



Facultad de Matemáticas

Departamento de Estadística e
Investigación Operativa

— TRABAJO FIN DE GRADO —

Tablas de Contingencia

Inmaculada Millán Díaz
Sevilla, Septiembre 2017

Abstract

Contingency tables show classifications of individuals under the possible combinations of the levels of two or more categorical variables. The analysis of these tables is fundamental to solve questions about the dependence or not of the variables observed, the measure of the association between both variables or the homogeneity of the populations.

The objective of this work is the study of the basic models for the analysis of contingency tables, as well as the practical implementation of these techniques in R.

Índice general

1. Introducción: Variables Categóricas	7
1.1. Clasificación de las Variables Categóricas	7
1.2. Distribuciones para Datos Categóricos	9
1.3. Inferencia Estadística para Datos Categóricos	10
1.3.1. Inferencia Estadística para Parámetros Binomiales	13
1.3.2. Inferencia Estadística para Parámetros Multinomiales	15
2. Descripción de Tablas de Contingencia	19
2.1. Estructura de Probabilidad para Tablas de Contingencia	19
2.2. Tipos de Estudios	27
2.3. Tablas Estratificadas	31
3. Inferencia para Tablas de Contingencia	41
3.1. Intervalos de Confianza	41
3.2. Test de Independencia en Tablas de Doble Entrada	45
3.3. Test de Independencia para Muestras Pequeñas	51
4. Tablas de Contingencia y R	57

Capítulo 1

Introducción: Variables Categóricas

Una *variable categórica* tiene una escala de medida que consiste en un conjunto de categorías. Por ejemplo, la ideología política, que a menudo se mide como liberal, moderada o conservadora.

El desarrollo de métodos para variables categóricas fue estimulado por estudios de investigación en las ciencias sociales y biomédicas, aunque no están restringidas a estas áreas. Aparecen de forma frecuente en las ciencias del comportamiento (ej. tipo de enfermedad mental, con las categorías de esquizofrenia, depresión y neurosis), genética (ej. tipo de alelo heredado por una descendencia) o educación (ej. respuesta de los estudiantes frente a una pregunta de un examen, con las categorías de correcto e incorrecto). Ocurren en otros campos como las ciencias de la ingeniería o el control de calidad industrial.

1.1. Clasificación de las Variables Categóricas

Podemos distinguir las variables categóricas entre *variables de respuesta* y *variables explicativas*. Por ejemplo, los modelos de regresión describen como varía la variable de respuesta frente a los distintos valores de la variable explicativa. Como ocurre con el precio de una vivienda según sus metros cuadrados o su localización.

Las variables categóricas tienen dos tipos principales de escala. Las varia-

bles que tienen categorías sin un ordenamiento natural se llaman *nominales*. Como por ejemplo la afiliación religiosa (Católica, Protestante, Judía, Musulmana, . . .), el modo de transporte (coche, bicicleta, autobús, metro, . . .) o el tipo favorito de música (clásica, jazz, pop, rock, . . .). Para variables nominales, el orden de las categorías es irrelevante. El análisis estadístico no depende de dicho ordenamiento.

Muchas variables categóricas tienen categorías ordenadas. Éstas variables son llamadas *ordinales*. Algunos ejemplos son el tamaño de un coche (subcompacto, compacto, mediano o grande), la clase social (alta, media o baja) o la condición de un paciente (buena, normal o grave). Las variables ordinales tienen categorías ordenadas, pero las distancias entre las distintas categorías son desconocidas. Aunque una persona categorizada como moderada es más liberal que una persona categorizada como conservadora, ningún valor numérico describe cuánto más liberal es una persona.

Una *variable intervalo* es aquella que tiene distancias numéricas entre dos valores cualesquiera. Por ejemplo, el nivel de presión arterial, los ingresos anuales o el tiempo de vida funcional de un electrodoméstico.

La forma en la que se mide una variable determina su clasificación. Por ejemplo, “educación” es una variable nominal cuando se mide sólo como escuela privada o pública; es una variable ordinal cuando se mide por el grado más alto alcanzado, usando las categorías de ninguno, escuela secundaria, grado, máster o doctorado; o es una variable intervalo cuando se mide por el número de años estudiados, usando los números enteros 0,1,2, . . .

Las variables se clasifican como *continuas* o *discretas* según el número de valores que puedan tomar. La medida real de todas las variables se produce de forma discreta, debido a las limitaciones de precisión en los instrumentos de medida. La clasificación continua-discreta, en la práctica, distingue entre variables que toman muchos valores y variables que toman pocos valores. Por ejemplo, los estadísticos a menudo tratan variables discretas que tienen un gran número de valores (tales como las puntuaciones de un examen) como variables continuas.

Las variables nominales son *cualitativas* —las distintas categorías difieren en cualidad, no en cantidad. Las variables intervalo son *cuantitativas* —los distintos niveles tienen diferentes características de interés. La posición de las variables ordinales es difusa en esta clasificación aunque se suelen clasificar como cuantitativas asignando puntuaciones numéricas a las distintas categorías.

1.2. Distribuciones para Datos Categóricos

Los análisis de datos requieren suposiciones sobre el mecanismo aleatorio que genera los datos. Para los modelos de regresión con respuestas continuas, la distribución normal juega el papel central. Ahora analizaremos las distribuciones clave para respuestas categóricas: *binomial*, *multinomial* y de *Poisson*.

1. Distribución Binomial:

Tenemos un número n de observaciones binarias. Sean y_1, y_2, \dots, y_n las respuestas para n ensayos independientes e idénticamente distribuidos tales que $P(Y_i = 1) = \pi$ y $P(Y_i = 0) = 1 - \pi$. Usamos la etiquetas genéricas “éxito” y “fracaso” para las respuestas 1 y 0 respectivamente. Ensayos idénticamente distribuidos significa que la probabilidad de éxito π es la misma para cada ensayo. Ensayos independientes significa que los $\{Y_i\}$ son variables aleatorias independientes que a menudo se llaman ensayos *Bernoulli*. El número total de éxitos, $Y = \sum_{i=1}^n Y_i$, tiene la distribución *binomial* con índice n y parámetro π .

La función de masa de probabilidad es:

$$p(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad y = 0, 1, 2, \dots, n$$

Se cumple que $E(Y_i) = E(Y_i^2) = 1 \times \pi + 0 \times (1 - \pi) = \pi$ y $\text{var}(Y_i) = \pi(1 - \pi)$. La distribución para $Y = \sum_i Y_i$ tiene media $\mu = E(Y) = n\pi$ y varianza $\sigma^2 = \text{var}(Y) = n\pi(1 - \pi)$.

2. Distribución Multinomial:

Supongamos ahora que cada uno de los n ensayos independientes e idénticamente distribuidos puede tener resultados en cualquiera de c categorías. Sea $y_{ij} = 1$ si el ensayo i tiene resultado en la categoría j y $y_{ij} = 0$ en caso contrario. Entonces el vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ic})$ representa un ensayo multinomial con $\sum_j y_{ij} = 1$. Nótese que y_{ic} es redundante, siendo linealmente dependiente de los otros. Dado $n_j = \sum_i y_{ij}$ representa el número de ensayos que tienen resultado en la categoría j . Los conteos (n_1, n_2, \dots, n_c) tienen la distribución multinomial.

Si $\pi_j = P(Y_{ij} = 1)$ denota la probabilidad de éxito en la categoría j para cada ensayo, la función de masa de probabilidad multinomial es:

$$p(n_1, n_2, \dots, n_{c-1}) = \left(\frac{n!}{n_1! n_2! \dots n_{c-1}!} \right) \pi_1^{n_1} \pi_2^{n_2} \dots \pi_c^{n_c} \quad (1.1)$$

Como $E(n_j) = n\pi_j$ y $\text{var}(n_j) = n\pi_j(1 - \pi_j)$ la distribución marginal de cada n_j es una binomial con parámetros n y π_j . Estas variables no son independientes puesto que $\text{cov}(n_i, n_j) = -n\pi_i\pi_j$.

3. Distribución de Poisson:

A veces los datos de recuento no resultan de un número fijo de ensayos. Por ejemplo, si $y =$ número de muertes debidas a un accidente automovilístico en Italia durante la próxima semana, no hay límite superior n fijo para y . Dado que y debe ser un número entero no negativo, su distribución debe colocar su masa en ese rango. La distribución más simple es la de Poisson. Sus probabilidades dependen de un sólo parámetro, la media μ . La función de masa de probabilidad de Poisson es:

$$p(y) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots$$

Satisface $E(Y) = \text{var}(Y) = \mu$. La distribución de Poisson se utiliza para los conteos de sucesos que ocurren aleatoriamente en el tiempo o en el espacio, cuando los resultados en periodos o regiones disjuntas son independientes.

1.3. Inferencia Estadística para Datos Categóricos

La elección de la distribución para la variable respuesta es sólo un paso del análisis de los datos. En la práctica, esta distribución tiene los valores de los parámetros desconocidos.

En esta sección veremos métodos para hacer inferencia sobre los parámetros.

Función de Probabilidad y Estimador de Máxima Verosimilitud

Para la estimación de los parámetros usaremos la *máxima verosimilitud*. Bajo condiciones de regularidad débil los estimadores de máxima verosimilitud tie-

nen propiedades deseables. Tienen distribuciones normales para grandes muestras, son asintóticamente consistentes, convergiendo al parámetro cuando n aumenta, y son asintóticamente eficientes, produciendo errores estándar para muestras grandes no mucho más grandes que los de los otros métodos de estimación.

Dado los datos, para una distribución de probabilidad elegida, la *función de verosimilitud* es la probabilidad de esos datos tratados como una función del parámetro desconocido. El estimador de máxima verosimilitud es el valor del parámetro que maximiza esta función. Este es el valor del parámetro bajo el cual los datos observados tienen la mayor probabilidad de ocurrencia. El valor del parámetro que maximiza la función de verosimilitud también maximiza el logaritmo de esa función. Es más fácil maximizar el logaritmo de verosimilitud ya que es una suma en lugar de un producto de términos.

Al parámetro lo denotamos por β y a su estimador de máxima verosimilitud por $\hat{\beta}$. La función de verosimilitud es $\ell(\beta)$ y el logaritmo de la función de verosimilitud es $L(\beta) = \log[\ell(\beta)]$.

Para muchos modelos $L(\beta)$ tiene una forma cóncava y $\hat{\beta}$ es el punto en el cual la derivada es igual a 0. El estimador de máxima verosimilitud es entonces la solución de la ecuación $\partial L(\beta)/\partial \beta = 0$.

Cuando β es multidimensional se denota por $\boldsymbol{\beta}$, y $\hat{\boldsymbol{\beta}}$ es la solución de un conjunto de ecuaciones de verosimilitud.

Dado SE (standard error) denota el error estándar de $\hat{\beta}$, y $\text{cov}(\hat{\boldsymbol{\beta}})$ la matriz de covarianza asintótica de $\hat{\boldsymbol{\beta}}$. Bajo condiciones de regularidad, $\text{cov}(\hat{\boldsymbol{\beta}})$ es la inversa de la matriz de información.

El elemento (j,k) de la matriz de información es:

$$- E \left(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} \right) \quad (1.2)$$

Los errores estándar son las raíces cuadradas de los elementos diagonales de la inversa de la matriz de información. Cuanto mayor es la curvatura del logaritmo de la verosimilitud, menores serán los errores estándar. Es razonable dado que la gran curvatura implica que el logaritmo de la verosimilitud disminuye rápidamente a medida que $\boldsymbol{\beta}$ se aleja de $\hat{\boldsymbol{\beta}}$; por lo tanto los valores tendrían más probabilidades de ocurrir si $\boldsymbol{\beta}$ toma un valor cercano a $\hat{\boldsymbol{\beta}}$.

Contrastes de Wald y de Razón de Verosimilitud

Estos tests usan la función de verosimilitud para hacer inferencia. Las introducimos para la hipótesis nula $H_0 : \beta = \beta_0$ y entonces discutimos su relación con la estimación de intervalos.

Dado el error estándar SE no nulo de $\hat{\beta}$, el *estadístico de Wald* es:

$$z = \frac{\hat{\beta} - \beta_0}{\text{SE}}$$

que tiene una distribución normal estándar aproximada cuando $\beta = \beta_0$.

Su extensión multivariante para la prueba $H_0 : \beta = \beta_0$ es:

$$W = (\hat{\beta} - \beta_0)' [\text{cov}(\hat{\beta})]^{-1} (\hat{\beta} - \beta_0)$$

La covarianza no nula está basada en la curvatura (1.2) del logaritmo de la verosimilitud de $\hat{\beta}$. La distribución normal multivariante asintótica para $\hat{\beta}$ implica una distribución chi-cuadrado asintótica para W . Los grados de libertad coinciden con el rango de $\text{cov}(\hat{\beta})$, que es el número de parámetros no redundantes en β .

La prueba de razón de verosimilitudes usa la función de verosimilitud a través de la razón de dos maximizaciones: el máximo sobre los posibles valores del parámetro bajo H_0 , y el máximo bajo el conjunto mayor de valores del parámetro permitiendo H_0 y una alternativa H_a que sea cierta.

