

Conflictes semiòtics en el càlcul de probabilitats a partir de taules de doble entrada¹

Carmen Díaz Batanero, Inmaculada de la Fuente

Dpt. de Metodologia de les Ciències del Comportament, Universitat de Granada,
mcdiaz@ugr.es, edfuente@ugr.es

Resum

En aquest article presentem un estudi exploratori sobre la capacitat d'interpretació de dades en una taula de doble entrada i de càlcul de probabilitats simples, compostes i condicionals a partir d'aquestes dades. Anàlitzem conflictes semiòtics d'una mostra de 154 estudiants de Psicologia en l'execució de les tasques i concloem que convé dedicar temps a la lectura i interpretació d'aquestes taules en l'ensenyament.

Introducció

Avui dia es defensa la necessitat de la “cultura estadística”, com a capacitat per a interpretar i avaluar críticament la informació estadística i prendre decisions fonamentades sobre els fenòmens estocàstics presents en diversos contextos personals i professionals (Gal, 2002). Els articles en la premsa, en revistes científiques o de divulgació i els llibres de text o de consulta en moltes disciplines inclouen taules i gràfics estadístics. A més, molts servidors d'Internet, com el de l'Institut Nacional de Estadística (www.ine.es) distribueixen des de les bases de dades informació estadística sobre temes molt variats, que l'usuari pot descarregar seleccionant les variables.

La possibilitat d'aconseguir dades amb tanta facilitat és molt important per als estudiants d'educació secundària i universitat en l'elaboració dels treballs de les diferents assignatures. També diversos professionals (metges, economistes, inversors, comerciants, etc.) poden necessitar dades d'aquest tipus. Però, per a fer un ús eficient d'aquesta informació, cal llegir i

interpretar correctament les taules estadístiques. La lectura no sempre és senzilla i intuïtiva, en contra del que sovint se suposa.

En aquest treball ens centrem específicament en la taula estadística de doble entrada, amb la finalitat de mostrar que, malgrat l'aparent simplicitat, és un objecte semiòtic complex. Anàlitzarem les respostes d'una mostra de 154 estudiants de Psicologia a preguntes elementals sobre càlcul de probabilitats simples, compostes i condicionals a partir d'una taula de doble entrada. Aquests tres tipus de probabilitats són fonamentals en les aplicacions de l'estadística, en l'estudi de la inferència estadística, i en el de la correlació i regressió entre variables. D'altra banda, en el terreny professional i a la vida quotidiana, la presa de decisions encertades en situacions d'incertesa es basa en gran mesura en el càlcul de probabilitats, especialment condicionals, que són difícils d'avaluar i comprendre (Díaz i de la Fuente, 2005).

Per a l'anàlisi, ens basem en idees de Godino (2003), per a qui l'activitat matemàtica és essencialment relacional, ja que els objectes matemàtics es relacionen per mitjà de funcions semiòtiques. Aquest autor ens recorda que en el treball matemàtic usem uns objectes en representació d'altres, en especial dels objectes abstractes, fent una correspondència, sovint implícita, entre l'objecte representant i el representat. Per exemple, la paraula “probabilitat” fa referència a un objecte matemàtic abstracte (el concepte de probabilitat, que compleix una sèrie d'axiomes). De vegades, però, els significats donats pel professor i l'estudiant a una mateixa expressió matemàtica no coincideixen i es pro-

¹ Traducció del castellà feta per Núria Planas.

dueix un *conflicte semiòtic*. El professor parla de la probabilitat, que no pot ser major que la unitat, i un estudiant pot interpretar el terme probabilitat com equivalent a “possibilitats”, és a dir, com el nombre de casos favorables d’un esdeveniment i , per tant, acceptar valors més grans que la unitat.

D’acord amb la teoria esmentada, que hem exposat molt breument, podem fer una anàlisi detallada dels objectes matemàtics que apareixen a la taula de doble entrada i de la forma en què es relacionen en les tasques de lectura i càlcul de probabilitats. En analitzar de la mateixa forma les respostes dels estudiants, podem identificar conflictes semiòtics o desajustaments entre els significats donats a una expressió per l’alumne i pel professor. A continuació descrivint investigacions sobre interpretació de taules de doble entrada i presentem l’anàlisi duta a terme.

