

## 7. Estadística no paramétrica

La estadística no paramétrica es una rama de la estadística que estudia las pruebas y modelos estadísticos cuya distribución subyacente no se ajusta a los llamados criterios paramétricos. Su distribución no puede ser definida a priori, pues son los datos observados los que la determinan. La utilización de estos métodos se hace recomendable cuando no se puede asumir que los datos se ajusten a una distribución conocida, cuando el nivel de medida empleado no sea, como mínimo, de intervalo.

Las principales pruebas no paramétricas son las siguientes:

Prueba  $\chi^2$  de Pearson

Prueba binomial

Prueba de Anderson-Darling

Prueba de Cochran

Prueba de Cohen kappa

Prueba de Fisher

Prueba de Friedman

Prueba de Kendall

Prueba de Kolmogórov-Smirnov

Prueba de Kruskal-Wallis

Prueba de Kuiper

Prueba de Mann-Whitney o prueba de Wilcoxon

Prueba de McNemar

Prueba de la mediana

Prueba de Siegel-Tukey

Prueba de los signos

Coefficiente de correlación de Spearman

Tablas de contingencia

## Prueba de Wald-Wolfowitz

### Prueba de los rangos con signo de Wilcoxon

La mayoría de estos test estadísticos están programados en los paquetes estadísticos más frecuentes, quedando para el investigador, simplemente, la tarea de decidir por cuál de todos ellos guiarse o qué hacer en caso de que dos test nos den resultados opuestos. Hay que decir que, para poder aplicar cada uno existen diversas hipótesis nulas y condiciones que deben cumplir nuestros datos para que los resultados de aplicar el test sean fiables. Esto es, no se puede aplicar todos los test y quedarse con el que mejor convenga para la investigación sin verificar si se cumplen las hipótesis y condiciones necesarias pues, si se violan, invalidan cualquier resultado posterior y son una de las causas más frecuentes de que un estudio sea estadísticamente incorrecto. Esto ocurre sobre todo cuando el investigador desconoce la naturaleza interna de los test y se limita a aplicarlos sistemáticamente.

Es importante mencionar que si la distribución de los datos se ajusta a un tipo de distribución conocida, existen otras [pruebas] que, en la práctica, son más aconsejables pero que así mismo requieren otros supuestos. En este caso, la estadística a emplear es la estadística paramétrica, dentro de la cual muchas veces podemos encontrar equivalencias entre pruebas pero con diferencias en la potencia entre ambas siendo siempre la potencia de las pruebas no paramétricas menor que la potencia de las pruebas paramétricas equivalentes. Aun así, el uso adecuado de los tamaños muestrales disminuye la posibilidad de cometer un [error tipo II], puesto que aumenta al mismo tiempo la eficacia de la prueba. Es decir, a medida que se aumenta el tamaño de la muestra, disminuye la posibilidad de cometer un error tipo II (un falso negativo: No rechazar la hipótesis nula cuando ésta en realidad es falsa).

### 7.1 Prueba $\chi^2$ de Pearson

La **prueba  $\chi^2$  de Pearson** es considerada como una prueba no paramétrica que mide la discrepancia entre una distribución observada y otra teórica (bondad de ajuste), indicando en qué medida las diferencias existentes entre ambas, de haberlas, se deben al azar en el contraste de hipótesis. También se utiliza para probar la independencia de dos variables entre sí, mediante la presentación de los datos en tablas de contingencia.

La fórmula que da el estadístico es la siguiente:

$$\chi^2 = \sum_i \frac{(\text{observada}_i - \text{teórica}_i)^2}{\text{teórica}_i}$$

Cuanto mayor sea el valor de  $\chi^2$ , menos verosímil es que la hipótesis sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones.

Los grados de libertad **gl** vienen dados por:

$gl = (r-1)(k-1)$ . Donde  $r$  es el número de filas y  $k$  el de columnas.

- Criterio de decisión:

No se rechaza  $H_0$  cuando  $\chi^2 < \chi_t^2(r-1)(k-1)$ . En caso contrario sí se rechaza.

Donde  $t$  representa el valor proporcionado por las tablas, según el nivel de significación estadística elegido.

