

15

Análisis de frecuencias

En los capítulos anteriores se han analizado los modelos teóricos sobre muestras, inferencia y decisión estadísticas aplicadas a magnitudes biológicas de tipo continuo en los modelos paramétricos. Luego, se mostraron los modelos no paramétricos para analizar los casos en donde es dudoso el cumplimiento de los supuestos, o bien cuando se trabaja con otro tipo de magnitudes como las cualitativas. Ahora se presentan nuevos modelos estadísticos aptos para magnitudes de tipo discreto, como las de recuento, de uso extendido en Medicina, Bioquímica y Farmacia. En este capítulo se muestran modelos para ensayar hipótesis cuando se trabaja con datos frecuenciales. Por ende, también se aplican a casos en donde, si bien la variable es continua, los datos se han agrupado en clases con sus frecuencias respectivas, o sea, en los hechos se ha “discretizado” a una magnitud de tipo continuo. Lo mismo para el caso de magnitudes cualitativas donde se miden las frecuencias de una clasificación de tipo ordinal o dicotómica. Cualquiera sea el caso, se pueden hacer los supuestos de que las muestras fueron extraídas de cierta población, donde las frecuencias obtenidas representan las proporciones verdaderas de la frecuencia total en la población. Con los modelos teóricos, como la Binomial, Poisson, Gauss, etc. se pueden obtener las frecuencias esperadas. Entonces el problema es comparar las *frecuencias observadas* contra las *esperadas* predichas por la teoría usada. Así se puede comparar una teoría científica contra la realidad en un experimento. Como la coincidencia entre ambas nunca es total, el problema se reduce a poder determinar si las diferencias observadas entre ellas se deben al azar y la teoría es correcta, o de lo contrario se debe rechazar el supuesto realizado. Un supuesto histórico es el de Student (1907) cuando demostró que el error en el conteo celular realizado con un hemocitómetro sigue mejor el modelo Poisson en vez del Normal usado tradicionalmente. Surge así un nuevo concepto, el de *bondad de ajuste*, que se verá mejor en el próximo capítulo junto con los tests de validación respectivos para los casos de ajuste de distribuciones. Aquí se comienza por mostrar la prueba de la Chi-cuadrado aplicado a casos donde los datos se ordenan según un único criterio de clasificación; luego con el G-test se moderniza el uso clásico de la Chi-cuadrado. A continuación se presentan los casos bidimensionales con los llamados *tests de independencia*, en las denominadas *Tablas de Contingencia*. Se profundiza el caso más usual de las tablas de 2 x 2 donde hay dos criterios de clasificación divididos en dos clases cada uno. Los criterios de decisión se basan en la prueba clásica de la Chi-cuadrado, el G-test y el Test exacto de Fisher. En particular, se presenta el concepto del “*odd-ratio*” para efectuar investigaciones clínicas, por su cada vez más difundido uso en Bioquímica y Farmacia, en pruebas de medicamentos y de infecciones por exposición a factores de riesgo, tanto clínicas como epidemiológicas.

15.1 La prueba de la Chi-cuadrado

Por lo general, los resultados de un muestreo casi nunca coinciden con los valores predichos por la teoría o modelo con el cual se está trabajando. Esto es así por las fluctuaciones aleatorias en las mediciones o error de medición. El problema es poder determinar si esas diferencias se deben al azar, o bien no se ajustan al modelo teórico estudiado, en cuyo caso este deberá ser modificado y vuelto a investigar. La *validación estadística* es la herramienta para poder decidir en estos casos. En este punto se verá este problema cuando los datos son de recuentos o de cantidad de casos observados y están agrupados en clases. En los dos casos se está trabajando con *frecuencias*. En la Tabla 15.1 siguiente se muestra el caso general donde ocurren n acontecimientos $A_1, A_2, A_3, \dots, A_n$. Las muestras extraídas de la población en estudio permiten determinar el número de veces en que ocurre cada acontecimiento y sus frecuencias observadas serán: $O_1, O_2, O_3, \dots, O_n$. Por ejemplo, si se está lanzando un dado, habrá 6 acontecimientos posibles, el número de ases observados será O_1 y así sucesivamente. El modelo teórico postulado permite obtener la frecuencia esperada: $E_1, E_2, E_3, \dots, E_n$. En el ejemplo del dado, si se efectuaron 600 lanzamientos, el número esperado de ases será $E_1 = 100$ bajo el postulado de que el dado no está cargado.

Tabla 15.1: Caso general.

Acontecimiento	A_1	A_2	A_3	...	A_n
Frecuencia esperada	E_1	E_2	E_3	...	E_n
Frecuencia observada	O_1	O_2	O_3	...	O_n

La manera clásica de estudiar las diferencias entre frecuencias esperadas y observadas es usar el estadígrafo χ^2 para cuantificar la discrepancia entre ambas. Este se calcula con la ecuación:

$$\chi^2 = [(O_1 - E_1)^2 / E_1] + [(O_2 - E_2)^2 / E_2] + [(O_3 - E_3)^2 / E_3] + \dots + [(O_n - E_n)^2 / E_n]$$

O sea,

$$\chi^2 = \sum_1^n \frac{(O_i - E_i)^2}{E_i}$$

De la fórmula surge que si $\chi^2 = 0$, la frecuencia esperada coincide exactamente con la observada y entonces la teoría predice perfectamente los acontecimientos. Si existe una diferencia cualquiera entre ambas frecuencias, será $\chi^2 > 0$, valor que irá aumentando si no es producto del azar, hasta alcanzar valores significativos que permitan rechazar la hipótesis nula de la igualdad entre la teoría y la realidad.

La distribución muestral del valor χ^2 se aproxima mucho a una del tipo Chi-cuadrado, modelo que se usa para validar hipótesis. Los grados de libertad se calculan para dos casos posibles:

- 1) *Hipótesis Extrínseca*: es una hipótesis externa a los datos. No se necesita de estos para obtener los parámetros poblacionales en el cálculo de las frecuencias observadas. Por ejemplo, el lanzamiento de una moneda o un dado en teoría de juegos, las leyes de Mendel, etc.

$$v = n - 1$$

- 2) *Hipótesis Intrínseca*: es una hipótesis interna a los datos. Se necesitan los datos para sacar los parámetros poblacionales. Por ejemplo, si se trabaja bajo el supuesto de normalidad: la media y la varianza de la población se estiman con los datos muestrales -ver el caso 1 del cuadro 4.2- o sea, habrá $r = 2$ grados de libertad perdidos.

$$v = n - r - 1$$

En el caso de las extrínsecas es $v = n - 1$ porque con las primeras $n - 1$ muestras, se puede determinar la restante, entonces como una de ellas se obtiene a partir de las demás, se pierde solo un grado de libertad. Para ilustrar las ideas anteriores se desarrolla el siguiente caso:

Ejemplo: se lanza una moneda al aire 100 veces y se obtienen 58 caras. Ensayar la hipótesis que la moneda está bien.

(H₀) $p = 0,5$ la moneda esta bien.

(H₁) $p \neq 0,5$ la moneda esta "cargada".

De acuerdo con la hipótesis nula, si la moneda está bien se espera un número de 50 caras:

$$\chi^2 = [(O_1 - E_1)^2 / E_1] + [(O_2 - E_2)^2 / E_2] = [(58 - 50)^2 / 50] + [(42 - 50)^2 / 50] = 2,56 \text{ (ns)}$$

Para hallar el valor crítico se toman: $v = n - 1 = 2 - 1 = 1$ grado de libertad y un nivel de confianza del 95%. Entonces de tablas es: $\chi^2_{0,95; 1} = 3,84$ que resulta ser mayor que el valor de comparación $\chi^2 = 2,56$ (valor no significativo), por lo tanto no se puede rechazar la hipótesis nula. Se dice: no hay prueba suficiente como para pensar que la moneda está mal hecha.

15.2 La prueba de G (G-test)

Esta es otra técnica para comprobar la concordancia entre las frecuencias observadas y las esperadas. El llamado *Test de la Razón de Verosimilitudes* (G - test) se ha empezado a usar recientemente y posee propiedades teóricas que lo hacen preferible a la clásica prueba de la Chi-cuadrado. La fórmula usada para el cálculo de la prueba Chi-cuadrado puede escribirse de la manera siguiente:

$$\chi^2 = \sum (O_i - E_i)^2 / E_i = [\sum (O_i)^2 / E_i] - n$$

Donde el *cociente de frecuencias* es el indicador de las diferencias. Cuando la coincidencia sea perfecta, el cociente será E_i y la suma de todos los casos será n , lo cual anulará el valor de comparación: $\chi^2 = n - n = 0$. A medida que aparezcan diferencias entre las frecuencias, cada uno de

estos irá aumentando el valor total alejándolo del cero hasta llegar a valores significativos desde el punto de vista estadístico. Ahora bien, este cociente de las frecuencias $L_i = O_i / E_i$ (o de las probabilidades, si se dividen ambos miembros por n) llamado cociente de *verosimilitudes*, puede emplearse como un indicador de concordancia entre frecuencias, y así se puede construir un test para validar hipótesis. La distribución teórica de este indicador se simplifica con la transformación del valor **G**, que se calcula con:

$$G = 2 \sum_1^k O_i \ln(O_i / E_i)$$

Donde k es la cantidad de clases o muestras. Este estadígrafo para muestras grandes ($n > 25$) se aproxima bastante a una Chi-cuadrado. Para muestras pequeñas, el valor más exacto se puede lograr a través de la Probabilidad Binomial. El número de grados de libertad, lo mismo que para la prueba de la Chi-cuadrado visto antes, se calcula para dos casos posibles con:

- 1) *Hipótesis Extrínsecas*: $v = k - 1$
- 2) *Hipótesis Intrínsecas*: $v = k - r - 1$

Ejemplo) Se efectúa un cruce genético y el supuesto mendeliano es esperar una proporción de 3:1 de fenotipos en la primera generación filial. Los resultados obtenidos muestran en la prole 178 casos del tipo normal y 22 casos de mutaciones. Para poder testear el grado de concordancia del experimento entre las frecuencias observadas y las esperadas (75% de normales y 25% de mutantes) se puede usar el G-test como sigue:

Fenotipo	Frec. Observada O_i	Frec. Esperada E_i	$O_i \ln(O_i/E_i)$	G
Normales	178	150	30,46	24,8***
Mutantes	22	50	-18,06	
Total	200	200	12,4	

El resultado final $G = 24,8$ debe ser comparado con el valor crítico de tablas, para $v = 1$ grado de libertad. Se obtiene $\chi^2_{0,999; 1} = 10,828$ mucho menor que G. Por lo tanto, se tienen evidencia altamente significativa como para rechazar la hipótesis nula. *No significa que las Leyes de Mendel dejen de cumplirse*, sino que deben buscarse las causas genéticas o experimentales que aclaren lo ocurrido en este caso. Como por ejemplo el concepto de gen ligado, o una falla de método en los cruces, etc. También se puede usar para este mismo caso la prueba de Chi-cuadrado en efecto:

Fenotipo	Frec. Observ. O_i	Frec. Esper. E_i	Diferencias ($O_i - E_i$)	($O_i - E_i$) ²	$\frac{(O_i - E_i)^2}{E_i}$
Normales	178	150	+ 28	784	5,23
Mutantes	22	50	- 28	784	15,68
Total	200	200	0		$\chi^2 = 20,91***$

$\chi^2 = 20,91 \gg \chi^2_{0,999;1} = 10,828$ y se rechaza la hipótesis.

