2011

UNED

DISEÑOS DE INVESTIGACIÓN Y ANÁLISIS DE DATOS

[TEMA 9]

Contrastes no paramétricos

Contrastes no paramétricos

Resumen

En este capítulo definiremos lo que entendemos por estadística no paramétrica en contraste con la estadística paramétrica. Aunque la distinción entre ambos conceptos no es clara y unívoca, pretendemos proporcionar suficientes criterios de clasificación como para que el estudiante pueda diferenciarlas sin problemas. A continuación presentamos cinco contrastes de hipótesis no paramétricas: el test de los signos y el test de Wilcoxon para una muestra, el test U de Mann-Whitney-Wilcoxon para dos muestras independientes y el test de Wilcoxon para dos muestras relacionadas. Por último presentamos el test de Kruskal-Wallis para más de dos muestras independientes. El criterio para elegir estos contrastes y no otros ha sido el de presentar los contrastes no paramétricos equivalentes a sus equivalentes paramétricos que han sido tratados previamente en otros capítulos.

Índice

9.1.- Introducción: ¿Qué es la estadística no paramétrica?

Todas las técnicas de contraste de hipótesis y de cálculo de intervalos de confianza que se han discutido en los capítulos anteriores se basan en el supuesto de que las muestras extraídas proceden de una población de puntuaciones que se distribuyen según una función de distribución conocida (*v.g.*, la curva normal). Estos tests se engloban en lo que se conoce como "métodos paramétricos" ya que todos los cálculos matemáticos que se realizan dependen del supuesto de que las puntuaciones proceden de una familia de distribuciones paramétrica particular. El objetivo de estos métodos es estimar un elemento de esta familia paramétrica como el candidato más probable del que proceden nuestras observaciones.

Veamos varios ejemplos de lo que estamos diciendo para intentar aclararlo. Si en un contraste de hipótesis realizamos el supuesto de que las puntuaciones se distribuyen según la curva normal en la población, tendríamos toda una familia o conjunto de funciones de densidad de probabilidad que tienen todos la misma forma (campana de Gauss). Por consiguiente, cada elemento de este conjunto se diferenciaría del resto no por su forma (todas son iguales a este respecto) sino por sus **parámetros** μ y σ (es decir, su media y su desviación típica). La estadística paramétrica funciona realizando supuestos sobre los valores de μ y σ. Por ejemplo, suponiendo la normalidad, puede asumirse además que la varianza de dos poblaciones de puntuaciones son idénticas o puede ponerse a prueba la hipótesis de que la μ de una población tiene un valor concreto. En este sentido, podemos decir que la estadística paramétrica toma en consideración un "espacio paramétrico". En el caso del supuesto de normalidad este espacio paramétrico vendría dado por todos los valores posibles para μ y σ. Este espacio engloba el área que resulta de la intersección entre los intervalos $(-\infty, +\infty)$ y $[0, +\infty)$, respectivamente (véase la Figura 1). Es decir, si el eje de abscisas (eje horizontal) representa las medias y el eje de ordenadas (vertical) representa las desviaciones típicas, y recordando que la desviación típica no puede ser negativa, entonces la media puede ser, teóricamente, cualquier valor entre +Infinito y –Infinito, que se representa como el intervalo abierto $(-\infty, +\infty)$, mientras que la varianza, al tener que ser igual o superior a cero vendría dada por el intervalo semi-cerrado de valores $[0, +\infty)$ en donde se incluye teóricamente el 0. En la Figura 1 se ha representado este "espacio de los parámetros μ y

σ", para este ejemplo particular, como el área en blanco que queda por encima del eje de abscisas. El área inferior en gris (sólo se ha dibujado una pequeña parte de la misma) está prohibida porque representaría σ negativos. En consecuencia, dado un conjunto de datos empíricos, el objetivo de la estadística paramétrica sería encontrar aquellos parámetros (μ y σ) de la curva normal teórica que mejor ajustan los datos. Es por ello que se dice que "buscamos" en el espacio de parámetros (área en blanco de la Figura 1) aquellos valores de μ y σ de los que resulta más probable que se hayan extraído nuestros datos.

