

UN ESTUDIO INICIAL DE SESGOS EN EL RAZONAMIENTO SOBRE PROBABILIDAD CONDICIONAL EN ALUMNOS UNIVERSITARIOS

Assumpta Estrada, Universitat de Lleida

Carmen Díaz, Universidad de Granada

Inmaculada de la Fuente, Universidad de Granada

RESUMEN

Presentamos un estudio de evaluación de los sesgos en el razonamiento condicional de una muestra de 159 estudiantes de Matemáticas, Magisterio y Psicología. Los resultados indican que el tema no resulta sencillo y evidencian la necesidad de potenciar la formación estadística sobre todo entre los futuros profesores.

ABSTRACT

In this paper we present the results from assessing biases in conditional reasoning, which was carried out on a sample of 159 future teachers, Mathematics and Psychology students'. The results suggest the need to improve the statistical training of these students, in particular the future teachers.

INTRODUCCIÓN

La probabilidad condicional es fundamental en las aplicaciones de la Estadística, porque permite incorporar cambios en nuestro grado de creencia sobre los sucesos aleatorios a medida que adquirimos nueva información. Es también un concepto teórico básico requerido en la construcción del espacio muestral producto. Por ello, su correcta comprensión y el razonamiento sobre la misma son requisitos indispensables en el estudio de la inferencia estadística, tanto clásica como bayesiana, así como en el estudio de la asociación entre variables, la regresión y los modelos lineales. En el terreno profesional e incluso en la vida cotidiana, la toma de decisiones acertadas en situaciones de incertidumbre se basa en gran medida en el razonamiento condicional.

Sin embargo, éste es un tema en que se observa la compleja interrelación entre intuición y conocimiento formal en el campo de la probabilidad (Batanero, Henry y Parzys, 2005). La Psicología del Razonamiento, y las investigaciones en Didáctica de la Probabilidad, muestran la existencia de intuiciones incorrectas, sesgos de razonamiento y errores de comprensión y aplicación de este concepto (Tarr y Lennin, 2005). Algunos de ellos están bastante extendidos y una enseñanza formal de la probabilidad es insuficiente para superarlos. Es necesario que el sujeto sea consciente de estas dificultades y aprenda a afrontar los problemas condicionales con unas herramientas adecuadas.

En este trabajo abordamos el estudio de sesgos en el razonamiento condicional en futuros profesores y estudiantes de Psicología, completando nuestras aportaciones anteriores al respecto (Díaz, 2005; Díaz y de la Fuente, 2006; Estrada y Díaz, 2006).

INVESTIGACIÓN PREVIA Y ANÁLISIS DE ÍTEMS

La investigación sobre probabilidad condicional ha sido muy extensa tanto en Psicología como en Educación Matemática (como ejemplo citamos los trabajos de Ojeda, 1995, Gras. y Totohasina, 1995; Sánchez, 1996, Martignon, y Wassner, 2002 y Huerta y Lonjedo, 2005). Debido a las limitaciones de espacio, no es posible describir todas estas investigaciones en este trabajo; por ello remitimos al lector a Díaz (2005) y Tarr y Lannin (2005) para una revisión sobre el tema.

Nos centraremos en particular en cuatro sesgos. El primero, denominado *falacia de la conjunción* (Tversky y Kahneman, 1982), es la creencia de que es más probable la intersección de dos sucesos que la de cada uno de ellos por separado. Los sujetos que lo manifiestan elegirían el distractor a) en el ítem 1. Según Tversky y Kahneman, el error se produce como resultado de considerar la conjunción como más representativa de la población generadora que cada evento separado, lo que ocurre cuando uno de los dos sucesos parece altamente probable. La falacia hace que los sujetos sólo tengan en cuenta este suceso y no sean conscientes de que su intersección con otro reduce su probabilidad.

Ítem 1. Supón que Carlos Ferrero alcanza la final de Roland Garros en 2004. ¿Cuál de los siguientes sucesos consideras más probable?