Investigacions sobre interpretació de taules de doble entrada

Una taula de doble entrada o de contingència serveix per a presentar resumidament la distribució de freqüències d’una població o mostra, classificada respecte a dues variables estadístiques. La Taula 1 presenta la forma més simple, quan les variables només tenen dues categories. En aquest treball ens restringim a aquesta forma, tot i que les dades estadístiques acostumen a aparèixer en taules amb més de dues fileres o columnes. Algunes de les fileres o columnes de les taules estadístiques que trobem a la premsa o Internet són totals parcials de diverses fileres o columnes anteriors, de manera que els conflictes semiòtics que identificarem seran possiblement més seriosos quan es demani als estudiants interpretar una taula més complexa.

Taula 1. *Format típic de la taula de doble entrada 2x2*

	<i>A</i>	<i>no A</i>	Total
B	<i>a</i>	<i>b</i>	<i>a + b</i>
no B	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>n = a + b + c + d</i>

La recerca didàctica sobre les taules de doble entrada s’ha centrat en l’anàlisi de la capacitat dels estudiants d’identificar l’associació o relació estadística entre les variables A i B a partir de les dades presentades seguint el format de la Taula 1. La Figura 1 mostra una tasca típica (Estepa, 1993) en aquestes investigacions.

Figura 1. *Tasca típica de les recerques sobre associació*

Ítem 1. Es vol estudiar si una medecina produeix trastorns digestius en els ancians. S’han observat 25 ancians durant un període suficient de temps obtenint-se els següents resultats:

	Molèsties digestives	No té molèsties	Total
Pren la medicina	9	3	17
No la pren	7	1	8
	16	9	25

Usant les dades de la taula, raona si, en aquests ancians, patir trastorns digestius depèn de la medecina.

Inhelder i Piaget (1955) van iniciar l’estudi del raonament sobre associació estadística. Aquests autors van considerar que la comprensió de la idea d’associació implica la comprensió prèvia de proporció i probabilitat. Per aquest motiu, només van estudiar aquest tipus de raonament amb nois a partir dels 13-14 anys, quan, segons la seva teoria, es troben a l’etapa de les operacions formals. Inhelder i Piaget indiquen que, en preguntar si les variables A i B de la Taula 1 estan relacionades, els més joves només analitzen la relació entre casos on es donen alhora els valors A i B (casella a , Taula 1) respecte als casos totals (valor n , Taula 1). A l’ítem 1, aquests alumnes no deduirien bé l’existència d’una associació directa entre les variables ja que el nombre d’ancians amb trastorns digestius que prenen la medicina és superior a les altres tres categories.

Una estratègia més avançada, però encara incorrecta, és comparar les cel·les de la Taula 1 dues

a dues. Un cop admès que també els casos a la cel·la (d) (no apareixen ni A ni B) són favorables a l'existència d'associació, aquests subjectes usarien només les dades d'una filera o d'una columna, sense calcular la relació entre el total de casos que confirmen l'associació ($a+d$) i la resta dels casos ($b+c$), que segons Piaget i Inhelder seria una estratègia correcta.

Hi ha estratègies similars en treballs amb estudiants adults (e.g., Smedslund, 1963 i Estepa, 1993), que, majoritàriament, basen el seu judici en l'existència d'associació entre A i B, bé a la casella (a) o comparant (a) amb (b). La dificultat de la tasca es posa de manifest en considerar que, com assenyalen Jenkins i Ward (1965), l'estratègia de comparació de diagonals, correcta segons Piaget i Inhelder per a resoldre aquests problemes, només és vàlida per a taules amb igual nombre de casos per filera (freqüències marginals). L'estratègia correcta consisteix a comparar les probabilitats condicionals de què ocorri A quan B és certa i de què ocorri A quan B és falsa, és a dir, a l'ítem 1, caldria comparar les raons $9/17$ i $7/8$ (freqüències condicionals).

Anàlisi semiòtica de la tasca

A les recerques de la secció anterior no es qüestiona la possible dificultat dels alumnes en llegir les dades de la Taula 1, que, en realitat és un objecte semiòtic complex. Tot i que les dades de les cel·les a , b , c , d es refereixen a freqüències absolutes, cadascuna d'una doble condició (valors de filera i columna), el significat no és equivalent. Una freqüència alta a les cel·les a (presència del caràcter A; presència del caràcter B) i la d (absència d'A, absència de B) indicaria una associació positiva entre variables, mentre que una freqüència alta a les altres dues cel·les suggeriria una associació negativa. Això, però, no és fàcilment percebut pels estudiants (Estepa, 1993). A més, en una cel·la donada (per exemple a) es poden deduir tres freqüències relatives diferents:

- Freqüència relativa doble: $\frac{a}{a+b+c+d}$
- Freqüència relativa respecte a la filera: $\frac{a}{a+b}$
- Freqüència relativa respecte a la columna: $\frac{a}{a+c}$

I també es poden calcular les freqüències relatives marginals de fileres i columnes:

$$\frac{a+b}{a+b+c+d} \quad \text{i} \quad \frac{a+c}{a+b+c+d}$$

Tots aquests objectes matemàtics coexisteixen i l'alumne els pot confondre en realitzar la interpretació de les dades d'una taula i estudiar l'associació entre les variables A i B. Per exemple, a la re-cerca de Batanero i altres (1996), sobre un 20% dels estudiants del curs preuniversitari confon les freqüències condicionals respecte a filera i columna.

És possible, per tant, que aquesta confusió aparegui en el càlcul de probabilitats a partir de la taula, on de cadascuna de les freqüències relatives anteriors es dedueix una probabilitat, en un context d'extracció a l'atzar d'un element de la població o mostra. Suposats equiprobables tots els subjectes, i aplicant la regla de Laplace com quocient entre casos favorables i possibles, podem obtenir una probabilitat diferent de cadascuna de les freqüències relatives anteriors:

- Probabilitat de què passin alhora A i B,
 $(A \cap B) = \frac{a}{a+b+c+d}$
- Probabilitat condicional d'A, sabent que ha passat B; $P(A/B) = \frac{a}{a+b}$
- Probabilitat condicional de B, sabent que ha passat A; $P(B/A) = \frac{a}{a+c}$
- Probabilitats simples d'A i B $P(A) = \frac{a+b}{a+b+c+d}$
i $P(B) = \frac{a+c}{a+b+c+d}$

Segons Falk (1986) molts estudiants no discriminen adequadament entre les dues direccions de la probabilitat condicional $P(A/B)$ i $P(B/A)$, denominant l'error *fal·làcia de la condicional transposta*. Aquest error s'ha observat també en el context d'emetre o interpretar un diagnòstic mèdic, ja que es confon la probabilitat de tenir una malaltia quan ha estat positiva una prova mèdica, que a vegades no és molt alta, amb la probabilitat d'obtenir un resultat positiu a la prova mèdica, si es té la malaltia, que és molt propera a la unitat (Eddy, 1982).

La facilitat de resolució de tasques que impliquen probabilitats condicionals depèn, segons Pollatsek

i altres (1987) de la redacció dels enuncisats. Per exemple, Einhorn i Hogarth (1986) van observar que alguns alumnes interpreten incorrectament la conjunció “i”, confonent probabilitat conjunta i condicional. La meitat dels subjectes a Ojeda (1995) van interpretar la intersecció com a condició. Recentment la influència del format en què es presenten les dades sobre la solució dels problemes simples de probabilitat condicional s’ha comprovat a Huerta i Lonjedo (2005).

Mètode i resultats

Tot seguit, analitzarem les respostes donades a cadascun dels quatre apartats de l’ítem 2 en una mostra de 154 estudiants de primer curs de Psicologia. La tasca és part d’un qüestionari més ampli que avalua la comprensió de la probabilitat condicional (Díaz, 2004) i que es va passar als estudiants després d’haver-se examinat de la part de l’assignatura que contenia el tema de probabilitat condicional.

Ítem 2. En una població s’ha realitzat una entrevista a un grup de persones, obtenint-se els següents resultats:

	55 anys o més jove	Més de 55 anys	Total
Ha sofert un atac de cor	29	75	104
No ha sofert atac	401	275	676
Total	430	350	780

Si escollim a l’atzar una d’aquestes persones:

- b) Quina és la probabilitat de què hagi tingut un atac de cor?
- b) Quina és la probabilitat de què tingui més de 55 anys i a més hagi tingut un atac de cor?
- c) Sabent que la persona escollida té més de 55 anys, quina és la probabilitat de què hagi tingut un atac de cor?
- d) Sabent que la persona escollida ha tingut un atac de cor, quina és probabilitat de què tingui més de 55 anys?

El qüestionari es va passar en una sessió de l’assignatura d’Anàlisi de dades d’aquests alumnes. Es va motivar la seva col·laboració en puntuar el nombre de respostes correctes com una part de la qualificació final de l’assignatura. Es va explicar als estudiants el propòsit de la investigació i participaren amb interès. Els resultats es van usar posteriorment per tal de discutir amb els estudiants les seves interpretacions incorrectes i motivar-los a aprofundir en l’estudi de la probabilitat.