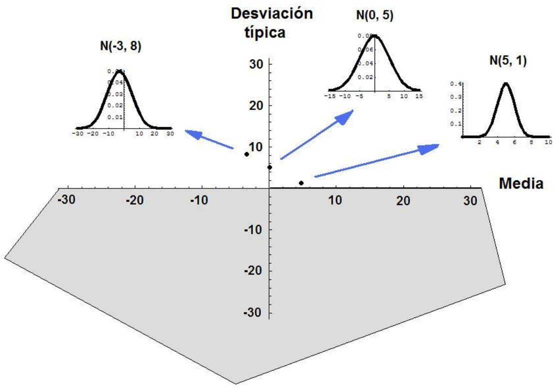


Figura 1: Espacio paramétrico de funciones normales. El eje de abscisas representa la media (μ) y el eje de ordenadas representa la desviación típica (σ) de funciones gaussianas. La mitad horizontal inferior no pertenece al espacio paramétrico debido a que representa valores negativos para la desviación típica. Cada punto del espacio paramétrico factible representa una curva normal concreta. En el gráfico se han representado tres puntos y sus correspondientes curvas.

Es importante señalar que valores como, en el ejemplo del espacio paramétrico para la curva normal, otros estadísticos como la mediana no son parámetros de esa distribución. Los únicos parámetros de la distribución normal son μ y σ ya que son los valores que aparecen en su definición analítica:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Sin embargo, toda función normal tiene mediana (o cualquier estadístico de sesgo, curtosis, etc.). De forma sucinta, la mediana no es un parámetro de la función normal (ni ningún otro estadístico de posición) porque no aparece en la expresión analítica que la define. En la definición analítica de la curva normal sólo aparecen constantes (la unidad, π o el número e), parámetros (μ y σ) y, por supuesto, la variable independiente (x).

Pongamos otro ejemplo de familia paramétrica. Si asumimos que el fenómeno que estamos estudiante se distribuye según una Binomial, entonces tendremos que realizar una búsqueda en el espacio de todas las funciones binomiales. Estas funciones tienen dos parámetros: *n* y *p* (el número de ensayos o veces que se repite un experimento de Bernouilli y la probabilidad de éxito en cada ensayo, respectivamente). Sabemos que *n* tiene que ser un número natural y es igual al número de elementos muestreados. Por su parte, *p* se encuentra en el intervalo cerrado [0, 1], es decir, que puede tomar cualquier valor entre 0 y 1, incluyendo estos valores. Por ello el espacio de búsqueda paramétrico viene dado por los intervalos [0, 1] para cada entero *n*. En la Figura 2 esto se ha representado mediante una línea vertical desde 0 hasta 1 (parámetro *p*) que parte de cada valor entero del eje de abcisas (parámetro *n*). Dado un valor concreto de *n*, el espacio de búsqueda sería el intervalo desde 0 hasta 1. En términos gráficos, las líneas verticales que se muestran en la Figura 2.

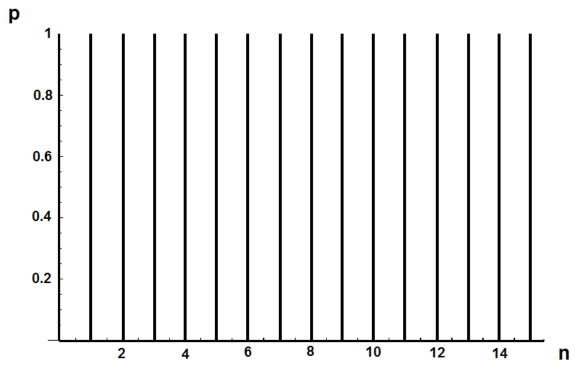


Figura 2: Espacio paramétrico de las funciones binomiales. Para cada valor de n (número de datos), el objetivo de la estadística paramétrica sería encontrar el valor *p* que mejor ajusta los datos. Este valor *p*, como probabilidad que es, se encuentra entre 0 y 1.

