

COMPRENSIÓN DE MEDIDAS DE ASOCIACIÓN EN TABLAS RxC POR ESTUDIANTES DE PSICOLOGÍA

Measures of association in tables RxC for student's in Psychology

Cañadas, G. R., Arteaga, P., Contreras, J. M. y Gea, M. M.

Universidad de Granada

Resumen

El objetivo de este trabajo fue evaluar la competencia de cálculo e interpretación de las medidas de asociación en tablas RxC en estudiantes universitarios, finalizada una experiencia de enseñanza. Se describe la experiencia desarrollada en un curso reglado de análisis de datos, y la evaluación del aprendizaje en 92 estudiantes de psicología. Se encuentra que la mayoría es capaz de calcular estas medidas en una tarea abierta, con ayuda de Excel, y una proporción alta de estudiantes interpreta el tipo de asociación estadística. Sin embargo, se observa una mayor dificultad en la interpretación probabilística del resultado y en la contextualización del mismo.

Palabras clave: *Tablas RxC, medidas de asociación, cálculo e interpretación, evaluación de la comprensión.*

Abstract

The objective of this research was to evaluate the competence of calculation and interpretation of the measures in association in RxC tables by college students after a teaching experience. The experience developed in a compulsory course of data analysis, and the assessments of learning by 92 students in Psychology are described. The majority of students were able to calculate these measures in an open task, with the help of Excel, and a high proportion of students provided a statistical interpretation of the kind of association. However, there was difficulty in the probabilistic interpretation of the results and its contextualization.

Keywords: *Tables' RxC, measures in association, calculation and interpretation, assessment of understanding.*

INTRODUCCIÓN

Las tablas de contingencia son un instrumento de gran utilidad en la presentación y análisis de datos cualitativos, que se utilizan con gran frecuencia en actividades profesionales, donde se estudia si existe asociación entre variables. A pesar de su importancia, el aprendizaje del tema no se encuentra exento de errores y dificultades, que persisten incluso después de la enseñanza (Batanero, Estepa y Godino, 1997). La investigación sobre la comprensión del tema después de la enseñanza es bien escasa y se centra más bien en las estrategias intuitivas de los estudiantes en la interpretación de dichas tablas (Inhelder y Piaget, 1955; Jenkins y Ward 1965).

El objetivo de esta investigación fue analizar la competencia en el cálculo e interpretación de las medidas de asociación en una tabla RxC en estudiantes de psicología, después de recibir enseñanza sobre el tema, completando la investigación precedente con resultados de un proceso de instrucción. El interés de realizar este estudio se debe a que no existen investigaciones previas sobre esta comprensión, a pesar de que el estudio de las medidas de asociación en tablas rxc se incluye en la mayoría de las universitarias.

Investigaciones previas

La investigación sobre la asociación estadística se inicia con los trabajos de Inhelder y Piaget (1955), que describen las estrategias usadas en los juicios de asociación en tablas 2x2 (como la mostrada en la Tabla 1) en diferentes edades. Cuando se pide estudiar la asociación entre dos variables A y B a partir de los datos de la tabla, los razonamientos más elementales se originan en la adolescencia, donde los sujetos usan sólo la celda a para juzgar la asociación; es decir, consideran que hay asociación sólo si el número de casos en que se presentan a la vez A y B es elevado.

Entre los 12 y 15 años, se observa que los alumnos comparan celdas dos a dos, por ejemplo, comparan la celda a con la celda b , considerando la presencia o ausencia de la variable B y manteniendo fija la presencia de A . Un nivel posterior sería comprender y distinguir las cuatro situaciones posibles que se muestran en la tabla, considerando como casos favorables a la presencia (ausencia) conjunta de las variables las celdas a y d , y como casos desfavorables a la asociación entre las variables las celdas b y c . Finalmente, un razonamiento avanzado establece las relaciones diagonales (a y d se perciben como favorables a la asociación, mientras que b y c serían contrarias), comparándolas entre sí o con el total ($a+b+c+d$). Jenkins y Ward (1965) indican que esta última estrategia sólo se puede usar con frecuencias marginales iguales para la variable independiente.

