Capítulo 12

CONTRASTES NO PARAMÉTRICOS

1. INTRODUCCIÓN

En ocasiones, el investigador no está seguro acerca del tipo de distribución de probabilidad de la que proceden las observaciones muestrales que ha recogido, y le parece muy arriesgado hacer un supuesto concreto acerca de la misma. Los procedimientos que hemos analizado en capítulos anteriores consideran la situación contraria, en que se hace un supuesto específico acerca de la familia de distribuciones de probabilidad que generó la muestra, y se procede a estimar los parámetros que caracterizan la distribución correspondiente: Poisson, Normal, etc., y posiblemente se lleven a cabo contrastes de hipótesis acerca de valores numéricos de dichos parámetros. Vamos a considerar en este capítulo situaciones en que el investigador no está interesado en los valores de parámetros concretos, entre otras razones, porque tampoco está muy seguro acerca de la familia de distribuciones de probabilidad con la que está trabajando, y quiere aprender algo acerca de la misma.

En tal situación, el investigador puede estar interesado, precisamente, en detectar el tipo de distribución de que procede su muestra, es decir, en llevar a cabo : a) contrastes de *ajuste de una distribución muestral a una distribución teórica* y, en particular, contrastes de Normalidad, que son el primer tipo de contrastes que veremos en este Capítulo. El segundo tipo de preguntas se refiere a si dos o más muestras independientes proceden de una misma distribución de probabilidad, sin necesidad de especificar de qué tipo es ésta: b) *contrastes de homogeneidad entre distribuciones*. El tercer tipo de cuestiones se refiere a la posible independencia de distintas características observadas en la muestra, con independencia del tipo de distribución que siga cada una de ellas: d) *contrastes de independencia entre características* muestrales.

Todos estos contrastes son denominados *no paramétricos* o *independientes de la distribución*, a diferencia de los *paramétricos*, aquellos que se basan en un supuesto específico acerca de la distribución de probabilidad poblacional, como son tanto los de tipo t de Student, como los de chi-cuadrado y los contrastes F que vimos en el Capítulo XX, y que requieren el conocimiento o la estimación de los valores paramétricos relevantes.

Estos contrastes son generalmente rápidos y fáciles de utilizar, y apenas precisan ningún supuesto. Muchos de ellos tienen la enorme ventaja de ser aplicables a datos cualitativos, cuya clasificación ha de ser necesariamente ordinal, situación típica en la que no podríamos aplicar muchos contrastes paramétricos, que se basan en la consideración de una distribución de probabilidad de tipo continuo: Normal, t de Student, etc. Tampoco es obvio que en las situaciones en las que parece suficientemente adecuado efectuar un supuesto concreto acerca de la distribución poblacional, como puede ser la Normalidad, sea generalmente preferible utilizar contrastes paramétricos, pues los contrastes no paramétricos tienen algunas ventajas, entre ellas el tener expresiones sencillas y bastante intuitivas, y son válidos en muestras muy cortas, en las que los procedimientos paramétricos pueden no serlo. Además, no requieren la estimación de parámetros de ninguna distribución de probabilidad.

2. CONTRASTES DE AJUSTE A UNA DISTRIBUCIÓN TEÓRICA

2.1 Contrastes basados en la distribución de frecuencias muestral

2.1.1 El contraste chi-cuadrado

La idea básica de este contraste consiste en comparar las frecuencias observadas en la muestra para cada suceso relevante, con las que deberían haberse obtenido en una población que perteneciese a una distribución de probabilidad específica. La hipótesis alternativa es que la muestra procede de una distribución de probabilidad diferente de la que utilizamos en el contraste. El contraste puede aplicarse tanto a distribuciones discretas como continuas, pero previamente el investigador debe establecer una partición del espacio muestral en k sucesos mutuamente exclusivos.

En este tipo de contraste, la hipótesis nula recoge una determinada distribución de probabilidad teórica, que implica una determinada *probabilidad* p_i para cada uno de los k sucesos considerados, con $\sum_{i=1}^{k} p_i = 1$. La probabilidad p_i, multiplicada por el tamaño muestral, n, proporciona la frecuencia teórica $T_i = np_i$ que debería haberse observado en la muestra, de ser la hipótesis nula cierta, para cada uno de los k sucesos, con: $\sum_{i=1}^{k} np_i = n\sum_{i=1}^{k} p_i = n$. Por otra parte, cada uno de tales sucesos tiene su propia frecuencia muestral observada, $O_i = n_i$, con $\sum_{i=1}^{k} n_i = n$. Si la muestra procediera efectivamente, de la distribución de probabilidad que hemos recogido en la hipótesis nula, ambas frecuencias, la teórica, T_i, y la empírica, O_i, es decir, la observada en la muestra, no deberían diferir mucho.

El estadístico:

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(O_i - T_i)^2}{T_i} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

propuesto por Pearson, sigue una distribución chi-cuadrado con un número de grados de libertad igual al número de sucesos diferentes considerados, k, menos uno [ver Apéndice 1]. No todas las frecuencias de una distribución multinomial son independientes entre sí, puesto que el tamaño muestral n está dado, y la suma de las k frecuencias debe ser igual a n. Existe, por tanto, una restricción entre todas ellas, y sólo k-1 son independientes. Por eso es que la distribución de probabilidad del estadístico de Pearson es χ_{k-1}^2. Para que la distribución chi-cuadrado sea una aproximación razonable, es preciso que las frecuencias observadas en la muestra para cada suceso sean mayores o iguales a 5. Si algún suceso tiene frecuencia inferior, debe agruparse con otros.

Como en algunas ocasiones anteriores, estamos utilizando nuevamente una distribución de tipo continuo para aproximar variables discretas. Es conveniente, por tanto, utilizar la *corrección de continuidad de Yates* en el cálculo del estadístico chi-cuadrado, cuando sólo existe un grado de libertad, es decir, cuando las dos variables en consideración pueden presentarse en la forma de tan sólo dos alternativas diferentes, y las frecuencias de casillas son pequeñas:

$$\chi^2 = \sum_{i=1}^{m} \frac{(|O_i - E_i| - 0{,}5)^2}{E_i}$$

En muestras pequeñas, en que cada frecuencia esperada se encuentra entre 5 y 10, es conveniente comparar ambos valores de la distribución chi-cuadrado, corregido y sin corregir, y confiar en obtener la concordancia entre los resultados que de ambas se derivan.

Por último, hay que tener en cuenta que si es preciso estimar algún parámetro para caracterizar la distribución teórica, entonces el número de grados de libertad de la distribución chi-cuadrado del estadístico de Pearson pierde un grado de libertad adicional por cada parámetro estimado. Generalmente, cuando se rechaza la hipótesis nula de ajuste a la distribución teórica especificada, se dispone de suficiente información como para precisar en qué sentido es el rechazo, como veremos en algunos ejemplos.

Ejemplo 12.1.- El ejemplo típico de esta situación considera un dado, del que se han obtenido 300 tiradas, y cuya perfección, entendida como ausencia de sesgos en los resultados, se pretende contrastar. La partición del espacio de sucesos es clara en este caso, pues consiste en los 6 resultados distintos que pueden observarse. La hipótesis nula H_0 recoge en tal caso una distribución de probabilidad uniforme. Es decir, la hipótesis nula es que el dado está, efectivamente, correctamente construido, y, al no tener sesgos, la probabilidad teórica con que deberíamos observar cada uno de los sucesos posibles: 1, 2, 3, 4, 5 y 6, es de 1/6, lo que aparece en la segunda línea del Cuadro 12.1:

	Cuadro 12.1 Resultado						
	1	2	3	4	5	6	Total
Observada (O_i)	44	62	52	45	50	47	300
Teórica (E_i)	50	50	50	50	50	50	300

mientras que la primera línea presenta las frecuencias que suponemos que se observaron en el experimento. Ello genera un valor del estadístico:

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - T_i)^2}{T_i} = \frac{(44-50)^2}{50} + \frac{(62-50)^2}{50} + \frac{(52-50)^2}{50} + \frac{(45-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(47-50)^2}{50} =$$

$$= \frac{36+144+4+25+0+9}{50} = \frac{218}{50} = 4{,}36$$

que debemos comparar con una chi-cuadrado con 6-1 = 5 grados de libertad.

Esta es una *tabla de clasificación simple* en la que la lista de frecuencias observadas de una característica se compara con la de una distribución teórica. A lo largo de este capítulo utilizaremos tablas de clasificación, simples y múltiples, es decir, de una o varias características, sobre las que definiremos contrastes del tipo chi-cuadrado, como el que acabamos de ver. En ellos, el número de grados de libertad en tablas como la anterior es igual a: (m-1)(n-1), siendo m y n el número de columnas y de filas. En nuestro caso: (m-1)(n-1) = (6-1)(2-1) = (5)(1) = 5, que es el valor de k-1 en este caso. El valor crítico para la distribución chi-cuadrado con 5 grados de libertad a niveles de significación del 0,05 y 0,01 es, respectivamente: 11,1 y 15,1. Nuestro estadístico es inferior a ambos valores, por lo que *no rechazamos* la hipótesis nula de que el dado es correcto, teniendo todos los resultados posibles, del 1 al 6, igual probabilidad.

Ejemplo 12.2.- Supongamos ahora que, en un examen de preguntas con múltiple elección, cada una con 4 opciones, un alumno contesta correctamente 40 de entre las 100 preguntas que se le formularon. ¿Cómo podemos contrastar la hipótesis nula de que el alumno conoce, efectivamente, la materia de la que se ha examinado? Para ello, construimos el cuadro:

	Cuadro 12.2 *Respuestas*		*Total*
	Correctas	*Incorrectas*	
Frecuencias observadas:	40	60	100

Frecuencias teóricas:	25	75	100

La hipótesis nula consiste en creer que el alumno contesta sin conocer la respuesta. Si así fuese, contestando al azar y puesto que hay cuatro opciones en cada pregunta, el alumno habría contestado correctamente una de cada cuatro preguntas, es decir, 25 de las 100 cuestiones formuladas. Por ello tenemos las frecuencias teóricas de la última columna de la tabla.

El estadístico chi-cuadrado toma el valor:

$$\chi^2 = \sum_{i=1}^{2} \frac{(O_i - T_i)^2}{T_i} = \sum_{i=1}^{2} \frac{(n_i - np_i)^2}{np_i} = \frac{(40-25)^2}{25} + \frac{(60-75)^2}{75} = \frac{225}{25} + \frac{225}{75} = 9 + 3 = 12$$

que excede sobradamente de los valores críticos de la distribución chi-cuadrado con 1 grado de libertad a los niveles de significación del 0,05 y 0,01, que son: 3,84 y 6,63. En consecuencia, rechazamos la hipótesis nula, habiendo encontrado evidencia en el sentido de que el alumno conoce la materia mejor que para simplemente creer que ha contestado las cuestiones al azar.

Ejemplo 12.3.- Un partido político manifiesta que el 75% de la población es favorables a sus propuestas políticas. Para contrastar tal hipótesis frente a la alternativa de que el porcentaje de simpatizantes es inferior, se efectúa una encuesta a 1.000 personas, a las que se les pregunta si son o no favorables a dicho partido político, obteniendo el resultado de que 680 lo son, mientras que las 320 restantes se manifiestan como no favorables al partido político. Ello conduce al cuadro:

Frecuencias	*Cuadro 12.3 Opinión del encuestado*		*Total*
	Favorable	*No favorable*	
Observadas	680	320	1.000
Teóricas	750	250	1.000

en la que aparecen las frecuencias teóricas que surgen de la opinión emitida por dicho partido político. En consecuencia, tenemos el valor del estadístico:

$$\chi^2 = \sum_{i=1}^{2} \frac{(O_i - T_i)^2}{T_i} = \sum_{i=1}^{2} \frac{(n_i - np_i)^2}{np_i} = \frac{(680-750)^2}{750} + \frac{(320-250)^2}{250} = 26{,}13$$

que excede con mucho del valor crítico de las tablas correspondientes a la distribución chi-cuadrado con 1 grado de libertad. Consecuentemente, rechazamos la hipótesis nula y, con ello, la opinión emitida por el partido político en consideración. Como el número de simpatizantes observado, 680, es inferior al que correspondería a la proporción anunciada, 0,75, al rechazar la hipótesis nula pasamos a creer que la verdadera proporción es inferior a ésta.

El mismo contraste puede resolverse utilizando la aproximación Normal a la distribución binomial. Recordemos que, para valores grandes de *n*, la distribución binomial *B(n,p)* puede aproximarse por una *N(np, np(1-p))*, por lo que la proporción, que no es sino el cociente de la binomial *B(n,p)* por *n*, se aproxima por una distribución *N(p, p(1-p)/n)*. Así, bajo la hipótesis nula de que la opinión del partido político es correcta, la proporción *muestral* sigue, aproximadamente, una distribución $N(0{,}75\ ;\sigma^2)$ donde la desviación típica es:

$$\sigma = \sqrt{\frac{(0{,}75)(0{,}25)}{1000}} = 0{,}0137$$

Por tanto, el intervalo de confianza del 0,95 de una cola comprende valores superiores a: (0,75)-(1,65)(0,0137) = 0,727. Por tanto, como la proporción muestral, 0,68, cae fuera del intervalo del 95%, constituye significativa evidencia en contra de la hipótesis nula, que postulaba la veracidad de la afirmación del partido político, y que habremos de rechazar, en favor de que la proporción de simpatizantes es inferior al 75%.

