Bibliografía

1. Bradley, J. V.: *Distribution-free Statistical Tests,* Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968, caps. 1-3, 5, 11 y 13.
2. Dixon, W. J., y F. J. Massey: *Introduction to Statistical Analysis* 3ª ed., McGraw-Hill Book Company, Nueva York, 1969, cap. 17.
3. Freund, J. E.: *Modern Elementary Statistics,* 3ª ed., Prentice-Hall, Inc., Englewood Cliffs, N. J., 1967, cap. 13.
4. Hays, W. L.: *Statistics,* Holt, Rinehart and Winston, Inc., Nueva York, 1963, cap. 18.
5. Pierce, Albert: *Fundamentals of Nonparametric Statistics,* Dickenson Publishing Company, Inc. Belmont, Cal., 1970, cap. 14.
6. Savage, I. R.: "Bibliography of Nonparametric Statistics and Related Topics", *Journal of the American Statistical Association,* vol. 48, pp. 844-906, 1953.
7. Siegel, S.: *Nonparametric Statistics for the Behavioral Sciences,* McGraw-Hill Book Company, Inc., Nueva York, 1956, caps. 5 y 6.
8. Smith, K.: "Distribution-free Statistical Methods and the Concept of Power Efficiency", en L. Festinger y D. Katz (eds.) *Research Methods in the Behavioral Sciences,* The Dryden Press, Inc., Nueva York, 1953, pp. 536-577.
9. Swed, F. S., y C. Eisenhart: "Tables for Testing Randomness of Grouping in a Sequence of Alternatives", *Annals of Mathematical Statistics,* vol. 14, pp. 66-87, 1943.
10. Walker, H. M. y J. Lev: *Statistical Inference,* Henry Holt and Company, Inc. Nueva York, 1953, cap. 18.

XV. ESCALAS NOMINALES: PROBLEMAS DE CONTINGENCIA

En el presente capítulo vamos a estudiar las relaciones entre dos o más escalas nominales. Ya vimos que el caso de dos escalas nominales dicotómicas podía tratarse como un problema que comportara una diferencia de proporciones. Resulta a menudo deseable servirse de un procedimiento de prueba más general, que nos ponga en condiciones de averiguar las diferencias que haya entre tres o más muestras, o de comparar dos (o más) muestras con respecto a una variable de más de dos categorías. La prueba de la χ-cuadrada que vamos a examinar en la próxima sección nos permite establecer relaciones entre escalas nominales con cualquier número de categorías. Se introducirán al propio tiempo algunos conceptos nuevos. Hasta aquí sólo nos hemos ocupado de pruebas acerca de la *existencia* de una relación entre dos variables. En este capítulo se presentarán algunas medidas indicativas de la *fuerza* o grado de relación. Se examinarán al propio tiempo procedimientos empleados para el control de una o más variables.

XV.1. *La prueba de la χ-cuadrada*

La prueba de la χ-cuadrada es una prueba muy general que puede emplearse cuando deseamos apreciar si unas frecuencias obtenidas empíricamente difieren significativamente o no de las que se esperarían bajo cierto conjunto de supuestos teóricos. La prueba general presenta muchas posibilidades de aplicación, la más común de las cuales, en ciencias sociales, es la relativa a los problemas de "contingencia" en los que dos variables de escala nominal se han clasificado por comparación de una con otra.[1] Supóngase, por ejemplo, que se han relacionado una con otra la confesión religiosa y la filiación política y que los datos se han resumido en el siguiente cuadro de contingencia de 3×3:

Partido	*Protestantes*	*Católicos*	*Judíos*	*Total*
Republicanos	126	61	38	225
Demócratas	71	93	69	233
Independientes	19	14	27	60
Total	216	168	134	518

1 En relación con otro empleo de la χ-cuadrada, véase el ejercicio 3 al final del capítulo.

Obsérvese que si las frecuencias se convirtieran en porcentajes, podríamos decir que, en tanto que el 58.3 por ciento de los protestantes son republicanos, sólo prefieren este partido el 36.3 por ciento de los católicos y el 28.4 por ciento de los judíos. Se nos podría entonces ocurrir preguntar si esas diferencias eran o no significativas desde el punto de vista estadístico. Como quiera que se tienen tres confesiones religiosas y tres categorías de preferencia política, no podemos servirnos de una simple prueba de las diferencias de las proporciones. Sin embargo, sirviéndonos de la prueba de la χ-cuadrada, podemos establecer esencialmente la misma clase de hipótesis nula que anteriormente. Podemos suponer, en efecto, que no existe diferencia alguna entre las tres confesiones religiosas. Esto equivale a decir que las proporciones de republicanos, de demócratas y de independientes deberían ser las mismas en cada uno de dichos grupos. Partiendo, pues, del supuesto de que la hipótesis nula es correcta y de que las muestras son aleatorias e independientes, podemos calcular un conjunto de frecuencias que podrían esperarse, dados los totales marginales en cuestión. En otros términos, podemos calcular el número de protestantes de los que esperaríamos fueran republicanos y comparar esta cifra con la que se ha obtenido en realidad. Si la diferencia y las diferencias correspondientes a las otras casillas son considerables, probablemente sospechemos de la hipótesis nula.

Hay que obtener, pues, alguna medida de la diferencia entre las frecuencias observadas y las esperadas. Existe, por supuesto, una gran cantidad de medidas, pero necesitamos una con respecto a la cual la distribución de muestras sea conocida y esté tabulada. Por ello nos servimos de una media designada como de la χ-cuadrada (χ^2), que se define como sigue:

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} \qquad \text{(XV.1)}$$

en lo que f_o y f_e se refieren respectivamente a las frecuencias observadas y esperadas para cada casilla.[2] O en otras palabras: la χ-cuadrada se obtiene tomando primero el cuadrado de la diferencia entre las frecuencias observadas y esperadas para cada casilla. Dividimos dicha cifra entre el número de casos *esperados* en cada casilla, con objeto de normalizarla, de modo que las mayores contribuciones no provengan siempre de las casillas mayores. Y la suma de todas esas cantidades no negativas para todas las casillas es el valor de la χ-cuadrada.

[2] Con objeto de reducir la confusión hemos abandonado el índice *i*, suponiéndose, con todo, que estamos sumando los resultados de todas las casillas.

Obsérvese que cuanto mayores son las diferencias entre las frecuencias observadas y las esperadas, tanto mayor es el valor de la χ-cuadrada. Ésta sólo será cero si todas las frecuencias observadas y esperadas son idénticas. Podemos proceder a una verificación de la hipótesis nula buscando la distribución de muestreo de la χ-cuadrada. Difícilmente anticiparemos que las frecuencias observadas y las esperadas sean exactamente las mismas. Sin embargo, si el valor de la χ-cuadrada resulta mayor de lo que al azar se anticiparía, estaremos en condiciones de descartar la hipótesis nula siguiendo el procedimiento habitual.

Problema. Podemos servirnos del ejemplo puesto anteriormente, pero simplificándolo, de manera que obtengamos una tabla de 2×2. La extensión del mismo al caso general resultará después muy sencilla. Supongamos, pues, que se han combinado los católicos y los judíos y que se ha prescindido de los independientes. Tenemos así el siguiente cuadro:

Partido	*Protestantes*	*Católicos y judíos*	*Total*
Republicanos	126	99	225
Demócratas	71	162	233
Total	197	261	458

Importa observar que las cifras de cada casilla son en realidad frecuencias y no porcentajes. Si las cifras dadas son porcentajes, hay que convertirlas en frecuencias, ya que, desde el punto de vista estadístico, la prueba de la χ-cuadrada comporta una comparación de frecuencias y no de porcentajes.

1. *Supuestos.*

   Nivel de medición: dos escalas nominales
   Modelo: muestras aleatorias independientes
   Hipótesis: no existen diferencias entre las poblaciones confesionales en relación con la preferencia política.

Por supuesto, el nivel de medición *puede* ser más elevado. En efecto, las pruebas de la χ-cuadrada se utilizan con frecuencia con escalas ordinales e inclusive, en ocasiones, con escalas de intervalo. Sin embargo, según vimos en los capítulos precedentes, se dispone en tales casos de pruebas más fuertes que se emplearán por lo regular con preferencia a la χ-cuadrada. Una vez más, hay que suponer independencia entre las muestras para servirse de la prueba de la χ-cuadrada. La magnitud de la mues-

tra ha de ser relativamente grande, porque la χ-cuadrada, según la define la fórmula, tiene una distribución de muestreo que sólo se aproxima a la del cuadro si N es grande.[3]

La hipótesis nula puede formularse en cierto número de modos equivalentes. Decir que no hay diferencia entre grupos confesionales en materia de preferencia política equivale esencialmente a decir que no hay diferencia alguna entre la filiación religiosa y la preferencia electoral. Hay que tener presente, sin embargo, que semejante afirmación sólo se aplicaría a las variables tales como se las haya definido operativamente; en este caso, por ejemplo, la preferencia política y la religión se definirían como variables dicotómicas. Podría también enunciarse la hipótesis nula enumerando las diversas proporciones que se suponen iguales. Si bien este último método sea tal vez el más preciso, puede resultar con todo muy embarazoso en el caso general.

2. *Nivel de significación.* Supongamos que queremos demostrar una diferencia y que deseamos ser extremadamente cautos. Nos serviremos, en consecuencia, del nivel de .001. Supóngase asimismo que no se ha anticipado la dirección de la diferencia.

3. *Distribución de muestreo.* Las distribuciones de muestreo de la χ-cuadrada están dadas en el cuadro I del Apéndice 2. Obsérvese que las distribuciones difieren de acuerdo con los grados de libertad implicados. La determinación de los grados de libertad se examinará más abajo. Como quiera que, independientemente de la dirección de la relación entre la confesión y la preferencia política, nuestro interés está en saber si la χ-cuadrada obtenida es o no *mayor* de lo que se esperaría al azar, sólo nos ocupamos de la cola mayor de la distribución. La cola menor, que consta de valores muy pequeños de la χ-cuadrada, no se suele emplear por lo regular en los problemas de contingencia.

4. *Cálculo de la estadística de la prueba.* Lo primero que hacemos en el cálculo de la χ-cuadrada es obtener las frecuencias esperadas. La hipótesis nula dice que no hay preferencias de la gente en cuanto a la votación. Por lo tanto, independientemente de cuál sea el verdadero número de republicanos en cada una de las *poblaciones* confesionales, esperaríamos que, a la larga, habría la misma proporción de aquéllos en ambas *muestras*. Como quiera que la proporción de republicanos en la muestra combinada es de 225/458, o sea .4913, esperaríamos la misma cifra en cada una de las dos muestras confesionales. Así, pues, anticiparíamos en cada uno de ellos los mismos porcentajes de republicanos y de demócratas. Podemos obtener luego el *número* esperado de republicanos entre los protestantes multiplicando .4913 por el número total de protestantes de la muestra. En esta forma, el número anticipado de protestantes republicanos sería (.4913)

[3] Para un examen más detallado de este problema véanse las pp. 299-301.

(197) = 96.8. Las demás frecuencias anticipadas pueden calcularse en forma análoga. Por lo regular se recomienda retener por lo menos una cifra decimal al calcular las frecuencias esperadas. De modo que en el caso anterior no redondearíamos a 97.

Antes de pasar adelante, conviene observar que las frecuencias esperadas también pueden obtenerse razonando en forma inversa, esto es, en términos de la proporción de republicanos que esperaríamos que fueran protestantes. Toda vez que la proporción de protestantes en la muestra combinada es de 197/458, o sea .4301, podemos obtener la frecuencia anticipada de republicanos protestantes como sigue: (.4301) (225) = 96.8. El lector ha de acostumbrarse a obtener las frecuencias esperadas en ambas formas, a título de control de los cálculos.

Una vez que nos hayamos acostumbrado al procedimiento, encontraremos probablemente más sencillo servirnos de una simple fórmula como la que se describe a continuación. Si designamos las casillas y los totales marginales como

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

entonces la frecuencia esperada puede obtenerse multiplicando los dos marginales correspondientes a la casilla en cuestión y dividiendo entre N. Así, por ejemplo, la cifra esperada para la casilla a sería

$$(a+b)\,(a+c)/N = (225)\,(197)/458 = 96.8$$

El empleo de este último procedimiento reduce todo error de redondeo que podría introducirse dividiendo primero (para obtener la proporción) y multiplicando luego.