Tabla 1. Esquema de una tabla 2x2

	B	No B	Total
A	a	b	a+b
No A	c	d	c+d
Total	a+c	b+d	

En relación a la enseñanza y aprendizaje del tema, Batanero, Estepa, Godino, y Green (1996) investigan las estrategias intuitivas de los estudiantes. Batanero, Estepa y Godino (1997) describen diferentes concepciones de los estudiantes destacando, entre otras, la concepción causal, por su resistencia incluso después de la enseñanza. Se presenta cuando sólo se considera asociación si puede atribuirse a la presencia de una relación de causa-efecto entre las variables. También se encuentran la concepción determinista (considerar asociación sólo si existe dependencia funcional), la concepción local (confirmar la asociación basándose sólo en un subconjunto de datos del estudio) o la concepción unidireccional (cuando no se admite la asociación inversa), por lo que la asociación negativa se llega a estimar como cero (Erlick y Mills, 1967).

Más recientemente, Cañadas, Díaz, Batanero y Estepa (2013) analizan los juicios de asociación de estudiantes en psicología, donde observan que muchos dan una estimación positiva de la intensidad de la asociación, incluso en el caso de independencia. Además, la mayor precisión de la estimación corresponde a la dependencia directa y a valores altos del coeficiente de asociación. También Cañadas, Batanero, Estepa y Arteaga (2013) estudian los juicios de asociación con datos ordinales donde la mayoría sobrevaloraron la asociación, debido a la correlación ilusoria ó la falta de familiaridad con datos ordinales.

El trabajo más relacionado con el que ahora presentamos fue realizado por Cañadas, Batanero, Arteaga, Gea (2014), donde analizamos la comprensión que los estudiantes de psicología tienen de las medidas de asociación en tablas 2x2. Se propuso a una muestra de estudiantes resolver una tarea abierta, con ayuda de Excel donde debían calcular e interpretar el coeficiente Phi de Pearson, los riesgos relativos y la razón de productos cruzado. La mayoría pudo calcular estos estadísticos adecuadamente y una proporción alta de estudiantes proporciona una interpretación estadística correcta del tipo de asociación entre las variables. Sin embargo se observa mayor dificultad en la interpretación probabilística del resultado y en la contextualización del mismo. En este trabajo continuamos el anterior, proponiendo a los estudiantes el cálculo e interpretación de los coeficientes adecuados en una tabla rxc.

MÉTODO

Muestra. Contexto educativo

La experiencia se llevó a cabo en el grado de Psicología en la Universidad de Granada, dentro del curso reglado *Técnicas de Análisis en la Investigación Psicológica*. Esta asignatura es de carácter obligatorio de primer curso de grado, y en ella se estudia la asociación estadística (tablas de contingencia, medidas de asociación, etc.). Los datos se recogen después de la enseñanza de este tema. Además los estudiantes han trabajado con las tablas de contingencia durante el bachillerato, aunque es la primera vez que estudian las medidas de asociación, que no volverán a estudiar posteriormente.

En el estudio participaron 92 estudiantes, divididos en dos grupos: 51 estudiantes en el primer grupo (8 hombres y 43 mujeres) y 53 en el segundo (16 hombres y 37 mujeres), mayoritariamente mujeres, como es frecuente en esta carrera. La mayoría se encontraba en un rango de edad comprendido entre 19 y 20 años, con algún estudiante aislado que superaba esta edad.

Diseño de la enseñanza

La enseñanza se implementó en el segundo cuatrimestre académico, casi al finalizar la asignatura. El diseño de la enseñanza comenzó fijando su contenido, a partir de un análisis previo del tema en libros de estadística orientados a Psicología o Educación (por ejemplo, Ato y López, 1996; Guàrdia, Freixa, Però y Turbany, 2007). De dicho análisis, se seleccionaron los objetos matemáticos para la construcción de la propuesta didáctica (problemas, conceptos, propiedades, procedimientos, representaciones y argumentos), organizándolos en un proceso de estudio pretendido. En el diseño de la enseñanza también se tuvieron en cuenta los principios de diseño instruccional propuestos por Cobb y McClain (2004) para el aprendizaje de la estadística.

El proceso de estudio se organiza en cuatro lecciones específicamente diseñadas (Cañadas, 2011), con soporte en una página web en que se desarrollan los bloques de contenido (www.ugr.es/~analisisdedatos/webcurso/presentacion.html): (1) tablas de contingencia, lectura e interpretación; (2) asociación estadística, dependencia funcional e independencia; (3) el estadístico Chi-cuadrado y contrastes asociados; y (4) medidas de asociación. Asimismo, se proporcionó al estudiante (en la web diseñada para la enseñanza) diferentes hojas de cálculo ya programadas en el programa Excel para facilitar los cálculos, con breves descripciones en cada hoja (Figura 1).