Si ℓ_0 denota el máximo valor de la función de verosimilitud bajo H_0 y ℓ_1 el máximo valor general, es decir, bajo $H_0 \cup H_a$, entonces ℓ_1 es siempre mayor que ℓ_0 , y por tanto la razón $\Lambda = \ell_0/\ell_1$ no puede exceder 1.

El *estadístico de razón de verosimilitudes* es igual a:

$$-2 \log \Lambda = -2 \log (\ell_0/\ell_1) = -2 (L_0 - L_1)$$

donde L_0 y L_1 denotan el logaritmo de las funciones de verosimilitud maximizadas.

El estadístico $-2 \log \Lambda$ tiene una distribución chi-cuadrado bajo la hipótesis nula cuando $n \rightarrow \infty$ cuyo grado de libertad es la diferencia entre las dimensiones de los espacios de parámetros $H_0 \cup H_a$ y H_0 .

Construcción de los Intervalos de Confianza

En la práctica es más informativo construir intervalos de confianza para los parámetros que probar hipótesis sobre sus valores. Un intervalo de confianza resulta de invertir un contraste de hipótesis. Por ejemplo, un intervalo de confianza al 95 % para β es el conjunto de β_0 para los cuales el contraste $H_0 : \beta = \beta_0$ tiene un p-valor mayor que 0.05.

Para $\alpha \in (0, 1)$, z_α denota el percentil $100(1 - \alpha)$ % de la distribución normal estándar, es decir, z_α verifica $P(Z > z_\alpha) = \alpha$, donde $Z \sim N(0,1)$. Los intervalos de confianza al $100(1 - \alpha)$ % basados en el uso de la normalidad asintótica usan $z_{\alpha/2}$. Por ejemplo, $z_{0.025} = 1.96$ para una confianza del 95 %.

De forma similar $\chi_{df}^2(\alpha)$ es el percentil $100(1 - \alpha)$ % de una distribución χ^2 con df grados de libertad.

El intervalo de confianza de Wald es el conjunto de β_0 para los cuales $|\hat{\beta} - \beta_0| / SE < z_{\alpha/2}$. Esto da el intervalo $\hat{\beta} \pm z_{\alpha/2}(SE)$.

El intervalo de confianza basado en la razón de verosimilitudes es el conjunto de β_0 para los cuales $-2(L(\beta_0) - L(\hat{\beta})) < \chi_1^2(\alpha)$. Recalcamos que $\chi_1^2(\alpha) = z_{\alpha/2}^2$.

Cuando $\hat{\beta}$ tiene una distribución normal el logaritmo de la función de verosimilitud tiene una forma parabólica. Para muestras pequeñas con datos categóricos, $\hat{\beta}$ puede estar lejos de la normalidad y el logaritmo de la función de verosimilitud puede estar lejos de una curva simétrica parabólica. Esto también puede ocurrir con muestras de moderadas a grandes cuando un modelo contiene muchos parámetros. En estos casos la inferencia basada en la normalidad asintótica de $\hat{\beta}$ puede tener un rendimiento inadecuado.

1.3.1. Inferencia Estadística para Parámetros Binomiales

En este apartado presentaremos pruebas e intervalos de confianza para el parámetro binomial π , basado en y éxitos en n ensayos independientes.

La función de masa de probabilidad era: $p(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$. Los coeficientes $\binom{n}{y}$ no influyen sobre donde ocurre el máximo con respecto a π . Un equivalente al logaritmo de la verosimilitud sería:

$$L(\pi) = \log[\pi^y(1 - \pi)^{n-y}] = y \log(\pi) + (n - y) \log(1 - \pi)$$

Derivando con respecto a π :

$$\frac{\partial L(\pi)}{\partial \pi} = \frac{y}{\pi} - \frac{n - y}{1 - \pi} = \frac{y - n\pi}{\pi(1 - \pi)}$$

Igualando esto a 0 obtenemos la ecuación de verosimilitud, cuya solución es el estimador de máxima verosimilitud $\hat{\pi} = y/n$, la proporción muestral de éxitos para los n ensayos.

Se tiene que:

$$E\left(\frac{\partial^2 L(\pi)}{\partial \pi^2}\right) = \frac{n}{\pi(1 - \pi)}$$

Por tanto, el error estándar de $\hat{\pi}$ es $\sigma(\hat{\pi}) = \sqrt{\frac{\pi(1 - \pi)}{n}}$.

Además, dado que $E(Y) = n\pi$ y $\hat{\pi} = Y/n$, la media es $E(\hat{\pi}) = \pi$.

Consideremos ahora $H_0 : \pi = \pi_0$. El estadístico de Wald es:

$$z_W = \frac{\hat{\pi} - \pi_0}{\text{SE}} = \frac{\hat{\pi} - \pi_0}{\sqrt{\hat{\pi}(1 - \hat{\pi})/n}}$$

Usa el error estándar evaluado en $\hat{\pi}$.

Una prueba de significación simplemente indica si un valor particular de π es probable. Es mejor usar un intervalo de confianza para determinar el rango de valores probables.

Invertir el estadístico de prueba de Wald da el intervalo de valores de π_0 para los cuales $|z_W| < z_{\alpha/2}$, es decir:

$$\hat{\pi} \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

Históricamente, este era uno de los primeros intervalos de confianza utilizados para cualquier parámetro. Desafortunadamente, funciona mal al menos que n sea muy grande. La probabilidad de cobertura real suele caer por debajo del coeficiente de confianza nominal, muy por debajo cuando el parámetro π está cerca de 0 ó 1.

Esto motiva un intervalo de Wald ajustado:

$$\tilde{\pi} \pm z_{\alpha/2} \sqrt{\frac{\tilde{\pi}(1 - \tilde{\pi})}{n^*}}, \quad \text{donde } n^* = n + z_{\alpha/2}^2.$$

que añade $\frac{1}{2}z_{\alpha/2}^2$ observaciones de cada tipo a la muestra.

1.3.2. Inferencia Estadística para Parámetros Multinomiales

Ahora presentamos la inferencia para parámetros multinomiales $\{\pi_j\}$. Como una función de $\{\pi_j\}$, la función de masa de probabilidad multinomial (1.1) es proporcional al núcleo:

$$\prod_j \pi_j^{n_j} \quad \text{donde } \pi_j \geq 0 \text{ y } \sum_j \pi_j = 1 \quad (1.3)$$

Los estimadores de máxima verosimilitud son los $\{\pi_j\}$ que maximizan (1.3).

El logaritmo de la función de verosimilitud multinomial es:

$$L(\boldsymbol{\pi}) = \sum_j n_j \log \pi_j$$

Para eliminar la redundancia tratamos a L como una función de $(\pi_1, \dots, \pi_{c-1})$, ya que $\pi_c = 1 - (\pi_1 + \dots + \pi_{c-1})$. Por tanto tenemos que $\partial \pi_c / \partial \pi_j = -1$, para $j = 1, \dots, c - 1$.

Entonces:

$$\frac{\partial \log \pi_c}{\partial \pi_j} = \frac{1}{\pi_c} \frac{\partial \pi_c}{\partial \pi_j} = -\frac{1}{\pi_c},$$

derivando $L(\boldsymbol{\pi})$ con respecto a π_j da la ecuación de verosimilitud:

$$\frac{\partial L(\boldsymbol{\pi})}{\partial \pi_j} = \frac{n_j}{\pi_j} - \frac{n_c}{\pi_c} = 0$$

La solución de máxima verosimilitud satisface $\hat{\pi}_j/\hat{\pi}_c = n_j/n_c$. Ahora,

$$\sum_j \hat{\pi}_j = 1 = \frac{\hat{\pi}_c \left(\sum_j n_j \right)}{n_c} = \frac{\hat{\pi}_c n}{n_c}$$

Así que $\hat{\pi}_c = n_c/n$ y entonces $\hat{\pi}_j = n_j/n \forall j$. Los estimadores de máxima verosimilitud de $\{\pi_j\}$ son las proporciones de la muestra.

Estadístico Chi-Cuadrado de Pearson

La prueba de Pearson evalúa si los parámetros multinomiales son iguales a ciertos valores específicos. Consideramos $H_0 : \pi_j = \pi_{j0}, j = 1, \dots, c$, donde $\sum_j \pi_{j0} = 1$. Cuando H_0 es cierto, los valores esperados de $\{n_j\}$, llamadas *frecuencias esperadas*, son $\mu_j = n\pi_{j0}, j = 1, \dots, c$.

Pearson propuso el estadístico:

$$X^2 = \sum_j \frac{(n_j - \mu_j)^2}{\mu_j}$$

Diferencias mayores de $\{n_j - \mu_j\}$ producen valores grandes de X^2 .

Si X_0^2 denota el valor observado de X^2 , el p-valor es el valor bajo la hipótesis nula de $P(X^2 \geq X_0^2)$. Esto es igual a la suma de las probabilidades multinomiales de todos los vectores de conteo con $X^2 \geq X_0^2$.

Para muestras grandes, X^2 tiene aproximadamente una distribución chi-cuadrado con $c - 1$ grados de libertad. El p-valor está aproximado por $P(\chi_{c-1}^2 \geq X_0^2)$, donde χ_{c-1}^2 denota una variable aleatoria chi-cuadrado con $c - 1$ grados de libertad.

Estadístico Chi-Cuadrado de Razón de Verosimilitudes

Una prueba alternativa para parámetros multinomiales usa el test de razón de verosimilitudes. El núcleo de la verosimilitud multinomial es (1.3). Bajo H_0 la verosimilitud es máxima cuando $\hat{\pi}_j = \pi_{j0}$. En un caso general, el máximo se obtiene cuando $\hat{\pi}_j = n_j/n$. La razón de verosimilitudes es igual a:

$$\Lambda = \frac{\prod_j (\pi_{j0})^{n_j}}{\prod_j (n_j/n)^{n_j}}$$

El estadístico de razón de verosimilitudes es:

$$G^2 = -2 \log \Lambda = 2 \sum n_j \log (n_j/n\pi_{j0})$$

Cuanto mayor sea el valor de G^2 , mayor será la evidencia contra H_0 .

En el caso general, el espacio de parámetros está formado por $\{\pi_j\}$ sujeto a $\sum_j \pi_j = 1$ con dimensionalidad $c-1$. Para n grande, G^2 tiene aproximadamente una distribución chi-cuadrado con $c-1$ grados de libertad.

Cuando H_0 es cierta, X^2 y G^2 tienen distribuciones asintóticas chi-cuadrado con $c-1$ grados de libertad. De hecho, son asintóticamente equivalentes ya que $X^2 - G^2$ converge en probabilidad a cero. Cuando H_0 es falsa, tienden a crecer proporcionalmente a n .

Para c fijo, a medida que n aumenta, la distribución de X^2 converge a una chi-cuadrado más rápidamente que la de G^2 . La aproximación chi-cuadrado es generalmente pobre para G^2 cuando $n/c < 5$. Cuando c es grande, puede ser decente para X^2 para n/c tan pequeño como 1 si la tabla no contiene frecuencias esperadas muy pequeñas y moderadamente grandes.

Capítulo 2

Descripción de Tablas de Contingencia

En este capítulo hablaremos de tablas que muestran relaciones entre variables categóricas. También definiremos parámetros que resumen su asociación. Nos centraremos en variables binarias, las cuales tienen dos categorías.

Primero, introduzcamos una terminología básica y notación.

2.1. Estructura de Probabilidad para Tablas de Contingencia

Una tabla de contingencia es una de las formas más comunes de resumir datos categóricos. En general, el interés se centra en estudiar si existe alguna asociación entre una variable *fila* y otra variable *columna* y/o calcular la intensidad de dicha asociación.

Sean X e Y dos variables categóricas de respuesta, X con I categorías e Y con J categorías. Un sujeto puede venir clasificado en una de las $I \times J$ categorías, que es el número posible de categorías que existe. Las respuestas (X, Y) de un sujeto elegido aleatoriamente de alguna población tiene una distribución de probabilidad. Una tabla rectangular que tiene I filas para las categorías de X y J columnas para las categorías de Y muestra esta distribución.

Cuando las casillas de la tabla contienen las frecuencias observadas, la tabla se

denomina *tabla de contingencia*, término que fue introducido por Karl Pearson en 1904.

Una tabla de contingencia (o *tabla de clasificación cruzada*), con I filas y J columnas se denomina tabla $I \times J$.

Por ejemplo, se considera la distribución conjunta de dos variables y la correspondiente tabla de contingencia en una muestra de pacientes de un hospital. Se tiene la siguiente tabla:

X : Se toma aspirina o placebo ($I=2$).

Y : No se sufre un ataque cardíaco o bien se sufre y es mortal o no ($J=3$).

	Infarto de Miocardio		
	Ataque Mortal	Ataque No Mortal	No Ataque
Placebo	18	171	10.845
Aspirina	5	99	10.933

Cuadro 2.1: Clasificación Cruzada de Uso de Aspirina y Ataque Cardíaco.

El Cuadro 2.1, una tabla de contingencia 2×3 , relaciona el uso de aspirina y los ataques al corazón. Esto fue un estudio aleatorizado sobre si la ingesta regular de aspirina reduce la mortalidad por enfermedad cardiovascular. Los pacientes del hospital tomaban una tableta de aspirina o placebo. El estudio era ciego, los pacientes no sabían si tomaban aspirina o no. De los 11.034 pacientes que tomaron placebo, 18 sufrieron un ataque mortal al corazón, mientras que de los 11.037 pacientes que tomaron aspirina, 5 sufrieron un ataque fatal al corazón.