- a) **Carlos Ferrero pierde el primer set.**
- b) Carlos Ferrero pierde el primer set pero gana el partido.
- c) Los dos sucesos son iguales de probables.

En segundo lugar, hemos estudiado la *falacia de la condicional transpuesta*, por la que muchos estudiantes no discriminan adecuadamente entre las dos direcciones de la probabilidad condicional $P(A/B)$ y $P(B/A)$ (Falk, 1986). Este error se ha observado en problemas de contextos médicos, como el ítem 2, donde se confunde la probabilidad de tener una enfermedad cuando ha sido positivo el test de diagnóstico, con la probabilidad de un resultado positivo en el test de diagnóstico, dado que se tiene la enfermedad (Eddy, 1982), dando la respuesta c). Puesto que una prueba diagnóstica se diseña para detectar una cierta enfermedad, la probabilidad de que una persona enferma obtenga un resultado positivo, es muy alta. La probabilidad de que una persona sana tenga un resultado positivo en la prueba es muy pequeña, pero no imposible y el alto número de personas sanas en la población hace que la probabilidad de estar enfermos si el test ha sido positivo no sea, en general, demasiado alta (Eddy, 1982).

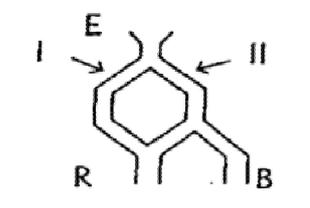
Ítem 2. Un test diagnóstico de cáncer fue administrado a todos los residentes en una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

- a) Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
- b) **Predecir un resultado positivo en el test de diagnóstico, si la persona tiene cáncer.**
- c) Tengo la misma confianza en ambas predicciones.

La comprensión de la relación de condicionalidad también se dificulta si la secuencia temporal de los sucesos no coincide con el orden dado en el condicionamiento. El problema 3 (Ojeda, 1995) sirve para evaluar este error (*falacia del eje temporal*), ya que el suceso condicionante (caer en el orificio R es posterior en el tiempo al suceso cuya probabilidad se evalúa (pasar por I). La secuencia temporal dificulta la identificación correcta del espacio muestral del experimento y los estudiantes dan erróneamente la respuesta 0,5, no teniendo en cuenta las bolas que caen en el orificio B. Un razonamiento muy simple permite encontrar la solución correcta: Si se dejan caer, por ejemplo, 100 bolas por E, 50 irán por I y otras 50 por II; De estas últimas, 25 caen, aproximadamente por R, así como las 50 que van por I. En total de cada 75 bolas que caen en R, aproximadamente 50 vienen por I, luego la probabilidad es $2/3$, es decir, 0,66.

Ítem 3. Una bola se suelta en E. Si sale por R, ¿cuál es la probabilidad de que haya pasado por el canal I?

- a) 0,50.
- b) 0,66.**
- c) 0,33.
- d) No se puede calcular.



Finalmente analizamos dos sesgos ligados a la idea de independencia, íntimamente ligada a la de probabilidad condicional. En el ítem 4 (Sánchez, 1996) la respuesta a) indica la confusión entre sucesos independientes y sucesos mutuamente excluyentes, lo cual es una idea incorrecta, porque precisamente dos sucesos excluyentes son dependientes ya que uno de ellos no puede ocurrir cuando ocurre el otro. La respuesta b) indica la creencia incorrecta de que dos sucesos simultáneos en el tiempo no pueden ser independientes. La única posible respuesta correcta es c), donde se presenta la definición formal de independencia de dos sucesos, que en este caso es simple de comprobar puesto que $P(\text{rey de oros}) = 1/40$; $P(\text{rey}) = 1/10$ y $P(\text{oros}) = 1/4$.

Ítem 4. Se extrae una carta al azar de una baraja española (40 cartas con 4 palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota, caballo y rey). Sea A el suceso "se extrae una carta de oros" y B el suceso "se extrae un rey", ¿los sucesos A y B son independientes?