Notar que en ambos casos el valor crítico de tablas es el mismo. Sin embargo, el valor del estadígrafo $G = 24.8 > \chi^2 = 20,91$. Esto significa que G pone más en evidencia la diferencia entre las frecuencias esperadas y observadas, es más eficiente para ello. Se dice que la prueba del G-test es más *sensible* que la prueba de la Chi-cuadrado (χ^2 -test) para detectar diferencias significativas. Viceversa, cuando la hipótesis nula no pueda ser rechazada, es de esperar que el valor G sea de menor valor que el χ^2 para el mismo caso. Se dice que el G-test es más *conservador* (le cuesta más rechazar una hipótesis) que a la prueba de la Chi-cuadrado (ver ejemplo 2).

Se puede generalizar a más casos como el siguiente:

Ejemplo 2) En los experimentos de Mendel con porotos se cosecharon 315 porotos lisos y amarillos, 108 lisos y verdes, 101 rugosos y amarillos y 32 rugosos y verdes. La proporción esperada para este caso es 9 : 3 : 3 : 1. Decidir si los datos observados se ajustan a la teoría mendeliana.

En este caso, el investigador puede considerar que tiene cuatro tipos de porotos diferentes y usar la relación mendeliana para obtener las frecuencias esperadas de cada uno. Las proporciones esperadas son 9/16 para amarillos y lisos, 3/16 para amarillos y rugosos como para verdes y lisos, y 1/16 para rugosos y verdes. Con tales proporciones puede calcular las frecuencias esperadas de las 4 casillas. Por ejemplo, el número esperado de porotos lisos y amarillos será $(9 / 16) N = 312,75$.

Porotos	O _i	E _i	O _i ln (O _i /E _i)	G
Lisos y amarillos	315	312,75	2,258	0,476 (ns)
Rugosos y amar.	101	104,25	-3,199	
Lisos y verdes	108	104,25	3,817	
Rugosos y verdes	32	34,75	-2,638	
Total	556	556	0,238	

Los grados de libertad son: $\nu = (4 - 1) = 3$ y entonces $\chi^2_{0,95; 3} = 7,815 \gg \gg G$ por lo que no se puede rechazar la H₀.

Luego si se usa el modelo menos potente de la Chi-cuadrado resulta:

$$\chi^2 = \frac{(315 - 312,75)^2}{312,75} + \frac{(108 - 104,25)^2}{104,25} + \frac{(101 - 104,25)^2}{104,25} + \frac{(32 - 34,75)^2}{34,75} = 0,47$$

Notar que ahora el modelo de la G es más conservador que el de la Chi pues:

$$G = 0,476 > \chi^2 = 0,47$$

Esto muestra que al modelo de la G le cuesta más rechazar una hipótesis nula que al modelo de la Chi-cuadrado, para un mismo caso. Por eso G es más conservador que Chi, y si a esto se le agrega que G es más sensible que Chi para detectar diferencias, se justifica el abandono por parte de los investigadores del clásico modelo de la Chi cuadrado para analizar estos casos.

15.3 Tablas de contingencia

Se tratará ahora el caso de los llamados *tests de independencia*. La noción de independencia estadística o probabilística se vio en el Capítulo 6.2. Allí se mostró que si dos sucesos son independientes, la probabilidad de que ocurran juntos se puede calcular con el producto de sus probabilidades individuales. Por ejemplo, en teoría de juegos la probabilidad de sacar dos secas al lanzar dos monedas al aire es de $1/4$. En Genética la probabilidad de que un grano de maíz sea de color rojo es de $1/2$ y de que sea rugoso $1/3$, por lo tanto la probabilidad de que sea a la vez rojo y rugoso es de $1/6$. Se puede plantear un experimento para verificar la hipótesis anterior. Con el modelo de la Chi-cuadrado se pueden comparar las frecuencias observadas y las esperadas. La proporción mendeliana es de $2 : 2 : 1 : 1$ y la prueba testea dos cosas a la vez: a) las probabilidades de cada una y b) si son independientes. La primera hipótesis es para verificar el modelo de Mendel, pero la segunda prueba si los genes son independientes o si están ligados, es decir, ubicados en el mismo cromosoma.

La Tabla 15.1 muestra un caso de una *tabla de contingencia* del tipo $1 \times n$ identificada como una *tabla de clasificación simple*, y en el mismo punto se muestra cómo se puede plantear la *prueba de la Chi-cuadrado* para resolver un test de hipótesis. Generalizando estas ideas, se llega a un caso más amplio: *Tabla de Contingencia $h \times k$* donde las frecuencias observadas O_{ij} ocupan una tabla de h filas y k columnas. En correspondencia con las mismas se pueden calcular, bajo algún modelo teórico planteado, sus respectivas frecuencias esperadas E_{ij} . Entonces se puede plantear una vez más la prueba de la Chi-cuadrado para este caso como:

$$\chi^2 = \sum \sum (O_{ij} - E_{ij})^2 / E_{ij}$$

Donde O_{ij} : es la frecuencia observada en la columna i ; fila j
 E_{ij} : es la frecuencia esperada en la columna i ; fila j

Este estadígrafo se contrasta con el valor crítico $\chi^2_{\alpha; n}$. Los grados de libertad se calculan para dos casos posibles:

1) *Hipótesis Extrínseca*: es una hipótesis externa a los datos, no se necesita de estos para obtener los parámetros poblacionales en el cálculo de las frecuencias esperadas:

$$v = (h - 1) (k - 1)$$

2) *Hipótesis Intrínseca*: es una hipótesis interna, se necesitan los datos para sacar los m parámetros poblacionales:

$$v = (h - 1) (k - 1) - m$$

Ejemplo 1) En cierta Universidad, los ingresantes fueron clasificados en tres grupos principales, de acuerdo con la orientación de sus estudios secundarios en: escuelas de comercio, normal e industrial. Se desea saber si la elección de sus carreras universitarias guarda alguna relación con su origen. Para ello, se tomaron datos de tres grandes facultades con orientaciones similares, tales como: económicas, humanista y ciencias exactas. Se formula la hipótesis nula que la elección de

carrera que hace el ingresante es independiente del tipo de su escuela secundaria. O sea, la proporción de estudiantes inscriptos en las carreras es la misma.

Frecuencias Observadas (Oij)					Frecuencias Esperadas (Eij)				
	Comercio	Normal	Indus.	TOTAL		Comercio	Normal	Indus.	TOTAL
Economía	524	325	200	1049	Economía	424,9	315,1	309,1	1049
Sociales	435	256	254	945	Sociales	382,7	283,8	278,5	945
Exactas	256	320	430	1006	Exactas	407,4	302,1	296,4	1006
TOTAL	1215	901	884	3000	TOTAL	1215	901	884	3000

Los datos obtenidos se vuelcan en una tabla de frecuencias observadas, como se muestra en la tabla de la izquierda arriba. Para obtener las frecuencias esperadas se procede como sigue:

Si los dos factores investigados (Tipo de escuela secundaria y Tipo de Universidad elegida) son independientes, entonces el valor esperado en la primera casilla E11 se puede imaginar como la probabilidad esperada P11 multiplicada por el número total de datos N. La probabilidad teórica P11, de que un alumno verifique las condiciones pedidas en tal casilla (estudió en una escuela de comercio en su secundaria y eligió la facultad de Económicas) será:

$P_{11} = E_{11} / N$ Por otra parte, puede imaginarse la tabla anterior como una partición del universo de alumnos. En particular, la celda $i = 1$ y $j = 1$ es la intersección de dos sucesos: **E**: facultad de económicas y **C**: escuela de comercio. Luego:

$P(E \cap C) = P(E) \cdot P(C)$ Si **E** y **C** son independientes, la probabilidad del suceso $E \cap C$ es igual al producto de sus probabilidades. Y a su vez, cada una es:

$$P(E) = N_E / N = 1049 / 3000 = 0,35 \quad \text{y} \quad P(C) = N_C / N = 1215 / 3000 = 0,405$$

Reemplazando en la primera ecuación es: $P_{11} = P(E \cap C) = E_{11} / N$ entonces,

$$E_{11} = N \cdot P_{11} = N P(E \cap C) = N \cdot P(E) \cdot P(C) = N \cdot (N_E / N) (N_C / N) \quad \text{o sea,}$$

$$E_{11} = (N_E \cdot N_C) / N = (1049 \cdot 1215) / 3000 = 424,9$$

La frecuencia esperada de la primera casilla es igual al producto de los totales marginales correspondientes, dividido el tamaño total de la muestra tomada. Análogamente para las demás casillas se puede concluir que: la frecuencia esperada de una casilla cualquiera Eij se puede obtener con la relación:

$$E_{ij} = N_i \cdot N_j / N$$

Aplicando esta relación a todas las casillas de las frecuencias observadas, se obtienen las frecuencias esperadas Eij que se muestran en la tabla de arriba, en su parte derecha. Una vez que se tienen todos los valores se puede calcular el estadígrafo:

$$\chi^2 = \sum \sum (O_{ij} - E_{ij})^2 / E_{ij} = [(524 - 424,9)/424,9] + [(325 - 315,1)/315,1] + \dots = 191,47 \quad ***$$

valor mucho mayor que el de tablas $\chi^2_{\alpha;v}$ con $v = (3 - 1) (3 - 1) = 4$. O sea, $\chi^2_{0,999;4} = 18,467$

Por lo tanto, se rechaza la hipótesis nula y se concluye que el tipo de escuela secundaria realizada por el ingresante influye en su elección de carrera universitaria, como era de esperar. Se deja al lector el cálculo del estadígrafo G para este caso, donde la evidencia es tan fuerte que no se necesita mayor sensibilidad para rechazar H_0 .

Esta prueba requiere que los datos observados en cada casilla no sean pequeños. Ninguna celda deberá tener frecuencia esperada nula y el 20% de las mismas no debe ser inferior a 20. Si los datos originales no verifican estas condiciones, la solución es agrupar categorías similares para lograr que se verifiquen estas condiciones.