FRECUENCIAS DOBLES							FRECUENCIAS RELATIVAS POR FILAS								
		Variable Y							Variable Y						
		Supervivencia							Supervivencia						
		y1	y2	y3	y4	y5	Total			y1	y2	y3	y4	y5	Total
		Si	No	0	0	0		Si	No	0	0	0	0	0	
Variable X	x1 Primera	194	128				322	Variable X	x1 Primera	0,60	0,40	0,00	0,00	0,00	1
Clase	x2 Segunda	119	161				280	Clase	x2 Segunda	0,43	0,58	0,00	0,00	0,00	1
	x3 Tercera	138	573				711		x3 Tercera	0,19	0,81	0,00	0,00	0,00	1
	x4						0		x4	0	0,00	0,00	0,00	0,00	0
	x5						0		x5	0	0,00	0,00	0,00	0,00	0
	Total	451	862	0	0	0	1313								
FRECUENCIAS RELATIVAS DOBLES							FRECUENCIAS RELATIVAS POR COLUMNAS								
		Variable Y							Variable Y						
		Supervivencia							Supervivencia						
		y1	y2	y3	y4	y5	Total			y1	y2	y3	y4	y5	Total
		Si	No	0	0	0		Si	No	0	0	0	0	0	
Variable X	x1 Primera	0,15	0,10	0,00	0,00	0,00	0,25	Variable X	x1 Primera	0,43	0,15	0,00	0,00	0,00	
Clase	x2 Segunda	0,09	0,12	0,00	0,00	0,00	0,21	Clase	x2 Segunda	0,26	0,19	0,00	0,00	0,00	
	x3 Tercera	0,11	0,44	0,00	0,00	0,00	0,54		x3 Tercera	0,31	0,66	0,00	0,00	0,00	
	x4	0	0,00	0,00	0,00	0,00	0,00		x4	0	0,00	0,00	0,00	0,00	0,00
	x5	0	0,00	0,00	0,00	0,00	0,00		x5	0	0,00	0,00	0,00	0,00	0,00
	Total	0,34	0,66	0,00	0,00	0,00	1,00	Total		1,00	1,00	0,00	0,00	0,00	

NOTA IMPORTANTE. SÓLO PUEDES ESCRIBIR EN LOS CUADROS VERDES

Figura 1. Hoja “Frecuencias”

Tarea de evaluación propuesta a los estudiantes

Al finalizar la enseñanza se realizaron varias pruebas de evaluación, que los estudiantes resolvieron individualmente. Los instrumentos utilizados fueron cuestionarios de opción múltiple, para evaluar el conocimiento teórico adquirido, junto a problemas abiertos que se resolvían con el apoyo de las hojas de cálculo Excel que se facilitaron a los estudiantes (e. g., Ver Figura 1).

En este trabajo se analizan los resultados en uno de estos problemas abiertos, dirigido a evaluar el aprendizaje sobre medidas de asociación en tablas $R \times C$. En el problema se plantea el cálculo de las medidas de asociación en una tabla 2×3 junto a su interpretación (Figura 2), donde existe una asociación entre las variables media-baja (coeficiente V de Cramer = 0,364) estadísticamente significativa ($p=0,000$). Se trata de variables de las que el estudiante no tiene teorías previas al respecto, donde la relación sería indirecta debido a la existencia de una tercera variable como puede ser que una variable A inicialmente parece que hace cambios en B , pero lo que ocurre es que la variable A produce cambios en una variable C , y los cambios en la variable C producen los cambios en la variable B .

Tarea. Se está realizando un estudio sobre la población de estudiantes que hicieron la prueba de selectividad en el curso 2000-2001. Dos de las variables registradas fueron “convocatoria” (junio-septiembre) y “tipo de estudios” por el que se inclinaba el estudiante (grado medio, facultades, escuelas técnicas). Los resultados encontrados en la Universidad Complutense fueron los expuestos en la tabla siguiente:

	Escuelas Técnicas	Facultades	Grado Medio
Septiembre	200	500	800
Junio	1500	1300	700

Calcule e interprete las medidas de asociación para tablas $R \times C$.

