

Errores en el cálculo de probabilidades en tablas de doble entrada en profesores en formación

Assumpta Estrada Roca, Universitat de Lleida
aestrada@matematica.udl.cat

Carmen Díaz Batanero, Universidad de Granada
mcdiaz@ugr.es

Resumen

La lectura e interpretación de tablas de doble entrada es un elemento básico de la cultura estadística de cualquier ciudadano y los futuros profesores, que serán los responsables de su enseñanza a los niños, deberían adquirir estas competencias durante su formación. Pero este hecho no tiene resonancia en España y la Estadística no forma parte de la mayoría de los currícula de los profesores en formación. En este trabajo presentamos los resultados de un estudio exploratorio sobre los conflictos semióticos de los profesores en formación en el cálculo de probabilidades en tablas de doble entrada.

Abstract

Being able of correctly reading and interpreting two-way tables is a basic component of statistical literacy for all citizens. Therefore future teachers who will be responsible of teaching statistics to children at school level should acquire these abilities along their training. However, this capacity is taken for granted in Spain and its teaching is not usually included in the curriculum for training teachers. In this study, we present the results of a small exploratory study that describes the future teachers' semiotic conflicts in solving elementary probability problems when data are given in a two-way table.

1. Introducción

La abundancia de información estadística presente en la vida cotidiana ha originado un interés por la formación estadística básica del ciudadano, independientemente de su profesión, clase social o nivel educativo. Es lo que se conoce como “*statistical literacy*” o “cultura estadística”, y se define como ‘capacidad para interpretar y evaluar críticamente la información estadística y tomar a partir de ella decisiones fundamentadas sobre los fenómenos estocásticos que las personas pueden encontrar en diversos contextos personales y profesionales’ (Gal, 2002).

Pero aunque en los últimos años la Estadística forma parte del currículo de Matemáticas y son cada vez más fuertes las voces que reclaman la cultura estadística para todos, no se percibe una respuesta consecuente en España, ni a nivel de la educación de los niños en las escuelas, ni en las facultades de Ciencias de la Educación, encargadas de formar al profesorado. Como señalan Silva y cols. (1999), el profesor en formación debiera estar motivado y contar con destrezas suficientes, para ir aprendiendo las técnicas estadísticas necesarias y así enseñar los contenidos estadísticos previstos en el currículo y ser capaz de evaluar la calidad de su trabajo profesional.

Sin embargo, en la práctica, los propios profesores reconocen que es la parte de las Matemáticas donde se saltan más materia, pues piensan que hay cosas más importantes sobre las que preparar a los alumnos para las evaluaciones finales, no siendo conscientes de sus propias dificultades en el tema (Gattuso y Panone, 2002). Incluso cuando la Estadística tiene una presencia innegable en los libros de texto, la formación específica de los profesores en Estadística y su didáctica es prácticamente inexistente y generalmente consiste en un aprendizaje rutinario de fórmulas, y no la perciben como una herramienta de trabajo multidisciplinar indispensable en su vida académica y profesional (Estrada, 2005).

Un último punto es la naturaleza interdisciplinar del tema, que hace que los conceptos estadísticos aparezcan en otras materias, como las Ciencias Sociales, la Biología, la Geografía, etc., donde los profesores, a veces se ven obligados a enseñar Estadística, lo que puede ocasionar conflictos cuando las definiciones o propiedades de los conceptos presentados no coinciden con las impartidas en la clase de Matemáticas.

Por otro lado, hay pocas investigaciones sobre las concepciones y conocimientos de los profesores en relación a la Estadística, y no se conocen las principales dificultades de los futuros profesores en muchos conceptos importantes. Creemos que todas estas razones nos llevan a la necesidad de potenciar la investigación específica sobre formación del profesorado en Estadística, y ello requiere también conocer cuáles son sus dificultades y errores en el aprendizaje de la materia.

Hemos centrado nuestro estudio en las tablas de doble entrada porque 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 cálculo de probabilidades, especialmente condicionales, que sin embargo son difíciles de evaluar y comprender (Díaz y de la Fuente, 2005). Pensamos también que los futuros profesores pueden presentar algunos de los conflictos descritos en Díaz (2005) en estudiantes de Psicología.

En primer lugar resumiremos las investigaciones previas y analizaremos la tarea consistente en la interpretación de una tabla estadística de doble entrada. Seguidamente completaremos nuestro trabajo previo (Estrada y Díaz, 2006), evaluando la capacidad de lectura e interpretación de probabilidades a partir de ella de una muestra de 117 profesores en formación de la Universitat de Lleida.