Distribución Conjunta

Si π_{ij} denota la probabilidad de que (X, Y) ocurra en la casilla de la fila i y la columna j , la distribución de probabilidad $\{\pi_{ij}\}$ es la *distribución conjunta* de X e Y .

La distribución conjunta viene dada por:

$$\pi_{ij} = P(X = i, Y = j) \quad \text{con } i = 1, \dots, I \text{ y } j = 1, \dots, J$$

Distribución Marginal

Las *distribuciones marginales* son los totales de filas (o columnas) que resultan al sumar las probabilidades conjuntas.

Denotamos los conteos marginales por $\{\pi_{i+}\}$ para la variable fila y $\{\pi_{+j}\}$ para la variable columna, donde el subíndice "+" denota la suma sobre ese índice. Esto es,

$$\pi_{i+} = P(X = i) = \sum_{j=1}^J P(X = i, Y = j) = \sum_{j=1}^J \pi_{ij}$$

$$\pi_{+j} = P(Y = j) = \sum_{i=1}^I P(X = i, Y = j) = \sum_{i=1}^I \pi_{ij}$$

Estos satisfacen:

$$\sum_i \pi_{i+} = \sum_j \pi_{+j} = \sum_i \sum_j \pi_{ij} = 1$$

Las distribuciones marginales proporcionan información acerca de una sola variable.

Distribución Condicional

En la mayor parte de las tablas de contingencia, como en el ejemplo anterior, una de las variables, digamos Y , es una variable de respuesta y la otra variable, X , es una variable explicativa o predictora. Cuando X es fijo en lugar de aleatorio, la noción de distribución conjunta para X e Y ya no es significativa. Sin embargo, para una categoría fija de X , Y tiene una distribución de probabilidad. Esto permite estudiar como cambia su distribución a media que cambia el valor de X .

Distribución Condicionada de Y respecto de X

Dado un sujeto que está clasificado en la fila i de X , $\pi_{j|i}$ denota la probabilidad de clasificación en la columna j de Y , $j = 1, \dots, J$. Nótese que:

$$\sum_j \pi_{j|i} = 1$$

Las probabilidades $\{\pi_{1|i}, \dots, \pi_{J|i}\}$ forman la *distribución condicional* de Y en la categoría i de X .

La mayor parte de los estudios se centran en la comparación de las distribuciones condicionadas de Y para varios niveles de la variable explicativa.

Sensibilidad y Especificidad

Supongamos un test de diagnóstico para una determinada enfermedad. Las siguientes dos características son de interés:

- La *sensibilidad* caracteriza la capacidad de la prueba para detectar la enfermedad en sujetos enfermos.
- La *especificidad* caracteriza la capacidad de la prueba para detectar la ausencia de la enfermedad en sujetos sanos.

Ilustremos estas definiciones con un ejemplo:

Cáncer de Mama	Diagnóstico		Total
	Positivo	Negativo	
Sí	0.82	0.18	1
No	0.01	0.99	1

Cuadro 2.2: Distribuciones Condicionadas Estimadas para el Diagnóstico de Cáncer de Mama.

Los resultados del Cuadro 2.2 sirven para ver el impacto del uso de la mamografía junto con el examen clínico de la mama.

En este caso X denota el estado verdadero de la enfermedad (si una mujer tiene realmente cáncer de mama) e Y el diagnóstico (positivo o negativo), donde un resultado positivo significa que la prueba sugiere la presencia de cáncer.

Las probabilidades estimadas en el Cuadro 2.2 son las probabilidades condicionadas de Y dado X .

Con las pruebas de diagnóstico para una enfermedad, los dos posibles resultados son uno positivo cuando el sujeto tiene la enfermedad y uno negativo

cuando el sujeto no la tiene. Dado que el sujeto tiene la enfermedad, la probabilidad condicionada de que el resultado del diagnóstico sea positivo es llamada *sensibilidad*. Dado que el sujeto no tiene la enfermedad, la probabilidad condicionada de que el resultado sea negativo se llama *especificidad*.

Para una tabla 2×2 con el formato del Cuadro 2.2, la sensibilidad es $\pi_{1|1}$ y la especificidad es $\pi_{2|2}$. En el Cuadro 2.2, la estimación de la sensibilidad de la combinación de una mamografía y un examen clínico es 0.82. De las mujeres con cáncer, el 82 % se diagnostica correctamente. La especificidad es 0.99. De las mujeres que no tienen cáncer, el 99 % se diagnostica correctamente.

Idealmente, estos valores deben ser altos.

Independencia de las Variables Categóricas

Cuando ambas variables son variables de respuesta, se pueden usar tanto las distribuciones conjuntas como las condicionales para describir la asociación entre ellas. La distribución condicional de Y dada X se relaciona con la distribución conjunta por:

$$\pi_{j|i} = \pi_{ij}/\pi_{i+} \quad \forall i, j$$

Dos variables categóricas de respuesta son *independientes* si las probabilidades conjuntas son iguales al producto de sus probabilidades marginales,

$$\pi_{ij} = \pi_{i+}\pi_{+j} \quad \text{para } i = 1, \dots, I \text{ y } j = 1, \dots, J$$

Cuando X e Y son independientes:

$$\pi_{j|i} = \pi_{ij}/\pi_{i+} = (\pi_{i+}\pi_{+j})/\pi_{i+} = \pi_{+j} \quad \text{para } i = 1, \dots, I$$

Cada distribución condicional de Y es idéntica a la distribución marginal de Y . Así, dos variables son independientes cuando $\{\pi_{j|1} = \dots = \pi_{j|I}, \text{ para } j = 1, \dots, J\}$; esto es, la probabilidad de que la respuesta de cada columna sea la misma en cada fila.

Cuando Y es una variable de respuesta y X es una variable explicativa, hablamos de *homogeneidad* en lugar de independencia.

Distribuciones de los Recuentos

Cuando el tamaño de la muestra es una variable aleatoria, los recuentos de las casillas $\{Y_{ij}\}$ se consideran variables aleatorias independientes de Poisson con parámetros $\{\mu_{ij}\}$. La función de masa de probabilidad para los resultados potenciales $\{n_{ij}\}$ es el producto de las probabilidades de Poisson $P(Y_{ij} = n_{ij})$ para las $I \times J$ casillas, o

$$\prod_i \prod_j \frac{e^{-\mu_{ij}} \mu_{ij}^{n_{ij}}}{n_{ij}!}$$

Cuando el tamaño de la muestra es fijo es habitual considerar que los recuentos de las casillas de las tablas de contingencia se distribuyen como una multinomial. En el *muestreo multinomial*, fijamos el tamaño total n pero no los totales de filas y columnas. Así, se modeliza la situación de que las filas se refieren a diferentes grupos. Si se tienen $I \times J$ casillas con observaciones, la distribución de probabilidad de los recuentos es:

$$\frac{n!}{n_{11}!, \dots, n_{IJ}!} \prod_i \prod_j \pi_{ij}^{n_{ij}}$$

En este caso, los parámetros desconocidos son las probabilidades conjuntas $\{\pi_{11}, \dots, \pi_{IJ}\}$.

A veces, las observaciones en una variable de respuesta Y aparecen de manera separada según cada nivel de la variable explicativa X . En este caso se tratan los totales de filas como fijos.

Se simplifica la notación de modo que $n_i = n_{i+}$. Los recuentos $\{n_{ij}, j = 1, \dots, J\}$ tales que $\sum_j n_{ij} = n_i$, se distribuyen para cada nivel i de X como:

$$\frac{n_i!}{\prod_j n_{ij}!} \prod_j \pi_{j|i}^{n_{ij}} \quad (2.1)$$

En este caso, los parámetros desconocidos son las probabilidades condicionadas $\{\pi_{1|i}, \dots, \pi_{J|i}\}$.

Este esquema se denomina *muestreo multinomial independiente* o *muestreo de producto de multinomiales*.

Comparación de Proporciones en Tablas 2×2

Muchos estudios se diseñan para comparar grupos basándose en una respuesta Y binaria. Con dos grupos tenemos una tabla de contingencia 2×2 .

	Éxito	Fracaso
Grupo 1	$\pi_{1 1}$	$\pi_{2 1}$
Grupo 2	$\pi_{1 2}$	$\pi_{2 2}$

Se denota:

$$\pi_{1|i} = \pi_i$$

$$\pi_{2|i} = 1 - \pi_{1|i} = 1 - \pi_i$$

De modo que la tabla se puede reescribir como:

	Éxito	Fracaso
Grupo 1	π_1	$1 - \pi_1$
Grupo 2	π_2	$1 - \pi_2$

Se quiere comparar π_1 y π_2 . Para ello, podemos estudiar:

1. La *diferencia de proporciones*:

$$\pi_1 - \pi_2$$

La diferencia de proporciones está entre -1.0 y $+1.0$. Es igual a cero cuando las filas tienen distribuciones condicionales idénticas. La variable de respuesta Y es independiente de X cuando $\pi_1 - \pi_2 = 0$.

2. El *riesgo relativo*:

$$\frac{\pi_1}{\pi_2}$$

Puede ser cualquier valor real no negativo. Un riesgo relativo igual a 1.0 corresponde a la independencia.

3. La *razón de posibilidades*:

$$\theta = \frac{\pi_1/(1 - \pi_1)}{\pi_2/(1 - \pi_2)}$$

Para una probabilidad π de éxito, la *posibilidad de éxito* se define como $\Omega = \pi/(1 - \pi)$. La posibilidad toma valores no negativos. Cuando $\Omega > 1$, la probabilidad de éxito es mayor que la de fracaso.

Dentro de la fila i tenemos $\Omega_i = \pi_i/(1 - \pi_i)$, y θ es, por tanto, Ω_1/Ω_2 .

Para las distribuciones conjuntas con probabilidades de cada casilla $\{\pi_{ij}\}$, la definición equivalente para la posibilidad de la fila i es $\Omega_i = \pi_{i1}/\pi_{i2}$, para $i = 1, 2$. Entonces la razón de posibilidades es:

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \quad (2.2)$$

La independencia entre X e Y se tiene cuando $\theta = 1$. Cuando $1 < \theta < \infty$, la probabilidad de éxito es mayor en la fila 1 que en la fila 2; esto es, $\pi_1 > \pi_2$. Cuando $0 < \theta < 1$, $\pi_1 < \pi_2$. Cuando una de las casillas tiene probabilidad 0, θ es igual a 0 o a ∞ .

Cuando cambia el orden de las filas, θ queda invertido. Por ello es mejor usar el *logaritmo de la razón de posibilidades*. Cuando cambia el orden de las filas $\log \theta$ cambia su signo. La independencia se tiene para $\log \theta = 0$. El logaritmo de la razón de posibilidades es simétrica respecto a este valor.

La razón de posibilidades no cambia de valor si las filas se convierten en las columnas y las columnas en las filas. Esto está muy cerca de la simetría para (2.2). No es necesario identificar una clasificación como la variable de respuesta para usar θ . De hecho, θ se definió en términos de probabilidades usando $\pi_i = P(Y = 1|X = i)$, y se podría definir también usando probabilidades condicionales inversas. Con una distribución conjunta, existen distribuciones condicionales en cada dirección, y

$$\begin{aligned} \theta &= \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{P(Y = 1|X = 1)/P(Y = 2|X = 1)}{P(Y = 1|X = 2)/P(Y = 2|X = 2)} \\ &= \frac{P(X = 1|Y = 1)/P(X = 2|Y = 1)}{P(X = 1|Y = 2)/P(X = 2|Y = 2)} \end{aligned} \quad (2.3)$$

De hecho, la razón de posibilidades es igual de válida para los distintos tipos de estudio, retrospectivos, prospectivos, y transversales, que veremos más adelante.

La razón de posibilidades y el riesgo relativo se relacionan de la siguiente manera:

$$\text{razón de posibilidades} = \text{riesgo relativo} \times \left(\frac{1 - \pi_2}{1 - \pi_1} \right)$$

Las magnitudes son similares cuando la probabilidad π_i esta cerca de cero en ambos grupos.

Para ilustrar estas tres asociaciones veamos el Cuadro 2.1 anterior, donde se relacionaba el uso de aspirinas con los ataques cardíacos. La tabla de contingencia diferencia entre ataques mortales y no mortales, combinemos estos dos por ahora. De los 11.034 pacientes que tomaron placebo, 189 sufrieron un ataque al corazón, una proporción de $189/11.034 = 0.0171$. De los 11.037 que tomaron aspirina, 104 sufrieron un ataque al corazón, una proporción de 0.0094. La diferencia de proporciones de la muestra es $0.0171 - 0.0094 = 0.0077$. El riesgo relativo es $0.0171/0.0094 = 1.82$. La proporción de los que sufrieron ataques al corazón y tomaron placebo es 1.82 veces la proporción de los que sufrieron ataques al corazón y tomaron aspirina. La razón de posibilidades es $(189 \times 10.933)/(10.845 \times 104) = 1.83$. La probabilidad de ataque al corazón para aquellos que tomaban placebo es 1.83 veces la probabilidad para aquellos que tomaban aspirina. Vemos que el riesgo relativo y la razón de posibilidades son bastante similares debido a que la proporción de ataque al corazón es menor que 0.02 en cada grupo.

