

¿EN QUE DIRECCION DEBEN ANOTARSE LOS PORCENTAJES?

HANS ZEISEL: *Dígalo con números* (México, D.F., Fondo de Cultura Económica, 2ª edición, 1980, pp. 43-54. Traducción de Roberto Helier):

La tabulación comparativa, cuadro que relaciona dos o más factores entre sí, ofrece para el analista un problema peculiar, a saber, al elaborar un cuadro: ¿en qué dirección deben calcularse las cifras de tanto por ciento, vertical u horizontalmente?

LA REGLA DE CAUSA Y EFECTO

Hay una regla que puede servirnos de guía para solucionar este problema, pero antes hace falta una aclaración: en una tabulación comparativa, siempre que uno de los factores pueda ser considerado la causa y el otro (o los otros) el efecto, *los porcentajes deben computarse en dirección del factor causal siempre y cuando la muestra sea representativa de esa dirección.*

Tómese por ejemplo la tabulación comparativa del Cuadro III-1.

CUADRO III-1. *Lectura de libros según el nivel educativo* *

	Univer- sidad	Escuela Secundaria	Escuela Primaria	Total
Leído durante el mes anterior				
Por lo menos un libro	163	208	30	401
Ningún libro	234	624	304	1.162
Total	397	832	334	1.563

* Tomado del Gallup Political Index, marzo de 1966.

Como lo ilustran los Cuadros III-2 y III-3, los porcentajes pueden anotarse en cualquier dirección.

Matemáticamente hablando, estos tres cuadros son equivalentes, pero como en ellos los porcentajes se calculan en direcciones diferentes se ponen de relieve distribuciones diferentes y para que sean examinadas con facilidad, se ofrecen a la vista comparaciones también diferentes. Es muy

instructivo examinar detenidamente los Cuadros III-2 y III-3. Los porcentajes del Cuadro III-3 dan por resultado una información sencilla y poco sorprendente: la lectura de libros aumenta con la educación; el 41% de todos los informantes con educación universitaria ha leído por lo menos un libro durante el último mes, en contraste con el 9% de entre quienes no pasaron de la escuela primaria.

CUADRO III-2. *Educación y niveles varios de lectura de libros*

	Univer- sidad	Escuela Secundaria	Escuela Primaria	Total
Leído durante el mes anterior				
Por lo menos un libro	41%	52	7	100%
Ningún libro	20%	54	26	100%
Total	25%	53	22	100%
(Número de casos)	(401)	(1.162)	(1.563)	

CUADRO III-3. *Niveles de lectura de libros según la educación*

	Univer- sidad %	Escuela Secundaria %	Escuela Primaria %	Total %
Leído durante el mes anterior				
Por lo menos un libro .	41	25	9	26
Ningún libro	59	75	91	74
Total	100	100	100	100
(Número de casos) . .	(397)	(832)	(334)	(1.563)

El contenido del Cuadro III-2 es de clase diferente: el 7% de las personas que leyeron un libro durante el último mes tuvo solamente educación primaria, comparado con el 26% de quienes no habían leído libro alguno. Este no es un contenido muy interesante. Sin embargo, el cuadro plantea un problema más intrincado aún. Para apreciar el significado de las cifras de porcentaje, habría que saber cuál es la proporción de personas con educación primaria solamente, dentro de la población total. Claro está que si únicamente el 10% de la población adulta no pasó de la escuela primaria, la cifra significa una cosa. Pero como todo el 22% (334 de 1.563)

cae dentro de ese grupo (que, de paso, es el promedio nacional correcto) la cifra significa algo más. No obstante, esto demuestra que lo primero que debe uno conocer es la distribución dada en el Cuadro III-3: el efecto de la educación sobre la lectura de libros.

El problema puede enfrentarse de manera distinta. Como veremos más adelante, la mayoría de los análisis de encuestas está planteada para reproducir un resultado experimental hipotético.