(Adaptado de San Martín y Pardo, 1989, p. 449)

Figura 2. Tarea propuesta

En las tablas $R \times C$ podemos diferenciar entre dependencia directa y dependencia inversa observando la frecuencia en algunas celdas. No es tan sencillo como en el caso 2×2 , donde las celdas a (presencia-presencia) y d (ausencia-ausencia) contribuyen a la asociación directa, mientras las otras dos celdas, donde se da un solo carácter y el otro no, apoyan la asociación inversa. En nuestro caso, habría que observar el crecimiento de una distribución marginal (1º fila de datos), y el decrecimiento de otra (2º fila de datos), para apoyar la asociación.

Las soluciones correctas para calcular los coeficientes de asociación, realizando las correspondientes interpretaciones, son las siguientes (Ruiz-Maya, Martín, Montero y Uriz, 1995; Ato y López, 1996; Batanero y Díaz, 2008):

- Para calcular el *coeficiente de contingencia de Pearson*, ($C = \sqrt{\chi^2 / (\chi^2 + n)}$ con máximo $\sqrt{\frac{\text{Min}\{r-1, c-1\}}{1 + \text{Min}\{r-1, c-1\}}}$) los estudiantes pueden utilizar el programa Excel según la hoja de cálculo programada, con el cual obtienen el valor 0,342, así como el valor máximo, que es 0,7071. Para interpretar este resultado el estudiante puede considerar que el valor del coeficiente es levemente inferior a la mitad de su valor máximo (0,3536), por lo que puede considerar que hay una dependencia moderada.

- Para calcular el valor de *Lambda de Goodman y Kruskal*, ($\lambda_x = \frac{(\sum f_{mj}) - f_{m+}}{N - f_{m+}}$) los estudiantes pueden introducir en la hoja de cálculo programada el número de columnas y número de filas, y obtendrán el valor 0,067, es decir, la reducción del error de predicción es sólo del 6,7%. Se puede, por tanto, concluir que el error apenas se ve reducido al predecir el valor de la variable dependiente *Y*, conocido el valor de la variable independiente *X*. En este caso, es difícil predecir de qué facultad es el estudiante si se sabe en qué convocatoria aprobó, aunque, si cambiamos filas y columnas, se obtiene un valor Lambda de Goodman y Kruskal = 0,167, por lo que es algo más fácil (pero no demasiado, pues solo se reduce el error el 16,7%) deducir en qué convocatoria aprueba el estudiante si se sabe qué estudios realiza.
- Para calcular el valor *V de Cramer*, ($V = \sqrt{\chi^2 / n(p-1)}$) los estudiantes pueden utilizar la hoja de cálculo programada en Excel y obtener el valor 0,364. Para interpretar el resultado han de concluir que está un poco por debajo de la mitad de su valor posible; por lo que también este coeficiente indica que hay una dependencia moderada.

Se considera parcialmente correcta cuando el estudiante realiza el cálculo y falla en la interpretación o viceversa.

Análisis de los datos

Recogidas las respuestas escritas a la tarea, se realizó un análisis de su contenido (Krippendorff, 1991), considerando cada apartado como una medida de asociación. Con la lectura atenta de las producciones de los estudiantes, se buscaron las componentes relacionadas con cada uno de los apartados del problema. Las respuestas de todos los estudiantes fueron transcritas en un fichero de texto, asignando al estudiante su número y a cada respuesta un código correspondiente al apartado.

Las respuestas en cada apartado se compararon entre sí, mediante un proceso inductivo y cíclico, hasta llegar a un número suficiente de categorías para mostrar la diversidad de soluciones. La codificación fue revisada por todos los autores hasta llegar a un acuerdo sobre las categorías que se utilizarían en el estudio. Somos conscientes de que a priori es posible otros tipos de respuestas, pero sólo se han considerado las que dieron los estudiantes, quienes mostraron en general poca capacidad de argumentación.

RESULTADOS

A continuación, se exponen los resultados en cada apartado de la tarea.

Coefficiente de contingencia de Pearson

Respuestas correctas:

RC.1. Se calcula e interpreta de forma correcta el valor del coeficiente de contingencia de Pearson según su valor máximo, que es necesario conocer para poder interpretar adecuadamente la intensidad de la asociación (Ruiz-Maya, Martín, Montero y Uriz, 1995). Para ello, se introducen correctamente los datos en el programa, así como el número de filas y de columnas.

