

# La distribución de datos bidimensionales en los libros de textos de matemáticas de Bachillerato

M. Magdalena Gea

Universidad de Granada

Carmen Batanero

Universidad de Granada

J. António Fernandes

Universidade do Minho

Emilse Gómez

Universidad Nacional de Colombia

## Introducción

Interpretar y delimitar la distribución de los datos en un estudio estadístico es primordial para 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; Batanero, Arteaga y Gea, 2011). 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). 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.

Los libros de texto son un recurso didáctico muy utilizado en la enseñanza de las matemáticas, aportando un apoyo al profesor y al alumno y determinando, en parte, la enseñanza. Su estudio permite observar los resultados de la transposición didáctica (Chevallard, 1991), esto es, los cambios que experimenta el conocimiento matemático cuando

es adaptado para la enseñanza. Desde el currículo propuesto en las directrices curriculares al implementado en el aula, una fase importante es el currículo escrito, que se plasma en los libros de texto (Herbel, 2007). Es por ello necesario el estudio del resultado de la transposición didáctica, plasmado en los libros de texto, para asegurar que no se producen desajustes respecto al significado institucional de los objetos matemáticos.

Cordero y Flores (2007) indican que el discurso matemático escolar es determinado con frecuencia por el libro de texto, que regula las acciones de enseñanza y aprendizaje, junto con las creencias de los profesores. Reys, Reys y Chavez (2004) sugieren que los libros de texto presentan las ideas matemáticas en diferentes contextos, a la vez que permiten a los estudiantes explorar diferentes ideas y facilitan el aprendizaje.

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

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 variable y distribución bidimensional, los tipos de frecuencias y distribuciones asociadas, o el tratamiento tabular y gráfico del lenguaje que se utiliza.

Sánchez Cobo (1999) analiza los objetivos de los libros, la metodología de exposición del tema, contenidos matemáticos presentados, número de ejercicios y ejemplos, presencia de consideraciones históricas y la función que las demostraciones realizan en el texto. Lavallo, Micheli y Rubio (2006) por su parte se centran en el enfoque con que se presentan las nociones tratadas (teoría-práctica o al contrario), el nivel de profundidad: si se deducen las fórmulas de cálculo, el tipo de situaciones problemáticas y si se utilizan herramientas tecnológicas.

El objetivo de este trabajo es completar los anteriores, pues por un lado, Sánchez-Cobo analiza libros de hace casi 20 años, y los documentos curriculares actuales (MEC, 2007) han reforzado el tema. Por otro lado, los libros analizados por Lavallo, Micheli y Rubio fueron publicados en Argentina. Además, analizamos aspectos no tratados por estos autores, pues nos centramos en la distribución bidimensional, analizando las definiciones de los conceptos asociados a ella, junto con su representación tabular y gráfica en una muestra de libros de texto de Bachillerato españoles actuales. En lo que sigue analizamos los fundamentos, métodos y resultados del estudio, finalizando con algunas conclusiones.

## Fundamentos

### Marco teórico

Un elemento fundamental en la construcción del conocimiento matemático son los conceptos involucrados en la resolución de los problemas. Aunque conocimiento conceptual y procedimental son polos de un continuo, 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 puede 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.

Skemp (1993) formuló dos principios que han de ser tenidos en cuenta en el aprendizaje de los conceptos matemáticos: los conceptos más avanzados que aquellos que una persona ya posee no pueden ser comunicados simplemente a través de una definición. En primer lugar, sería necesario proporcionar una colección apropiada de ejemplos, apoyados en los conceptos que ya se conocen; y en segundo lugar es necesario estar seguros que éstos ya están presentes en la mente del aprendiz en el acto de enseñanza y aprendizaje. En nuestro análisis estudiaremos si los libros aplican el primero de estos principios, introduciendo ejemplos apropiados, antes de proponer la definición de los conceptos.

Es importante tener en cuenta que los conceptos no se reducen a su definición; Vergnaud (1990) incluye en ellos sus propiedades esenciales invariantes, las situaciones en las que se presenta y las representaciones que le son asociadas. Además de analizar si la definición de los conceptos presenta sus propiedades invariantes, también analizaremos las representaciones tabulares y gráficas asociadas a la distribución de datos bidimensionales.

Por su parte, Joyce y Weil (1996) consideran varios aspectos a tener en cuenta en un concepto: (a) su nombre o etiqueta que se asocia al concepto y que puede ser una palabra o símbolo que nosotros asimilamos a las representaciones verbales o simbólicas; (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. Todos estos puntos son considerados en el estudio.

Godino (2002) indica que un objeto matemático se caracteriza por su definición y el enunciado de sus propiedades esenciales; 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 los libros de texto, que se utilizan en la enseñanza, ya que la progresiva construcción del significado de estas nociones depende directamente de los conceptos que se definan y utilicen.

## 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 (1999), Ortiz, Batanero y Serrano (2001), Azcárate y Serradó (2006) y Cobo y Batanero (2004).

Entre estos trabajos resaltamos el de Ortiz (1999), quien estudia la probabilidad en los libros de textos españoles, y del cual hemos tomado algunas variables para nuestro estudio. En concreto, el autor estudia las representaciones tabulares y gráficas asociadas a la probabilidad en los libros de texto, punto que también tendremos en cuenta en nuestro trabajo. Asimismo estudia las definiciones presentadas de los conceptos asociados en el tema de probabilidad, aunque únicamente diferencia si la definición se lleva a cabo en forma explícita o implícita. En nuestro caso consideraremos que la definición es implícita, si el concepto se utiliza en un ejemplo sin llegar a definirlo.

El primer antecedente directamente 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, Lavallo *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.

Nuestra finalidad es completar estos estudios con puntos no tratados por los autores.

## Metodología

En este estudio se analizaron doce libros de texto, todos ellos correspondientes al currículo actual de Bachillerato en las dos modalidades en que se incluye el tema: seis libros de la modalidad de *Humanidades y Ciencias Sociales* (H1, H2, H3, H4, H5 y H6) y otros seis de la modalidad de *Ciencias y Tecnología* (T1, T2, T3, T4, T5 y T6), donde  $H_i$  e  $T_i$ , con  $i=1, 2, 3, 4, 5$  e  $6$ , son de la misma editorial. Todos los textos son de editoriales de gran prestigio y tradición, son posteriores al actual decreto en vigor en España (MEC, 2007), se usan actualmente y se presentan en el Anexo.

En ellos se analizaron las definiciones de los conceptos asociados a la distribución de datos bidimensionales, así como el tipo de representación que se utiliza para presentar esta distribución. En el análisis hemos considerado las siguientes variables:

V1: *Concepto definido*. En primer lugar se han clasificado todos los conceptos 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. Siguiendo

a Vergnaud (1990) consideramos correctas las definiciones cuando describen con detalle sus características esenciales, incluyendo también sus representaciones simbólicas, pues estos son elementos fundamentales. El enunciado de sus propiedades características, como parte de la definición del concepto, es también considerado por Godino (2002) y por Joyce y Weil (1996), que añaden también la representación verbal o simbólica del concepto. Sobre cada una de estas categorías de la variable primaria se analizan tres variables adicionales en las definiciones:

V2: *Forma* de presentación del concepto, que puede ser mediante ejemplos (E), usualmente utilizando implícitamente una definición operacional, 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).