Recollides les respostes, es varen categoritzar. Usant la nomenclatura de la Taula 1, la notació a les Taules 3, 4 i 5 és: A=“Ha sofert un atac de cor”; no A=“No ha sofert atac de cor”; B=“Més de 55 anys”, no B=“55 anys o més jove”. A la Taula 3 presentem els percentatges de respostes correctes per a les quatre tasques de l’ítem 2. Observem que, tot i que molts estudiants responen bé al càlcul de la probabilitat simple, hi ha més d’un 40% d’errors a les tasques de càlcul de probabilitats compostes i condicionals.

Taula 3. Percentatges de respostes correctes ($n=154$)

Tasca	Percentatge
a) Càlcul de $P(A)$	90
b) Càlcul de $(A \cap B)$	44
c) Càlcul de $P(A/B)$	53
d) Càlcul de $P(B/A)$	56

A la Taula 4 classifiquem les respostes. Els únics errors en el càlcul de la probabilitat simple han estat la confusió amb la probabilitat conjunta (respondre 75/780) o amb una condicional (75/350), a més d’alguns errors de càlcul o de donar casos possibles (104), error que suposa acceptar que la probabilitat és major que la unitat. A les altres tasques hi ha més varietat d’errors, amb la confusió de probabilitats condicionals i conjuntes ja destacada per Einhorn i Hogarth (1986) i Ojeda (1995).

Taula 4. Freqüències de respostes a les quatre tasques

Resposta de l'estudiant (n=154)	Tasca demanada			
	$P(A)$	$P(A \cap B)$	$P(A/B)$	$P(B/A)$
$P(A)$	138*	2	4	1
$P(A \cap B)$	2	67*	3	3
$P(A/B)$	1	19	81*	25
$P(B/A)$		36	20	86*
$P(\text{no } A \cap B)$		2	7	9
Suposa independència		3		
$P(A/\text{no } B)$	1		1	
$P(\text{no } A/\text{no } B)$		1	1	
$P(\text{no } B/A)$				1
$P(A \cup B)$		3	1	
Probabilitat major que 1		1	1	1
Calcula passos possibles	3	4	5	3
$P(A)/P(B)$			1	4
Error de càlcul o no identifica dades	4	11	12	3
Blanc	5	5	17	18

* Resposta correcta

Però, alhora, es produeixen errors no descrits a les investigacions anteriors:

- Confusió d'un esdeveniment i el seu complementari, responent, per exemple, donar la solució $275/676$ ($P(B/\text{no } A)$), quan es pregunta $P(B/A)$.
- Confusió de probabilitats amb casos possibles (freqüències absolutes), donant, per exemple, el valor 75 per a $P(A \cap B)$.
- Obtenció de probabilitats majors que la unitat; com respondre $P(A \cap B) = 780/75$, invertint en aquest cas la fórmula de Laplace.
- Confusió de la unió i la intersecció, donant la solució $P(A \cap B) = \frac{104}{780} + \frac{350}{780} - \frac{75}{780}$. Aquests alumnes (tres casos) han notat que els esdeveniments A i B no són mútuament excloents i han aplicat la fórmula correcta de la unió d'esdeveniments,

però han donat a la paraula "i" de l'enunciat un significat d'unió, habitual en el llenguatge quotidià, però diferent en aquest cas del significat matemàtic.

- Supòsit d'independència en les dades, tot i la dependència patent a la taula. Aquests alumnes calculen la probabilitat de la intersecció, com a producte de les probabilitats $P(A)$ i $P(B)$, calculades correctament, donant la solució

$$P(A \cap B) = \frac{104}{780} \times \frac{350}{780}$$

- Quocient $P(B)/P(A)$ quan es demana probabilitat condicional (respondre $350/104$); l'alumne associa la probabilitat condicional a un quocient, però hi ha conflicte en la confusió del numerador d'aquest quocient.

A la Taula 5 hem classificat aquests errors pels diferents conflictes semiòtics entre objectes matemàtics involucrats en les tasques demanades.