## 7.2 Prueba de Anderson-Darling

En estadística, la **prueba de Anderson-Darling** es una prueba no paramétrica sobre si los datos de una muestra provienen de una distribución específica. La fórmula para el estadístico  $A$  determina si los datos  $\{Y_1 < \dots < Y_N\}$  (observar que los datos se deben ordenar) vienen de una distribución con función acumulativa  $F$

$$A^2 = -N - S$$

donde

$$S = \sum_{k=1}^N \frac{2k-1}{N} [\ln F(Y_k) + \ln (1 - F(Y_{N+1-k}))]$$

El estadístico de la prueba se puede entonces comparar contra las distribuciones del estadístico de prueba (dependiendo que  $F$  se utiliza) para determinar el P-valor.

La prueba de Anderson-Darling es una prueba estadística que permite determinar si una muestra de datos se extrae de una distribución de probabilidad. En su forma básica, la prueba asume que no existen parámetros a estimar en la distribución que se está probando, en cuyo caso la prueba y su conjunto de valores críticos siguen una distribución libre. Sin embargo, la prueba se utiliza con mayor frecuencia en

contextos en los que se está probando una familia de distribuciones, en cuyo caso deben ser estimados los parámetros de esa familia y debe tenerse estos en cuenta a la hora de ajustar la prueba estadística y sus valores críticos. Cuando se aplica para probar si una distribución normal describe adecuadamente un conjunto de datos, es una de las herramientas estadísticas más potentes para la detección de la mayoría de las desviaciones de la normalidad.

### 7.3 Coeficiente kappa de Cohen

El **Coeficiente kappa de Cohen** es una medida estadística que ajusta el efecto del azar en la proporción de la concordancia observada., para los elementos cualitativos (variables categóricas). En general se cree que es una medida más robusta que el simple cálculo del porcentaje de concordancia, ya que  $\kappa$  tiene en cuenta el acuerdo que ocurren por casualidad. Algunos investigadores<sup>2</sup> han expresado su preocupación por la tendencia de  $\kappa$  tomar frecuencias las categorías observadas 'como dados, lo que puede tener el efecto de subestimar a un acuerdo para una categoría que es también de uso general; por esta razón,  $\kappa$  es considerado una medida excesivamente conservadora de acuerdo.

Otros <sup>3</sup> impugnan la afirmación de que el kappa "tiene en cuenta" la posibilidad de acuerdo. Para hacerlo con eficacia requeriría un modelo explícito de cómo la oportunidad afecta las decisiones de calificadores. El llamado ajuste oportunidad de las estadísticas kappa supone que, cuando no completamente cierto, los evaluadores simplemente suponen un escenario muy poco realista.

#### Cálculo

El Coeficiente kappa de Cohen mide el acuerdo entre dos evaluadores que cada clasificar los elementos de N en C categorías mutuamente excluyentes. La primera mención de una estadística kappa como se atribuye a Galton (1892), ver Smeeton (1985).

La ecuación para  $\kappa$  es:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)},$$

donde  $\Pr(a)$  es el acuerdo observado relación entre los evaluadores, y  $\Pr(e)$  es la probabilidad hipotética de la posibilidad de acuerdo, utilizando los datos observados para calcular las probabilidades de cada observador diciendo aleatoriamente cada categoría. Si los evaluadores son completamente de acuerdo entonces  $\kappa = 1$ . Si no

hay acuerdo entre los calificadores distintos de lo que cabría esperar por azar (según lo definido por  $\Pr(e)$ ),  $\kappa = 0$ .

El artículo seminal introducir kappa como una nueva técnica fue publicado por Jacob Cohen en la revista Educación y Psicología de medición en 1960.<sup>6</sup>

Una estadística similar, llamada pi, fue propuesta por Scott (1955). Kappa de Cohen y pi de Scott difieren en términos de cómo se calcula  $\Pr(e)$ .

Tenga en cuenta que la kappa de Cohen mide el acuerdo entre sólo dos evaluadores. Para que una medida similar de acuerdo (kappa de Fleiss) utiliza cuando hay más de dos evaluadores, ver Fleiss (1971). El Fleiss Kappa, sin embargo, es una generalización múltiples evaluadores de pi de Scott estadística, no kappa de Cohen.

