

ESTIMACIÓN DE LA ASOCIACIÓN POR ESTUDIANTES DE PSICOLOGÍA

PSYCHOLOGY STUDENTS' ESTIMATION OF ASSOCIATION

Cañadas, G.¹, Díaz, C.², Batanero, C.¹, Arteaga, P.¹

¹Universidad de Granada

²Universidad de Huelva.

Resumen. *Las tablas de contingencia son un instrumento necesario en el trabajo de los profesionales de psicología y, en general, en ciencias de la salud, pues la emisión de un diagnóstico requiere del juicio de asociación sobre dos variables cuyos datos se presentan en una tabla de contingencia. En este trabajo describimos un estudio exploratorio de la percepción de la asociación y estimación de su intensidad en tablas de contingencia en una muestra de estudiantes de psicología (62 estudiantes de la*

re.ac.uk

resultados con un estudio previo de Estepa (1993) con estudiantes de Bachillerato.

Palabras clave: Tablas de contingencia, asociación, estimación de la asociación, teorías previas, enseñanza universitaria.

Abstract. *Contingency tables are a necessary tool for the work of psychology professionals and, in general, for health sciences, since providing a diagnostic requires from the professional an association judgment in a contingency table. This paper describes an exploratory study of the perception of the association and estimation of its intensity in contingency tables in a sample of psychology students (62 students from the University of Granada and 62 from the University of Huelva). We compare the obtained results between the two universities, and we also compare our results with a previous study of Estepa (1993) carried out with pre-university students.*

Key words: contingency tables, association, estimating association, previous theories, university teaching.

INTRODUCCIÓN

Un modo común de presentar la información estadística es en forma de tabla de doble entrada o tabla de contingencia, a la que se presta poca atención en la enseñanza universitaria, suponiendo que su lectura e interpretación son sencillas. Dichas tablas se presentan con frecuencia en las actividades de diagnóstico y evaluación psicológica, donde el especialista se enfrenta a diferentes síntomas que pueden estar asociados o no con una patología. Con esta información, se podrá evaluar si existe o no asociación entre las correspondientes variables (Díaz y Gallego, 2006). Por otro lado, el razonamiento sobre datos bivariados es uno de los temas de aprendizaje prioritarios en los cursos universitarios de estadística (Zieffler, 2006).

El objetivo de este trabajo es evaluar la comprensión intuitiva de los estudiantes que ingresan en la Licenciatura de Psicología sobre la asociación en tablas de contingencia. Más concretamente pretendemos estudiar sus juicios de asociación en función de diversas variables y la precisión con que se estiman los coeficientes de asociación. Otro objetivo es comparar los resultados con los de un trabajo previo de Estepa (1993) con estudiantes del curso preuniversitario.

INVESTIGACIONES PREVIAS

Las investigaciones iniciales sobre este tema las inician Inhelder y Piaget (1955), quienes conciben la asociación como último paso en el desarrollo del razonamiento probabilístico y describen las estrategias usadas en los juicios de asociación en tablas 2x2 similares a la Tabla 1 a diferentes edades. Los estudios sobre asociación estadística son numerosos en Psicología, algo lógico considerando su importancia en el diagnóstico y evaluación, donde se han descrito estrategias erróneas en los juicios de asociación y estimación incorrecta de la asociación. Sin embargo este tipo de estudios son muy escasos en Didáctica de la Matemática.

Síntoma	Con enfermedad	Sin la enfermedad	Total
Presente	A	B	A+B
Ausente	C	D	C+D
Total	A+C	B+D	

Tabla 1. Ejemplo de tabla de contingencia en una prueba médica

El estudio de la precisión en el juicio de asociación ha sido llevado a cabo por muchos investigadores. Un ejemplo, es la investigación de Crocker (1981), quien indica que los juicios de asociación incrementan su exactitud cuando están presentes diferentes condiciones: Cuando los datos se presentan simultáneamente y las frecuencias son bajas; se presentan en forma de tablas; los sucesos que covarían lo hacen juntos en el tiempo; se elimina de las tareas la necesidad de obtener muestras; los datos no son ambiguos y se elimina la necesidad de recordar casos o estimar frecuencias.