2.1.2 Contraste de Kolmogorov-Smirnov.

El contraste de Kolmogorov-Smirnov sigue una idea similar al anterior, pero en vez de comparar las probabilidades de diversos sucesos, compara los valores de las funciones de distribución, tanto en la muestra, como la que teóricamente se derivaría de la población que se ha explicitado en la hipótesis nula. El estadístico de Kolmogorov-Smirnov consiste en la máxima distancia observada entre ambas funciones de distribución:

$$D_n = \underset{x}{Supremo} \mid F_n(x) - F(x) \mid$$

donde $F_n(x)$ denota la función de distribución muestral, y hay que comparar su valor con unas tablas específicas de este estadístico, pues no sigue ninguna distribución conocida. Para tamaños muestrales, n, superiores a 100, el valor crítico puede obtenerse mediante: $\sqrt{-\ln(\alpha/2)/2n}$, siendo 1-α el nivel de confianza. Lógicamente, se rechaza la hipótesis nula si el estadístico toma un valor superior al de las tablas. Bajo H_0, la muestra fue extraída de la población considerada en dicha hipótesis nula, por lo que las funciones de distribución muestral y teórica serían tan similares que, incluso tomando su máxima distancia, ésta sería *suficientemente reducida*. Cuando el valor numérico del estadístico excede del valor crítico de las tablas, se considera que no es suficientemente reducido, constituyendo evidencia en el sentido de que las funciones de distribución difieren una de otra y, por ello, la hipótesis nula debe rechazarse.

La distribución del estadístico de Kolmogorov-Smirnov es independiente del tipo de distribución de la que fue extraída la muestra, lo cual es interesante, pues nos permite utilizar una única tabla de valores críticos para este estadístico; de lo contrario, deberíamos tener una tabla para cada tipo de distribución de probabilidad F incluida en H_0. Este contraste puede utilizarse asimismo con distribuciones de tipo discreto, pero entonces sólo podemos decir que α es el *máximo* nivel de significación del contraste que hayamos diseñado. Para aplicar el contraste con distribuciones continuas es preciso agrupar sus valores en clases o intervalos, con lo que se pierde cierta información. Al utilizar únicamente la información muestral incorporada en la máxima distancia entre las funciones de distribución, este estadístico ignora mucha información muestral, a diferencia del contraste basado en el estadístico χ^2 de Pearson.

El estadístico de Kolmogorov-Smirnov puede utilizarse para construir bandas de confianza para una distribución teórica desconocida $F(x)$, a partir de una distribución empírica. Para ello, fijado un valor de α, tomamos de la tabla el valor crítico D_n correspondiente a α y n, el tamaño muestral del que se dispone. El extremo superior $F_s(x)$ de la banda para $F(x)$ se construye sumando D_n a la función de distribución empírica, hasta que se alcanza el nivel 1, permaneciendo entonces en éste. El nivel inferior $F_i(x)$ es igual a cero hasta que la distribución empírica llega a ser igual o mayor a D_n. A partir de entonces, $F_i(x) = F(x) - D_n$.

Ejemplo 12.4.- Volviendo al ejemplo del lanzamiento del dado, podemos construir el cuadro de ambas funciones de distribución:

	Cuadro 12.4 *Resultado*					
	1	2	3	4	5	6
Función de distribución muestral	0,145	0,353	0,527	0,677	0,843	1,0
Función de distribución teórica	0,167	0,333	0,500	0,667	0,833	1,0
Diferencias en valor absoluto	0,022	0,020	0,027	0,010	0,010	0

con una máxima diferencia de 0,027. En nuestro caso, con $n = 300$, tenemos, para $\alpha = 0,05$, un valor crítico de 0,035. Como el valor del estadístico de Kolmogorov-Smirnov que hemos obtenido es inferior a este nivel crítico, no rechazamos la hipótesis nula de que la muestra procede de la población teórica, que asignaba probabilidad 1/6 a cada resultado posible.

2.2 Contrastes basados en estadísticos de posición.

Otro enfoque para contrastar el grado de ajuste de la distribución muestral a una distribución teórica se basa en el uso de medidas de posición, como la mediana, los cuartiles, etc.. La idea es que si la muestra procede de la distribución hipotética que se considera, entonces sus medianas no deberían ser muy diferentes, pero tampoco sus cuartiles, etc.. Por tanto, estos contrastes se basan en la distancia entre los estadísticos de posición de la muestra y los teóricos. Distintos contrastes utilizan distintos estadísticos de posición.

2.2.1 Contraste de los signos

Este contraste considera un percentil q, $0 < q < 99$ en la población que se considera bajo la hipótesis nula. Denotemos por θ el valor numérico del percentil q en la distribución teórica. A continuación, definimos para cada elemento muestral una función indicatriz X_i que toma el valor 1 si el dato está por debajo de θ, y 0 si la observación muestral está por encima de θ. Este es un fenómeno Bernouilli, con probabilidad $p = \frac{q}{100}$ puesto que, por definición de *percentil*, bajo la hipótesis nula, exactamente $q\%$ de los elementos poblacionales están por debajo de θ, y $(100-q)\%$ está por encima de θ. Si consideramos ahora la suma de tales funciones indicatriz:

$$X = \sum_{i=1}^{n} X_i$$

esta variable seguirá una distribución binomial $B(n,p)$. Se recoge evidencia en contra de la hipótesis nula cuando el porcentaje de elementos muestrales por debajo de θ, el teórico percentil de orden q, es muy superior o muy inferior a $q\%$. Deberemos hacer, por tanto, un contraste de dos colas, y determinar los valores críticos λ_i y λ_s, tales que:

$$P[X \geq \lambda_s] = \sum_{x=\lambda_s}^{n} \binom{n}{x} \left(\frac{q}{100}\right)^x \left(1 - \frac{q}{100}\right)^{n-x} = \frac{\alpha}{2}$$

$$P[X \leq \lambda_i] = \sum_{x=0}^{\lambda_i} \binom{n}{x} \left(\frac{q}{100}\right)^x \left(1 - \frac{q}{100}\right)^{n-x} = \frac{\alpha}{2}$$

El contraste se denomina de los *signos* porque considera el número de signos positivos y negativos de las diferencias x_i-θ, siendo x_i cada una de las observaciones muestrales.

Cuando el tamaño muestral es suficientemente grande, podemos utilizar la aproximación Normal a la distribución binomial, con lo que el cálculo de estos valores críticos se simplifica enormemente.

Ejemplo 12.5.- Si tomamos el *primer cuartil* como criterio de ajuste de la distribución muestral a una distribución teórica, la probabilidad de que un elemento muestral esté por debajo de él es 0,25, siendo 0,75 la probabilidad de que sea superior. Supongamos que disponemos de una muestra de tamaño $n = 20$. El número de observaciones muestrales por debajo del primer cuartil sigue, por tanto, una distribución *B(20;0,25)*. De acuerdo con las tablas, esta variable toma un valor igual o inferior a 1 con probabilidad 0,0243, y un valor igual o inferior a 9 con probabilidad 0,9861. En consecuencia, tenemos: $P(2 \leq X \leq 9) = 0{,}9618$, próxima a un nivel de confianza del 95%. Supongamos que θ es el primer cuartil bajo la hipótesis nula; deberíamos esperar tener en la muestra un número entre 2 y 9 de datos inferiores a θ. Uno ó 2, o bien 10 ó más datos inferiores a θ constituirían evidencia contra H_0.

Alternativamente, si tuviéramos como hipótesis alternativa que el primer cuartil es superior a θ, utilizaríamos el hecho de que una variable *B(20;0,25)* toma un valor igual o inferior a 8 con probabilidad 0,9591. Por tanto, 9 o más observaciones muestrales por debajo de θ sería un número excesivo bajo H_0, lo que nos llevaría a rechazar dicha hipótesis en favor de la alternativa.

2.2.2 El contraste de la mediana

Un caso particular del anterior, de especial interés, consiste en el uso de la mediana como estadístico de ajuste. Una vez calculada la mediana *m* de la distribución teórica, entonces la mitad de los elementos muestrales deberían estar por encima de *m* y la mitad restante por debajo de *m*, si la muestra procede realmente de la población teórica. Evidentemente, no se espera que esto sea exactamente cierto en todas las muestras, pero si la división que *m* introduce en la muestra deja porcentajes de elementos por encima y por debajo muy diferentes de 50%, entonces tendremos evidencia significativa en contra de la población teórica que hayamos especificado bajo H_0, que deberemos rechazar.

Ahora habrá que determinar un valor crítico λ tal que:

$$P[X \geq \lambda] = \sum_{X=\lambda}^{n} \binom{n}{X} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$$

y puede utilizarse nuevamente la aproximación Normal si el tamaño muestral es suficientemente grande. El otro valor crítico que delimita el intervalo de confianza del 100(1-α)% es n-λ

2.2.3. Contraste de Wilcoxon de rangos con signos

Los contrastes de signos anteriores consideran si las observaciones muestrales están por encima o por debajo de un determinado estadístico de referencia, pero no tienen en cuenta la magnitud de sus distancias a dicho estadístico. El contraste de Wilcoxon salva tal limitación, considerando los valores numéricos de las diferencias x_i - m, donde m es la mediana de la distribución teórica. Una vez calculadas, se ordenan crecientemente los valores absolutos de dichas diferencias, y se asigna a cada una de ellas su *rango*, siendo éste el orden que ocupa cada diferencia en dicha secuencia creciente.

El estadístico T^+ de Wilcoxon consiste en la suma de los rangos que corresponden a las diferencias que eran positivas, antes de tomar valores absolutos, mientras que el estadístico T^- consiste en la suma de los rangos que corresponden a las diferencias que eran negativas. Un valor elevado de T^+ indicará que las mayores diferencias tienen signo positivo, de modo que la población no parece ser simétrica alrededor de m, estando desplazada hacia la derecha. Un valor reducido indica que las mayores diferencias se producen con datos por debajo de m, sugiriendo que está desplazada hacia la derecha. Por último, el estadístico T de Wilcoxon consiste en la suma de todos los rangos, habiendo asignado esta vez a cada uno de ellos el signo de la diferencia x_i - m. Dicho de otra forma: $T = T^+ - T^-$.

Los rangos asignados de este modo a una muestra de tamaño n constituyen una secuencia formada por los primeros n números naturales. Bajo la hipótesis nula de simetría alrededor de m, deberíamos esperar que la mitad de los rangos proviniese de diferencias de signo positivo, y la otra mitad, de diferencias de signo negativo. Además, cada uno de ellos debería ser positivo o negativo con probabilidad 1/2. Por tanto, el rango k-ésimo toma valores k y -k con probabilidad 1/2 en cada caso, por lo que tiene esperanza cero, y lo mismo ocurrirá con el estadístico T. En consecuencia, la varianza de T coincide con la suma de los cuadrados de los n primeros números naturales, es decir:

$$Var(T) = \frac{n(n+1)(2n+1)}{6}$$

Una primera forma de utilización del contraste de Wilcoxon consiste en utilizar el estadístico T, con la aproximación Normal a su distribución, con esperanza cero y la varianza que acabamos de ver. Una segunda forma se basa en el uso del estadístico T^+. Como este estadístico utiliza sólo los rangos procedentes de diferencias x_i- m positivas, su suma debería ser igual, en promedio, a la mitad de la suma de los n primeros números naturales, de modo que:

$$E(T^+) = \frac{n(n+1)}{4}$$

mientras que su varianza es la cuarta parte de la del estadístico T:

$$Var(T^+) = \frac{n(n+1)(2n+1)}{24}$$

Existe una tabla de valores críticos del estadístico T de rangos de Wilcoxon con la cual puede compararse el estadístico si no se quiere efectuar la aproximación Normal, aunque ésta es muy conveniente. Si el estadístico muestral de Wilcoxon está por debajo del valor crítico de las tablas, no podremos mantener la hipótesis nula de simetría alrededor de m.

Ejemplo 12.6.- Consideremos los resultados que en un examen de inglés ha obtenido un grupo de 20 alumnos. Queremos contrastar que las calificaciones proceden de una población simétrica alrededor de una mediana igual a 70.