15.4 Tablas de contingencias 2 x 2

Es un caso particular del visto en el punto anterior, pero es el más popular entre los investigadores y por ese motivo se lo tratará con más detalle. Los modelos estadísticos propuestos para analizar este tipo de tablas son tres: el test de la Chi-cuadrado, el G-test y el test exacto de Fisher. Es conveniente, antes de seguir adelante, identificar a tres tipos de casos que se pueden encontrar en la práctica, de acuerdo a como el investigador desea planificar su experimento:

Modelo I: se fija de antemano únicamente el tamaño muestral.

Modelo II: se fija de antemano el tamaño de los totales marginales en uno de los factores.

Modelo III: se fijan de antemano todos los totales marginales de la tabla.

El modelo de la Chi-cuadrado no fue diseñado específicamente para ninguno de los tres modelos y los investigadores lo suelen usar sin hacer distinción entre ellos. Sin embargo, estudios recientes aconsejan lo siguiente:

- Usar el G-test de independencia para los modelos *I* y *II*.
- Usar el test exacto de Fisher para el modelo *III*.

Si un bioquímico elige al azar una cantidad de 100 pacientes para estudiar la incidencia del sexo en la preferencia del paciente por optar por uno de los dos turnos de extracción de la muestra. Tiene dos factores a investigar: Sexo y Turno. No controla la cantidad de hombres y mujeres que concurren a su laboratorio, ni tampoco si se reparten en el turno de la mañana o de la tarde. Por lo tanto este caso será tratado como Modelo *I*. En cambio, si fija de antemano que el 50% de su muestra serán hombres, entonces cambia la situación y ahora tiene un Modelo *II*. Si además elige que el 70% de los seleccionados sea del turno mañana y el resto de la tarde, ya ha definido los cuatro totales marginales y por tanto debe tratar a este caso como Modelo *III*.

Ejemplo 1) En los experimentos de Mendel con porotos detallado más arriba, se puede realizar otro tipo de análisis. Ahora se desea verificar los dos factores: Color y Textura, son independientes (esto es, no pueden estar ligados). Entonces, se plantea este mismo problema como si fuera una tabla de 2x2 (en lugar de 1x4 como antes). En este caso, el investigador no puede fijar ninguno de los totales marginales, a lo sumo selecciona la cantidad N de porotos para la muestra. Las proporciones esperadas en la Tabla de 1 x 4 eran: 9/16 para amarillos y lisos, 3/16 para ama-

rillos y rugosos como para verdes y lisos, y 1/16 para rugosos y verdes. Pero ahora para testear independencia serán las frecuencias esperadas de las 4 casillas. Por ejemplo, el número esperado de porotos lisos y amarillos será $(416 \times 423 / 556) = 316,5$ en lugar de 312,75 cuando se usa la teoría mendeliana. Planteado de esta manera el problema queda:

Frecuencias observadas: O_{ij}

Frecuencias esperadas: E_{ij}

FACTOR B: tersura

FACTOR	Lisos	Rugosos	Total
A: color			
Amarillos	315	101	416
Verdes	108	32	140
Total	423	133	556

	Lisos	Rugosos	Total
Amarillos	316,5	99,5	416
Verdes	106,5	33,5	140
Total	423	133	556

Se trata de un **Modelo I**. Luego se calcula el estadígrafo con:

$$\chi^2 = \frac{(315 - 316,5)^2}{316,5} + \frac{(108 - 106,5)^2}{106,5} + \frac{(101 - 99,5)^2}{99,5} + \frac{(32 - 33,5)^2}{33,5} = 0,118$$

Los grados de libertad son: $\nu = (2 - 1)(2 - 1) = 1$. De tablas se encuentran valores críticos mayores que el estadígrafo calculado y por lo tanto no se puede rechazar la hipótesis nula.

Aplicando el G-test al mismo caso se obtiene:

$$G = 2 \sum O_i \cdot \ln(O_i / E_i)$$

$$G = 2 [315 \ln(315/316,5) + 108 \ln(108/106,5) + 101 \ln(101/99,5) + 32 \ln(32/33,5)]$$

$$G = 2 (-1,497 + 1,511 + 1,511 - 1,466) = 0,059 \text{ (no significativo)}. \text{ Acá } G < \chi^2 \text{ (robustez).}$$

Ejemplo 2) Se eligen 400 pacientes al azar, de entre aquellos que sufren cierta enfermedad. A la mitad se le aplica una nueva droga curativa, y a la otra mitad un placebo. Decidir si la nueva droga es efectiva con los datos encontrados que se muestran en la tabla siguiente.

Frecuencias observadas: O_{ij}

Frecuencias esperadas: E_{ij}

FACTOR B: Curación

FACTOR	SI	NO	Total
A: droga			
SI	185	15	200
NO(placebo)	95	105	200
Total	280	120	400

	SI	NO	Total
SI	140	60	200
NO	140	60	200
Total	280	120	400

Se trata de un Modelo II. Por razones didácticas, se lo resuelve primero con el modelo de la Chi-cuadrado, sin y con la corrección de Yates por continuidad -se trata de datos discretos y la distribución teórica es continua-; esta corrección consiste en restarle la mitad de la menor unidad

empleada en valor absoluto, en este caso 0,5. Luego, se aplica el modelo aconsejado el G-test y por ultimo el G-test con la corrección de Williams por continuidad. En total se resuelve el problema de cuatro formas diferentes, para poder ver las ventajas y desventajas de cada uno:

Caso 1) Empleando el modelo de la Chi-cuadrado es:

$$\chi^2 = \frac{(185 - 140)^2}{140} + \frac{(95 - 140)^2}{140} + \frac{(15 - 60)^2}{60} + \frac{(105 - 60)^2}{60} = 96,43 *** >>> \chi^2_{0,999; 1} = 10,83$$

Caso 2) Empleando el modelo de la Chi-cuadrado con la corrección de Yates es:

$$\chi^2 = \frac{(|185 - 140| - 0,5)^2}{140} + \frac{(|95 - 140| - 0,5)^2}{140} + \frac{(|15 - 60| - 0,5)^2}{60} + \frac{(|105 - 60| - 0,5)^2}{60}$$

$$\chi^2 = \frac{(44,5)^2}{140} + \frac{(44,5)^2}{140} + \frac{(44,5)^2}{60} + \frac{(44,5)^2}{60} = 94,3 ***$$

Hay evidencia altamente significativa como para rechazar la hipótesis nula. Se tiene prueba científica de que la droga es buena. Tanto con el modelo clásico de la Chi-cuadrado como con este mismo modelo aplicándole la corrección por continuidad (que la hace menos sensible).

Caso 3) Empleando la prueba de G resulta:

$$G = 2 \sum O_i \ln (O_i / E_i) = 2 [185 \ln (185/140) + 95 \ln (95/140) + 15 \ln (15/60) + 105 \ln (105/60)]$$

$$G = 2 (51,56 - 36,84 - 20,79 - 58,76) = 105,38 ***$$

Una vez más, se ve la mayor sensibilidad de este modelo frente al Chi-cuadrado $G > \chi^2$

Caso 4) Empleando la prueba de G con la corrección de Williams por continuidad, la cual es preferible a la de Yates, resulta:

$$G_{\text{corr.}} = G / Q \quad \text{con} \quad Q = 1 + (A \cdot B / 6n)$$

Para ello se calculan los siguientes factores

$$A = [n (\text{total marginal de fila 1})^{-1} + n (\text{total marginal de fila 2})^{-1} - 1]$$

$$B = [n (\text{total marginal de columna 1})^{-1} + n (\text{total marginal de columna 2})^{-1} - 1]$$

O sea,

$$A = [(400/200) + (400/200) - 1] = 3$$

$$B = [(400/280) + (400/120) - 1] = 3,76$$

$$Q = 1 + 3 (3,76) / 2400 = 1 + 0,0047 = 1,0047$$

$$G_{\text{corr.}} = G / Q = 105,38 / 1,0047 = 104,89 \text{ ***}$$

El valor ajustado es ligeramente menor que G, pero igualmente significativo. Conviene destacar que esta corrección se impone cuando el valor de G esta cerca del valor crítico de tablas, para despejar cualquier duda. Pero, en el presente ejemplo la diferencia es tan grande que no se hace necesario realizar la corrección. Aquí se hizo por razones didácticas.

Otra vez se puede notar que el G-test pone más en evidencia las diferencias significativas que el test del la Chi-cuadrado. Esto se debe a la mayor sensibilidad del G-test, que se convierte en la opción recomendable, para tratar las tablas de contingencia 2 x 2 diseñada como un Modelo II experimental. Con la aclaración de que se debe usar la corrección de Williams cuando el valor del estadígrafo esté cercano al límite de la zona de rechazo al 95% de confianza.

Los valores críticos usados en las tablas de 2 x 2 son:

$$\chi^2_{1,095} = 3,841 \quad ; \quad \chi^2_{1,099} = 6,635 \quad \text{y} \quad \chi^2_{1,0999} = 10,83$$

En resumen: en una Tabla 2 x 2, diseñada como Modelo I o Modelo II, conviene usar la prueba de la G (G-test). Cuando el valor obtenido experimentalmente esté cercano a 3,841 se debe usar la corrección de Williams para salir de dudas. Cuando supere 6,635 no hace falta tal corrección, salvo para la sutileza de determinar la probabilidad asociada. Y cuando sea mayor de 11 ya no hay dudas de la fuerte evidencia obtenida.

Ejemplo 3) Un cardiólogo decide investigar si los problemas de hipertensión de sus pacientes son independientes de los antecedentes familiares de los mismos. Para ello, escoge 12 historias clínicas que cuentan con esos datos; del total son: 5 con antecedentes y 7 sin, mientras que del total solo 5 no son hipertensos. Clasifica sus datos en una tabla y aplica el test exacto de Fisher para determinar si ambos factores son independientes.

Casos observados

Factor B: Hipertensos

Factor A :		SI	NO	Total
Con	SI	2	5	7
antecedentes	NO	3	2	5
	Total	5	7	12

En términos generales

	SI	NO	Total
SI	a	b	a+b
NO	c	d	c+d
Total	a+c	b+d	N

Es un **Modelo III**. La probabilidad exacta se calcula con la probabilidad multinomial. Escribiendo la tabla de datos observados, en términos generales (la tabla de la derecha) se puede expresar:

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

Tomando logaritmos para facilitar el calculo queda:

$$\ln P = \ln (a+b)! + \ln (c+d)! + \ln (a+c)! + \ln (b+d)! - \ln n! - \ln a! - \ln b! - \ln c! - \ln d!$$

Se puede usar cualquiera de ambas relaciones para resolver el cálculo. Para este caso no se usan los logaritmos porque los valores son chicos. Para usar el Modelo de Fisher se procede así:

Paso 1) Se calcula la probabilidad multinomial de que a lo sumo haya **a** casos en la primera celda. Esto es, calcular $P(r \leq a) = P(r = 0) + P(r = 1) + \dots + P(r = a)$. Cada uno de estos términos es una Multinomial que se calcula con la fórmula de más arriba.