V3: *Lugar* en que se presenta la definición en el libro de texto, que puede ser al inicio del tema (I), al final del mismo, después de haberlo utilizado en ejemplos o problemas (F), o bien se define varias veces a lo largo del tema (VM). Esta variable se considera un indicador de la orientación del libro, bien al enfoque teoría-práctica, dando más predominio a la primera, o al contrario (Lavalle *et al.*, 2006).

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). Para nosotros este punto indica la importancia dada el concepto en el texto.

Además se han considerado dos variables adicionales, relacionadas con la representación de la distribución bidimensional, puesto que Vergnaud (1990) incluye dichas representaciones como elementos constituyentes de los conceptos. Además, una característica de las matemáticas es la diversidad de representaciones relativas a un concepto; dichas representaciones contribuyen a su comprensión por parte de los estudiantes (Duval, 1993). Aunque en nuestro análisis también se han encontrado diferentes representaciones verbales y simbólicas, en este trabajo no se hace un estudio detallado, debido a la limitación de espacio. No obstante haremos referencia al uso inadecuado de representaciones simbólicas en algunos conceptos.

V5: *Representaciones tabulares utilizadas*. De acuerdo a Ortiz (1999), las tablas estadísticas ofrecen una estructuración particular de la información aportada por el conjunto de datos, presentando no sólo números (valores de la variable, frecuencias o porcentajes) sino relaciones entre diferentes objetos matemáticos. Nosotros analizamos estas representaciones tabulares, que junto con las gráficas sirven para resumir los datos bidimensionales y como ayuda en su interpretación. Dentro de las representaciones tabulares hemos encontrado en los textos listados simples de datos bidimensionales, listados de datos acompañados por frecuencias (tablas simples) y tablas de doble entrada.

V6: *Representaciones gráficas*. Además del diagrama de dispersión, que es la representación más utilizada para representar datos bidimensionales, se han hallado en los textos analizados gráficos de burbujas, histogramas, diagramas de barras y pictogramas tridimensionales.

## Resultados y discusión

### Conceptos considerados (V1)

En primer lugar clasificamos los conceptos identificados en nuestro análisis; algunos 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.* Todos los textos, salvo [H1] y [T1], definen este concepto, y en estos dos textos se sustituye la definición por la de distribución bidimensional. La definición se introduce después de algunos ejemplos en 6 de los 12 textos, como sugiere Skemp (1993), y en otros dos se introducen ejemplos después de una definición estructural (Sfard, 1991). Finalmente otros dos textos no incluyen ejemplos. De acuerdo a Vergnaud (1990) y Joyce y Weil (1996), consideramos correctas estas definiciones cuando describen con detalle sus características, incluyendo también la notación  $(X, Y)$ , como el ejemplo mostrado en la Figura 1.

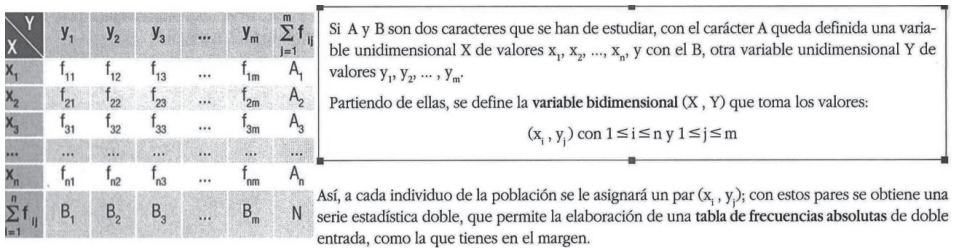


Figura 1 — Definición de variables estadística bidimensional ([H4], p. 218).

En este sentido, todas las definiciones que muestran en los textos son correctas, salvo la que se ofrece en los textos [H6] y [T6], que limita el concepto de variable estadística bidimensional a variables cuantitativas ([T6], p. 320):

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

*C12. Frecuencias dobles.* Introducida la variable estadística bidimensional, se definen los diferentes tipos de frecuencia. Todos los textos, salvo [H4], [T4] y [T5], incluyen la definición de frecuencia doble, con ejemplos, generalmente antes de introducir la definición

(Skemp, 1993). En dos textos, no se realiza una definición formal, sino simplemente se presentan ejemplos ([H1], [T1]) y en el resto, en cinco textos se introduce describiendo su cálculo, es decir, dando una definición operacional del concepto, mientras en dos se presenta en forma estructural (Sfard, 1991). Un ejemplo de definición operacional se muestra en [H2], p. 217:

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.

Algunas definiciones suelen ser parcialmente correctas ([H2] y [T2]), al no incluir la notación, aunque se describan sus propiedades (definición estructural) o su forma de cálculo (definición operacional). [H3] y [T3] combinan estas dos definiciones (Figura 2), siendo estos textos, junto con [H5] 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  $(f_r)_{ij} = \frac{f_{ij}}{N}$  y se cumple que la suma de las frecuencias relativas de todos los pares de observaciones es 1. ([H3], p. 218)

En cambio, en la misma editorial el texto [T5] dirigido al Bachillerato de Ciencias y Tecnología omite esta definición. Suponemos que esta omisión se debe a la necesidad de reducir espacio en el tema, ya que [T5] incluye conceptos sobre la variable unidimensional (variable unidimensional, medidas de centralización y dispersión) y presta menos atención a la distribución de datos bidimensionales, aunque sí utiliza la definición en un ejercicio de representación gráfica de una variable bidimensional mediante un histograma ([T5], p. 314), ejemplo que también se muestra en [H5].

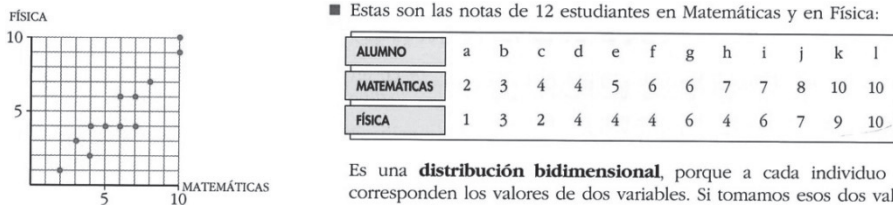
*C13. Distribución bidimensional.* Todos los textos analizados hacen un uso explícito de esta noción, a pesar de que 8 de los 12 no la definan, posiblemente por considerarse una noción equivalente a la de variable estadística bidimensional ([H3], [H4], [H5], [H6] y [T3], [T4], [T5], [T6]). Esta hipótesis fue sugerida por Sánchez Cobo (1999), en cuyo análisis sólo tres de los once textos estudiados definen 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.