2.2. Tipos de Estudios

$$\begin{array}{l} \text{Tipos} \\ \text{de Estudios} \end{array} \left\{ \begin{array}{l} \text{- Retrospectivos} \left\{ \begin{array}{l} \text{- Estudios de Caso-Control} \end{array} \right. \\ \text{- Prospectivos} \left\{ \begin{array}{l} \text{- Estudios de Cohorte} \\ \text{- Ensayos Clínicos} \end{array} \right. \\ \text{- Transversales} \end{array} \right.$$

En los *estudios de caso-control* los sujetos incluidos proceden de dos grupos, según sean *casos* (con la enfermedad o daño en estudio) o *controles* (sin el daño en estudio). Este tipo de diseño hizo su aparición a mediados del siglo XX cuando, en los países desarrollados, el interés de la Salud Pública comenzó a centrarse en las enfermedades crónicas. La idea básica es comparar los antecedentes de los “enfermos” de una población con los antecedentes de los “sanos” de la misma población, tratando de descubrir diferencias en las exposiciones que expliquen, al menos parcialmente, la razón por la que unos enfermaron y otros no.

En el análisis se comparan las exposiciones de los casos con las de los controles, y los resultados son presentados usando las posibilidades (cociente entre la probabilidad de enfermar y la probabilidad de no enfermar) y la razón de posibilidades de adquirir una enfermedad entre expuestos y entre no expuestos.

	Casos	Controles	Total
Expuestos	a	b	a+b
No Expuestos	c	d	c+d
Total	a+c	b+d	a+b+c+d

La razón de posibilidades $\theta = (a \times d)/(c \times b)$ será mayor cuanto mayor sea el número de casos expuestos y el de controles no expuestos y menor cuanto mayor sea el número de casos no expuestos y el de controles expuestos.

El Cuadro 2.3 que aparece a continuación viene de uno de los primeros estudios del vínculo entre el cáncer de pulmón y fumar.

	Casos	Controles
Fumar	688	650
No Fumar	21	59
Total	709	709

Cuadro 2.3: Clasificación Cruzada de Fumar y el Cáncer de Pulmón.

En 20 hospitales en Londres, pacientes ingresados con cáncer de pulmón en el año anterior fueron entrevistados sobre su comportamiento de fumar. Para cada uno de los 709 pacientes, los investigadores estudiaban el comportamiento de fumar en un paciente sin cáncer de pulmón, del mismo hospital, del mismo sexo y dentro del mismo grupo de edad. Los 709 *casos* en la primera columna del Cuadro 2.3 son los que tienen cáncer de pulmón y los 709 *controles* en la segunda columna son los que no lo tienen. Un fumador fue definido como una persona que había fumado al menos un cigarrillo al día durante el pasado año.

Normalmente, tener cáncer de pulmón es una variable de respuesta y el comportamiento de fumar es una variable explicativa. En este estudio, sin embargo, la distribución marginal del cáncer es fijada por el diseño de muestreo, y el resultado medido es si el sujeto fue alguna vez fumador. Este tipo de estudio tiene un diseño *retrospectivo* dado que “mira hacía el pasado”.

Tales estudios son comunes en aplicaciones relacionadas con la salud. A menudo, las dos muestras son iguales. A veces también las muestras de casos y controles son independientes y no coincidentes. Por ejemplo, otro estudio de caso-control sobre el cáncer de pulmón y el consumo de tabaco fue seleccionar una muestra de pacientes que habían fallecido debido al cáncer y luego clasificarlas según el tipo de cáncer y su comportamiento de fumar.

Se podría querer comparar los fumadores con los no fumadores en términos de la proporción que sufrió cáncer de pulmón. Estas proporciones se refieren a la distribución condicional de cáncer de pulmón dado el comportamiento de fumar. En cambio, el estudio de caso-control proporciona proporciones en la dirección inversa, para la distribución condicional del comportamiento de fumar dado el estado del cáncer de pulmón. Para aquellos en el Cuadro 2.3 con cáncer de pulmón, la proporción para los que fumaban era $688/709 = 0.970$, mientras que era $650/709 = 0.917$ para los controles.

Cuando sabemos la proporción de la población que tiene cáncer, podemos usar el Teorema de Bayes para calcular las distribuciones condicionales en la dirección de interés. De lo contrario, utilizando una muestra retrospectiva, no podremos estimar la probabilidad del cáncer en cada categoría del comportamiento de fumar. Para el Cuadro 2.3 no se conoce la prevalencia de la población de cáncer de pulmón, y los pacientes que lo sufrieron fueron probablemente muestreados con una tasa muy superior a su ocurrencia en la población general.

Con los muestreos con diseños retrospectivos, tales como los estudios de caso-control, es posible estimar las probabilidades condicionales $P(X = i|Y = j)$. Normalmente no es posible estimar la probabilidad $P(Y = j|X = i)$ de un resultado de interés, la diferencia de proporciones o el riesgo relativo para ese resultado. Sin embargo, es posible estimar la razón de posibilidades ya que en (2.3) se determina por probabilidades condicionales en cualquier dirección.

Para ilustrarlo volvemos al ejemplo del Cuadro 2.3 con $X =$ comportamiento de fumar e $Y =$ cáncer de pulmón. Los datos fueron dos muestras binomiales de X a niveles fijos de Y . Esto es, podemos estimar la probabilidad de que un sujeto fue fumador dado el resultado de si el sujeto tenía cáncer; esto era $688/709$ para los casos y $650/709$ para los controles. No podemos estimar la

probabilidad de cáncer dado si era fumador, que es más relevante. Por lo tanto, no podemos estimar la diferencia de proporciones o el riesgo relativo, ya que se limitan a las comparaciones de las probabilidades de ser fumador. Sin embargo, podemos calcular la razón de posibilidades usando (2.3),

$$\frac{(688/709)/(21/709)}{(650/709)/(59/709)} = \frac{688 \times 59}{650 \times 21} = 3.0$$

Por otra parte, las interpretaciones pueden usar la dirección de interés, a pesar de que el estudio fue retrospectivo: las posibilidades estimadas del cáncer de pulmón para fumadores eran 3.0 veces las posibilidades estimadas para no fumadores.

Por el contrario, imaginemos un estudio que muestree sujetos de una población de adolescentes y 60 años después mida las tasas de cáncer de pulmón para fumadores y no fumadores. Este diseño de muestreo es *prospectivo*. Hay dos tipos de estudios prospectivos, los *ensayos clínicos* y los *estudios de cohorte*.

Los ensayos clínicos asignan aleatoriamente sujetos a los grupos de fumadores y no fumadores.

Los estudios de cohorte sustentan su estrategia de análisis en el seguimiento en el tiempo de dos o más grupos de individuos que han sido divididos según el grado de exposición a un determinado factor (corrientemente en 2 grupos: expuestos y no expuestos). Al inicio ninguno de los individuos incluidos en ambos grupos tiene la enfermedad o daño en estudio y el objetivo es comparar la incidencia de “nuevos casos” entre ambos grupos.

El último tipo de estudio es el que tiene *diseño transversal*. Muestra los sujetos y los clasifica simultáneamente en ambas variables.

Los estudios prospectivos normalmente condicionan los totales $\{n_i = \sum_j n_{ij}\}$ para las categorías de X y consideran cada fila de J recuentos como una muestra multinomial independiente en Y . Los estudios retrospectivos normalmente tratan los totales $\{n_{+j}\}$ para Y como fijos y consideran cada columna de I recuentos como una muestra multinomial de X . En los estudios de diseño transversal el tamaño total de la muestra es fijo pero no los totales de filas y columnas, y los recuentos de las casillas $I \times J$ son muestras multinomiales.

Los estudios de caso-control, cohorte y transversales son estudios *observacionales*. Simplemente observan quién está seleccionado en cada grupo y quién

tiene el resultado de interés. Por el contrario, los ensayos clínicos son estudios *experimentales*, el investigador tiene la ventaja de control experimental sobre qué sujetos recibe cada tratamiento. Tales estudios pueden usar el poder de aleatorización para hacer que los grupos se equilibren más o menos con otras variables que pueden estar asociadas con la respuesta.

2.3. Tablas Estratificadas

Una parte importante de muchos estudios, estudios observacionales especialmente, es la elección de las variables de control. Al estudiar el efecto de X en Y , se debe controlar cualquier covariable que pueda influir en esta relación. Esto implica usar algún mecanismo para mantener la covariable constante. De lo contrario, un efecto observado de X en Y puede reflejar efectos de esta covariable. La relación entre X e Y muestra entonces *confusión*. Los estudios experimentales pueden eliminar efectos de las covariables de confusión al asignar los sujetos aleatoriamente a diferentes niveles de X , pero esto no es posible con estudios observacionales.

Existen diferentes estrategias para “controlar” este efecto y una de ellas es la *estratificación*. Por ejemplo, supongamos que un estudio de casos y controles arrojó una asociación positiva entre el consumo de café y el cáncer de páncreas con los datos del Cuadro 2.4.

Café	Casos	Controles
Sí	196	104
No	89	106

Cuadro 2.4: Razón de Posibilidades = 2.24.

Café	Casos	Controles	Café	Casos	Controles
Sí	32	64	Sí	164	40
No	48	96	No	41	10

Cuadro 2.5: No Fumadores.

Cuadro 2.6: Fumadores.

Sin embargo, al considerar un tercer factor como el tabaco y dividir los individuos del estudio en fumadores (Cuadro 2.6) y no fumadores (Cuadro 2.5) no parece existir relación entre café y cáncer de páncreas en los no fumadores y tampoco en los fumadores (razón de posibilidades = 1 en ambos grupos).

El análisis por estratos hace evidente que el consumo de tabaco ha distorsiona-

do la relación entre el consumo de café y el cáncer de páncreas. En esta relación es el tabaco el que estaría incrementando el riesgo de cáncer de páncreas, y como entre los fumadores el consumo de café es más frecuente, la tabla simple mostraba una asociación entre el consumo de café y el cáncer de páncreas.

La existencia de diferencias entre los resultados de un análisis no estratificado y uno estratificado estará mostrando que el factor por el que se estratificó ejerce un efecto de confusión en la relación que exhiben los factores estudiados.

Si bien es posible la estratificación conjunta de varios factores con el objetivo de controlarlos o ajustarlos simultáneamente (ej. varones fumadores, varones no fumadores, mujeres fumadoras, mujeres no fumadoras), la generación de varios estratos disminuye notablemente el tamaño muestral de cada estrato, lo que hace en extremo inestables las estimaciones realizadas al interior de cada estrato.

Supongamos ahora que un estudio considera los efectos del tabaquismo pasivo, los efectos de un no fumador de vivir con un fumador. Para analizar si el tabaquismo pasivo está asociado con el cáncer de pulmón, un estudio de diseño transversal podría comparar las tasas de cáncer entre no fumadores cuyos cónyuges fuman y no fumadores cuyos cónyuges no fuman.

El estudio debe intentar controlar la edad, el estatus socioeconómico, u otros factores que puedan relacionarse tanto con el cónyuge fumador como con el desarrollo del cáncer. De lo contrario los resultados tendrán una utilidad limitada. Los cónyuges de no fumadores tienden a ser más jóvenes que los cónyuges de fumadores, y las personas más jóvenes son menos propensas a tener cáncer. Entonces, una menor proporción de casos de cáncer entre cónyuges de no fumadores puede reflejar su menor edad.

En esta sección hablaremos de la asociación entre variables categóricas X e Y controlando una posible variable de confusión Z .

Controlamos Z estudiando la relación XY a niveles fijos de Z . Las secciones transversales bidireccionales de la tabla de contingencia de tres vías clasifican X e Y en las categorías separadas de Z . Estas secciones se denominan *tablas parciales*. Muestran la relación XY mientras eliminan los efectos de Z manteniendo su valor constante.

La tabla de contingencia bidireccional obtenida combinando las tablas parciales se denomina *tabla marginal XY*. Cada recuento de casillas en la tabla marginal es una suma de recuentos de la misma localización en las tablas parciales. La tabla marginal, en lugar de controlar Z , la ignora. La tabla marginal no contiene información sobre Z . Es simplemente una tabla bidireccional que

relaciona X e Y pero puede reflejar los efectos de Z en X e Y .

Las asociaciones en tablas parciales se denominan *asociaciones condicionales*, debido a que se refieren al efecto condicional de X en Y con Z fijo en algún nivel. Las asociaciones condicionales en tablas parciales pueden ser bastante diferentes a la asociación en la tabla marginal. De hecho, puede ser engañoso analizar sólo la tabla marginal. El siguiente ejemplo lo ilustra.

Ejemplo Pena de Muerte

El Cuadro 2.7 es una tabla de contingencia $2 \times 2 \times 2$ –dos filas, dos columnas, y dos capas– de un artículo que estudió los efectos de las características raciales de si las personas condenadas por homicidio recibieron la pena de muerte. Los 674 sujetos clasificados en el Cuadro 2.7 fueron acusados en casos de asesinato múltiple en Florida.