Paso 2) Se compara esta probabilidad acumulada $P(r \leq a)$ con el nivel de significación $\alpha = 0,05$, para aceptar o rechazar la hipótesis nula como es usual.

Para facilitar la visualización del problema se presentan los casos para obtener la acumulada $r \leq 2$

r = 2

2	5
3	2

r = 1

1	6
4	1

r = 0

0	7
5	0

$$P = \frac{7! \cdot 5! \cdot 7! \cdot 5!}{12! \cdot 2! \cdot 5! \cdot 3! \cdot 2!} = 0,26515 = P(r = 2) \quad : \text{ celda de la izquierda.}$$

$$P = \frac{7! \cdot 5! \cdot 7! \cdot 5!}{12! \cdot 1! \cdot 6! \cdot 4! \cdot 1!} = 0,04399 = P(r = 1) \quad : \text{ celda del medio.}$$

$$P = \frac{7! \cdot 5! \cdot 7! \cdot 5!}{12! \cdot 0! \cdot 7! \cdot 5! \cdot 0!} = 0,00126 = P(r = 0) \quad : \text{ celda de la derecha.}$$

$$P(r \leq 2) = P(r = 0) + P(r = 1) + P(r = 2) = 0,3104 \gg \alpha = 0,05$$

Por lo tanto, no se puede rechazar la hipótesis nula. Los antecedentes familiares son independientes de la hipertensión. Como los datos son muy pocos, el investigador puede aumentar el número de casos para tener un panorama más acorde a sus sospechas.

15.5 Aplicaciones en Farmacia y Bioquímica

En la últimas décadas la aplicación de estadística en Medicina se ha incrementado vertiginosamente. En un comienzo, sus métodos para el diseño y análisis de experimentos comenzaron a usarse en el área de los estudios sobre enfermedades del corazón y cáncer. La incorporación de nuevos índices clínicos -como los vistos en el capítulo 4- empezó a mediados de la década del 70. Hay muchas razones para que haya ocurrido esto; entre ellas se pueden citar a dos: cada vez es más aceptado que la mayoría de las enfermedades son causadas por múltiples factores y el desarrollo de cada enfermedad involucra muchos cofactores diferentes. El análisis multivariado, especial para estos casos, que requiere un nivel matemático no acorde con el general en carreras relacionadas con Medicina, se ha vuelto accesible gracias a la aparición de software de computadora que soluciona los problemas a los que no son especialistas. Además, el hardware necesario para poder usarlo ya está disponible en el mercado, a precios de ensueño en el pasado.

Más recientemente, los estudios epidemiológicos se han incorporado a esta corriente innovadora, con la aplicación de técnicas más sofisticadas. En la problemática del contagio cabe mencionar a dos áreas en particular: los estudios sobre infecciones de tipo hospitalaria y sobre la inmunodeficiencia humana con el virus tipo 1 (HIV-1). Muchas investigaciones se están realizando para estudiar la historia natural de la HIV-1 y los factores asociados con el desarrollo de la

adquisición del síndrome de esta (SIDA). Por su parte, ya es aceptado universalmente que las infecciones hospitalarias no son causadas por un único factor de riesgo, sino que se originan en una compleja interacción entre factores relacionados con el infectado, el agente infeccioso y el medio ambiente que lo rodea. En este capítulo se presentan las medidas de asociación más usuales en este tipo de estudios: el “*riesgo relativo*” (*relative risk*) y la “*eficacia*” (odds ratio), en su forma más elemental, aplicadas a diseños con tablas de 2x2, dejando la regresión logística para estudios más avanzados.

Una enfermedad puede ser estudiada desde dos perspectivas. Una de tipo “observacional” donde el investigador no interviene ni manipula los factores involucrados y deja que la enfermedad siga su curso natural. La otra de tipo “experimental” donde el investigador puede controlar de alguna manera factores que pueden ser de importancia en el proceso de la enfermedad, y se usan para examinar relaciones “causa-efecto” entre los factores involucrados. Los estudios poblacionales por cohortes y los estudios caso-control son del tipo observacional. Los estudios en pruebas clínicas son de tipo experimental, donde un grupo está expuesto a un factor de riesgo o a un tratamiento, y el otro grupo no lo está. El investigador, usando una tabla de números al azar, reparte a los pacientes seleccionados para el experimento, en forma aleatoria, dentro de cada grupo. Un grupo actuará como placebo o control mientras que el otro recibirá el tratamiento o una protección. Si se homogenizan los grupos antes de comenzar el tratamiento, cuidando factores tales como edad, sexo, etc., se supone que se evita el sesgo que produciría alguno de esos factores, si no es tomado en cuenta. Si los grupos son lo suficientemente grandes, se espera que dentro de los resultados el grupo protegido sea más grande que el otro, y aparecerá una significación estadística que le permitirá al investigador tener una prueba científica de la relación causa-efecto entre los factores estudiados. Conviene destacar que esta prueba es de tipo estadístico, lo cual no siempre tiene significado en lo clínico, *que exista significación estadística no siempre implica que haya una significación clínica.*

Hay que ser cuidadoso con la homogeneización de la muestra. Si es “demasiado” homogénea, se tornará difícil extender las conclusiones del estudio a la población en general. Por otra parte, si no lo es, pueden aparecer factores importantes en el desarrollo de la enfermedad, que introduzcan variabilidades engañosas. Por ejemplo, si se hace una investigación sobre infecciones post-operatorias en enfermedades coronarias, estudiando el efecto de determinado antibiótico para prevenirlas, se hace poco creíble extender las conclusiones de este estudio a todas las infecciones de tipo hospitalario. Por otra parte, si la muestra se toma eligiendo a pacientes provenientes de cualquier cirugía, sin importar edad o sexo, ocurre otro tanto. Por ejemplo, incluir muchas cesáreas en el estudio, sin hacer discriminación de las mismas en el total, implicaría un sesgo al sexo femenino y a cierta escala de edades. Esto invalidaría extender las conclusiones del estudio a ambos sexos. El otro punto a tener muy en cuenta, es realizar este tipo de experimentos al estilo “doble-ciego”. Esto es, ni los que intervienen en la investigación, ni los pacientes, deben saber quiénes reciben las drogas y quiénes los placebos. De esta forma se evita cualquier tipo de subjetividad o favoritismo. La diferencia entre este tipo experimental de trabajo (“clinical trials”) y los estudios poblacionales del tipo caso-control, es que la enfermedad ya se ha presentado en este último, a diferencia del primero. Y el investigador busca en la historia pasada del paciente las causas que originaron la enfermedad, los factores que pueden haber influenciado. Su grupo control será otro grupo de individuos que no la padezcan y que tengan características similares a los enfermos, y buscando en la historia pasada la existencia de los mismos factores en estos últimos, podrá sacar conclusiones.

15.6 Odds-ratio y riesgo relativo

En Medicina el *riesgo* puede ser definido como la probabilidad de que ocurra un determinado suceso que implique un peligro para la salud de un paciente. Tal como el riesgo de contraer una septicemia luego de ser operado.

El *factor de riesgo* es una magnitud cualitativa, de tipo dicotómica, donde sus dos resultados posibles indican si el factor está presente o no. Usualmente esto se denota con: SÍ-NO. Son ejemplos de factor de riesgo características tales como: edad, sexo, tratamiento profiláctico antes de una cirugía, inmunización, medicación preventiva, etc. Cuando un grupo de pacientes recibe una droga que se está probando, será el caso SÍ del factor de riesgo, y el otro grupo usado como control o placebo será el caso NO. En cambio, cuando el primer grupo es cateterizado en su cirugía se piensa que está expuesto a una infección (SÍ) y otro grupo donde no fue necesaria (NO) será el de control.

La comparación del riesgo es uno de los objetivos principales de los estudios epidemiológicos, y su cuantificación implica la medición experimental de sus dos índices asociados: *Riesgo Relativo* (RR) y “*Odds-Ratio*” (OR).

El diseño experimental para medir ambos índices es el planteo de una Tabla de Contingencia del tipo 2x2. Como el investigador es quien decide el tamaño muestral y cuántos serán expuestos al factor de riesgo, es un caso de *Modelo II*. Suponiendo que toda la población estudiada es homogénea en todas sus características excepto en una: una parte de esta se expone a un factor de riesgo, que se cree es de importancia en causar una determinada enfermedad, y que ninguno de los cuatro casos posibles tiene una ocurrencia nula; el diseño experimental usado generalmente se muestra esquemáticamente en la Tabla 15.2 siguiente:

Tabla 15.2: Frecuencias observadas entre la exposición a un factor de riesgo y la contracción de una enfermedad cualquiera.

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
O	a	b	$n_1 = a+b$
Característica	c	d	$n_2 = c+d$
Grupal	Total $n_3 = a+c$	$n_4 = b+d$	n

$$RR = (a \cdot n_2) / (c \cdot n_1)$$

$$OR = (a \cdot d) / (b \cdot c)$$

Riesgo relativo (RR) se define como el cociente entre la probabilidad de contraer la enfermedad de la población expuesta y la probabilidad de contraerla de los no expuestos.

$$RR = P(E) / P(noE) = (a/n_1) / (c/n_2) = (a \cdot n_2) / (c \cdot n_1)$$

que es la expresión de más arriba.

Cuando no hay pacientes de la población expuesta que contraigan la enfermedad ($a = 0$) el RR se anula. En cambio, si ningún paciente de los no expuestos contrajo la enfermedad ($c = 0$) entonces el RR se vuelve infinito. Si los factores son *independientes* el $RR = 1$. El RR solo puede ser estimado con estudios prospectivos. La interpretación se puede ver mejor con un ejemplo: se sabe que entre los que sufrieron infarto de miocardio el nivel de colesterol medio es de 300

mg/dl. En la población general el nivel es de 200 mg/dl. Se conoce la distribución de colesterol en ambas poblaciones, la de los infartados y la de la población humana usadas como referencia. Desde el punto de vista médico se toma como valor referencial 250 mg/dl, se considera a un paciente con “colesterol alto” cuando supera dicho valor. Luego se cuentan los casos de encontrados con colesterol alto en los infartados y entre los no infartados. Sea por caso, una relación de $120/180 = 2/3$ entre los infartados y de $1000/9000 = 1/9$ entre los no infartados, entonces se calcula el $RR = (2/3) / (1/9) = 6$. Eso se interpreta así: “Si un paciente tiene un colesterol mayor que 250, su chance de infarto es 6 veces mayor que si tuviese un valor de 200 o menos”. Así se construyen las simplificaciones en los diarios y revistas de divulgación.