Respuestas parcialmente correctas:

RP.1. El estudiante calcula bien los valores necesarios, es decir, el del coeficiente y su valor máximo, pero no interpreta la intensidad de la asociación. Por ejemplo, un alumno deduce correctamente que hay dependencia al obtener un coeficiente distinto de cero, pero no es capaz de decidir qué intensidad indica el valor hallado, a pesar de haber indicado explícitamente el rango posible de valores que, para la tabla dada, puede tomar el coeficiente.

RP.2. Este es un caso particular del anterior. El estudiante calcula bien el valor del coeficiente de contingencia de Pearson y el máximo que este valor puede alcanzar, pero surge un problema, pues la interpretación del valor obtenido es incorrecta, ya que indica que la asociación es moderada-alta, en lugar de una asociación moderada o moderada-baja. Un problema en esta interpretación es que no hay una regla fija que el estudiante pueda aplicar (por ejemplo, usar unos valores fijos a partir de los cuáles la asociación sea alta o moderada), pues la decisión sobre la intensidad que debe considerarse alta es hasta cierto punto subjetiva. No obstante, tradicionalmente sólo los valores mayores que 0,7 se consideran altos, por lo que es posible que el estudiante haya extrapolado indebidamente esta propiedad (aplicable al coeficiente) a su valor máximo, que si es superior a 0,7.

RP.3. El estudiante calcula bien el coeficiente de contingencia de Pearson, aunque comete un error en el valor máximo que puede tener este coeficiente, posiblemente debido a confundir el número de columnas y de filas o no introducirlo correctamente en el programa. Estos estudiantes dan una interpretación correcta de los valores que han obtenido, considerando una asociación moderada o moderada-baja, ya que el coeficiente no supera la mitad del valor máximo.

RP.4. El estudiante calcula correctamente el coeficiente de contingencia de Pearson, pero comete un error en el cálculo del su valor máximo, y además, realiza una interpretación errónea de su magnitud, considerando una asociación moderada-alta, en lugar de una asociación moderada o moderada-baja. Un ejemplo sería extender indebidamente propiedades del coeficiente a su máximo.

RP.5. Al igual que en los dos casos anteriores, el estudiante calcula bien el valor del coeficiente de contingencia de Pearson y confunde el valor máximo, que además, no interpreta.

RP.6. Se calcula bien el coeficiente de contingencia de Pearson, lo que implica una buena construcción de la tabla, pero no el máximo, y la interpretación es incorrecta o no interpreta. Los errores más comunes son no considerar el número de filas y columnas (ya que el valor máximo es incorrecto) o el significado del valor obtenido (producido por no interpretar o hacerlo de forma errónea).

RP.7. El estudiante calcula mal el coeficiente de contingencia de Pearson, pues introduce con errores la tabla en la hoja de cálculo, aunque calcula bien el máximo posible de este coeficiente. No interpreta los valores que obtiene.

RP.8. Aquí el estudiante comete un error de cálculo del coeficiente como anteriormente y calcula bien su máximo como el caso anterior, pero hace una interpretación correcta de los valores que ha obtenido.

Respuestas incorrectas:

RI.1. Calcula o interpreta incorrectamente los dos valores. En ocasiones el estudiante calcula coeficientes de asociación de tablas 2x2 (coeficiente Phi de Pearson, riesgo relativo y razón de productos cruzados), en lugar de los adecuados para tablas de contingencia $R \times C$ (coeficiente de contingencia de Pearson, V de Cramer y Lambda de Goodman y Kruskal).

Tabla 2. Porcentajes de respuestas en el coeficiente de contingencia de Pearson

	Respuesta	Cálculo correcto	Cálculo Max correcto	Interpretación correcta
RC.1	39,1	39,1	39,1	39,1
RP.1	2,2	2,2	2,2	
RP.2	14,1	14,1	14,1	
RP.3	7,6	7,6		7,6
RP.4	5,4	5,4		
RP.5	1,1	1,1		
RP.6	4,3	4,3		
RP.7	1,1		1,1	

RP.8	2,2		2,2	2,2
RI.1	6,5			
No responde	16,3			
Total	100	74,8	58,7	48,9

Podemos observar en la Tabla 2 que sólo 36 estudiantes (39,1%) responden correctamente al cálculo e interpretación del coeficiente de contingencia de Pearson. La respuesta correcta es la más frecuente, seguida de RP.2, en la que se calculan bien los valores (del coeficiente y su máximo posible), pero la interpretación de estos resultados obtenidos es incorrecta. También hacemos notar que el coeficiente de contingencia de Pearson ha sido calculado correctamente por el 74,8% de los estudiantes, su máximo es realizado correctamente por el 58,7% y la interpretación la realiza correctamente el 48,9%. Observamos que la mayor dificultad se da en la interpretación.