Raza de la Víctima	Raza del Acusado	Pena de Muerte		Porcentaje de Pena de Muerte
		Sí	No	
Blanco	Blanco	53	414	11.3
	Negro	11	37	22.9
Negro	Blanco	0	16	0.0
	Negro	4	139	2.8
Total	Blanco	53	430	11.0
	Negro	15	176	7.9

Cuadro 2.7: Sentencia de Pena de Muerte por la Raza del Acusado y la Raza de la Víctima.

Las variables son Y = sentencia de pena de muerte, con las categorías (sí, no), X = raza del acusado, y Z = raza de la víctima, cada una de las cuales con las categorías (blanco, negro).

Se estudia el efecto de la raza del acusado en la sentencia de pena de muerte, tratando la raza de la víctima como una variable de control.

Para cada combinación de la raza del acusado y la raza de la víctima la Figura 2.1 muestra el porcentaje de acusados que recibieron la pena de muerte. Esto describe las asociaciones condicionales. Cuando las víctimas eran blancas, la pena de muerte fue impuesta a un $22.9\% - 11.3\% = 11.6\%$ más para acusados negros que para acusados blancos. Cuando las víctimas eran negras, la pena de muerte fue impuesta a un $2,8\%$ más para acusados negros que para acusados blancos. Controlando la raza de la víctima manteniendola fija, la pena de muerte fue impuesta más a menudo a acusados negros que a acusados blancos.

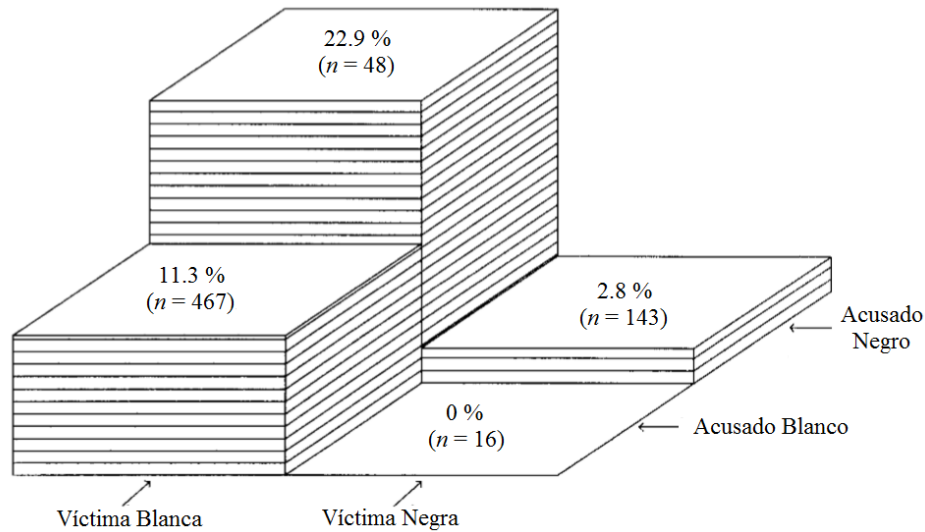


Figura 2.1: Porcentaje que recibió la Pena de Muerte

La parte inferior del Cuadro 2.7 muestra la tabla marginal. Se obtiene de la suma de los recuentos de las casillas sobre las dos categorías de la víctima, combinando así las dos tablas parciales (ej, $11+4=15$). En general, el 11 % de los acusados blancos y el 7.9 % de los acusados negros recibieron la pena de muerte. Ignorando la raza de la víctima, la pena de muerte fue impuesta de forma menos frecuente a acusados negros que a acusados blancos. La asociación invierte la dirección en comparación con las tablas parciales.

¿Por qué la asociación cambia tanto cuando ignoramos la raza de la víctima? Esto se relaciona con la naturaleza de la asociación entre la raza de la víctima y cada una de las otras variables. Primero, la asociación entre la raza de la víctima y la raza del acusado es extremadamente fuerte. La tabla marginal que relaciona estas variables tiene razón de posibilidades:

$$(467 \times 143)/(48 \times 16) = 87.0.$$

Segundo, el Cuadro 2.7 muestra esto, independientemente de la raza del acusado, la pena de muerte era mucho más probable cuando las víctimas eran blancas que cuando las víctimas eran negras. Así que los blancos tienden a asesinar a blancos, y asesinar a blancos es más probable que resulte en pena de muerte.

Esto sugiere que la asociación marginal debería mostrar una mayor tendencia que las asociaciones condicionales a que los acusados blancos reciban la pena de muerte. De hecho, el Cuadro 2.7 tiene ese patrón.

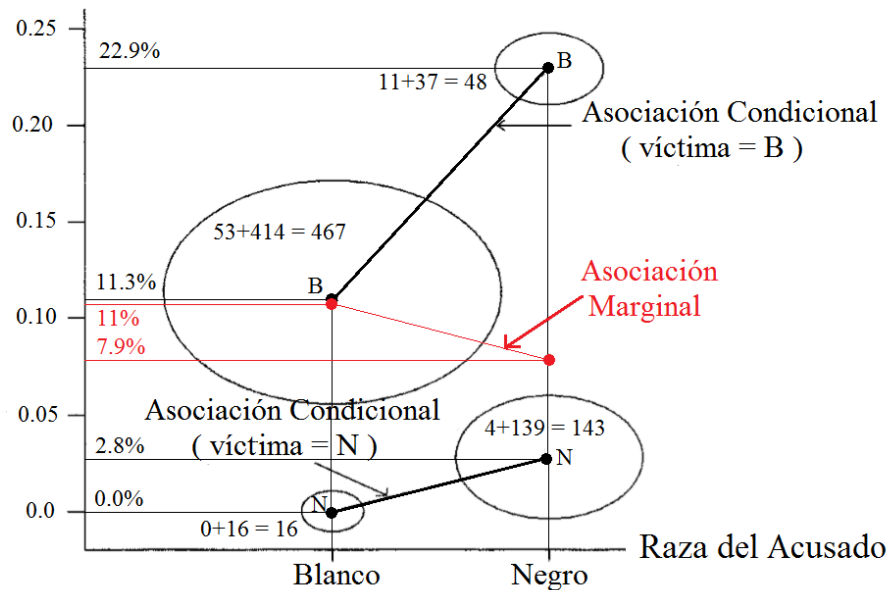


Figura 2.2: Proporción que recibió la Pena de Muerte por la Raza del Acusado, controlando e ignorando la Raza de la Víctima

La Figura 2.2 ilustra porqué la asociación marginal difiere de las asociaciones condicionales. Para cada raza del acusado, la figura dibuja la proporción que recibe la pena de muerte en cada categoría de la raza de la víctima. Cada proporción está etiquetada por un símbolo de la letra que da la categoría de la raza de la víctima. Alrededor de cada observación hay un círculo que tiene un área proporcional al número de observaciones en esa combinación de la raza del acusado y la raza de la víctima.

Por ejemplo, la B en el círculo más grande representa la proporción de 0.113 que recibe la pena de muerte para casos con acusados blancos y víctimas blancas. Ese círculo es el más grande porque el número de casos de esta combinación (53 + 414 = 467) es el mayor. El siguiente círculo más grande es el de los casos donde los negros asesinan a negros.

Controlamos la raza de las víctimas comparando los círculos que tienen la misma letra de la raza de la víctima en sus centros. La línea que conecta los dos círculos con la letra B tiene una pendiente positiva, como la línea que une los dos círculos con la letra N. Al controlar la raza de la víctima se refleja que la pena de muerte es más probable para acusados negros que para acusados

blancos. Cuando agregamos resultados a través de la raza de la víctima para obtener un resultado de resumen para el efecto marginal de la raza del acusado sobre la pena de muerte, los círculos más grandes, que tienen un mayor número de casos, tienen mayor influencia. Así, las proporciones de resumen para cada raza del acusado, marcadas en la figura en rojo, se acercan más al centro de los círculos mayores que al centro de los círculos más pequeños. La línea que conecta las proporciones marginales de resumen tiene una pendiente negativa, indicando que, en general, la pena de muerte es más probable para acusados blancos que para acusados negros.

El resultado de que la asociación marginal puede tener una dirección diferente de las asociaciones condicionales se llama *Paradoja de Simpson*.

Razón de Posibilidades

Las razones de posibilidades pueden describir asociaciones condicionales y marginales. Vamos a ilustrar esto para tablas $2 \times 2 \times K$, donde K denota el número de categorías de una variable de control, Z . Dado $\{\mu_{ijk}\}$ denota las frecuencias esperadas de las casillas para algún modelo de muestreo, tales como una muestra binomial, multinomial o de Poisson.

Dentro de una categoría k fija de Z , la razón de posibilidades:

$$\theta_{XY^{(k)}} = \frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}} \quad (2.4)$$

describe la asociación XY en la tabla parcial k . Las razones de posibilidades para las K tablas se denominan *razones de posibilidades XY condicionales*. Estas pueden ser bastante diferentes de las razones de posibilidades marginales. La tabla XY marginal tiene frecuencias esperadas $\{\mu_{ij+} = \sum_k \mu_{ijk}\}$. La razón de posibilidades XY marginal es:

$$\theta_{XY} = \frac{\mu_{11+}\mu_{22+}}{\mu_{12+}\mu_{21+}}$$

Los valores de la muestra de $\theta_{XY^{(k)}}$ y θ_{XY} usan fórmulas similares con los recuentos de las casillas sustituidos por las frecuencias esperadas.

En el ejemplo anterior, que ilustraba la asociación entre la raza del acusado con la pena de muerte, en la primera tabla parcial, con víctimas blancas, tenemos:

$$\hat{\theta}_{XY(1)} = \frac{53 \times 37}{414 \times 11} = 0.43$$

Las posibilidades de la muestra para los acusados blancos que recibieron pena de muerte fueron el 43% de las posibilidades de la muestra para acusados negros.

En la segunda tabla parcial, con víctimas negras, la razón de posibilidades estimada es:

$$\hat{\theta}_{XY(2)} = \frac{0 \times 139}{16 \times 4} = 0.0$$

ya que la pena de muerte nunca se dió a los acusados blancos con víctimas negras.

La estimación de la razón de posibilidades marginal usa la tabla marginal 2×2 dentro del Cuadro 2.7, ignorando la raza de la víctima, esto es:

$$\hat{\theta}_{XY} = \frac{53 \times 176}{430 \times 15} = 1.45$$

La posibilidad de pena de muerte era el 45% más grande para acusados blancos que para acusados negros. Sin embargo, dentro de cada categoría de la raza de la víctima, esas posibilidades fueron más pequeñas para acusados blancos. Esta reversión en la asociación después de controlar la raza de la víctima muestra la Paradoja de Simpson.

Independencia

Supongamos, de forma más general, que X tiene I categorías e Y tiene J categorías. Una tabla $I \times J \times K$ describe la relación entre X e Y , controlada por la variable Z . Si X e Y son independientes en la tabla parcial k , entonces X e Y son *independientes condicionalmente a nivel k* de Z . Cuando Y es una respuesta, esto significa que:

$$P(Y = j|X = i, Z = k) = P(Y = j|Z = k), \quad \text{para todo } i, j \quad (2.5)$$

Más generalmente, X e Y son *independientes condicionalmente dado Z* cuando son independientes condicionalmente para todos los niveles de Z , esto es,

cuando (2.5) se cumple para todo k . Entonces, dado Z , Y no depende de X .

Supongamos que se aplica un modelo multinomial a toda tabla de tres vías con probabilidades conjuntas $\{\pi_{ijk} = P(X = i, Y = j, Z = k)\}$. Entonces:

$$\pi_{ijk} = P(X = i, Z = k)P(Y = j|X = i, Z = k)$$

que bajo la independencia condicional de X e Y , dada Z , es igual a:

$$\pi_{ijk} = \pi_{i+k}P(Y = j|Z = k) = \pi_{i+k}P(Y = j, Z = k)/P(Z = k)$$

La independencia condicional es entonces equivalente a:

$$\pi_{ijk} = \pi_{i+k}\pi_{+jk}/\pi_{++k}, \quad \text{para todo } i, j \text{ y } k \quad (2.6)$$

Vamos a ver que la independencia condicional no implica la independencia marginal.

Sumando (2.6) sobre k :

$$\pi_{ij+} = \sum_k \pi_{i+k}\pi_{+jk}/\pi_{++k}$$

Los tres términos en la suma implican k , y esto no significa $\pi_{ij+} = \pi_{i++}\pi_{+j+}$, la independencia marginal.

(Z) Clínica	(X) Tratamiento	(Y) Respuesta	
		Éxito	Fracaso
1	A	18	12
	B	12	8
2	A	2	8
	B	8	32
Total	A	20	20
	B	20	40

Cuadro 2.8: Frecuencias Esperadas para mostrar que la Independencia Condicional no implica la Independencia Marginal.

Para tablas $2 \times 2 \times K$, X e Y son independientes condicionalmente cuando la razón de posibilidades entre X e Y es igual a 1 para cada categoría de Z . Las frecuencias esperadas $\{\mu_{ijk}\}$ del Cuadro 2.8 ilustran esta relación para $Y =$ respuesta (éxito, fracaso), $X =$ tratamiento (A, B), y $Z =$ clínica (1, 2). De (2.4), las razones de posibilidades XY condicionales son:

$$\theta_{XY(1)} = \frac{18 \times 8}{12 \times 12} = 1.0, \quad \theta_{XY(2)} = \frac{2 \times 32}{8 \times 8} = 1.0$$

Dada la clínica, la respuesta y el tratamiento son independientes condicionalmente. La tabla marginal combina las tablas para las dos clínicas. Su razón de posibilidades es $\theta_{XY} = (20 \times 40)/(20 \times 20) = 2.0$, así que las variables no son independientes marginalmente.

