

ENSEÑANZA Y COMPRENSIÓN FORMAL DE LAS TABLAS DE CONTINGENCIA

Gustavo R. Cañadas, Juan J. Ortiz, José M. Contreras, María M. Gea
Facultad de Educación, Universidad de Granada
grcanadas@ugr.es

España

Resumen. En este trabajo, se describe una experiencia de enseñanza de las tablas de contingencia, evaluando el aprendizaje de uno de los temas (Tablas de contingencia, lectura e interpretación) en una muestra de 93 estudiantes del primer año de la licenciatura de Psicología. Se presentan los resultados en 6 ítems de opción múltiple y un problema abierto, que indican un aprendizaje satisfactorio, adquisición de estrategias formales y superación de la concepción local y causal sobre la asociación.

Palabras clave: tablas de contingencia, dificultades, materiales de enseñanza

Abstract. In this research, we describe an experience of teaching contingency tables, and assess the learning of a lesson (Contingency tables, reading and interpretation) in a sample of 93 students in the first year of Psychology. Results are presented for six multiple-choice items and an open problem, indicating a successful learning, acquisition of formal strategies and overcoming the local and causal conception of the association.

Key words: contingency tables, difficulties, teaching materials

Introducción

Las tablas de contingencia son una forma común de resumir datos categóricos. En general, el interés se centra en estudiar si existe asociación entre una variable fila y otra variable columna y/o calcular la intensidad de dicha asociación. A pesar de su importancia, se presta poca atención a este tema, suponiendo que su comprensión es sencilla.

En este trabajo se describe una experiencia de enseñanza; se analiza también la evaluación del aprendizaje del tema: “Tablas de contingencia, lectura e interpretación”, utilizando, como marco teórico el enfoque onto-semiótico (Godino, Batanero y Font, 2007) en que la comprensión formal se concibe como correspondencia entre el significado institucional de un concepto (en matemáticas) y el significado personal adquirido por el estudiante. Se comienza analizando la investigación previa y se describe el método y resultados.

En los antecedentes, así como en nuestra investigación, se entiende como estrategia la información que usan los sujetos para llegar al juicio de asociación y la forma en que la combinan. Distinguimos entre estrategias intuitivas (las usadas antes de la enseñanza) y formales (basadas en conceptos aprendidos después de la enseñanza).

Investigaciones previas

Inhelder y Piaget (1955) inician la investigación sobre juicios de asociación, describiendo las estrategias de los chicos a partir de 13 años. Para explicar sus resultados utilizan una la Tabla I, indicando que en una primera etapa, sólo se usa la celda a (presencia de los caracteres A y B)

sin comprender que la celda d (ausencia de los dos caracteres) tiene el mismo peso en relación a la asociación. En una segunda etapa (entre los 13 y 15 años), se comparan las celdas dos a dos (por ejemplo a con b , o a con c). El tercer paso, sería comprender cuales son los casos favorables (a y d) y desfavorables (b y c) de la asociación, sin compararlos y finalmente, el sujeto establece las relaciones diagonales, comparándolas entre sí o con el total ($a+b+c+d$).

Posteriormente autores, como Smedlund (1963) y Shaklee y Mins (1982) han analizado las estrategias de sujetos adultos, concluyendo que, en contraposición con lo supuesto por Piaget e Inhelder, algunos continúan con estrategias propias de niños, por lo que el desarrollo del concepto de asociación no sería espontáneo en los sujetos. La dificultad de la tarea la muestran Jenkins y Ward (1965), indicando que incluso la estrategia de comparar las diagonales, considerada como correcta por Inhelder y Piaget, sólo llevaría a un juicio de asociación correcto, si las frecuencias marginales por filas o por columnas en la Tabla I fuesen iguales, pero no en el caso general. La estrategia correcta para el caso general consiste en comparar la diferencia entre las probabilidades $P(B|A)$ y $P(B|no A)$.

