LAS TABLAS DE CONTINGENCIA: ALGUNAS IDEAS PARA EL AULA

Gustavo R. Cañadas, *Universidad de Granada*

Pedro Arteaga, *Universidad de Granada*

María M. Gea, *Universidad de Granada*

RESUMEN.

En los últimos años asistimos a un aumento de la necesidad de una mayor comprensión de tablas de contingencia, debido a su alta presencia en la sociedad que vivimos, como por ejemplo prensa e Internet. Sería necesario mejorar la formación de los estudiantes, lo que requiere el conocimiento de los resultados de las investigaciones didácticas. Recurriendo a los datos de un problema, se presenta un ejemplo que permitirá introducir distribuciones conjuntas, cálculo de probabilidades, representaciones gráficas y procedimientos que ayudan a analizar la asociación entre variables. El objetivo de este trabajo es resumir algunas de dichas investigaciones didácticas y mostrar formas de introducir el análisis de asociación en tablas de contingencia.

Nivel educativo: Secundaria y Universidad.

1. INTRODUCCIÓN.

Una de los modos más comunes de presentar la información estadística es en forma de tabla de doble entrada o tabla de contingencia, ya que es una forma de transmitir mucha información en poco espacio. Arteaga et al., (2011) indican que la correcta interpretación de las tablas es parte de la cultura estadística, añadiendo que todo ciudadano ha de saber utilizar para moverse con éxito en la sociedad de la información.

Según el MECD (2015), los alumnos de Educación Secundaria y Bachillerato, han de tener amplios conocimientos en el manejo de este objeto matemático.

Por tanto, presentamos una investigación que puede trabajarse a distintos niveles de profundidad, variando esta profundización según estemos trabajando en Educación Secundaria Obligatoria, con un nivel más básico; hasta niveles Universitarios más abstracto y completos.

2. ANTECEDENTES.

Los estudios relacionados con las tablas de contingencia son amplios, comenzando con Inhelder y Piaget (1955), quienes describen estrategias en el análisis de este objeto matemático.

El estudio de la precisión en el juicio de asociación ha sido llevado a cabo, entre otros, por Croker (1981), que indica que estos juicios mejoran si las frecuencias son bajas, se presentan en forma de tablas, los sucesos covarían simultáneamente, las variables son continuas y la correlación es positiva fuerte.

Allan y Jenkins (1983) indican que se tiende a basar los juicios por un lado en la diferencia entre casos confirmatorios y no confirmatorios, y por otro lado en la compatibilidad causal entre las variables. Erlick y Mills (1967) indican que la asociación negativa hace pensar al sujeto que interpreta que no existe asociación. Otros factores que influyen en los juicios de asociación (Arkes y Harkness, 1983) pueden ser: la frecuencia de la primera casilla parece tener mayor importancia, la forma de etiquetar las filas y columnas, y la presencia de números pequeños en algunas casillas. Cañadas, Díaz, Batanero y Estepa (2013) en una investigación con estudiantes de primero de psicología, observan respuestas de estimación positiva, incluso en el caso de independencia

La asociación puede ser debida a una relación causa-efecto unilateral, pero también según Barbancho (1973), a la interdependencia, dependencia indirecta, concordancia y covariación espúrea. La comprensión de la asociación implicaría, además de la exactitud en el juicio, comprender estos tipos de relaciones. Sin embargo, Estepa (1994) describe la concepción causal, según la cuál el sujeto sólo considera la asociación entre variables si puede adjudicarse a la presencia de una relación causal entre las mismas. También define la concepción unidireccional donde el estudiante no admite la asociación inversa, considerándose como independencia.

Cañadas, Batanero, Contreras y Arteaga (2011), encuentran estrategias empleadas por estudiantes de primero de psicología no descritas anteriormente. Entre ellas aparecen: utilizar una representación gráfica de las frecuencias relativas condicionales, ó calcular las frecuencias esperadas en las tablas, argumentando que debe haber equiprobabilidad en todas las celdas para que haya independencia.