Cuadro 12.8
Calificaciones en el examen de inglés

Calificaciones	75 62 80 71 67 69 73 83 71 74 73 79 65 73 64 67 76 72 68 74
Diferencias x_i - m	5 -8 10 1 -3 -1 3 13 1 4 3 9 -5 3 -6 -3 6 2 -2 4
Rangos	13,5 -17 19 2 -8 -2 8 20 2 11,5 8 18 -13,5 8 -15,5 -8 15,5 4,5 -4,5 11,5

Al asignar rangos, cuando una determinada magnitud para la diferencia x_i - m aparece repetida, se asigna a cada una de ellas el promedio de los rangos que corresponderían a dichas observaciones. Así, la menor diferencia observada es de magnitud 1, que aparece en tres ocasiones, a las que corresponderían rangos 1, 2 y 3. Su promedio es 2, que asignamos a cada una de las tres observaciones. La siguiente magnitud, que es 2, aparece en dos ocasiones, a las que corresponderían rangos 4 y 5, cuyo promedio, 4,5, es el rango que asignamos a ambas observaciones.

El estadístico T^+ es igual a 141,5, mientras que T^- = 68,5, por lo que $T = T^+ - T^-$ = 141,5 - 68,5 = 73, con esperanza nula, y varianza: (20)(21)(41)/6 = 2870. Por tanto, el ratio: $[T-E(T)]/\sqrt{Var(T)}$ que se distribuye N(0,1) en muestras grandes, toma el valor: 73/(53,57) = 1,363, que está dentro del intervalo de confianza del 95% de dicha distribución. Por tanto, en la medida en que dicha aproximación sea correcta, no rechazamos la hipótesis nula de que la distribución es simétrica alrededor de su mediana.

Si utilizamos el estadístico T^+ tipificado: $[T^+ - E(T^+)]/\sqrt{Var(T^+)}$, que tiene como distribución aproximada $N(0,1)$ en muestras grandes, toma en esta muestra un valor: [141,5 - 105]/(26,79) =
-1,362 que está nuevamente dentro de la banda de confianza del 95% para la $N(0,1)$. El lector puede comprobar que el estadístico que podría obtener con T^- conduce al mismo resultado numérico.

2.4 Contraste de Normalidad

Un caso especial de ajuste a una distribución teórica se produce cuando la hipótesis nula especifica que dicha distribución es Normal. Para efectuar el contraste, se calculan la media, así como la varianza muestral, S_x^2, y se ordena la muestra en sentido creciente. A continuación, se toman las diferencias entre mayor y menor elementos, segundo mayor y menor, terceros elementos menor y mayor, etc.. Cada una de estas diferencias viene corregida por un coeficiente que se toma de una tabla construida al efecto por Shapiro y Wilk, y se agregan los resultados en una suma D. El estadístico de Shapiro-Wilk es el cociente:

$$W = \frac{D^2}{n S_x^2}$$

y el valor crítico se toma de una segunda tabla propuesta por estos autores. El valor crítico que proporciona esta tabla para un nivel de significación α es el mínimo que puede tomar el estadístico *W* si la población de la que se extrajo la muestra es Normal.

Ejemplo 12.XX.- Las calificaciones del examen de inglés, ordenadas, son:

62 < 64 < 65 < 67 < 67 < 68 < 69 < 71 < 71 < 72 < 73 < 73 < 73 < 74 <
74 < 75 < 76 < 79 < 80 < 83

y el estadístico es:

$$D = (0{,}473)(83-62) + (0{,}321)(80-64) + (0{,}257)(79-65) + (0{,}209)(76-67) + (0{,}169)(75-67) +$$

$$(0{,}133)(74-68) + (0{,}101)(74-69) + (0{,}071)(73-(71) + (0{,}042)(73-71) + (0{,}014)(73-72) = 23{,}44$$

que, genera un estadístico:

$$W = \frac{(23{,}44)^2}{20(27{,}96)} = 0{,}983$$

que, casualmente, coincide con el valor crítico de las tablas al 95% de confianza, y es inferior al valor crítico al 99%, que es de 0,988. El valor crítico al 90% es 0,979, por lo que no hay mucha evidencia ni a favor ni en contra de la Normalidad en este caso.

Otro contraste de Normalidad, muy popular en el trabajo econométrico, utiliza el hecho, ya mencionado en el Capítulo XX de que toda población Normal tiene coeficientes de asimetría y de curtosis igual a cero. C.M.Jarque y A.K.Bera probaron que, si la distribución de probabilidad es Normal, entonces el estadístico:

$$JB = n\left[\frac{AS^2}{6} + \frac{(K-3)^2}{24}\right]$$

se distribuye como una chi-cuadrado con 2 grados de libertad.

Para los mismos datos del ejemplo anterior, la hoja de cálculo NOPARAM.WQ1 contiene los cálculos, que conducen a un valor numérico de 0,2434 para este estadístico, muy inferior a los valores críticos, a los niveles de significación habituales, de la distribución χ_2^2, por lo que no rechazamos la hipótesis nula de Normalidad.

3. CONTRASTES DE HOMOGENEIDAD ENTRE DISTRIBUCIONES

Los contrastes de homogeneidad de entre distribuciones tratan de discernir si dos muestras independientes pueden haber sido extraídas de la misma distribución de probabilidad poblacional. La hipótesis nula es que las distribuciones de probabilidad de las que se extrajeron ambas muestras son idénticas aun siendo desconocidas, siendo la hipótesis alternativa que ambas distribuciones son diferentes.

3.1 Contrastes de homogeneidad en muestras bidimensionales apareadas

Una clase importante de problemas cubre los casos en que tenemos, bien observaciones de varias características de un conjunto de individuos, o bien mediciones de una misma característica en distintos instantes de tiempo. En ambos casos tenemos varias observaciones para cada individuo en la muestra, y queremos contrastar la homogeneidad de las distribuciones de cada una de dichas observaciones. Estas son *muestras pareadas*, en las que tendremos el mismo número de observaciones de cada una de las características, a diferencia de los casos que analizaremos posteriormente.

Los contrastes de rangos que vamos a examinar se basan en la idea de que, si las dos muestras proceden de igual población, entonces no debería haber mucha diferencia entre los rangos de sus elementos, una vez que estos se hubieran ordenado creciente o decrecientemente. Todos estos contrastes de rangos producen los mismos resultados con independencia de que los elementos muestrales se ordenen creciente o decrecientemente. Los *contrastes denominados de signo* se basan únicamente en la diferencia entre los elementos ordenados de ambas muestras, mientras que los *contrastes de rangos* más generales utilizan asimismo información acerca de la magnitud de las distancias que separan los elementos que ocupan igual rango en las dos muestras ordenadas. Son similares a los que ya analizamos en la Sección XX para el caso de ajuste a una distribución teórica.

3.1.1 Contraste del signo

El contraste del signo se utiliza para contrastar la homogeneidad de dos *poblaciones continuas*, utilizando como información muestras de igual tamaño, seleccionadas de cada una de las poblaciones. El contraste se basa en ordenar de acuerdo con algún criterio las observaciones en cada muestra, y comparar cada par de ellas, substituyendo los dos números por un signo más, un signo menos o un cero, según que el elemento de la primera muestra sea superior, inferior o igual al elemento correspondiente de la segunda muestra.

***Ejemplo 12.7*.-** Supongamos que el grupo anterior de alumnos realiza un segundo examen de inglés a la vuelta de una estancia en el Reino Unido. Este examen es de idéntica dificultad al primero, y se pretende con él analizar la efectividad de la estancia en el extranjero en cuanto al aprendizaje del inglés. Los resultados en ambos exámenes fueron:

Cuadro 12.6

Alumno	*Calificación antes de la estancia*	*Calificación después de la estancia*		
1	75	78	3	10,0
2	62	66	4	14,5
3	80	78	-2	-5,5
4	71	71	0	

5	67	69	2	5,5
6	69	71	2	5,5
7	73	76	3	10,0
8	83	82	-1	-2,0
9	71	75	4	14,5
10	74	72	-2	-5,5
11	73	74	1	2,0
12	79	71	-8	-18,0
13	65	68	3	10,0
14	73	76	3	10,0
15	64	67	3	10,0
16	67	71	4	14,5
17	76	72	-4	-14,5
18	72	72	0	
19	68	73	5	17,0
20	74	75	1	2,0

La hipótesis nula es: H_0: *la estancia no ayudó a mejorar el nivel de inglés de los alumnos*. En esta aplicación no tiene sentido pensar que la estancia en el Reino Unido fue perjudicial para el nivel de inglés de los alumnos, por lo que el contraste será de una sola cola. La hipótesis alternativa es: H_1: *la estancia contribuyó a mejorar el nivel de inglés de los alumnos*. En un contraste de dos colas la hipótesis alternativa consideraría tanto la posibilidad de que la estancia haya, efectivamente, contribuido a mejorar el nivel de inglés de los alumnos, como que haya contribuido a deteriorarlo.

Podemos observar que la calificación del examen realizado al regreso de la estancia es, en 13 alumnos, superior a la calificación obtenida antes de la estancia, mientras que para dos alumnos fue idéntica. Si la hipótesis nula fuese cierta, esperaríamos que, a la vuelta de la estancia en el Reino Unido, la probabilidad de obtener un mejor resultado en la prueba de inglés fuese igual a la probabilidad de obtener un resultado inferior, ambas igual a 1/2. Estaríamos entonces ante una variable aleatoria binomial para cada alumno, que se materializa en dos sucesos: calificación superior, calificación inferior, con probabilidades: $p = 1\text{-}p = 0{,}5$.

Eliminando de la muestra los 2 alumnos cuya calificación no varía, tenemos un total de 18 alumnos, con 13 signos positivos. Bajo H_0, estaríamos ante una variable binomial: $B(18; 0{,}5)$, con esperanza matemática: $E(X) = np = (18)(0{,}5) = 9$ y varianza igual a: $np(1\text{-}p) = 18(0{,}5)^2 = 4{,}5$, por lo que la desviación típica sería 2,12 .

Utilizando *la aproximación Normal a la distribución binomial*, podemos calcular (aproximadamente) la probabilidad de obtener 13 o más signos positivos:

$$P(X \geq 12{,}5) = P\left(\frac{X-\mu}{\sigma} \geq \frac{12{,}5-\mu}{\sigma}\right) = P\left(Z \geq \frac{12{,}5-9}{2{,}12}\right) = P(Z \geq 1{,}651)$$

donde hemos incorporado la habitual corrección por ser la distribución que se aproxima de tipo discreto. Generalmente se utiliza esta aproximación, porque el cálculo de las probabilidades binomiales es sumamente complejo.

Trabajando al nivel de significación $\alpha = 0{,}05$, en un contraste de una cola, tendríamos un nivel crítico para la variable *N(0,1)* de 1,645. Es decir, bajo la hipótesis nula, estaríamos en presencia (aproximadamente) de una variable $N(9; 2{,}12)$, y con probabilidad 0,95, el resultado de la prueba estandarizada debería ser inferior a 1,645. Nuestro resultado es superior, si bien ligeramente, por lo que el supuesto contenido en H_0, que nos ha servido para obtener la distribución Normal no puede mantenerse. Rechazaríamos la hipótesis nula y pasaríamos a aceptar la hipótesis alternativa de que la estancia en UK ha servido para mejorar el nivel de inglés de los alumnos en el grupo de estudio.

El nivel crítico para el contraste de una cola al nivel de confianza del 0,99 (nivel de significación de 0,01) es de 2,33, en cuyo caso no rechazaríamos la hipótesis nula y mantendríamos la hipótesis de que la estancia en *UK* no ha servido para mejorar el nivel de inglés de los alumnos en el grupo de estudio, mientras que al 90%, el rechazo es bastante claro.

En casos en que el contraste a un determinado nivel de significación tiene una resolución tan marginal como aquí ha ocurrido trabajando al 5%, el investigador debe decidir si se decanta por aumentar o disminuir dicho nivel, y resolver el contraste al nuevo nivel escogido. En este ejemplo podríamos habernos desplazado al 0,10 o al 0,01, pero ha parecido más interesantes la segunda opción, que nos ha llevado a no rechazar la hipótesis nula, muy posiblemente porque quisiéramos estar seguros de que las diferencias realmente existen si es que creemos detectarlas.

Este contraste tiene la misma limitación que el similar que ya examinamos en la Sección 12.2.2, de tener en cuenta únicamente los *signos* de las diferencias, pero no su magnitud, a diferencia del contraste de Wilcoxon que presentamos a continuación. A diferencia de éste, sin embargo, el contraste del signo tiene la ventaja de poder ser utilizado con distribuciones cualitativas.

3.1.2. Contraste de Wilcoxon

El contraste de Wilcoxon para la homogeneidad de distribuciones es muy similar al contraste de igual nombre que ya vimos para el ajuste de una distribución muestral a una distribución teórica.

Ejemplo 12.8.- Si tomamos nuevamente el ejemplo de los alumnos de inglés, el Cuadro 2 presenta todos los cálculos necesarios para obtener el estadístico T de Wilcoxon: en la cuarta columna aparece no sólo el signo de cada diferencia, sino también su magnitud. Al igual que en el contraste de los signos, ignoramos los casos en que las observaciones numéricas en ambas muestras coinciden. La quinta columna asigna *rangos* a tales diferencias, asignando a las observaciones con igual rango, el promedio de estos, de igual modo que hicimos en el contraste descrito en la Sección 2.2.3. De acuerdo con la hipótesis nula, las diferencias provienen de una distribución simétrica con mediana cero.