Se observará que este procedimiento de multiplicar marginales para dividirlos entre el número total de casos, viene a ser básicamente el mismo que se examinó en el capítulo IX en relación con la independencia de dos variables. Esto pone de relieve el hecho de que las frecuencias esperadas son computadas sobre la base del *supuesto* de que las variables no están relacionadas, en tanto que las frecuencias observadas nos muestran el grado en que se viola este supuesto. Recuérdese que si los eventos (o variables) A y B son estadísticamente independientes, el conocer el valor de uno no nos ayudará a predecir el otro. Si las frecuencias observadas y las esperadas son exactamente iguales, ello significaría, en nuestro ejemplo, que el conocer las diferencias religiosas de una persona no nos permitiría predecir sus inclinaciones políticas.

Por convención, ponemos por lo regular las frecuencias esperadas entre paréntesis, debajo de las frecuencias realmente obtenidas para cada casilla, tal como se indica a continuación:

Partido	*Protestantes*	*Católicos y judíos*	*Total*
Republicanos	126	99	225
	(96.8)	(128.2)	
Demócratas	71	162	233
	(100.2)	(132.8)	
Total	197	261	458

Los cálculos para la χ-cuadrada pueden resumirse en un cuadro como el XV.1. Obsérvese que la cantidad $f_o - f_e$ tiene el mismo valor para cada casilla. El lector debería convencerse por sí mismo de que esto será siempre así en el caso de tablas de 2×2, pero que no se deja con todo generalizar a otros casos. El hecho de elevar este valor al cuadrado tiene por efecto la eliminación de las cantidades negativas. Importa que se empleen en el denominador las frecuencias *esperadas*, y no las observadas. En efecto, estas últimas variarán de una muestra a otra, y pueden incluso ser iguales a cero.

CUADRO XV.1. *Cálculos de la χ-cuadrada*

Casilla	f_0	f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$(f_0 - f_e)^2/f_e$
a	126	96.8	29.2	852.64	8.808
b	99	128.2	−29.2	852.64	6.651
c	71	100.2	−29.2	852.64	8.509
d	162	132.8	29.2	852.64	6.420
Total	458	458.0			30.388

Resulta a menudo más conveniente servirse de una fórmula de cálculo que no requiera la sustracción efectiva de cada frecuencia esperada de su correspondiente observada. Desarrollando el numerador en la expresión de la χ-cuadrada y uniendo los términos obtenemos:

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = \Sigma \frac{f_o^2 - 2f_o f_e + f_e^2}{f_e}$$

$$= \Sigma \frac{f_o^2}{f_e} - 2\Sigma f_o + \Sigma f_e$$

Pero, toda vez que tanto Σf_o como Σf_e son iguales a N, tenemos:

$$\chi^2 = \Sigma \frac{f_o^2}{f_e} - N \qquad \text{(XV.2)}$$

Sirviéndonos de esta fórmula, que comporta una sola sustracción, obtenemos el mismo resultado que anteriormente (véase cuadro XV.2).

CUADRO XV.2. *Cálculo de la χ-cuadrada sirviéndose de la fórmula*

Casilla	f_o^2	f_o^2/f_e
a	15 876	164.008
b	9 801	76.451
c	5 041	50.309
d	26 244	197.620
Total		488.388

$$\chi^2 = 488.388 - 458$$
$$= 30.388$$

En el caso de una tabla de solamente 2×2, resulta posible expresar la χ-cuadrada como simple función de las frecuencias de las casillas y de los totales marginales. Si se designan las casillas como anteriormente, tenemos:

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad \text{(XV.3)}$$

Si bien este cálculo requiere la multiplicación de números grandes, el empleo de los logaritmos lo simplificará con todo considerablemente. Vemos el paso, de la ecuación (XV.3), que la χ-cuadrada será cero cuando el producto diagonal *ad* sea exactamente igual al producto *bc*. Este hecho puede emplearse como método rápido para saber si es o no necesario seguir adelante con la prueba de significación. Si los productos diagonales son casi iguales, la χ-cuadrada será demasiado pequeña para proporcionar significación. Estos productos diagonales sirven asimismo para determinar la dirección de la relación sin que tengamos que molestarnos en calcular los porcentajes. El mayor de los dos productos indica, en efecto, cuál de las diagonales contiene la mayoría de los casos.

* Tanto las anteriores fórmulas para χ (chi) al cuadrado, como el procedimiento para calcular frecuencias esperadas, son sufi-

cientes en la mayoría de los casos, pero puede resultar útil conocer una versión algo distinta, aplicable al caso $r \times c$ en general, conveniente para quienes deseen proseguir el tema de la χ al cuadrado en otros textos más avanzados. Esta formulación alternativa será utilizada más adelante para obtener el límite superior de χ al cuadrado en el caso general $r \times c$. Por otra parte, esta forma alternativa para la fórmula no requiere el cálculo explícito de las frecuencias esperadas.

Sea N_{ij} = número observado en (i, j)-ésima casilla del cuadro, y
e_{ij} = número esperado (bajo H_0) en la casilla (i, j),

para $i = 1, 2, \ldots, r$; y $j = 1, 2, \ldots, c$.

Sea $N_{i.} = \sum_{j=1}^{c} N_{ij}$, para $i = 1, 2, \ldots, r$ (total de filas), y

$N_{.j} = \sum_{i=1}^{r} N_{ij}$, para $j = 1, 2, \ldots, c$ (total de columnas).

Así podremos expresar χ al cuadrado como sigue

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(N_{ij} - e_{ij})^2}{e_{ij}}$$

pero puesto que

$$e_{ij} = N \frac{N_{i.}}{N} \frac{N_{.j}}{N} = \frac{N_{i.}\, N_{.j}}{N}$$

la fórmula computadora (XV.2) pasa a ser

$$\chi^2 = N \left[\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{N_{ij}^2}{N_{i.}\, N_{.j}} - 1 \right]$$

y así vemos que no hay necesidad de computar explícitamente las frecuencias esperadas.

5. *Decisión.* Antes de servirnos del cuadro de la χ-cuadrada, hemos de determinar los grados de libertad asociados a esta estadística de prueba. En los problemas anteriores, los grados de libertad dependían siempre del número de los casos seleccionados. En los problemas de contingencia, en cambio, dichos grados sólo dependen del número de *casillas* del cuadro. Al calcular las frecuencias esperadas, pudo observarse que no es necesario calcular valores para cada casilla, ya que la mayoría de ellas podían obtenerse por sustracción. Y de hecho, en la tabla de 2×2 sólo

necesitamos calcular una de las frecuencias esperadas, y las otras quedan automáticamente determinadas. Esto es así porque, para calcular las frecuencias esperadas, nos servimos de los totales marginales de nuestra muestra. En otros términos: si ponemos el valor de una casilla cualquiera, los demás valores están perfectamente determinados, ya que las frecuencias esperadas han de tener los mismos totales marginales que las observadas. Por lo tanto, sólo tenemos un grado de libertad.

Habiendo, pues, averiguado que en la tabla de 2×2 sólo hay un grado de libertad, buscamos en el cuadro de la χ-cuadrada a lo largo de la hilera correspondiente a un grado de libertad hasta encontrar el nivel de significación deseado. Vemos en esta forma que al nivel de .001 le corresponde una χ-cuadrada de 10.827. Esto significa que, si todos los supuestos son efectivamente correctos, obtendremos un valor de la χ-cuadrada igual o mayor que ése una vez entre mil. En otros términos: sólo muy raramente diferirán las frecuencias observadas y las esperadas en una cantidad que dé una χ-cuadrada $\geq$ 10.827, *si* no hubiera relación alguna entre la confesión religiosa y la preferencia en cuanto al voto (tal como se ha definido operativamente en este problema). Y como quiera que hemos obtenido para la χ-cuadrada un valor igual a 30.388, concluimos que la hipótesis nula puede descartarse al nivel de .001. Vemos, de paso, que, si N es grande, no es nada difícil llegar a obtener significación al nivel de .001.

Pese a que sólo nos ocupáramos de valores grandes de la χ-cuadrada, la dirección de la relación no se anticipó en el ejemplo anterior. Independientemente de si los protestantes presentaban más probabilidades de ser republicanos o demócratas, el resultado habría sido una χ-cuadrada grande si los porcentajes eran también grandes. En otros términos, la estadística de la prueba es aquí indiferente a la dirección de la relación, ya que comporta los cuadrados de las desviaciones y, por consiguiente, no puede ser negativa. Podemos sacar partido de las predicciones relativas a la dirección partiendo simplemente por la mitad el nivel de significación obtenido. En efecto, si la χ-cuadrada es lo bastante grande para dar significación al nivel de .10 sin anticipar dirección, el resultado será también significativo al nivel de .05, a condición, por supuesto, que la dirección de la relación se haya fijado de antemano.

Si el nivel de significación deseado no puede obtenerse exactamente de la tabla de la χ-cuadrada, se conseguirá una aproximación satisfactoria extrayendo la raíz cuadrada de la χ-cuadrada y recurriendo a la tabla normal. Así, por ejemplo, sabemos que una χ-cuadrada de 3.841 con un grado de libertad corresponde al nivel de .05 si no se ha adivinado la dirección. La raíz cuadrada de esta cifra es 1.96, que es el valor de Z necesario para obtener

significación con la tabla normal. Ésta, sin embargo, sólo puede emplearse en el caso de problemas de contingencia de 2×2.

Caso general. En el caso general de la tabla de contingencia con r hileras y c columnas, los supuestos y cálculos para la χ-cuadrada sólo requieren una ligera modificación. La hipótesis nula de "ausencia de diferencias" o "ausencia de relación" implica ahora que cada población tendrá las mismas proporciones para cada una de las categorías de la segunda variable. Las frecuencias esperadas pueden obtenerse exactamente en la misma forma que anteriormente, pero tendremos ahora rc casillas, y los grados de libertad serán distintos.

Supóngase que nos servimos del mismo problema anterior, pero en su forma original, o sea la de una tabla de 3×3. Observemos de paso que esta tabla nos proporciona mayor información que la de 2×2, en la que los católicos y los judíos se combinaron en una sola categoría. Podemos, por lo tanto, esperar resultados que difieran algo de aquellos obtenidos anteriormente. Calculando las frecuencias esperadas por uno cualquiera de los métodos anteriormente sugeridos, obtenemos:

Partido	*Protestantes*	*Católicos*	*Judíos*	*Total*
Republicanos	126 (93.8)	61 (73.0)	38 (58.2)	225
Demócratas	71 (97.2)	93 (75.6)	69 (60.2)	233
Independientes	19 (25.0)	14 (19.4)	27 (15.6)	60
Total	216	168	134	518

Puede construirse una tabla de cálculo lo mismo que anteriormente (véase cuadro XV.3).

Para determinar los grados apropiados de libertad, observamos que, una vez las dos primeras frecuencias esperadas inscritas en la primera columna, la tercera se halla determinada por sustracción. Y lo mismo es cierto de la segunda. Todas las frecuencias esperadas de la tercera columna estarán determinadas a partir de los totales de la hilera. En términos generales: para cada una de las primeras $c-1$ columnas será posible llenar todas las casillas menos una, o $r-1$. La columna final estará, pues, siempre perfectamente determinada. Por lo tanto, el número de los grados de libertad de la tabla de contingencia de $r \times c$ puede expresarse por medio de la fórmula

$$df = (r-1)(c-1)$$

CUADRO XV.3. *Cálculo de la χ-cuadrada para una tabla de contingencia de 3×3*

Casilla	f_o	f_e	f_o^2	f_o^2/f_e
a	126	93.8	15 876	169.254
b	61	73.0	3 721	50.973
c	38	58.2	1 444	24.811
d	71	97.2	5 041	51.862
e	93	75.6	8 649	114.405
f	69	60.2	4 761	79.086
g	19	25.0	361	14.440
h	14	19.4	196	10.103
i	27	15.6	729	46.731
Total	518	518.0		561.665

$$\chi^2 = 561.665 - 518 = 43.665$$

Obsérvese que esta fórmula da un grado de libertad en el caso especial en que $r = c = 2$.

Toda vez que son 4 los grados de libertad asociados a nuestra tabla de 3×3, vemos que para el rechazo al nivel de .001 se requiere una χ-cuadrada de 18.465. Rechazamos, por consiguiente, la hipótesis nula. Obsérvese que si para rechazar se requiere un valor mayor de la χ-cuadrada, es porque hay muchas más casillas que contribuyen a dicho valor. Como quiera que la χ-cuadrada representa una suma y no un promedio, esperaríamos que, en igualdad de condiciones, cuanto mayor sea el número de casillas, tanto mayor será la χ-cuadrada. El hecho de que el valor de la χ-cuadrada requerido para obtener significación aumente con los grados de libertad no debería sorprendernos.[4]

Corrección de continuidad. Ya se indicó que la prueba de la χ-cuadrada requiere una N relativamente grande debido al hecho de que la distribución de muestreo de la estadística de la prueba sólo se aproxima a la distribución de muestreo dada en la tabla de la χ-cuadrada si N es grande. Plantéase, pues, naturalmente la cuestión de cuán grande debe ser N para que podamos servirnos de dicha prueba. La respuesta depende del número de casillas y de los totales marginales. Generalmente, cuanto menor sea el número de casillas y cuanto más aproximadamente iguales sean todos los totales marginales, tanto menor podrá ser N. Los criterios normalmente utilizados para decidir si el número de casos es o no suficiente, implican las frecuencias *esperadas* de cada casilla. Siempre que cualquiera de estas frecuencias sea

[4] Obsérvese que esto era al revés en el caso de la distribución t. ¿Por qué?

aproximadamente de cinco o menor, se recomienda proceder a alguna clase de modificación, como se indica a continuación.