Allan y Jenkins (1983) se centran en dos principios en los que se basan los patrones de los juicios de asociación: (1) Se tiende a basar los juicios en la diferencia entre casos confirmatorios (celda A en la tabla 1) y contradictorios (celda D en la tabla 1), y (2) la compatibilidad causal entre las variables independiente y dependiente juega un papel crítico. En el trabajo de Erlick y Mills (1967) se expone que la asociación negativa se estima como muy próxima a cero. Además aparecen tres factores que influyen en los juicios de asociación en el trabajo de Arkes y Harkness (1983), que son: (1) El valor de la frecuencia en la casilla "A" (tabla 1) parece tener mayor impacto en las contingencias estimadas; (2) la etiquetación de las filas y columnas puede drásticamente influenciar las contingencias estimadas, y (3) la presencia de números pequeños en las casillas puede influir en una sobreestimación.

Otros autores han estudiado la influencia de las teorías previas sobre el contexto del problema en la exactitud de la estimación de la asociación (Jennings, Amabile y Ross, 1982; Wright y Murphy, 1984; Alloy y Tabacnick, 1984; Meiser y Hewstone, 2006). La estimación de la asociación es más precisa si las personas no tienen teorías respecto al tipo de asociación sobre los datos. Si las teorías previas del sujeto indican el mismo

tipo de asociación que reflejan los datos empíricos, los sujetos tienden a sobre estimar el valor del coeficiente de asociación. Pero cuando los datos no reflejan los resultados esperados por estas teorías, aparece en los sujetos un conflicto cognitivo y se suelen guiar por sus teorías, más que por los datos. Así puede resultar una estimación menor o mayor del coeficiente de asociación, dependiendo de los conocimientos previos sobre la situación que tengan los sujetos y de los datos proporcionados.

Chapman y Chapman (1969) describen un razonamiento común, que se denomina *correlación ilusoria*. Consiste en formarse teorías sobre la relación entre variables que impide evaluar correctamente las contingencias empíricas. Lleva a la percepción de una relación donde no existe ninguna, o bien a la percepción de una relación más fuerte que la que existe en realidad.

La asociación entre variables puede ser debida a la existencia de una relación causa-efecto unilateral (una variable es causa de la otra), pero también según Barbancho (1973), a la interdependencia (cada variable afecta a la otra), dependencia indirecta (una tercera variable afecta a otras dos), concordancia (coincidencia en preferencia u ordenación de la misma serie de datos por dos jueces) y covariación espúrea o casual. La comprensión de la asociación implicaría, además de la exactitud en el juicio, comprender estos tipos de relaciones entre las variables. Sin embargo, Estepa (1993), quien estudia las concepciones que muestran los sujetos respecto a la asociación, describe la *concepción causal*, según la cual el sujeto sólo considera la asociación entre variables si puede adjudicarse a la presencia de una relación de causa - efecto entre las mismas. También define la *concepción unidireccional* como el caso en que el estudiante no admite la asociación inversa, considerándose la intensidad de la asociación, pero no su signo y considera la asociación inversa como independencia. En un estudio posterior (Batanero, Estepa y Godino, 1997) muestran que la concepción causal no mejora con la enseñanza.

MÉTODO

La muestra estuvo formada por un total de 124 alumnos de primer año de la Licenciatura en Psicología de las Universidades de Granada (62 alumnos) y Huelva (62 alumnos), que cursaban una asignatura de Análisis de Datos I. Aunque todavía no habían estudiado el tema, los alumnos han estudiado estadística y probabilidad en la Educación Secundaria y Bachillerato. El cuestionario utilizado está adaptado de Estepa y Batanero (1995). En la Figura 1 se presenta el primero de los ítems. El resto de los ítems se presentan como anexo, exceptuando las preguntas b) y c) que son idénticas en todos ellos.

Las variables de tarea tenidas en cuenta en el cuestionario (Tabla 2) fueron las siguientes:

1. *Signo de la asociación entre las variables*, considerándose los tres casos posibles: Dependencia directa, dependencia inversa e independencia.
2. *Intensidad de la dependencia*, medida mediante el coeficiente Phi de Pearson en tablas 2x2 y con el coeficiente V de Cramer en tablas 2x3. Se eligió un ítem de intensidad moderada-baja y dos de intensidad moderada-alta.
3. *Concordancia entre los datos y las teorías previas sugeridas por el contexto*. Se usaron dos ítems en que coincide la asociación empírica en los datos con las

expectativas previas, otro donde no coincide y uno donde no hay teorías previas.

4. *Tipo de covariación.* Usamos las tres categorías de Barbancho (1973): Dependencia causal unilateral, interdependencia y dependencia indirecta.