El contraste de Wilcoxon se basa en que, si las dos muestras fuesen homogéneas, entonces las diferencias entre cada dos observaciones correspondientes no sólo se distribuirían uniformemente entre valores positivos y negativos, sino que también sus magnitudes se distribuirían simétricamente: habría, en promedio, tantas diferencias pequeñas de signo positivo como de signo negativo, y tantas diferencias grandes de un signo como de otro.

En nuestro ejemplo, los rangos positivos suman 125,5, mientras que los rangos negativos suman 45,5, lo cual parece, a simple vista, bastante heterogéneo. Si el valor del estadístico T de Wilcoxon es suficientemente pequeño, es decir inferior al nivel crítico de las tablas, es que esta suma menor es *significativamente* inferior a la otra suma mayor, rompiendo la simetría que debe cumplirse bajo la hipótesis nula. En consecuencia, rechazamos ésta.

Ya hemos argumentado que nuestro ejemplo se presta a un contraste de una sola cola. La idea es que la estancia en el Reino Unido puede haber aumentado el nivel de inglés de los alumnos, por lo que la mediana sería ahora superior a la de antes. Por tanto, si el valor del estadístico excede del valor crítico de las tablas, tendremos evidencia significativa en contra de la hipótesis de que ambas medianas son iguales. Los valores de la tabla de Wilcoxon para T^{-} [al final de este texto] para n = 18, y niveles de significación de 0,025 y 0,01 son 40 y 33, respectivamente, que son superados por el valor de nuestro estadístico, 45,5. En consecuencia, no rechazamos la hipótesis nula de *homogeneidad de resultados* en los exámenes de inglés de antes y después de la estancia en *UK*, manteniendo por tanto que dicha estancia *no contribuyó* a mejorar *significativamente* el nivel de inglés del grupo de alumnos.

A partir de n = 25, es mejor utilizar el hecho de que la suma de los rangos positivos, $T+$, se distribuye aproximadamente Normal, con: $E(T) = \frac{n(n+1)}{4}$; $Var(T) = \frac{n(n+1)(2n+1)}{24}$; . En nuestro ejemplo, con n = 18, aun no siendo la situación más recomendable para la aproximación, tendríamos una distribución Normal con esperanza matemática 85,5 y desviación típica igual a 22,96, de modo que el valor de nuestro estadístico, estandarizado, sería:

$$Z = \frac{T - E(T)}{\sqrt{Var(T)}} = \frac{45{,}5 - 85{,}5}{22{,}96} = -1{,}742$$

que conduciría a rechazar la hipótesis nula al 99% de confianza, aunque no al 99%, ya que los valores críticos de la *N(0,1)* para contrastes de una cola, a niveles de significación de 0,05 y 0,01 son, respectivamente, 1,645 y 2,33. A pesar de no ser totalmente apropiada en nuestro caso la aproximación Normal conduce, al 95% de confianza, a los mismos resultados que obtuvimos sin dicha aproximación. El resultado numérico del contraste es el mismo, sólo que cambiado de signo, si en vez de T^{+}, se utiliza T^{-}.

3.2 Contrastes de homogeneidad generales

Examinamos en esta sección contrastes de homogeneidad entre muestras cualesquiera, no necesariamente procedentes de un mismo conjunto de individuos, a diferencia de los casos comprendidos en la sección anterior. Por tanto, vamos a comparar muestras de tamaños generalmente diferentes entre sí.

3.2.1 Contraste de Kolmogorov-Smirnov para dos muestras

Este contraste puede utilizarse para la hipótesis nula de que las distribuciones continuas de las que han sido extraídas dos muestras aleatorias simples de tamaños n_1 y n_2, son iguales. Para llevarlo a cabo se divide el espacio muestral en k intervalos (o sucesos) disjuntos, se calculan ambas funciones de distribución empíricas, $F_{n_1}(x)$ y $F_{n_2}(x)$, y se toma el estadístico:

$$D_{n_1,n_2} = \underset{x}{supremo} \mid F_{n_1}(x) - F_{n_2}(x) \mid$$

Si las dos muestras proceden de la misma población, sus funciones de distribución empíricas no pueden ser muy distintas, por lo que el contraste es *siempre de una cola*, y se rechaza la hipótesis nula de igual distribución si el estadístico toma un valor *suficientemente grande*. Para ello, se utilizan las tablas de este estadístico, para encontrar el umbral λ tal que: $P(D_{n_1,n_2} \geq \lambda) = \alpha$. Dicho valor crítico puede aproximarse, cuando ambas muestras son grandes, por:

$$\lambda = k \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

siendo k = 1,22; 1,36; 1,63 a niveles de confianza del 90%, 95% y 99%, para el contraste de dos colas, y de k = 1,07; 1,22 y 1,52 para contrastes de una sola cola.

El contraste de Kolmogorov-Smirnov tiene más potencia para el caso en que las distribuciones tienen diferentes medianas que para el caso en que, teniendo la misma posición central, difieren en su dispersión.

Ejemplo.- Supongamos que disponemos de dos muestras, ambas sobre el intervalo [0,1]:

Muestra 1: 0,65 0,31 0,42 0,81 0,12 0,91 0,72 0,41 0,94 0,61 0,52 0,21 0,16 0,74 0,65 0,83 0,44 0,42 0,56 0,73

Muestra 2: 0,15 0,94 0,81 0,72 0,21 0,35 0,18 0,74 0,73 0,62 0,85 0,91 0,18 0,23 0,65

a partir de las cuales obtenemos las siguientes distribuciones empíricas sobre intervalos de longitud 0,20:

Funciones de distribución empíricas					
$F_n(x)$	0≤x<0,20	0,20≤x<0,40	0,40≤x<0,60	0,60≤x<0,80	0,80≤x<1,00
Muestra 1: $n_1 = 20$	0,10	0,20	0,50	0,80	1,0
Muestra 2: $n_2 = 15$	0,20	0,40	0,40	0,73	1,0

La máxima diferencia entre ambas funciones de distribución se produce en el segundo intervalo, y es de: 0,20 - 0,40 = -0,20. El valor crítico de las tablas del estadístico de Kolmogorov-Smirnov para estos tamaños muestrales es de 13/30 = 0,433, por lo que *no rechazamos* la hipótesis nula de igualdad de distribuciones de probabilidad. La aproximación: $\sqrt{(n_1 + n_2)/n_1 n_2}$ es, en este caso, igual a 0,464, no muy diferente del valor de las tablas. Tampoco en esta comparación rechazaríamos la hipótesis de igualdad de distribuciones.

3.2.2 El contraste de Mann-Whitney de sumas de rangos

Este estadístico, introducido simultáneamente por Mann y Whitney en 1947, se utiliza para *contrastar si dos muestras, extraídas independientemente, proceden de la misma población.* El único supuesto preciso es que la población o poblaciones de que se han extraído las muestras, sean de tipo continuo, pero no requiere simetría. *La hipótesis nula H_0 es que las esperanzas matemáticas de ambas poblaciones son iguales,* mientras que la alternativa puede establecer que las esperanzas matemáticas son diferentes (contraste de dos colas), o que una de ellas, previamente escogida, es superior a la otra (contraste de una cola). Este contraste resume, por tanto, las características de la distribución de probabilidad en su esperanza matemática.

El planteamiento lógico del contraste es, por tanto, igual al del contraste de igualdad de esperanzas que ya resolvimos mediante un estadístico t de Student cuando suponíamos que ambas poblaciones eran Normales. El estadístico U se construye, al igual que los anteriores, a partir de las sumas de rangos. Supongamos que disponemos de dos muestras de tamaño n_1 y n_2, tomadas de dos poblaciones potencialmente diferentes. Comenzamos considerando ambas como una muestra global de tamaño n_1+n_2, y asignamos rangos a todos estos elementos, resolviendo casos de igualdad de rangos del mismo modo que en los estadísticos previos, ignorando dichas observaciones. Después, calculamos las sumas de rangos, R_1 y R_2, de ambas muestras. A continuación, calculamos los estadísticos:

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

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

y *se compara el menor de ellos $U_m = min\ (U_1\ ,\ U_2)$, con los valores críticos de las tablas* específicamente calculadas para este estadístico, que aparecen en el Apéndice XX. Si el valor del estadístico muestral es inferior al de las tablas, se rechaza la hipótesis nula de igualdad de distribuciones de probabilidad.

Ejemplo 10.- Una determinada universidad cree que es posible que los alumnos que proceden de instituciones privadas de enseñanza media tengan distinto rendimiento académico en las licenciaturas universitarias que aquellos que provienen de instituciones públicas. Con el objeto de contrastar esta hipótesis, se recogen aleatoriamente los siguientes resultados de las pruebas de selectividad:

	Cuadro 12.7 Puntuación																	
Privada	97	68	75	95	75	84	86	94	73	69	85	93	85	88	91	74	60	74
Pública	84	65	87	93	78	77	67	95	68	74	78	74	67	91				

que genera las dos primeras filas en la siguiente tabla de rangos, obtenidos a partir de la muestra agregada:

	Cuadro 12.8 Rangos en la muestra global																		
Privada	32	5,5	13,5	30,5	13,5	18,5	22	29	8	7	20,5	27,5	20,5	24	25,5	10,5	1	10,5	319,5

Pública	18,5 2 23 27,5 16,5 15 3,5 30,5 5,5 10,5 16,5 10,5 3,5 25,5	208,5
Privada < Pública	0 10 8 0 8 4 4 1 10 10 4 1 4 3 2 8 14	99

Las sumas de los rangos son 319,5 y 208,5, respectivamente, y tenemos n_1=18 observaciones para alumnos de escuelas privadas y n_2 = 14 para los que provienen de escuelas públicas, de modo que tenemos:

$$U_1 = (18)(14) + (18)(19)/2 - 319{,}5 = 252 + 171 - 319{,}5 = 103{,}5$$
$$U_2 = (18)(14) + (14)(15)/2 - 208{,}5 = 252 + 105 - 208{,}5 = 148{,}5$$

El menor estadístico es U_1, con un valor de 103,5 que excede del valor crítico al nivel de significación de 0,05 para el contraste de una cola, que es igual a 82 (ver Tabla XX en el Apéndice XX). En consecuencia, inferimos que la menor suma de rangos no es substancialmente inferior a la suma mayor y, en consecuencia, las dos poblaciones no parecen diferentes, en base a las dos muestras de que disponemos. No rechazamos la hipótesis nula de homogeneidad poblacional, por lo que los alumnos provenientes de ambos tipos de instituciones parecen obtener resultados académicos similares.

Conviene tener en cuenta que a través de la relación: $U_1 + U_2 = n_1 n_2$, puede obtenerse un estadístico a partir del otro, de modo que basta calcular uno de ellos. Si el que calculemos es inferior a $(n_1 n_2)/2$, sabremos que hemos obtenido el estadístico más bajo, es decir, el que queremos comparar con los valores críticos de las tablas, y si no, su "complemento": $U_j = n_1 n_2 - U_i$, $i,j = 1,2$, $i \neq j$,es el estadístico que deseamos utilizar, sin necesidad de tener que utilizar la expresión que aparece en su definición.

Bajo la hipótesis nula de igualdad de distribución de probabilidad para ambas muestras, el estadístico U_m no debe tomar valores excesivamente grandes. Hay que obtener valores críticos λ_i y λ_s, tales que:

$$P\left(\lambda_i < U_m < \lambda_s\right) = 1 - \alpha$$

teniendo en cuenta que la tabla al final del texto proporciona las probabilidades *p*, $p = P(U \leq U_m)$.

Existe también para este estadístico una aproximación Normal, obtenida por Mann-Whitney para cuando tanto n_1 como n_2 aumentan de tamaño. Dicha distribución Normal, que aproxima el comportamiento del estadístico U tiene por esperanza matemática y varianza:

$$E(U_m) = \frac{n_1 n_2}{2} \ ; \qquad Var(U_m) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12};$$

y parece proporcionar aproximaciones admisibles cuando ambos tamaños muestrales son superiores a 20. En nuestro ejemplo, tenemos:

$$E(U_m) = \frac{(18)(14)}{2} = 126$$

$$Desviación \quad típica(U_m) = \sqrt{\frac{(18)(14)(18+14+1)}{12}} = 26{,}32$$

por lo que el valor estandarizado de nuestro estadístico, 103,5, es igual a:

$$Z = \frac{103{,}5 - 126}{26{,}32} = 0{,}85$$

que está notablemente por debajo de los valores críticos para el contraste de una cola en una población Normal(0,1), a los niveles de significación habituales: 0,10, 0,05 ó 0,01. En consecuencia, no rechazamos la hipótesis nula, llegando así a la misma conclusión que en el análisis anterior.