- a) No son independientes, porque en la baraja hay un rey de oros.
- b) Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda carta para mirar si es oros.
- c) Sí, porque $P(\text{rey de oros}) = P(\text{rey}) \times P(\text{oros})$.**
- d) No, porque $P(\text{rey} / \text{oros}) \neq P(\text{rey})$.

MÉTODO Y RESULTADOS

Los cuatro ítems anteriores forman parte de un cuestionario más amplio sobre evaluación de la comprensión y sesgos relacionados con la probabilidad condicional (Díaz, 2004). Este cuestionario fue construido con un riguroso proceso metodológico (Díaz y de la Fuente, 2005). Cada uno de los ítems fue elegidos entre varios posibles, para evaluar un cierto sesgo o conocimiento, a partir de un juicio de expertos en el que participaron 10 investigadores, españoles y latinoamericanos, todos los cuales habían realizado una investigación, sobre didáctica de la probabilidad o de la probabilidad condicional. También se tuvo en cuenta los resultados de pruebas piloto de ítems realizadas con estudiantes de psicología (entre 50 y 177 estudiantes según ítem).. La fiabilidad de consistencia interna del cuestionario completo (estimada con una muestra de 591 estudiantes de psicología dio un valor del coeficiente Alfa de Crombach igual a 0,796 y la fiabilidad de prueba repetida, estimada mediante una muestra de 106 estudiantes a los que se pasó dos veces el cuestionario, con un mes de intervalo dio un valor igual a 0,871.

La muestra participante en el trabajo que ahora presentamos está formada por un total de 159 estudiantes, distribuidos en la forma siguiente: 65 estudiantes de Magisterio que seguían una asignatura optativa de Estadística Aplicada a la Educación, 37 estudiantes de 5º curso de Matemáticas, que seguían un curso optativo de Didáctica de la Matemática y 57 estudiantes de Psicología que seguían un curso obligatorio de Análisis de Datos. Todos ellos habían estudiado probabilidad, aunque con diferente profundidad en las asignaturas citadas. La decisión de tomar estos estudiantes para la muestra vino motivada por nuestro interés en analizar el razonamiento condicional de futuros profesores de Matemáticas, que posiblemente tengan que enseñar Probabilidad en su vida profesional (bien en la Educación Primaria, bien en la Secundaria).

La muestra de alumnos de Psicología sirve de muestra de comparación, además de ser el área de enseñanza de dos de las autoras, y tener, por tanto un especial interés en evaluar posibles errores en el razonamiento de los estudiantes. Los cuestionarios fueron pasados durante una de las sesiones de clase en cada una de las asignaturas citadas. Se informó a los estudiantes de la finalidad de la investigación, pidiéndoles que completaran con cuidado las preguntas. Todos ellos colaboraron con interés. A continuación se presentan los resultados expresados, no en valores absolutos, sino en porcentajes.

Tabla 1. Resultados en el ítem 1

	Magisterio (n=65)	Matemáticas (n=37)	Psicología (n=57)
a) Carlos Ferrero pierde el primer set.	18,5	27,0	33,3
b) Carlos Ferrero pierde el primer set pero gana el partido.	13,8	16,2	10,5
c) Los dos sucesos son iguales de probables.	67,7	51,4	54,4
En blanco.	0,0	5,4	1,8

Los resultados de la Tabla 1 indican que la falacia de la conjunción no es tan frecuente como en la investigación de Tversky y Kahneman (1982), y en su lugar aparece el sesgo

de equiprobabilidad (Lecoutre y Durand, 1988), que consiste en considerar los dos sucesos equiprobables, en más de la mitad de los estudiantes de todos los grupos. No obstante, es claro que los estudiantes olvidan que el producto de probabilidades ha de ser menor que la probabilidad de cada suceso simple o no aplican esta propiedad, cuando se les presenta un problema en un contexto cotidiano

En la Tabla 2, presentamos los resultados del ítem 2, donde todos los grupos han presentado un patrón similar de respuesta: mayor elección de la alternativa c), falacia de la condicional transpuesta. Este resultado está de acuerdo con los obtenidos por Eddy (1982), así como en las investigaciones de Batanero y cols. (1996), y Díaz y de la Fuente (2006) sobre cálculo de probabilidades condicionales a partir de tablas de contingencia