## 7.4 Prueba de Friedman

En estadística la **prueba de Friedman** es una prueba no paramétrica desarrollado por el economista Milton Friedman. Equivalente a la prueba ANOVA para medidas repetidas en la versión no paramétrica, el método consiste en ordenar los datos por filas o bloques, reemplazándolos por su respectivo orden. Al ordenarlos, debemos considerar la existencia de datos idénticos.

### 7.4.1 Método

Sea  $\{x_{ij}\}_{m \times n}$  una tabla de datos, donde  $m$  son las filas (*bloques*) y  $n$  las columnas (*tratamientos*). Una vez calculado el orden de cada dato en su bloque, reemplazamos al tabla original con otra  $\{r_{ij}\}_{m \times n}$  donde el valor  $r_{ij}$  es el orden de  $x_{ij}$  en cada bloque  $i$ .

1. Cálculo de las varianzas intra e inter grupo:

$$\begin{aligned} SS_t &= n \sum_{j=1}^m (\bar{r}_j - \bar{r})^2 \\ SS_e &= \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (r_{ij} - \bar{r})^2 \\ \bar{r}_j &= \frac{1}{m} \sum_{i=1}^m r_{ij} \end{aligned}$$

$$\bar{r} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n r_{ij}$$

2. El estadístico viene dado por  $Q = \frac{SS_t}{SS_e}$ .
3. El criterio de decisión es  $\mathbf{P}(\chi_{n-1}^2 \geq Q)$ .

## 7.5 Prueba de Kolmogórov-Smirnov

En estadística, la prueba de **Kolmogórov-Smirnov** (también prueba **K-S**) es una prueba no paramétrica que se utiliza para determinar la bondad de ajuste de dos distribuciones de probabilidad entre sí.

En el caso de que queramos verificar la normalidad de una distribución, la prueba de Lilliefors conlleva algunas mejoras con respecto a la de Kolmogórov-Smirnov; y, en general, el test de Shapiro–Wilk o la prueba de Anderson-Darling son alternativas más potentes.

Conviene tener en cuenta que la prueba Kolmogórov-Smirnov es más sensible a los valores cercanos a la mediana que a los extremos de la distribución. La prueba de Anderson-Darling proporciona igual sensibilidad con valores extremos.

### 7.5.1 Estadístico

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{si } y_i \leq x, \\ 0 & \text{alternativa.} \end{cases}$$

Para dos colas el estadístico viene dado por

$$D_n^+ = \max(F_n(x) - F(x))$$

$$D_n^- = \max(F(x) - F_n(x))$$

donde  $F(x)$  es la distribución presentada como hipótesis.

## 7.6 Prueba de Kruskal-Wallis

En estadística, la **prueba de Kruskal-Wallis** (de William Kruskal y W. Allen Wallis) es un método no paramétrico para probar si un grupo de datos proviene de la misma población. Intuitivamente, es idéntico al ANOVA con los datos reemplazados por categorías. Es una extensión de la prueba de la U de Mann-Whitney para 3 o más grupos.

Ya que es una prueba no paramétrica, la prueba de Kruskal-Wallis no asume normalidad en los datos, en oposición al tradicional ANOVA. Sí asume, bajo la hipótesis nula, que los datos vienen de la misma distribución. Una forma común en que se viola este supuesto es con datos heterocedásticos.

### 7.6.1 Método

$$K = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_{i\cdot} - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2},$$

El estadístico está dado por:

- $n_i$  es el número de observaciones en el grupo  $i$
- $r_{ij}$  es el rango (entre todas las observaciones) de la observación  $j$  en el grupo  $i$
- $N$  es el número total de observaciones entre todos los grupos
- $\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$
- $\bar{r} = (N + 1)/2$  es el promedio de  $r_{ij}$ .

Note que el denominador de la expresión para  $K$  es exactamente  $\frac{(N - 1)N(N + 1)}{12}$ . Luego  $K = \frac{12}{N(N + 1)} \sum_{i=1}^g n_i (\bar{r}_{i\cdot} - \bar{r})^2$ .

