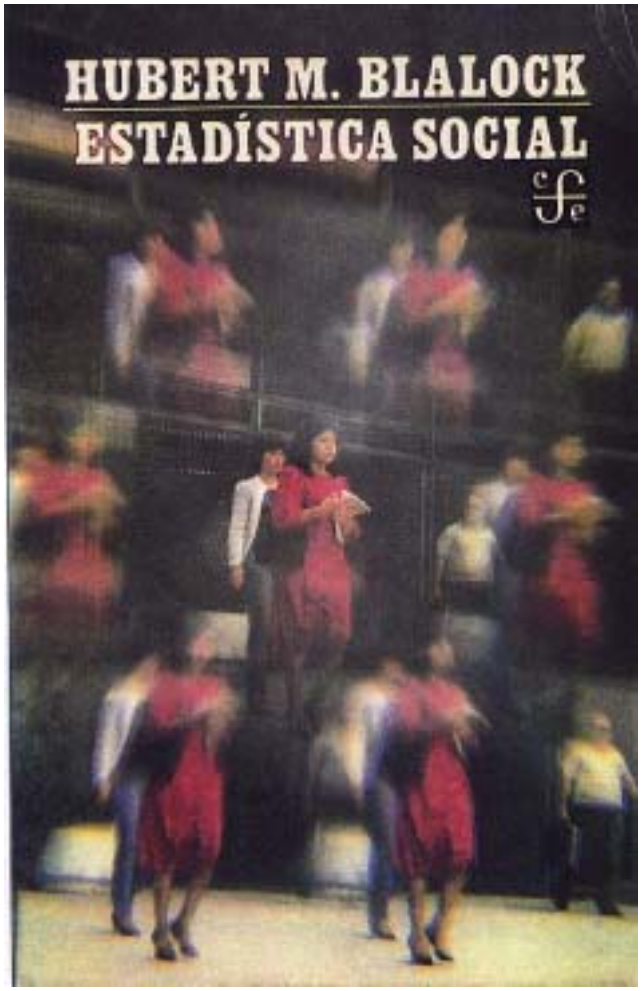
# ESTADÍSTICA SOCIAL

## HUBERT M. BLALOCK




FONDO DE CULTURA ECONÓMICA

México, 1986

# ÍNDICE

Tercera Parte
ESTADÍSTICA INDUCTIVA



Cuarta Parte
ESTADISTICAS BIVARIADAS Y MULTIVARIADAS

Quinta Parte
MUESTREO

APÉNDICES

# III. ESCALAS, NOMINALES: PROPORCIONES, PORCENTAJES Y RAZONES

Es mucho más sencillo resumir los datos que comportan escalas nominales que en el caso en que se emplean escalas de intervalo. La operación aritmética básica es, en el primer supuesto, la de contar el número de los casos al interior de cada categoría y de anotar sus tamaños relativos. Un grupo determinado puede constar de 36 varones y 24 mujeres, o de 25 protestantes, 20 católicos y 15 judíos. Sin embargo, para poder establecer comparaciones con otros grupos, hay que tener en cuenta el número de casos en cada uno de los grupos considerados. Las medidas que se examinan en el presente capítulo permiten establecer comparaciones entre diversos grupos, mediante normalización esencialmente en relación con el tamaño. Sin duda alguna, dos de las medidas en cuestión, la de las proporciones y la de los porcentajes, son ya conocidas de todos.

## III.l. Proporciones

Con objeto de poder servirnos de las proporciones, hemos de presumir que el método de clasificación ha sido tal que las categorías son mutuamente exclusivas y exhaustivas. En otros términos: cada individuo ha sido puesto en una categoría y en una sola. Con fines de simplificación, tomemos una escala nominal que conste de cuatro categorías, con $N_1$, $N_2$, $N_3$ y $N_4$ casos respectivamente. Supongamos que el número total de los casos sea N. La proporción de casos en cualquier categoría dada está definida como el número en la categoría dividido entre el número total de casos. Por lo tanto, la proporción de individuos de la primera categoría se halla dada por la cantidad $N_1/N$, y las proporciones de las demás categorías son respectivamente de $N_2/N$, $N_3/N$, $N_4/N$ . Es obvio que el valor de una proporción no puede ser mayor que la unidad. En efecto, como quiera que

$$N_1 + N_2 + N_3 + N_4 = \mathrm{N}$$

Tenemos que

$$\frac{N_1}{N} + \frac{N_2}{N} + \frac{N_3}{N} + \frac{N_4}{N} = \frac{N}{N} = 1$$

Así pues, si adicionamos las proporciones de los casos en todas las categorías (mutuamente exclusivas), el resultado es la unidad.