Si se trata de un estudio del tipo caso-control, generalmente el RR no puede calcularse y se necesita del OR como una medida de asociación entre ambos factores analizados. Entonces, el *Odds Ratio* (OR) se define como el cociente entre dos “odds” posibles. Un “*odd*” es la relación entre la cantidad de “enfermos” y los “no enfermos” de una población dada. Como hay dos poblaciones, la expuesta y la no expuesta al factor de riesgo, hay dos tipos de “*odd*” posibles y la tasa entre ambos es el valor de OR. Sea:

$P_1 = a / n_1$: La probabilidad de contraer la enfermedad de la población expuesta.

$(1 - P_1) = b / n_1$: la probabilidad de no contraer la enfermedad de la población expuesta

El “*odd*” de la población expuesta es $O_1 = P_1 / (1 - P_1) = a / b$. A su vez,

$P_2 = c / n_2$: la probabilidad de contraer la enfermedad de la población no expuesta.

$(1 - P_2) = d / n_2$: la probabilidad de no contraer la enfermedad de la población no expuesta.

El “*odd*” de la población no expuesta es $O_2 = P_2 / (1 - P_2) = c / d$.

Por lo tanto, su cociente es $OR = O_1 / O_2 = (a \cdot d) / (c \cdot b)$ que es la expresión vista en la Tabla 15.2 de más arriba. Cuando el $OR = 1$ significa que ambos factores estudiados son independientes entre sí. Si a o d se anulan entonces el $OR = 0$; en cambio, si se anulan c o b , se hace infinito. Para el caso usado como ejemplo para el RR, se puede calcular:

$$OR = (120 / 1000) / (60 / 8000) = 16$$

Y se interpreta así: “cuanto más alto el OR, peor”. Otra forma de ver este índice es como el cociente de dos pares de probabilidades. Una es el RR entre los expuestos y el otro es el RR entre los no expuestos:

$$RR1 = Pe / (1 - Pe) = \frac{a / (a+b)}{b / (a+b)} = a/b \quad \text{y} \quad RR2 = Pno e / (1 - Pno e) = \frac{c / (c+d)}{d / (c+d)} = c/d$$

Luego es: $OR = RR1 / RR2$

La diferencia en el uso de uno u otro índice reside en el tipo de investigación que se está realizando. *RR* se emplea en un estudio de *incidencia*, donde la frecuencia de algún resultado se calcula entre dos grupos determinados por la presencia o ausencia de alguna característica. En

cambio, *OR* se usa en un *estudio de control* donde los resultados se obtienen en dos grupos, uno expuesto a un factor de riesgo y el otro usado como placebo o control.

Tanto el *RR* como el *OR* son estadísticos cuantitativos, no se asocian a pruebas de inferencia que terminan en un nivel de significación; en ese sentido no importa la distribución, lo que se estima es el intervalo de confianza de ambos valores al 95% o 99%, obtenidos con las relaciones que se muestran a continuación. Luego se trata de ver si el valor de *RR* = 1, o *OR* = 1 cae dentro de dichos intervalos, en cuyo caso se piensa que hay independencia entre los factores:

Test para OR:

Valor esperado: $E(OR) = \mu_{OR} = OR$

Varianza: $E\{(OR^2)\} - \mu_{OR}^2 = \sigma_{OR}^2$. Para simplificar el cálculo se toman logaritmos:
 $\sigma^2(\ln OR) = a^{-1} + b^{-1} + c^{-1} + d^{-1}$

O sea, $\sigma^* = (a^{-1} + b^{-1} + c^{-1} + d^{-1})^{1/2} = SE(\ln OR)$

Luego, los límites del intervalo de confianza de *OR* (95%) son:

Límite superior: $OR_s = OR e^{1,96 \sigma^*}$ 95% CI (*OR*_i ; *OR*_s)
Límite inferior: $OR_i = OR e^{-1,96 \sigma^*}$

Cuando el valor *OR* = 1 cae fuera del intervalo, hay evidencia como para rechazar la *H*₀. Para el 99% de confianza el exponente es $\pm 3,09$ (en lugar de $\pm 1,96$ para el 95%).

Test para RR:

Valor esperado: $E(RR) = \mu_{RR} = RR$

Varianza: $E\{(RR^2)\} - \mu_{RR}^2 = \sigma_{RR}^2$. Para simplificar el cálculo se toman logaritmos:
 $\sigma^2(\ln RR) = a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1}$

O sea, $\sigma'' = [a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1}]^{1/2} = SE(\ln RR)$

Luego, los límites del intervalo de confianza de *OR* (95%) son:

Límite superior: $RR_s = RR e^{1,96 \sigma''}$ 95% CI (*RR*_i ; *RR*_s)
Límite inferior: $RR_i = RR e^{-1,96 \sigma''}$

Cuando el valor *RR* = 1 cae fuera del intervalo, hay evidencia como para rechazar la *H*₀. Para el 99% de confianza el exponente es $\pm 3,09$ (en lugar de $\pm 1,96$ para el 95%).

Para ver mejor estas ideas se desarrollan los ejemplos siguientes:

Ejemplo 1) En un hospital pediátrico se tomaron datos de contagio bacteriano post-operatorio, con relación al uso de una cateterización de la arteria umbilical en 340 cesáreas realizadas. Los datos tomados de la obra de Muñoz y Townsend (pág. 963) son ficticios.

Cateterización:	Bacteriemia		Total	OR = $\frac{130 \cdot 80}{30 \cdot 100}$
	SI	NO		
SÍ	130	100	230	
NO	30	80	110	OR = 3,47
Total	160	180	340	

$$\sigma^2 (\ln OR) = a^{-1} + b^{-1} + c^{-1} + d^{-1} = (130)^{-1} + (100)^{-1} + (30)^{-1} + (80)^{-1} = 0,06353$$

$$\sigma^* = (a^{-1} + b^{-1} + c^{-1} + d^{-1})^{1/2} = (0,06353)^{1/2} = 0,252$$

Luego, los límites del intervalo de confianza de OR (95%) son:

$$\text{Límite superior: ORs} = OR \cdot e^{1,96 \sigma^*} = 3,47 \cdot e^{1,96 (0,252)} = 5,686$$

$$\text{Límite inferior: ORi} = OR \cdot e^{-1,96 \sigma^*} = 3,47 \cdot e^{-1,96 (0,252)} = 2,116$$

Como el valor: OR = 1 cae fuera de este intervalo 95% CI (2,116 ; 5,686), se tiene prueba estadística de una relación entre la infección y el haber tenido una cateterización operatoria.

Sin embargo, este es un caso de una Tabla de contingencia 2x2, Modelo II. Por lo tanto, conviene más aplicar la prueba de G a los mismos datos. Resulta: $G = 26,33^{***}$ y corrigiendo con Williams es un poco menor $G_{corr} = 26,192^{***}$. Pero ahora se tiene una prueba científica de la falta de independencia entre ambos factores analizados: hay asociación estadística entre el factor analizado y la enfermedad.

Ejemplo 2) En una investigación poblacional se obtuvieron 38.152 datos que fueron clasificados según si sufrieron un infarto o no, y dentro de estos fueron desagrupados de acuerdo a si tenían un valor de glucosa alto (mayor a 1,1). Se trata de establecer si un valor alto de glucosa tiene incidencia en sufrir infartos. Los datos mostrados son ficticios. Los resultados se vuelcan en la tabla siguiente:

Glucosa Alta	Infartados		Total	RR = $\frac{161 \cdot 34768}{2636 \cdot 748}$
	SÍ	NO		
SÍ	161	2475	2636	
NO	748	34020	34768	RR = 2,839
Total	909	36495	37404	

$$\sigma^2 (\ln RR) = a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1} = (161)^{-1} - (2636)^{-1} + (748)^{-1} - (34768)^{-1} = 0,0072$$

$$\text{O sea, } \sigma'' = [a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1}]^{1/2} = 0,0848 = SE(\ln RR)$$

Luego, los límites del intervalo de confianza de OR (95%) son:

$$\text{Límite superior: RRs} = RR \cdot e^{1,96 \sigma''} = 2,839 \cdot e^{1,96 (0,0845)} = 3,350$$

$$\text{Límite inferior: RRi} = RR \cdot e^{-1,96 \sigma''} = 2,839 \cdot e^{-1,96 (0,0845)} = 2,406$$

Como RR = 1 cae fuera del intervalo (2,406 ; 3,350), se rechaza la hipótesis nula de independencia entre ambos factores. Se concluye que las personas con glucosa alta tienen un riesgo

de infarto 284% mayor que los que tienen glucosa normal. Notar que si se hubiese hecho un G-test, el resultado sería $G_{corr} = 118,1^{***}$ y se tiene una evidencia muy fuerte para rechazar la hipótesis nula. Con el test de la Chi cuadrado ocurriría otro tanto.

15.7 Modelos en Epidemiología

Existen dos maneras de tomar las muestras en Epidemiología una es la de hacer un muestreo basándose en la enfermedad y la otra en la exposición. En la primera el investigador selecciona al azar una cierta cantidad de enfermos y otra de no enfermos, mientras que en la segunda selecciona una cierta cantidad de individuos expuestos al factor de riesgo (o de protección) y otra de no expuestos (o no protegidos). Conviene definir dos modelos genéricos para los estudios epidemiológicos:

Modelo D: Este diseño del estudio es específico para estudiar la enfermedad porque se hace un muestreo sobre la base del factor de riesgo. El investigador decide los valores de TE (total de individuos expuestos) y TnE (total de individuos no expuestos).

Modelo E: Este diseño del estudio es específico para estudiar el factor de riesgo (exposición o protección) porque se hace un muestreo sobre la base de la enfermedad. El investigador decide los valores de TD (total de individuos enfermos) y TnD (total de individuos sanos).