Se supone que la distribución de la χ-cuadrada es continua. En realidad, sin embargo, si el número de casos es relativamente pequeño, resulta imposible que el valor calculado de la χ-cuadrada tome muchos valores distintos. Esto es así porque las frecuencias observadas han de ser siempre números enteros. Al corregir con fines de continuidad, nos imaginamos que las frecuencias observadas pueden tomar efectivamente todos los valores posibles y nos servimos de los que quedan a una distancia de media unidad a uno y otro lado del entero obtenido, lo que dará los resultados más conservadores. En el caso de la tabla de 2×2, la corrección de continuidad puede hacerse muy fácilmente. Esta corrección consiste ya sea en añadir o sustraer .5 de las frecuencias observadas, con objeto de reducir el tamaño de la χ-cuadrada. La versión corregida de la ecuación (XV.3) es la siguiente:

$$\chi^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Para apreciar el efecto de la corrección de continuidad, podemos ver los siguientes cuadros:

(*A*)	7 (10)	13 (10)	20
	8 (5)	2 (5)	10
	15	15	30
	$\chi^2 = 5.40$		

(*B*)	7.5 (10)	12.5 (10)	20
	7.5 (5)	2.5 (5)	10
	15	15	30
	$\chi^2 = 3.75$		

En el cuadro *B* hemos corregido por razones de continuidad reduciendo las diferencias entre las frecuencias observadas y esperadas en media unidad. Hemos supuesto que había entre 6.5 y 7.5 casos en la casilla superior de la izquierda, y hemos tomado el número de 7.5, porque es el valor más cercano, al interior de dicho intervalo, de la frecuencia esperada de 10.0. En este ejemplo, la corrección de continuidad reduce el nivel de significación de aproximadamente .02 a algo más de .05. Es obvio, por lo demás, que las correcciones de continuidad producirán menos efecto cuando las frecuencias esperadas sean mayores. Toda vez que semejante corrección comporta en realidad un esfuerzo adicional muy pequeño y que, por otra parte, al proceder así actuamos en sentido conservador, se recomienda efectuarla siempre que en cualquier casilla la frecuencia esperada descienda por debajo de 10. Con muestras muy pequeñas, incluso esta corrección produce resultados engañosos. Para las tablas de 2×2 se dispone de una prueba alternativa que se examina en la sección siguiente.

En el caso de la tabla general de contingencia, las correcciones de continuidad no son fáciles de hacer. Si el número de casillas es relativamente grande y si solamente una o dos de las casillas tienen frecuencias esperadas de 5 o menos, entonces recomiéndase, por lo general, seguir adelante con las pruebas de la χ-cuadrada, sin preocuparse mucho por tales correcciones. En cambio, si el número de casillas es pequeño, la única alternativa práctica consistirá tal vez en combinar las categorías de modo que dichas casillas resulten eliminadas. Por supuesto, las categorías sólo pueden combinarse si ello posee teóricamente algún sentido. Así, por ejemplo, si hubiera una categoría "de otras confesiones" que constara de un número tan grande de grupos confesionales que la categoría no tuviera teóricamente sentido alguno, tal vez sería preferible excluir a dichas personas por completo del análisis aunque, como regla general, no es buen sistema el de excluir datos de un análisis.

*XV.2. *La prueba exacta de Fisher*

En el caso de tablas de 2×2 en las que N es muy pequeña, es posible servirse de una prueba desarrollada por R. A. Fisher, que nos da probabilidades exactas, y no aproximadas. Si designamos las casillas y los marginales de la tabla de 2×2 de la siguiente manera:

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

podemos conseguir la probabilidad de obtener *exactamente* esas frecuencias en la hipótesis nula de que no hay diferencias en las proporciones de las poblaciones. Esta probabilidad nos está dada por la fórmula:

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!}$$

Esta fórmula de probabilidad puede obtenerse utilizando la distribución hipergeométrica para el cálculo de probabilidades sobre la base de muestreo *sin reposición*. En esta prueba, como en

algunas otras pruebas no paramétricas, podremos entender el problema como si éste contuviera repetidas muestras de una "población" de tamaño N. Tratamos así nuestra muestra obtenida como si se tratara de una población real, e imaginamos en este ejemplo que las categorías de nuestros casos les dan cabida en una de las cuatro casillas. Como hay $a+c$ individuos en la primera columna, $a+b$ en la primera fila, y así sucesivamente, ¿cuál será la probabilidad de que de los $a+b$ individuos de la primera fila correspondan exactamente a a la primera columna y b a la segunda? Nos imaginamos haber muestreado $a+b$ individuos al azar pero sin reposición, colocándolos en la primera fila, con los restantes cayendo por necesidad en la segunda fila. En efecto, resulta que imaginamos que llenamos las casillas por un proceso esencialmente al azar, y preguntamos cuál hubiera sido la exactitud de los resultados si hubiese sido seguido tal proceso.

Aplicando la fórmula para la distribución hipergeométrica dada en la sección X.4, veremos que la probabilidad de obtener exactamente a y b casos en las dos casillas de la fila superior vendría dada por

$$P(a,b)=\frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{N}{a+b}}$$

Escribiendo cada uno de los términos en función de factoriales, y simplificando, obtenemos:

$$P(a,b)=\frac{\dfrac{(a+c)!}{a!(a+c-a)!}\dfrac{(b+d)!}{b!(b+d-b)!}}{\dfrac{N!}{(a+b)!\,(N-a-b)!}}=\frac{\dfrac{(a+c)!\,(b+d)!}{a!c!\quad b!d!}}{\dfrac{N!}{(a+b)!\,(c+d)!}}=$$

$$=\frac{(a+c)!(b+d)!(a+b)!(c+d)!}{N!a!b!c!d!}$$

Puede comprobarse fácilmente que se habría conseguido el mismo resultado si hubiéramos concebido el problema como orientado a seleccionar una muestra de $a+c$ individuos, asignándolos a continuación a la primera *columna*.

Obsérvese que hay nueve factoriales en esta fórmula de P. Por lo tanto, la tarea de calcularla sería formidable. Por otra parte, como quiera que normalmente se está interesado en obtener la

cola entera de la distribución de muestreo y no la probabilidad de averiguar exactamente los resultados obtenidos, habría que añadir, a esta probabilidad primera, las probabilidades de obtener incluso más resultados poco corrientes en la misma dirección.

Un sencillo ejemplo numérico ilustrará lo que esto significa. Supóngase que hemos obtenido la siguiente tabla de 2×2:

3	9	12
12	5	17
15	14	29

Si suponemos que los marginales permanecen fijos, vemos inmediatamente que hay tres resultados (en la misma dirección) que son incluso más difíciles de obtenerse. Son los siguientes:

2	10	12
13	4	17
15	14	29

1	11	12
14	3	17
15	14	29

0	12	12
15	2	17
15	14	29

Obsérvese que podemos llegar a las tablas sucesivas reduciendo cada vez en uno las casillas a y d y aumentando en uno las casillas b y c, hasta llegar a la tabla final, en la que la casilla a está vacía.

Supongamos que la casilla a es siempre la que contiene el menor número de casos, ya que siempre tendremos la posibilidad de disponer las tablas en tal forma.[5] Sirvámonos del símbolo P_0 para designar la probabilidad de obtener exactamente cero casos en la casilla a (dados los marginales en cuestión), en la hipótesis nula; pongamos que P_1 representa la probabilidad de obtener exactamente un caso en la casilla a, P_2 la de obtener exactamente dos casos, etcétera. Así, pues, en este problema particular hemos de obtener la suma de las probabilidades

$$P_0+P_1+P_2+P_3$$

para calcular la probabilidad de obtener tres o menos casos en la casilla a. Y ya que nos estamos sirviendo de una prueba de

[5] En raros casos cambiará la dirección de la relación si se sigue la regla de que la casilla a sea siempre la más pequeña. Por ejemplo, si las dos distribuciones marginales son muy desiguales, la regla tal vez no se aplique. Así, si a, b, c y d son 1, 2, 3 y 7, respectivamente, el producto ad $(=7)$ es mayor que el producto bc $(=6)$. Si uno reduce entonces a hasta 0, las casillas resultantes serán 0, 3, 4 y 6, y se producirá una inversión de dirección, puesto que $bc > ad$. Deben ser vigiladas tales inversiones y, en caso de que se produzcan, deberá denominarse como a la casilla más pequeña en la menor de las dos diagonales.

una sola cola, habremos de doblar el nivel de significación obtenido, si no estuviéramos en condiciones de poder predecir la dirección.[6]

Será mucho más conveniente que calcular cada una de las P_i de la fórmula anterior, que comporta productos de factoriales, obtener P_0 directamente y obtener luego las probabilidades restantes como funciones de P_0. Con objeto de distinguir entre las varias combinaciones posibles de los valores numéricos de a, b, c y d en el caso de marginales fijos, sirvámonos de un subíndice k para designar la magnitud de la casilla más pequeña a. Así, por ejemplo, si hay k individuos en la casilla a, designaremos las cantidades de las diversas casillas como $a_k\,(=k)$, b_k, c_k y d_k. Toda vez que se supone que los marginales permanecen fijos, si disminuimos a_k y d_k en uno, hemos de aumentar b_k y c_k también en uno. Podemos ahora simplificar la fórmula de P_0, ya que $a_0 = 0$ y, por consiguiente, $a_0! = 1$ (por definición), $(a_0 + b_0)! = b_0!$, y $(a_0 + c_0)! = c_0!$. O sea que cierto número de factoriales se eliminan, dejándonos con:

$$P_0 = \frac{(c_0 + d_0)!(b_0 + d_0)!}{N!d_0!}$$

El numerador consta ahora solamente de los factoriales de dos de los marginales, en lugar de los cuatro, y el denominador sólo comporta $N!$ y $d_0!$. El valor de d_0 puede obtenerse de la última de las tablas anteriores. Por lo tanto, en este ejemplo, $(c_0 + d_0) = 17$, $(b_0 + d_0) = 14$, $N = 29$, y $d_0 = 2$. P_0 puede calcularse ahora sirviéndonos de una tabla de logaritmos de factoriales, o bien escribiendo los factoriales y simplificando.

Con objeto de calcular los valores de P_1, P_2 y P_3 necesitamos ahora una fórmula general de P_{k+1} en función de P_k. Ya que los marginales se suponen fijos, tenemos:

$$P_{k+1} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!(a_k+1)!(b_k-1)!(c_k-1)!(d_k+1)!}$$

debido al hecho de que, al añadir uno a la casilla a, lo añadimos también a la casilla d y lo sustraemos tanto de b como de c. Si dividimos ahora P_{k+1} entre P_k, prácticamente todos los términos desaparecen. En efecto, los numeradores de ambas probabilidades son idénticos, ya que todos ellos comportan los mismos marginales. El factorial de N se elimina. Y nos queda:

[6] En un sentido estricto, la prueba de Fisher deberá ser usada probablemente sólo en el caso en que previamente se hubiera predicho la dirección, ya que las dos colas casi nunca serán perfectamente simétricas.

$$\frac{P_{k+1}}{P_k} = \frac{a_k!b_k!c_k!d_k!}{(a_k+1)!(b_k-1)!(c_k-1)!(d_k+1)!}$$

Pero $a_k!/(a_k+1)!$ es igual a $1/(a_k+1)$, y lo mismo por lo que se refiere a $d_k!/(d_k+1)!$. O sea, pues, $b_k!/(b_k-1)! = b_k$, y $c_k!/(c_k-1)! = c_k$. Por consiguiente:

$$\frac{P_{k+1}}{P_k} = \frac{b_k c_k}{(a_k+1)(d_k+1)}$$

o sea

$$P_{k+1} = \frac{b_k c_k}{(a_k+1)(d_k+1)} P_k$$

y los factoriales fastidiosos han desaparecido. Por lo tanto, podemos servirnos de esta fórmula para obtener P_1 a partir de P_0. Una vez obtenida P_1 podemos calcular P_2, y así sucesivamente.