En el Cuadro III-3 puede verse la respuesta a un experimento pensado para determinar en qué grado aumenta la afición por la lectura cuando la gente asiste a la universidad. Es indudable que un cuadro de encuesta retrospectiva difiere de muchas maneras de un experimento controlado¹, pero la forma del razonamiento es en esencia igual.

A VECES LOS PORCENTAJES PUEDEN ANOTARSE EN CUALQUIER DIRECCIÓN

A veces, por la ambigüedad de la situación, la regla de causa y efecto no será útil como guía.

Los factores de una tabulación comparativa no siempre encajan dentro de una relación directa de causa y efecto; a menudo su relación será indirecta. Además, la regla de causa y efecto para el cálculo de porcentajes puede prestarse a varias interpretaciones. La "causa" puede ser lo que el investigador domine. Por consiguiente, ocurrirá frecuentemente que cualquiera de los factores de una tabulación comparativa pueda tomarse como el causal, en el sentido de que controla la variación del otro factor.

CUADRO III-4. *Remedios para cuatro enfermedades seleccionadas*

Enfermedades	Alka-Seltzer	Anacín	Aspirina	Otros	Total
Neuralgia	107	47	198	24	376
Resfriado	98	41	401	30	570
Malestar gástrico	302	60	—	23	385
Dolores de cabeza	—	242	210	26	478
(Número de informes) ...	507	390	809	103	1.809

A este respecto, un buen ejemplo es el Cuadro III-4, que contiene las respuestas de 1.809 informantes a la pregunta "¿Qué remedio prefiere usted para la neuralgia, el resfriado, la acidez gástrica y los dolores de cabeza?".

¹ Véase el Capítulo VII.

En el Cuadro III-5 se considera que el malestar es la causa que mueve a seleccionar la droga respectiva y, de acuerdo con ello, los porcentajes se calculan en la dirección horizontal del cuadro III-4.

CUADRO III-5. *Remedios preferidos para cuatro enfermedades*

Remedios	Neural- gia %	Res- friado %	Malestar gástrico %	Dolores de cabeza %
Alka-Seltzer	28	17	78	—
Anacín	13	7	16	50
Aspirina	53	71	—	44
Otros	6	5	6	6
Total	100	100	100	100
(Número de informantes) ..	(376)	(570)	(385)	(478)

El Cuadro III-5 responde a la pregunta de cuál es la importancia relativa de estos remedios con respecto a los cuatro padecimientos. En el 53% de los casos de neuralgia la gente elige aspirina. Entre los que sufren de resfriados, la aspirina como remedio muestra tener una preferencia aun más marcada: el 71% la toma. Para la acidez gástrica el 78% prefiere Alka-Seltzer y para los dolores de cabeza las preferencias se distribuyen uniformemente entre el anacín y la aspirina.

Pero el Cuadro III-4 puede contemplarse desde otro punto de vista. Puede uno preguntarse hasta qué punto los afectados de estas cuatro indisposiciones se sienten atraídos por las promesas curativas de cada uno de los remedios. El Cuadro III-6 da las respuestas calculando los porcentajes en otra dirección.

CUADRO III-6. *Importancia de las enfermedades con respecto a los remedios*

	Alka- Seltzer %	Anacín %	Aspi- rina %	Otros %
Neuralgia	21	12	24	23
Resfriado	19	12	50	29
Malestar gástrico	60	15	—	22
Dolores de cabeza	—	62	26	26
Total	100	100	100	100
(Número de informes) ..	(507)	(390)	(809)	(103)

LA CONDICIÓN DE MUESTREO REPRESENTATIVO

A veces puede abandonarse la regla sobre la dirección en que deben calcularse los porcentajes debido a las limitaciones estadísticas de la muestra. Al considerar un problema específico sería aconsejable calcular los porcentajes en una cierta dirección, pero esto podría vedarlo el carácter de la muestra.