2. Se puede realizar una corrección para los valores repetidos dividiendo  $K$  por  $1 - \frac{\sum_{i=1}^G (t_i^3 - t_i)}{N^3 - N}$ , donde  $G$  es el número de grupos de diferentes rangos repetidos, y  $t_i$  es el número de observaciones repetidas dentro del grupo  $i$  que tiene observaciones repetidas para un determinado valor. Esta corrección hace cambiar a  $K$  muy poco al menos que existan un gran número de observaciones repetidas.
3. Finalmente, el  $p$ -value es aproximado por  $\Pr(\chi_{g-1}^2 \geq K)$ . Si algún  $n_i$  es pequeño ( $< 5$ ) la distribución de  $K$  puede ser distinta de la chi-cuadrado.

### 7.7 Prueba U de Mann-Whitney

En estadística la prueba U de Mann-Whitney (también llamada de Mann-Whitney-Wilcoxon, prueba de suma de rangos Wilcoxon, o prueba de Wilcoxon-Mann-Whitney) es una prueba no paramétrica aplicada a dos muestras independientes. Es, de hecho, la versión no paramétrica de la habitual prueba t de Student.

Fue propuesto inicialmente en 1945 por Frank Wilcoxon para muestras de igual tamaños y extendido a muestras de tamaño arbitrario como en otros sentidos por Henry B. Mann y D. R. Whitney en 1947.

### 7.7.1 Planteamiento de la prueba

La prueba de Mann-Whitney se usa para comprobar la heterogeneidad de dos muestras ordinales. El planteamiento de partida es:

1. Las observaciones de ambos grupos son independientes
2. Las observaciones son variables ordinales o continuas.
3. Bajo la hipótesis nula, la distribución de partida de ambos grupos es la misma y,
4. Bajo la hipótesis alternativa, los valores de una de las muestras *tienden a exceder* a los de la otra:  $P(X > Y) + 0.05 P(X = Y) > 0.05$ .

### 7.7.2 Cálculo del estadístico

Para calcular el estadístico  $U$  se asigna a cada uno de los valores de las dos muestras su rango para construir

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$
$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

donde  $n_1$  y  $n_2$  son los tamaños respectivos de cada muestra;  $R_1$  y  $R_2$  es la suma de los rangos de las observaciones de las muestras 1 y 2 respectivamente.

El estadístico  $U$  se define como el mínimo de  $U_1$  y  $U_2$ .

Los cálculos tienen que tener en cuenta la presencia de observaciones idénticas a la hora de ordenarlas. No obstante, si su número es pequeño, se puede ignorar esa circunstancia.

### 7.7.3 Distribución del estadístico

La prueba calcula el llamado estadístico  $U$ , cuya distribución para muestras con más de 20 observaciones se aproxima bastante bien a la distribución normal.



La aproximación a la normal,  $z$ , cuando tenemos muestras lo suficientemente grandes viene dada por la expresión:

$$z = (U - m_U) / \sigma_U$$

Donde  $m_U$  y  $\sigma_U$  son la media y la desviación estándar de  $U$  si la hipótesis nula es cierta, y vienen dadas por las siguientes fórmulas:

$$m_U = n_1 n_2 / 2.$$
$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}.$$

#### 7.7.4 Implementaciones

- R tiene una implementación del test (al que se refiere como el *Wilcoxon two-sample test*) mediante `wilcox.test` (y para el caso de datos pareados, `wilcox.exact` en el paquete `exactRankTests` o con la opción `exact=FALSE`).
- Existe una librería de java para realizar este Test y otros muchos más. Su nombre es "Commons Math" y está dentro de un programa de trabajo de Apache.