De la misma forma que antes, *p* y *n* son parámetros de la distribución binomial porque aparecen en su expresión analítica:

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

En esta expresión podemos ver que, de nuevo, sólo tenemos parámetros (p y n) y la variable independiente x. El valor de q es redundanate ya que se puede expresar como

1-p. En este caso, no existen constantes. Podemos decir que la aparición de una variable en la expresión analítica de una f.d.p., distinta de la variable indendiente, la caracteriza como un parámetro de esa distribución porque define a la distribución. A su vez, los parámetros caracterizan unívocamente a la función.

En definitiva, la denominación de "técnicas paramétricas" procede de la búsqueda de los parámetros subyacentes a unos datos muestrales asumiendo que estos se distribuyen según una función de probabilidad o de densidad de probabilidad concreta (hablaremos genéricamente de f.d.p. o *función de densidad de probabilidad*). Todos los tests paramétricos asumen una determinada forma para la distribución poblacional de la variable dependiente (normal, binomial, etc.) y esta forma depende de unos parámetros, distintos y propios de cada f.d.p.

Pero a veces nos encontramos con datos, poblaciones o situaciones en que no podemos asumir los supuestos subyacentes a los tests paramétricos y necesitamos procedimientos inferenciales cuya validez no dependa de esos supuestos. En este caso se nos hace necesario acudir a otro conjunto de técnicas que no exijan buscar los parámetros de la distribución poblacional de los datos. Es por ello que en contraposición a los anteriores métodos, se los conoce como "métodos no paramétricos". Los contrastes de hipótesis no paramétricos se realizan con datos procedentes de una distribución de probabilidad desconocida y, por tanto, cuya forma no viene especificada en el desarrollo del test.

Sin embargo, aunque la explicación anterior pueda parecer directa y simple, resulta difícil proporcionar una definición precisa del concepto de "técnicas no paramétricas" ya que no existe un consenso absoluto al respecto. La explicación anterior ha tratado aclarar pedagógicamente la diferencia entre ambos tipos de técnicas, incidiendo en la aclaración de los términos utilizados (paramétrico vs. no paramétrico). Pero la literatura técnica diferencia estas técnicas de formas muy diversas. Así, por ejemplo, Ross (2004) las define como aquellos tests que se aplican cuando los datos proceden de una distribución de probabilidad cuya forma no viene especificada, es decir, cuando no podemos **asumir** una forma concreta de la distribución poblacional para los datos (normal, exponencial, binomial, etc.). Es por ello que muchos autores denominan a estos tests como "tests sin supuesto distribucional" (*distribution-free*) en vez de "no paramétricos".

No obstante, sería equivocado afirmar que los tests no paramétricos no realizan ningún tipo de supuestos (*assumption-free*), solo que los supuestos realizados por los tests no paramétricos son menos rígidos que los realizados por los tests paramétricos (Marascuilo y McSweeney, 1977). Es decir, los tests paramétricos **sí** realizan supuestos, pero son supuestos más generales, no relativos a una f.d.p. particular. Por ejemplo, un supuesto muy utilizado en tests no paramétricos es el supuesto de distribución simétrica de la variable dependiente en la población. Debido a que existen otras muchas distribuciones distintas de la normal que también son simétricas, este supuesto no nos obliga a trabajar con una distribución normal (que es simétrica) sino con todo el espacio de distribuciones simétricas. No obstante, sí nos elimina del área de búsqueda todas las distribuciones asimétricas (*v.g.*, la F o la χ^2). Esta es, quizás, una de las razones por las que algunos autores han introducido otros términos que pretenden establecer matizaciones a medio camino entre las técnicas paramétricas y las no paramétricas (*v.g.*, técnicas semi-paramétricas).