Taula 5. Freqüència de tipus de conflictes semiòtics identificats

	N=154
Confon probabilitat condicional i conjunta	77
Confon una probabilitat condicional i la seva inversa	45
Confon un esdeveniment i el seu complementari	23
Probabilitat major que 1	18
Calcula casos possibles	15
Confon probabilitat simple amb condicional o conjunta	11
Confon unió i intersecció	4
Suposa independència	3
Confon la fórmula de la probabilitat condicional amb $P(A)/P(B)$	5
Error de càlcul o no identifica dades	30
No respon	45
Total conflictes a les quatre tasques i total de mostra	262

Les freqüències de la taula corresponen a la suma de conflictes identificats en el total de les quatre tasques, tenint en compte que, a vegades, una resposta comporta més d'un conflicte semiòtic. Quan en el càlcul de la probabilitat condicional $P(A/B)$ l'alumne respon la probabilitat conjunta $P(\text{no } A \cap B)$, un primer conflicte és confondre condició i intersecció, i un segon és confondre un esdeveniment amb el seu complementari. En dividir el total de conflictes entre el nombre d'estudiants, trobem un promig de 1,7 a les quatre tasques, de manera que els estudiants troben conflicte a la meitat de les tasques.

Els errors consistents a confondre probabilitat condicional amb conjunta o confondre probabilitat condicional i inversa, citats a les recerques anteriors, són els més freqüents, però no expliquen per ells mateixos els elevats percentatges (propers al 40%) d'errors en les tasques de càlcul de probabilitat condicional i composta. Hi ha un nombre apreciable d'estudiants que confonen un esdeveniment amb el seu complementari i que, donant bé la fórmula de càlcul, tenen errors de càlcul o no identifiquen les dades a l'enunciat. Són menys freqüents d'altres conflictes relatius a la confusió de freqüència i probabilitat o a la inversió de la fórmula de càlcul

de probabilitats, amb valors majors que 1. Finalment alguns estudiants no són capaços de donar una resposta.

Conclusions

L'anàlisi i els resultats anteriors indiquen que la lectura de taules estadístiques de doble entrada i el càlcul de probabilitats a partir d'elles no és una tasca trivial per als alumnes universitaris. Coincidim amb Falk (1986) en què el llenguatge ordinari, usat en l'enunciat de problemes de probabilitat condicional no té la suficient precisió i és, per tant, ambigu. Quan usem la notació matemàtica per tal de referir-nos a una probabilitat condicional és clar quin és l'esdeveniment condicionant i quin el condicionat, però en el llenguatge ordinari la probabilitat condicional i la inversa no sempre es distingeixen entre elles o respecte a la probabilitat conjunta. Donada la freqüència d'aparició d'aquestes taules a la premsa, Internet i en el material professional, i la importància d'una correcta interpretació per a la presa de decisions, l'ensenyament ha de parar més atenció al desenvolupament de la capacitat de lectura de taules estadístiques.

Referències

- 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.
- DÍAZ, C. (2004). *Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar*. Treball de recerca. Universitat de Granada.
- DÍAZ, C.; DE LA FUENTE, I. (2005). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Epsilon*, 59, 245-260.
- EDDY, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. A D. Kahneman, P. Slovic i Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. Nova York: Cambridge University Press.
- EINHORN, H. J.; 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*. Tesi doctoral. Universitat de Granada.
- FALK, R. (1986). Conditional Probabilities: insights and difficulties. A R. Davidson i J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. (292 - 297). Victòria, Canadà: International Statistical Institute.
- GAL, I (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- GODINO, J. D. (2003). *Teoría de las funciones semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática*. Dpt. de Didàctica de la Matemàtica. Universitat de Granada. www.ugr.es/local/jgodino/indice_tfs.htm.
- INHELDER, B.; PIAGET, J. (1955). *De la logique de l'enfant a la logique de l'adolescent* París: P.U.F.
- HUERTA, P.; LONJEDO, M. A. (2005). *The nature of the quantities in a conditional probability problem. Its influence in the problem resolution*. Treball presentat al IV CERME Conference. Sant Feliu de Guíxols, Girona.
- JENKINS, H. M.; WARD, W. C. (1965). Judgement of contingency between responses and outcomes. *Psychological Monographs*, 79.
- OJEDA, A. M. (1995). Dificultades del alumnao respecto a la probabilidad condicional. *UNO*, 5, 37-55.
- POLLATSEK, A.; WELL, A. D.; KONOLD, C.; HARDIMAN, P. (1987). Understanding Conditional Probabilities. *Organization, Behavior and Human Decision Processes*. 40, 255 - 269.
- SMEDSLUND, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165-173.

Nota: El treball és part del Projecte BSO2002-0334-7 i de la Beca FPU: AP2003-5130