Es ésta una propiedad importante de las proporciones que se deja extender fácilmente a cualquier número de categorías.

Ilustremos el empleo de las proporciones con los datos dados en el cuadro III.l.

CUADRO III.1 *Número de delincuentes y de no delincuentes en dos localidades hipotéticas*

| Sujetos | Localidad 1 | Localidad 2 |
|---|---|---|
| Delincuentes | | |
| Primer delito | 58 | 68 |
| Reincidentes | 43 | 137 |
| No delincuentes | 481 | 1081 |
| Total | 582 | 1286 |

Resulta más bien difícil decir cuál de las dos localidades cuenta con mayor número de delincuentes, porque son diversamente grandes. En cambio, si expresamos los datos en términos de proporciones, podemos establecer una comparación directa. En efecto, la proporción de primeros delitos es, en la comunidad 1, de 58/582, o .100; la de la localidad 2, en cambio, es de 68/1286 o sea .053. Las demás proporciones pueden calcularse en forma análoga, resumiendo los resultados en forma de cuadro (cuadro III.2). El cuadro en cuestión nos permite apreciar que los números relativos de delincuentes son muy parecidos en las dos localidades, pero que la segunda de ellas contiene un número considerablemente más bajo de primeros delitos y una proporción más alta de reincidentes.

CUADRO III.2. *Proporciones de delincuentes y de no delincuentes en dos localidades hipotéticas*

| Sujetos | Localidad 1 | Localidad 2 |
|---|---|---|
| Delincuentes | | |
| Primer delito | .100 | .053 |
| Reincidentes | .074 | .107 |
| No delincuentes | .826 | .841 |
| Total | 1.000 | 1.000 |

La suma de las proporciones de la localidad 2 no exactamente la unidad, debido a los errores de redondeo. En ocasiones es conveniente presentar los datos de tal modo que las sumas sean exactamente igual a 1.000. Esto puede acaso exigir el ajuste de algunas proporciones de las categorías que comprenden el mayor número de casos. [1] El argumento a favor de este procedimiento está en que un cambio en la última cifra decimal de una proporción mayor es relativamente menos importante que el mismo cambio en una cifra menor. Así, por ejemplo, podría cambiarse la proporción de los no delincuentes de la localidad 2 en .840, de modo que la suma resultante sea igual a la unidad.

El cuadro 111.2 comprende proporciones del número total de casos en cada una de las comunidades. Supóngase, sin embargo, que el interés se centraba sobre todo en los delincuentes, y que deseábamos conocer la proporción de los reincidentes *entre los delincuentes*. El número total de delincuentes en las dos localidades es respectivamente de 101 y 205. Por lo tanto, entre los delincuentes, las proporciones de los reincidentes son respectivamente de 43/101, o .426 y 137/205, o .668. A primera vista estas cifras pueden proporcionar una impresión ligeramente diferente de la del primer conjunto de proporciones. Habríamos de guardarnos especialmente de concluir que el segundo espécimen es "más delictivo" que el primero. Por supuesto, este último conjunto de proporciones nada nos dice en absoluto acerca de las cifras relativas de no delincuentes en los dos especímenes considerados. Es obvio que no existe sustitutivo alguno de la lectura atenta de los cuadros. Constituye un buen principio acostumbrarse a determinar siempre las categorías que se hallan comprendidas en el número total de casos que sirve de denominador de la proporción. El lector debiera siempre preguntar: "*¿de qué* es esto la proporción?" Y la respuesta resultará clara del conjunto.