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

Tabla I. Tabla de contingencia 2x2

Pérez Echeverría (1990) clasifica dichas estrategias en niveles de dificultad (niveles 1 a 3 si se usan sólo 1 a 3 de las 4 celdas de la tabla; nivel 4 si se usan las 4 celdas con estrategias aditivas y nivel 5 si se usan las 4 celdas con estrategias multiplicativas).

Otro problema es que muchas personas forman sus propias teorías sobre la relación entre variables en la tabla de contingencia que les impide evaluar correctamente la asociación (Chapman y Chapman, 1969), fenómeno conocido como “correlación ilusoria”. Por otro lado, Estepa (1993) describe la concepción causal, según la cual 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 y la local, cuando el sujeto basa su juicio en sólo una parte de los datos.

Una experiencia de enseñanza

Para contribuir a superar los problemas anteriores, se desarrollaron materiales didácticos para facilitar la comprensión de las tablas de contingencia en estudiantes universitarios de ciencias sociales, que fueron experimentados (Cañadas, 2011) dentro de dos grupos de clase regulares

de primer curso de Psicología (en total 94 alumnos). Se dedicó a la enseñanza 6 sesiones de 1 hora; cuatro de ellas en el aula tradicional en los grupos mencionados, dedicadas a la presentación de los temas. Otras dos sesiones fueron prácticas en el laboratorio de informática, donde cada alumno trabajó independientemente con el ordenador utilizando unas hojas de Microsoft Excel, preparadas para las prácticas. Cada grupo de teoría se dividió en tres de 15 alumnos, todos ellos impartidos por el mismo profesor. El material del curso con las actividades detalladas se encuentra disponible en la página web: <http://www.ugr.es/~analisisdedatos/webcurso/presentacion.html>. El contenido se ha organizado en cuatro lecciones:

1. *Tablas de contingencia, lectura e interpretación.* Se trata que los alumnos aprendiesen a: (a) Resumir datos sobre dos variables estadística en una tabla de contingencia; (b) Identificar las frecuencias dobles (c) Calcular las frecuencias relativas dobles, marginales y condicionales e interpretarlas; (d) Representar gráficamente los datos mediante diagrama de barras adosadas, diagrama de barras apiladas y gráfico tridimensional y (e) Calcular probabilidades simples, compuestas y condicionales a partir de datos de una tabla de contingencia.
2. *Asociación estadística, dependencia funcional e independencia.* Se trató de que los alumnos aprendiesen a: (a) Diferenciar la asociación estadística, dependencia funcional e independencia; (b) Reconocer el tipo de relación entre dos variables comparando las frecuencias condicionales; (c) Calcular las frecuencias esperadas en caso de independencia y (d) Analizar posibles explicaciones de una asociación estadística: relación causal, interdependencia, tercera variable explicativa o asociación espuria.
3. *El estadístico Chi-cuadrado y contrastes asociados.* La tercera lección, se dedicó a: (a) Dar una medida de la diferencia entre frecuencias observadas y esperadas en caso de independencia; (b) Calcular e interpretar el estadístico Chi-cuadrado y sus grados de libertad; (c) Comprender los pasos para llevar a cabo el contraste de independencia y el contraste de homogeneidad y (e) Comprender los supuestos del contraste Chi-cuadrado.
4. *Medidas de asociación.* Se trató de que los alumnos aprendiesen a: (a) Interpretar la intensidad de la dependencia entre dos variables en una tabla de contingencia; (b) Calcular e interpretar medidas de asociación en tablas 2x2: Coeficiente Phi de Pearson, Riesgo relativo y Razón de productos cruzados; (c) Calcular e interpretar medidas de asociación en tablas rxc; Coeficiente de contingencia de Pearson y V de Cramer y (d) Calcular e interpretar medidas de reducción del error de predicción de una variable, cuando se conoce el valor de la otra: Lambda de Goodman y Kruskal.

Para asegurar la validez del estudio las clases fueron observadas. Las interacciones en la clase también fueron grabadas en audio, para posteriormente poder comparar con la observación y anotar las principales incidencias y dudas planteadas por los estudiantes. Los profesores habituales de los cursos también asistieron a las sesiones.