Considérese el ejemplo siguiente: durante dos días (jueves y sábado) de una semana dada se realizaron observaciones en la sección de calcetería de un almacén de Nueva York. Se registraron los precios de las medias que compraron las clientes así como las edades de éstas. Los resultados, en números absolutos, se exponen en el Cuadro III-7.

CUADRO III-7. *Precio de las medias según la edad de las clientes **

Edad	Escala de Precios			Total
	\$ 0.59 - \$ 0.99	\$ 1.00 - \$ 1.29	\$ 1.30 - \$ 1.79	
Hasta 34 años	265	12	130	407
De 35 a 49 años	240	140	208	588
De 50 años o más	32	110	25	167
Total	537	262	363	1.162

* Las observaciones se hicieron en jueves y sábado.

El departamento de ventas deseaba saber a cuáles de los varios grupos de edad se vendían mejor las medias de diferentes categorías de precios. El precio se consideró, pues, como causa (de atraer a varios grupos de edad), y en consecuencia los porcentajes se calcularon en dirección vertical:

CUADRO III-8. *Precio de las medias y edad de las clientes*

Edad	\$ 0.59 - \$ 0.99	\$ 1.00 - \$ 1.29	\$ 1.30 - \$ 1.79
	%	%	%
Hasta 34 años	49	5	36
De 35 a 49 años	45	53	57
De 50 años o más	6	42	7
Total	100	100	100
(Número de casos)	(537)	(262)	(363)

Basándose en el Cuadro III-8, el gerente de ventas dedujo que sólo el 6% de la calcetería de precio más bajo lo compraban mujeres de 50 años

para arriba, y que el 33% de las medias más caras lo compraban mujeres hasta de 34 años, y así sucesivamente. Olvidó, sin embargo, que estas observaciones de la clientela se habían hecho en jueves y sábado, dos días en los cuales la distribución de edades de los clientes no es, de ninguna manera, representativa de toda la clientela. En estos dos días, la edad del cliente promedio es menor que la usual y son precisamente las mujeres jóvenes que trabajan quienes compran en tales días: en jueves, porque en Nueva York tal día los almacenes permanecen abiertos hasta las nueve de la noche y en sábado porque es su día libre. Por tal razón, el Cuadro III-8 es engañoso. Por otra parte, el Cuadro III-9 es correcto si se supone que las mujeres jóvenes y las de más edad se comportan los jueves y los sábados de la misma manera, en lo que respecta a la compra de medias, que los restantes cuatro días de la semana. Análisis posteriores demostraron que tal era el caso.

CUADRO III-9. *Precio de las medias según la edad de las clientes*

Escala de precios	Hasta 34 años %	De 35 a 49 años %	De 50 años o más %
\$ 0.59 - \$ 0.99	65	41	19
\$ 1.00 - \$ 1.29	3	24	66
\$ 1.30 - \$ 1.79	32	35	15
Total	100	100	100
(Número de casos)	(407)	(508)	(167)

Las mujeres de edad avanzada compran, pues, en su mayoría, las medias de precio medio, el grupo de mujeres más jóvenes prefiere las medias más baratas, y el grupo de edad madura prefiere las más caras.

Si se conociera la distribución por edades de las clientes durante *toda* la semana, entonces podría elaborarse una versión más correcta del Cuadro III-8, tomando en cuenta los diversos grupos de edad, de modo que la distribución de ellos en el cuadro pudiera equipararse con la distribución en el almacén durante toda la semana.

En el siguiente ejemplo se presenta con más claridad la falta de representatividad como obstáculo para aplicar la regla de causa y efecto. Para estimar la fuerza relativa de dos partidos políticos antes de cierta elección se hizo en dos Estados una votación preliminar y, como parte del estudio, en cada uno de ellos se realizaron dos mil entrevistas. Los resultados, redondeados para exponer el problema en su forma más simple, se muestran en el Cuadro III-10.