## III.2 Porcentajes

Los porcentajes pueden obtenerse de las proporciones multiplicando simplemente por 100. La palabra porcentaje significa *por ciento*. Por lo tanto, al servirnos de los porcentajes normalizamos en relación con el volumen, calculando el número de individuos que habría en una categoría determinada si el total de los casos fuera 100, permaneciendo inalterada la proporción en cada categoría. Y como quiera que las proporciones sumadas dan la unidad, es obvio que los porcentajes sumarán 100, a menos que las categorías no sean mutuamente exclusivas o exhaustivas.

Al reproducir resultados, los porcentajes se emplean con mucha mayor frecuencia que las proporciones. Las cifras del cuadro III. 2 habrían podido expresarse lo mismo en términos de porcentajes. Mejor que servirnos de los mismos datos, tomemos otro cuadro que puede servir para ilustrar otros diversos aspectos. Supongamos que tenemos tres agencias de servicios domésticos con una distribución de casos como la que se indica en el cuadro III.3.

Como es usual, los porcentajes se han dado hasta el primer decimal y se han operado los ajustes de los últimos dígitos, de modo que los totales den exactamente 100. Aquí el número de casos de cada agencia es lo suficientemente grande corno para justificar el empleo de porcentajes. Sin embargo, si el número de casos hubiera sido menor, el empleo de aquéllos habría resultado equívoco. En efecto, supóngase que la agencia C había tratado sólo 25 casos en total. Si hubiera habido cuatro madres solteras y siete parejas de novios, los porcentajes en dichas categorías habrían sido respectivamente del 16 y del 28 por ciento. Y como quiera que muchas personas acostumbran mirar sólo los porcentajes y no el número efectivo de casos comprendidos, podría fácilmente obtenerse la impresión de que había muchas más parejas de novios que de madres solteras. Como se verá cuando lleguemos a la estadística inductiva, la diferencia entre cuatro y siete casos puede deberse perfectamente a factores puramente casuales. El empleo de los porcentajes y las proporciones comporta por lo regular una estabilidad mucho mayor de las cifras. Por lo tanto, he aquí dos

[1] Puede utilizarse exactamente el mismo procedimiento en el caso de porcentaje.

reglas generales importantes: 1)*indíquese siempre el número de casos juntamente con los porcentajes o las proporciones, y* 2) *no se calcule nunca un porcentaje, a menos que el número de casos en que está basado se halle a proximidad de los 50 o más*. Si el número de casos es muy pequeño, será preferible indicar el número efectivo de ellos en cada categoría, sin recurrir a los porcentajes. En el caso anterior, por ejemplo, indicaríamos simplemente que la agencia C había tratado cuatro madres solteras y siete parejas de novios.

CUADRO III.3. *Distribución de los números y porcentajes de casos tratados por tres agencias hipotéticas de servicios domésticos*

| Clase de casos | Agencia A | | Agencia B | | Agencia C | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Nº | % | Nº | % | Nº | % | Nº | % |
| Matrimonios | 63 | 47.3 | 88 | 45.5 | 41 | 36.6 | 192 | 43.8 |
| Divorciados | 19 | 14.3 | 37 | 19.2 | 26 | 23.2 | 82 | 18.7 |
| Novios | 27 | 20.3 | 20 | 10.4 | 15 | 13.4 | 62 | 14.2 |
| Madres solteras | 13 | 9.8 | 32 | 16.6 | 21 | 18.8 | 66 | 15.1 |
| Otros | 11 | 8.3 | 16 | 8.3 | 9 | 8.0 | 36 | 8.2 |
| Total | 133 | 100.0 | 193 | 100.0 | 112 | 100.0 | 438 | 100.0 |

Véase ahora la columna del total que indica la distribución en porcentajes de las tres agencias juntas. Esas cifras se han obtenido sumando el número de casos de cada tipo y el número total de casos tratados por las tres agencias juntas. Para el cálculo de los porcentajes totales se utilizó, pues, como base un *N* de 438. Supóngase, sin embargo, que el número de casos no nos hubiera sido dado en el cuerpo del cuadro, sino que se hubiera presentado como en el cuadro III.4. En tal caso podría darse la tentación de obtener los porcentajes totales tomando directamente la media aritmética de los tres porcentajes de cada hilera. Semejante procedimiento no tendría en cuenta el hecho de que las tres agencias habían tratado números diferentes de casos; sólo se justificaría si los números de éstos fueran efectivamente iguales. El procedimiento correcto consistiría en ponderar cada porcentaje por el número correspondiente de casos. Uno de los medios para hacerlo consistiría en calcular hacia atrás para obtener el número efectivo de casos de cada casilla. Lo que podría efectuarse multiplicando el número total de casos tratados por la agencia por la *proporción* de una categoría determinada. Por ejemplo, (133) (.473) =63.