Lambda de Goodman y Kruskal

Respuestas correctas:

RC.1. El estudiante calcula e interpreta correctamente el valor de Lambda de Goodman y Kruskal, como proporción del error reducido. Usa una notación adecuada, e introduce la tabla de contingencia de forma correcta en el programa, informando en la interpretación que es el porcentaje de error que se ve reducido al predecir el valor de la variable dependiente Y , conocido el valor de la variable independiente X , en nuestro caso el 6,7% (Ruiz-Maya, Martín, Montero y Uriz, 1995).

RC.2. Otra posible solución correcta, aunque más pobre que la anterior, es que el estudiante calcule correctamente el valor de Lambda de Goodman y Kruskal, y lo interprete en función de si hay asociación o no. Una solución de este tipo ocurre cuando el estudiante interpreta el coeficiente únicamente como coeficiente de asociación y no lo relaciona con la reducción en el porcentaje de error en la predicción.

Respuestas parcialmente correctas:

RP.1. En este caso, el estudiante calcula bien el valor del coeficiente Lambda de Goodman y Kruskal, pero no realizan una interpretación de este resultado, por lo que pensamos que no ha comprendido la forma de interpretarlo.

RP.2. Como en el caso anterior, el estudiante calcula bien el valor de Lambda de Goodman y Kruskal, pero hace una interpretación errónea, en términos de asociación, considerando una asociación moderada, o bien en términos de predicción de error, indicando que no mejora la predicción con el conocimiento de una variable.

Respuestas incorrectas:

RI.1. El estudiante realiza un cálculo incorrecto del valor de Lambda de Goodman y Kruskal, calcula e interpreta las medidas de asociación para tablas 2x2, o no calcula Lambda de Goodman y Kruskal, debido a que consideraron que no es una medida de asociación, pues en los apuntes se presentaba como una medida basada en la reducción proporcional del error.

Tabla 3. Porcentajes de respuestas en el coeficiente Lambda

	Porcentaje	Cálculo correcto	Interpretación correcta
RC.1	31,6	31,6	31,6
RC.2	16,3	16,3	16,3
RP.1	11,9	11,9	
RP.2	11,9	11,9	
RI.1	8,7		
No responde	19,6		
Total	100	71,7	47,9

En la Tabla 3 presentamos los resultados, observando que 44 estudiantes (47,9%) dan la respuesta correcta. Como se ha indicado, algunos estudiantes calcularon únicamente el coeficiente de contingencia de Pearson y el valor V de Cramer, ya que se trata de medidas basadas en la reducción proporcional del error, y no la consideraron como coeficiente de asociación. Entre las respuestas más frecuentes destacan las correctas y, posteriormente, las parcialmente correctas, lo cual muestra unos resultados buenos. El cálculo correcto de Lambda ha sido alto (71,7% de la muestra), pero como en otros coeficientes, la interpretación correcta sigue siendo baja (47,9% de la muestra).

V de Cramer

Respuestas correctas:

RC.1. El estudiante interpreta y calcula el valor del coeficiente correctamente. Utiliza una notación adecuada en su explicación, construye correctamente la tabla de contingencia haciendo uso de la hoja de cálculo programada, tiene en consideración el número de filas y de columnas, y además considera correctamente su máximo como uno (Ruiz-Maya, Martín, Montero y Uriz, 1995).

Respuestas parcialmente correctas:

RP.1. Se calcula bien el coeficiente V de Cramer, pero la interpretación es errónea. Por ejemplo, el estudiante considera independencia, en lugar de una asociación moderada o moderada-baja.

RP.2. Esta respuesta es un caso particular de la anterior, pues el estudiante, calcula bien el valor del coeficiente de V de Cramer, pero no interpreta el resultado.