Resaltamos el texto [T2], que no sólo la omite, sino que elimina de la tabla de doble entrada los totales de fila y columna, que sí se incluyen en [H2], y sirven de base para la definición de este concepto.

Esta definición en los textos en que se incluye es parcialmente correcta porque no se establecen relaciones apropiadas con el concepto de variable estadística unidimensional, pues sólo se considera el caso en que las variables unidimensionales que la forman tengan igual número de categorías. Es decir, no se considera el caso de un conjunto de  $m \times n$

modalidades ( $m$  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, hay poco uso de las tablas de doble entrada; 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.

En los textos [T1] y [H1] las definiciones van acompañadas de ejemplos, aprovechando para introducir una tabla o un diagrama de dispersión, como en [T1] (ver Figura 2), que combina tres representaciones diferentes del concepto, a la vez que define la distribución bidimensional. Posteriormente se ofrece una definición estructural (Sfard, 1991) como conjunto de valores de dos variables estadísticas unidimensionales. La misma definición y ejemplos se repiten en el texto de esta editorial [H1], dirigido a alumnos de Humanidades y Ciencias Sociales.



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

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

Figura 2 — Definición de distribución bidimensional ([T1], p. 332).

*C2. Distribución marginal y condicional.* De la distribución bidimensional 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 hemos analizado la definición de la frecuencia y distribución marginal, y de la frecuencia y distribución condicional.

*C21. Frecuencia marginal.* La frecuencia marginal absoluta o relativa de un valor  $X = x_i$  se obtiene sumando la frecuencia absoluta o relativa doble de todos los pares de valores en la distribución bidimensional que tengan este valor de  $X$ . Sánchez Cobo (1999) encuentra este concepto tan sólo en cuatro de los once textos que analizó. En nuestro caso en algo más de la mitad de los textos, aunque en uno de ellos ([H2]) tan sólo mediante ejemplos. En otros dos se incluyen ejemplos, en uno ([H5]) después de la definición y en el de la misma editorial dirigido al Bachillerato en Ciencias y Tecnología ([T5]) antes de ella. Excepto en este texto, el resto que la define utiliza una definición operacional, complementada con la estructural en cuatro textos.

Las definiciones son parcialmente correctas de acuerdo a Vergnaud (1990) y Joyce y Weil (1996), porque no ofrecen la notación asociada a esta noción, aunque los textos



[H3], [H5], [H6], [T3] y [T6] se aproximan a ella, bien con el uso de sumatorios ([H5]) o de letras ([H3], [T3]). En este último la notación de la frecuencia marginal ( $A_i$ ) es confusa pues no es la habitual, ni se explica el significado del subíndice:

Frecuencia absoluta marginal de la variable  $X_i$ ,  $A_i = \sum_{j=1}^m f_{ij}$  que representa el número de elementos en los que  $X$  toma el valor  $x_i$ .

Frecuencia absoluta marginal de la variable  $Y_j$ ,  $B_j = \sum_{i=1}^n f_{ij}$  que representa el número de elementos en los que  $Y$  toma el valor  $y_j$  ([H3], p. 218).

El caso más llamativo son los textos [H6] y [T6], que incluyen la notación adecuada:  $f_{\bullet m}$  y  $f_{n \bullet}$  en la representación tabular de la variable, y al definir el concepto la cambian por esta otra:  $f_i$  donde además, no se emplea terminología verbal “*frecuencia absoluta marginal*”, sino únicamente lenguaje simbólico.

*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 ella algunas variables estadísticas unidimensionales. Además, Sánchez Cobo equipara esta noción de frecuencia marginal a la distribución bidimensional, 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.

Esta definición se presenta en 9 de los 12 textos, aunque en 5 de ellos, sólo mediante ejemplos. En el resto, siempre en forma operacional; incluso en dos textos ([H6], [T6]) sólo se da la definición operacional, sin mostrar ejemplos de la misma.

Por ejemplo, en [H1] y [T1] se usa esta noción sin una definición formal, como etiquetado en una tabla de frecuencias. Por su parte, el texto [H2] la define al margen y como complemento a la definición de frecuencia marginal, del siguiente modo ([H2], 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.

Las definiciones que se presentan en los textos analizados suelen ser parcialmente correctas, ya que no se ofrece la notación asociada, siendo tan sólo correcta en [H3] y [T3] (ver Figura 3).

*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  $Y = y_j$ , coincide con la frecuencia absoluta doble del par de valores  $(x_i, y_j)$ . La frecuencia condicional relativa sería igual al cociente entre la frecuencia absoluta doble y la marginal del valor que sirve de condición. Es importante que el estudiante comprenda que la condición aporta nueva información y, por ello, la frecuencia relativa condicional es, en general, diferente de la frecuencia relativa doble condicional. Esta comprensión será de gran ayuda para dar

sentido a la regresión, como modelo de ajuste a los datos, cuyo objeto es predecir una variable a partir de la otra.

Sánchez Cobo (1999) sólo encuentra la frecuencia condicional en uno de los once textos que analiza. En nuestro caso, la definición sólo se incluye en [H3] y [T3], de modo operacional y seguida con ejemplos. Estos textos también presentan la definición operacional de las frecuencias relativas condicionales, apoyándose en la representación tabular de la distribución bidimensional (ver Figura 3) y también con ayuda de ejemplos. Este concepto se orienta principalmente a la definición de distribución condicional, que describimos a continuación. Al igual que en conceptos anteriores, la definición es parcialmente correcta, pues no se incluye la notación simbólica asociada.

*C24. Distribución condicional.* Sería el conjunto de valores de una de las variables junto con las frecuencias condicionadas respecto a un valor de la otra. De una distribución bidimensional se pueden deducir diferentes distribuciones condicionales.

## 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 3 — Definición de distribución marginal y condicionada ([H3], p. 220).

Tan sólo los textos [H3] y [T3] presentan una definición de esta noción (ver Figura 3), que de algún modo engloba la definición de la frecuencia condicional, aunque, como hemos

indicado, sería importante diferenciar estos conceptos. Su definición es de tipo operacional, seguida de ejemplos y se apoya en la representación tabular, y aunque tiene algunas imprecisiones, es correcta. No se aclara que el hecho de tomar la primera columna (fila), además de aquella por la que se condiciona, es porque contiene las categorías de la variable que se condiciona, pues esta primera fila no se requiere. No obstante, después de la definición, el texto propone un ejemplo para clarificar la notación y funcionalidad.

Una vez descritas con detalle todas las categorías de la primera variable (concepto definido) pasamos a presentar y discutir los resultados respecto a las tres variables que se cruzan con la anterior.