Evaluación

También se prepararon diferentes pruebas de evaluación, para ser utilizadas en tres momentos del proceso de estudio, en cuya construcción se han seguido las recomendaciones psicométricas habituales para asegurar su calidad. El aprendizaje se evaluó con un cuestionario, compuesto de 6 ítems para cada uno de los temas y algunos problemas abiertos. En este trabajo presentamos los resultados obtenidos en relación con el tema 2 “asociación estadística, dependencia funcional y aleatoria”. Su comprensión fue evaluada con los 6 ítems que se reproducen a continuación, marcando en negrita las respuestas correctas en cada ítem y un problema abierto.

El ítem 1 evalúa el conocimiento que el estudiante adquiere de propiedades sencillas que permiten evaluar si dos variables de una tabla de contingencia son independientes: Las opciones a) y b) son correctas, pues en caso de independencia, todas las distribuciones condicionales por fila o columna coinciden. La c) es igualmente correcta, ya que la independencia implica esta relación de igualdad entre las frecuencias condicionales y las frecuencias marginales.

Ítem 1. Para que dos variables de una tabla de contingencia sean independientes, han de ser iguales:

- a. *Las frecuencias relativas condicionales por columnas.*
- b. *Las frecuencias relativas condicionales por filas.*
- c. *Las frecuencias relativas condicionales y frecuencias relativas marginales.*

El ítem 2 evalúa el conocimiento del procedimiento de cálculo de las frecuencias esperadas en una tabla de contingencia. La respuesta correcta es la c), pues $e_{ij} = \frac{f_{i.} \times f_{.j}}{n}$. En la respuesta a) se confunde las frecuencias absolutas dobles con las frecuencias absolutas marginales, en el cálculo de las frecuencias esperadas; en la b) se cambia las frecuencias relativas dobles por las frecuencias absolutas marginales, en el proceso de cálculo y en la d) se confunde las frecuencias relativas marginales con las frecuencias absolutas marginales, además de que falta el total de la muestra, en el cálculo.

Ítem 2. Las frecuencias esperadas se calculan mediante:

- Las frecuencias absolutas dobles y el total de la muestra.
- Las frecuencias relativas dobles y el total de la muestra.
- Las frecuencias absolutas marginales y el total de la muestra.
- Las frecuencias relativas marginales

Evalúa el conocimiento de la relación entre frecuencias relativas y marginales en caso de independencia. La respuesta correcta es la b), pues en caso de asociación podría darse la igualdad en un caso. La respuesta a) es incorrecta, ya que describe una propiedad de la independencia y no de la asociación. Asimismo lo es la respuesta c), puesto que puede ocurrir que en caso de asociación en unos casos se cumpla esta igualdad y en otros no.

Ítem 3. En caso de que haya asociación entre variables, las frecuencias relativas dobles:

- En todas las celdas son iguales al producto del total por fila y columna que le corresponda, es decir $h_{i,j} = h_i \cdot h_j$
- Puede ocurrir que coincida en alguna celda al producto del total por fila y columna que le corresponda, es decir $h_{i,j} = h_i \cdot h_j$
- Nunca son iguales al producto del total por fila y columna que le corresponda, es decir nunca se cumple $h_{i,j} = h_i \cdot h_j$

En el ítem 4 se desea conocer la interpretación que dan los alumnos a la relación existente entre diferentes celdas en la tabla 2x2 y el signo de la asociación. La respuesta adecuada es la b), pues en la diagonal principal están los valores que informan de dependencia directa, referidos a la presencia-presencia (A-B), y ausencia-ausencia (no A-no B). Por el contrario en las otras dos celdas, se da un solo carácter y el otro no y serían las celdas favorables a una asociación inversa. Las otras opciones son incorrectas.

Ítem 4. En las siguientes tablas 2x2 indicamos el tipo de asociación que informan las diferentes celdas ¿Cuáles de las siguientes tablas es correcta?

a.

	B	No B
A	Dep. directa	Dep. directa
No A	Dep. directa	Dep. inversa

b.

	B	No B
A	Dep. directa	Dep. inversa
No A	Dep. inversa	Dep. directa

c.