Tabla 2. Resultados en el ítem 2

	Magisterio (n=65)	Matemáticas (n=37)	Psicología (n=57)
a) Predecir que una persona tiene cáncer, si ha dado positivo en el test de diagnóstico.	12,3	10,8	15,8
b) Predecir un resultado positivo en el test de diagnóstico, si la persona tiene cáncer.	40,0	37,8	31,6
c) Tengo la misma confianza en ambas predicciones.	46,2	45,9	49,1
En blanco.	1,5	5,5	3,5

El ítem 3 (tabla 3) fue especialmente difícil, siendo los resultados incluso peores que los que obtuvo Ojeda (1995) con estudiantes mexicanos de educación secundaria (40% resolvió correctamente el problema). La mayoría de los alumnos de todos los grupos ha escogido la alternativa a), que no tiene en cuenta las bolas que caen por B, haciendo una restricción indebida del espacio muestral. Han sido pocos los alumnos que han escogido la alternativa d), por lo que, en general, la falacia del eje temporal es inconsciente.

Tabla 3. Resultados en el ítem 3

	Magisterio (n=65)	Matemáticas (n=37)	Psicología (n=57)
a) 0'50.	47,7	73,0	75,4
b) 0'66.	27,7	8,1	7,0
c) 0'33.	15,4	13,5	5,3
d) No se puede calcular.	3,1	5,4	8,8
En blanco.	6,1	0	3,5

Para finalizar, en el ítem 4 (Tabla 4), son muy pocos los alumnos que escogen la respuesta correcta,. Un porcentaje apreciable de estudiantes considera que los sucesos no son independientes al ser compatibles (distractor a, error citado por Sánchez, 1996) o bien al ser simultáneos (distractor b, que corresponde a la concepción cronológica de la independencia según Gras y Totohasina, 1995).

Tabla 4. Resultados en el ítem 4

	Magisterio (n=65)	Matemáticas (n=37)	Psicología (n=57)
a) No son independientes porque en la baraja hay un rey de oros.	22,2	32,4	26,3
b) Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja. Luego sacamos una segunda carta para mirar si es oros.	37,3	48,6	15,8
c) Sí, porque $P(\text{rey oros}) = P(\text{rey}) \times P(\text{oros})$	36,1	10,8	31,6
d) No, porque $P(\text{rey/oros}) \neq P(\text{rey})$	8,3	8,1	17,5
En blanco.	0	0	8,8

En general la distribución de respuestas en los grupos es similar en todos los ítems, a pesar de la mayor formación en el caso de los estudiantes de matemáticas.

CONCLUSIONES

Aunque los resultados han de interpretarse con precaución debido al tamaño limitado de las muestras, la tendencia se repite en las nuevas muestras de estudiantes de Psicología (591 alumnos). Ello nos hace reflexionar sobre la manera en que preparamos a los estudiantes sobre este tema, en particular a los que serán profesores de Matemáticas en el futuro, quienes pueden transmitir las ideas incorrectas a sus estudiantes.

Como sugiere tanto nuestra investigación, como las que hemos citado en los antecedentes, el estudio de esta tema no es sencillo, ya que tiene una amplia variedad de matices y los alumnos lo asocian con la problemática de la causalidad y la temporalidad, teniendo dificultad en la visualización del espacio muestral, confundiendo independencia y exclusión, cambiando los términos de la probabilidad condicional, y asignando a la probabilidad conjunta un valor mayor que a la probabilidad simple, es decir, violando las reglas lógicas del cálculo de probabilidades condicionales. El profesor de matemáticas que debe enseñar probabilidad a sus alumnos ha de ser consciente de esta problemática, para comprender algunas dificultades de sus estudiantes, quienes encuentran a lo largo de su aprendizaje las mismas paradojas y situaciones contraintuitivas que aparecieron en el desarrollo histórico del cálculo de probabilidades (Batanero, Henry y Parzysz, 2005).