Una segunda forma de calcular el estadístico consiste en comparar cada dato de una de las dos muestras, con todos los de la otra, para contabilizar el número de ocasiones en que se tiene que el primero es inferior al segundo. Si efectuamos este ejercicio con los datos del ejemplo, la tercera fila del cuadro muestra que cada puntuación de un alumno de enseñanza privada resulta inferior a las puntuaciones obtenidas por los alumnos de enseñanza pública en 99 ocasiones. Además, puede comprobarse que se producen 9 casos de empates, con igual calificación entre alumnos de ambos tipos de enseñanza. Sumando la mitad de los empates, 4,5, al total de 99, tenemos 103,5, el mismo valor del estadístico que antes ya calculamos.

3.2.3 Contraste de Siegel-Tukey de igualdad de varianzas

El procedimiento de Mann-Whitney fue adaptado por S.Siegel y J.Tukey puede adaptarse para *contrastar si dos muestras independientes han sido extraídas de poblaciones con igual varianza*, frente a la hipótesis alternativa de que han sido extraídas de poblaciones con varianzas diferentes. Para ello, una vez ordenados todos los elementos de ambas muestras, combinados, se asignan rangos comenzando desde el menor y el mayor, hacia el centro: al menor valor se le asocia el rango 1; al valor más elevado y al que le precede se asignan los rangos 2 y 3 ; al segundo y tercer valores más bajos se asignan los rangos 4 y 5, y así sucesivamente. Si el número total de observaciones en ambas muestras es par, una de ellas se quedará sin rango. Las expresiones anteriores se utilizan para calcular el estadístico R_m, que es la suma de rangos de la muestra de menor tamaño. La interpretación del contraste estriba en que si una de las dos muestras procede de una población con mayor dispersión, recibirá los rangos menores, mientras que la que procede de una muestra de menor variabilidad recibirá los rangos mayores. Puede apreciarse que el contraste tiene interés cuando condicionamos en que ambas distribuciones tienen una media de posición central similar.

El estadístico R_m puede aproximarse, para $n_1+n_2 > 20$, por una distribución Normal:

$$R_m \sim N\left(\frac{n_m(n+1)}{2}; \frac{n_1 n_2(n+1)}{2}\right)$$

donde $n_m = \min(n_1,n_2)$, y $n = n_1+n_2$.

Ejemplo 12.7 (NOPARAM.WQ1).- Supongamos que, para tratar de evaluar las diferentes capacidades de 16 hombres y 14 mujeres, se somete a un examen lógico a una muestra de 16 estudiantes varones y 14 estudiantes mujeres, obteniendo los siguientes resultados:

	Puntuación															
Hombres	92	76	85	70	84	66	72	68	74	90	84	82	74	70	76	82
Mujeres	84	88	94	90	78	82	64	70	82	74	76	80	78	90		

Las calificaciones ordenadas son:

64m; 66h; 68h; 70h; 70m; 70h; 72h; 74m; 74h; 74h; 76m; 76h; 76h; 78m; 78m; 80m; 82m; 82m; 82h; 82h; 84h; 84h; 84m; 85h; 88m; 90m; 90h; 90m; 92h; 94m;

donde hemos anotado si corresponden a hombres o mujeres. Los rangos correspondientes a este contraste son:

64m; 66h; 68h; 70h; 70m; 70h; 72h; 74m; 74h; 74h; 76m; 76h; 76h; 78m; 78m; 80m;
82m; 82m; 82h; 82h; 84h; 84h; 84m; 85h; 88m; 90m; 90h; 90m; 92h; 94m;

con rangos:

1m 4h 5h 8h 9m 12h 13h 16m 17h 20h 21m 24h 25h 28m 29m --
27m 26m 23h 22h 19h 18h 15m 14h 11m 10m 7h 6m 3h 2m

que producen sumas de rangos 201 para hombres y 234 para mujeres. La distribución aproximada es:

$$N\left(\frac{(14)(31)}{2};\frac{(14)(16)(31)}{12}\right) = N(217;578{,}67)$$

con desviación típica: $\sigma = 24{,}06$. Como el estadístico procedente de la muestra de menor tamaño, la de mujeres, es 234, que no resulta significativamente distinto del valor esperado en una población $N(217;578{,}7)$, no rechazamos la hipótesis nula de que ambas poblaciones tienen igual dispersión.

3.2.4 El contraste de Kruskal-Wallis

Este estadístico, propuesto por W.H.Kruskal y W.A.Wallis en 1952, generaliza el estadístico *U*, cuando se trabaja con más de 2 muestras independientes y se pretende contrastar la hipótesis nula de que todas ellas proceden de la misma población. El procedimiento es idéntico al seguido con el estadístico *U*, sólo que ahora utilizamos el estadístico *H*:

$$H = \frac{12}{n(n+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} + \ldots + \frac{R_k^2}{n_k} \right) - 3(N+1)$$

siendo n_i el tamaño de cada una de las *k* muestras independientes de que se dispone, y R_i las respectivas sumas de rangos. *n* denota la suma de los *k* tamaños muestrales: $n = \sum_{i=1}^{k} n_i$. Todo lo que se requiere es que la población o poblaciones de las que se extrajeron las muestras, sean de tipo continuo. La hipótesis nula es que las medias de las *k* poblaciones de donde se extrajeron las muestras son idénticas. Para ello, se ordenan los valores combinados de todas las *k* muestras en orden creciente, y se asignan rangos.

La distribución del estadístico *H* depende del número de muestras consideradas y de sus tamaños, por lo que es preciso disponer de un amplio número de tablas de valores críticos. Kruskal y Wallis probaron que si la hipótesis nula es cierta, y si cada muestra consta de al menos 5 observaciones, entonces el estadístico *H* sigue una distribución chi-cuadrado con *k*-1 grados de libertad, y rechazamos la hipótesis nula de homogeneidad de las poblaciones cuando el valor del estadístico *H* que hemos obtenido en nuestra aplicación excede del valor de las tablas de dicha distribución. Las tablas de Kruskal-Wallis dependen del número de muestras distintas, *k*, por lo que hay que disponer de varias de ellas. No las presentamos aquí, y sugerimos la aproximación χ_{k-1}
mencionada.

El contraste de Kruskal-Wallis es más potente que un contraste no paramétrico semejante, como el contraste de la mediana que a continuación presentamos, cuando se utiliza para varios grupos. Es algo menos potente que el contraste paramétrico *F*, pero más sencillo de utilizar.

Ejemplo 12.8.- Una agencia oficial está preocupada acerca de la posibilidad de que las prácticas de precios que se siguen en ciudades pequeñas para un determinado producto, reflejan prácticas de monopolio. Para contrastar dicha posibilidad toma muestras de tamaño 6 en tres ciudades: una de menos de 50.000 habitantes, otra de menos de 500.000 habitantes, y otra muestra en una gran ciudad. Si detecta que los precios son mayores en las ciudades pequeñas, tendrá cierta evidencia acerca de práctica monopolista en las mismas, por la poca presencia de competencia. Las observaciones muestrales fueron:

Precios en ciudad pequeña: 186, 162, 146, 174, 192, 178
Precios en ciudad media: 166, 142, 180, 135, 162, 174
Precios en gran ciudad: 142, 126, 118, 135, 127, 156

que generan los rangos en la muestra global:

Rangos en ciudad pequeña: 17; 10,5; 8; 13,5; 18; 15 Suma = 82
Rangos en ciudad media: 12; 6;5, 16; 4,5; 10,5; 13,5 Suma = 63
Rangos en gran ciudad: 6,5; 2; 1; 4,5; 3; 9 Suma = 26

que producen un valor del estadístico de Kruskal-Wallis de 9,94, superior a los valores críticos de de una χ_2^2, que son de 5,99 y 7,38 al 95% y 97,5%, por lo que rechazamos la hipótesis nula de ausencia de diferencias en las prácticas de fijación de precios, obteniendo cierta evidencia a favor de prácticas de monopolio.

3.2.5 Contrastes chi-cuadrado de Pearson

Este contraste se diferencia del de ajuste teórico por el hecho de que, en los contrastes de homogeneidad, no especificamos cuál es la población estadística hipotética de la que las muestras de que disponemos habrían sido seleccionadas, sino que contrastamos tan sólo que estos han sido extraídas de la misma población, sin especificar cuál sea ésta.

Ejemplo 12.9.- Supongamos que se pretende contrastar si la distribución de tipos sanguíneos es igual en hombres que en mujeres, para lo que se efectúa un análisis de sangre a 12.000 personas, 5.000 hombres y 7.000 mujeres, teniendo los siguientes resultados:

	Tipo A	Tipo B	Tipo 0	Total
Hombres	2.400	1.900	700	5.000
Mujeres	3.100	2.700	1.200	7.000
TOTAL	5.500	4.600	1.900	12.000

Si las muestras fuesen homogéneas, como afirma la hipótesis nula, la mejor estimación de la proporción de personas con sangre tipo A en la población, sería la obtenida a partir de todas las observaciones muestrales, con independencia de si la sangre fue extraída de un hombre o de una mujer. Así, las proporciones estimadas de personas con sangre tipo A, tipo B y tipo 0, serían, respectivamente, 55/120, 46/120 y 19/120, es decir: 0,4583, 0,3833 y 0,1583. Entonces, esperaríamos encontrar en nuestra muestra (teóricamente) que un número de (0,4583)(5.000) = 2.292 hombres tiene sangre tipo A, (0,3833)(5.000) = 1.917 tienen sangre tipo B, y (0,1583)(5.000) = 792 tienen sangre tipo 0. Haciendo un cálculo similar para las mujeres, obtendríamos el cuadro de frecuencias teóricas (bajo la hipótesis nula):

	Tipo A	Tipo B	Tipo 0	Total
Hombres	2.292	1.917	792	5.000
Mujeres	3.208	2.683	1.108	7.000

TOTAL	5.500	4.600	1.900	12.000

y el estadístico chi-cuadrado:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - T_i)^2}{T_i} = \frac{(108{,}3)^2}{2292} + \frac{(-16{,}7)^2}{1917} + \frac{(-91{,}7)^2}{792} +$$

$$+ \frac{(-108{,}3)^2}{3208} + \frac{(16{,}7)^2}{2683} + \frac{(91{,}7)^2}{1108} = 27{,}2$$

que compara las frecuencias observadas O_i con las frecuencias teóricas T_i, excede de 5,991 y 9,210, que son los valores críticos de una chi-cuadrado con (3-1)(2-1) = 2 grados de libertad, lo que nos lleva a rechazar la hipótesis de homogeneidad en la distribución de tipos sanguíneos entre hombres y mujeres.

3.2.6 El contraste de rachas

El contraste de rachas sirve para discutir la homogeneidad de las poblaciones de las que fueron extraídas dos muestras independientes. Se basa en la ordenación natural, creciente o decreciente de los elementos de ambas muestras, pero en vez de utilizar los rangos, utiliza las *rachas*, secuencias de valores sucesivos de una de las muestras. Si las muestras proceden de la misma población, entonces los valores de ambas se alternarán en la lista global, de modo puramente aleatorio. Si, por el contrario, una de las muestras procede de una población *situada*, en algún sentido, a la derecha de la otra, entonces los valores de ambas muestras tenderán a aparecer en la forma de unas pocas rachas amplias. El contraste se resuelve, por tanto, utilizando el número de rachas, del modo que vamos a ver con unos ejemplos. De esta discusión se intuye que el contraste de rachas es útil para analizar la posible heterogeneidad de dos poblaciones cuando la alternativa es que la función de distribución de una de ellas está desplazada a la derecha con respecto a la otra. No es excesivamente potente cuando las distribuciones tienen similar posición central, y difieren en su varianza.

Ejemplo 12.10.- Los dividendos distribuidos por una muestra que engloba a empresas eléctricas, de comunicaciones y transportes, aumentaron un determinado año con respecto al año previo: 5,2%; 2,3%, 6,5%; 3,4%, 4,5%, 3,5%, 6,2%, 4,0%, 4,7%, 3,2%, 5,1%, 4,6%, mientras que los repartido por una muestra de bancos crecieron: 8;4%, 4,2%, 3,6%, 6,7%, 7,2%, 7,4%, 10,5%, 6,6%, 6,3%, 7,0%. Al establecer una única secuencia, tenemos:

2,3; 3,2; 3,4; 3,5; 3,6; 4,0; 4,2; 4,5; 4,6; 4,7; 5,1; 5,2; 6,2; 6,3; 6,5; 6,6; 6,7; 7,0; 7,2;
<u>E E E E</u> <u>B</u> <u>E</u> <u>B</u> <u>E E E E E E</u> <u>B</u> <u>E</u> <u>B B B B</u>

7,4; 8,4; 10,5
<u>B B B</u>

que comprende una primera racha de datos de dividendos de 4 empresas no bancarias, una racha de 1 banco, una racha de 1 empresa, una racha de 1 banco, una racha de 6 empresas, un banco, una empresa no bancaria y una última racha de 7 bancos. Hay un total de 8 rachas.