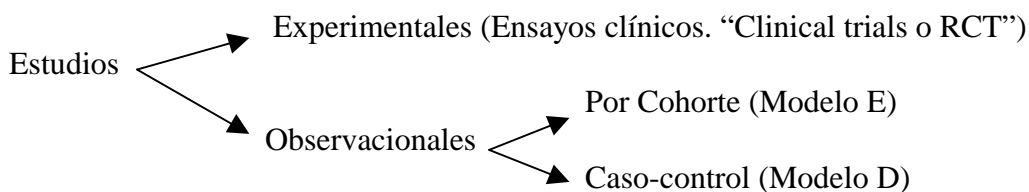
Tabla 15.3 Tabla de contingencia 2 x 2 para estudios epidemiológicos

Factor de Riesgo	Enfermedad		Total	<i>Modelo D:</i> El investigador decide el valor de los totales marginales TE y TnE
	SÍ	NO		
o Protección	SÍ	a	b	TE
	NO	c	d	TnE
Total	TD	TnD	N	<i>Modelo E:</i> El investigador decide el valor de los totales marginales TD y TnD

Por lo tanto, los valores de las cuatro celdas de la Tabla (a, b, c y d) junto con los otros dos totales marginales que no son elegidos por el investigador, serán “naturales” es decir libres de cualquier manipulación. Y lo mismo ocurrirá con los índices calculados a partir de ellos. Esto no se tiene en cuenta siempre, debido a los usos y costumbres habituales. Una mala costumbre es por ejemplo, primero hacer una selección al azar de los individuos enfermos y no enfermos, y luego eliminar a aquellos que no se ajustan a ciertos requerimientos del investigador. Por ejemplo, en un artículo publicado en *The Lancet*, 340: 1074-78, 1992, sobre el análisis de la eficacia de una vacuna contra la meningitis, los investigadores procedieron de esa manera y no hay forma de demostrar que la muestra tomada sea representativa de la población estudiada. Es evidente que no hubo mala intención por parte de los investigadores, porque explican este procedimiento en detalle, pero no pueden demostrar que la selección efectuada final sea representativa ni aleatoria, como tampoco que su criterio no haya influenciado los resultados. Esto es lo mismo que cuando se elige con la mano a una rata de una jaula que contiene muchas para hacer un experimento, es evidente que no hay preferencia por una u otra, pero no hay manera de demostrar que la extracción sea aleatoria. En realidad se atrapó, a la que se dejó agarrar, ya sea porque no es muy veloz, porque esta gorda, enferma, etc. Este problema es como el de la “casta Susana”, no es suficiente

serlo sino que además, hay que parecerlo. Y para ello la única manera es: 1º) De entre todos los individuos disponibles hay que descartar a los que no verifican los requisitos del investigador. 2º) Y luego hacer la selección al azar de entre los remanentes.

Los estudios epidemiológicos pueden ser clasificados como sigue:



En los estudios experimentales, el investigador controla ciertos factores escogidos que podrían tener alguna importancia en el desarrollo de la enfermedad. Lo más común es que manipule el factor (por ejemplo la aplicación de una vacuna en estudios inmunológicos, etc.). En los estudios observacionales el investigador no manipula los factores, estos ocurren en forma la natural. Los estudios por Cohorte enrolan a todos los individuos que estuvieron expuestos a un agente de exposición (por ejemplo HIV-1), tanto como el seguimiento de la historia natural que tuvo la enfermedad en los individuos (por ejemplo casos de SIDA). Se trata del uso de un diseño del estudio como un Modelo E. En cambio, en los estudios del tipo Caso-control, los individuos son seleccionados en base a la exposición (Modelo D) para ver si resultaron enfermos o no. El objetivo central de todos estos estudios, es la *comparación del riesgo* y las cuantificaciones de este concepto se hacen a través de dos índices el Riesgo Relativo y el Odds Ratio. En los estudios por Cohorte y Ensayos clínicos, el RR puede calcularse directamente. En cambio, el OR solo puede estimarse con estudios prospectivos como en lo Caso-control, donde RR no puede ser calculado.

Ambos índices se usan para ver si hay asociación estadística entre el factor y la enfermedad, cuando $OR = 1$ o bien $RR = 1$, se interpreta como que hay independencia estadística. El problema es que no son índices epidemiológicos propiamente dichos. Por ejemplo:

Caso-control (Modelo E)

Situación A

Factor	Enfermo		
	Si	No	
Si	10	10	20
No	90	90	180
	100	100	200

OR = 1 y OR 95% CI (0,4 ; 2,5)

Situación B

Factor	Enfermo		
	Si	No	
Si	90	90	180
No	10	10	20
	100	100	200

OR = 1 y OR 95% CI (0,4 ; 2,5)

Mirando los valores de OR en ambas situaciones, se deduce que no hay asociación estadística entre el factor y la enfermedad. No hay diferencia entre ambas desde el punto de vista estadístico. Sin embargo el porcentaje de expuestos en uno es claramente diferente del otro, porque no es lo mismo tener 20 que 180 individuos expuestos, sobre 200 estudiados. Si se efectúa una comparación de ambos OR siempre sería no significativa, en cambio si se comparasen la proporción de expuestos, o bien los Odds de Exposición, sí sería significativa. Esto muestra que OR no tiene un significado epidemiológico estricto, sino que solo estudia si el factor analizado tuvo algo que ver

con el desarrollo de la enfermedad. En un Caso-control, la enfermedad ya se ha producido en los Casos y no ha ocurrido en los controles. Por lo tanto la conclusión epidemiológica en estos ejemplos es: “Hay muchos factores que pueden ocasionar la enfermedad (dependencia multifactorial) y de acuerdo a los resultados obtenidos, el factor analizado no esta relacionado con la enfermedad, se trata de un “confounding factor”. Ahora bien, si se analizan las tablas anteriores con el test de la Chi-cuadrado resulta $\chi^2 \approx 0$ en ambos casos. La pregunta es ¿ Por que motivo se usa el OR-test en lugar de Chi para hacer la misma cosa ? La respuesta de los epidemiólogos es que OR tiene un significado clínico y el otro no. Pero lo mismo ocurre si hay asociación, como sigue

Situación C

Factor	Enfermo		
	Si	No	
Si	20	8	28
No	80	92	172
	100	100	200

OR = 2,9 y OR 95% CI (1,2 ; 6,9)

Situación D

Factor	Enfermo		
	Si	No	
Si	92	80	172
No	8	20	28
	100	100	200

OR = 2,9 y OR 95% CI (1,2 ; 6,9)

Mirando los valores de OR en ambos casos, la conclusión es que hay asociación entre el factor y la enfermedad. Lo mismo si se usa la Chi cuadrado, porque $\chi^2 \approx 6 > 3,841$. Ahora no es lo mismo tener 28 que 172 expuestos en 200 muestras, desde el punto de vista epidemiológico. Una rápida mirada a los resultados permite ver que la proporción de expuestos cambia de 14% a 86%. Imaginando que la Situación C ocurre después de cierta vacuna aplicada a la población, y la otra situación era antes de la vacuna. Entonces, esta vacuna debe ser relevante porque produjo el cambio en los expuestos (vacunados / no vacunados). Sin embargo, si se comparan ambas situaciones con el índice OR, no se podrá demostrar la eficacia de la vacuna Lo que sí hace OR es mostrar que la vacuna y la enfermedad están asociados (lo mismo que Chi). La pregunta es ahora: ¿ Cual es el objeto de usar OR en lugar de Chi en estos estudios ? La respuesta se deja a cargo del lector.

Ensayos clínicos y estudios por Cohorte (Modelo D)

Situación E

Factor	Enfermo		
	Si	No	
Si	10	90	100
No	10	90	100
	20	180	200

RR = 1 y OR 95% CI (0,4 ; 2,5)

Situación F

Factor	Enfermo		
	Si	No	
Si	90	10	100
No	90	10	100
	180	20	200

RR = 1 y OR 95% CI (0,4 ; 2,5)

El principal objetivo de estos estudios es detectar al factor de riesgo para la enfermedad. La diferencia entre ambos es que en caso por Cohortes el factor de riesgo ocurre naturalmente, en cambio en los Ensayos clínicos es controlado por el investigador. Cuando un grupo es expuesto (o protegido) a cierto factor, se lo compara con otro no expuesto (placebo) y se mide el RR. En los dos ejemplos anteriores el valor de RR = 1 y esto muestra que el factor no está asociado a la enfermedad. Lo mismo si se hace el test de la Chi-cuadrado ($\chi^2 \approx 0$). Sin embargo, no es lo mismo haber encontrado 20 que 180 enfermos en 200 casos. Si la situación E ocurre antes de cierto hecho y la F después, este hecho deber ser relevante, aunque estadísticamente no lo sea. El índice

RR dice únicamente que no hay asociación estadística. La pregunta es ¿ Si RR no está relacionado con la prevalencia encontrada de la enfermedad, y el objeto de estos estudios es ver que pasa con la enfermedad, para que se los usa en lugar de la Chi ? La respuesta de los epidemiólogos es que la prevalencia se puede ver directamente, en cambio el RR muestra que el factor analizado no es responsable de la enfermedad, sino que debe ser otro factor el causante.

Situación G

Factor	Enfermo		
	Si	No	
Si	20	80	100
No	9	91	100
	29	171	200

RR = 2,2 y OR 95% CI (1,1 ; 4,6)

Situación H

Factor	Enfermo		
	Si	No	
Si	99	1	100
No	45	55	100
	144	56	200

RR = 2,2 y OR 95% CI (1,8 ; 2,7)

En los dos casos anteriores se deduce que hay asociación estadística entre el factor y la enfermedad porque RR = 2,2 y el valor RR = 1 no cae dentro del intervalo de confianza (lo mismo si se usa Chi). Pero no es lo mismo obtener 29 que 144 enfermos en 200 casos. Si la Situación G se presenta antes de cierto hecho, y la H después, entonces este hecho debe ser importante porque la prevalencia subió del 14,5% al 72%. Sin embargo, si se comparan los RR obtenidos en ambos casos, estadísticamente no se podrá detectar diferencia alguna y la influencia del hecho no podrá ser validada. Todo lo que hace RR es decirnos que hay asociación entre el factor y la enfermedad, pero no puede detectar un cambio en la prevalencia, a pesar de que se supone que estos estudios se diseñan para analizar la enfermedad. Nuevamente la pregunta es: ¿ Por cual motivo se usa el índice RR si se puede hacer lo mismo con el Chi-test ? La respuesta se la deja al lector.