Opcional:

En relación a las distinciones anteriores, el estudiante debe diferenciar entre los supuestos realizados por el método de inferencia de los tests (criterio que diferencia las técnicas paramétricas de las no paramétricas) de las distribuciones muestrales utilizadas en los cálculos para el contraste de hipótesis en cualquiera de estas técnicas (tanto paramétricas como no paramétricas). Aunque la estadística no paramétrica no haga referencia entre sus supuestos a una f.d.p. concreta como la distribución poblacional de la variable dependiente, no obstante, los estadísticos que se calculan en la estadística no paramétrica (v.g., U, el número de signos, etc.) sí se distribuirán según una u otra distribución paramétrica concreta. Por ejemplo, cuando veamos el test de los signos comprobaremos que este test no hace ningún supuesto sobre la forma de la distribución poblacional de la variable dependiente (en este sentido es un test no paramétrico) pero, sin embargo, utilizará la distribución binomial (como distribución muestral del estadístico) para realizar los cálculos de los niveles de probabilidad asociados a H_0. La distribución binomial es una distribución paramétrica pero el test de los signos no la utiliza como un supuesto necesario para aplicar el test sino como una herramienta para calcular niveles de probabilidad.

Otros autores que se enfrentan a la tarea de diferenciar entre técnicas paramétricas y no paramétricas hacen hincapié simplemente en que, de forma genérica, estas últimas realizan supuestos menos restrictivos o rígidos que las técnicas paramétricas (Daniel, 1990). Desde esta perspectiva, la distinción entre paramétrico *vs.* no paramétrico no es una distinción cualitativa sino cuantitativa. Hay tests que se acercan más al extremo paramétrico y otros al extremo no paramétrico, existiendo otros en puntos intermedios del continuo. Aunque esta clasificación cuantitativa de los tests es una definición más vaga que las anteriores, resulta muy útil ya que hay técnicas estadísticas entre cuyos objetivos iniciales explícitos se encuentra el realizar los menos supuestos posibles sobre la procedencia de los datos.

Una tercera opinión insiste en que lo que caracteriza a las técnicas no paramétricas es el nivel de medida de los datos. Las técnicas no paramétrica se suelen utilizar cuando las escalas utilizadas para medir la variable dependiente son de tipo nominal u ordinal, o bien cuando las escalas sean de tipo de intervalo/razón pero han sido recodificadas en variables de tipo nominales u ordinales. Esto no quiere decir que no puedan utilizarse técnicas no paramétricas para variables medidas en escalas de intervalo o razón pero la disminución de potencia que esto produciría en relación a las técnicas paramétricas no recomienda esta opción.

Nosotros estamos de acuerdo con Wasserman (2006) cuando subraya que el punto esencial de las técnicas no paramétricas consiste en que los métodos estadísticos desarrollados en este área tratan de mantener los supuestos lo menos restrictivos o rígidos posibles. Hemos de señalar que cuanto menos restrictivos sean los supuestos que se realicen para el método de contraste de hipótesis, más amplias serán las posibilidades que tendremos que contemplar. Esto se comprende fácilmente si se observa que cuando hacemos el supuesto de que la distribución de las puntuaciones poblacionales se distribuye según la curva normal, sólo tendremos que buscar dentro de un espacio paramétrico de valores μ y σ relativamente restringido (véase Figura 1). Sin embargo, si no hacemos restricción alguna en la forma de la distribución de las puntuaciones, como sucede en las técnicas no paramétricas, tenemos que admitir que estamos buscando en un espacio mucho más amplio que incluye no solo funciones normales sino exponenciales, logarítmicas, etc. Este espacio puede llegar a ser infinito por lo que también se conoce a las técnicas no paramétricas como “técnicas paramétricas de dimensionalidad infinita” (*infinite-dimensional parameter* -Arkadi Nemirovski, 2000-).

Esta definición suaviza la opinión de que la denominación de “tests no paramétricos” ha sido desafortunada (Noether, 1991). Lo que sí es cierto es que el concepto de “no paramétrico” no tiene una definición precisa y universalmente aceptada.