Cuadro III. 4. *Distribución en porcentajes de los casos tratados por tres agencias hipotéticas de servicios domésticos, con los porcentajes dispuestos verticalmente*

| Clase de casos | Agencia A (N=133) % | Agencia B (N=193) % | Agencia C (N=112) % |
|---|---|---|---|
| Matrimonios | 47.3 | 45.5 | 36.6 |
| Divorciados | 14.3 | 19.2 | 23.2 |
| Novios | 20.3 | 10.4 | 13.4 |
| Madres solteras | 9.8 | 16.6 | 18.8 |
| Otros | 8.3 | 8.3 | 8.0 |
| Total | 100.0 | 100.0 | 100.0 |

Obsérvese que los porcentajes dados en los cuadros III.3 y III.4 tienen por objeto contestar a ciertas preguntas y no otras. Nos permiten examinar cada agencia por separado y ver la distribución de los casos tratados. Permiten además la comparación de las agencias entre sí en relación con los casos tratados. Así, por ejemplo, la agencia B y C trataron relativamente más madres solteras y personas divorciadas de las que trató la agencia A. Supóngase, sin embargo, que nos interesaban ante todo los casos de cierto tipo y el número relativo de ellos tratados por cada agencia. Así, por ejemplo, podría eventualmente interesarnos saber el porcentaje de todos los matrimonios que pasaron por la agencia B. En estas condicione resultaría más conveniente calcular los porcentajes a través del cuadro. En efecto, podríamos tomar el número total de matrimonios y ver cuáles porcentajes de dicha categoría fueron tratados respectivamente por las agencias A, B y C. Los porcentajes sumarían entonces 100 en el sentido horizontal de cuadro, y no en el vertical, y los resultados se resumirían como el cuadro III. 5.

CUADRO III.5. *Distribución en porcentajes de los casos. Tratados por tres agencias hipotéticas de servicios domésticos, con los porcentajes calculados horizontalmente*

| Clase de casos | Agencia A (N=133) % | Agencia B (N=193) % | Agencia C (N=112) % | Total (N=438) % |
|---|---|---|---|---|
| Matrimonios(N=: 192) | 32.8 | 45.8 | 21.4 | 100.0 |
| Divorciados(N = 82) | 23.2 | 45.1 | 31.7 | 100.0 |
| Novios(N= 62) | 43.5 | 32.3 | 24.2 | 100.0 |
| Madres solteras(N = 66) | 19.7 | 48.5 | 31.8 | 100.0 |
| Otros(N= 36) | –* | –* | –* | –* |

* Los porcentajes no se calculan cuando la base es inferior a 50.

De modo que los porcentajes pueden calcularse tanto en sentido vertical como en sentido horizontal. Por lo tanto, los cuadros han de examinarse siempre cuidadosamente para ver exactamente cómo se han calculado aquéllos. Para los casos en que la propia teoría nos dicta cuál es la variable que debe ser considerada causalmente primaria o independiente, podrá bastarnos una simple regla empírica. Si tenemos la costumbre de situar la variable independiente en la parte alta del cuadro, y la variable dependiente al lado izquierdo, los porcentajes sumarán 100 hacia abajo, y las comparaciones se harán de *izquierda* a *derecha*. En el ejemplo relativo a la comparación de niveles de delincuencias en dos localidades, cabría normalmente suponer que ciertas características locales pueden tener influencia sobre la delincuencia, más bien que a la inversa.

