto propone un ejemplo que clarifica su notación y funcionalidad.

## 2. Distribuciones marginales y condicionadas

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

| $X$   | Frecuencias absolutas marginales | $X$   | Frecuencias absolutas condicionadas por $y_j$ | Frecuencias relativas condicionadas por $y_j$ |
|-------|----------------------------------|-------|---|---|
| $x_1$ | $A_1$                            | $x_1$ | $f_{1j}$                                      | $f_{1j}/B_j$                                  |
| $x_2$ | $A_2$                            | $x_2$ | $f_{2j}$                                      | $f_{2j}/B_j$                                  |
| ...   | ...                              | ...   | ...   | ...   |
| $x_n$ | $A_n$                            | $x_n$ | $f_{nj}$                                      | $f_{nj}/B_j$                                  |
|       |                                  |       | $B_j$   |   |

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

Figura 5. Definición de distribución marginal y condicionada ([H4], p. 220).

Una vez descritas con detalle todas las categorías de la variable principal (concepto definido) pasamos a presentar y discutir los resultados respecto al resto de las variables, cada una de las cuáles se cruza con la anterior.

### 4.2. Forma de introducción

En la Tabla 1 presentamos la forma en que se introduce la definición del concepto en los textos analizados. El que más aparece definido es la frecuencia doble, seguido por los de variable estadística y frecuencia o distribución marginal. Apenas se define la distribución o frecuencia condicional, a pesar de su importancia para el desarrollo de nociones como la regresión; para poder comprender bien el significado de la recta de regresión, es importante entender que la función de regresión es el lugar geométrico de la media de las distribuciones condicionales.

Hay también inconsistencia pues, a veces, se define la frecuencia de un cierto tipo (doble, marginal y condicional), pero no la correspondiente distribución, y viceversa. Podemos destacar también la variabilidad de libros, desde los muy completos, como [H4], hasta los que apenas incluyen definiciones de estos conceptos, como [H2] o [H6].

Tabla 1. Forma de introducción

| Conceptos   | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 |
|---|----|----|----|----|----|----|----|----|
| <i>C1.Variable estadística y distribución bidimensional</i> |    |    |    |    |    |    |    |    |
| C11.Variable estadística bidimensional                      |    |    | EF | EF |    | F  | FE | EF |
| C12.Frecuencia doble  | E  | E  | E  | FE | E  |    | FE | EF |
| C13.Distribución bidimensional                              | EF | FE | F  |    | F  |    |    |    |
| <i>C2.Distribución marginal y condicional</i>               |    |    |    |    |    |    |    |    |
| C21.Frecuencia marginal                                     |    |    | E  | F  | F  |    | FE | F  |
| C22.Distribución marginal                                   | E  |    | E  | FE |    | E  |    | F  |
| C23.Frecuencia condicional                                  |    |    |    | FE |    |    |    |    |
| C24.Distribución condicional                                |    |    |    | FE |    |    |    |    |

Nota: E – ejemplos; F – formal; EF – ejemplos luego definición formal; FE – definición formal luego ejemplos.

La presentación se realiza generalmente a partir de ejemplos (en muchos casos en forma exclusiva), pero también, a veces, se comienza por una definición formal estructural, que en algunos casos es exclusiva (no siguen ejemplos). Lo ideal sería la presentación que comienza por ejemplos, seguida de la definición formal, pero son pocos los textos que la hacen de este modo y no en todos estos conceptos; sería también admisible la presentación formal seguida por ejemplos, que aparece en muchos textos.

#### 4.3. Lugar de introducción

En cuanto al lugar de introducción de estas definiciones, suele ser la primera vez que aparece el concepto (Tabla 2), lo que indica una orientación teoría-práctica. Únicamente en un texto se presenta la frecuencia y distribución marginal después de haberlo utilizado, y en otro la distribución bidimensional en varios momentos.