2. Investigaciones previas

En España los primeros trabajos sobre errores y concepciones de los profesores en ejercicio y en formación en el campo de la Probabilidad son los de Azcarate (1995), Serrano (1996) y Cardeñoso (1998), así como las publicaciones posteriores de estos autores.

No nos centraremos en el análisis de los resultados de estos trabajos que, en general, indican la necesidad de una mejor formación de los profesores en el terreno de la Probabilidad, sino que haremos un resumen de las investigaciones sobre interpretación de tablas de doble entrada, aunque éstas no se ocupen específicamente del colectivo de los profesores en formación, sino de estudiantes universitarios en general.

Investigación sobre tablas de contingencia

Una tabla de doble entrada sirve para presentar la distribución conjunta de dos variables estadísticas. En su forma más simple, cuando las variables poseen sólo dos categorías, toma la forma de la Tabla 1, que es la que usamos en este trabajo.

Tabla 1. Formato típico de la tabla de contingencia 2x2

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

La investigación didáctica sobre las tablas de contingencia se ha centrado preferentemente en analizar la capacidad de diversos tipos de estudiantes para identificar la asociación o relación estadística entre las variables A y B a partir de los datos presentados en este tipo de tabla. Una tarea típica en estas investigaciones se muestra en la tabla 2 (Estepa, 1993).

El estudio del razonamiento sobre la asociación estadística fue iniciado por Inhelder y Piaget (1955), quienes consideraron la asociación como la última etapa en la comprensión de la probabilidad, y por ello sólo la estudian con chicos a partir de los 13-14 años.

Tabla 2. Tarea típica de las investigaciones sobre asociación

Se quiere estudiar si el fumar produce trastornos respiratorios a un grupo de personas. Para ello, se han observado durante un periodo suficiente de tiempo a 250 personas obteniendo los siguientes resultados:

	Molestias respiratorias	No tiene molestias	Total
Fuma	90	60	150
No fuma	60	40	100
Total	160	100	250

Utilizando los datos de la tabla, razona si en esta muestra, el padecer trastornos respiratorios depende o no de fumar.

Una primera estrategia incorrecta para resolver el problema es analizar solamente la relación entre los casos favorables positivos (casilla *a* en la tabla 1) en relación a los casos totales (valor *n* en la tabla 1). En nuestro ejemplo, estos alumnos deducirían incorrectamente la existencia de una asociación directa entre las variables, ya que el

número de fumadores con trastornos es superior a cualquiera de las otras tres categorías.

Una estrategia algo más avanzada, pero todavía incorrecta es comparar las casillas dos a dos. En la tabla 1, una vez admitido que también los casos d (ausencia-ausencia) son favorables a la existencia de asociación, usarían sólo los datos de una fila o una columna. Por ejemplo compararían los 90 fumadores con trastornos con los 60 sin ellos, deduciendo erróneamente la existencia de asociación.

Piager e Inhelder consideraron correcto calcular la relación entre los casos que confirman la asociación ($a+d$) y el resto de los casos ($b+c$), es decir, en este caso comparar 130 personas en la diagonal principal con 120 en la diagonal secundaria. Esta estrategia es de nuevo incorrecta como señalan Jenkins y Ward (1965).

La estrategia correcta general consiste en comparar las dos probabilidades condicionales de que ocurra A cuando B es cierta y de que ocurra A cuando B es falsa, es decir, en la tabla 2, sería necesario comparar las razones 90/160 con 60/100 (frecuencias condicionales), que son iguales en este problema y por tanto muestran la independencia.

Análisis semiótico de la tabla de doble entrada

En Díaz (2005) y Estrada y Díaz (2006) analizamos la tabla 1, mostrando su complejidad. Nos basamos para nuestro análisis en ideas tomadas de Godino (2003), para quien la actividad matemática es esencialmente relacional y en ella los objetos matemáticos se ponen en relación por medio de funciones semióticas. Utilizando esta teoría, podemos hacer un análisis más detallado de los objetos matemáticos implícitos al interpretar una tabla doble y la forma en que son relacionados en el cálculo de probabilidades, así como de los principales conflictos semióticos de los futuros profesores en la tarea.