CUADRO III-10. *Votos por partido en dos Estados*

Estado	Republicanos	Demócratas	Total
A	625	1.375	2.000
B	875	1.125	2.000
Total	1.500	2.500	4.000

En el Cuadro III-11 los porcentajes están calculados en la dirección horizontal del Cuadro III-10. Aquí se supone que la causa de la fuerza relativa de los diferentes partidos políticos es el Estado mismo, es decir, la particular estructura social y económica del Estado y de la actitud política de sus habitantes.

CUADRO III-11. *Importancia del partido para el Estado*

	Estado A %	Estado B %
Republicanos	31	44
Demócratas	69	56
Total	100	100
(Número de entrevistas)	(2.000)	(2.000)

El cuadro muestra que el Partido Demócrata es el más fuerte en ambos Estados, pero que su preponderancia es mayor en el Estado A.

Si calculamos los porcentajes en la dirección vertical del Cuadro III-10 obtenemos el Cuadro III-12.

Este cuadro parece indicar que el 42% de los votos a favor de los republicanos proviene del estado A, mientras que el 55% de los votos a

CUADRO III-12. *Importancia del Estado para el partido*

Estado	Republicanos %	Demócratas %
A	42	55
B	58	45
Total	100	100
(Número de entrevistas)	(1.500)	(2.000)

favor de los demócratas proviene del mismo Estado. Pero una observación sencilla revela que el Cuadro III-12 no puede leerse de esa manera.

Es cierto que el 42% de los votos *de la muestra* a favor del Partido Republicano procede del Estado A. Pero el 42%, y en ese sentido los cuatro porcentajes del Cuadro III-12, son el resultado de dos componentes: la fuerza relativa de los partidos dentro de cada Estado y la decisión arbitraria de realizar un número igual de entrevistas en cada uno de ellos a pesar de que en el Estado B hay aproximadamente dos veces más votantes que en el Estado A².

Si equilibramos adecuadamente las cifras del Cuadro III-12 duplicando el número de entrevistas en el Estado B, obteniendo 4.000 en este Estado y 2.000 en el Estado A, podremos elaborar el Cuadro III-13.

CUADRO III-13. *Importancia del Estado para el partido*

Estado	Republicanos		Demócratas		Total de votantes en cada estado
	Número	Porcentaje	Número	Porcentaje	
A	(625)	26	(1.375)	38	(2.000)
B	(1.750)	74	(2.250)	62	(4.000)
Total	(2.375)	100	(3.625)	100	(6.000)

Los porcentajes son absolutamente diferentes de los del Cuadro III-12 sencillamente porque hicimos que el número de entrevistas realizadas en cada Estado estuviese en relación al número de sus votantes. En realidad, el Cuadro III-13, a diferencia del Cuadro III-12 sí representa lo que indica su título.

LA COLUMNA DEL TOTAL

En trabajos de investigación se acostumbra incluir en cualquier cuadro una columna que totalice las varias divisiones. Tal columna puede ser la primera o la última. En términos matemáticos esta columna del total es sencillamente la suma de todas las subdivisiones, y como tal es un conjunto exacto de números. Pero, para el lector, la columna del total tiene un significado más general que suele originarse en el título que ostenta el cuadro. Sirva de ilustración el Cuadro III-14, sobre la asistencia a la escuela dominical de los niños norteamericanos.

² Como el Cuadro III-11 era el objetivo principal de la votación primaria fue correcto teóricamente hacer una muestra de un número igual de entrevistas en cada Estado.

CUADRO III-14. *Asistencia a la escuela dominical de niños norteamericanos de diferentes niveles socioeconómicos **

Asistencia	Nivel económico							Total %
	I %	II %	III %	IV %	V %	VI %	VII %	
Regular	49.1	45.5	43.3	52.3	37.3	36.0	28.4	43.3
Irregular	50.9	54.5	56.7	47.7	62.7	64	71.6	56.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
(Número de casos)	(275)	(246)	(568)	(461)	(467)	(150)	(162)	(2.329)

* Tomado de John E. Anderson y otros, *The Young Child in the Home*. (Conferencia de la Casa Blanca sobre la salud y protección de los niños, Nueva York, 1936). La escala económica fue elaborada por F. L. Goodenough y J. E. Anderson, *Experimental Child Study* (Nueva York, 1931. I = Profesional; II = Semiprofesional; III = Administrativo; IV = Agricultores; V = Semicualificados; VI = Escasamente calificados; VII = Jornaleros.