**Forma de introducción (V2)**

En la Tabla 1 presentamos la forma en que se define cada uno de los conceptos descritos anteriormente en los textos analizados. El que más aparece definido es el de variable estadística bidimensional, seguida por la frecuencia doble y la distribución marginal. Por otro lado, apenas se define la distribución o frecuencia condicional (sólo [H3] y [T3]), a pesar de su importancia para el desarrollo de nociones como la regresión, ya que, para poder comprender bien el significado de la recta de regresión, es importante entender que la función de regresión es el lugar geométrico de la media de las distribuciones condicionales.

Tabla 1 — Forma de definir el concepto

Conceptos	H1	H2	H3	H4	H5	H6	T1	T2	T3	T4	T5	T6
<i>C1. Variable estadística y distribución bidimensional</i>												
C11. Variable estadística bidimensional		ES	ESO	SO	SE	ESO		ES	ESO	SO	SE	ESO
C12. Frecuencia doble	E	OE	SE		OE	EO	E	OE	SE			EO
C13. Distribución bidimensional	OES	S					OES	S				
<i>C2. Distribución marginal y condicional</i>												
C21. Frecuencia marginal		E	OS		OSE	O			OS		ES	O
C22. Distribución marginal	E	EO	OE	E		O	E		OE	E		O
C23. Frecuencia condicional			OE						OE			
C24. Distribución condicional			OE						OE			

Nota: E=mediante ejemplos; O=definición operacional; S=definición estructural.

No se observan grandes diferencias en los textos de la misma editorial dirigidos a las dos modalidades de Bachillerato, pero sí entre editoriales en cuanto al número de conceptos considerados. En el Bachillerato en Ciencias y Tecnología encontramos un interés algo menor por las definiciones de los tipos de frecuencias, ya que en [T2] se omite la definición de frecuencia y distribución marginal, y en [T5] la de frecuencia doble, mientras que las mismas editoriales sí definen estos conceptos en la modalidad de Humanidades y Ciencias Sociales ([H2] y [H5]).

En cuanto a los textos de modalidades distintas de Bachillerato, también se observan diferencias, encontrando desde textos muy completos como [H3] y [T3], que incluyen casi todos los conceptos estudiados, hasta los que sólo se incluyen definiciones de dos conceptos como [H4] y [T4]. Además, se observa inconsistencia en cuanto al tratamiento de algunos conceptos pues, a veces, se define la frecuencia de un cierto tipo (doble, marginal y condicional), pero no la correspondiente distribución, o viceversa.

Los conceptos se suelen definir acompañados de ejemplos, en incluso en muchos casos sólo se presentan ejemplos del concepto, sin proponer una definición. Es decir, se realiza una definición implícita (en el sentido de Ortiz, 1999) por lo que es posible que el estudiante que sigue exclusivamente el texto, no llegue a formalizarlo convenientemente. Los textos [H1], [T1], [T2] y [H2] suelen presentar los ejemplos antes de la definición, como recomienda Skemp (1993). Por el contrario [H3], [H5], [T3] y [T5] comienzan por una definición operacional o estructural (Sfard, 1991) a la que siguen ejemplos. El resto de los textos no sigue un patrón a este respecto. Lo ideal sería la presentación que comienza por ejemplos, seguida de la definición formal, pero son pocos los textos que siguen este procedimiento en todos estos conceptos; sería también admisible la presentación formal seguida por ejemplos, que aparece en otros textos.

Observamos un predominio de la presentación operacional de los conceptos relacionados con la distribución marginal y condicional, que es pocas veces seguida por la definición estructural, con la que el estudiante podría adquirir un significado más completo del concepto. Por el contrario, en la presentación de conceptos relacionados con la variable estadística y distribución bidimensional, o bien se presentan los dos tipos de definición formal o se da predominio a la definición estructural.

### **Lugar de introducción (V3)**

En cuanto al lugar del texto donde se introducen las definiciones de los conceptos analizados, suele ser la primera vez que se utiliza el concepto dentro del texto (Tabla 2). Ello indica una orientación teoría-práctica en estos textos para el tema, pues indica la concepción del autor de que primero hay que introducir teóricamente un concepto, antes de proponer ejemplos o problemas que permitan aplicarlo.

Esta es una tendencia general en los textos analizados y para todos los conceptos que hemos estudiado. Únicamente en [H1] y [T1] se define la distribución bidimensional en varios momentos del desarrollo del tema, y en otros casos como [H2], [H6], [T5] y [T6] se define la frecuencia (doble o marginal) o la distribución marginal después de haberlas utilizado, al final del tema. En estos casos, primero se ha propuesto ejercicios o

ejemplos en que se utiliza el contexto y se define como síntesis de lo aprendido al finalizar el tema.

Cabe mencionar una editorial donde la definición de frecuencia marginal en las dos modalidades de Bachillerato es diferente: en [H5] se lleva a cabo al inicio del tema, mientras que en [T5] la frecuencia marginal se define tras haber sido usada en un ejemplo; en este último caso, la definición se realiza simplemente con una nota al margen.

Tabla 2 — Lugar de introducción

Conceptos	H1	H2	H3	H4	H5	H6	T1	T2	T3	T4	T5	T6
<i>C1. Variable estadística y distribución bidimensional</i>												
C11. Variable estadística bidimensional		I	I	I	I	I		I	I	I	I	I
C12. Frecuencia doble	I	I	I		I	F	I	I	I			F
C13. Distribución bidimensional	VM	I					VM	I				
<i>C2. Distribución marginal y condicional</i>												
C21. Frecuencia marginal		F	I		I	I			I		F	I
C22. Distribución marginal	I	F	I	I		I	I		I	I		I
C23. Frecuencia condicional			I						I			
C24. Distribución condicional			I						I			

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

### Uso de las definiciones (V4)

Se ha analizado igualmente la intensidad de uso de cada concepto a lo largo del tema, como un indicador de la importancia que los autores dan al mismo. En la Tabla 3 se muestra el uso de las definiciones a lo largo del tema, observando que se suelen utilizar en forma continua todos los conceptos definidos en la mayoría de los libros, no encontrando diferencias en los textos de la misma editorial dirigidos a las dos modalidades de Bachillerato.

Lo que sí encontramos es diferencias entre los diferentes textos de distintas editoriales. Por ejemplo, en los textos [H1] y [T1], los pocos conceptos que define se usan poco o nada, mientras [H3], [H5], [T3] y [T5] hacen mucho mayor uso.

Igualmente hay diferencias en el uso de un mismo concepto en diferentes editoriales. Así la variable estadística bidimensional que, a pesar de que tenga un uso destacado en los

libros, se utiliza poco en [H3] y [T3], y en su lugar se usa la noción de frecuencia doble sobre todo en los enunciados de tareas.