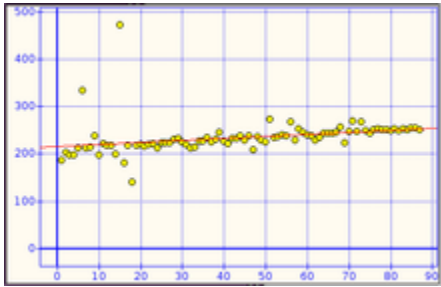
#### 7.8 Prueba de la mediana

La prueba de la mediana es una prueba no paramétrica que podemos considerar un caso especial de la prueba de chi-cuadrado, pues se basa en esta última. Su objetivo es comparar las medianas de dos muestras y determinar si pertenecen a la misma población o no.

Para ello, se calcula la mediana de todos los datos conjuntamente. Después, se divide cada muestra en dos subgrupos: uno para aquellos datos que se sitúen por encima de la mediana y otro para los que se sitúen por debajo. La prueba de chi-cuadrado determinará si las frecuencias observadas en cada grupo difieren de las esperadas con respecto a una distribución de frecuencias que combine ambas muestras.

Esta prueba está especialmente indicada cuando los datos sean extremos o estén sesgados.

## 7.9 Coeficiente de correlación de Spearman



El coeficiente de correlación de Spearman es menos sensible que el de Pearson para los valores muy lejos de lo esperado. En este ejemplo: Pearson = 0.30706 Spearman = 0.76270

En estadística, el **coeficiente de correlación de Spearman**,  $\rho$  (rho) es una medida de la correlación (la asociación o interdependencia) entre dos variables aleatorias continuas. Para calcular  $\rho$ , los datos son ordenados y reemplazados por su respectivo orden.

El estadístico  $\rho$  viene dado por la expresión:

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

donde D es la diferencia entre los correspondientes estadísticos de orden de x - y. N es el número de parejas.

Se tiene que considerar la existencia de datos idénticos a la hora de ordenarlos, aunque si éstos son pocos, se puede ignorar tal circunstancia

Para muestras mayores de 20 observaciones, podemos utilizar la siguiente aproximación a la distribución t de Student

$$t = \frac{\rho}{\sqrt{(1 - \rho^2)/(n - 2)}}$$

La interpretación de coeficiente de Spearman es igual que la del coeficiente de correlación de Pearson. Oscila entre -1 y +1, indicándonos asociaciones negativas o positivas respectivamente, 0 cero, significa no correlación pero no independencia.

La tau de Kendall es un coeficiente de correlación por rangos, inversiones entre dos ordenaciones de una distribución normal bivariante.

### 7.9.1 Ejemplo

Los datos brutos usados en este ejemplo se ven debajo.

CI Horas de TV a la semana

106 7

86 0

100 28

100 50

99 28

103 28

97 20

113 12

113 7

110 17

El primer paso es ordenar los datos de la primera columna. Después, se crean dos columnas más. Ambas son para ordenar (establecer un lugar en la lista) de las dos primeras columnas. Después se crea una columna "d" que muestra las diferencias entre las dos columnas de orden. Finalmente, se crea otra columna "d<sup>2</sup>". Esta última es sólo la columna "d" al cuadrado.

Después de realizar todo esto con los datos del ejemplo, se debería acabar con algo como lo siguiente:

CI (i)	Horas de TV a la semana (t)	orden(i)	orden(t)	d	d <sup>2</sup>
86	0	1	1	0	0
97	20	2	6	4	16
99	28	3	8	5	25
100	50	4.5	10	5.5	30.25

100	28	4.5	8	3.5	12.25
103	28	6	8	2	4
106	7	7	2.5	4.5	20.25
110	17	8	5	3	9
113	7	9.5	2.5	7	49
113	12	9.5	4	5.5	30.25

Nótese como el número de orden de los valores que son idénticos es la media de los números de orden que les corresponderían si no lo fueran.

Los valores de la columna  $d^2$  pueden ser sumados para averiguar  $\sum d_i^2 = 196$ . El valor de  $n$  es 10. Así que esos valores pueden ser sustituidos en la fórmula.

$$\rho = 1 - \frac{6 \times 196}{10(10^2 - 1)}$$

De lo que resulta  $\rho = -0.187878787879$ .