Algunas investigaciones se han realizado con un proceso de enseñanza. Los resultados de Cañadas, Batanero, Gea y Contreras, (2013a) indican que, a pesar de una enseñanza planificada con cuidado, no comprendieron bien que aunque exista asociación en alguna celda aislada la frecuencia esperada puede coincidir con la observada, recordar el cálculo de las frecuencias relativas condicionales, ó comprender que la independencia implica igualdad entre las frecuencias condicionales y las frecuencias marginales, entre otras. Los resultados de un problema abierto (Cañadas, Batanero, Gea y Contreras, 2013b), indican una mejora en los juicios de asociación y las estrategias utilizadas. En un problema donde hay que aplicar un contraste Chi-cuadrado de homogeneidad en una tabla de contingencia (Cañadas, Batanero, Diaz y Roa, 2012), se observan que cerca del 80% de los estudiantes plantean correctamente; 64,1% determinan correctamente el valor del estadístico Chi-cuadrado y el valor *p*. La interpretación es más difícil pues sólo el 52% toman la decisión correcta y el 43% son capaces de interpretar los resultados en el contexto del problema.

Todas estas investigaciones nos alertan de que el tema puede no ser tan sencillo como pensamos para los estudiantes. En lo que sigue se presenta una investigación a partir de la cual se puede facilitar su enseñanza.

3. UNA INVESTIGACIÓN EN EL AULA

Una posible investigación en el aula parte de los datos de la Figura 1. Si trabajamos con alumnos de Educación Secundaria Obligatoria, se le facilitarán las tablas de frecuencia formada, sin embargo en el caso de Bachillerato, los alumnos podrán trabajar directamente con un ordenador, utilizando un software estadísticos donde formarán ellos mismos las tablas de contingencia. Intentaremos promover un aprendizaje por descubrimiento, y el profesor les ayudaría en las posibles dudas.

Más concretamente, en este proyecto se pretenden los siguientes objetivos:

- Resumir datos sobre dos variables estadísticas en una tabla de contingencia;
- Identificar las frecuencias que corresponde a cada par de valores de las variables;
- Calcular las frecuencias relativas dobles e interpretarlas;
- Calcular las frecuencias marginales, e interpretarlas;
- Calcular las frecuencias condicionales e interpretarlas;
- Representar gráficamente los datos mediante diagrama de barras adosadas, diagrama de barras apiladas y gráfico tridimensional; y
- Calcular probabilidades simples, compuestas y condicionales a partir de datos de una tabla de contingencia.

En un colegio se hace un estudio en el que participan 525 estudiantes con cuatro tipos de profesores (A, B, C y D), queremos comparar los suspensos con los aprobados. ¿Cómo podemos representar esta situación? ¿Cómo podemos ver si el número de aprobados depende del profesor?

Profesores	NOTAS		Total
	Aprobados	Suspensos	
A	68	57	125
B	90	60	150
C	70	30	100
D	120	30	150
Total	348	177	525

Figura 1. Ejercicio introductorio

Se indicaría a los estudiantes que, para representar la situación se usarían dos variables estadísticas: la variable *X*, que representaría el profesor, con cuatro valores (x_1, x_2, x_3, x_4) y la variable *Y* para hacer referencia a las calificaciones (y_1, y_2). Se indicaría que el tipo de tabla mostrado en el ejemplo (Tabla 1) corresponde a la tabla de contingencia, más concretamente, de una tabla 4x2.

Se leería la Tabla 1de forma comprensiva, haciendo observar, por ejemplo, que hay 120 alumnos aprobados con el profesor D, interpretando del mismo modo las otras frecuencias dobles. Respecto a las frecuencias marginales, se calcularían, observando, por ejemplo, que hay 150 estudiantes con el profesor B (90+60) y 100 alumnos con el profesor C (70+30). Se calcularían las frecuencias relativas dobles y marginales, interpretándolas; por ejemplo, el 17% de los alumnos tienen al profesor B y están aprobados (90/525)x100 y el 34% de los alumnos están suspensos (177/525)x100.

A continuación se representarían los datos en un diagrama de barras adosado (que para el ejemplo sería la Figura 2).

Profesores (X)	NOTAS (y)	
	Aprobados (y_1)	Suspensos (y_2)
A (x_1)	68	57
B (x_2)	90	60
C (x_3)	70	30
D (x_4)	120	30