Las frecuencias absolutas dobles a , b , c , d son matemáticamente equivalentes en una tabla 2×2 , pero no psicológicamente. La celda a (presencia del carácter A; presencia del carácter B) y la d (ausencia de A; ausencia de B) serían las que indicarían (si su frecuencia es alta) una asociación positiva entre las variables; y las otras dos una asociación negativa. Además, de una celda dada (por ejemplo a) se pueden deducir tres frecuencias relativas diferentes:

- Frecuencia relativa doble: $\frac{a}{a+b+c+d}$
- Frecuencia relativa respecto a su fila: $\frac{a}{a+b}$
- Frecuencia relativa respecto a su columna: $\frac{a}{c+d}$

Más aún, podemos calcular las frecuencias relativas marginales de filas y columnas:

$$\frac{a+b}{a+b+c+d} \text{ y } \frac{a+c}{a+b+c+d}$$

Supuestos equiprobables todos los casos en la muestra y al preguntarnos por la probabilidad de obtener un sujeto al azar de la misma, podemos obtener también una probabilidad diferente de cada una de las frecuencias relativas anteriores:

- Probabilidad de que ocurran simultáneamente A y B :

$$P(A \cap B) = \frac{a}{a+b+c+d}$$

- Probabilidad condicional de A , sabiendo que ha ocurrido B :

$$P(A/B) = \frac{a}{a+b}$$

- Probabilidad condicional de B , sabiendo que ha ocurrido A :

$$P(B/A) = \frac{a}{c+d}$$

- Probabilidades simples de A y B :

$$P(A) = \frac{a+b}{a+b+c+d} \text{ y } P(B) = \frac{a+c}{a+b+c+d}$$

Todos estos objetos matemáticos co-existen y pueden ser confundidos por el alumno al realizar la interpretación de los datos de una tabla, bien en el estudio de la asociación entre las variables A y B o al calcular probabilidades a partir de ellas (Batanero y cols., 1996). Hemos de tener también en cuenta que algunos estudiantes no discriminan adecuadamente entre las dos direcciones de la probabilidad condicional $P(A/B)$ y $P(B/A)$. Falk (1986) denomina a este error *falacia de la condicional transpuesta*.

Pollatsek, Well, Konold y Hardiman (1987) indican que la ejecución de tareas que implican probabilidades condicionales depende de cómo se redacten los enunciados. Por ejemplo, Einhorn y Hogarth (1986) observaron que algunos alumnos interpretan incorrectamente la conjunción *y*, confundiendo la probabilidad conjunta y la probabilidad condicional. También en la investigación de Ojeda (1995), la mitad de los sujetos del estudio interpretaron la intersección como condicionamiento. Finalmente Huerta y Lonjedo (2005) muestran la influencia del formato en que se presentan los datos sobre la solución de los problemas simples de probabilidad condicional.

3. Diseño del estudio

Con objeto de analizar los posibles conflictos semióticos de los estudiantes al interpretar los datos de una tabla doble y calcular a partir de ella probabilidades simples, compuestas y condicionales, analizaremos los resultados del ítem 1 de una muestra de 117 estudiantes de diferentes especialidades de la diplomatura de Magisterio. La tarea es parte de un cuestionario global de comprensión de la probabilidad condicional (Díaz, 2004) y se pasó a los estudiantes en una de las horas de clase de la asignatura de Matemáticas y su didáctica, antes de iniciar el bloque temático sobre Probabilidad y Estadística.

Ítem 1. En una población se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	más de 55 años	Total
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ningún ataque	401	275	676
Total	430	350	780

Si elegimos al azar a una de estas personas:

- ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
 - ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido un ataque al corazón?
 - Sabiendo que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?
 - Sabiendo que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?
-

Se intentó motivar su colaboración explicándoles el objeto de la investigación y los resultados fueron usados posteriormente para discutir con los estudiantes sus lagunas, sus interpretaciones incorrectas y motivarlos en profundizar en el estudio de la probabilidad.

4. Análisis de resultados

Recogidas las respuestas, se procedió a categorizarlas. En la Tabla 3 presentamos los resultados de las respuestas correctas y sus porcentajes con las abreviaturas siguientes: A ="Ha sufrido un ataque al corazón"; no A ="No ha sufrido ataques al corazón"; B ="Más de 55 años", no B ="Menos de 55 años".

Podemos observar que, aunque la mayoría responde correctamente al cálculo de la probabilidad simple, hay alrededor de un 40% de errores en las tareas de cálculo de probabilidades compuestas y condicionales. Los porcentajes de respuestas correctas son menores que los de la investigación de Díaz (2005) con estudiantes de Psicología en el cálculo de la probabilidad simple y similares en los otros apartados.