### 7.9.2 Determinando la significación estadística

La aproximación moderna al problema de averiguar si un valor observado de  $\rho$  es significativamente diferente de cero (siempre tendremos  $-1 \leq \rho \leq 1$ ) es calcular la probabilidad de que sea mayor o igual que el  $\rho$  esperado, dada la hipótesis nula, utilizando un test de permutación. Esta aproximación es casi siempre superior a los métodos tradicionales, a no ser que el conjunto de datos sea tan grande que la potencia informática no sea suficiente para generar permutaciones (poco probable con la informática moderna), o a no ser que sea difícil crear un algoritmo para crear permutaciones que sean lógicas bajo la hipótesis nula en el caso particular de que se trate (aunque normalmente estos algoritmos no ofrecen dificultad).

Aunque el test de permutación es a menudo trivial para cualquiera con recursos informáticos y experiencia en programación, todavía se usan ampliamente los métodos tradicionales para obtener significación. La aproximación más básica es comparar el  $\rho$  observado con tablas publicadas para varios niveles de significación. Es una solución simple si la significación sólo necesita saberse dentro de cierto rango, o ser menor de un determinado valor, mientras haya tablas disponibles que especifiquen los rangos adecuados. Más abajo hay una referencia a una tabla semejante. Sin embargo, generar estas tablas es computacionalmente intensivo y a

lo largo de los años se han usado complicados trucos matemáticos para generar tablas para tamaños de muestra cada vez mayores, de modo que no es práctico para la mayoría extender las tablas existentes.

Una aproximación alternativa para tamaños de muestra suficientemente grandes es una aproximación a la distribución t de Student. Para tamaños de muestra más grandes que unos 20 individuos, la variable

$$t = \frac{\rho}{\sqrt{(1 - \rho^2)/(n - 2)}}$$

tiene una distribución t de Student en el caso nulo (correlación cero). En el caso no nulo (ej: para averiguar si un  $\rho$  observado es significativamente diferente a un valor teórico o si dos  $\rho$ s observados difieren significativamente), los tests son mucho menos potentes, pero puede utilizarse de nuevo la distribución t.

Una generalización del coeficiente de Spearman es útil en la situación en la cual hay tres o más condiciones, varios individuos son observados en cada una de ellas, y predecimos que las observaciones tendrán un orden en particular. Por ejemplo, un conjunto de individuos pueden tener tres oportunidades para intentar cierta tarea, y predecimos que su habilidad mejorará de intento en intento. Un test de la significación de la tendencia entre las condiciones en esta situación fue desarrollado por E. B. Page y normalmente suele conocerse como Page's trend test para alternativas ordenadas.

### 7.10 Tabla de contingencia

En estadística las **tablas de contingencia** se emplean para registrar y analizar la relación entre dos o más variables, habitualmente de naturaleza cualitativa (nominales u ordinales).

Supóngase que se dispone de dos variables, la primera el sexo (hombre o mujer) y la segunda recoge si el individuo es zurdo o diestro. Se ha observado esta pareja de variables en una muestra aleatoria de 100 individuos. Se puede emplear una tabla de contingencia para expresar la relación entre estas dos variables:

	<b>Diestro</b>	<b>Zurdo</b>	<b>TOTAL</b>
<b>Hombre</b>	43	9	52
<b>Mujer</b>	44	4	48
<b>TOTAL</b>	87	13	100

Las cifras en la columna de la derecha y en la fila inferior reciben el nombre de **frecuencias marginales** y la cifra situada en la esquina inferior derecha es el **gran total**.

La tabla nos permite ver de un vistazo que la proporción de hombres diestros es aproximadamente igual a la proporción de mujeres diestras. Sin embargo, ambas proporciones no son idénticas y la significación estadística de la diferencia entre ellas puede ser evaluada con la prueba  $\chi^2$  de Pearson, supuesto que las cifras de la tabla son una muestra aleatoria de una población. Si la proporción de individuos en cada columna varía entre las diversas filas y viceversa, se dice que existe asociación entre las dos variables. Si no existe asociación se dice que ambas variables son independientes.