Ítem 1. Se pretende estudiar si el sufrir insomnio tiene relación con los trastornos de estrés. En una muestra de 250 personas observadas se obtuvieron los siguientes resultados:

	Padece estrés	No padece estrés
Tener insomnio	90	60
No tener insomnio	60	40

a. Utilizando estos datos razona si sufrir insomnio tiene relación con padecer estrés.

b. Indica como has usado los datos de la tabla para llegar a tu conclusión

c. Asigna una puntuación entre 0 (mínimo) y 1 (máximo) según la intensidad de esta relación, marcando una cruz en el punto de esta escala que creas adecuado:

Figura 14. Ejemplo de ítem

	Tabla 2x2			Tabla 2x3
	Item 1	Item 2	Item 3	Item 4
Dependencia	Independencia	Inversa	Directa	Directa
Valor del coeficiente de asociación	0	-0,62	0,67	0,37
Concuerta con teoría previa	No	Si	No hay teoría	Si
Tipo covariación	Interdependencia	Causal unilateral	Dependencia indirecta	Causal unilateral

Tabla 2. Variables de tarea en los ítems

RESULTADOS Y DISCUSIÓN

Juicio de asociación

En el apartado (a) de cada ítem se pide a los alumnos un juicio sobre la posible asociación entre las variables presentadas en la tabla. Para evaluar la competencia de los estudiantes, se presenta en la Tabla 3 el porcentaje de estudiantes que considera (o no) la existencia de relación entre las variables. Añadimos en la cuarta columna el coeficiente de asociación real calculado con los datos del ítem.

Estimación de la asociación por estudiantes de psicología

Ítem	Consideran relación	Consideran independencia	No contestan	Coefficiente de asociación	Teorías previas
1	42 (67,7)	19 (30,6)	1 (1,6)	0	En contra
2	60 (96,8)	2 (3,2)	0 (0)	-0,62	A favor
3	53 (85,5)	7 (11,3)	2 (3,2)	0,67	No hay
4	57 (91,9)	0(0)	5 (8,1)	0,37	A favor

Tabla 3. Porcentaje de sujetos según asociación considerada en cada ítem

La mayoría de los alumnos indica la existencia de asociación en todos los ítems, con mayor frecuencia cuando ésta se confirma con los datos, pero también en el ítem 1, donde los datos presentan independencia perfecta. Este resultado puede explicarse por la correlación ilusoria, descrita por Chapman y Chapman (1969). En el estudio de Estepa (1993) sólo el 31,4% de estudiantes indican que hay asociación en el caso de independencia perfecta, aunque los datos numéricos del ítem son exactamente los mismos en las dos investigaciones. La diferencia entre nuestro ítem y el de Estepa es el contexto, más relacionado con el campo de la psicología en nuestra investigación. Nuestros estudiantes muestran un mayor efecto de las teorías previas en este ítem, así como la concepción causal de la asociación, ligando los conceptos de asociación y causalidad.

Nuestros resultados son mucho mejores en el ítem 2 (asociación inversa), que en el estudio de Estepa, donde sólo el 47,1% de estudiantes consideran la existencia de asociación. El cambio en este ítem es el contexto, aumento de la intensidad en nuestro caso y del tamaño de muestra. Ello ha hecho que apenas se haya presentado en nuestros estudiantes la concepción unidireccional de la asociación descrita por Estepa.

Los resultados en los dos estudios en el ítem 3 son similares (88,2% en Estepa), donde sólo se aumentó las frecuencias, sin variar ni el contexto ni la intensidad de asociación. Finalmente mejoran un poco los resultados en el ítem 4 (88,2% de estudiantes consideran asociación en el estudio de Estepa), donde se aumentó un poco la intensidad de la asociación manteniendo fijas el resto de variables.

Estimación del coeficiente de asociación

En el apartado (c) de cada ítem se pedía al estudiante una puntuación entre 0 y 1 según la intensidad que ellos percibiesen en la asociación las variables, punto no analizado por Estepa (1993). En la tabla 4 presentamos el valor medio obtenido en el conjunto de la muestra, y en cada una de las universidades. Dicho valor sería una estimación del coeficiente de asociación (no se solicitó especificar el signo).

La mayor precisión se da en la estimación es el ítem 3, donde los alumnos no tienen teoría al respecto; el valor medio estimado es prácticamente idéntico al empírico en ambas provincias, y en la muestra global. Hay una sobre-estimación del coeficiente en los otros tres ítems para ambas provincias y la muestra global, observándose el efecto de las teorías previas de los estudiantes.