Tabla 2. Lugar de introducción

| Conceptos   | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 |
|---|----|----|----|----|----|----|----|----|
| <i>C1.Variable estadística y distribución bidimensional</i> |    |    |    |    |    |    |    |    |
| C11.Variable estadística bidimensional                      |    |    | I  | I  |    | I  | I  | I  |
| C12.Frecuencia doble  | I  | I  | I  | I  | I  |    | I  | F  |
| C13.Distribución bidimensional                              | VM | I  | I  |    | I  |    |    |    |
| <i>C2.Distribución marginal y condicional</i>               |    |    |    |    |    |    |    |    |
| C21.Frecuencia marginal                                     |    |    | F  | I  | I  |    | I  | I  |
| C22.Distribución marginal                                   | I  |    | F  | I  |    | I  |    | I  |
| C23.Frecuencia condicional                                  |    |    |    | I  |    |    |    |    |
| C24.Distribución condicional                                |    |    |    | I  |    |    |    |    |

Nota: I – inicio del tema; F – final del tema; VM – varias veces a lo largo del tema.

Cabe mencionar que, en general, estas definiciones se ubican en torno a la organización de datos bidimensionales (gráfica y tabular), y en particular, en el tratamiento de la representación tabular de la distribución bidimensional.

#### 4.4. Uso de las definiciones

En la Tabla 3 se muestra el uso de las definiciones a lo largo del tema, observando que generalmente se suelen utilizar con frecuencia en la mayoría de los libros, aunque también hay diferencias. En el texto [H1], los pocos conceptos que define respecto a la distribución estadística bidimensional se usan poco o nada. Además, las nociones de frecuencia y distribución marginal se usan menos que las de frecuencia o distribución doble. La frecuencia y distribución condicional se usa en el único texto que la define.

Tabla 3. Uso en el tema

| Conceptos  | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 |
|--|----|----|----|----|----|----|----|----|
| <i>C1. Variable estadística y distribución bidimensional</i> |    |    |    |    |    |    |    |    |
| C11. Variable estadística bidimensional                      |    |    | S  | P  |    | S  | S  | S  |
| C12. Frecuencia doble  | P  | S  | S  | S  | N  |    | S  | S  |
| C13. Distribución bidimensional                              | P  | S  | S  |    | S  |    |    |    |
| <i>C2. Distribución marginal y condicional</i>               |    |    |    |    |    |    |    |    |
| C21. Frecuencia marginal                                     |    |    | N  | S  | P  |    | S  | P  |
| C22. Distribución marginal                                   | N  |    | N  | S  |    | S  |    | S  |
| C23. Frecuencia condicional                                  |    |    |    | S  |    |    |    |    |
| C24. Distribución condicional                                |    |    |    | S  |    |    |    |    |

Nota: S – uso continuado; P – poco uso; N – no se usa.

El uso que se hace de la variable estadística bidimensional a lo largo del tema es relevante en todos los libros, salvo en el texto [H4] donde, a pesar de ofrecer una definición correcta y formal, utiliza la noción de distribución como sinónimo de variable (cuando en realidad son dos conceptos diferentes), principalmente en los enunciados de tareas. La definición de distribución bidimensional suele presentarse de modo formal, antes de su uso, es decir, de forma estructural, describiendo sus condiciones o propiedades (Sfard, 1991).

## 5. Conclusiones

El análisis de las definiciones de conceptos relacionados con la distribución de datos bidimensionales en los textos analizados, sugiere un escaso tratamiento, a pesar de su relevancia en el desarrollo de otros conceptos que se introducen después en el tema, como las nociones de correlación y regresión. Esto puede influir en la comprensión conceptual de estos temas, que debe incluir la de los principios del dominio dado y sus interrelaciones (Rittle-Johnson et al., 2001), en este caso, los diferentes tipos de frecuencias, distribuciones y variables.

En general, se definen pocos conceptos al respecto, siendo las nociones de frecuencia doble, seguida de las de variable estadística bidimensional, y frecuencia y distribución

marginal las que principalmente se incluyen. Este resultado no sería tan preocupante, si al menos las definiciones que se aportaran fuesen apropiadas.

Cuatro de los siete textos que definen la frecuencia doble no utilizan el lenguaje simbólico apropiado; además sólo dos textos establecen su relación con la frecuencia relativa doble. La noción de variable estadística bidimensional se define correctamente, pero se trata de modo implícito, como sinónimo de distribución estadística, cuando no lo son. En consecuencia, salvo en uno de los textos, sólo se define uno de estos conceptos. Por otro lado se asume el mismo número de categorías en las dos variables unidimensionales o al menos no se indica explícitamente que este número de categorías puede ser diferente en las dos variables.