Tabla 1. Tipo de calificación según el profesor

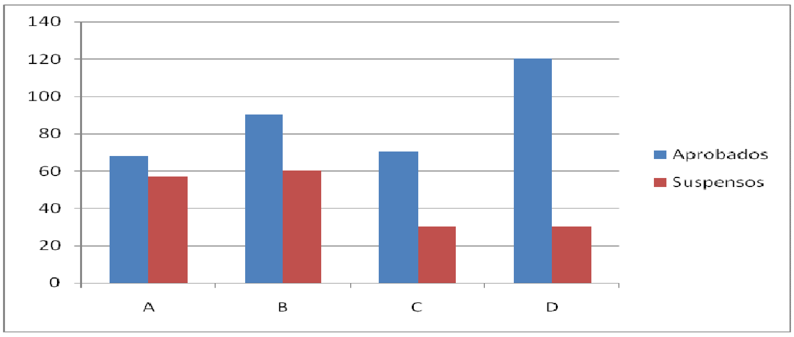

Figura 2. Diagrama de barras adosadas

Siguiendo a Batanero y Díaz (2008), se introducirían la notación y concepto de frecuencia absoluta doble, destacando que cada uno de los diferentes pares de valores (x_i, y_j) pertenece a la misma unidad estadística (el mismo estudiante), realizando preguntas a algunos estudiantes para que interpretasen las frecuencias de la Tabla 1. Se introduciría la notación para las frecuencias relativas absolutas dobles f_{ij} y relativas dobles h_{ij} y la fórmula $h_{ij}=\frac{f_{ij}}{n}$ que relaciona los dos tipos de frecuencias. Se calcularían e interpretarían las frecuencias relativas dobles para el ejemplo, resultando la Tabla 2.

		Notas (Y)	
		Aprobados (y_1)	Suspensos (y_2)
Profesores (X)	A (x_1)	0,13 (h_{11})	0,11 (h_{12})
	B (x_2)	0,17 (h_{21})	0,11 (h_{22})
	C (x_3)	0,13 (h_{31})	0,06 (h_{32})
	D (x_4)	0,23 (h_{41})	0,06 (h_{42})

Tabla 2. Frecuencias relativas dobles

Se indicaría que, a partir de la tabla de contingencia, pueden obtenerse diferentes distribuciones de una variable: las frecuencias absolutas y relativas marginales por filas y columnas, calculándose e interpretándose para el ejemplo. Se haría observar que la distribución marginal es suma de las frecuencias dobles, en el ejemplo $f_{11}+f_{12}=f_{1.}$, $f_{21}+f_{22}=f_{2.}$, $f_{31}+f_{32}=f_{3.}$, $f_{41}+f_{42}=f_{4.}$. También se observaría que, al sumar la columna del total por filas, o sumando el total por columnas se obtiene el total de la muestra: $f_{1.}+f_{2.}+f_{3.}+f_{4.} = f_{.1}+f_{.2} = n$.

A continuación se introducirían las distribuciones condicionales, motivando el tema, e indicando que, en el ejemplo, sería posible centrarse en solamente una parte de los datos, por ejemplo en los alumnos que tiene un profesor. Se introduciría para ello las *distribuciones condicionales.* Se introduciría la notación $h(x_i|y_j)$ para la frecuencia relativa condicional del valor x_i entre los individuos que presentan el carácter y_j, y su cálculo (Amón, 1993):

$$h(x_i \mid y_j) = \frac{f_{ij}}{f_{.j}} = \frac{h_{ij}}{h_{.j}}$$

Igualmente se propondría la notación y cálculo de la frecuencia relativa condicional:

$$h(y_j \mid x_i) = \frac{f_{ij}}{f_{i.}} = \frac{h_{ij}}{h_{i.}}$$

Se obtendrían las distribuciones condicionales por filas y columnas en el ejemplo obteniendo las Tablas 3 y 4, a partir de las cuáles se harían preguntas como, por ejemplo, si hay el mismo porcentaje de aprobados y suspensos con el profesor B.

Profesores (*X*)	Notas (*Y*)		
	Aprobados (y_1)	Suspensos (y_2)	TOTAL
B (x_2)	0,6 ($h(y_1 \mid B)$)	0,4 ($h(y_2 \mid B)$)	1

Tabla 3. Distribución de calificación de alumnos con el profesor B

		Notas (*Y*)
		Suspensos (y_2)
Profesor (*X*)	A (x_1)	0,32 ($h(x_1 \mid Y=y_2)$)
	B (x_2)	0,34 ($h(x_2 \mid Y=y_2)$)
	C (x_3)	0,17 ($h(x_3 \mid Y=y_2)$)
	D (x_4)	0,17 ($h(x_4 \mid Y=y_2)$)
	Total	1

Tabla 4. Distribución de X condicionada a Y_2

		Variable *Y*			
		Y_1	...	Y_c	TOTAL
Variable *X*	X_1	f_{11}	...	f_{1c}	$f_{1.}$
	...	...		...	...
	X_r	f_{r1}	...	f_{rc}	$f_{r.}$
	TOTAL	$f_{.1}$	...	$f_{.c}$	N

Tabla 5. Tabla de contingencia para el caso general

Se haría ver que la suma de las frecuencias relativas por filas y columnas es igual a la unidad (Nortes, 1993). Se generalizaría la notación al caso de las tablas rxc (r filas y c columnas) (Tabla 5).