Ignorando la clínica, ¿por qué las probabilidades de éxito para el tratamiento A son el doble que para el tratamiento B? Las razones de posibilidades XZ e YZ condicionales da una pista. Las razones de posibilidades entre Z y cualquiera de los dos X o Y , para cualquier categoría fija de la otra variable, es igual a 6.0.

Por ejemplo, la razón de posibilidades de XZ para la primera categoría de Y es igual a $(18 \times 8)/(12 \times 2) = 6.0$. Las probabilidades condicionales (dada la respuesta) de recibir el tratamiento A en la clínica 1 son seis veces las de la clínica 2, y las probabilidades condicionales (dado el tratamiento) de éxito en la clínica 1 son seis veces las de la clínica 2. La clínica 1 tiende a usar el tratamiento A, y la clínica 1 tiende a tener más éxito también.

Por ejemplo, si los pacientes de la clínica 1 tienden a ser más jóvenes y tener mejor estado de salud que los pacientes de la clínica 2, quizás tuvieran una mayor tasa de éxito independientemente del tratamiento escogido.

Es engañoso estudiar sólo la tabla marginal, concluyendo que los éxitos son más probables con el tratamiento A.

Los pacientes de una misma clínica son probablemente más homogéneos que la muestra total, y la respuesta es independiente del tratamiento en cada clínica.

Homogeneidad

Una tabla $2 \times 2 \times K$ tiene una *asociación homogénea* XY cuando

$$\theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)}$$

Entonces el efecto de X en Y es el mismo para cada categoría de Z . La independencia condicional de X e Y es el caso especial donde $\theta_{XY(k)} = 1.0$.

Bajo asociación homogénea XY , la homogeneidad también es válida para las otras asociaciones. Por ejemplo, la razón de posibilidades condicional entre dos categorías de X y dos categorías de Z es idéntica en cada categoría de Y . Para la razón de posibilidades, la asociación homogénea es una propiedad simétrica. Se aplica a cualquier par de variables vistas a través de las categorías de la tercera. Cuando esto ocurre, se dice que no hay interacción entre dos variables en sus efectos sobre la otra variable.

Cuando la interacción existe, la razón de posibilidades condicional para cualquier par de variables cambia entre las categorías de la tercera.

Para $X = \text{fumar (sí, no)}$, $Y = \text{cáncer de pulmón (sí, no)}$, y $Z = \text{edad (<45, 45-65, >65)}$, suponemos que $\theta_{XY(1)} = 1.2$, $\theta_{XY(2)} = 3.9$, y $\theta_{XY(3)} = 8.8$. Entonces, fumar tiene un efecto débil en el cáncer de pulmón para personas jóvenes, pero el efecto se fortalece considerablemente con la edad. En este caso la “edad” se llama *efecto modificador*; el efecto de fumar se modifica en función de su valor.

Para la pena de muerte (Cuadro 2.7), $\hat{\theta}_{XY(1)} = 0.43$ y $\hat{\theta}_{XY(2)} = 0.0$. Los valores no están cerca, pero la segunda estimación es inestable debido a que el conteo de casillas es cero. Al agregar $1/2$ a cada recuento de casillas, $\hat{\theta}_{XY(2)} = 0.94$. Debido a que $\hat{\theta}_{XY(2)}$ es inestable y debido a que se produce una mayor variación de la variabilidad de muestreo, estas tablas parciales no necesariamente contradicen la asociación homogénea en una población.

Capítulo 3

Inferencia para Tablas de Contingencia

En este capítulo introduciremos métodos de inferencia para tablas de contingencia. Los métodos asumen muestras de Poisson, multinomiales o binomiales independientes.

Primero asumiremos muestras grandes. Presentaremos los intervalos de confianza para tablas 2×2 para las medidas de asociación tales como la razón de posibilidades. Describiremos las pruebas chi-cuadrado de la hipótesis de independencia entre dos variables categóricas, y mostraremos como seguir la prueba usando residuos o la propiedad de partición de chi-cuadrado para extraer componentes que describen la evidencia acerca de la asociación. Por último introduciremos algunos métodos para muestras pequeñas.

3.1. Intervalos de Confianza

La precisión de los estimadores de los parámetros de asociación se caracteriza por los errores estándar de sus distribuciones de muestreo. En esta sección presentaremos errores estándar e intervalos de confianza para muestras grandes.

Estimación del Intervalo para la Razón de Posibilidades

El estimador que se utiliza para la razón de posibilidades es:

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

Este estimador puede ser 0 ó ∞ dependiendo de los recuentos que se tengan. Una posible opción entonces es trabajar con el estimador *corregido*:

$$\tilde{\theta} = \frac{(n_{11} + 0.5)(n_{22} + 0.5)}{(n_{12} + 0.5)(n_{21} + 0.5)}$$

o bien con la transformación $\log \hat{\theta}$.

Una estimación del error estándar de $\log \hat{\theta}$ es:

$$\hat{\sigma}_{\log \hat{\theta}} = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right)^{1/2}$$

de modo que el correspondiente intervalo de confianza de Wald es:

$$\log \hat{\theta} \pm z_{\alpha/2} \hat{\sigma}_{\log \hat{\theta}}$$

Si se toman las exponenciales (antilogaritmo) de los extremos se obtiene el intervalo correspondiente para $\hat{\theta}$.

El test es algo conservador ya que la probabilidad de cubrimiento es algo mayor que el nivel nominal.

Estimación del Intervalo para la Diferencia de Proporciones

Supongamos que tenemos muestras binomiales independientes, de modo que en el grupo i tenemos $Y_i \sim \text{Binom}(n_i, \pi_i)$, $i = 1, 2$, de modo que el estimador es:

$$\hat{\pi}_i = \frac{Y_i}{n_i}$$

y la media y el error estándar son:

$$E(\hat{\pi}_1 - \hat{\pi}_2) = \pi_1 - \pi_2$$

$$\sigma(\hat{\pi}_1 - \hat{\pi}_2) = \left[\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \right]^{1/2}$$

Sustituyendo π_i por $\hat{\pi}_i$ obtenemos el estimador del error estándar:

$$\hat{\sigma}_{\hat{\pi}_1 - \hat{\pi}_2} = \left[\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2} \right]^{1/2}$$

Entonces, el intervalo de confianza de Wald es:

$$\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \hat{\sigma}_{\hat{\pi}_1 - \hat{\pi}_2}$$

Cuando los valores de π_1 y π_2 están próximos a 0 ó a 1 este intervalo tiene una probabilidad de cubrimiento menor que la teórica que se considera.

Estimación del Intervalo para el Riesgo Relativo

El riesgo relativo muestral viene dado por:

$$r = \frac{\hat{\pi}_1}{\hat{\pi}_2}$$

Se prefiere usar mejor el logaritmo ya que converge más rápido a la distribución normal. El estimador del correspondiente error estándar es:

$$\hat{\sigma}_{\log r} = \left[\frac{1 - \hat{\pi}_1}{n_1 \hat{\pi}_1} + \frac{1 - \hat{\pi}_2}{n_2 \hat{\pi}_2} \right]^{1/2}$$

El intervalo de confianza de Wald para $\log r$ es:

$$\log r \pm z_{\alpha/2} \hat{\sigma}_{\log r}$$

Ejemplo

Por ejemplo, en la tabla de contingencia sobre el uso de la aspirina y el infarto de miocardio.

	Infarto Miocardio		Total
	Sí	No	
Placebo	28	656	684
Aspirina	18	658	676

Cuadro 3.1: Estudio sobre el Uso de Aspirina y el Infarto de Miocardio.

El estudio asignó al azar 1360 pacientes que ya habían sufrido un accidente cerebrovascular a un tratamiento con aspirina (un comprimido de dosis baja al día) o un tratamiento con placebo. El Cuadro 3.1 informa del número de muertes por infarto de miocardio durante un periodo de seguimiento de aproximadamente 3 años.

La razón de posibilidades muestral $\hat{\theta} = 1.56$ está cerca de $\tilde{\theta} = 1.55$, ya que ningún recuento de casillas es nulo. El estimador del error estándar de $\log \hat{\theta} = 0.445$ es $\hat{\sigma}(\log \hat{\theta}) = 0.307$.

Un intervalo de confianza al 95 % para $\log \hat{\theta}$ en la población que representa esta muestra es $0.445 \pm 1.96(0.307)$, ó $(-0.157, 1.047)$. El correspondiente intervalo para θ es $[\exp(-0.157), \exp(1.047)]$, ó $(0.85, 2.85)$. La estimación de la razón de posibilidades real es bastante imprecisa.

Dado que el intervalo de confianza para $\hat{\theta}$ contiene el valor 1.0, es posible que las probabilidades reales de muerte por infarto de miocardio sean iguales para la aspirina y el placebo. Si realmente hay un efecto beneficioso en la aspirina pero la razón de posibilidades no es grande, puede requerir un gran tamaño de muestra para mostrar ese beneficio debido al número relativamente pequeño de casos de infarto de miocardio.

La diferencia de proporciones muestral es igual a:

$$\hat{\pi}_1 - \hat{\pi}_2 = 28/684 - 18/676 = 0.04093 - 0.02662 = 0.01431.$$

El estimador del error estándar es $\hat{\sigma}_{\hat{\pi}_1 - \hat{\pi}_2} = 0.00978$.

Un intervalo de confianza al 95 % sería $0.01431 \pm 1.96(0.00978)$, que es lo mismo que $(-0.00486, 0.03348)$.

El riesgo relativo muestral es $r = 0.04093/0.02662 = 1.53757$, el estimador del error estándar del logaritmo de r es $\hat{\sigma}_{\log r} = 0.2972$ y su intervalo de confianza al 95 % $\log(1.53757) \pm 1.96(0.2972)$, que se traduce a $(0.86, 2.75)$.

3.2. Test de Independencia en Tablas de Doble Entrada

Los contrastes de independencia se pueden aplicar tanto para muestreo multinomial (con $I \times J$ categorías) como para muestreo multinomial independiente (para las distintas filas). En el primer caso se contrasta la *independencia* y en el segundo la *homogeneidad*.

El contraste que se plantea es:

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j} \quad \text{para todo } i, j$$

$$H_1 : \pi_{ij} \neq \pi_{i+}\pi_{+j} \quad \text{para algún } i, j$$

Se utiliza el contraste χ^2 de Pearson. Si es cierta H_0 entonces el número esperado de observaciones en cada casilla es:

$$E(n_{ij}) = n\pi_{i+}\pi_{+j} = \mu_{ij}$$

y los estimadores de máxima verosimilitud son:

$$\hat{\mu}_{ij} = n\hat{\pi}_{i+}\hat{\pi}_{+j} = n\frac{n_{i+}}{n}\frac{n_{+j}}{n} = \frac{n_{i+}n_{+j}}{n}$$

Se utiliza el siguiente estadístico:

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$$

Si es cierta H_0 , entonces:

$$X^2 \sim \chi_{(I-1)(J-1)}^2$$

es decir, asintóticamente es un test de la chi-cuadrado.

Otra alternativa es usar el test de razón de verosimilitudes:

$$G^2 = 2 \sum_i \sum_j n_{ij} \log \left(\frac{n_{ij}}{\hat{\mu}_{ij}} \right)$$

también se distribuye asintóticamente como $\chi_{(I-1)(J-1)}^2$ cuando H_0 es cierta.

Se rechaza H_0 para valores grandes de X^2 . La convergencia a la distribución chi-cuadrado es más rápida para X^2 que para G^2 . La aproximación para X^2 puede ser razonablemente buena si las frecuencias esperadas son mayores que 1 y la mayor parte son mayores que 5. Cuando no se pueden aplicar, se pueden utilizar métodos para muestras pequeñas.

Estas pruebas tienen utilidad limitada dado que un p-valor pequeño indica una fuerte asociación pero proporciona poca información sobre la naturaleza o la fuerza de asociación.

Una comparación entre las frecuencias observadas y estimadas de las casillas ayuda a mostrar la naturaleza de la dependencia. Bajo H_0 , grandes diferencias de $(n_{ij} - \hat{\mu}_{ij})$ ocurren en casillas con valores grandes de μ_{ij} .

El *residuo de Pearson* que está definido para cada casilla como:

$$e_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}}{\hat{\mu}_{ij}^{1/2}}$$

intentó adaptarse a esto.

El residuo de Pearson y el estadístico de Pearson se relacionan así:

$$\sum_i \sum_j e_{ij}^2 = X^2$$

Bajo H_0 , los $\{e_{ij}\}$ son asintóticamente normales con media 0 y varianza menor que 1. Un residuo de Pearson estandarizado, que es asintóticamente una normal estándar, resulta de dividirlo por su error estándar. Para H_0 : independencia, esto es:

$$\frac{n_{ij} - \hat{\mu}_{ij}}{[\hat{\mu}_{ij}(1 - p_{i+})(1 - p_{+j})]^{1/2}}$$

Un residuo de Pearson estandarizado que excede aproximadamente 2 ó 3 en valor absoluto indica falta de ajuste de H_0 en esa casilla. Valores grandes son más relevantes cuando los grados de libertad son mayores y es más probable que al menos uno sea grande simplemente por casualidad.