Volviendo a nuestro ejemplo numérico, obtenemos P_0 como sigue:

$$P_0 = \frac{14!17!}{29!2!} = .17535 \times 10^{-5}$$

Y por consiguiente:

$$P_1 = \frac{b_0 c_0}{(a_0+1)(d_0+1)} P_0 = \frac{12(15)}{1(3)}(.17535 \times 10^{-5}) = 10.521 \times 10^{-5}$$

Al calcular P_2 hemos de cuidar de servirnos de a_1, b_1, c_1 y d_1, y no de las cifras empleadas para obtener P_1. Tenemos, así:

$$P_2 = \frac{b_1 c_1}{(a_1+1)(d_1+1)} P_1 = \frac{11(14)}{2(4)}(10.521 \times 10^{-5}) = 202.529 \times 10^{-5}$$

Y análogamente:

$$P_3 = \frac{b_2 c_2}{(a_2+1)(d_2+1)} P_2 = \frac{10(13)}{3(5)}(202.529 \times 10^{-5}) = 1\,755.252 \times 10^{-5}$$

Obsérvese que cada uno de los factores del numerador va disminuyendo en 1, al calcular P_{k+1} a partir de P_k, en tanto que los del denominador van aumentando cada vez en una unidad. Sumando las probabilidades tenemos, pues:

$$P_0 + P_1 + P_2 + P_3 = (.175 + 10.521 + 202.529 + 1\,755.252) \times 10^{-5}$$
$$= 1\,968.48 \times 10^{-5} = .0197$$

Por lo tanto, la probabilidad de obtener tres o menos individuos en la casilla a es, con la hipótesis nula, de .02, y tomaremos nuestra decisión de rechazar o no la hipótesis nula en consecuencia.

Debido a que la prueba de Fisher es exacta, merece preferencia respecto de la prueba de la χ-cuadrada corregida con fines de continuidad. Y como quiera que por lo regular la prueba de la χ-cuadrada dará probabilidades algo más bajas que la prueba de Fisher, si lo que se desea en realidad es rechazar la hipótesis nula, obraremos, al servirnos de ésta, en sentido conservador. En otros términos, si nos servimos de la prueba de la χ-cuadrada, puede ser que lleguemos a probabilidades que en realidad sean demasiado pequeñas, lo que nos llevaría acaso a la conclusión de que la hipótesis nula deba descartarse cuando en realidad no sea así. Si la frecuencia mínima esperada es sensiblemente superior a 5 y si se emplea la corrección de continuidad, las dos pruebas darán aproximadamente los mismos resultados. Aun logrando evitar el empleo de factoriales en el caso de la prueba de Fisher, se echa de ver que, si la frecuencia menor de la casilla es mayor que 5, los cálculos necesarios podrán resultar muy fastidiosos. De ahí que se encuentre que dicha prueba resulta más práctica en el caso de N muy pequeñas, o siempre que el tamaño de la muestra sea moderado y uno o más de los marginales sean muy pequeños. En los casos en que ambos, $(a+b)$ y $(c+d)$ son ≤ 30, existen tablas en (3) que simplifican considerablemente el empleo de esa prueba exacta.

XV.3. *Medidas de la fuerza de la relación*

Hasta aquí sólo nos hemos ocupado de la cuestión de saber si existía o no una relación entre variables. Hemos establecido hipótesis nulas en el sentido de que no se daba relación alguna, y hemos tratado de descartarlas. Pero, cuando estamos en condiciones de descartar, ¿qué es lo que hemos logrado? Designamos una relación como estadísticamente significativa cuando hemos establecido, bajo el riesgo de error de tipo I, que *sí existe* una relación entre las dos variables. Sin embargo, ¿quiere esto decir que la relación es significativa en el sentido de ser una relación fuerte o importante? No necesariamente. En efecto, la cuestión de la fuerza de la relación es totalmente distinta de la de su existencia. En esta sección vamos a ocuparnos de diversas medidas de grados de asociación que ayudan a contestar la segunda de las preguntas.

A primera vista podría parecer razonable tratar de establecer la fuerza de la relación observando simplemente el nivel de significación conseguido con una prueba. Así, por ejemplo, podría discurrirse en el sentido de que si una prueba es significativa al nivel de .001 y otra al nivel de .05, la primera sería la más fuerte

de las dos. Pero, ¿es esto necesariamente así? El examen de los dos niveles de significación nos dirá en cuál caso podemos estar más seguros de que la relación existe. Así, en el primero de los dos casos citados estaríamos casi seguros de que *existe* efectivamente una relación, pero no lo estaríamos tanto en el segundo. Hemos de recordar, no obstante, que el nivel de significación alcanzado depende del tamaño de las muestras usadas. En efecto, como se indicó anteriormente, si las muestras son muy grandes, resulta por lo regular muy fácil establecer significación, aun en el caso de una relación muy superficial. Esto significa, de hecho, que, cuando las muestras son grandes, decimos en realidad muy poca cosa al afirmar que hemos establecido una relación "significativa". En el caso de muestras grandes, es mucho más importante preguntar, "dado que existe una relación, ¿cuál es su fuerza?"

Con objeto de ilustrar lo que se acaba de decir, veamos un poco más de cerca cierta propiedad de la χ-cuadrada. Al hacerlo, el lector deberá tener presente que los mismos principios se aplican exactamente a otras clases de pruebas de significación. Preguntémonos qué sucede con la χ-cuadrada cuando el número de casos aumenta. Con fines de ilustración podemos tomar la siguiente tabla de 2×2.

30	20	50
20	30	50
50	50	100

La χ-cuadrada de esta tabla resulta ser exactamente 4.0. Supongamos ahora que se duplican los tamaños de las muestras, manteniendo las mismas proporciones en cada casilla. Obtendríamos así:

60	40	100
40	60	100
100	100	200

y la χ-cuadrada sería 8.0, o sea una cifra exactamente doble de la anterior. Examinando la fórmula de la χ-cuadrada, resulta muy fácil demostrar que, si las proporciones de las casillas permanecen inalteradas, la χ-cuadrada varía directamente con el número de casos. Si duplicamos el número de éstos, duplicamos aquélla, y si triplicamos los primeros, triplicamos la segunda. Supóngase que el número de casos inicial se multiplica por el factor k. Entonces, como quiera que las proporciones de las casillas permanecen inalteradas, toda nueva frecuencia observada será exactamente k veces la anterior, y lo mismo por lo que se refiere a las

frecuencias esperadas. La nueva χ-cuadrada puede, pues, expresarse como:

$$(\chi^2)' = \Sigma \frac{(kf_o - kf_e)^2}{kf_e} = \Sigma \frac{k^2(f_o - f_e)^2}{kf_e} = k \Sigma \frac{(f_o - f_e)^2}{f_e}$$

Así, pues, el valor de la nueva χ-cuadrada es exactamente k veces el de la primitiva.

Las implicaciones de este hecho pueden destacarse por medio de otra ilustración. Supóngase que obtenemos los siguientes resultados al relacionar las diferencias de sexo con la tolerancia respecto de conductas anómalas:

Tolerancia	*Varones*	*Mujeres*
Alta	26	24
Baja	24	26

En este caso la χ-cuadrada es 0.16, y estaremos en lo cierto informando que la relación no es significativa. Supóngase, sin embargo, que el estudio fue muy ambicioso y que se reunieron datos correspondientes a 10 000 casos, con los siguientes resultados:

Tolerancia	*Varones*	*Mujeres*
Alta	2 600	2 400
Baja	2 400	2 600

La χ-cuadrada es ahora 16.0, o sea un valor altamente significativo desde el punto de vista estadístico. Sin embargo, si hubiéramos expresado los resultados en términos de porcentajes, la cosa se habría presentado como mucho menos interesante. Si dijéramos que el 52 por ciento de los varones era altamente tolerante, en tanto que *sólo* correspondía a dicha categoría el 48 por ciento de las mujeres, nos criticarían con razón por destacar las diferencias aparentemente insignificantes tanto desde el punto de vista teórico como del significado práctico. Este ejemplo ilustra un punto muy importante. En efecto, una diferencia puede ser interesante estadísticamente sin serlo en ningún otro sentido. En el caso en que se seleccionaron 10 000 casos, podemos estar bien seguros que hay cierta relación superficial, que produciría una relación significativa desde el punto de vista estadístico.

Vemos, pues, que si una muestra es pequeña, se requiere una relación mucho más manifiesta para obtener significación. Por lo tanto, con las muestras pequeñas las pruebas de significación son mucho más importantes. En tales casos es posible que digamos mucho cuando podemos establecer significación. El nivel de significación depende de dos factores, a saber: de la fuerza

o grado de la relación y del tamaño de las muestras. Puede obtenerse significación con una relación muy fuerte y muestras muy pequeñas o, inversamente, con una relación muy débil y muestras muy grandes. En la mayor parte de la investigación social, nuestro interés primordial está no tanto en hallar variables relacionadas unas con otras, sino en localizar relaciones importantes. Aunque conviene recalcar que no todas las relaciones fuertes son importantes (*v.gr.* la relación entre las edades respectivas del marido y la mujer), para que una relación sea de alguna importancia práctica ha de ser por lo menos moderadamente fuerte. Una vez que ha sido establecida la existencia de una relación, el investigador debería preguntarse siempre, "¿cuán fuerte es?"

¿Cómo se mide, pues, la fuerza de una relación? Estamos buscando una medida descriptiva que nos ayude a resumir la relación de tal modo que podamos comparar varias relaciones y llegar a una conclusión respecto de cuál sea la más fuerte. Desde el punto de vista ideal, nos gustaría tener alguna clase de interpretación operativa de la medida que nos atrae intuitivamente. Por convención, los estadígrafos han adoptado la costumbre de concebir medidas que tengan la unidad por límite superior, y cero o bien menos uno (-1.0) como límite inferior. Muchas relaciones sólo pueden alcanzar su límite de 1.0 (o -1.0) cuando la relación es perfecta, y adoptan el valor de cero cuando entre las variables no existe relación alguna, o sea cuando son independientes. Vamos a examinar a continuación algunas medidas que pueden utilizarse con las tablas de contingencia, procediendo a apreciar sus propiedades.

Antes de entrar en el examen de varias medidas de asociación que pueden emplearse con las tablas de contingencia, habría que mencionar, por lo menos, el procedimiento relativamente sencillo y obvio de indicar diferencias en términos de porcentajes. Es posible, sin la menor duda, obtener una indicación muy buena del grado de relación entre dos variables dicotómicas comparando porcentajes. Así, por ejemplo, si el 60 por ciento de los varones seleccionados se clasifican como altamente tolerantes, en tanto que sólo se pone en tal categoría el 30 por ciento de las mujeres, tenemos una diferencia del 30 por ciento entre los dos grupos. ¿Por qué, pues, no servirnos de una medida semejante como medida de la fuerza de la relación? Si comparamos individuos de las clases media e inferior, por ejemplo, desde el punto de vista de la tolerancia, y sólo obtenemos una diferencia del 20 por ciento, podemos afirmar una relación más fuerte entre el sexo y la tolerancia que entre ésta y la clase.

En el caso especial de la tabla de 2×2, los porcentajes pueden efectivamente compararse en tal forma, y la extensa familiarización con los porcentajes, en contraste con otros tipos de

medidas, hablaría ciertamente en favor de estas comparaciones.[7] Pero, ¿qué pasará con la tabla general de $r \times c$? Aquí el uso de los porcentajes puede dificultarle al lector apreciar a primera vista cuán fuerte sea la relación. Supóngase, por ejemplo, que se utilizaban tres clases con los siguientes resultados: clase superior, 70 por ciento altamente tolerante; clase media, 50 por ciento altamente tolerante, y clase inferior, 30 por ciento altamente tolerante. Tenemos ahora una distancia del 40 por ciento entre las clases superior e inferior, o sea una diferencia numéricamente mayor que la que existe entre los varones y las mujeres. Por otra parte, por lo regular esperaremos una diferencia mayor cuando sólo se consideran los extremos. Supóngase que se hubieran tenido cinco clases, ¿qué clase de diferencias de porcentajes esperaríamos ahora, y cómo compararíamos los resultados con los de la tabla de 2×2? Y para introducir una idea más, supóngase que nos sirviéramos de cuatro categorías de tolerancia. Es obvio que se hace difícil establecer comparaciones de una tabla a otra. Necesitamos, pues, una medida única de resumen, que tenga los mismos límites superior e inferior, independientemente del número de casillas.