Tabla 3 — Uso en el tema

Conceptos	H1	H2	H3	H4	H5	H6	T1	T2	T3	T4	T5	T6
<i>C1. Variable estadística y distribución bidimensional</i>												
C11. Variable estadística bidimensional		S	P	S	S	S		S	P	S	S	S
C12. Frecuencia doble	P	S	S		S	S	P	S	S			S
C13. Distribución bidimensional	P	S					P	S				
<i>C2. Distribución marginal y condicional</i>												
C21. Frecuencia marginal		N	S		S	P			S		S	P
C22. Distribución marginal	N	N	S	S		S	N		S	S		S
C23. Frecuencia condicional			S						S			
C24. Distribución condicional			S						S			

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

De los pocos textos que definen la distribución bidimensional ([H1], [T1] y [H2], [T2]), la mitad apenas lo usan poco y la otra mitad hacen un uso constante. En general, las nociones de frecuencia y distribución marginal se definen menos que las de frecuencia o distribución doble, aunque, cuando se hace, se suele usar bastante. Igualmente la frecuencia y distribución condicional se emplea en los únicos textos que la definen ([H3] y [T3]). Al respecto, mencionar que, aunque en [T5] no se define la noción de frecuencia doble, éste texto sí la utiliza.

Finalizado el análisis de las definiciones, pasamos a estudiar las representaciones tabulares y gráficas de la distribución bidimensional pues las representaciones son consideradas por Vergnaud (1990) como elementos constituyentes de los conceptos, y por ser la variedad de representaciones una característica de las matemáticas (Duval, 1993).

### Representación tabular de datos bidimensionales (V5)

Como se ha indicado, Ortiz (1999) sugirió que las tablas estadísticas ofrecen una estructuración particular de la información aportada por el conjunto de datos, presentando no sólo valores numéricos, sino relaciones entre diferentes objetos matemáticos. En nuestro

tema, ponen en relación los valores de la variable con diferentes tipos de frecuencia. Todos los libros analizados reconocen su importancia en la organización de datos, aunque su tratamiento varía de un texto a otro. Hemos encontrado tres tipos diferenciados que se describen a continuación:

*Listado de datos.* La representación tabular más utilizada es el listado en que cada fila (o columna) de la tabla representa un par ordenado de los dos valores correspondientes a las variables observadas en cada individuo de la muestra (ver Figura 2). Este listado es denominado en algunos textos *tabla bidimensional simple* ([H6] y [T6]). Es común, avanzado el tema, que se vayan añadiendo columnas según la necesidad de los cálculos, por ejemplo, columnas con los cuadrados del valor de la variable para el cálculo de la covarianza.

*Tabla bidimensional simple con frecuencias.* Es una extensión del listado de datos, donde se añade la frecuencia correspondiente a cada par de valores  $(x_i, y_j)$  representado en la tabla. La Figura 4 muestra una tarea que utiliza esta representación:

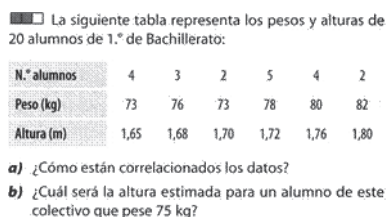


Figura 4 — Datos bidimensionales en una tarea mediante representación tabular ([T4], p. 320).

*Tabla de doble entrada.* Es la representación tabular más habitual para organizar los datos de un estudio bidimensional, principalmente cuando se dispone de una muestra de gran tamaño. Se compone de tantas filas y columnas como modalidades presenten las variables que conforman la distribución bidimensional; y en cada una de las intersecciones que se realicen entre fila (por ejemplo  $Y = y_j$ ) y columna (por ejemplo,  $X = x_i$ ) se encuentra la frecuencia de aparición de dicha modalidad  $(x_i, y_j)$ . Sánchez Cobo indica que estas tablas se presentan asociadas a la definición de la variable estadística bidimensional.

El tratamiento de cada una de estas representaciones difiere según la modalidad de Bachillerato en que nos encontremos, así como los textos analizados en cada una de ellas. En la Tabla 4 presentamos un resumen de la presencia que estas representaciones adquieren en el desarrollo del tema. Se ha diferenciado tres tipos de presentación, las dos primeras de las cuales a veces se combinan:

- a) Si se explica teóricamente el propósito y contenido de la tabla, tipos de datos y sus representaciones;
- b) Si la tabla se usa en la práctica, bien para presentar ejemplos o en ejercicios o problemas;
- c) Hay una presencia anecdótica, se presentando una o dos tablas de este tipo sin una explicación de su construcción o su utilización.

En la Tabla 4 se resume los tipos de tablas encontrados y su presentación. Todos los textos presentan los listados de datos y tablas de doble entrada, y algunos de ellos añaden la tabla bidimensional simple con frecuencias.

Tabla 4 — Representación tabular en los textos

	Presencia en el tema	H1	H2	H3	H4	H5	H6	T1	T2	T3	T4	T5	T6
Listado de datos	Desarrollo teórico y práctico	x			x			x			x		
	Uso eminentemente práctico		x	x		x	x		x	x		x	x
Tabla bidimensional simple con frecuencias	Desarrollo teórico y práctico			x			x			x			
	Uso eminentemente práctico				x						x		
	Presencia anecdótica		x			x			x			x	x
Tabla de doble entrada	Desarrollo teórico y práctico		x	x		x	x		x	x		x	x
	Presencia anecdótica	x			x			x			x		

En general, se presenta la tabla de doble entrada al comienzo del tema como medio de organización de la información de los datos bidimensionales, acompañando a la definición de variable y/o distribución bidimensional (ver Figura 1), y los diferentes tipos de frecuencias; incluso los textos [H2], [H5] y [T2], [T5] tratan la agrupación de datos en intervalos para la construcción de la tabla de doble entrada. Se suele desarrollar su descripción en forma teórica y además se acompaña de algún ejemplo o ejercicio que permita apreciar su utilidad práctica. Sólo en [H1], [H4] y [T1], [T4] estas tablas tienen una presencia anecdótica; pues sólo se incluyen al final del tema dentro de algunos ejercicios resueltos.

Por el contrario, el listado de datos en la mayoría de los textos ([H2], [H3], [H5], [H6] y [T2], [T3], [T5], [T6]) se restringe a la práctica; pues en los cálculos de diversos tipos de coeficientes destaca su uso generalizado. También algunos textos ([H1], [H4] y [T1], [T4]) basan el desarrollo teórico del tema en el listado de datos; como se ha visto, estos textos hacen un uso anecdótico de la tabla de doble entrada o la tabla bidimensional simple con frecuencias. Esta última solo se desarrolla teóricamente en algunos textos.

Este resultado es a nuestro parecer preocupante, por la baja idoneidad didáctica de la enseñanza del tema, basada sólo en los listados de datos. Ello es debido a que, de acuerdo



a Arteaga (2011), las representaciones de datos que se limitan a un listado de los mismos, no llegan a representar explícitamente la distribución de la variable bidimensional, y tendrían menor complejidad semiótica que aquellas en que se han resumido las frecuencias, hecho que ocurre en la tabla bidimensional simple con frecuencias o la tabla de doble entrada.