RP.3. El estudiante calcula mal el valor del coeficiente a pesar de interpretar correctamente los valores que ha obtenido. Una respuesta muy frecuente es considerar el valor 0,515 del coeficiente V de Cramer, que se obtiene al considerar 3 filas y 3 columnas, en lugar de 2 filas y 3 columnas acompañada de una interpretación adecuada del mismo.

Respuestas incorrectas:

RI.1. Calcula e interpreta incorrectamente el valor del coeficiente o calcula e interpreta las medidas de asociación para tablas 2x2.

Tabla 4. Porcentajes de respuestas en el coeficiente V de Cramer

	Porcentaje	Cálculo correcto	Interpretación correcta
RC.1	44,5	44,5	44,5
RP.1	15,2	15,2	
RP.2	2,2	2,2	
RP.3	9,8		9,8
RI.1	1,1		
RI.2	6,5		
RI.3	1,1		
No responde	19,6		
Total	100	61,9	54,3

Podemos observar, en la Tabla 4, como únicamente 41 estudiantes (44,5%) calculan e interpretan correctamente del valor V de Cramer, a pesar de disponer del programa de Excel para facilitar su cálculo. Entre las respuestas parcialmente correctas destaca la respuesta RP.1, en la que los estudiantes calculan bien el coeficiente de la V de Cramer, pero la interpretación que realizan es errónea. De entre los errores que cometen los estudiantes, destacamos la dificultad que algunos muestran al introducir la función $\text{Min}\{\}$ o al introducir el número de filas y columnas en la hoja de cálculo. Consideramos que la tarea propuesta fue de dificultad media, pues la obtención de la V de Cramer la realiza correctamente el 61,9% de estudiantes aunque su interpretación tan sólo el 54,3%.

Al comparar los tres coeficientes observamos que el coeficiente de contingencia de Pearson es calculado e interpretado correctamente por el 39,1%, parcialmente por el 38,1% y de forma incorrecta por el 6,5%; el valor V de Cramer correctamente por el 44,5%, parcialmente por el 27,2% y de forma incorrecta por el 8,7%; y el valor Lambda de Goodman y Kruskal correctamente por el 47,9%, parcialmente por el 29,2%, y de forma incorrecta por el 3,3%. En resumen, podemos concluir que el cálculo de estas medidas de asociación resultan de dificultad moderada para los estudiantes, donde el índice de no respuesta es inferior al 20%, el cálculo e interpretación correcto de los mismos alrededor del 40%, y de manera parcialmente correcta entre un 27% y un 40%. Resultó más sencillo el cálculo e interpretación del valor Lambda de Goodman y Kruskal, seguida por la V de Cramer y finalmente el coeficiente de contingencia de Pearson.

DISCUSIÓN Y CONSECUENCIAS PARA LA ENSEÑANZA

Nuestros resultados muestran que el porcentaje de estudiantes que obtienen una respuesta correcta a la tarea del cálculo e interpretación de alguno de los coeficientes de asociación para tablas $R \times C$ estudiados, es mucho mayor respecto al encontrado en investigaciones previas con estudiantes que no han recibido instrucción sobre el tema (Cañadas, Batanero, Estepa, y Arteaga, 2013; Cañadas, Díaz, Batanero, Estepa, 2013). Consideramos que la experiencia de enseñanza resultó beneficiosa en este sentido, contribuyendo a la mejora del razonamiento de los estudiantes sobre la asociación en tablas de contingencia $R \times C$. El uso del programa Excel, a través de las hojas de cálculo que se programaron para servir de apoyo al proceso de enseñanza del tema, supusieron una gran ayuda a los estudiantes para el estudio del tema, además de servir de herramienta de cálculo de los coeficientes estudiados.

A pesar de estos resultados, y en comparación con investigaciones previas con medidas de asociación en tablas 2×2 (Cañadas, Batanero, Arteaga, Gea, 2014), observamos que el porcentaje de no respuesta a la tarea es muy superior en nuestro trabajo con medidas de asociación en tablas $R \times C$, que en otros trabajos. Estos resultados pueden indicar la mayor dificultad de estas medidas de asociación en relación a las utilizadas más usualmente como es el caso de las medidas de asociación en tablas 2×2 .