Medidas tradicionales basadas en la χ-cuadrada. Ya se observó que la χ-cuadrada es directamente proporcional a N. Podemos servirnos de este hecho para construir varias medidas de asociación. En el caso de las dos tablas de contingencia

30	20	50		60	40	100
20	30	50	y	40	60	100
50	50	100		100	100	200

deseamos una medida que tenga el mismo valor para cada una de las tablas, ya que, cuando expresamos los resultados en términos de porcentajes, éstos son los mismos en ambos casos. En otros términos: diríamos probablemente que los grados o fuerzas de la relación son idénticos en los dos grupos de datos, y que la única diferencia está en la magnitud de las muestras. Aunque el valor de la χ-cuadrada sea el doble en la segunda tabla de lo que es en la primera, observamos, con todo, que, si se la divide en cada caso entre el número total de los casos, los resultados son idénticos. Esto sugiere que la expresión χ^2/N o algún múltiplo de la misma nos daría una de las propiedades que buscamos en nuestra medida, o sea la de dar el mismo resultado cuando las proporciones de casillas comparables son idénticas.

[7] Veremos otra ventaja de los porcentajes cuando estudiemos declives en el capítulo XVII. Como ya se indicó en el caso de las pruebas para diferencias de diferencias en proporciones, una diferencia de proporciones puede ser considerada como un caso especial de declive.

Obsérvese que el valor de χ^2/N, o ϕ^2 según se la escribe comúnmente, es 0 cuando entre las variables no existe relación en absoluto. Resulta que, en el caso de tablas de 2×2 (o $2 \times k$), ϕ^2 tiene también la unidad por límite superior cuando la relación entre las dos variables es perfecta. Supóngase, en efecto, que hubiéramos obtenido la siguiente tabla:

50	0	50
0	50	50
50	50	100

Puede verificarse fácilmente que, en este caso, la χ-cuadrada es 100 y, por consiguiente, ϕ^2 es 100/100, o sea 1.0. Ocurrirá siempre que, cuando dos casillas opuestas diagonalmente sean *ambas* cero, el valor de la χ-cuadrada en una tabla de 2×2 sería N, y por lo tanto ϕ^2 será la unidad. Es obvio que, en el ejemplo considerado, la relación es perfecta. Si el sexo se relacionara en él con la tolerancia, podríamos decir que todos los varones son altamente tolerantes y todas las mujeres altamente intolerantes. En una terminología con la que no habremos de tardar en familiarizarnos, podemos decir que el todo de la variación en materia de tolerancia se explica por el sexo o está asociado con él.[8]

En la tabla general de $r \times c$, ϕ^2 puede alcanzar un valor considerablemente mayor que la unidad. Por lo tanto, se han desarrollado diversas otras medidas que son asimismo simples funciones de χ^2/N, pero que tienen también como límite superior la unidad. La primera de éstas, designada como la T de Tschuprov, se define como:

$$T^2 = \frac{\chi^2}{N\sqrt{(r-1)(c-1)}} = \frac{\phi^2}{\sqrt{(r-1)(c-1)}}$$

Aunque el límite superior de T sea la unidad, este límite sólo puede alcanzarse cuando los números de hileras y columnas son iguales. En otros términos: T ha de ser siempre menor que la unidad en una tabla de 2×3 o de 3×5. Si hay considerablemente más hileras que columnas (o viceversa), el límite superior de T puede quedar muy por debajo de la unidad. Para corregir este hecho, podemos siempre dividir el valor obtenido de T entre la máxima T posible para números dados de hileras y columnas. Sin embargo, como quiera que disponemos de medidas más satisfactorias, no necesitamos examinar este procedimiento de corrección.

[8] Esto supone, por descontado, que la tolerancia se toma como variable dicotómica.

* Podemos mostrar que el límite superior de ϕ^2 es Min $(r-1, c-1)$, utilizando la fórmula:

$$\chi^2 = N\left[\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{N_{ij}^2}{N_{i.}N_{.j}} - 1\right]$$

Obsérvese que:

$$\frac{N_{ij}^2}{N_{i.}N_{.j}} \leq \frac{N_{ij}}{N_{i.}} \quad \text{para} \quad i = 1, 2, \ldots, r$$

y

$$\frac{N_{ij}^2}{N_{i.}N_{.j}} \leq \frac{N_{ij}}{N_{.j}} \quad \text{para} \quad j = 1, 2, \ldots, c$$

Por tanto

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{N_{ij}^2}{N_{i.}N_{.j}} \leq \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{N_{ij}}{N_{i.}} = \sum_{i=1}^{r} 1 = r$$

y

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{N_{ij}^2}{N_{i.}N_{.j}} \leq \sum_{j=1}^{c}\sum_{i=1}^{r}\frac{N_{ij}}{N_{.j}} = \sum_{j=1}^{c} 1 = c$$

Así:

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{N_{ij}^2}{N_{i.}N_{.j}} \leq \text{Min}\,(r, c)$$

y de allí:

$$\chi^2 \leq N[\text{Min}\,(r, c) - 1] = N[\text{Min}\,(r-1, c-1)]$$

Por tanto:

$$\phi^2 \leq \text{Min}\,(r-1, c-1)$$

Hay otra medida, introducida por Cramér y que designaremos como V, que se define como sigue:

$$V^2 = \frac{\chi^2}{N\,\text{Min}\,(r-1, c-1)} = \frac{\phi^2}{\text{Min}\,(r-1, c-1)}$$

en donde Min$(r-1, c-1)$ designa $r-1$ o $c-1$, según cuál de ellas sea menor (valor mínimo de $r-1$ y $c-1$). Si bien V no se utiliza corrientemente en la bibliografía social, con todo parece ser preferible a T, en cuanto puede alcanzar la unidad aun cuando los números de hileras y columnas no sean iguales. Como puede

verificarse fácilmente, V y T son equivalentes siempre que $r = c$. De otra forma, siempre será V algo mayor que T. Por supuesto, ambas medidas son equivalentes de ϕ en el caso de 2×2. Y vemos también que V y ϕ serán idénticas en el caso de $2 \times k$.

Otra medida de asociación basada en la χ-cuadrada es el coeficiente de contingencia de Pearson, C, que está dado por:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

Al igual que las otras medidas, C se hace cero cuando las variables son independientes. Sin embargo, el límite superior de C depende del número de hileras y columnas. En el caso de 2×2, el límite superior de C^2 se convierte en $N/(N+N)$, ya que χ^2 puede alcanzar un valor máximo de N. Por lo tanto, el límite superior de C es .707. Si bien el límite superior aumenta a medida que aumenta el número de hileras y columnas, dicho límite siempre es menor que la unidad. De ahí que C sea algo más difícil de interpretar que las otras medidas, a menos que se introduzca una corrección dividiendo entre el valor máximo de C para números particulares de hileras y columnas. En el caso de la tabla 2×2, por ejemplo, la C obtenida habría de dividirse entre .707.

Las medidas anteriores de la fuerza de la relación se basan todas ellas en la χ-cuadrada. Como quiera que por lo regular el valor de la χ-cuadrada se habrá calculado previamente con objeto de verificar el significado, todas las medidas en cuestión requieren en realidad muy poco cálculo adicional. Pero por otra parte, no existe razón particular alguna en cuya virtud una medida de asociación haya de basarse en la estadística de la prueba correspondiente. En efecto, puede demostrarse que todas las medidas basadas en la χ-cuadrada son algo arbitrarias en su esencia y sus interpretaciones dejan mucho que desear. Así, por ejemplo, todas ellas confieren mayor peso a las columnas o hileras de marginales más pequeños que a las de marginales mayores [2]. Sin embargo, como quiera que tanto la prueba T como la C se encuentran con frecuencia en la bibliografía, el lector debería familiarizarse con sus propiedades.

La Q de Yule. Otra medida de uso corriente es la Q de Yule, que es también un caso especial de la medida γ (gamma) que se discutirá en el capítulo XVIII en relación con las escalas ordinales. Esta medida sólo puede emplearse con la tabla de 2×2 y se define como sigue:

$$Q = \frac{ad - bc}{ad + bc}$$

en donde a, b, c y d se refieren a las frecuencias de las casillas. Obsérvese que, una vez elevado al cuadrado y multiplicado por N, el numerador es el mismo que en la expresión de la χ-cuadrada. Lo mismo que en el caso de las demás medidas, Q desaparece cuando las variables son independientes, o sea, cuando los productos diagonales ad y bc son iguales. A diferencia de ϕ^2, sin embargo, Q alcanza sus límites de $\pm$ 1.0 cuando *una* cualquiera de las casillas es igual a cero. Con objeto de comprender el carácter de las circunstancias en cuya virtud Q pueda ser igual a la unidad en tanto que ϕ^2 queda por debajo de dicho valor, tomemos los siguientes ejemplos:

30	0	30	40	0	40
20	50	70	10	50	60
50	50	100	50	50	100

Mientras Q adopta el valor de la unidad en estas dos tablas, los valores correspondientes de ϕ^2, en cambio, son de .429 y .667 respectivamente. En ambos casos sería imposible que desaparecieran *dos* casillas diagonalmente opuestas, debido al carácter de los marginales. De ahí que ϕ^2 sólo pueda adoptar el valor de uno cuando se verifican determinadas condiciones en relación con los marginales. En la tabla de 2 × 2, los marginales de la primera variable han de ser idénticos a los de la segunda.[9] Cuanto mayor sea, pues, la discrepancia entre los marginales de las hileras y las columnas, tanto menor es el límite superior de ϕ^2.

Plantéase ahora la cuestión de saber si queremos o no considerar una relación como "perfecta" cuando sólo desaparece una de las casillas. Al parecer, la respuesta a esta cuestión debería depender, entre otras cosas, de la manera como estén formadas las categorías de las dos variables. Por lo regular es posible concebir un problema en términos de una variable independiente y una variable dependiente. Parecería, pues, razonable sostener que, para que una relación sea perfecta, los marginales de la variable dependiente habrían de "convenir" naturalmente a los de la variable independiente. Supóngase, por ejemplo, que hubiera 60 protestantes y sólo 40 católicos y judíos. En este caso, para que la relación fuera perfecta, esperaríamos que todos los 60 protestantes votaran republicano y todos los 40 restantes votaran en favor de los demócratas. Los marginales serían así los mismos para ambas variables, y tanto ϕ^2 como Q serían iguales a la unidad. Por otra parte, si la mitad de la muestra votara republicano y la otra mitad demócrata, entonces, aunque todos los votos republicanos provinieran de los protestantes, no podríamos decir que la relación era perfecta, ya que 10 de los protestantes habrían votado demócrata. En tal caso, los marginales de la variable dependiente no coincidirían con los de la independiente, y ϕ^2 sería inferior a la unidad. Por lo tanto, en tal caso ϕ^2 parecería ser la medida más apropiada, ya que Q tomaría el valor de la unidad a pesar de la relación imperfecta entre las dos variables.

[9] Esto no significa que los marginales hayan de comportar una partición de 50-50. Significa, en efecto, que si uno de los marginales se parte en 70 y 30, el otro ha de estar también partido de 70 y 30. Las correcciones de marginales desiguales son asimismo posibles, pero, como se desprende del examen que sigue, habrá que ser cauto en el empleo de tales correcciones.

Ocurre en ocasiones que los marginales de la variable dependiente son fijos, en virtud del método empleado al establecer las categorías. Así, por ejemplo, si la variable dependiente fuera en realidad continua pero se hubiera hecho dicotómica en la mediana, entonces los dos grupos de marginales no podrían ser idénticos, a no ser que los marginales de las variables independientes estuvieran también partidos en 50 y 50. Por ejemplo: si la preferencia confesional se hubiera referido a las marcas del conservadurismo político dividiendo en dos a la mediana, entonces ϕ^2 no podría alcanzar la unidad (en el supuesto de la misma partición confesional anterior). En tal caso, Q podría resultar una medida más apropiada, ya que tiene en cuenta el hecho de que los marginales de la variable dependiente se han fijado por completo en virtud del método de investigación.

La tau de Goodman y Kruskal. Cierto número de otras medidas de asociación susceptibles de emplearse con las tablas de contingencia han sido presentadas por Goodman y Kruskal [5], [6] y [7]. La mayoría de ellas comportan lo que se ha designado como interpretaciones probabilistas. Como quiera que tienen un sentido intuitivo que permite interpretar valores intermedios entre cero y uno, estas medidas podrán parecer superiores a las que se basan en la χ-cuadrada.

Con objeto de ilustrar una de estas medidas, la τ_b (tau), tomamos un ejemplo numérico. Designaremos las escalas nominales relacionadas una con otra como A y B, y tomaremos a B como variable dependiente.

	B_1	B_2	B_3	*Total*
A_1	300	600	300	1 200
A_2	600	100	100	800
Total	900	700	400	2 000

Supongamos ahora que se nos da una muestra (o población) de 2 000 personas y se nos pide clasificarlas en una de las tres categorías B_1, B_2 o B_3, de tal modo que terminemos exacta-

mente con 900 casos en B_1, 700 en B_2 y 400 en B_3. Supóngase primero que no sabemos nada acerca de los individuos que nos van a ayudar en esta tarea. Si los individuos nos son dados en un orden totalmente al azar, podemos calcular muy fácilmente el número de errores que podemos esperar cometer al asignar los individuos a una de las tres categorías en cuestión.