Aunque hay tablas para este contraste de rachas, tampoco las vamos a presentar, porque la aproximación Normal es suficientemente buena en la mayoría de los casos. Cuando el tamaño de ambas muestras es al menos igual a 10, el número de rachas sigue una distribución que puede aproximarse por:

$$R \sim N\left(\frac{2n_1n_2}{n_1+n_2}+1\,;\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}\right) \simeq N\left(\frac{2n_1n_2}{n_1+n_2}+1\,;\frac{4n_1^2n_2^2}{(n_1+n_2)^3}\right)$$

donde la última expresión para la varianza, más simple, es válida para tamaños muestrales grandes.

En nuestro ejemplo, el número esperado de rachas es 11,9, con desviación típica de 2,269, por lo que el intervalo del 95% de confianza es (11,9-4,45 ; 11,9+4,45) = (7,45 ; 16,35). Como el número de rachas en la muestra está dentro del intervalo, no rechazamos la hipótesis nula de aleatoriedad pura en la generación de los incrementos de dividendos. En consecuencia, no detectamos evidencia en contra de que las políticas de dividendos de empresas bancarias y no bancarias sean iguales. Además, el intervalo del 90% de confianza es: (11,9-3,73 ; 11,9+3,73) = (8,16 ; 15,63), que conduce a rechazar la hipótesis de igualdad. El bajo número de observaciones sugiere utilizar un nivel de confianza en línea con el 90%, por lo que la hipótesis de igualdad debería rechazarse.

El resultado de este contraste es algo sorprendente, pues la estructura de rachas al principio y final de la muestra sugiere todo lo contrario, que las políticas de incremento de dividendos no son iguales. Cabe hacer varios comentarios respecto al resultado: 1) con una racha menos, habríamos rechazado la hipótesis nula, 2) si sospechamos, como sugiere la evidencia muestral, que los bancos han aumentado sus dividendos más que las empresas no bancarias, entonces tenderíamos a encontrara pocas rachas, con las empresas no bancarias concentradas al comienzo de la lista ordenada, y los bancos concentrados al final de dicha relación. El contraste debería ser entonces de una cola, rechazando la homogeneidad si el número de rachas es inferior a un cierto número crítico. Dicho nivel crítico se obtiene, al 95%, utilizando el umbral de 1,645 para la normal: 11,9 - (1,645)(2,27) = 11,9 - 3,73 = 8,17. Como el número de rachas en la muestra, 8, es inferior a este umbral, rechazamos la hipótesis nula de homogeneidad.

4. CONTRASTES DE ALEATORIEDAD

Quizá el contraste más básico se refiere al hecho de que las observaciones disponibles proceden de una población realmente aleatoria, frente a la alternativa de que hay cierto determinismo en ellas, lo que consideramos en esta sección.

4.1 Contraste de rachas

El contraste que presentamos en esta sección se refiere a una situación en que las observaciones muestrales pueden presentar una de dos determinadas características. Si éstas se producen de manera puramente aleatoria, al observar la secuencia de extracciones, ambas características se alternarán de modo totalmente impredecible. Imaginemos, sin embargo, que al extraer secuencialmente los elementos muestrales tenemos una *racha* larga en la que se produce la primera característica, seguida de otra larga *racha* de elementos en los que aparece la segunda característica. Pensaremos entonces que hemos extraído elementos de dos poblaciones diferentes. En este caso, un número reducido de rachas nos ha proporcionado evidencia en contra de la aleatoriedad de la muestra.

Los contrastes de rachas pueden ser útiles, pero hay que diseñarlos específicamente en función de cuál sea la hipótesis alternativa. Por ejemplo, supongamos que la hipótesis nula es muestra aleatoria, frente a una alternativa de valores crecientes con cierta tendencia, y que examinamos la relación de elementos muestrales *en el orden en que fueron extraídos*. Asignamos a cada elemento muestral un indicador E o D, según que esté por encima o por debajo de la mediana muestral m. En este caso, obtendremos evidencia en contra de H_0 si observásemos un número reducido de rachas, con indicadores D concentrados al comienzo de la relación, e indicadores E concentrados hacia el final de la relación. El contraste, que debería ser de una cola en este caso, se lleva a cabo como describimos en la Sección 3.2.6.

Podría ocurrir que sospechemos que los valores altos y bajos se alternan sistemáticamente en la extracción de la muestra, siendo ésta la alternativa a la aleatoriedad pura. En tal caso, tendríamos evidencia en contra de H_0 si el número de rachas fuese elevado, por lo que deberíamos efectuar un contraste de una cola, pero esta vez rechazando H_0 si el numero de rachas es *suficientemente elevado*. Si quisiéramos contrastar frente a ambas posibilidades simultáneamente, entonces elaboraríamos un contraste de dos colas.

4.2 Contraste de diferencias sucesivas

En datos temporales, podemos estar interesados en contrastar que las observaciones muestrales no son independientes, en el sentido de que un valor elevado condiciona que la siguiente extracción muestral, haciendo más probable que ésta sea asimismo elevada. Decimos entonces que tenemos una situación de *correlación serial positiva*. Si un valor elevado hace más probable que la siguiente extracción muestral sea reducida, decimos que tenemos *correlación serial negativa*. Si consideramos el cuadrado medio de las diferencias sucesivas de los valores x_t observados:

$$d^2 = \frac{\sum_{i=2}^{n} (x_t - x_{t-1})^2}{2(n-1)}$$

que, bajo la hipótesis nula de independencia, tiene un valor esperado de σ^2. Por tanto, si tomamos el cociente entre este estadístico y la cuasivarianza muestral:

$$r = \frac{d^2}{s^2} = \frac{\sum_{t=2}^{n} (x_t - x_{t-1})^2}{2\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

bajo H_0, este estadístico será próximo a 1, puesto que la cuasivarianza es un estimador insesgado de la varianza poblacional. Si existe correlación serial positiva, cada elemento muestral tenderá a ser similar a su predecesor, con lo que r se hará pequeño. Si existe correlación negativa, r se hará grande. Para muestras grandes, la distribución de probabilidad de r puede aproximarse por:

$$r \sim N\left(1\,;\frac{n-2}{n^2-1}\right)$$

5. CONTRASTES DE ASOCIACIÓN ENTRE DISTRIBUCIONES

Los contrastes que analizamos en esta sección están diseñados para analizar si dos características de una población están o no relacionadas. Posteriormente, aprenderemos cómo cuantificar el grado de correlación entre variables utilizando modelos econométricos, pero los contrastes que aquí introducimos tienen dos ventajas: 1) no están condicionados por ningún supuesto acerca del tipo de distribución que sigue la población de la que se extrajeron las observaciones muestrales, y 2) no está condicionado en una hipótesis de causalidad acerca de la dirección en que se produce la asociación entre dos variables.

En tal sentido, los estadísticos que aquí introducimos son semejantes al coeficiente de correlación lineal de los Capítulos XX y XX, pero tienen frente a éste la enorme ventaja de ser utilizables con características cualitativas.

5.1 Contraste de Spearman de correlación por rangos

La *correlación por rangos* consiste en examinar el grado de correspondencia que existe entre los rangos que asignamos a las características cuyo grado de asociación queremos medir. Como siempre, los rangos pueden asignarse en orden creciente o decreciente. Cuando las observaciones muestrales no son cualitativas, sino que son cuantificables, o se pueden ordenar de acuerdo con un determinado criterio, entonces pueden definirse unos pseudo-coeficientes de correlación, que cuantifican el grado de asociación entre distribuciones, a partir de los rangos.

El *coeficiente de correlación de rangos*, propuesto por Carl Spearman en 1904, es:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

donde d_i es la diferencia entre los rangos correspondientes a los valores de las dos características correspondientes a un mismo elemento u observación muestral.

Ejemplo 12.12.- Para examinar el grado de relación que existe entre las calificaciones de Macroeconomía y Matemáticas, se tomaron las notas de 15 alumnos, teniendo los datos del Cuadro XX. Tras las calificaciones de las materias de Macroeconomía y Matemáticas, hemos obtenido las dos columnas de rangos de cada una de ellas, y la columna 6 tiene precisamente las diferencias de rangos, d_i, que no es sino la diferencia de las dos columnas anteriores. El *coeficiente de correlación de rangos de Spearman* toma valores entre -1 y +1. Al igual que con el coeficiente de correlación lineal, un valor próximo a cero indicaría que ambas variables apenas tienen relación. Valores próximos a la unidad indican que ambas variables tienen una relación estrecha. Si es positivo, la relación es directa, es decir, valores elevados de Y vienen asociados a valores elevados de X , y valores bajos de X, con valores bajos de Y. Si el coeficiente de correlación por rangos es negativo, valores elevados de una variable vienen asociados con valores reducidos de la otra. Al igual que con los otros estadísticos de rangos antes estudiados, cuando varias observaciones son iguales se les asigna el promedio de los rangos que les corresponderían.

El valor del coeficiente de correlación de rangos en este ejemplo es:

	Calificaciones		Rangos			
Alumno	Macroeconomía	Matemáticas	Macroeconomía	Matemáticas	d_i	d_i^2
1	70	64	7	9	-2,0	4,0
2	82	90	4	1	3,0	9,0
3	54	36	11	15	-4,0	16,0
4	91	86	2	2,5	-0,5	0,25
5	64	52	10	12,5	-2,5	6,25
6	87	76	3	6	-3,0	9,0
7	32	40	14	14	0,0	0,0
8	74	82	6	5	1,0	1,0
9	66	54	9	11	-2,0	4,0
10	42	58	13	10	3,0	9,0
11	52	66	12	8	4,0	16,0
12	30	52	15	12,5	2,5	6,25
13	78	84	5	4	1,0	1,0
14	92	86	1	2,5	-1,5	2,25
15	68	72	8	7	1,0	1,0
			120	120	0	85

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{(6)\ (85)}{(15)\ (15^2-1)} = 1 - 0{,}1518 = 0{,}8482$$

mostrando una *relación muy estrecha* entre las calificaciones que los alumnos obtienen en ambas asignaturas.

La significatividad del coeficiente de correlación de rangos de Spearman puede calcularse aplicando la relación:

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

para comparar el valor de t con la tabla de la distribución t de Student con n-2 grados de libertad. Esta aproximación es válida para $n \geq 10$, que en este ejemplo conduce a una valor numérico de 5,774 que está fuera del intervalo de confianza de la t de Student con 13 grados de libertad, a niveles del 95% ó 99%. En consecuencia, *rechazamos la hipótesis nula de ausencia de asociación entre ambas calificaciones*, y pasamos a mantener la alternativa de asociación positiva entre ambas, ya que el coeficiente de correlación de rangos es positivo.

Alternativamente a este tipo de análisis, los procedimientos de regresión que analizaremos en los Capítulos 13 a 15 permiten cuantificar la relación entre las calificaciones obtenidas en ambas materias, si bien una justificación totalmente rigurosa descansa en el supuesto de distribución Normal de las observaciones muestrales, que, en ocasiones el investigador no está en condiciones de asegurar.

Debe notarse que, aunque nuestro ejemplo caracterizaba la relación entre dos variables de naturaleza cuantitativa, el mismo tipo de análisis puede llevarse a cabo para variables cualitativas, siempre que éstas puedan ordenarse con una determinada escala ordinal. De este modo, el coeficiente de correlación de rangos de Spearman puede interpretarse como pseudo-coeficientes de correlación de rangos entre dichas variables.

5.2 Contraste de Kendall

El estadístico τ de Kendall es otro coeficiente de correlación de rangos. Para utilizarlo en la medición de la asociación entre dos atributos, se obtienen primero los rangos de ambas variables X e Y. A continuación, se ordenan las observaciones de una de ellas, X, por ejemplo, de acuerdo con su orden natural, y se considera la lista de los rangos de Y ordenada del modo que así resulta. Se compara cada rango de Y con cada uno de los que le siguen, y se define una función indicatriz que toma el valor +1 cuando los dos rangos de Y en la comparación están ordenados, y el valor -1 cuando están invertidos. El rango de Y que aparece en primer lugar, es decir, el que corresponde con el menor[1] rango de X, se compara con *todos* los que le siguen. El rango de Y correspondiente al segundo rango de X se compara con los n-1 rangos de Y que le siguen, y así hasta el correspondiente al penúltimo rango de X, que se compara con el rango de Y correspondiente al último rango de X.

El número total de comparaciones es: $(n-1)+(n-2)+...+1 = \frac{(n-1)n}{2}$, y el estadístico de Kendall es el cociente entre la suma S de todos los valores de la función indicatriz y el número total de comparaciones:

$$\tau = \frac{S}{(n-1)n/2} = \frac{2S}{(n-1)n}$$

Si los rangos de Y resultasen quedar ordenados cuando hemos seguido el orden natural de las observaciones de X, entonces la suma S tomaría su máximo valor numérico, que con el número total de comparaciones, y el estadístico τ sería igual a 1. Si los rangos de Y quedasen en un orden exactamente inverso al natural, el estadístico τ tomaría entonces su mínimo valor posible, que es -1. En general: $-1 \leq \tau \leq 1$. La hipótesis nula de ausencia de asociación entre ambas variables hará que en cada comparación de rangos de Y hay la misma probabilidad de que sus rangos estén o no invertidos, por lo que habrá aproximadamente tantos 1 como -1, y la suma

[1] O mayor, según se haya ordenado la muestra, lo cual es irrelevante.