En cuanto a las tareas que se presentan al estudiante se basan principalmente en listados de datos, quizá por la necesidad de agilizar los cálculos, dado el tiempo de que se dispone para impartir el tema. En algunos casos, y sobre todo para el cálculo de la covarianza o el coeficiente de correlación lineal, se utiliza la tabla bidimensional simple con frecuencias como conversión en filas o columnas de la tabla de doble entrada, pero este uso llega a ser escaso ya que, por lo general, los textos presentan un único ejercicio resuelto con el uso de esta representación ([H2], [H5] y [T2], [T5], [T6]).

No se encuentran diferencias significativas en el tratamiento de las representaciones tabulares entre las modalidades de Bachillerato para la misma editorial. Tan sólo [H6] y [T6] muestran un tratamiento diferenciado de la tabla bidimensional simple con frecuencias, siendo paradójico que la modalidad científico tecnológica posea un tratamiento anecdótico de ésta.

### Representación gráfica de datos bidimensionales (V6)

La representación gráfica fue también analizada por Ortiz (1999), aunque en su estudio sólo encontró representaciones de datos univariantes, como gráficos de barras o líneas. Este tipo de representación alcanza un estatus privilegiado en el estudio de la distribución bidimensional, con una doble finalidad. En primer lugar, permite visualizar dicha distribución, facilitando su interpretación, y en segundo lugar constituye un recurso didáctico para la enseñanza posterior de la correlación y regresión. Estos temas se suelen introducir a partir de un diagrama de dispersión, en que se visualiza fácilmente la intensidad y signo de la correlación y la tendencia lineal o no lineal. Se presentan a continuación las diferentes representaciones gráficas que hemos encontrado.

*Diagrama de barras tridimensional* (ver Figura 5). Es una representación espacial de la distribución bidimensional, cuyos valores se sitúan en un plano, utilizándose la tercera dimensión para representar la frecuencia de cada par de valores, con una barra de altura proporcional a su frecuencia.

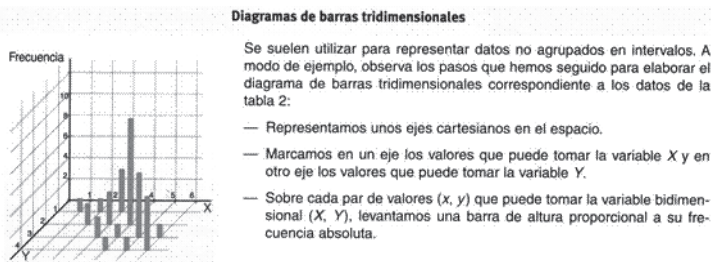


Figura 5 — Diagrama de barras tridimensional ([H2], p. 219).

*Diagrama de dispersión y gráfico de burbujas.* La representación gráfica más utilizada de la distribución bidimensional es el diagrama de dispersión o nube de puntos, donde cada par de valores se presenta en un sistema de coordenadas cartesianas. Cuando todos los datos poseen frecuencia absoluta igual a la unidad cada dato se corresponde con un punto en el plano. En otro caso, se puede optar por dibujar circunferencias con área proporcional a la frecuencia de cada par de valores (gráfico de burbujas) (ver Figura 6). Como se ha indicado, las dos representaciones son muy útiles para interpretar la relación entre las variables de estudio, ya que permite visualizar su intensidad (a través de la mayor o menor dispersión de la nube de puntos), su sentido (si la relación es directa o inversa) y el tipo (lineal o no), observando su tendencia (Sánchez Cobo, 1999). Además, el diagrama de burbujas es muy útil para visualizar en el plano, simultáneamente, hasta tres variables (el diámetro) o incluso cuatro, si mediante el color pudiera representarse una cuarta variable cualitativa.

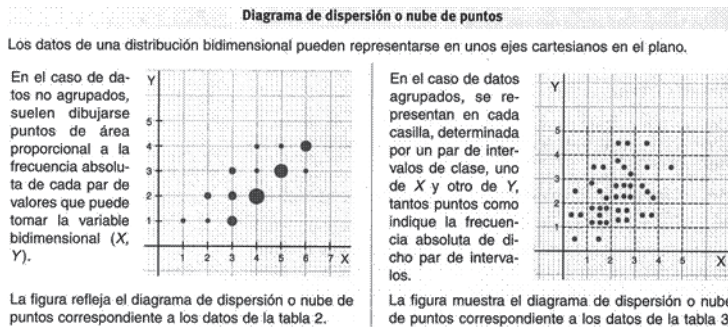


Figura 6 — Diagrama de burbujas y diagrama de dispersión ([T2], p. 270).

*Histograma tridimensional.* Este tipo de representación se utiliza principalmente cuando las variables que conforman la variable estadística bidimensional son continuas. Permite visualizar en el espacio la distribución bidimensional, situando en el plano  $XY$  cada pareja de intervalos de clase, y levantando sobre el rectángulo resultante del cruce de estos intervalos un prisma de volumen proporcional a su frecuencia absoluta (ver Figura 7).

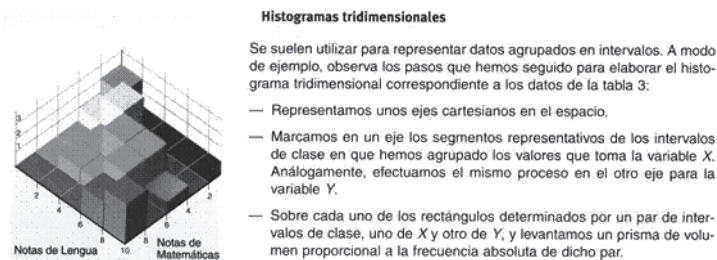


Figura 7 — Histograma tridimensional ([H2], p. 220).



Como resultado del análisis, se concluye que los textos sólo describen teóricamente el diagrama de dispersión. Aunque los alumnos han estudiado anteriormente el histograma, gráfico de barras y pictograma, siempre ha sido para datos univariantes, por lo cual algunos pudieran no generalizar fácilmente a los bivariantes. El gráfico de burbujas, por otro lado, es completamente nuevo para los estudiantes, y no se describe teóricamente.

Por otro lado, no se suele distinguir el diagrama de barras del histograma ([H1], [H5] y [T1], [T5]); ninguno de los textos que utilizan el gráfico de burbuja lo identifican como tal, siendo tratado como una extensión del diagrama de dispersión (ver Figura 8). Además, [H5] y [T5] señalan la importancia del grosor de los puntos dependiendo de la frecuencia de los datos, pero no se indica que éste debe ser proporcional al radio con que se representan. Estas imprecisiones debieran ser tenidas en cuenta por el profesor, ya que es importante que el estudiante identifique correctamente los objetos matemáticos que utiliza.

También podemos observar la destacada presencia del diagrama de dispersión en todos los textos, siendo en algunos la única representación gráfica utilizada ([H4], [H6] y [T4], [T6]). Sólo la mitad de los textos presenta el gráfico de burbujas, generalmente de modo anecdótico, a pesar de que está especialmente indicada para distribuciones en que los datos posean frecuencia distinta a la unidad, que son las situaciones más comunes que encontramos en la realidad.