Llama poderosamente la atención el alto porcentaje de respuestas en blanco en las diferentes tareas propuestas, bastante mayor que el obtenido en Díaz (2005). Al averiguar por qué tantos estudiantes no son capaces de responder; ellos lo justifican con respuestas tales como "No me acuerdo", "No entendía lo que me pedían", "No me lo han explicado nunca", "Hace mucho tiempo que lo estudié", "Lo sabía, pero lo he olvidado" y, por consiguiente, nos indican la necesidad de una mejor formación de los futuros profesores en el terreno de la Probabilidad.

Tabla 3. Frecuencias (y porcentajes) de respuesta en las cuatro tareas

Respuesta	Tarea pedida			
	$P(A)$	$P(A \cap B)$	$P(A/B)$	$P(B/A)$
$P(A)$	87 (74.4)		1	1
$P(B)$	1			
$P(A \cap B)$	1	62 (52.9)		1
$P(A/B)$		8	67 (57.2)	11
$P(B/A)$		15	7	66 (56.4)
$P(\text{no } A \cap B)$		2	2	3
Supone independencia		2		
$P(A/\text{no } B)$	1		1	
$P(\text{no } A/\text{no } B)$			1	
$P(A \cup B)$		2	1	1
Calcula casos posibles		1	2	
Otros errores	1	1	2	
Blanco	26 (22.0)	24 (20.5)	33 (28.3)	36 (30.8)

Entre los errores cometidos destacamos la confusión entre probabilidades condicionales y conjuntas, y podemos observar que se repiten en nuestros estudiantes al igual que en los estudios previos, la confusión entre los diferentes tipos de probabilidades: la probabilidad simple con la condicional. Pero, al mismo tiempo, y coincidiendo con Díaz (2005), se producen otros errores no descritos en las investigaciones anteriores, consistentes en la confusión de un suceso y su complementario, la confusión de probabilidades con casos posibles (frecuencias absolutas), la confusión de la unión y la intersección, e incluso la suposición en los datos independencia, aunque la dependencia quede patente en la tabla.

Tabla 4. Tipos de conflictos semióticos identificados en la lectura de las tablas

	N
Confunde una probabilidad condicional y su inversa	18
Confunde probabilidad condicional y conjunta	26
Confunden un suceso y su complementario	10
Supone independencia	2
Confunde probabilidad simple con condicional o conjunta	4
Confunde unión e intersección	2
Calcula casos posibles (f. absolutas)	3
Otros errores	4

En la Tabla 4 hemos clasificado estos errores según los diferentes conflictos semióticos entre objetos matemáticos involucrados en la tarea.

Las frecuencias de dicha tabla corresponden a la suma de los conflictos identificados en el total de las cuatro tareas propuestas a los profesores en formación, teniendo también en cuenta que, a veces, una respuesta conlleva más de un conflicto semiótico. Por ejemplo, cuando al pedirle calcular la probabilidad $P(B/A)$, el alumno da como respuesta $P(\text{no } A \cap B)$, un primer conflicto radica en confundir el condicionamiento con intersección, y un segundo en confundir un suceso con su complementario.

Observamos que los errores consistentes en confundir una probabilidad condicional con una conjunta o confundir una probabilidad condicional con su inversa, citados en las investigaciones anteriores, son los más frecuentes. Hay un cierto número de estudiantes que confunden un suceso con su complementario y con menor frecuencia se presentan otros conflictos referidos a la confusión de frecuencia y probabilidad.

5. Conclusiones

Los resultados complementan y reafirman los estudios previos (Díaz, 2005; Estrada y Díaz, 2006) y nos indican que la lectura de tablas dobles no es una tarea trivial para los futuros profesores, al menos en el contexto de cálculo de probabilidades. Prueba de ello es el alto porcentaje de respuestas en blanco justificadas según los estudiantes por un desconocimiento de la materia o una falta de comprensión de lo que se les pedía, así como el número de errores en aquellos que completan la tarea. También se repiten en nuestros alumnos la confusión entre los diferentes tipos de probabilidades.

Dada la frecuente aparición de este tipo de tablas en la prensa, en internet y en el material profesional del profesor, así como la necesidad de toma de decisiones a partir de ellas, consideramos que la enseñanza ha de prestar más atención al desarrollo de la capacidad de lectura correcta de tablas estadísticas en los futuros profesores. Además, el profesor requiere unos conocimientos básicos sobre probabilidad y sobre los errores frecuentes en los juicios probabilísticos, por lo que sería deseable que se conciencie de que él mismo los puede presentar y conozca los instrumentos de evaluación que les permita diagnosticarlos en sus alumnos. Por ello pensamos que los diferentes planes de estudios escolares y universitarios deberían incluir unos programas más completos de Estadística y su correspondiente base probabilística.