	B	No B
A	Dep. directa	Dep. inversa
No A	Dep. inversa	Dep. inversa

Este ítem trata de detectar la concepción causal de la asociación (Estepa, 1993). La respuesta adecuada es la c), pues si A es causa de B, entonces habrá asociación. La respuesta a) es incorrecta, pues causa si implica asociación, pero puede ser positiva o negativa. Asimismo es errónea la respuesta b); aquí aparece la falacia “correlación implica causalidad”.

Ítem 5. Indica cuál de las siguientes frases es cierta:

- Si hay una relación causal entre A y B, entonces habrá asociación positiva entre A y B
- Si al tomar datos de A y B encontramos asociación entre las variables, entonces habrá una relación causal entre A y B
- Si hay una relación causal entre A y B, entonces habrá asociación, que puede ser positiva o negativa.

Este ítem analizar la comprensión de la diferencia entre dependencia funcional y aleatoria. La respuesta correcta es la b), pues la dependencia aleatoria no implica un único valor de la variable dependiente al variar la independiente. La respuesta a) es incorrecta, pues da la definición de dependencia funcional. La respuesta c) es asimismo incorrecta, puesto que la dependencia funcional no es siempre directa.

Ítem 6. La diferencia entre la dependencia funcional y la dependencia aleatoria consiste en:

- En la dependencia aleatoria a cada valor de la variable independiente X le corresponde sólo un valor de la variable dependiente Y
- En la dependencia aleatoria, al variar X suele variar Y, pero no siempre
- La dependencia aleatoria puede ser directa o inversa, pero la funcional siempre es directa

También se propuso sobre este tema el siguiente problema abierto, que trata de evaluar la competencia del estudiante para establecer un juicio de asociación y sus estrategias

Problema. Un grupo de 200 personas aquejadas de ansiedad fue dividido aleatoriamente en dos subgrupos. Al primer grupo se ofreció un medicamento realmente efectivo y al otro se ofreció un placebo (medicamento sin efecto). Al cabo de un mes fueron interrogados sobre la eficacia conseguida, con el siguiente resultado:

	Su ansiedad ha disminuido	Siguen con mucha ansiedad
Medicamento	50	15
Placebo	96	39

¿Hay asociación entre el tipo de tratamiento (medicamento o placebo) y el efecto producido (la ansiedad disminuye o no)? ¿O son las variables independientes? Indica cómo has llegado a esta conclusión (puedes usar el método que prefieras)

Resultados y discusión

En la Tabla 2 se resumen las respuestas a los ítems correspondientes al Tema 2. Estos resultados muestran un alto grado de aprendizaje, que se refleja en las pocas respuestas en blanco y el alto porcentaje de respuestas correctas. Los mejores resultados se obtienen en el ítem 4 (identificación de celdas que informan del signo de la asociación en la tabla 2x2); ítem 5

(diferencia entre asociación y causalidad) y opción a) del ítem 6 (dependencia funcional). Por otro lado, los mayores errores se producen en la opción b) del ítem 3 (no se reconoce que en caso de independencia en alguna celda puede ser igual las frecuencias esperadas y observadas y la opción c) del ítem 1 (confusión de una propiedad de independencia).

	Apartado	Correcto	Incorrecto	En blanco
Ítem 1	a (Verdadera)	67 (71,3)	24 (25,5)	3 (3,2)
	b (Verdadera)	69 (73,4)	22 (23,4)	3 (3,2)
	c (Verdadera)	27 (28,7)	64 (68,1)	3 (3,2)
Ítem 2	a (Falsa)	74 (78,7)	18 (19,1)	2 (2,1)
	b (Falsa)	80 (85,1)	12 (12,8)	2 (2,1)
	c (Verdadera)	59 (62,8)	33 (35,1)	2 (2,1)
	d (Falsa)	87 (92,6)	5 (5,3)	2 (2,1)
Ítem 3	a (Falsa)	39 (41,5)	51 (54,3)	4 (4,3)
	b (Verdadera)	17 (18,1)	73 (77,7)	4 (4,3)
	c (Falsa)	59 (62,8)	31 (33)	4 (4,3)
Ítem 4	a (Falsa)	91 (96,8)	1 (1,1)	2 (2,1)
	b (Verdadera)	90 (95,7)	2 (2,1)	2 (2,1)
	c (Falsa)	91 (96,8)	1 (1,1)	2 (2,1)
Ítem 5	a (Falsa)	89 (94,7)	5 (5,3)	0 (0)
	b (Falsa)	80 (85,1)	13 (13,8)	1 (1,1)
	c (Verdadera)	79 (84)	15 (16)	0 (0)
Ítem 6	a (Falsa)	77 (81,9)	12 (12,8)	5 (5,3)
	b (Verdadera)	43 (45,7)	46 (48,9)	5 (5,3)
	c (Falsa)	60 (63,8)	29 (30,9)	5 (5,3)