Por otra parte, sólo algunos textos incluyen las definiciones de distribuciones condicionales y marginales, haciendo, los que la incluye, un uso limitado, ya que, por lo general, se presenta tan sólo la frecuencia y distribución marginal, con definiciones que no suelen incluir notación simbólica. Este resultado creemos que se debe al escaso uso de la tabla de doble entrada en los textos analizados. De este modo, si todos los datos de la distribución bidimensional poseen frecuencia uno, no tiene mucho interés determinar la distribución condicional de una de las variables a un valor de la otra.

Todos estos resultados han de interpretarse con precaución, pues, de acuerdo a Lowe y Pimm (1996), el impacto del libro de texto depende no sólo del mismo libro, sino del lector, y del profesor, así como de las interacciones que determinan su uso en el aula. Por ello, el desempeño de todas las cuestiones planteadas en el presente estudio quedan abiertas al diseño de situaciones problema específicas al tratamiento de datos bidimensionales, que permitan al estudiante alcanzar el nivel de abstracción adecuado para el desarrollo de nociones posteriores como la correlación y regresión.

**Agradecimientos:** Proyecto EDU2010-14947, FPI-BES-2011-044684 (MICINN-FEDER) y grupo FQM126 (Junta de Andalucía).

## **Referencias**

- Azcárate, P. y Serradó, A. (2006). Tendencias didácticas en los libros de texto de matemáticas para la Eso. *Revista de Educación*, 340, 341-378.
- 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.
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90 (2), 272-292.

- Estepa, A. (2007). Caracterización del significado de la correlación y regresión de estudiantes de Educación Secundaria. *Zetetiké*, 15 (28), 119-151.
- 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., 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.
- Gea, M. M. (2012). *Fundamentos para un estudio sobre la didáctica de la correlación y regresión*. Tesis de Máster. Universidad de Granada.
- Godino, J. D. (2003). *Teoría de las Funciones Semióticas. Un Enfoque Ontológico-Semiótico de la Cognición e Instrucción Matemática*. Granada: Departamento de Didáctica de la Matemática.
- Lavalle, A. L., Micheli, E. B. y Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *RELIME*, 9 (3), 383-406.
- 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 (Eds.), *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 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.
- Ortiz, J. J., Batanero, C. y Serrano, L. (2001). El lenguaje probabilístico en los libros de texto. *Suma*, 38, 5-14.
- 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), 346-362.
- 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), 1-36.
- Sánchez Cobo, F. T. (1999). *Significado de la correlación y regresión para los estudiantes universitarios*. Tesis doctoral no publicada. 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.

## Anexo

Tabla A1. Libros de texto utilizados en el análisis

| Código | Referencia  |
|--------|---|
| H1     | Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). <i>Matemáticas aplicadas a las Ciencias Sociales I</i> . Madrid: Anaya.   |
| H2     | Arias, J. M. y Maza, I. (2011). <i>Matemáticas aplicadas a las Ciencias Sociales I</i> . Madrid: Bruño.   |
| H3     | Anguera, J., Biosca, A., Espinet, M. J., Fandos, M.J., Gimeno, M. y Rey, J. (2008). <i>Matemáticas I aplicadas a las Ciencias Sociales</i> . Barcelona: Guadiel.                  |
| H4     | Monteagudo, M. F. y Paz, J. (2008). <i>1º Bachillerato. Matemáticas aplicadas a las Ciencias Sociales</i> . Zaragoza: Luis Vives.   |
| H5     | Martínez, J. M., Cuadra, R., Heras, A. (2008). <i>Matemáticas aplicadas a las Ciencias Sociales. 1º Bachillerato</i> . Madrid: McGraw-Hill.                                       |
| H6     | Bescós, E. y Pena, Z. (2008). <i>Matemáticas aplicadas a las Ciencias Sociales</i> . Vizcaya: Oxford University Press.  |
| H7     | Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santos, D. y de Vicente, M. (2009). <i>Matemáticas aplicadas a las Ciencias Sociales I</i> . Madrid: Santillana. |
| H8     | Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008). <i>Matemáticas aplicadas a las Ciencias Sociales I</i> . Madrid: SM.                                |