Se realizarían otras representaciones gráficas de los datos, mediante el diagrama de barras apilado (Figura 3) y gráfico tridimensional (Figura 4), en los cuales los estudiantes compararían la información que proporciona cada uno de estos gráficos y el diagrama de barras adosado. También se les indicaría que las gráficas se pueden construir con frecuencias absolutas, relativas o porcentajes. Se haría observar que el porcentaje de aprobados no es igual en los diferentes profesores.

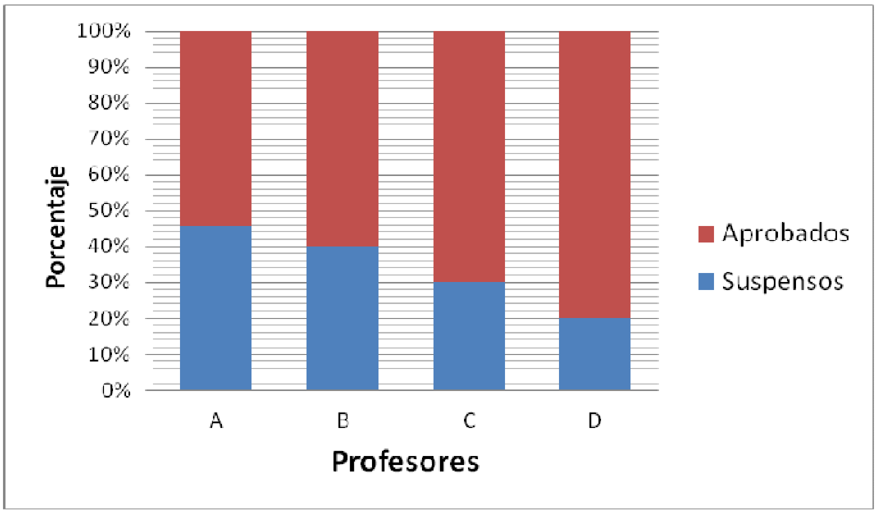

Figura 3. Diagrama de barras apilado

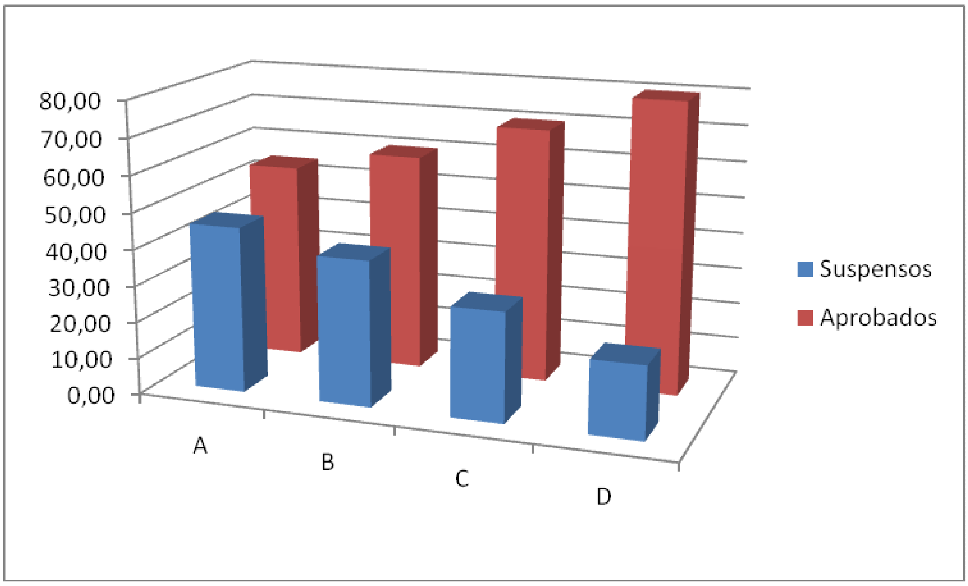

Figura 3. Diagrama de barras apilado

Esto motivaría el cálculo de frecuencias relativas condicionales respecto a cada fila y cada columna. A partir de la tabla, los alumnos pueden identificar las probabilidades simples de aprobar o suspender y las probabilidades condicionales de aprobar con cada profesor. Se concluye que es más probable estar aprobado entre los alumnos que tienen al profesor D y que los porcentajes en cada fila suman 100. Igualmente se interpretarían los resultados por columnas para resolver problemas como ver si se ha aprobado, cuál es el profesor más probable del alumno. Finalmente se mostraría en la Tabla 6 todos los resultados anteriores.