La conclusión final con estos ocho ejemplos es que: *ni OR, ni RR tienen un verdadero significado epidemiológico, sino que solo sirven para ver la asociación estadística.*

15.8 Análisis de factores ocultos

Existen muchos factores que pueden desencadenar una enfermedad, por eso se habla de fenómenos multifactoriales. Cuando el problema se lo reduce a solo uno, como se hace en una Tabla de 2 x 2, puede ocurrir que ese único factor analizado no explique adecuadamente el fenómeno analizado, por ejemplo

Ejemplo 3) Otro investigador, al revisar el Ejemplo 1 del punto anterior, sospecha que en realidad hay un factor oculto que no se tuvo en cuenta en el caso. Piensa que el tiempo de gestación debió haber sido considerado porque no es lo mismo un nacimiento prematuro que uno normal. Para verificar sus sospechas reorganiza los datos como se muestra a continuación, donde arma dos tablas de contingencia, modificando los factores. Su conclusión ante estos datos es que la asociación entre bacteriemia y cateterización es *aparente*. En realidad, hay una muy fuerte relación entre bacteriemia y tiempo de gestación (OR = 75), y a su vez el tiempo de gestación está muy relacionado (OR = 5) con las cateterizaciones. Su planteo es algo como:

Tiempo de gestación → Cateterización → Bacteriemia
 En lugar del primer ejemplo donde: Cateterización → Bacteriemia

		Bacteriemia		Total	OR = $\frac{150 \cdot 150}{10 \cdot 30}$
		SÍ	NO		
Prematuro:	SÍ	150	30	180	OR = 75,0
	NO	10	150	160	
Total		160	180	340	

		Cateterización		Total	OR = $\frac{150 \cdot 160}{180 \cdot 80}$
		SÍ	NO		
Prematuro:	SÍ	150	30	180	OR = 5,0
	NO	80	80	160	
Total		230	110	340	

Aplicando el G-test a los dos casos anteriores resulta para el primero, entre Bacteriemia y Tiempo de gestación, un valor $G = 232,11^{***}$. Mientras que para el segundo, entre Cateterización y Tiempo de gestación, es $G = 43,821^{***}$. En correspondencia con los valores de OR respectivos, pero con demasiada evidencia para mostrar la relación.

Ejemplo 4) Un tercer investigador sospecha que no hay relación entre Cateterización y Bacteriemia, sino que es aparente. Hay una especie de enmascaramiento entre ambas por la influencia del factor Tiempo de gestación. Entonces, para remover tal influencia reorganiza los datos de la manera siguiente:

Cateterización en Prematuros:	Bacteriemia		Total	OR = $\frac{127 \cdot 7}{23 \cdot 23}$
	SÍ	NO		
SÍ	127	23	150	OR = 1,68
NO	23	7	30	
Total	150	30	180	

Cateterización en no premat.:	Bacteriemia		Total	OR = $\frac{5 \cdot 75}{5 \cdot 75}$
	SÍ	NO		
SÍ	5	75	80	OR = 1
NO	5	75	80	
Total	10	150	160	

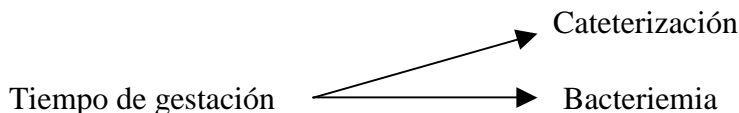
Al examinar los nuevos valores de OR cuando se remueve el factor Tiempo de gestación, se ve una caída brusca de los mismos hasta un valor $OR = 1$ que muestra la independencia de los factores en el caso de los nacimientos en tiempo normal. En efecto, si se piensa en una tabla de independencia, los valores de las frecuencias esperadas de OR se calculan con:

		Enfermedad y/o Contagio		Total
		SÍ	NO	
Expuesta:	SÍ	$(a+b)(a+c)/N$	$(a+c)(c+d)/N$	$a + c$
	NO	$(a+b)(b+d)/N$	$(b+d)(c+d)/N$	$b + d$
Total		$a + b$	$c + d$	N

Luego, cuando los factores son independientes resulta:

$$\mu (OR) = [(a+b)(a+c) \cdot (b+d)(c+d)] / [(a+b)(b+d) \cdot (a+c)(c+d)] = 1$$

Si se examinan las dos tablas de más arriba con el G-test, resulta para la de arriba $G_{corr.} = 1,035$ y para la de abajo $G = 0$, ambos valores no significativos. Por lo que no se puede demostrar una asociación entre el factor Cateterización y Bacteriemia. Su planteo es de esta forma:



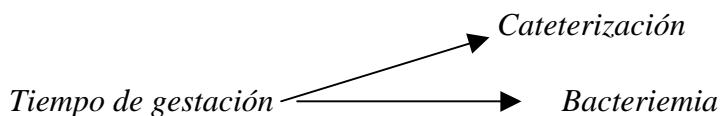
Hay varias cosas para destacar en estos ejemplos (1), (3) y (4). Lo más notable es que el desarrollo del último ejemplo demuestra el análisis incorrectamente efectuado en el primero. La relación real es entre los factores Tiempo de gestación y Bacteriemia. La Cateterización es independiente de la Bacteriemia cuando se remueve el Tiempo de gestación. Lo segundo es que cuando $OR = 1$, entonces el G se anula. Lo tercero es que a medida que OR crece, mucho más rápido crece G. Por último, si se tienen en cuenta los valores críticos de la Chi-cuadrado con un grado de libertad, se deduce que cuando $G \geq 3,841$ hay diferencia significativa y se puede probar asociación entre los dos factores analizados. El problema final es para cuál valor de OR el estadígrafo G comienza a ser significativo. Si se tuviese tal valor, se dispondría de un valor crítico para OR. La conclusión final es:

Se debe tener mucho cuidado en la elección de los factores para evitar enmascaramientos.

Los estadígrafos OR y RR se emplean como números índices, más que para armar un test de hipótesis del tipo G-test o χ^2 -test. Por eso, se estila armar intervalos de confianza de los mismos. De esta manera se puede cuantificar la relación entre los factores de riesgo. Luego, con ver si el valor unitario cae dentro del intervalo, se tiene el test respectivo.

15.9 Factores encajados o jerárquicos

A veces puede ocurrir que un factor de riesgo tenga mucha más importancia para una enfermedad que otro factor incidente. Como una especie de "categorización" o jerarquía de los factores. Imaginando, es como que uno está dentro de otro. Cuando esto ocurra se puede tener la tentación de quitar al menos importante a fin de ver mejor las cosas. En el ejemplo 3, de más arriba, el investigador separa el factor cateterización en dos partes para compararlo con la Bacteriemia, y así no puede comprobar la asociación entre estos dos factores. Su conclusión es que ambos factores son independientes y los separa. Pero como en el ejemplo 2, del mismo punto, otro investigador probó una asociación entre Tiempo de gestación y Cateterización, hace un planteo como:



El problema es que al desglosar los datos para forzarlos a entrar en una tabla de 2x2, el tamaño muestral se achica y eso va contra la primer regla de oro estadística. En un caso como el visto en los ejemplos, con datos hipotéticos donde las significaciones son tan extremas, no parece ser de importancia. Pero en casos reales donde no se tenga la suerte de encontrar valores muy grandes de G, para rechazar hipótesis, el disminuir N puede ser de importancia.

Si se miran los tres ejemplos mencionados con el OR, se puede notar que el máximo valor de $OR = 75$ se da para la combinación de Bacteriemia (B) con Tiempo de gestación (T). El factor Cateterización (C) con (B), tiene un $OR = 3,47$ y con (T) es de $OR = 5$. Mientras que al subdividir la muestra para separar el factor (T) y poder hacer dos comparaciones de (C) y (T), en ambos casos el OR es prácticamente uno. Es claro en el primer ejemplo que la *causa* es (C) y el *efecto* es (B). Cuando se hace la segunda comparación la *causa* es (T) y el *efecto* es (B). En la tercera se plantea como causa (T) y como efecto (C) lo que ya no es tan claro. Otro tanto ocurre en los dos últimos ejemplos. Si hubiera más factores todo sería más confuso. Por lo tanto, conviene ver las cosas de una manera diferente, antes que con la perspectiva de una tabla 2x2. En principio, la causa original, el factor de riesgo real es la cesárea. Esto se subdivide en dos casos: parto prematuro y parto normal (factor T). Por otra parte, también se puede subdividir en otros dos casos, teniendo en cuenta si fue necesario o no efectuar una cateterización. Se conforma así una partición del universo en cuatro casos mutuamente excluyentes entre sí. Si se efectúa una Tabla 2x2 con ellos, se encuentra significación -segunda tabla del ejemplo 2: (T) con (C). Pero parecería mejor imaginar a estos dos factores como uno encajado en el otro. Uno de primer nivel, en este ejemplo sería (T) porque produce mayor significación, y el otro (C) en un segundo nivel (subfactor). Entonces, se podría presentar una Tabla 2x4 para representar mejor las cosas como se muestra a continuación en el ejemplo siguiente:

Ejemplo) Usando los datos de los ejemplos anteriores de Bacteriemia en neo-natos, otro investigador decide presentar los datos con un modelo encajado de tablas y analizar los datos:

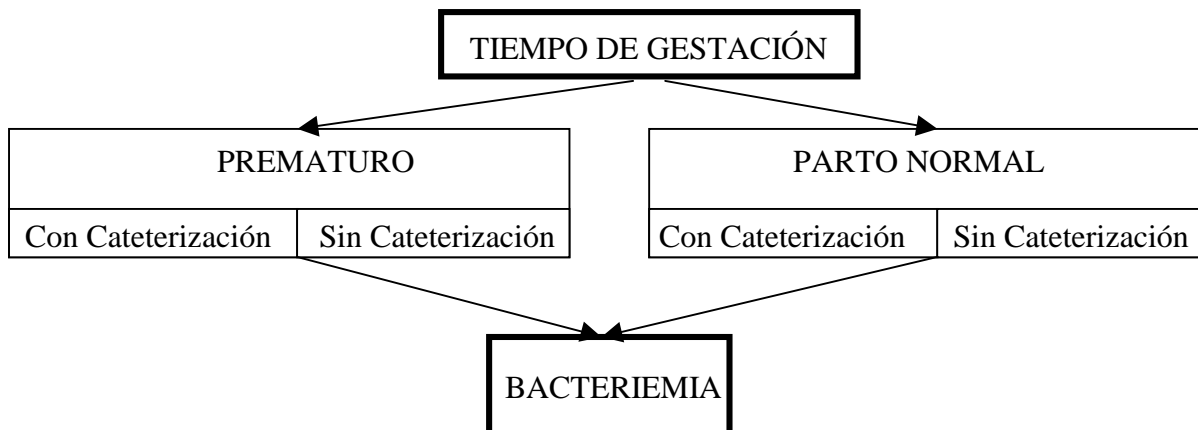
Frecuencias Observadas			Oij		BACTERIEMIA		Total
					SI	NO	
TIEMPO de GESTACION	Prematuro	Cateteriz.	SÍ	127	23	150	
			NO	23	7	30	
	Normal	Cateteriz.	SÍ	5	75	80	
			NO	5	75	80	
Total			160	180	340		

Frecuencias Esperadas			Eij		Total	
					SI	NO
TIEMPO de GESTACION	Prematuro	Cateteriz.	SÍ	70,59	79,41	150
			NO	14,12	15,88	30
	Normal	Cateteriz.	SI	37,65	42,35	80
			NO	37,65	42,35	80
Total			160	180	340	

Cálculo del estadígrafo G		
$G_{ij} = 2 O_{ij} \ln(O_{ij}/E_{ij})$		
j = 1	j = 2	i
149,18	-57,001	1
22,4512	-11,4702	2
-20,1882	85,7175	3
-20,1882	85,7175	4
Gj		
131,255	102,964	
$G = G_1 + G_2 = 234,219$		

Al efectuar una tabla de 2x4 con todos los datos obtenidos, usando una clasificación jerárquica, el resultado final es $G = 234,22^{***}$ que es el máximo encontrado. En la tabla 2x2 hecha en el ejemplo 2, con los mismos factores (T) y (B) el valor era ligeramente menor $G = 232$. Notar que la tabla de 4x2 es una yuxtaposición de las dos tablas no significativas, vistas en el ejemplo 3. Con todos los datos juntos se tiene una mejor visualización del problema.