Ejemplo: Educación y Creencias Religiosas

El Cuadro 3.2 clasifica una muestra de 2726 personas de acuerdo a su creencia religiosa y el nivel de educación alcanzado.

Educación	Creencia Religiosa			Total
	Extrema	Moderada	Liberal	
Primaria	178	138	108	424
Secundaria	570	648	442	1660
Bachiller	138	252	252	642
Total	886	1038	802	2726

Cuadro 3.2: Creencias Religiosas y Educación.

El cuadro 3.3 contiene las frecuencias esperadas para H_0 : independencia. Por ejemplo, $\hat{\mu}_{11} = n_{1+}n_{+1}/n = (424 \times 886)/2726 = 137.8$. Los estadísticos chi-cuadrado son $X^2 = 69.2$ y $G^2 = 69.8$, con grados de libertad $(3 - 1)(3 - 1) = 4$. Los p-valores son < 0.0001 . Estos estadísticos proporcionan una evidencia muy fuerte de asociación.

Educación	Creencia Religiosa			Total
	Extrema	Moderada	Liberal	
Primaria	137.8	161.5	124.7	424
Secundaria	539.5	632.1	488.4	1660
Bachiller	208.7	244.5	188.9	642
Total	886	1038	802	2726

Cuadro 3.3: Frecuencias esperadas estimadas.

El Cuadro 3.4 muestra los residuos de Pearson estandarizados para probar la independencia. Por ejemplo, $n_{11} = 178$ y $\hat{\mu}_{11} = 137.8$. Las proporciones marginales son $p_{1+} = 424/2726 = 0.156$ y $p_{+1} = 886/2726 = 0.325$. El residuo de Pearson estandarizado para esta casilla es:

$$(178 - 137.8) / [(137.8)(1 - 0.156)(1 - 0.325)]^{1/2} = 4.5$$

Esta casilla muestra una discrepancia mucho mayor entre n_{11} y $\hat{\mu}_{11}$ de lo esperado si las variables fueran realmente independientes.

La tabla muestra un residuo positivo grande para los sujetos con educación primaria y con creencia religiosa extrema. Lo mismo se observa para los sujetos con bachillerato y creencia religiosa liberal. Lo anterior significa que hay más sujetos en esas combinaciones de lo que predice el supuesto de independencia. Similarmente, existen menos estudiantes con bachillerato y creencia religiosa extrema y menos estudiantes con secundaria y creencia religiosa liberal de lo que predice el supuesto de independencia.

Educación	Creencia Religiosa		
	Extrema	Moderada	Liberal
Primaria	4.5	-2.6	-1.9
Secundaria	2.6	1.3	-4.0
Bachiller	-6.8	0.7	6.3

Cuadro 3.4: Residuos estandarizados de Pearson.

Partición Chi-Cuadrado

Si Z es una normal estándar, Z^2 sigue una distribución chi-cuadrado con un grado de libertad. Una variable aleatoria con distribución chi-cuadrado con v grados de libertad tiene representación $Z_1^2 + Z_2^2 + \dots + Z_v^2$, con Z_i independientes. Luego, el estadístico chi-cuadrado podría partitionarse en v componentes chi-cuadrado con 1 grado de libertad.

Además, si X_1^2 y X_2^2 son variables aleatorias independientes con distribuciones chi-cuadrado con v_1 y v_2 grados de libertad respectivamente, entonces:

$$X^2 = X_1^2 + X_2^2 \sim \chi_{v_1+v_2}^2.$$

Un particionamiento puede mostrar que una asociación refleja básicamente diferencias entre ciertas categorías o grupos de categorías.

Ilustremos el particionamiento con un ejemplo sobre el origen de la esquizofrenia.

La siguiente tabla clasifica una muestra de psiquiatras de acuerdo con el pensamiento de la escuela a la que pertenecen y su opinión sobre el origen de la esquizofrenia.

Escuela	Opinión sobre el Origen		
	Biogénico	Ambiental	Combinación
Ecléctica	90	12	78
Médica	13	1	6
Psico	19	13	50

Cuadro 3.5: Origen de la Esquizofrenia.

Se tiene que $G^2 = 23.04$ con 4 grados de libertad.

Para comprender mejor esta asociación se particiona G^2 en cuatro componentes independientes.

Para la primera subtabla $G_1^2 = 0.29$ con 1 grado de libertad. Para la segunda subtabla $G_2^2 = 1.36$ con 1 grado de libertad.

	Bio	Amb		Bio+Amb	Comb
Ecl	90	12	Ecl	102	78
Med	13	1	Med	14	6

Cuadro 3.6: Subtablas 1 y 2.

Existe poca evidencia de tener diferencias entre el pensamiento de las escuelas Ecléctica y Médica sobre el origen de la esquizofrenia.

Para la tercera subtabla $G_3^2 = 12.95$ con 1 grado de libertad. Para la cuarta subtabla $G_4^2 = 8.43$ con 1 grado de libertad.

	Bio	Amb		Bio+Amb	Comb
Ecl+Med	103	13	Ecl+Med	116	84
Psico	19	13	Psico	32	50

Cuadro 3.7: Subtablas 3 y 4.

Los resultados obtenidos en el análisis de las subtablas 3 y 4 indican que existen diferencias entre el pensamiento de la escuela Psicoanalítica y las otras escuelas (Ecl + Med). Los miembros de la escuela Psicoanalítica son mucho más contundentes que las otras escuelas en atribuir el origen de la esquizofrenia

al Ambiente (subtabla 3) y a una Combinación (subtabla 4). La suma de los cuatro componentes es $G^2 = 23.04$, que es el valor obtenido en la prueba de independencia del Cuadro 3.5.

Reglas del Particionamiento

- La suma de los grados de libertad de las subtablas deben ser igual a los grados de libertad de la tabla completa.
- Cada frecuencia observada en la tabla completa debe aparecer en solamente una de las subtablas.
- Cada total marginal en la tabla completa debe ser un total marginal en solamente una de las subtablas.
- La suma del G^2 en cada subtabla es igual al que se obtiene con la tabla completa. No siempre ocurre lo mismo cuando se utiliza el estadístico X^2 .

Limitaciones de las Pruebas Chi-Cuadrado

Las pruebas de independencia chi-cuadrado sólo indican el grado de evidencia que tiene la hipótesis alternativa de asociación entre las variables. Requieren un tamaño de muestra grande. Los estadísticos X^2 y G^2 no cambian de valor cuando se reordenan las filas y columnas, lo cual indica que las variables son tratadas como nominales. Además, el valor del X^2 depende del tamaño de muestra ya que:

$$X^2 = n \sum_i \sum_j \frac{(\hat{\pi}_{ij} - \hat{\pi}_{i+}\hat{\pi}_{+j})^2}{\hat{\pi}_{i+}\hat{\pi}_{+j}}$$

Cuando la variable fila X y la variable columna Y son ordinales es de esperar una tendencia positiva o negativa. El análisis más popular en este caso asigna puntuaciones a las categorías de X , $u_1 \leq \dots \leq u_I$, y a las categorías de Y , $v_1 \leq \dots \leq v_J$, tales que conservan la ordenación, y resume la *tendencia lineal*. Sea r el coeficiente de correlación entre las puntuaciones. Consideremos el estadístico:

$$M^2 = (n - 1)r^2 \sim \chi^2(1)$$

para muestras grandes. Se rechaza la hipótesis de independencia para valores grandes de M^2 . Un p-valor pequeño no implica una asociación lineal aunque sí que la componente lineal permite rechazar la hipótesis.

3.3. Test de Independencia para Muestras Pequeñas

Todos los procedimientos vistos hasta ahora se basan en distribuciones asintóticas. Si tenemos muestras grandes no hay problemas, pero con muestras pequeñas es preferible usar contrastes exactos.

Test Exacto de Fisher para Tablas 2×2

Consideramos una tabla 2×2 donde la hipótesis nula es de independencia entre las dos variables. Esto corresponde a que la razón de posibilidades es igual a uno, $\theta = 1$. Supongamos que los recuentos de las casillas $\{n_{ij}\}$ provienen de dos muestras aleatorias independientes, o de una única distribución multinomial definida sobre las 4 casillas de la tabla. Se consideran todas las posibles tablas que tienen como totales de fila y columna los valores observados en los datos reales. Entonces, la distribución de los recuentos de las casillas sigue una distribución hipergeométrica.

Si condicionamos los totales de fila y columna, solamente nos queda libre o sin fijar un recuento, por ejemplo, n_{11} , de modo que éste determina los otros tres recuentos de las casillas. Así, la distribución hipergeométrica expresa la probabilidad de los recuentos en términos de sólo n_{11} :

$$p(t) = P(n_{11} = t) = \frac{\binom{n_{1+}}{t} \binom{n_{2+}}{n_{+1} - t}}{\binom{n}{n_{+1}}}$$

donde los posibles valores son:

$$m_- \leq n_{11} \leq m_+$$

con:

$$\begin{aligned} m_- &= \text{máx} \{0, n_{1+} + n_{+1} - n\} \\ m_+ &= \text{mín} \{n_{1+}, n_{+1}\} \end{aligned}$$

Dados los valores totales de las marginales, las tablas que tienen mayores valores n_{11} también tienen mayores valores de la razón de posibilidades.

$$\theta = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

El contraste de independencia entre las dos variables se puede formular como:

$$H_0 : \theta \leq 1$$

$$H_1 : \theta > 1$$

El p-valor es igual a la probabilidad de la distribución hipergeométrica de que n_{11} sea al menos tan grande como el observado, digamos t_0 ,

$$P(n_{11} \geq t_0)$$

de modo que se tiene una evidencia más fuerte a favor de H_1 .

Ejemplo del té

Consideremos un ejemplo clásico de Fisher. Un amigo de Fisher, Bristol, afirmaba que era capaz de distinguir en una taza de té con leche, qué se había echado primero, si té o leche.

Para comprobarlo diseñó un experimento donde se probaban 8 tazas de té. De ellas, en 4 se había echado primero la leche y en las otras 4 primero el té. Se trataba de adivinar en que orden se había echado la leche y el té.

Las tasas se presentaron de manera aleatoria obteniendo los resultados que se muestran en el Cuadro 3.8.

Primero	Predicción		Total
	Leche	Té	
Leche	3	1	4
Té	1	3	4
Total	4	4	

Cuadro 3.8: Experimento de Fisher.

Evidentemente las distribuciones marginales están fijadas en 4.

La distribución bajo la hipótesis nula de n_{11} es una hipergeométrica definida para todas las tablas 2×2 que tienen como marginales $(4, 4)$. Los posibles valores para n_{11} son $(0, 1, 2, 3, 4)$.

En la tabla observada se adivinan 3 de las 4 tazas donde se ha echado primero leche, la probabilidad es:

$$P(3) = \frac{\binom{4}{3}\binom{4}{1}}{\binom{8}{4}} = 0.23$$

Así, para que se apoye la hipótesis $H_1 : \theta > 1$ el único valor extremo que se obtiene es con $n_{11} = 4$, es decir, que se adivinen todas.

$$P(4) = \frac{\binom{4}{4}\binom{4}{0}}{\binom{8}{4}} = 0.014$$

Alternativa Bilateral

La definición de p-valor depende de cómo ordenamos las tablas. Lo que suele ir programado en software es, si se cumple que $p(t) = P(n_{11} = t)$,

$$p = P(p(n_{11}) \leq p(t_0))$$

Sumamos la probabilidad de todas aquellas tablas que son tan probables o menos que la tabla observada.

Otra opción es:

$$p = P(|n_{11} - E(n_{11})| \geq |t_0 - E(n_{11})|)$$

teniendo en cuenta que para la hipergeométrica $E(n_{11}) = n_{1+}n_{+1}/n$.

Este procedimiento equivale a:

$$p = P(X^2 \geq X_0^2)$$

siendo X_0^2 el valor observado de X^2 .

En el ejemplo de las tazas de té, se tiene que:

$$\begin{aligned}
P(n_{11} = 0) &= 0.014 \\
P(n_{11} = 1) &= 0.229 \\
P(n_{11} = 2) &= 0.514 \\
P(n_{11} = 3) &= 0.229 \\
P(n_{11} = 4) &= 0.014
\end{aligned}$$

Por tanto, se suman todas las probabilidades que son menores o iguales que la probabilidad $P(3) = 0.229$ de la tabla observada. Se obtiene,

$$P(0) + P(1) + P(3) + P(4) = 0.486$$

Distribución Condicionada Exacta

Consideremos una tabla $I \times J$ y supongamos que las filas son muestras multinomiales independientes. En consecuencia, los totales de fila $\{n_{i+}\}$ están fijados.

Asumimos la hipótesis de que la distribución condicionada de cada fila es la misma:

$$H_0 : \pi_{j|1} = \dots = \pi_{j|I} = \pi_{+j} \quad \text{para } j = 1, \dots, J$$

y condicionamos ahora los totales de columna.

La distribución condicionada de los conteos es:

$$\frac{\prod_i n_{i+}! \prod_j n_{+j}!}{n! \prod_i \prod_j n_{ij}!}$$

la distribución *hipergeométrica múltiple*.