Tabla 2. Frecuencias (y porcentajes) de respuestas en los ítems (n=94)

		Frecuencia	Porcentaje
Correcta	Contraste Chi-cuadrado	16	17,2
Correcta	Cálculo del coeficiente Phi de Pearson	14	15,1
Correcta	Cálculo de otros coeficientes	3	3,2
Correcta	Cálculo de dos coeficientes	3	3,2
Correcta	Compara las distribuciones condicionales	12	12,9
Correcta	Compara frecuencias observadas y esperadas	2	2,1
Parcialmente correcta	Comparar la suma de las diagonales	2	2,1
	No responde	41	44,1
Total		93	100

Tabla 3. Frecuencias y porcentajes de estrategias en el problema.

El problema planteado presenta una asociación moderada y los estudiantes en su mayoría lo han reconocido (74 estudiantes, 79,6%). La mayoría también indica que la dependencia es directa (60 estudiantes, es decir 64,5%), mostrando de nuevo un buen aprendizaje. En la Tabla 3 se presentan las estrategias utilizadas, la mayoría correctas; un 21% utiliza estrategias formales fruto del aprendizaje, un 17,1% estrategias intuitivas (estrategias producidas antes de la enseñanza sin formalidad matemática), pero todas ellas de nivel 5 en la clasificación de Pérez Echeverría; un 44,1% no indica la estrategia, posiblemente debido a falta de capacidad de argumentación, pues como se ha indicado la mayoría de estudiantes responde correctamente a la asociación y su signo.

Conclusiones

En este trabajo hemos descrito una experiencia de enseñanza de la asociación en tablas de contingencia, presentando un breve resumen de la evaluación del aprendizaje en uno de los temas. Los resultados indican un aprendizaje satisfactorio, y competencia en la detección de asociación en un problema abierto, con uso tanto de estrategias formales (estrategias producidas posteriores a la enseñanza), como de estrategias de nivel 5 en la clasificación de Pérez Echeverría. Se superan las concepciones causal y local, descritas por Estepa (1993). Ha sido menor la capacidad de argumentación de los estudiantes, muchos de los cuáles no fueron capaces de explicar su estrategia en el problema abierto. Como consecuencia, aunque el material se ha relevado de interés al usarlo con alumnos de Psicología, se plantea, así el reto de continuar este trabajo con nuevas investigaciones sobre la enseñanza y aprendizaje de las tablas de contingencia.

Agradecimiento: Proyecto EDU2010-1494; Beca FPU-AP2009-2807 y grupo FQM126 (Junta de Andalucía).

Referencias bibliográficas

- Cañadas, G. R. (2011). *Las tablas de contingencia para Psicología*. Departamento de Didáctica de la Matemática, Universidad de Granada. Granada.
- Chapman, L. J., & Chapman, J.P. (1969). Illusory correlation as an obstacle to the use of valid Psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.
- Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis de doctorado no publicada. Universidad de Granada.

- Godino, J. D., Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education*, 39 (1-2), 127-135.
- Inhelder, B., & 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.
- Pérez Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- Shaklee, H. y Mins, M. (1982). Sources of error in judging event covariations: Effects of memory demands, *Journal of Experimental Psychology Learning, Memory and Cognition*, 8(3), 208-224.
- Smedlund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165-174.