El esquema final de la interacción de los factores se puede presentar con:



Y mirando todos los caminos posibles para ir de T a B, el de mayor riesgo es aquel que combina: Prematuro- Con Cateterización, para contraer la enfermedad. Su contribución particular al valor total del estadígrafo G es netamente superior a los demás.

15.10 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|---|---|
| 1) El análisis de frecuencias se puede aplicar a cualquier magnitud, dividida en clases. | V | F |
| 2) La validación estadística se puede hacer con: | | |
| 3) El G-test es mejor que la prueba con Chi-cuadrado porque es más sensible. | V | F |
| 4) El G-test es mejor que la prueba con Chi-cuadrado porque es menos robusto. | V | F |
| 5) La expresión matemática para calcular χ^2 es: | | |
| 6) Las hipótesis extrínsecas toman en cuenta el número de parámetros poblacionales r. | V | F |
| 7) Las hipótesis intrínsecas hacen que los grados de libertad sean: n-1. | V | F |
| 8) La corrección de Yates se debe hacer para corregir la continuidad. | V | F |
| 9) Cuando se usa una distribución continua, para testear una variable discreta hay que hacer: | | |
| 10) El G-test se llama test de la razón de verosimilitudes (<i>likelihood</i>). | V | F |
| 11) El cociente de frecuencias es el cociente de verosimilitud. | V | F |
| 12) La distribución teórica de G es la probabilidad multinomial. | V | F |
| 13) No hay que hacer distinciones en G entre hipótesis intrínsecas y extrínsecas. | V | F |
| 14) La función distribución para efectuar el G-test es la Chi-cuadrado. | V | F |
| 15) Las tablas de contingencia se usan para validar la independencia de los factores. | V | F |
| 16) Para calcular el número de grados de libertad en una tabla se usa: (h-1)(k-1)-m | V | F |
| 17) Para el cálculo de frecuencias esperadas en una tabla de contingencia se usa el concepto | | |
| 18) Una frecuencia esperada E_{ij} en una tabla se calcula con: | | |
| 19) Una tabla 2x2 puede ser clasificada en tres modelos diferentes. | V | F |
| 20) La clasificación se basa en cómo fija el investigador el tamaño muestral. | V | F |
| 21) Para un Modelo I y II conviene usar el test exacto de Fisher. | V | F |
| 22) Para corregir el G-test por continuidad se usa la y se calcula con | | |
| 23) Los índices: RR y OR se obtienen de una tabla cualquiera de 2x2. | V | F |
| 24) El RR es un cociente de probabilidades entre los expuestos y los no expuestos. | V | F |
| 25) Si $OR = 1$ los factores analizados son independientes. | V | F |

- 26) El mejor diseño experimental para determinar OR es el “doble-ciego”. **V F**
 27) La eficacia se calcula con el OR. **V F**
 28) Si el OR crece, aumenta el valor del estadígrafo G. **V F**
 29) Cuando $OR = 1$ resulta $G = 0$. **V F**
 30) El OR permite comparar dos factores con validación para ver si hay relación entre ellos. **V F**
 31) Se debe tener cuidado de evitar enmascaramientos. **V F**
 32) Para estudiar la concordancia entre dos métodos clínicos las muestras son independientes **V F**
 33) Cuando se realizan dos mediciones a cada individuo se puede usar el test de Student **V F**

2) En tres turnos de producción en una fábrica: Mañana, Tarde y Noche, se producen accidentes laborales. De acuerdo con los datos siguientes, probar la H_0 que la proporción de accidentes es la misma para los tres turnos laborales.

	Mañana	Tarde	Noche	Total
Empleados	150	180	120	450
Accidentes	5	6	5	16
Total	155	186	125	466

3) Un investigador midió la forma en que se repartían las visitas de ratones a trampas tratadas previamente con orina de ratones, con otras limpias de olor. Probar que el tratamiento efectuado a las trampas no afecta la frecuencia de visita.

	Limpias	Con orina	Total
Visitadas	3	17	20
No visitadas	9	23	32
Total	12	40	52

Datos de Perrota, C.A. Un. Wisconsin, 1954.

4) Un total de 130 ratones fue seleccionado para probar la penicilina, luego de haber sido inyectados con un *staphylococcus ureus* cultivado en un caldo enriquecido. La tabla de supervivencia obtenida fue la siguiente. Verifica si la penicilina tuvo incidencia en la cura.

	Vivos	Muertos	Total
Control	8	12	20
Penicilina	48	62	110
Total	56	74	130

5) Luego de dos años de investigaciones sobre una nueva medicación y tratamiento para un tipo de cáncer, el grupo obtuvo los datos de 80 pacientes elegidos:

	Sin Tratar	Tratados	Total
Sin tratar	11	29	40
Tratados	20	20	40
Total	31	49	80

Hace la prueba de Chi- cuadrado y no le da significativa si hace la corrección de Yates. Se le ocurren tres opciones:

- 1) Presentar el informe sin la corrección.
- 2) Mentir modificando los datos, bajando de 11 a 10 para hacerla significativa.
- 3) Usar el G-test y ver qué pasa.

Decidir cuál de las opciones es la que debería elegir y dar las razones para la misma.

6) Para verificar la asociación entre dos métodos A y B se realizaron 200 mediciones y los resultados fueron los siguientes:

Método B	Método A		Total
	(+)	(-)	
(+)	80	15	95
(-)	10	95	105
Total	90	110	200

Aplicar el concepto de independencia: Chi-cuadrado y G-test para ver si están asociados.

7) Lo mismo que el problema anterior pero con los datos siguientes:

Método B	Método A		Total
	(+)	(-)	
(+)	9978	22	10000
(-)	10	9990	10000
Total	9988	10012	20000

8) Lo mismo que antes pero con los datos

Método B	Método A		Total
	(+)	(-)	
(+)	10	95	105
(-)	80	15	95
Total	90	110	200

Comparar estos resultados con los del problema (6) anterior.

9) Aplicar el Modelo de Cochran (Q-test) para el caso especial de tablas de 2x2 con la simplificación de Conover (1999). Usando la relación

$$Q = (c-d)^2 / (c+d) \text{ versus el valor crítico de Chi-cuadrado } 3,841 \text{ para } 1 \text{ grado de libertad}$$

Aplicar este test a los problemas (6), (7) y (8) y comparar con las conclusiones anteriores

Nota: c y d son los valores de la diagonal donde los métodos no concuerdan.

10) Discutir el concepto del método de Cochran (ver tema 14) con los conceptos de independencia y asociación estadística vistos en este tema.

Aéndice 1: Reproducibilidad en casos binarios con el método de Guttman.

El estudio de la reproducibilidad en casos binarios se realiza con un diseño experimental llamado test-retest. Consiste en medir dos veces al mismo individuo, con el mismo método clínico en magnitudes binarias. Los resultados se pueden presentar en una tabla como sigue:

Retest	Test		Total
	(+)	(-)	
(+)	a	b	a + b
(-)	c	d	c + d
Total	a + c	b + d	N

N individuos al azar fueron sometidos al mismo test dos veces. Los resultados se agrupan según si hay concordancias (a y d) o si hay discordancias (c y b) en el estudio. Una perfecta reproducibilidad implica que c y b son nulos.

Para analizar este tipo de estudio en forma estadística se desarrolló el método de Guttman (1947) para ciencias humanas. Allí los tests son cuestionarios y las respuestas son generalmente del tipo verdadero/falso. No se espera que los cuestionarios varíen entre una y otra aplicación, salvo alguna pregunta redactada en forma engañosa. Por lo tanto, la variabilidad encontrada en el estudio es atribuida al sujeto analizado. Sin embargo, en análisis clínicos se puede suponer a la muestra constante (adoptando las precauciones necesarias para que el analito no varíe entre el test y el retest); y por eso la variabilidad se puede atribuir a la inestabilidad del método clínico. Las causas de variabilidad pueden ser: fluctuación de los instrumentos, diferentes reacciones químicas, cambios en los observadores, etc. Este método consiste en calcular dos límites, dentro de los cuales se espera encontrar el porcentaje de reproducibilidad del método.

Límite inferior = $(k/k-1) (F-1/k) = 2F - 1$

Límite superior = $(k/k-1) (h - 1/k) = 2h - 1$

Donde: F es la suma de los valores máximos de cada columna (suma de frecuencias relativas).

G es la suma de las frecuencias relativas de la diagonal principal: $[(a + d)/N]$

$h = (1/k) \{ 1 + \sqrt{(k-1)(kG-1)} \} = (1/2) \{ 1 + \sqrt{(2G-1)} \}$

k = 2 (porque hay dos clases en cada resultado posible (positivo y negativo))

Para ilustrar el procedimiento se presenta el siguiente ejemplo:

Ejemplo: Se aplicó la determinación de VDRL a 872 pacientes dos veces seguidas. A cada muestra del paciente se la dividió en dos alícuotas, usando la primera para el test y la segunda para el retest. Los resultados se muestran a continuación:

Retest	Test		Total
	(+)	(-)	
(+)	400	12	412
(-)	50	410	460
Total	450	422	872

Los valores máximos de cada columna sumados son:
 $F = (400 + 410) / 872 = 0,929$
 $G = (400 + 410) / 872 = 0,929$ luego será
 $h = (1/2) \{ 1 + \sqrt{(2 \times 0,929 - 1)} \} = 0,963$
 Límite inferior = $[2 (0,929)] - 1 = 0,858$
 Límite superior = $[2 (0,963)] - 1 = 0,926$

Donde la reproducibilidad se halla entre el 85,8% y el 92,6%, valor considerado como muy bueno. Por lo tanto se concluye que la reproducibilidad de la VDRL empleada es aceptable desde un punto de vista clínico.