Por ello estamos de acuerdo con Rossman y Short (1995), que sugieren que en este tema debe presentarse a los estudiantes una variedad de aplicaciones en problemas reales, proponiendo situaciones interactivas y usando la tecnología para facilitar el aprendizaje. En concordancia con Nisbett y Ross (1980), la enseñanza de la probabilidad condicional

requiere también que los estudiantes confronten con sus propios razonamientos incorrectos y la integración con la enseñanza de la lógica condicional.

Nota: Trabajo apoyado por el Grupo PAI FQM-126, Junta de Andalucía y el Proyecto SEJ2004-00789, MEC y FEDER.

REFERENCIAS

- Batanero, C., Estepa, A., Godino, J. y Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27 (2), 151 – 169.
- Batanero, C., Henry, M. y Parzysz, B. (2005). The nature of chance and probability. En G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 16-42). Nueva York: Springer,
- Borovcnik, M. y Peard, R. (1996). Probability. En A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick y C. Laborde (Eds.), *International handbook in mathematics education* (Parte 1, pp. 239-288). Dordrecht: Kluwer.
- Díaz, C. (2004). *Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar*. Trabajo inédito de Investigación Tutelada. Departamento de Psicología Social y Metodología de las Ciencias del Comportamiento Universidad de Granada.
- Díaz, C. (2005). Evaluación de la falacia de la conjunción en alumnos universitarios. *Suma*, 48, 45-50.
- Díaz, C. y de la Fuente, I. (2006). Assessing psychology students' difficulties with conditional probability and bayesian reasoning. En A. Rossman y B. Chance (Editores). *Proceedings of ICOTS-7*. Salvador (Bahia, Brasil): International Association for Statistical Education. CD Rom
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. En D. Kahneman, P. Slovic y Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Estrada, A. y Díaz, C. (2006). Computing probabilities from two way tables. An exploratory study with future teachers. En A. Rossman y B. Chance (Editores). *Proceedings of ICOTS-7*. Salvador (Bahia, Brasil): International Association for Statistical Education. CD Rom
- Falk, R. (1986). Conditional Probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. (pp. 292 – 297). Victoria, Canada: International Statistical Institute.
- Gras, R. y Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle *Recherches en Didactique des Mathématiques*, 15(1), 49-95.
- Lecoutre, M. P. y Durand, J. L. (1988). Judgements probabilistes et modèles cognitifs: Étude d'une situation aleatoire. *Educational Studies in Mathematics*, 19, 357- 368.
- Huerta, P. y Lonjedo, M. A. (2005). The nature of the quantities in a conditional probability problem. Its influence in the problem resolution. Trabajo presentado en el IV CERME, European Research in Mathematics Education. Sant- Feliu de Guitxols, Gerona. Versión provisional on line: http://cerme4.crm.es/Papers%20definitius/5/HuertaLonjedo_CERME.doc.pdf
- Martignon, L. y Wassner, C. (2002). Teaching decision making and statistical thinking with natural frequencies. En B. Phillips (Ed.), *Proceedings of the Sixth International*

- Conference on Teaching of Statistics*. Ciudad del Cabo: IASE. CD ROM.
- Nisbett, R. y Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgments*. Englewood Cliffs, NJ: Prentice Hall
- Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO*, 5, 37-55.
- Rossmann, A. y Short, T. (1995). Conditional probability and education reform: Are they compatible? *Journal of Statistics Education*, 3(2). On line: <http://www.amstat.org/publications/jse/v3n2/rossman.html>
- Sánchez, E. (1996). Dificultades en la comprensión del concepto de eventos independientes. En F. Hitt (Ed.), *Investigaciones en Educación Matemática* (pp. 389-404). México: Grupo Editorial Iberoamericano.
- Tarr, J. E. y Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence? En G. A. Jones (ed.), *Exploring probability in school: Challenges for teaching and learning*, (pp. 215-239). Nueva York: Springer
- Tversky, A. y Kahneman, D. (1982). On the psychology of prediction. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 69-83). Cambridge, MA: Cambridge University Press.