En el ítem 1, que corresponde al caso de independencia perfecta, el valor medio estimado fue 0,44, tanto en cada muestra, como en el global. Los alumnos mostraron en este ítem sus teorías previas al respecto, como mostramos a continuación en algunos ejemplos de respuestas: *“Si puede tener algo de relación, puesto que padecer estrés puede ser una causa de tener insomnio ya que las preocupaciones y de más elementos influyen”, “En mi opinión tener insomnio si tiene relación con padecer estrés, ya que la*

mayoría de personas que tienen insomnio padecen estrés”, o “*Si tiene relación, pues las personas que padecen insomnio no descansan bien y les causa un estrés extra al que sumar al que puede ocasionarte otra sería de factores externos*”. Por otro lado, en este ítem la mayor frecuencia aparece en la celda (A), es decir la primera de las frecuencias absolutas de la tabla que corresponde a la presencia de los dos caracteres y que, según Arkes y Harkness (1983), tiene mayor impacto sobre la atención que las otra celdas. Todo ello podría haber influido en la estimación de una asociación no existente en los datos.

Ítem	Valor medio estimado en Granada	Valor medio estimado en Huelva	Valor medio estimado en Total	Coefficiente de asociación	Teorías previas
1	0,44	0,44	0,44	0	En contra
2	0,68	0,73	0,71	-0,62	A favor
3	0,65	0,68	0,67	0,67	No hay
4	0,78	0,81	0,80	0,37	A favor

Tabla 12. Estimación del coeficiente de asociación

Para el ítem 2 (dependencia inversa), se estima un coeficiente de asociación próximo al empírico en ambas Universidades, algo inferior en Granada. De este modo en nuestros alumnos no aparece un porcentaje apreciable de concepción unidireccional descrita por Estepa (1993) ni en cada muestra parcial ni en el global, pues además, la mayoría de los alumnos indica que hay asociación (no se pidió el signo de la asociación).

En el ítem 4, correspondiendo a una tabla 2x3 con una asociación positiva, se puede observar la gran diferencia que hay entre la estimación y su valor verdadero. Los estudiantes sobre estiman la asociación, guiándose por sus teorías previas que coinciden con el tipo de asociación en los datos. Estas teorías se manifestaron en los estudiantes al ser un contexto tan cercano a ellos (horas de estudio-aprobar o suspender), se han dejado llevar por su experiencia personal. Esto ocurre en ambas Universidades, aunque la diferencia con el valor verdadero es menor en los estudiantes de Granada.

Los estudiantes de Granada en todos los ítems estiman en promedio un coeficiente más bajo que los estudiantes de Huelva. Al realizar un contraste *t* de comparación de medias relacionadas, no obtuvimos diferencias estadísticamente significativa para ninguno de los ítems y los intervalos de confianza del 95% cubrieron el valor 0. Asimismo, se realizaron contrastes *F* para comparar las desviaciones típicas de cada par de variable, y tampoco se obtuvo diferencia estadísticamente significativa ni valores apreciables en el intervalo de confianza del 95% para el cociente de varianzas. De todo ello se deduce que las respuestas de los estudiantes fueron similares, a pesar de la diferencia de contexto educativo.

CONCLUSIÓN

Los resultados muestran que en los estudiantes en su mayoría juzgan la asociación, incluso en los casos en que ésta no existe, observándose el fenómeno de la correlación ilusoria y el efecto de las teorías previas sobre la precisión en la estimación del coeficiente de asociación. Respecto a las concepciones descritas por Estepa (1993) se ha observado la concepción causal, pero no así la unidireccional. La estimación del

coeficiente de asociación mejora en nuestro caso, comparando con los resultados de Estepa (1993), en todos los ítems salvo en este caso de independencia. Los resultados son similares en las dos universidades participantes.

Según Schield (2006), una persona culta debiera poder leer críticamente las tablas que encuentra en la prensa, Internet, medios de comunicación, y trabajo profesional. Esto supone no sólo la lectura literal, sino identificar las tendencias y variabilidad de los datos, lo que incluye la emisión correcta de un juicio de asociación. Cazorla (2002) indica que las tablas 2x2 pueden utilizarse para comunicar información y como instrumento de análisis de datos, así como para retener en la memoria una gran cantidad de información en forma eficiente. Tienen también un papel esencial en la organización, descripción y análisis de datos, al ser un instrumento de transnumeración, forma básica de razonamiento estadístico que proporciona nueva información, al cambiar de un sistema de representación a otro (Wild y Pfannkuch, 1999).

Todas estas razones, y los resultados de nuestro trabajo indican la necesidad de continuar la investigación didáctica al respecto. Puesto que la concepción causal y la correlación ilusoria no parecen mejorar con una enseñanza tradicional (Batanero, Estepa y Godino, 1997) tenemos el proyecto de continuar este trabajo, diseñando una propuesta alternativa de enseñanza que incluyan actividades en que se enfrente a los estudiantes con estos sesgos y se les ayude a superarlos. Esta propuesta se experimentará y se evaluarán a los alumnos para comparar las concepciones intuitivas con las que adquieran como consecuencia de la enseñanza.