	Notas			Interpretación
Profesor	Aprobados	Suspensos	TOTAL	
A	68	57	125	Frecuencia absoluta
	0,13	0,11		Frecuencias relativas
	0,54	0,46		Respecto a profesor
	0,2	0,32	0,24	Respecto a nota
B	90	60	150	Frecuencia absoluta
	0,17	0,11		Frecuencias relativas
	0,6	0,4		Respecto a profesor
	0,26	0,34	0,29	Respecto a nota
C	70	30	100	Frecuencia absoluta
	0,13	0,06		Frecuencias relativas
	0,7	0,3		Respecto a profesor
	0,2	0,17	0,19	Respecto a nota
D	120	30	150	Frecuencia absoluta
	0,23	0,06		Frecuencias relativas
	0,8	0,2		Respecto a profesor
	0,34	0,17	0,29	Respecto a nota
Total	348	177	525	
	0,66	0,34		Respecto a fila

Tabla 6. Tabla completa

Los alumnos pueden realizar también un gráfico de mosaicos a mano con ayuda de una regla graduada (Figura 5). Este gráfico es una representación mediante áreas de las proporciones: la longitud del eje vertical se divide proporcionalmente al número de alumnos según el tipo de profesor (A, B, C, D), formando rectángulos de la misma base. Cada uno se divide en proporción al número de aprobados y suspensos, usando dos colores para diferenciarlos. En este gráfico representamos la distribución marginal de cada profesor (eje vertical), las distribuciones condicionales de cada profesor aprobados o no y la distribución conjunta (dada por los rectángulos pequeños).

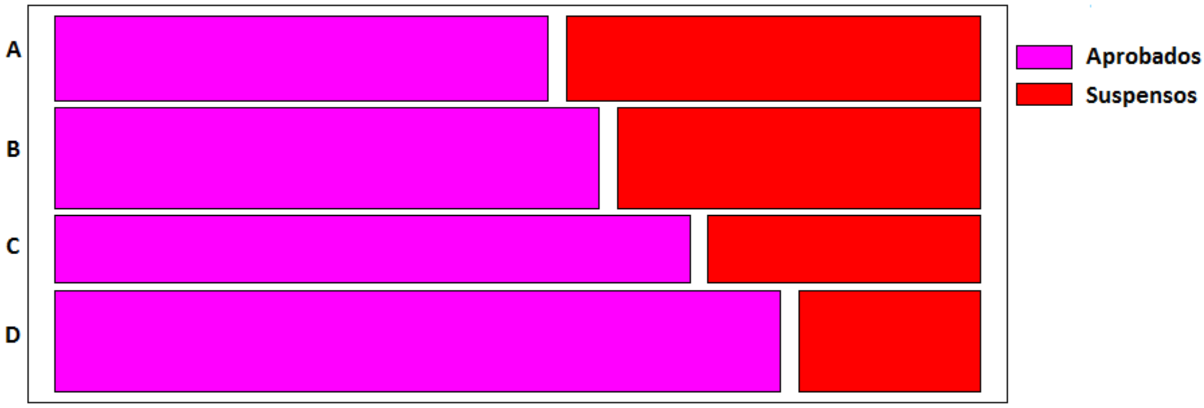

Figura 5. Gráfico de mosaicos

Se finalizaría el tema con la resolución del Ejercicio de la Figura 6 sobre idealismo, tomado de Sánchez (1995). El método más eficaz para completar la tabla sería buscar las filas y columnas que únicamente les falte un valor (incluido el valor del total). Por ejemplo, la columna "muy de acuerdo" tendría que cumplir: 73+36+X=167, por lo tanto X=58.

Ejercicio: Idealismo

A continuación mostramos una tabla con el cruce idealismo, medida esta variable a partir del grado de acuerdo-desacuerdo de los entrevistados con la siguiente frase: "*La vida sólo tiene sentido cuando una persona se dedica plenamente a una causa o ideal*" y filiación política

Filiación política	La vida sólo tiene sentido por un ideal				Total	%
	muy en desacuerdo	más bien desacuerdo	más bien en desacuerdo	muy de acuerdo		
Izquierda			420	73	954	58,1
Centro		106		36	298	
Derecha	20		181		389	23,7
Total	143	598		167	1641	
%	8,7	36,4	46,7	10,2	100	

1. Rellena los huecos de la tabla, con las frecuencias y porcentajes que faltan
2. Interpreta los resultados obtenidos.

(Sánchez, 1995, p. 300).