Guiándose por el título del cuadro, el lector supone que la columna del total es una muestra representativa de los niños norteamericanos y, por tanto, interpreta que el 43,3% de todos los niños norteamericanos asiste regularmente a la escuela dominical. Pero su interpretación será errónea, porque el número de entrevistas realizadas en cada nivel socioeconómico no corresponde a la proporción relativa de estos niveles en el país. El Cuadro III-15 indica el grado en que se desvía la distribución de la muestra con respecto a los datos de población reales.

CUADRO III-15. *Tamaño relativo de grupos socioeconómicos en los EE. UU.*

Nivel económico	Muestra empleada en el Cuadro III-14 %	Población de EE.UU. * %	Diferencia %
I	11.8	2.7	+ 9.1
II	10.6	5.2	+ 5.4
III	24.4	15.7	+ 8.7
IV	19.8	18.0	+ 1.8
V	20.1	29.2	- 9.1
VI	6.4	12.3	5.9
VII	6.9	16.9	- 10.0
Total	100.0	100.0	

* U. S. Census of Population, 1930.

Los niveles I, II y III de la clase alta están representados excesivamente en la muestra, y los niveles inferiores V, VI y VII, insuficientemente. Siendo mayor la asistencia a la escuela dominical en las clases superiores, como lo muestra el Cuadro III-14, sabemos que el 43,3% de la columna del total es una cifra demasiado alta. Si equilibramos los porcentajes de asistencia del Cuadro III-14, no conforme al tamaño de la muestra tal como está, sino como debiera estar de acuerdo con el Cuadro III-15, obtendremos, en lugar del 43,3%, el 40% como promedio de todos los niños norteamericanos que asiste regularmente a la escuela dominical.

De paso, pueden observarse dos cosas. En primer lugar, que la mala distribución de la muestra afecta a la corrección de la columna del total únicamente si las divisiones varían con respecto a la variable más importante. En nuestro caso, si los niveles socioeconómicos no mostrasen variaciones con respecto a la asistencia a la escuela dominical, la columna del total habría sido correcta a pesar de no ser representativa de todos los

niveles económicos. Obsérvese, en segundo lugar, que si la columna del total no es representativa del tamaño relativo de las subdivisiones, pueden y deben equilibrarse sus componentes como lo hicimos en el Cuadro III-15.

Hay indudablemente situaciones en que la columna del total es lo más importante; en otras ocasiones, carece de significado.

Cuando el único fin que persigue un estudio es descubrir la diferencia que hay entre ciertos grupos, digamos, entre la población universitaria y la no universitaria, como se vio en el Cuadro III-1, entonces la columna del total es de poco interés. Pero en una votación preliminar, por ejemplo, cuya finalidad es predecir el voto de la población electoral, entonces la columna del total puede ser decisiva.

RESUMEN

En la mayoría de las tabulaciones comparativas surge la interrogante relativa a la dirección en que deben calcularse las cifras de porcentaje. La regla general consiste en que deben calcularse en dirección del factor causal (la variable cuyo efecto se desea determinar), siempre y cuando la muestra sea representativa de esa dirección. Hay situaciones en que cualquiera de los factores puede considerarse como el causal. Afín a este problema está la propiedad de presentar una columna de total: se justifica únicamente cuando forma una muestra verdadera de la población, según lo indique el título de esa columna. Si no constituye una verdadera muestra, a menudo es posible hacer un ajuste para eliminar la distorsión, equilibrando adecuadamente los diferentes subgrupos.