9.2.- Aspectos positivos y negativos de los tests no paramétricos

Uno de los aspectos positivos de los tests no paramétricos reside en que pueden aplicarse sin hacer ningún supuesto sobre la forma específica de la distribución poblacional subyacente a la variable dependiente. Esto implica que si existe alguna justificación, teórica o empírica, para asumir una forma paramétrica concreta (*v.g.*, si tenemos información o datos para asumir que la distribución de la variable dependiente poblacional se distribuye según la normal) entonces debería utilizarse el test paramétrico apropiado.

Otra de sus ventajas es que los datos o puntuaciones a las que se aplican estos procedimientos no necesariamente tienen que ser cuantitativos sino que también puede ser categóricos (*v.g.*, variables medidas como Si/No, defectuoso/no defectuoso, etc.) o los datos pueden consistir en puntuaciones de orden (es decir, en vez de puntuaciones numéricas, valores de ordenaciones como primero, segundo, tercero, …).

Como hemos comentado previamente, la mayor parte de los tests no paramétricos dependen de pocos supuestos y estos suelen ser poco rígidos. Cuanto menos supuestos se utilicen en una técnica estadística significa que habrán más conjuntos de datos que puedan analizarse con esa técnica. Esto supone que tenemos menos probabilidad de equivocarnos si utilizamos una técnica no paramétrica que con la correspondiente paramétrica.

Aunque en esta época de fácil acceso a recursos informáticos no tiene mucho sentido considerar la facilidad/dificultad del cálculo como un aspecto positivo o negativo, sí es de observar que los cálculos de los procedimientos no paramétricos suelen ser más rápidos y fáciles que los propios de las técnicas paramétricas. Esto significa que las técnicas no paramétricas descansan en conceptos y métodos matemáticos más fáciles de entender que sus correspondientes paramétricos.

Pero no todo son ventajas en los tests no paramétricos. Existen desventajas y son importantes. La más relevante es que su utilización disminuye la potencia del test, es decir, la probabilidad de aceptar la H_0 cuando esta es falsa. De forma genérica, dados

dos tests equivalentes, uno paramétrico y otro no paramétrico, el test paramétrico es más potente que su equivalente no paramétrico. Debido a que existen procedimientos paramétricos y no paramétricos aplicables a los mismos conjuntos de datos y sólo los supuestos que hagamos sobre la distribución poblacional de los datos nos llevarán a utilizar unos u otros, la decisión puede ser difícil. Solo la mejor o peor ejecución de unos procedimientos en relación a los otros en unas condiciones determinadas (nivel de medida de la variable dependiente, existencia de valores atípicos, supuestos sobre la forma de los datos, etc.) podrán ser los criterios que utilicemos para preferir unos u otros. Los investigadores suelen poner a prueba ambos procedimientos en simulaciones con poblaciones normales u otras más generales. Los resultados de estos estudios han mostrado que los procedimientos no paramétricos, al no utilizar toda la información proporcionada en la muestra, son menos eficientes y necesitan tamaños muestrales mayores para alcanzar la misma potencia que el procedimiento paramétrico correspondiente. No obstante, esta pérdida de eficiencia de los métodos no paramétricos en relación a los paramétricos no suele ser elevada y, por tanto, tampoco la diferencia en el tamaño muestral necesario. Esta pequeña diferencia del tamaño muestral es especialmente saliente cuando las distribuciones subyacentes no son normales pero es más acusada cuando sí se cumple la normalidad de la variable en la población. La conclusión de esta desventaja de los estadísticos no paramétricos consiste en que se deben utilizar los procedimientos paramétricos siempre que los supuestos de los mismos se cumplan. Hacerlo de otro modo sería una pérdida de información.

Siendo a Daniel (1900) recomendamos la utilización de los procedimientos no paramétricos cuando:

1. La hipótesis que se pone a prueba no implica un parámetro de la distribución poblacional. Así, por ejemplo, incluso en el caso de que asumamos una distribución normal, si ponemos a prueba una hipótesis sobre la mediana deberíamos aplicar un test no paramétrico.
2. Los datos se han medido en una escala de nivel inferior al exigido por el procedimiento paramétrico correspondiente. Este caso se puede dar cuando los observaciones solo informan sobre el orden en el que han quedado las unidades de observación (primero, segundo, tercero, …) o solo cuentan el número de unidades que cumplen ciertos requisitos.