Además, puesto que una enseñanza enfocada sólo a la técnica es difícil que le permita al profesor superar estos errores sería importante pensar en situaciones didácticas que permitiesen superar la brecha abierta entre el conocimiento matemático y el conocimiento profesional (Ball, 2000). Dado que el conocimiento se construye en forma activa, debiéramos presentar a los profesores problemas y actividades de tipo constructivistas en las que interactuasen con sus compañeros y se enfrentasen a sus concepciones erróneas (Jaworski, 2001). La tarea que hemos presentado en este trabajo podría servir de base para construir una situación de este tipo en la enseñanza de la probabilidad a los futuros profesores.

6. Referencias

- Azcárate, P. (1995). *El conocimiento profesional de los profesores sobre las nociones de aleatoriedad y probabilidad. Su estudio en el caso de la educación primaria*. Tesis doctoral. Universidad de Granada.
- Ball, D. L. (2000). [Bridging practices: Intertwining content and pedagogy in teaching and learning to teach](#). *Journal of Teacher Education*, 51, 241-247.
- Batanero, C., Estepa, A., Godino, J. D. y Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151-169.
- Cardeñoso, J. M. (1998). *Las creencias y conocimientos de los profesores de primaria andaluces sobre la matemática escolar. Modelización de las concepciones sobre aleatoriedad y probabilidad*. Tesis doctoral. Universidad de Cádiz.
- Díaz, C. (2004). *Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar*. Trabajo de Investigación Tutelada. Universidad de Granada.
- Díaz, C. (2005) Conflictos semióticos en el cálculo de probabilidades a partir de tablas de doble entrada. *Biaix*, 24, 84-91.
- Díaz, C. y De la Fuente, I. (2005). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Epsilon*, 59, 245-260.
- Einhorn, H. J. y Hogart, R. M. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3-19.
- 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. Departamento de Didáctica de la matemática. Universidad de Granada.
- Estrada, A. (2005). A structural study of future teachers' attitudes towards statistics. Comunicación presentada en *CERME IV Conference (European Research in Mathematics Education Conference)*. Sant Feliu de Guíxols, Gerona, España.
- Estrada, A. y Díaz, C. (2006). Computing probabilities from two way tables. An exploratory study with future teachers. En A. Rossman y B. Chance (eds.), *Proceedings of ICOTS-7*. Salvador (Bahia): 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.
- Gal, I. (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- Gattuso, L. y Pannone, M. A. (2002). Teacher's training in a statistics teaching experiment. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*. Ciudad del Cabo: IASE. CD-ROM.

- Godino, J. D. (2003). *Teoría de las funciones semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática* (Theory of semiotic functions. An ontological and semiotic approach to mathematics cognition). Departamento de Didáctica de la Matemática. Universidad de Granada. Disponible en: http://www.ugr.es/local/jgodino/indice_tfs.htm.
- Huerta, P. y Lonjedo, M. A. (2005). The nature of the quantities in a conditional probability problem. Its influence in the problem resolution. Comunicación presentada en *CERME IV Conference (European Research in Mathematics Education)*. Sant Feliu de Guíxols, Gerona, España.
- Inhelder, B. y Piaget, J. (1955). *De la logique de l'enfant a la logique de l'adolescent* Paris: P.U.F.
- Jaworski, B. (2001). Developing mathematics teaching: teachers, teacher educators and researchers as co-learners. En L. Lin & T. J. Cooney (eds.), *Making sense of mathematics teacher education* (pp. 295-320). Dordrecht: Kluwer.
- Jenkins, H. M. y Ward, W. C. (1965). Judgement of contingency between responses and outcomes. *Psychological Monographs*, 79.
- Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO*, 5, 37-55.
- Pollatsek, A., Well, A. D., Konold, C. y Hardiman, P. (1987). Understanding Conditional Probabilities. *Organization, Behavior and Human Decision Processes*. 40, 255-269.
- Serrano, L. (1996). *Significados institucionales y personales de conceptos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad*. Tesis doctoral. Universidad de Granada.
- Silva, C. B. da, Cazorla, I. M. y Brito, M. R. F. (1999). Concepções e attitudes em relação a estatística. *Conferência Internacional: Experiências e Perspectivas do Ensino da Estatística* (pp. 18-29). Florianópolis.