El grado de asociación entre dos variables se puede evaluar empleando distintos coeficientes: el más simple es el coeficiente phi que se define por

$$\phi = \sqrt{\frac{\chi^2}{N}},$$

donde  $\chi^2$  se deriva del test de Pearson, y N es el total de observaciones -el gran total-.  $\Phi$  puede oscilar entre 0 (que indica que no existe asociación entre las variables) e infinito. A diferencia de otras medidas de asociación, el coeficiente  $\Phi$  de Cramer no está acotado.

### 7.10.1 Estudio de diferencia de proporciones

Hay situaciones en las que tenemos probabilidades de éxito cercanas al cero o del uno en donde las proporciones pueden ser poco representativas sobre el comportamiento dentro de los grupos. Por ejemplo:

- Si  $\pi_1 = 0.01$   $\pi_2 = 0.001$

$$d = \pi_1 - \pi_2 = 0.009$$

- Si  $\pi_1 = 0.41$   $\pi_2 = 0.401$

$$d = \pi_1 - \pi_2 = 0.009$$

Vamos a definir el **riesgo relativo** como  $r = \pi_1 / \pi_2$

Para los ejemplos anteriores:  $r = 0.01 / 0.001 = 10$   $r = 0.41 / 0.401 = 1.02$

En el primer caso el éxito dentro de los grupos es 10 veces mayor que en el otro.

Si X e Y independientes  $\rightarrow \pi_1 = \pi_2$  con lo que su riesgo relativo es  $r = \pi_1 / \pi_2 = 1$

Ahora bien, ¿cómo estimar r?

$$r' = p_1/p_2$$

En el ejemplo de más arriba:

$r' = (43/52) / (44/48) = 0.902$   $\rightarrow$  la proporción de éxito (diestro) dentro de las mujeres es alrededor de un 10% mayor que dentro del grupo de los hombres.

## 7.11 Prueba de los rangos con signo de Wilcoxon

La **prueba de los rangos con signo de Wilcoxon** es una prueba no paramétrica para comparar la media de dos muestras relacionadas y determinar si existen diferencias entre ellas. Se utiliza como alternativa a la prueba t de Student cuando no se puede suponer la normalidad de dichas muestras. Debe su nombre a Frank Wilcoxon, que la publicó en 1945.

Se utiliza cuando la variable subyacente es continua pero no se presupone ningún tipo de distribución particular.

### 7.11.1 Planteamiento

Suponga que se dispone de n pares de observaciones, denominadas  $(x_i, y_i)$ . El objetivo del test es comprobar si puede dictaminarse que los valores  $x_i$  e  $y_i$  son o no iguales.

### 7.11.2 Suposiciones

1. Si  $z_i = y_i - x_i$ , entonces los valores  $z_i$  son independientes.
2. Los valores  $z_i$  tienen una misma distribución continua y simétrica respecto a una mediana común  $\theta$ .

### 7.11.3 Método

La hipótesis nula es  $H_0: \theta = 0$ . Retrotrayendo dicha hipótesis a los valores  $x_i, y_i$  originales, ésta vendría a decir que son en cierto sentido del mismo tamaño.

Para verificar la hipótesis, en primer lugar, se ordenan los valores absolutos  $|z_1|, \dots, |z_n|$  y se les asigna su rango  $R_i$ . Entonces, el estadístico de la prueba de los signos de Wilcoxon,  $W^+$ , es

$$W^+ = \sum_{z_i > 0} R_i,$$

es decir, la suma de los rangos  $R_i$  correspondientes a los valores positivos de  $z_i$ .

La distribución del estadístico  $W^+$  puede consultarse en tablas para determinar si se acepta o no la hipótesis nula.

En ocasiones, esta prueba se usa para comparar las diferencias entre dos muestras de datos tomados antes y después del tratamiento, cuyo valor central se espera que sea cero. Las diferencias iguales a cero son eliminadas y el valor absoluto de las desviaciones con respecto al valor central son ordenadas de menor a mayor. A los datos idénticos se les asigna el lugar medio en la serie. La suma de los rangos se hace por separado para los signos positivos y los negativos.  $S$  representa la menor de esas dos sumas. Comparamos  $S$  con el valor proporcionado por las tablas estadísticas al efecto para determinar si rechazamos o no la hipótesis nula, según el nivel de significación elegido.