Como quiera que hemos de asignar 900 individuos a B_1, en tanto que 1 100 de cada 2 000 no corresponden en realidad a dicha clase, podemos esperar cometer a la larga 900(1 100/2 000), o sean 495 errores. En forma análoga, hemos de asignar 700 individuos a B_2, en tanto que de cada 2 000 los 1 300 no corresponden a ella. De ahí, pues, que al colocar a los individuos en B_2 podamos esperar cometer 700(1 300/2 000), o sea 455 errores. En otros términos, de los 700 que ponemos en dicha categoría sólo podemos esperar que se clasifiquen correctamente $700 - 455$, o sean 245 individuos. Por supuesto, no esperamos cometer exactamente 455 errores, pero ésta es, con todo, la cifra que obtendríamos si promediáramos nuestros errores a la larga. Finalmente, esperaríamos cometer 400(1 600/2 000) o 320 errores al asignar los individuos a B_3. Obsérvese que, pese a que hagamos a esta categoría una asignación menor, nuestro riesgo de error es superior al de las dos categorías precedentes, ya que sólo el 20 por ciento de los individuos corresponde a ella. Por lo tanto, en conjunto, al colocar los 2 000 individuos, esperaríamos cometer:

$$495 + 455 + 320 = 1\,270$$

errores. Nuestro promedio no sería muy bueno.

Pero supóngase ahora que se nos proporcionaba alguna información adicional acerca de cada individuo, diciéndonos si está en A_1 o en A_2. Y nos preguntamos si el hecho de conocer las clases A nos ayudará a reducir el número de errores cometidos al asignar los individuos a las categorías B. Si las variables A y B son estadísticamente independientes, sabemos que el conocimiento de A no nos ayudará a predecir B. En este caso, pues, esperaremos cometer exactamente los mismos errores en que incurrimos cuando no poseíamos información alguna acerca de A. Por otra parte, si la relación entre A y B fuera perfecta, estaríamos en condiciones de anticipar B con perfecta precisión conociendo A. La medida que vamos a desarrollar nos indica la reducción proporcional de errores siendo A conocida.

Veamos cómo calculamos el número de errores anticipados conociendo A. Si se nos da el hecho de que el individuo corresponde a la A_1, podemos servirnos de las cifras de la primera columna. Hemos de poner ahora exactamente 300 de los 1 200 individuos en B_1, los 600 restantes proviniendo de A_2. Ya que de los 1 200 individuos de A_1 900 no corresponden en realidad a B_1, podemos

esperar cometer 300(900/1 200) o 225 errores. Y en forma análoga, con los 600 individuos de A_1 que ponemos en B_2 podemos esperar cometer 300 errores, siendo el número de errores correspondiente a $B_3 = 225$. Tomamos ahora los 800 individuos de A_2 y asignamos 600 de ellos a B_1 y 100 de los 200 restantes a cada una de las categorías B_2 y B_3. Al proceder así, podemos esperar cometer 150, 87.5 y 87.5 errores respectivamente. Adicionando las dos cantidades de A_1 y A_2, vemos que podemos esperar cometer un total de 1 075 errores, si A es conocida.

Definimos la medida τ_b como reducción proporcional de errores. Así, pues:

$$\tau_b = \frac{\text{número de errores con } A \text{ desconocida} - \text{número de errores con } A \text{ conocida}}{\text{número de errores con } A \text{ desconocida}}$$

$$\tau_b = \frac{1\,270 - 1\,075}{1\,270} = \frac{195}{1\,270} = .154$$

En otros términos: nos hemos evitado 195 errores del número total esperado de 1 270, y los hemos reducido en un 15.4 por ciento. Si τ_b hubiera resultado ser .50, podríamos dar así la interpretación muy simple de que el conocimiento de A reduciría el número de errores a la mitad, en tanto que un valor de .75 equivaldría a reducir el número de los errores a un cuarto, y así sucesivamente. En el caso de ϕ^2 en cambio, semejante interpretación sencilla no es posible (*véase* [2]). Si hubiéramos querido interpretar las clases B a partir de las A, habríamos designado la medida correspondiente como τ_a. Por lo general, τ_a y τ_b no tendrán los mismos valores numéricos. ¿Por qué?

En el caso del cuadro 2×2 puede demostrarse que $\tau_a = \tau_b = \phi^2$. Esto nos indica que se dan dos tipos de dificultades en la anotación. Obsérvese que algunas de nuestras medidas (C, Q, T y V) vienen indicadas mediante letras latinas, en tanto que otras (ϕ y τ) lo son mediante letras griegas. Si fuéramos consecuentes deberíamos reservar las letras griegas para los parámetros de población calculados mediante muestras estadísticas. Por desgracia, una vez que los signos vienen siendo usados en forma generalizada, resulta difícil normalizar su empleo, y lo mejor que el lector puede hacer es tomar nota de la inconsistencia. Por otra parte, ciertas medidas aparecen elevadas al cuadrado, en tanto otras no lo están. Vemos especialmente en el caso 2×2 que el símbolo τ, no elevado al cuadrado, es equivalente a ϕ^2, el que en este caso es igual a T^2 y V^2. Así, en el caso del cuadro más general puede parecer razonable comparar τ con los otros coeficientes al cuadrado, aunque observando que no serán idénticos. En general

puede esperarse que los valores numéricos de τ sean *menores que* los coeficientes no elevados al cuadrado ϕ, T y V. Si hubiera que pensar en función de ciertas magnitudes absolutas, considerándolas pequeñas, medianas o grandes (por ejemplo: un valor inferior a .3 es "pequeño"), fácilmente podría incurrirse en error a menos que se reconocieran claramente las diferencias entre las medidas.

Lambda. Existe otra medida, lambda (λ) que es muy semejante a τ y que igualmente es asimétrica con respecto a A y B. Tomando a B como la variable dependiente con la que se hacen predicciones, obsérvese que el número esperado de errores se reducirá si se nos permite colocar a *todos* los individuos en la mayor de las categorías B_i (véase ejercicio 5, capítulo IX). En el ejemplo anterior esto habría supuesto colocar los 2 000 casos en B_1 en lugar de limitarnos a 900. Si lo hiciéramos así cometeríamos 1 100 errores, ya que hay un total de 1 100 casos en B_2 y B_3. Obsérvese que éstos son menos errores que los que hicimos en el caso del denominador de τ_b. Supongamos que sabemos la categoría de A a la que pertenece el individuo. Si se nos permite asignar la totalidad de los 1 200 individuos de A_1 a B_2, la fila que contiene el mayor número de individuos A_1, cometeremos solamente $300 + 300 = 600$ errores. De manera análoga, si colocamos a todos los 800 individuos A_2 en la categoría B_1, cometeremos sólo 200 errores. Conociendo, pues, la categoría A, y si se nos permite hacer estas distribuciones menos restrictivas, podremos esperar cometer 800 errores. Formaremos una medida λ_b, de "reducción proporcional en el error", como sigue:

$$\lambda_b = \frac{1\,100 - 800}{1\,100} = .273$$

Vemos que lambda es más fácil de calcular que tau; que supone una reducción no restrictiva de errores, y que en este ejemplo tiene un valor numérico considerablemente mayor que el de tau. Tiene sin embargo la indeseable propiedad de poder dar un valor numérico igual a cero en casos en que todas las demás medidas consideradas no serán cero, y cuando no desearíamos referirnos a las variables como no correlacionadas o estadísticamente independientes. Tal cosa puede ocurrir simplemente porque una de las B marginales sea mucho mayor que el resto, de tal manera que cualquiera que sea la categoría A, la decisión será siempre de colocar todos los individuos (para todo A_i), en la misma categoría B. Si por ejemplo las categorías B_1 y B_2 hubiesen sido combinadas en el anterior ejemplo hipotético, la decisión hubiera sido siempre la de colocar a todos los individuos en la categoría B_1 y B_2 y no en la de B_3, de tal manera que la

resultante λ_b hubiera sido cero. Por la misma razón, aun cuando un simple total marginal (por ejemplo, B_1) no domina al resto, es probable que algunas de las categorías menos numerosas no entren en absoluto en el círculo de lambda. En el ejemplo anterior la decisión nunca resulta en la asignación de individuos a B_3. Si se hubiera contado con una fila más, B_4, también con un número relativamente pequeño de casos, la lambda medida podría haber sido indiferente a la distribución de casos entre B_3 y B_4. Por estas razones se prefiere a tau sobre lambda en aquellos casos en que los totales marginales no son de aproximadamente la misma magnitud.

XV.4. *Control de otras variables*

Hasta aquí el examen de las pruebas de significación y de medidas de asociación sólo han comportado dos variables a la vez. En la mayoría de los problemas prácticos, en cambio, es necesario controlar una o más variables adicionales, que pueden ya sea enturbiar una relación o crear una relación espuria. Si bien es a menudo cierto que las generalizaciones en materia de ciencias sociales suelen establecerse en términos de sólo dos variables, se supone con todo casi siempre, implícitamente, que las variables relevantes se consideran como controladas. Con objeto de subrayar este hecho se emplea a menudo la frase "en igualdad de condiciones". Desde el punto de vista ideal, una hipótesis habría de enunciarse en forma que se entienda claramente cuáles variables han de controlarse. A medida que una disciplina va progresando hacia su madurez, las generalizaciones se hacen más calificadas, indicando las condiciones exactas en las que puede esperarse que se realicen. En las etapas iniciales de su desarrollo, sin embargo, resulta a menudo imposible saber cuáles son las variables relevantes que se necesita controlar. Ésta es la razón de que en ciencias sociales las proposiciones no se enuncien a menudo en forma que sugieran cuáles variables deban controlarse. No obstante, el lector debería acostumbrarse a buscar siempre las variables eventualmente posibles de controlar, aunque no se le haya invitado expresamente a hacerlo.

Según veremos más adelante, hay varios métodos posibles de control estadístico. El que se examina en el presente capítulo es tal vez el más directo y el que más se parece al experimento de laboratorio, en el que las variables de control se mantienen efectivamente constantes por medios físicos. En los experimentos de laboratorio se mantiene una variable de control a un valor constante, en tanto que las otras variables se relacionan entre sí. Así, por ejemplo, mientras se examina la relación entre la presión y el volumen, la temperatura se mantiene acaso a 70° F. Y si se encuentra una relación entre estas variables, puede resultar

posible enunciar su carácter con mucha mayor precisión que si la temperatura no se hubiera controlado. Sin embargo, el científico no estará autorizado a enunciar una generalización como de realización constante, a menos que la misma relación se verifique exactamente para todas las temperaturas. Realizará, sin duda, toda una serie de experimentos, cada uno de ellos a una temperatura diferente. Es muy probable que encuentre que la relación en cuestión sólo tiene lugar dentro de cierto margen de temperaturas. En estas condiciones habrá de especificar su generalización de modo que diga: "La relación entre la presión y el volumen es tal y cual, *a condición* que la temperatura se mantenga entre −100 y 600° F." Con suerte podrá hallar un factor de corrección que le permita enunciar de nuevo su proposición en forma que se aplique a un margen mayor de temperaturas. Y exactamente el mismo tipo de razonamiento se aplicará al control de variables adicionales. Podrían efectuarse controles simultáneos de diversas variables, manteniendo cada una de ellas a un valor fijo, y efectuando luego experimentos ulteriores con distintas *combinaciones* de valores de las variables de control. Si varios controles actuaran simultáneamente, se requerirá un número mucho mayor de experimentos análogos.

Existe cierta semejanza entre el procedimiento para lograr el control estadístico, que vamos a examinar a continuación, y un experimento de laboratorio en el curso del cual las variables son manipuladas físicamente y mantenidas constantes en diferentes niveles. Existe sin embargo una diferencia fundamental, que resulta vital, relacionada con la forma en que el observador *interpreta* los resultados. Cuando controlamos estadísticamente, llevamos a cabo manipulaciones con lápiz y papel, en el curso de las cuales ajustamos puntuaciones, o hacemos pasar a los individuos de uno a otro cuadro, pero en realidad no estamos manejando sus puntuaciones reales. Cuando, por ejemplo, estamos "controlando" estadísticamente un IQ, esto no significa que manejemos las constantes de inteligencia del individuo afectado. Podemos ajustar las puntuaciones de los IQ, restando de unas y sumando a otras, de manera que podamos pretender que son iguales entre sí, pero no podremos manipular la inteligencia real de una persona en forma que pueda compararse con los controles que gobiernan la temperatura o la presión en un experimento de laboratorio.