S será próxima a cero.

Para muestras de tamaño superior a 10, la distribución del estadístico puede aproximarse por:

$$\tau \sim N\left(0;\frac{2(2n+5)}{9n(n-1)}\right)$$

Por tanto, se trata de construir un intervalo de confianza para τ al nivel deseado, 100(1-α), utilizando sus tablas, y rechazar la hipótesis de ausencia de asociación si el valor numérico de τ cae fuera de dicho intervalo.

Ejemplo 12.13.- Volviendo a examinar las calificaciones de Macroeconomía y Matemáticas del ejemplo anterior, si ordenamos las primeras, y las acompañamos de la calificación en Matemáticas, tenemos:

Macroeconomía: 30 32 42 52 54 64 66 68 70 74 78 82 87 91 92
Matemáticas: 52 40 58 66 36 52 54 72 64 82 84 90 76 86 86

donde se producen 91 comparaciones en el orden correcto, y 14 inversiones en las calificaciones de Matemáticas. Es decir, hay 14 ocasiones en que una calificación de Matemáticas viene seguida de una calificación más baja. El número total de comparaciones es (15)(14)/2 = 105. Por tanto, la distribución del estadístico τ puede aproximarse por una $N(0;0{,}0307)$, con desviación típica σ = 0,192. El estadístico τ toma valor: τ = (91-14)/105 = 0,733, cuyo cociente por la desviación típica: 0,733/0,192 = 3,82 excede de los valores críticos de la $N(0,1)$ a los niveles de significación habituales, por lo que rechazamos la hipótesis nula de independencia, y pasamos a creer que existe asociación entre las calificaciones recibidas por los alumnos en ambas materias.

4.3. Tablas de contingencia

Las tablas de contingencia se utilizan para contrastar la posible independencia de dos características de los elementos muestrales. Por ejemplo, podemos contrastar si la edad de un votante está relacionada con su opinión política, el sexo de una persona con la opción profesional que escoge, etc.. Como en los contrastes anteriores, la hipótesis nula es la ausencia de asociación entre ambas características, y la tabla de contingencia puede utilizarse con variables discretas de cualquier tipo, cualitativo o cuantitativo. De hecho, uno de los aspectos de interés de estas tablas es su aplicabilidad al contraste de ausencia de asociación entre variables aleatorias de tipo cualitativo.

Ejemplo 12.14.- Supongamos que se celebra un referéndum que considera la posibilidad de introducir una legislación de carácter ecológico, que lleva asociada importantes restricciones de acceso a la zona afectada. En una muestra de 5.000 votantes, 3.000 mujeres y 2.000 hombres, observamos que votaron a favor de introducir la legislación: 1.200 de los 1.800 votantes menores de 25 años, 900 de los 2.100 votantes entre 25 y 35 años, y 400 de los 1.100 votantes mayores de 35 años. Queremos contrastar la hipótesis de que la opinión de los votantes no es independiente de su edad, para lo que construimos la tabla:

	Menores de 25 años	Entre 25 y 35 años	Mayores de 35 años	Total
A favor	1.200	1.200	400	2.800
En contra	600	900	700	2.200
Total	1.800	2.100	1.100	5.000

en la que debe apreciarse que hemos calculado la última columna a partir de la información muestral, agregando, por ejemplo, el número de votantes favorables a la propuesta, con independencia de su edad. La distribución por edades de los votantes favorables a la propuesta no es uniforme, pero tampoco lo es la composición por edades de la muestra total. Precisamente, se trata de contrastar si la distribución por edades de las 2.500 personas favorables a la propuesta no es significativamente diferente de la composición por edades de la muestra. Ello sería evidencia a favor de la ausencia de asociación entre opiniones y edades.

Los totales parciales que aparecen, tanto en la última fila de la tabla como en la última columna, son las distribuciones de probabilidad marginales de las dos variables que configuran la distribución bivariante objeto de análisis: edad y opinión hacia el referéndum.

La hipótesis nula en este tipo de contrastes es siempre la de independencia entre las dos variables consideradas. Bajo independencia, la probabilidad de encontrar en la muestra una persona inferior a 25 años que estuviese a favor de la propuesta sería igual al producto de las probabilidades de dos sucesos independientes: 1) que al considerar una persona al azar de entre la muestra, resulte ser un joven con edad inferior a 25 años, y 2) que dicha persona resulte tener una opinión favorable a la legislación. Hay 1.800 personas con edad inferior a 25 años en la muestra, por lo que la probabilidad del primer suceso es 1.800/5.000; por otra parte, hay 2.800 personas a favor de la legislación, por lo que la probabilidad del segundo suceso es: 2.800/5.000. Así, la probabilidad del suceso compuesto: obtener una persona inferior a 25 años, favorable a introducir la legislación, sería:

$$P(menor\ de\ 25\ años,\ favorable\ a\ la\ legislación) =$$

$$=P(menor\ de\ 25\ años).P(favorable\ a\ la\ legislación)=$$

$$= \frac{1.800}{5.000}.\frac{2.800}{5.000} = 0{,}2016$$

por lo que, bajo la hipótesis nula, la frecuencia muestral observada para este suceso (casilla (1,1) en el Cuadro) debería haber sido 0,2016 por el tamaño muestral (5.000), es decir, 1.008 personas. Razonando de modo similar, puede completarse el cuadro de frecuencias teóricas:

	Menores de 25 años	Entre 25 y 35 años	Mayores de 35 años	Total
A favor	1.008	1.176	616	2.800
En contra	792	924	484	2.200

Total	1.800	2.100	1.100	5.000

El contraste de independencia se basa en un estadístico chi-cuadrado del tipo que hemos visto en contrastes previos, que compara las frecuencias realmente observadas en la muestra para cada suceso compuesto, con las que teóricamente deberían haberse observado en el caso de ser cierta la hipótesis nula de independencia:

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{(O_{ij} - T_{ij})^2}{T_{ij}}$$

donde O_i y E_i denotan, respectivamente, las frecuencias observadas (Cuadro XX) y teóricas (Cuadro XX). Se trata, por consiguiente, de comparar las frecuencias que aparecen en cada casilla en las dos tablas anteriores. Si las frecuencias fuesen similares, ello querría decir que las frecuencias observadas en la muestra no difieren significativamente de las que teóricamente deberían haberse obtenido bajo independencia, por lo que no rechazaríamos la hipótesis nula de independencia. Ello vendrá indicado por un valor reducido del estadístico chi-cuadrado. Alternativamente, un valor "suficientemente" elevado de dicho estadístico nos llevará a rechazar la hipótesis nula, habiendo encontrado evidencia en el sentido de que las dos variables consideradas están relacionadas.

El número de grados de libertad del estadístico chi-cuadrado es el número de casillas independientes en la tabla de contingencia. Ocurre que no todas ellas lo son porque, una vez que se tiene la muestra, el número de personas de cada tramo de edad está dado. En nuestro ejemplo, puesto que hay en la muestra 2.800 personas a favor de la proposición de ley, siendo 1.200 de edad inferior a 25 años y otras 1.200 entre 25 y 35 años, entonces debe haber necesariamente 400 personas a favor de la propuesta de ley con edad superior a los 35 años. De igual modo, puesto que hay 2.200 personas en contra de la proposición, siendo 600 de ellas de edad inferior a los 25 años y 900 de edad entre 25 y 35 años, entonces hay, necesariamente 700 personas de edad superior a 35 años, en contra de la propuesta de ley. Hay sólo dos frecuencias independientes, coincidiendo con el producto
(m-1)(n-1) que ya hemos propuesto en ocasiones previas, donde m denota el número de columnas y n el número de filas de la tabla de frecuencias.

En este ejemplo, el valor numérico del estadístico chi-cuadrado:

$$\chi^2 = \frac{(1200-1008)^2}{1008} + \frac{(1200-1176)^2}{1176} + \frac{(400-616)^2}{616} + \frac{(600-792)^2}{792} + \frac{(900-924)^2}{924} + \frac{(700-484)^2}{484}$$

excede con mucho del valor crítico de las tablas de la distribución chi-cuadrado a los niveles de significación habituales, 0,05 ó 0,01, por lo que rechazamos la hipótesis nula de independencia, manteniendo la hipótesis alternativa de que existe asociación entre edades y opinión frente a la propuesta. A continuación el investigador debería examinar la dirección de tal asociación. para ello, compararía la distribución por edades de personas a favor de la propuesta, que es: un 42,9% de las personas favorables tiene menos de 25 años, otro 42,9% está entre 25 y 35 años, y el 14,3% restante es mayor de 35 años. La concentración de la opinión favorable entre los jóvenes es mayor que la distribución por edades de la muestra global, que es de: un 36% de personas menores de 25 años, un 42% entre 25 y 35 años, y un 22% mayores de 35 años, lo que sugiere que la opinión favorable a la legislación está concentrada de modo significativo en las personas menores de 25 años.

El contraste de independencia puede aplicarse con irrelevancia del número de opciones a que da lugar cada una de las variables consideradas. La distribución chi-cuadrado de este estadístico es tan sólo aproximada, por lo que no debe utilizarse salvo cuando las frecuencias de todas las casillas de la tabla de contingencia excedan de 5.

En general, el contraste de independencia con proporciones muestrales en una tabla de contingencia 2x2 es equivalente a un contraste de significación de diferencias de proporciones mediante la aproximación Normal, como vemos en el siguiente ejemplo:

Ejemplo 12.14.- Consideramos la posible dependencia existente entre el tipo de carrera escogida (técnica, humanística) por los alumnos que acceden por primera vez a la universidad y su sexo. Tras encuestar a 400 alumnos, obtuvimos la tabla de contingencia:

	Técnica	Humanista	
Masculino	80	110	190
Femenino	125	85	210
Total	205	195	400

que, como el lector puede comprobar fácilmente, conduce a rechazar la hipótesis nula, en favor de la alternativa, que afirma que una proporción superior de mujeres que de hombres prefiere las carreras técnicas. El valor numérico del estadístico chi-cuadrado es de 11,59, muy por encinma de los valores críticos de la distribución χ_1^2 a los niveles de confianza habituales. Por eso se rechaza la hipótesis de independencia.

Este contraste puede asimismo plantearse como un contraste de igualdad de proporciones entre dos poblaciones distintas, las de estudiantes de distinto sexo que acceden a un tipo de estudios. Para cada persona encuestada, disponemos de una variable tipo Bernouilli: X=1 si se eligió una carrera técnica, X=0 si se escogió una carrera humanista, por lo que el número de personas de un determinado sexo que escogió carrera técnica es una variable *B(n,p)*. Recordando que la varianza de una proporción es igual a $p(1-p)/n$, tendríamos las estimaciones de las proporciones por sexo, y sus varianzas:

$$p_m = \frac{8}{19} = 0{,}42; \quad Var(p_m) = \frac{\frac{8}{19}\frac{11}{19}}{190} = 0{,}001283;$$

$$p_f = \frac{125}{210} = 0{,}60; \quad Var(p_f) = \frac{\frac{125}{210}\frac{85}{210}}{210} = 0{,}001147;$$

El estadístico para el contraste de igualdad de proporciones se basa en comparar su diferencia numérica: $p_m - p_f$ con la desviación típica de dicha diferencia, bajo el supuesto de independencia de ambas proporciones, es decir:

$$Var(p_m - p_f) = Var(p_m) + Var(p_f) = 0{,}001283 + 0{,}001147 = 0{,}002430$$

Por tanto, el valor estandarizado de nuestro estadístico es:

$$\frac{0{,}42-0{,}60}{\sqrt{0{,}002430}}=\frac{0{,}42-0{,}60}{0{,}0493}=3{,}651$$

que excede del valor crítico de una Normal(0,1) a los niveles de significación habituales de 0,05 ó 0,01. Podemos utilizar la distribución Normal porque cada una de las dos proporciones se ha estimado con un número elevado de observaciones, 190 hombres y 210 mujeres, de modo que cada una de las dos proporciones muestrales es aproximadamente Normal, y también lo será su diferencia.