En la Tabla 5, el descriptor “*Utiliza frecuencias mayores que uno*” nos permite ver que algunos textos como [H3], [H6] y [T3], [T6] omiten el tratamiento del gráfico de burbujas, pero en su lugar representan en un diagrama de dispersión tantos puntos como frecuencia presenten los datos alrededor de las coordenadas de éstos (representación poco precisa); otros no usan estas situaciones ([H4] y [T4]).

Por otro lado, los textos [H5] y [T5] plantean una tarea de representación del diagrama de dispersión sin haber explicado al estudiante cómo representar los datos cuya frecuencia es distinta a la unidad. Sólo al final del tema muestran un ejemplo del gráfico de burbuja. Una situación parecida la encontramos en [T6] que no explica cómo representar datos con frecuencia mayor a la unidad, algo que [H6] sí hace, mientras que incluye una tarea en que el estudiante debe representar en un diagrama de dispersión una distribución bidimensional donde uno dato tiene frecuencia dos.

El uso de otras representaciones es pobre, siendo su presencia prácticamente anecdótica, y reducida a un ejemplo o un ejercicio práctico. El gráfico de barras es el que más se utiliza seguido del histograma, siendo [H2] el único que incluye el pictograma a modo de ejemplo.

Sin grandes diferencias, los textos de la modalidad de Ciencias y Tecnología tienen un uso más pobre de representaciones gráficas; como dijimos, quizá porque algunos ([T4], [T5] y [T6]) incluyen el estudio de la variable estadística unidimensional, y tienen menos espacio para el tratamiento de datos bidimensionales, algo que quizá también ocurre en [H4].

## Conclusiones

Del análisis de las definiciones de conceptos relacionados con la distribución de datos bidimensionales junto a sus representaciones tabular y gráfica en la muestra de textos de Bachillerato se pueden deducir algunas implicaciones para la mejora de los libros de texto futuros y para su uso en el aula, por parte del profesor.

En primer lugar, nuestros resultados sugieren una escasa presencia de definiciones de conceptos básicos en este tema, como son los de frecuencia y distribución condicional, necesarias para poder comprender más tarde el significado de la regresión.

En este sentido, recordamos que para conseguir una comprensión conceptual adecuada de cualquier tópico se deben incluir los principios del dominio dado y sus interrelaciones (Rittle-Johnson *et al.*, 2001); en este caso, los diferentes tipos de frecuencias, distribuciones y variables, junto a sus distintas representaciones. El resto de conceptos no siempre se definen explícitamente, aunque se utilizan en el tema; por tanto hay una definición implícita (Ortiz, 1999). Será responsabilidad del profesor evaluar el conocimiento de estos conceptos en sus estudiantes y proporcionarles una definición explícita adecuada, si fuese necesario.

Los conceptos incluidos en los textos suelen definir acompañados de ejemplos, aunque no siempre presentan los ejemplos antes de la definición, como recomienda Skemp (1993). Sería también admisible la presentación formal seguida por ejemplos, que aparece en otros textos, aunque es preferible lo contrario.

En los casos en que hay una definición explícita de los conceptos, observamos un predominio de la presentación operacional (Sfard, 1991) de los conceptos relacionados con la distribución marginal y condicional, que es pocas veces seguida por la definición estructural, con la que el estudiante podría adquirir un significado más completo del concepto. Por el contrario, en los conceptos relacionados con la variable estadística y distribución bidimensional, o bien se presentan los dos tipos de definición formal o se da predominio a la definición estructural (Sfard, 1991).

Hay también un uso sesgado de las representaciones de la distribución bidimensional, pues los textos muestran tendencia a la representación gráfica sin prestar atención, en su mayoría, al proceso de construcción de estos gráficos. Además, hay poca variedad en los tipos de representación gráfica, reducida casi exclusivamente al diagrama de dispersión, sin contemplar en algunos textos situaciones en que los datos bidimensionales posean frecuencia mayor a uno.

También hay un uso muy limitado de la tabla de doble entrada a favor del listado simple de datos, con objeto de agilizar los cálculos posteriores de la covarianza o el coeficiente de correlación lineal. Este listado tiene una complejidad semiótica insuficiente para visualizar la tendencia y estructura de los datos de acuerdo a Arteaga (2011). Se olvida, además, la fuerte presencia de tablas de doble entrada en la prensa y otros medios de comunicación (Arteaga, Batanero, Cañadas y Contreras, 2011), así como las dificultades de interpretación de las mismas descritas por diferentes autores (ver, por ejemplo, Cañadas, Batanero, Contreras y Arteaga, 2011). Sería importante proporcionar a los estudiantes

una mayor experiencia en la interpretación de estos tipos de tablas y las diferentes frecuencias asociadas a la distribución bidimensional.

Pensamos que los problemas anteriores pueden resolverse con recurso a la tecnología, cuyo interés para la enseñanza de la estadística ha sido analizado por Chance, Ben-Zvi, Garfield y Medina (2007) y Pratt, Davies y Connor (2011). Por ejemplo, en Excel se presenta unas herramientas muy sencillas de usar que permiten representar diagramas de dispersión y ajustar funciones lineales y de otros tipos a los datos, obteniendo sus expresiones algebraicas y el coeficiente de correlación. Todo ello simplifica los cálculos, y por ello los textos deberían utilizar no sólo listados simples de datos, sino tablas de doble entrada y presentar colecciones de datos más amplias en cuanto al número de elementos.

La tecnología también podría contribuir a presentar a los estudiantes una mayor variedad de representaciones gráficas, incluso con carácter dinámico, pues los estudiantes podrían pasar de una a otra para comparar la información aportada por cada una de ellas o cambiar las escalas para ver el efecto del cambio sobre la representación.

En el caso de la hoja de cálculo, el interés que tiene para los alumnos su aprendizaje, la relativa simplicidad de manejo y de experimentación de varias soluciones y la reducción o eliminación de errores de cálculo mediante su uso ha sido descrita por sus potencialidades en el aprendizaje de la construcción de tablas y gráficos estadísticos (Vasconcelos y Fernandes, 2013).

Otro resultado del estudio es que, aunque dirigidos a alumnos con diferente preparación y motivación, no se han observado grandes diferencias en los libros de texto de las dos modalidades de Bachillerato. De hecho, la misma editorial suele hacer una presentación muy parecida del tema a ambos tipos de alumnos; e incluso se incluyen menos definiciones en los textos de la modalidad de Bachillerato en Ciencias y Tecnología, a pesar que para estos alumnos las directrices curriculares oficiales (MEC, 2007) sugieren presentar el estudio de las matemáticas con mayor grado de formalidad que en el Bachillerato de Humanidades y Ciencias Sociales.

Para finalizar, hacemos notar que 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.