Hacemos también notar que una gran parte de los alumnos se limitan a dar un valor numérico, sin realizar una interpretación del resultado, y sin especificar en el contexto del problema lo que quiere decir ese resultado obtenido correctamente. Además, encontramos que algunos estudiantes (entre un 3% y un 9% de estudiantes) no lograron comprender el cálculo de algunos de los coeficientes al confundir filas y columnas en su determinación así como no comprender su significado pues basaban sus interpretaciones en coeficientes asociados a tablas 2×2 . Henry (1997) señala la dificultad de interpretación de los modelos matemáticos.

Todo ello sugiere la necesidad de poner mayor atención en la enseñanza del tema y en proponer mayor cantidad de actividades interpretativas en el proceso de enseñanza. Posiblemente, la facilidad de calcular e interpretar el valor de Lambda se debe a que este valor no se veía afectado por el número de filas y columnas, y no requiere el cálculo de un valor máximo de forma independiente.

Consideramos de gran utilidad el análisis realizado sobre el desempeño de los estudiantes en la tarea propuesta, que informa de las dificultades de algunos estudiantes en la interpretación y el cálculo de medidas de asociación. Como indican Cañadas, Batanero, Arteaga, Gea, (2014) las tablas de contingencia son un instrumento útil en nuestra vida diaria (como resumen de información) y profesional, y este trabajo avanza en esta línea de investigación, mostrando que aún queda mucho camino por recorrer para la mejora de la enseñanza de este objeto matemático.

Agradecimientos: Proyecto EDU2013-41141-P (MEC) y Grupo FQM-126 (Junta de Andalucía).

REFERENCIAS

- Ato, M. y López, J. J. (1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis
- Batanero, C. y Díaz, C. (2008). *Análisis de datos con Statgraphics*. Granada: Departamento de Didáctica de la Matemática.
- Batanero, C., Estepa, A. y Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer based teaching environment. En J. B. Garfield y G. Burrill (Ed.), *Research on the role of technology in teaching and learning statistics* (pp. 191-206). Minnesota, MN: International Statistical Institute.
- Batanero, C., Estepa, A., Godino, J. D., & Green D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151-169.
- Cañadas, G. R. (2011). *Las tablas de contingencia para psicología*. Granada: Departamento de Didáctica de la Matemática.
- Cañadas, G. R., Batanero, C., Arteaga, P., Gea, M. M. (2014). Medidas de asociación en tablas 2x2: evaluación de una experiencia de enseñanza con estudiantes universitarios. En M. T. González, M. Codes, D. Arnau y T. Ortega (Eds.), *Investigación en Educación Matemática XVIII* (pp. 207-216). Salamanca: SEIEM.
- Cañadas, G. R., Batanero, C., Estepa, A. y Arteaga, P. (2013). Juicios de asociación en tablas de contingencia con datos ordinales. En A. Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.), *Investigación en Educación Matemática XVII* (pp. 209-218). Bilbao: SEIEM.
- Cañadas, G. R., Díaz, C., Batanero, C. y Estepa, A. (2013). Precisión de los estudiantes de psicología en la estimación de la asociación, *Bolema*, 27(47), 759-778.
- Cobb, P. y Mc Clain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking*. Netherlands: Kluwer Academic Publishers.
- Erlick, D. E. y Mills, R. G. (1967). Perceptual quantification of conditional dependency, *Journal of Experimental Psychology*, 73 (1), 9-14.
- Guàrdia, J., Freixa, M., Però, M. y Turbany, J. (2007). *Análisis de datos en psicología*, Delta, Publicaciones Universitarias.
- Henry, M. (1997). Notion de modèle et modélisation en l'enseignement. En M. Henry (Ed.), *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM.
- Inhelder, B. y Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris: Presses Universitaires de France.
- Jenkins, H. M. y Ward, W. C. (1965). Judgment of the contingency between responses and outcomes, *Psychological Monographs*, 79, 1-17.
- Krippendorff, K. (1991). *Metodología de análisis de contenido*, Barcelona, Paidós.
- Meiser, T. y Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, 36(3), 315-336.
- Ruiz-Maya, L., Martín, F. J., Montero, J. M. y Uriz, P. (1995). *Análisis estadístico de encuestas: datos cualitativos*. Madrid: AC.
- San Martín, R. y Pardo, A. (1989). *Psicoestadística Contrastes paramétricos y no paramétricos*, Pirámide.