Cuando computamos los porcentajes para que sumen 100 hacia abajo, lo que en realidad hacemos es normalizar los tamaños de las localidades, ya que reconocemos que los factores que se refieren a sus tamaños relativos, o los muestreos realizado dentro de cada localidad, *no* dependen causalmente de sus niveles de delincuencia. Al computar hacia abajo los porcentajes controlando aquellos factores que afectan al tamaño de los dos muestreos. Este punto quedará más en claro una vez que hayamos considerado el concepto de inclinación de una línea recta en la que una de las variables figura como dependiente de la otra (ver capítulo XVII).

Resultará que los porcentajes computados en la dirección sugerida pueden ser considerados como casos especiales de dichos declives.

## III.3. Razones

La razón de un número A con respecto a otro número B se define como A dividido entre B. La cantidad que precede se pone en el numerador, en tanto que la que sigue forma el denominador. Supóngase que en una elección local se hallan inscriptos 365 republicanos, 420 demócratas y 130 independientes en calidad de votantes. En este caso la razón de los republicanos a los demócratas es de 365/420, y la de los republicanos y los demócratas a los independientes es de (365+420)/130. Obsérvese que, a diferencia de la proporción, la razón puede tomar un valor superior a la unidad. Vemos asimismo que la expresión que precede o que sigue pueden constar, uno y otra, de cantidades distintas (v.gr. republicanos y demócratas). Generalmente la razón se reduce a su expresión más simple eliminando en el numerador y el denominador los factores comunes. Así, pues, la razón de los demócratas a los independientes se escribirá como 42/13 o bien, en forma equivalente, como 42:13. En ocasiones es conveniente expresar la razón en términos de un denominador formado por la unidad. Por ejemplo, la razón de los demócratas a los independientes puede escribirse como 3.23 a 1.

Es obvio que las proporciones representan un tipo especial de razón en la que el denominador es el número total de los casos y el numerador una cierta fracción del aquél. Sin embargo, el término de razón se emplea por lo regular para referirse a casos en los que A y B representan categorías separadas y distintas. Podríamos, por ejemplo, establecer la razón de los delincuentes a los no delincuentes, o de los matrimonios a los novios. Es evidente que con cuatro o cinco categorías el número de razones posibles susceptible de calcularse es muy grande. En consecuencia, a menos que el interés se centre ante todo en uno o varios pares de categorías, será en general más económico y menos sujeto a confusión por parte del lector servirse de los porcentajes y las proporciones. Obsérvese que, si las categorías sólo son dos, será posible calcular la proporción directamente a partir de la razón y viceversa. Así, por ejemplo, si sabemos que la razón de los varones a las mujeres es de 3:2, entonces en cada cinco personas ha de darse un promedio de tres varones y dos mujeres. La proporción de los varones es, pues, de 3/5, o 6.

Las razones pueden expresarse en términos de cualquier base que resulte conveniente. La base de la razón está indicada por la magnitud del denominador. Así, por ejemplo, las razones relativas al sexo se indican convencionalmente en términos del número de varones por 100 mujeres. Por lo tanto, una razón de 94 en materia de sexo indicará que el número de los varones es ligeramente inferior al de las mujeres, en tanto que una razón de sexos de 108 significaría una ligera preponderancia de los primeros. Las bases que comportan números grandes, tales como 1000 o 100000, se emplean a menudo al calcular *cuotas*, otro tipo de razón, cuando el empleo de las proporciones o los porcentajes conduciría a valores decimales pequeños. Las cuotas de natalidad, por ejemplo, suelen darse en términos del número de nacimientos vivos por 1000 mujeres en edad de procrear. Las cuotas de asesinatos pueden darse en términos del número de asesinos por 100000 habitantes.

Las cuotas de crecimiento constituyen otro tipo corriente de razón. Al calcular una de estas cuotas, tomamos el crecimiento efectivo durante el período considerado, dividido entre el volumen al *principio* del período. Así, por ejemplo, si la población de una ciudad aumenta de 50000 a 65000 entre 1940 y 1950, la cuota de crecimiento durante el decenio en cuestión será de

$$\frac{65000 - 50000}{50000} = .30$$

o 30 por ciento. En el caso de cuotas de crecimiento, es obvio, que los porcentajes se prestan bien más allá del 100 por ciento, en tanto que serán negativos si la ciudad ha experimentado un descenso de población.