Este tipo de control y ajuste hipotético es muy conveniente, y no deberemos desconcertarnos *si* el mundo real coincide con lo que estamos haciendo. Si un cambio real en la inteligencia pudiera afectar nuestra relación en un sentido determinado, pero al mantenerla constante en un experimento nos fuera posible deducir la relación verdadera entre otras dos variables "con la inteligencia mantenida en nivel constante", resultarían justificadas nuestras manipulaciones con papel y lápiz. Debe reconocerse claramente que tales "controles" a base de lápiz y papel pueden ser realizados sobre *cualquier* variable de la que tengamos medidas (y categorías), incluso aquellas que son causalmente *dependientes* de las variables que estamos estudiando y aquellas que de manera espuria estén relacionadas, por razones extrañas, con alguna variable.

Los controles estadísticos son básicamente mucho más fáciles de realizar que los verdaderos controles, por lo que el margen de flexibilidad para su aplicación razonable es mucho mayor. Se requiere fundamentalmente una *teoría* que justifique la aplicación de tales controles, teoría en la que están implícitos supuestos acerca de la estructura causal del sistema de variables. Aunque el tema escapa al interés de un texto general sobre estadística, resulta necesario formular aquí unas palabras de cautela, ya que muchos malos entendidos, en relación con las operaciones de control estadístico, se han traducido en una aplicación ciega de variables de control sin apoyo en una teoría que lo justifique.

Volviendo al ejemplo de la relación entre las preferencias religiosas y los partidos políticos, se pueden controlar estadísticamente variables tales como el sexo y la clase social. Para mantener constante el sexo pueden, por ejemplo, ser considerados solamente los votantes varones. Si se observa que la relación se da en el caso de los varones y por separado en el de las hembras, podrá decirse que es aplicable al sexo, ya que habremos examinado ambas categorías de la variable "sexo". Es posible sin embargo que se observe la relación en el caso de los varones pero no en el de las hembras; en tales circunstancias habrá que calificar la generalización, volviendo nuestra atención a las causas por las cuales la relación existe para un sexo y no para el otro. Puede verse que el controlar las variables relevantes no sólo nos permite una prueba más rigurosa de una hipótesis, sino que nos suministra una mayor penetración en el caso en que se encuentre que la relación difiere de una categoría de la variable de control a la otra.

Algunas veces será conveniente controlar diversas variables a la vez. Debido a la escasez de los casos, se hace necesario con frecuencia controlar las variables relevantes una por una, perdiéndose, sin embargo, en esta forma cierta cantidad de información. Supóngase, por ejemplo, que se hubiera prescindido del sexo y se hubiera introducido un control en relación con la clase social de los electores. Examinaríamos, pues, cada clase social, para ver si la relación subsistía siempre. En contraste con este procedimiento, pudimos haber controlado simultáneamente desde los puntos de vista de la clase y del sexo, tomando todas las combinaciones posibles de las variables de control (*v.gr.* va-

rón de la clase inferior, mujer de la clase inferior, varón de la clase media, etcétera) y estudiando la relación en cada combinación de las categorías de control. Se concibe que la relación pueda verificarse acaso para todas las combinaciones, con excepción de la correspondiente a las mujeres de la clase inferior. Si esto fuera así, nos veríamos conducidos a investigar las peculiaridades de este subgrupo particular.

Con objeto de ilustrar el proceso, tomemos otro ejemplo concreto. Supóngase que tenemos los siguientes datos correspondientes a escolares: ambiente de la clase, cuota de inteligencia, grado escolar y la aplicación de cada niño. Convendrá resumir los datos en términos de una tabla maestra como la del cuadro XV.4.

CUADRO XV.4. *Cuadro maestro para correlacionar cuatro variables*

Inteligencia	*Grados*	*Clase media*		*Clase baja*		*Totales*
		Aplicación elevada	*Aplicación baja*	*Aplicación elevada*	*Aplicación baja*	
Alta	Alto	60	40	40	18	158
	Bajo	20	24	16	38	98
Baja	Alto	40	24	6	2	72
	Bajo	24	12	32	54	122
Totales		144	100	94	112	450

Obsérvese que un cuadro como éste contiene las casillas suficientes para que los cuatro tipos de información (clase, IQ, grados y aplicación) puedan ser, si así conviene, reconstruidos para cada individuo, es decir, que sabemos cuántas son las personas en las que se da la misma combinación de rasgos (por ejemplo: clase baja, IQ elevado, aplicación baja y grados altos). Si deseamos una información menos detallada podremos combinar los datos formando agrupaciones más amplias. Podemos por ejemplo reunir a los estudiantes de la clase media con los de la clase baja, manteniendo tan sólo la distinción relativa al IQ, la aplicación y los grados. Pero si se nos facilitase tan sólo una información menos detallada no nos sería posible recobrar el total de la información más que volviendo a hacer el análisis. Por tal razón un cuadro maestro tal como el XV.4 debe ser utilizado como cuadro de trabajo, sacando de él los datos para preparar una serie de otros cuadros separados.

Será en general más conveniente hacer el cuadro maestro de tal manera que la variable dependiente aparezca en la columna extrema de la izquierda, en tanto que la variable independiente más interesante aparezca en la fila baja del encabezamiento, lo que se traducirá en subcuadros con las frecuencias que están siendo comparadas directamente. En el cuadro XV.4, por ejemplo, tenemos cuatro subcuadros en cada uno de los cuales se relacionan las aplicaciones y los grados. Todos los individuos del subcuadro de la parte superior izquierda son de la clase media y tienen elevado IQ, y así sucesivamente. La exacta distribución de filas y columnas no tiene una importancia decisiva, ya que es bien claro que se las puede ordenar de acuerdo con la relación de intereses (tal como se hace en el cuadro XV.5).

CUADRO XV.5. *Serie de tablas de contingencia que relacionan dos variables con dos controles simultáneos*

Grados	*Aplicación elevada*		*Aplicación baja*	
	IQ alto	*IQ bajo*	*IQ alto*	*IQ bajo*
	Clase media			
Alto	60	40	40	24
Bajo	20	24	24	12
	Clase baja			
Alto	40	6	18	2
Bajo	16	32	38	54

Supóngase que sospechamos una propensión de los maestros en favor de la clase media, que se traduciría en la tendencia a dar buenas notas a los niños de la clase media, independientemente de su capacidad y aplicación, y buenas notas a los niños de la clase inferior solamente cuando muestran capacidad y aplicación a la vez. Anticiparíamos, en tal caso, que las notas habrían de ser por lo regular mejores para los niños de la clase media, controlando la inteligencia y el esfuerzo a la vez, excepto, posiblemente, en el caso de niños de gran capacidad y aplicación. Anticiparíamos asimismo que las relaciones entre las notas por una parte y la capacidad y la aplicación por la otra habrían de ser más fuertes en la clase inferior que en la media. En otros términos, si los niños de la clase media reciben siempre buenas notas, no debería haber relación (o sólo una relación superficial), en esta clase, entre las notas por una parte y la capacidad o la aplicación por la otra. Fijémonos en la relación entre las notas y la capacidad y averigüemos si es o no más fuerte en la clase

inferior. En este caso necesitaremos controlar el esfuerzo. En ambas clases habrá estudiantes aplicados y no tan aplicados. Por lo tanto, podemos construir cuatro tablas de contingencia como las del cuadro XV.5.

Comparamos ahora las dos clases con respecto a la existencia y la fuerza de la relación, considerando separadamente a los alumnos de aplicación elevada y baja respectivamente. La dirección de la relación puede también observarse en cada caso, ya sea calculando los porcentajes o comparando los productos diagonales. Calculando la χ-cuadrada y la ϕ para cada tabla, obtenemos los resultados del cuadro XV.6. Vemos en esta forma que las relaciones no son significativas por lo que se refiere a los niños de la clase media, pero que en cuanto a los de la clase inferior, en cambio, existe una relación positiva moderadamente fuerte en ambas categorías de aplicación entre la capacidad y las notas. Observamos asimismo que la relación es algo más fuerte en el caso de los estudiantes más aplicados.

CUADRO XV.6

Clase	*Aplicación*	*χ-cuadrada*	*Nivel de significación*	*ϕ*
Media	Alta	2.565	no significativa	.133
	Baja	.188	no significativa	.043
Baja	Alta	28.064	$p < .001$	.546
	Baja	15.582	$p < .001$	.373

El lector habrá sin duda observado el efecto pronunciado del control sobre el número de casos que figuran en cada casilla. En lugar de tener sólo cuatro casillas, en efecto, tenemos cuatro veces dicho número al servirnos de dos variables de control dicotómicas. Si se hubiera añadido un tercer control simultáneo, por ejemplo, el sexo, habríamos tenido 32 casillas en lugar de 16. Y si cualquiera de las variables hubiera comportado más de dos categorías, el número de las casillas habría aumentado. Así, pues, si bien los controles simultáneos pueden en teoría añadirse indefinidamente, el número de casos ha de ser muy grande para controlarse con este método. Una alternativa consistiría en reducir simplemente el carácter de la población y generalizar sólo respecto de los varones de la clase media de educación universitaria, o de algún otro subgrupo correspondiente. Podría seleccionarse luego una muestra mucho mayor de este subgrupo. Por lo general, si se ha de emplear el control simultáneo, resulta necesario seleccionar aquellos dos o tres controles que se presentan como más prometedores. Es posible, por supuesto, servirse de la prueba exacta de Fisher cuando el número de casos de cada casilla se hace muy pequeño; pero hay que recordar que será en tal caso necesario tener un alto grado de relación para obtener significación. Debido a esta atenuación de los casos, el mero hecho de que una relación se haga no significativa al introducir controles no constituye una prueba suficiente de que la variable de control produce efecto. Habría que calcular y comparar siempre medidas del grado de relación.

En los casos en que difieran las *relaciones* entre una categoría de una variable de control y la siguiente, tendremos un ejemplo de lo que se denomina *no aditividad* o *interacción estadística.* Ya se examinó esta posibilidad al tratar de la prueba para una diferencia de diferencias en las proporciones, y volveremos al tema de manera más detallada en los capítulos XVI y XX. Siempre que se sospeche la posibilidad de una interacción, *deberá* hacerse una prueba estadística que la localice, antes de seguir adelante. Como inevitablemente habrá algunas diferencias leves en las relaciones entre una muestra y la siguiente, la pregunta básica por formular en tales pruebas será la de si las muestras de interacción son lo suficientemente grandes como para que aquélla haya ocurrido por casualidad, incluso en ausencia de interacción entre la población. En este ejemplo, y dado el caso de que todas las variables han sido dicotomizadas, podrá hacerse una prueba sencilla de una diferencia de diferencias en proporciones, tal como sugiere el capítulo XIII. Como están siendo consideradas simultáneamente dos variables de control, puede incluso darse el caso de que se produzca lo que se denomina una interacción de segundo orden, o una diferencia de diferencias de diferencias. Por ejemplo: la diferencia entre las relaciones de aplicación elevada y aplicación baja puede ser mayor entre los niños de la clase baja que entre los de clase media.

Si se observa que la interacción tiene significación estadística, y es además lo bastante grande como para tener significación sustantiva, resultará necesario cualificar las generalizaciones haciendo una referencia específica a la categoría de control. Habría que decir, por ejemplo: "Se encontró una relación entre grados y habilidad en el caso de los niños de clase baja, pero no en los de clase media." A partir de dicho punto deberán estudiarse separadamente las restantes relaciones entre los dos niveles de clase. Si la interacción es por el contrario estadísticamente insignificante, o tan pequeña que pueda ser ignorada, aun siendo estadísticamente significativa, podrá deducirse razonablemente que las relaciones son básicamente similares entre las categorías de control. Estaremos en tal caso en la posibilidad de simplificar considerablemente el análisis, reuniendo los resultados separados. Veamos a continuación qué tipos específicos de simplificación resultan posibles en el caso de datos categorizados.

Podemos en primer lugar reunir las pruebas de chi al cuadrado en una sola prueba global, a condición de que aquéllas estén basadas en muestras al azar seleccionadas independientemente. El procedimiento es extremadamente sencillo, bastando sumar los distintos valores de chi al cuadrado y también los grados de libertad, evaluando el resultado de la manera habitual. Supongamos por ejemplo que en el caso de cuatro cuadros 2 × 2, las chi cuadradas resultantes fueron 2.1, 3.3, 2.7 y 2.9. La suma de estos valores es 11.0, y la de los grados de libertad, 4. En el cuadro vemos que una chi cuadrada de 11.0, con 4 grados de libertad resulta significativa al nivel de .05. Así, aun cuando ninguno de los valores separados de chi al cuadrado fuera significativo, podemos hacer uso del hecho de que el reunir los resultados tiene significación teórica. Estamos en efecto diciendo que si una relación se repite aproximadamente cada vez, pero la probabilidad de los resultados separados es en cada caso mayor de .05, podremos preguntarnos cuál sería el resultado de tal combinación de resultados si no hubiese relación en cualquiera de los cuatro cuadros.