Figura 6. Ejercicio sobre idealismo

Paso a paso se iría completando la tabla, finalizando con el cálculo del "Total", llegando a la Tabla 7. En la segunda parte del ejercicio se realizaría la interpretación de la tabla, indicando por ejemplo, que entre los que están muy de acuerdo el 43% es de izquierdas, mientras que los que están muy en desacuerdo el 69% son de izquierdas.

Ideología Recodificada	Vida sólo tiene sentido Muy en Desacuerdo	Más bien Desacuerdo	Más bien Desacuerdo	Muy de acuerdo	Total	%
Izquierda	99	362	420	73	954	58,1
Centro	24	106	132	36	298	18,2
Derecha	20	130	181	58	389	23,7
Total	143	598	733	167	1641	
%	8,7	36,4	46,7	10,2	100	

Tabla 7. Solución del ejercicio sobre idealismo

4. CONCLUSIONES

En la investigación de aula propuesta se ha trabajado con muchas de las ideas fundamentales en estadística: partimos de datos bivariantes, consideramos varios tipos de distribución (conjunta, condicional, marginal) y sus relaciones. Nos introducimos en las ideas de asociación e independencia, utilizando varios procedimientos para analizarla (comparación de probabilidades y proporciones).

El razonamiento estadístico se fomenta con la transnumeración, al pasar de la tabla de datos a diferentes tipos de gráficos que nos permite visualizar las relaciones en los datos y las fuentes de variabilidad. Además se ha recorrido todo el ciclo de un estudio estadístico, comenzando con un problema, eligiendo y utilizando modelos matemáticos y estadísticos para resolverlo y enfatizando la interpretación de los resultados. Finalmente, el tema es motivador para los alumnos ya que les lleva a observar la utilidad de la estadística para realizar investigaciones y responder preguntas y fomenta su actitud crítica frente a la información. El profesor interesado puede ampliar lo expuesto sobre tablas de contingencia en Batanero, Cañadas, Contreras & Arteaga (2012).

REFERENCIAS.

ALLAN, L. G. y JENKINS H. M. (1983). *The effect of representations of binary variables on judgment of influence*, Learning and Motivation, 14, 381-405.

ARKES, H. R. y HARKNESS, A. R. (1983). *Estimates of contingency between two dichotomous variables*, Journal of Experimental Psychology: General, 112 (1), 117-135.

ARTEAGA, P., BATANERO, C., CAÑADAS, G. R. y CONTRERAS, J. M. (2011). *Las tablas y gráficos estadísticos como objetos culturales*, Números, 76, 55-67

BARBANCHO, A. G. (1973). *Estadística elemental moderna.* Ariel. Cuarta edición Barcelona (España).

CAÑADAS, G. R., BATANERO, C., DIAZ, C., & ROA, R. (2012). *Psychology students' understanding of the Chi-squared test*, Statistique et Enseignement, 3 (1), 3-18.

CAÑADAS, G. R., BATANERO, C., GEA, M. M. y CONTRERAS, J. M. (2013a). *Comprensión de frecuencias asociadas a las tablas de contingencia por estudiantes de psicología*, Uni-pluri/versidad, 13 (3), 97-108.

CAÑADAS, G. R., BATANERO, C., GEA, M. M. y CONTRERAS, J. M. (2013b). *Comprensión de asociación en tablas de contingencia por estudiantes de psicología*, Revista de Investigación y Divulgación en Psicología y Logopedia, 3 (1), 19-24.

CAÑADAS, G. R., DIAZ, C., BATANERO, C. y ESTEPA, A. (2013). *Precisión de los Estudiantes de Psicología en la Estimación de la Asociación*, Bolema, 27 (47), 759-778.

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

ERLICK, D. E. y MILLS, R. G. (1967). *Perceptual quantification of condicional dependency*, Journal of Experimental Psychology, 73 (1), 9-14.

ESTEPA, A. (1994). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis Doctoral. Universidad de Granada, Granada (España).

INHELDER, B. y PIAGET, J. (1955). *De la logique de l´enfant à la logique de l´adolescent*. Presses Universitaires de France, Francia (París).

MECD (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.* Madrid: Autor.