GLOSARIO

Porcentaje
Proporción
Tasa
Razón

EJERCICIOS

1. Supóngase que se da el siguiente cuadro que muestra la relación entre la asistencia a la iglesia y el año de clase en una determinada universidad:

| *Asistencia* | *Año de clase* | | | | *Total* |
|---|---|---|---|---|---|
| *A la Iglesia* | *1er. Año* | *2º. Año* | *Inferior* | *Superior* | |
| Asistencia regular | 83 | 71 | 82 | 59 | 295 |
| Asistencia irregular | 31 | 44 | 61 | 78 | 214 |
| Total | 114 | 115 | 143 | 137 | 509 |

a) ¿Cuál es el porcentaje de asistencia regular en el conjunto? Respuesta, 57.96 %.

b) ¿Cuál es la razón de los estudiantes de primer año a los del año superior?

c) Entre los asistentes regulares, ¿cuál es la razón de los años inferiores a los superiores (de los 1º y 2º años a los años inferior y superior)? Respuesta, 1.09 a 1.

d) ¿Cuál es la proporción de los asistentes irregulares entre los estudiantes del año superior? ¿La proporción de estudiantes de año superior entre los asistentes irregulares? Respuesta .364; .569:

e) ¿Hay relativamente más asistentes irregulares entre los estudiantes de lº y 2º años que entre los de las clases inferior y superior? Exprésense los resultados en porcentajes.,

f) Resúmanse los datos en varías proposiciones.

2. Al estudiar la relación entre la productividad industrial y el tipo de líder de los grupos, un psicólogo social obtiene los siguientes datos, que muestran los niveles de productividad agrupados en tres tipos distintos de dirección:

| Productividad | Grupo de tipo de líder del grupo | | | Total |
|---|---|---|---|---|
| | Democrático | Liberal | Autoritario | |
| Alta | 37 | 36 | 13 | 86 |
| Mediana | 26 | 12 | 71 | 109 |
| Baja | 24 | 20 | 29 | 73 |
| Total | 87 | 68 | 113 | 268 |

a) ¿En qué dirección preferiría el lector calcular los porcentajes? ¿Por qué?
b) Calcúlense los porcentajes y resúmanse los datos en forma breve.
c) ¿Cuál es la razón de los productores de nivel alto a los de nivel bajo en cada uno de los grupos? En relación con estos datos particulares, ¿resumen las tres razones la situación de modo adecuado? Explíquese.

3. Si la razón de los blancos a los no blancos es de. 8/5 en una determinada localidad, ¿cuál es la proporción de los no blancos? Supóngase que la razón de los blancos a los negros fuera de 8/5, ¿podría obtenerse la proporción de negros en la misma forma? ¿Por qué, o por qué no?

4. Si una ciudad tenía una población de 153468 habitantes en 1940 y de 176118 en 1950, ¿Cuál fue la tasa de crecimiento (expresada en porcentaje) entre 1940 y 1950? Respuesta, 14.76%.

5. Si en un determinado condado hay 12160 varones y 11913 mujeres, ¿cuál es la razón entre los sexos (expresada en términos del número de varones por 100 mujeres)?

# BIBLIOGRAFÍA

**Anderson, T.R. y M. Zelditch**: *A Basic Course in Statistics*, 2ª ed, Holt, Rinehart and Winston, Inc. Nueva York, 1968, pp. 24 a31

**Freeman, L.C.** : *Elementary Applied Statistics*, John Wiley & Sons, Inc., Nueva York, 1965, cap. 4

**Hagood, M. J y D.O. Price**: *Statistics for Sociologists*, Henry Holt and Company, Inc., Nueva York, 1952, cap. 7.

**Weiss, R.S**. : *Statistics in Social Research*, John Wiley & Sons, Inc., Nueva York, 1968, cap. 4.

**Zeisel, Hans**: *Say It with* Figures, 5ª edición, Harper and Row, Publishers, Incorporated, Nueva York, 1968 caps. 1 y 2.