Cuando hay una única muestra multinomial, los parámetros desconocidos son $\{\pi_{ij}\}$. Para probar la independencia:

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall i, j$$

La distribución resulta del acondicionamiento de los totales de fila y columna. Estos son los estadísticos suficientes $\{\pi_{i+}\}$ y $\{\pi_{+j}\}$, que determinan la distribución bajo la hipótesis nula.

Para ambos modelos de la muestra, los conjuntos marginales se fijan después del acondicionamiento. El resultado final no depende de los parámetros desconocidos y, por lo tanto, permite cálculos de probabilidad exacta.

Test Exacto de Independencia para Tablas $I \times J$

El test exacto para tablas $I \times J$ utiliza la distribución hipergeométrica múltiple. Se define el p-valor como la probabilidad del conjunto de tablas con los márgenes dados que no son más probables de ocurrir que la tabla observada. Otras pruebas exactas ordenan las tablas usando un estadístico que describe la distancia desde H_0 , como por ejemplo X^2 . El p-valor es entonces el valor bajo la hipótesis nula de $P(X^2 \geq X_0^2)$ para el valor observado X_0^2 .

Cuando las clasificaciones tienen las categorías ordenadas un estadístico ordinal es más relevante.

Para la hipótesis alternativa de una asociación positiva, podríamos usar $P(r \geq t_0)$, donde r es la correlación y t_0 el valor observado.

Vamos a ilustrar el test exacto para categorías ordenadas con el Cuadro 3.9, que clasifica de forma cruzada el nivel de tabaquismo y el infarto de miocardio para una muestra de mujeres jóvenes en un estudio de caso-control.

	Nivel de Tabaquismo (cigarros por día)		
	Ninguno	1-24	> 25
Control	25	25	12
Infarto de Miocardio	0	1	3

Cuadro 3.9: Ejemplo para la Prueba Exacta

La segunda fila contiene pequeños recuentos, y las pruebas de muestras grandes pueden ser inapropiadas. Dado los recuentos marginales, la única tabla que tiene mayor evidencia de asociación positiva entre el tabaquismo y el infarto de miocardio tiene recuentos (25,26,11) para la fila 1 y (0,0,4) para la fila 2. Condicionada a ambos conjuntos de márgenes, la probabilidad nula de la tabla observada y esta tabla más extrema es igual a 0.018. Aunque la muestra contiene sólo cuatro pacientes con infarto de miocardio, existe evidencia de asociación positiva. La evidencia es más fuerte que el uso de X^2 , que ignora el ordenamiento de las categorías. El exacto es $P(X^2 \geq X_0^2) = P(X^2 \geq 6.96) = 0.052$.

Capítulo 4

Tablas de Contingencia y R

En este capítulo introduciremos algunas funciones del programa R relacionadas con las tablas de contingencia. Para ello hemos utilizados dos paquetes del programa, “gmodels” y “vcd”.

A continuación, utilizaremos algunos de los ejemplos ya vistos para ilustrar dichas funciones.

Cuadro 2.1: Influencia de la toma de Aspirina respecto a los Ataques Cardíacos

La tabla de contingencia diferencia entre ataques mortales y no mortales. Vamos a combinar estos dos por ahora.

X : Se toma placebo o aspirina.

Y : Se sufre o no ataque cardíaco.

	Ataque	No Ataque
Placebo	189	10.845
Aspirina	104	10.933

Para dibujar la tabla usamos 'ftable'.

```
- datos = c(189,104,10845,10933)
- tabla = cbind(expand.grid(list(Tratamiento = c("Placebo",
"Aspirina"), Resultado = c("Ataque", "No ataque"))),
count = datos)
- ftable(xtabs(count~Tratamiento+Resultado,tabla))
```

Obtenemos:

	Resultado	Ataque	No Ataque
Tratamiento			
Placebo		189	10.845
Aspirina		104	10.933

Para contrastar la igualdad de probabilidades de ataque al corazón por grupo $H_0 : \pi_1 = \pi_2$, donde π_1 es la probabilidad de ataque al corazón tomando aspirina y π_2 la probabilidad de ataque al corazón tomando placebo, se usa 'prop.test'.

```
- x = c(104,189)
- n = c((189+10845),(104+10933))
- prop.test(x,n)
```

Obtenemos:

2-sample test for equality of proportions with continuity correction

```
data: x out of n
X-squared = 24.383, df = 1, p-value = 7.897e-07
alternative hypothesis: two.sided
95 percent confidence interval:
-0.010807464 -0.004590148
sample estimates:
prop 1      prop 2
0.009425412 0.017124219
```

Donde prop 1 y prop 2 son π_1 y π_2 respectivamente.

En el caso de contrastar $H_0 : \pi_a \geq \pi_p$ y $H_1 : \pi_a < \pi_p$, usamos la opción 'alternative'.

```
- prop.test(x,n,alt = "less")
```

Se pueden obtener las proporciones a partir de la componente ‘estimate’ que, en este caso, es un vector numérico de longitud 2. Así, la diferencia de las proporciones se calcula como:

```
- temp = prop.test(x,n)
- names(temp$estimate) = NULL
- temp$estimate[1]-temp$estimate[2]
```

El resultado sería -0.007698806 .

Se puede calcular también el riesgo relativo y la razón de posibilidades.

Riesgo relativo:

```
- temp$estimate[2]/temp$estimate[1]
```

1.816814

Razón de posibilidades:

```
- x[2]*(n[1]-x[1])/(x[1]*(n[2]-x[2]))
```

1.831045

Cuadro 2.7: Sentencia de Pena de Muerte por la Raza del Acusado y la Raza de la Víctima

Raza de la Víctima	Raza del Acusado	Pena de Muerte		Porcentaje de Pena de Muerte
		Sí	No	
Blanco	Blanco	53	414	11.3
	Negro	11	37	22.9
Negro	Blanco	0	16	0.0
	Negro	4	139	2.8
Total	Blanco	53	430	11.0
	Negro	15	176	7.9

Procesamientos por asesinatos múltiples en Florida entre los años 1976 y 1987. Esta es una tabla de contingencia $2 \times 2 \times 2$. El ejemplo de la pena de muerte sirve para ilustrar las razones de posibilidades condicionales. Se estudia el efecto de

la raza del acusado (X) en el veredicto de culpabilidad (Y), tratando a la raza de la víctima (Z) como una variable control.

```
- vic.raza = c("blanca", "negra")
- def.raza = vic.raza
- pena.muerte = c("SI", "NO")
- datalabel = list(acusado = def.raza, muerte = pena.muerte,
victima = vic.raza)
- tabla = expand.grid(acusado = def.raza, muerte = pena.muerte,
victima = vic.raza)
- data = c(53, 11, 414, 37, 0, 4, 16, 139)
- tabla = cbind(tabla, recuento = data)
- xtabs(recuento ~ acusado + muerte + victima, data = tabla)
```

Obtenemos las tablas condicionales según la raza de la víctima:

```
, , victima = blanca
```

	muerte	
acusado	SI	NO
blanca	53	414
negra	11	37

```
, , victima = negra
```

	muerte	
acusado	SI	NO
blanca	0	16
negra	4	139

Calculamos ahora las razones de posibilidades condicionales:

```
- temp = xtabs(recuento ~ acusado + muerte + victima, data = tabla)
- apply(temp, 3, function(x) x[1,1]*x[2,2]/(x[2,1]*x[1,2]))
```

Obtenemos:

blanca	negra
0.4306105	0.0000000

Hay una casilla nula. La función ‘oddsratio’ añade 0.5 a cada casilla de la tabla. Para poder usar esta función usamos el paquete ‘vcd’.

```
- summary(oddsratio(temp,log = F,stratum = 3))
```

Y obtenemos:

```
blanca      negra
0.4208843   0.9393939
```

Cuadro 3.1: Uso de Aspirina e Infarto de Miocardio

Vamos a calcular los intervalos de confianza de la razón de posibilidades, el riesgo relativo y la diferencia de proporciones. Para ello usamos este ejemplo visto en el capítulo 3.

	Infarto	No Infarto
Placebo	28	656
Aspirina	18	658

```
- Medicina <- c("Placebo","Aspirina")
- Infarto <- c("SI","NO")
- tabla <- expand.grid(Medicina = Medicina,Infarto = Infarto)
- datos <- c(28,18,656,658)
- tabla <- cbind(tabla,count = datos)
- temp = tapply(tabla$count,tabla[,1:2],sum)
```

Obtendríamos la tabla anterior.

Vamos a usar ahora la siguiente función que hemos tenido que definir:

```
- Wald.ci <- function(Table,aff.response,alpha = .05) {
- pow <- function(x,a = -1) x^a
- z.alpha <- qnorm(1-alpha/2)
- if(is.character(aff.response))
- where <- eval(parse(text = aff.response))
- else where <- aff.response
- Next <- as.numeric(where == 1) + 1
```

Razón de Posibilidades:

- odds.ratio <- Table[where[1],where[2]]*Table[Next[1],Next[2]]/
(Table[where[1],Next[2]]*Table[Next[1],where[2]])
- se.OR <- sqrt(sum(pow(Table)))
- ci.OR <- exp(log(odds.ratio) + c(-1,1)*z.alpha*se.OR)

Diferencia de Proporciones:

- p1 <- Table[where[1],where[2]]/(n1 <- Table[where[1],Next[2]] +
Table[where[1],where[2]])
- p2 <- Table[Next[1],where[2]]/(n2 <- Table[Next[1],where[2]] +
Table[Next[1],Next[2]])
- se.diff <- sqrt(p1*(1 - p1)/n1 + p2*(1 - p2)/n2)
- ci.diff <- (p1 - p2) + c(-1,1)*z.alpha*se.diff

Riesgo Relativo

- RR <- p1/p2
- se.RR <- sqrt((1 - p1)/(p1*n1) + (1 - p2)/(p2*n2))
- ci.RR <- exp(log(RR) + c(-1,1)*z.alpha*se.RR)
- list(OR = list(odds.ratio = odds.ratio,CI = ci.OR),
proportion.difference = list(diff = p1 - p2,CI = ci.diff),
relative.risk = list(relative.risk = RR,CI = ci.RR)) }

Aplicando la función anterior a nuestros datos se obtienen las estimaciones y los intervalos de confianza para la razón de posibilidades, la diferencia de proporciones y el riesgo relativo.

- Wald.ci(temp,c(1,1))

odds.ratio

1.560298

odds.ratio - CI

0.8546703 2.8485020

proportion.difference

0.01430845

proportion.difference - CI

-0.004868983 0.033485890

relative.risk

1.537362

relative.risk - CI

0.858614 2.752671

Cuadro 3.2: Creencias Religiosas y Educación

Utilizamos este ejemplo para obtener los recuentos esperados que corresponden con el Cuadro 3.3.

Educación	Creencia Religiosa		
	Extrema	Moderada	Liberal
Primaria	178	138	108
Secundaria	570	648	442
Bachiller	138	252	252
Total	886	1038	802

```
- religion.recuento <- c(178,138,108,570,648,442,138,252,252)
- tabla = cbind(expand.grid(list(Religiosidad=c("Extrema",
"Moderada","Liberal"), Grado = c("Primaria","Secundaria",
"Bachiller"))),count = religion.recuento)
- tabla.array = tapply(tabla$count,tabla[,1:2],sum)
(res = chisq.test(tabla.array))
```

Obtenemos:

Pearson's Chi-squared test

data: tabla.array

X-squared = 69.157, df = 4, p-value = 3.42e-14

Recuentos esperados

res\$expected

Religiosidad	Grado		
	Primaria	Secundaria	Bachiller
Extrema	137.8078	539.5304	208.6618
Moderada	161.4497	632.0910	244.4593
Liberal	124.7425	488.3786	188.8789

Cuadro 3.8: Ejemplo del té

El experimento consistía en probar 8 tazas de té. En 4 de ellas se había echado primero la leche y en las otras 4 primero el té. Se trataba de adivinar en que orden se había echado la leche y el té.

Primero	Predicción	
	Leche	Té
Leche	3	1
Té	1	3

Test exacto de Fisher:

```
- fisher.test(matrix(c(3,1,1,3),byrow = T,ncol = 2))
```

Fisher's Exact Test for Count Data

```
data: matrix(c(3,1,1,3), byrow = T, ncol = 2)
p-value = 0.4857
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.2117329 621.9337505
sample estimates:
odds ratio
6.408309
```

Obtenemos el p-valor 0.4857 que corresponde con el calculado anteriormente:
 $P(0) + P(1) + P(3) + P(4) = 0.486$.

Bibliografía

- [1] Alan Agresti. *Categorical Data Analysis*. Second Edition, Gainesville, Florida (2002).
- [2] Erling B. Andersen. *Introduction to the Statistical Analysis os Categorical Data*. New York (1997).
- [3] Chris J. LLoyd. *Statistical Analysis of Categorical Data*. New York (1999).
- [4] George Casella, Roger L. Berger *Statistical Inference*. Second Edition, Pacific Grove, California (2002).
- [5] Vijay K. Rohatgi, A. K. MD. Ehsanes Saleh. *An introduction to probability and statistic*. Wiley (2001).
- [6] W. N. Venables, D. M. Smith, the R Development Core Team. *An Introduction to R*. (2009).