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

## Referencias

- Arteaga, P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores*. Tesis Doctoral. Universidad de Granada.
- Arteaga, P., Batanero, C., Cañadas, G. y Contreras, J. M. (2011). Las tablas y gráficos estadísticos como objetos culturales. *Números*, 76, 55–67.

- Azcárate, P. y Serradó, A. (2006). Tendencias didácticas en los libros de texto de matemáticas para la Eso. *Revista de Educación*, 340, 341–378.
- Batanero, C., Arteaga, P. y Gea, M. M. (2011). El currículo de estadística: Reflexiones desde una perspectiva internacional. *UNO*, 59, 9–17.
- Cañadas, G., Batanero, C., Contreras, J. M. y Arteaga, P. (2011). Estrategias en el estudio de la asociación en tablas de contingencia por estudiantes de psicología. *Educación Matemática*, 23(2), 5–32.
- Chance, B., Ben-Zvi, D., Garfield, J. y Medina, E. (2007). The role of technology in improving student learning in statistics. *Technology Innovations in Statistics Education*, 1(1). Online: repositories.cdlib.org/uclastat/cts/tise/.
- Chevallard, Y. (1991). *La transposición didáctica. Del saber sabio al saber enseñado*. Buenos Aires: Aique.
- 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.
- Cordero, F. y Flores, R. (2007). El uso de las gráficas en el discurso matemático escolar. Un estudio socioepistemológico en el nivel básico a través de los libros de texto. *Revista Latinoamericana de Matemática Educativa*, 10(1), 7–38.
- Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90(2), 272–292.
- Duval (1993). *Semiosis et Noesis*. Lecturas en Didáctica de la Matemática: Escuela francesa. México: Sección de Matemática Educativa del CINVESTAV-IPN.
- 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. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 22(2–3), 237–284.
- Herbel, B. A. (2007). From intended curriculum to written curriculum: Examining the “voice” of a mathematics textbook. *Journal for Research in Mathematics Education*, 38(4), 344–369.
- 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. *RELIME*, 9(3), 383–406.
- Lowe, E. y Pimm, D. (1996). ‘This is so’: a text on texts. In 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. In D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 221–255). Dordrecht: Kluwer.
- Ortiz, J. J. (1999). *Significado de los conceptos probabilísticos elementales en los textos de Bachillerato*. Tesis Doctoral. Universidad de Granada.

- Ortiz, J. J., Batanero, C. y Serrano, L. (2001). El lenguaje probabilístico en los libros de texto. *Suma*, 38, 5–14.
- Pratt, D., Davies, N., & Connor, C. (2011). The role of technology in teaching and learning statistics. In C. Batanero, G. Burrill y C. Reading (Rds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 97–107). New York: Springer.
- Reys, B. J., Reys, R. E. y Chavez, O. (2004). Why mathematics textbooks matter. *Educational Leadership*, 61(5), 61–66.
- 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.
- 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.
- 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.
- Skemp, R. R. (1993). *The psychology of learning mathematics*. London: Penguin Books.
- Vasconcelos, A. P. y Fernandes, J. A. (2013). O uso da folha de cálculo na construção de gráficos estatísticos por alunos do 7º ano. In J. A. Fernandes, F. Viseu, M. H. Martinho y P. F. Correia (Orgs.), *Atas do III Encontro de Probabilidades e Estatística na Escola* (pp. 127–143). Braga: Centro de Investigação em Educação da Universidade do Minho.
- Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques*, 10(2–3), 133–170.



**Anexo. 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: Grupo Anaya.
H2	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 — Grupo Edebé.
H3	Monteagudo, M. F. y Paz, J. (2008). <i>1º Bachillerato. Matemáticas aplicadas a las Ciencias Sociales</i> . Zaragoza: Edelvives (Editorial Luis Vives).
H4	Bescós, E. y Pena, Z. (2008). <i>Matemáticas aplicadas a las Ciencias Sociales</i> . Vizcaya: Oxford University Press España.
H5	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.
H6	Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008). <i>Matemáticas aplicadas a las Ciencias Sociales I</i> . Madrid: Ediciones SM.
T1	Colera, J., Oliveira, M. J., García, R. y Santaella, E. (2008). <i>Matemáticas I</i> . Madrid: Grupo Anaya.
T2	Biosca, A., Doménech, M., Espinet, M. J., Fandos, M. J. y Jimeno, M. (2008). <i>Matemáticas I</i> . Barcelona: Guadiel — Grupo Edebé.
T3	Monteagudo, M. F. y Paz, J. (2008). <i>1º Bachillerato. Matemáticas. Ciencias y Tecnología</i> . Zaragoza: Edelvives (Editorial Luis Vives).
T4	Bescós, E. y Pena, Z. (2009). <i>Matemáticas. 1º Bachillerato</i> . Navarra: Oxford University Press España.
T5	Antonio, M., González, L., Lorenzo, J., Molano, A., del Río, J., Santos, D. y de Vicente, M. (2008). <i>Matemáticas I. 1º Bachillerato</i> . Madrid: Santillana Educación.
T6	Vizmanos, J. R., Hernández, J., Alcaide, F., Moreno, M. y Serrano, E. (2008). <i>Matemáticas I</i> . Madrid: Ediciones SM.

**Resumen.** En este trabajo se analizan los siguientes conceptos asociados con la distribución de datos bidimensionales en doce libros de texto españoles de matemática del bachillerato: variable estadística y distribución bidimensional, frecuencia y distribución marginal y condicional. Para cada uno de ellos se analiza el lugar y forma de presentación, y su uso en el tema. También se examinan las representaciones tabulares y gráficas de la distribución bidimensional. Se concluye que la presentación es, a veces, incompleta o parcialmente incorrecta; otras veces se equiparan conceptos no equivalentes. Esta información puede ayudar al profesor en su elección del libro de texto y para prever las dificultades que con su uso puedan tener los estudiantes.

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

**Abstract.** In this paper we analyze the following concepts related to the two-dimensional data distribution in twelve Spanish high school mathematics textbooks: two-dimensional statistical variable and distribution; marginal and conditional frequency and distribution. For each of these concepts we analyze its location, way of presentation; and use along the lesson. We also examine the tabular and graphical representations of the two-dimensional distribution. We conclude that the presentation is sometimes incomplete or partly incorrect; other times non equivalent concepts are used as synonymous. This information can help the teacher when selecting the textbook and to predict the students' difficulties when using the book in this topic.

*Keywords:* two-dimensional data distribution; textbooks; high school; definition; representation.

■■■

M. MAGDALENA GEA  
Universidad de Granada  
mmgea@ugr.es

CARMEN BATANERO  
Universidad de Granada  
batanero@ugr.es

J. ANTÓNIO FERNANDES  
Universidade do Minho  
jfernandes@ie.uminho.pt

EMILSE GÓMEZ  
Universidad Nacional de Colombia  
egomez@unal.edu.co

(Recebido em setembro de 2013; aceite para publicação em março 2014)