Obsérvese que los resultados de semejante operación de reunión podrían muy bien diferir de la relación total entre dos variables *sin control alguno*. Al juntar los resultados, obtenemos esencialmente una relación *promedia dentro* de las categorías de la variable o las variables de control. Si hubiéramos prescindido simplemente de la variable o las variables de control, los efectos de semejantes controles habrían permanecido oscuros por completo. En tanto que, al unificar, efectuamos una sola prueba de χ-cuadrada de la relación conjunta entre dos variables, controlando en relación con las variables adicionales.

Y en forma análoga, podríamos desear obtener una sola medida de asociación calculando un promedio ponderado de las medidas basado en las cuatro tablas separadas. Un método que se ha sugerido para tal objeto consiste en el empleo de ponderaciones que sean proporcionales al número de los casos de cada tabla. Así, por ejemplo, podríamos multiplicar cada τ_b por el número de casos de la tabla, sumar los resultados y dividir, finalmente, entre el número total de casos de las cuatro tablas. Terminaríamos así con una sola prueba de significación y una sola medida de asociación que representarían un promedio de los resultados de las cuatro tablas.

Otro simple procedimiento para obtener una media ponderada es el que describiremos brevemente. (Para mayores detalles véase Rosenberg [12].) El procedimiento consiste, básicamente, en estandarizar todas las categorías de control, mediante la obtención de un promedio ponderado de proporciones (o porcentajes). Supongamos haber obtenido separadamente los resultados siguientes, para hombres y mujeres:

	Varones				*Hembras*			
	Protestantes	*Católicos*	*Judíos*	*Total*	*Protestantes*	*Católicas*	*Judías*	*Total*
Republicanos	180	80	20	280	100	50	10	160
Demócratas	90	80	50	220	60	30	70	160
Independientes	30	40	30	100	40	20	20	80
Total	300	200	100	600	200	100	100	400

Comenzaremos por transformar las cifras anteriores en proporciones, totalizando a 1.00, ya que la variable independiente aparece en la parte alta de cada cuadro. Los resultados serán los siguientes:

	Varones			*Hembras*		
	Protestantes	*Católicos*	*Judíos*	*Protestantes*	*Católicas*	*Judías*
Republicanos	.60	.40	.20	.50	.50	.10
Demócratas	.30	.40	.50	.30	.30	.70
Independientes	.10	.20	.30	.20	.20	.20
Total	1.00	1.00	1.00	1.00	1.00	1.00

Aceptando que deseamos oscurecer las diferencias entre estos dos cuadros, utilizando para ello un promediado, podremos formar un promedio ponderado, multiplicando cada proporción de las contenidas en el cuadro de varones por .6, ya que son 600 los varones en un total de 1 000 individuos en la muestra. De manera análoga podemos ponderar cada cifra en el cuadro de las hembras, multiplicándola por .4. Los resultados serán los siguientes:

	Protestantes	*Católicos*	*Judíos*
Republicanos	.56	.44	.16
	(.36 + .20)	(.24 + .20)	(.12 + .04)
Demócratas	.30	.36	.58
	(.18 + .12)	(.24 + .12)	(.30 + .28)
Independientes	.14	.20	.26
	(.06 + .08)	(.12 + .08)	(.18 + .08)
Total	1.00	1.00	1.00

en el que cada proporción de las que aparecen en el cuadro derivado es igual a la suma de las dos proporciones ponderadas (como se indica en los paréntesis), que a su vez figuraban en los

cuadros anteriores. Como la suma de las ponderaciones es de 1.0, también lo será la de las proporciones en cada columna del cuadro derivado. Los resultados pueden ser presentados también bajo la forma de porcentajes.

Este procedimiento para controlar mediante la obtención de promedios ponderados es, como se verá, muy generalizado. Hemos estandarizado el número de protestantes, católicos y judíos, de tal manera que sus tamaños relativos en las muestras de varones y de hembras pierdan significación. Si hubiese habido controles simultáneos para variables adicionales, habríamos podido ampliar este procedimiento de manera directa. Así, si hubiéramos deseado controlar según clases sociales, usando tres niveles, habríamos obtenido seis cuadros, uno para cada categoría sexo-clase. Después de haber vigilado si se produce interacción, y habiendo resuelto que ninguna diferencia importante podrá resultar oscurecida por la aplicación del procedimiento, podríamos asignar de nuevo gravámenes W_i a cada uno de los cuadros de control, haciendo $\Sigma W_i = 1.0$, obteniendo así un solo cuadro combinado, como en el ejemplo anterior.

Al sustituir así varias medidas y pruebas separadas por una sola medida y una sola prueba, nos enfrentamos a los problemas que se encuentran siempre que se emplean estadísticas de resumen. Concentramos nuestros datos, de modo que resulten menos estadísticos, pero, por otra parte, corremos el riesgo de distorsionar nuestros resultados. Por ejemplo: si una de las cuatro tablas en cuestión diera una χ-cuadrada grande y un grado de relación muy alto, en comparación con las demás, entonces el combinar los resultados, con lo que dicho hecho resulta oscurecido, puede revelarse como sumamente engañoso. O sea que, como siempre, las manipulaciones estadísticas no pueden constituir nunca un sustituto del sentido común.

Algunas de las ideas examinadas en esta sección, en particular las relativas a la reunión de los resultados de tablas separadas, son indudablemente nuevas y podrán parecer algo confusas de momento. Será útil, por lo tanto, volver a repasar esta sección, una vez que el lector se haya enfrentado al material de los capítulos XVI al XX. En dicho momento, en efecto, se habrán examinado ya y comparado diversos tipos de procedimientos de control.

EJERCICIOS

1. Calcúlese la χ-cuadrada para los datos del ejercicio 5 del capítulo IX. Tomando las aspiraciones profesionales como variable dependiente B, ¿cuál es el valor de τ_b? ¿Cómo se compara el valor de τ_b con el de la medida que se calculó en la parte *d*) del ejercicio 5?

2. En el ejercicio 3 del capítulo XIV nos servimos de la prueba de Smirnov. Tomando los mismos datos, ¿a qué conclusión llegamos al servirnos de la prueba de la χ-cuadrada? En relación con esos datos

particulares, ¿cuál prueba se preferirá? ¿Por qué? Calcúlense ϕ, T, V, C, τ_b y λ_b.

*3. La prueba de la χ-cuadrada puede emplearse en general para comparar frecuencias observadas y teóricas. En particular, puede utilizarse para verificar la hipótesis nula de que los datos de la muestra se han seleccionado al azar de una población normal. Las frecuencias observadas se comparan con las que se habrían anticipado en caso de ser la distribución efectivamente normal, con la misma media y desviación estándar que se han calculado de los datos de la muestra. Una vez obtenidos los valores de $\bar{X}$ y de s, podemos servirnos de los verdaderos límites y de la tabla normal para dar las frecuencias esperadas dentro de cada intervalo. Los grados de libertad serán $k-3$, en donde k representa el número de intervalos. Se perderá un grado de libertad, ya que el total de las frecuencias esperadas ha de ser N; los otros dos grados de libertad que se han perdido se deben a la necesidad de utilizar $\bar{X}$ y s a título de apreciaciones de los parámetros reales μ y σ. Teniendo estos hechos presentes, verifíquese si los siguientes datos se apartan o no significativamente de la normalidad: Respuesta $\chi^2 = 2.53$, sin rechazar.

Intervalo	*Frecuencia*
0.0- 9.9	7
10.0-19.9	24
20.0-29.9	43
30.0-39.9	56
40.0-49.9	38
50.0-59.9	27
60.0-69.9	13
	208

4. En un estudio reciente, H. L. Wilensky [14] encontró, al controlar la condición socioeconómica, una relación general entre la actividad sindical por una parte y la orientación política y la preferencia electoral por la otra. Los datos de 15 miembros negros tendían a apoyar este hallazgo general en relación con la preferencia electoral. Siete de los ocho negros que eran miembros inactivos del sindicato no siguieron la "línea" de éste al votar en 1948, en tanto que, de los siete miembros sindicalmente activos, cinco votaron de acuerdo con la sugerencia del sindicato. Averígüese si se da o no una relación significativa, sirviéndose: *a)* de la prueba exacta de Fisher, con dirección anticipada, y *b)* de la χ-cuadrada corregida con fines de continuidad con dirección anticipada. Respuesta: *a)* $p = .035$; *b)* $\chi^2 = 3.22$, $p < .05$.

5. Utilice los datos que siguen (disponiendo los cuadros en otra forma, si es necesario) para obtener información acerca de la precisión de los enunciados *a), b)* y *c)*. Allí donde sea adecuado, calcúlense medidas del grado de relación y control de las variables relevantes.

a) Las mujeres tienen menos prejuicios que los hombres, independientemente de la religión que profesen o de la clase social a que pertenezcan.

b) Los grados de relación entre la confesión y el prejuicio contra los negros dependerán de la clase social de la persona "afectada de prejuicio".

c) La razón de que los judíos aparezcan como menos afectados de prejuicio, en la tabla, que los no judíos se debe al alto porcentaje de mujeres y de personas de la clase superior en la muestra relativa a los judíos.

Religión	*Sexo*	*Grado del prejuicio contra los negros*				*Totales*
		Elevado		*Bajo*		
		Clase superior	*Clase inferior*	*Clase superior*	*Clase inferior*	
No judíos	Varones	14	30	15	16	75
	Mujeres	8	13	9	7	37
Judíos	Varones	13	7	22	15	57
	Mujeres	18	9	33	21	81
Total						250

6. Utilizando los datos del anterior ejercicio 5, constrúyanse cuadros que relacionen la religión con los prejuicios, con controles simultáneos para sexo y clase social. Suponiendo despreciable la posible interacción, normalícense estos resultados de forma que la relación entre religión y prejuicio, con controles, pueda ser presentada en un solo cuadro 2×2.

*7. Supongamos que se espera llevar a cabo una prueba chi al cuadrado con un cuadro 2×2, en que se relaciona la preferencia religiosa (protestante-católico), con la preferencia política (republicano-demócrata). Se planea tomar muestras al azar, del mismo tamaño, de protestantes y católicos, y se predice la dirección, esperando que la proporción de protestantes que son republicanos resultara de .60 aproximadamente, en tanto que la proporción de católicos que son republicanos será a su vez de .40, más o menos.

¿Cuántos casos resultarán necesarios si se requiere establecer significación al nivel de .05?

Bibliografía

1. Anderson, T. R., y M. Zelditch: *A Basic Course in Statistics*, 2ª ed., Holt, Rinehart and Winston, Inc., Nueva York, 1968, cap. 9.
2. Blalock, H. M.: "Probabilistic Interpretations for the Mean Square Contingency", *Journal of the American Statistical Association*, vol. 53, pp. 102-105, 1958.
3. Bradley, J. V.: *Distribution-free Statistical Tests*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968, cap. 8.
4. Downie, N. M., y R. W. Heath: *Basic Statistical Methods*, 2ª ed., Harper and Row, Publishers, Incorporated, Nueva York, 1965, cap. 14.

5. Goodman, L. A., y W. H. Kruskal: "Measures of Association for Cross Classifications", *Journal of the American Statistical Association*, vol. 49, pp. 732-764, 1954.
6. Goodman, L. A., y W. H. Kruskal: "Measures of Association for Cross Classifications, II: Further Discussion and References", *Journal of the American Statistical Association*, vol. 54, pp. 123-163, 1959.
7. Goodman, L. A., y W. H. Kruskal: "Measures of Association for Cross Classifications, III: Aproximate Sampling Theory", *Journal of American Statistical Association*, vol. 58, pp. 310-364, 1963.
8. Hagood, M. J., y D. O. Price: *Statistics for Sociologist*, Henry Holt and Company, Inc., Nueva York, 1952, cap. 21.
9. Hays, W. L.: *Statistics*, Holt, Rinehart and Winston, Inc., Nueva York, 1963, cap. 17.
10. McCarthy, P. J.: *Introduction to Statistical Reasoning*, McGraw-Hill Book Company, Nueva York, 1957, cap. 11.
11. Mueller, J. H., K. Schuessler, y H. L. Costner: *Statistical Reasoning in Sociology*, 2ª ed. Houghton Mifflin Company, Boston, 1970, cap. 9.
12. Rosenberg, Morris: "Test Factor Standardization as a Method of Interpretation", *Social Forces*, vol. 41, pp. 53-61, 1962.
13. Siegel, Sidney: *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, Nueva York, 1956, pp. 96-111.
14. Wilensky, H. L.: "The Labor Vote: A Local Union's Impact on the Political Conduct of its Members", *Social Forces*, vol. 35, pp. 111-120, 1956.