Agradecimientos

Proyecto EDU2010-14947 y becas FPU-AP2009-2807 (MCINN- FEDER), y FPU-AP2007-03222 (MEC-FEDER) y grupo FQM126 (Junta de Andalucía).

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.
- Alloy, L.B. y Tabachnik, N. (1984). Assessment of covariation by humans and animals: The Joint influence of prior expectations and current situational information, *Psychological Review*, 91, 112-149.
- Arkes, H.R. y Harkness, A.R. (1983). Estimates of contingency between two dichotomous variables, *Journal of Experimental Psychology: General*, 112, 1, 117-135.
- Barbancho, A. G. (1973). *Estadística elemental moderna*. Barcelona. Ariel cuarta edición, reimpresión de 1975.
- 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.
- Cazorla, I. (2002). *A relação entre a habilidades viso-pictóricas e o domínio de conceitos estatísticos na leitura de gráficos*. Tesis Doctoral. Universidad de

Campinas.

- Chapman, L. J. y Chapman, J.P. (1969). Illusory correlation as an obstacle to the use of valid Psychodiagnostic signs, *Journal of Abnormal Psychology*, 74, 271-280.
- Crocker, J. (1981). Judgment of covariation by social perceivers, *Psychological Bulletin*, 90, 2, 272-292.
- Díaz, J. y Gallego, B. (2006). Algunas medidas de utilidad en el diagnóstico, *Revista Cubana de Medicina General Integrada*, 22(1).
- Erlick, D.E. y Mills, R.G. (1967). Perceptual quantification of conditional dependency, *Journal of Experimental Psychology*, 73, 1, 9-14.
- Estepa, A. (1993). *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.
- Estepa, A. y Batanero, C. (1995). Concepciones iniciales sobre la asociación estadística, *Enseñanza de las Ciencias*, 13(2), 155-170.
- Inhelder, B. y Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris: Presses Universitaires de France.
- Jennings, D.L., Amabile, T.M. y Ross, L. (1982). Informal covariation assessment: Data-based versus theory-based judgments, en D. Kahneman, P. Slovic y A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press: Nueva York, pp. 211-230.
- 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.
- Schild, M. (2006). Statistical literacy survey analysis: reading graphs and tables of rates percentages. *Proceedings of the Sixth International Conference on Teaching Statistics*. Ed B. Phillips. Cape Town: International Statistical Institute and International Association for Statistical Education. Online: <http://www.stat.auckland.ac.nz/~iase>.
- Wright, J.C. y Murphy, G.L. (1984). The utility of theories in intuitive statistics: the robustness of theory-based judgments, *Journal of Experimental Psychology General*, 113(2), 301-322.
- Zieffler, A. (2006). *A longitudinal investigation of the development of college students' reasoning about bivariate data during an introductory statistics course*. Tesis doctoral. Universidad de Minnesota.

ANEXO. CUESTIONARIO

Item 2. Un psicólogo infantil estudia si existe relación entre ser hijo único (sin hermanos) y ser un niño problemático. Para ello se han observado 250 niños obteniendo los resultados siguientes

	Niño problemático	Niño no problemático
Tiene hermanos	40	100
Hijo único	100	10

- a. Utilizando estos datos razona si hay alguna relación entre ser un niño problemático y tener o no hermanos.

Item 3. Se quiere saber si sufrir o no de alergia tiene relación con llevar una vida sedentaria (llevar una vida sin realizar ningún tipo de ejercicio físico). Para ello obtuvimos los siguientes resultados de 300 sujetos:

	Sufre alergia	No sufre alergia
Forma de vida sedentaria	130	30
Forma de vida no sedentaria	20	120

- a. Utilizando estos datos razona si padecer alergia tiene o no relación con llevar una vida sedentaria en estos sujetos.

Item 4. La siguiente tabla nos indica el número de estudiantes que aprobaron o suspendieron un examen teniendo en cuenta el tiempo que cada estudiante dedicó a prepararlo

	Menos de 5 h	Entre 5 y 10 h	Más de 10 h	Total
Aprobados	5	15	51	71
Suspensos	20	7	2	29
Total	25	22	53	100

- a. Utilizando estos datos razona si aprobar o suspender el examen tiene relación con el tiempo dedicado a prepararlo.