4.4. Coeficientes de correlación para datos cualitativos

El grado de relación existente entre las dos variables que se recogen en las tablas de contingencia que analizamos en secciones anteriores puede evaluarse a partir de una medida que tiene la interpretación de un coeficiente de correlación. Para ello, se divide el valor del estadístico chi-cuadrado por el número n de observaciones, y se extrae la raíz cuadrada:

$$\rho=\sqrt{\frac{\chi^2}{n}}$$

o, más generalmente:

$$\rho=\sqrt{\frac{\chi^2}{n(k-1)}}$$

para tablas cuadradas kxk, con $k>2$, que toma valores positivos, entre cero y uno, y se denomina *coeficiente de correlación de atributos*. Este coeficiente es un complemento al contraste de independencia, y cobra sentido cuando la hipótesis nula de independencia es rechazada. Es entonces cuando el investigador estará interesado en aportar un indicador del grado de dependencia entre los atributos considerados, y este estadístico es una herramienta útil a tal fin.

Existe otro procedimiento de normalizar la variable chi-cuadrado del contraste de contingencia, que permite comparar su valor numérico entre aplicaciones diferentes. Se trata del *coeficiente de contingencia C*, que se define:

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

por lo que:

$$C^2=\frac{\chi^2}{n+\chi^2}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{O_{ij}^2}{T_{ij}}-n}{\sum\sum\frac{O_{ij}^2}{T_{ij}}}$$

En el caso del ejemplo 12.XX, el estadístico d eindpendencia toma un valor de 11,59, por lo que el coeficiente de contingencia es:

$$C = \frac{11{,}59}{411{,}59} = 0{,}17$$

mostrando cierto grado de relación entre Sexo y preferencia por estudios técnicos frente a humanistas.

El máximo valor de este estadístico en tablas con igual número de filas que de columnas es: $C_{máx} = \sqrt{\frac{k-1}{k}}$, siendo k el número de filas o columnas, por lo que es siempre inferior a la unidad. Comparando el valor numérico del estadístico C en una determinada aplicación con su máximo valor posible, se tiene una idea del grado de correspondencia entre ambas variables. La hipótesis nula del contraste es, en este caso: H_0: $C = 0$, o, lo que es lo mismo: $O_{ij} = T_{ij}$ $\forall i,j$, es decir, que no hay asociación entre las variables, siendo la hipótesis alternativa que existe asociación, y el coeficiente de contingencia C es estadísticamente diferente de cero. Para contrastar esta hipótesis se utiliza la tabla de la distribución chi-cuadrado, con un número de grados de libertad igual a $(m-1)(n-1)$.

Una limitación del coeficiente de contingencia es que, si bien toma el valor cero cuando las variables no están relacionadas, no alcanza necesariamente el valor 1 cuando están perfectamente relacionadas.

APÉNDICE 12.1: **Derivación del coeficiente de correlación de rangos de Spearman.**

Dada una variable cualitativa, supongamos que hemos calculado los rangos x_i e y_i de sus respectivas observaciones, una vez ordenadas de acuerdo con algún criterio. Si quisiéramos calcular el coeficiente de correlación lineal de dichos rangos, tendríamos:

$$r = \frac{\frac{1}{n}\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum (x_i-\bar{x})^2 \frac{1}{n}\sum (y_i-\bar{y})^2}}$$

pero siendo x_i e y_i rangos, tenemos que su suma coincide con la de los n primeros números naturales, es decir: $n(n+1)/2$, y la de sus cuadrados, con la de los cuadrados de estos: $n(n+1)(2n+1)/6$, por lo que sus medias: $\bar{x}$ e $\bar{y}$ son igual a: $(n+1)/2$ y sus varianzas:

$$Var(x) = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2 = \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2 = \frac{n^3-n}{12}$$

Por otra parte:

$$\sum_i (x_i-y_i)^2 = \sum_{i=1}^{n}\left[(x_i-\bar{x})-(y_i-\bar{y})\right]^2 = \sigma_x^2 + \sigma_y^2 - 2\sum_i (x_i-\bar{x})(y_i-\bar{y})$$

de modo que, teniendo en cuenta que $x_i - y_i$ es la diferencia de rangos, d_i, tenemos:

$$\sum_i (x_i-\bar{x})(y_i-\bar{y}) = \frac{\sigma_x^2+\sigma_y^2-\sum_{i=1}^{n} d_i^2}{2} = \frac{\frac{n^3-n}{12}+\frac{n^3-n}{12}-\sum_{i=1}^{n} d_i^2}{2} = \frac{n^3-n}{12}-\frac{\sum_{i=1}^{n} d_i^2}{2}$$

que llevada a la expresión analítica del coeficiente de correlación lineal r anterior, deriva en el coeficiente de correlación de rangos de Spearman.

Este coeficiente oscila entre -1 y +1. Si la concordancia de rangos es perfecta, entonces: $d_i = x_i-y_i = 0$ para todo $i = 1,..., n$, y, en consecuencia, $r_s = 1$. Si la discordancia es perfecta, entonces los pares de rangos son: $(1,n)$, $(2,n-1)$,...$(1,n)$ y la suma de las diferencias al cuadrado es igual a $(n^3-n)/3$, lo que implica que: $r_s = -1$.

APÉNDICE 12.B: Distribución del estadístico de Pearson.

La demostración, para el caso general es algo complicada, pero obedece a la misma lógica que en el caso de dos alternativas, $k = 2$, para el que presentamos aquí la demostración. En primer lugar, notemos que, al examinar cada elemento muestral, estamos ante una *distribución multinomial*, en la que dicho elemento puede presentarse en la forma de uno cualquiera de los k sucesos considerados, cada uno con una probabilidad diferente, p_i. Como sabemos, la distribución marginal de cada una de las k frecuencias multinomiales, O_i, sigue una distribución una binomial. Por tanto, su valor esperado es igual al producto de frecuencia total por probabilidad: np_i, y su varianza: $np_i(1-p_i)$. Además, si el tamaño muestral es suficientemente grande, entonces la distribución binomial de la frecuencia n_i puede aproximarse por una Normal, con la esperanza y varianza mencionadas: $O_i \sim N(np_i, np_i(1-p_i))$ y, en consecuencia: $(O_i-np_i)/\sqrt{np_i(1-p_i)} \sim N(0,1)$:

$$\frac{(O_1-np_1)^2}{np_1}+\frac{(O_2-np_2)^2}{np_2} = \frac{(O_1-np_1)^2}{np_1}+\frac{\left[n-O_1-n(1-p_1)\right]^2}{n_1(1-p_1)} =$$

$$\frac{(O_1-np_1)^2(1-p_1)+(-O_1+np_1)^2p_1}{np_1(1-p_1)} = \frac{(O_1-np_1)^2}{np_1(1-p_1)}$$

Ahora bien, con $k = 2$, tenemos:

que es el cuadrado de una $N(0,1)$ y, por tanto, una variable χ_1^2.

3.2.3 Contraste de la mediana

Este contraste trata de evaluar la hipótesis nula de que las medidas de centramiento de las poblaciones de que proceden nuestras muestras son las mismas. Es similar a los anteriores, pero se basa en las medianas y es, por tanto más aconsejable cuando se cree que la mediana es una medida de posición más representativa que la esperanza matemática. Este contraste no exige ninguna hipótesis acerca de la distribución o distribuciones de las que se extrajeron las muestras, excepto que sean de tipo continuo. Comenzamos ordenando todos los datos que integran las distintas muestras de mayor a menor, y obteniendo la mediana de esta muestra agregada. A continuación, calculamos el número de elementos que en cada muestra se hallan por debajo y por encima de dicha mediana global. Bajo la hipótesis nula de homogeneidad poblacional, esperaríamos que aproximadamente la mitad de los elementos de cada submuestra se hallase por encima o por debajo de la mediana global. Disponemos las frecuencias así obtenidas en una tabla de doble entrada, del tipo utilizado en los contrastes de contingencia, y utilizamos precisamente el mismo estadístico que el de dichos contrastes, que sigue nuevamente una distribución chi-cuadrado, al igual que en aquellos análisis.

Ejemplo 4.- Supongamos que, para tratar de evaluar las diferentes capacidades de hombres y mujeres, se somete a un examen lógico a una muestra de 16 estudiantes varones y 14 estudiantes mujeres, obteniendo los siguientes resultados:

	Puntuación															
Hombres	92	76	85	70	84	66	72	68	74	90	84	82	74	70	76	82
Mujeres	84	88	94	90	78	82	64	70	82	74	76	80	78	90		

Al tomar la muestra global se obtiene una mediana de 79, y tenemos 30 elementos, y podemos construir la tabla:

Grupo	Hombres	Mujeres	Total
Nº de observaciones por encima de la mediana	7	8	15
Nº de observaciones por debajo de la mediana	9	6	15
Total	16	14	30

del que podemos calcular las frecuencias teóricas o esperadas E_i:

Grupo	Hombres	Mujeres	Total

Nº de observaciones por encima de la mediana	8	7	15
Nº de observaciones por debajo de la mediana	8	7	15
Total	16	14	30

por lo que el estadístico chi-cuadrado tiene un valor numérico de: $(7\text{-}8)^2+(8\text{-}7)^2+((9\text{-}8)^2+(6\text{-}7)^2 = 4$, que cae dentro del intervalo de dos colas de la variable chi-cuadrado con (2-1)(2-1) = 1 grado de libertad, al 0,05 de significación, que es (0,001;5,02) por lo que no rechazamos la hipótesis nula de homogeneidad de medianas en las poblaciones que generaron las dos muestras. En tal sentido, rechazamos la homogeneidad de ambas poblaciones, habiendo obtenido evidencia que sugiere que el grupo con mayor porcentaje de observaciones por encima de la mediana global (mujeres) tiene una mediana (medida de centramiento) superior al correspondiente a la muestra con un mayor número de observaciones por debajo de la mediana global.

Para contrastar dicha hipótesis, se podría proceder a un contraste de una cola, elaborando la hipótesis alternativa utilizando ya la información que acabamos de reseñar. Es decir, podríamos ahora efectuar un contraste de una cola para la igualdad de proporciones. Sin embargo, en este caso parece que el resultado podría ser debido a un número excesivamente bajo de categorías. Contrastando al 0,01 de significación (99% de confianza), el valor crítico para el contraste de dos colas es de:5,02 (?? 7,88 ??), y no rechazamos la hipótesis nula de homogeneidad de ambas poblaciones.

Para el contraste, suele utilizarse el estadístico chi-cuadrado con la corrección de Yates:

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

que tiene un grado de libertad.

Cuando alguna observación coincide con la mediana, suele omitirse, siempre que no sean muy numerosos, o que los tamaños muestrales n1 y n2 sean suficientemente elevados. Este contraste es muy sencillo de utilizar, aunque tiene poca potencia, especialmente en comparación con el contraste paramétrico comparable. Por ello, es aconsejable que se disponga de tamaños muestrales grandes.

El estadístico H puede utilizarse para contrastar la homogeneidad de medianas en más de dos poblaciones, elaborando un cuadro como el anterior, pero con más de dos columnas, es decir, para más de dos muestras, si bien puede sustituirse por el contraste de Kruskal-Wallis, que parece ser más potente. En tal caso, las frecuencias esperadas de elementos por debajo y por encima de la mediana global en cada muestra pueden aproximarse por la mitad del tamaño muestral respectivo.

<u>Nota 1</u>: Utilizando la propiedad de que, tanto la suma de las frecuencias observadas como las teóricas es igual al tamaño muestral, n, el estadístico chi-cuadrado puede también escribirse:

$$\chi^2 = \sum_{i=1}^{n} \frac{O_i^2}{E_i} - n$$

<u>Nota 2</u>: Hay procedimientos alternativos, generalmente más sencillos, para calcular el valor del estadístico para el contraste de independencia en tablas de contingencia. En el caso de tablas 2x2 dicha expresión es:

$$\chi^2 = \frac{N(a_1 b_2 - a_2 b_1)^2}{(a_1+b_1)(a_2+b_2)(a_1+a_2)(b_1+b_2)} = \frac{N\ \Delta^2}{N_1 N_2 N_A N_B}$$

donde Δ denota el "determinante": $a_1b_2-a_2b_1$, N es el tamaño muestral, N_1 y N_2 son las frecuencias de la distribución marginal de una variable, y N_A, N_B las frecuencias marginales de la otra. Nótese que, por simetría, no importa que variable está representada en cada caso.

Con la corrección de Yates, esta expresión resulta:

$$\chi^2 = \frac{N\left(|a_1 b_2 - a_2 b_1| - \frac{1}{2}N\right)^2}{(a_1+b_1)(a_2+b_2)(a_1+a_2)(b_1+b_2)} = \frac{N\ (|\Delta| - \frac{1}{2}N)^2}{N_1 N_2 N_A N_B}$$

En el caso de tablas de contingencia 2x3, se tiene:

$$\chi^2 = \frac{N}{N_A}\left[\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \frac{a_3^2}{N_3}\right] + \frac{N}{N_b}\left[\frac{b_1^2}{N_1} + \frac{b_2^2}{N_2} + \frac{b_3^2}{N_3}\right] - N$$

donde a_i y b_i, i=1,2,3, denotan los elementos de las dos filas de la tabla de contingencia. Para llegar al cual hemos utilizado la expresión alternativa que antes vimos para el estadístico chi-cuadrado en la nota 1 anterior.