3. No se cumplen los supuestos necesarios para utilizar un procedimiento paramétrico. Por ejemplo, cuando la variable muestra un sesgo o una curtosis elevada.

Opcional (no es materia de examen)

Clasificación de los tests no paramétricos

A efectos prácticos se muestra en la Tabla 1 una clasificación no exhaustiva de los tests no paramétricos que consideramos de utilidad para la identificación del test apropiado a cada ocasión. En este capítulo no podemos revisarlos todos ellos por lo que hemos seleccionado los que consideramos más útiles en una primera aproximación al área (marcados en fondo gris). El alumno tendrá que acudir a la bibliografía especializada si necesita utilizar algún otro contraste.

Tipo de problema	Número y tipo de muestras				
	Una muestra	Dos o más muestras		Tres o más muestras	
		Independientes	Dependientes	Independientes	Dependientes
Localización (v.g., mediana)	Test de los signos Test de Wilcoxon	Test de Man-Whitney	Test de los Signos Test de Wilcoxon	Test de Kruskal-Wallis Test de Jonckheere	Test de Friedman
Dispersión		Test de Moses Test de Ansari-Bradley			
Bondad de ajuste	Test χ^2 Test de Kolmogorov-Smirnov Test de Lilliefors	Test de Kolmogorov-Smirnov			
Asociación	Coeficiente de spearman τ de Kendall Otros…	Test de χ^2 independencia		Coeficiente de Kendall Test de χ^2 independencia	
Regresión	Test de Brown-Mood Test de Theil	Test del paralelismo			

Tabla 1

Hemos de citar que, aunque no tengan cabida en la anterior clasificación, si atendemos a la definición de métodos paramétricos a la que nos hemos adherido (Wasserman, 2006), existe una enorme variedad de este tipo de procedimientos que no aparece reflejada en la Tabla 1. El estudiante podrá encontrar como procedimientos no paramétricos técnicas de remuestreo como el *bootstrap* o el *jackniffe*, técnicas de reducción de la dimensionalidad como ICA, técnicas de análisis de la señal como las ondículas (*wavelets*), suavizado de funciones mediante *splines*, etc.

Aunque parezca una trivialidad recordarlo, hemos de advertir al estudiante que paramétrico es un adjetivo que califica al método de análisis, no a los datos. Los datos no son "no paramétricos" sino el método mediante el que se los analiza.

A continuación veremos detalladamente cinco métodos no paramétricos. Los dos primeros se utilizan en el análisis de una muestra (el test de los signos y el test de Wilcoxon para una muestra). Los dos siguientes permiten el análisis de dos muestras relacionadas (test de Wilcoxon para dos muestras relacionadas) y no relacionadas (U de Mann-Whitney-Wilcoxon). Por último se presenta el test de Wilcoxon para más de dos muestras independientes.

Contrastes de hipótesis no paramétricos para una muestra

El test de los signos para una muestra

Supongamos que nos disponemos a recoger un conjunto de *n* datos procedentes de una muestra (*n* será el tamaño muestral definido, entre otros factores, a partir del tamaño del efecto y del procedimiento de análisis a seguir –paramétrico/no paramétrico– como se ha visto anteriormente). Las unidades de observación, es decir, los elementos a los que someteremos a medición, pueden ser personas, colegios, ratas, amebas, etc. dependiendo solamente de nuestra área de investigación. Para cada unidad de observación realizaremos la medición de una variable (*v.g.*, CI en personas, nivel de ruido ambiental en colegios, número de ensayos para aprender una tarea en ratas, movilidad en amebas, etc.). A la variable que midamos la denotaremos simbólicamente por X.

Antes de realizar las mediciones no sabremos los resultados de las mismas por lo que las denotaremos simbólicamente como X_i, indicando X la variable a medir y el subíndice *i* la unidad de observación concreta a la que se refiere aquel valor (*i.e.*, el niño, el colegio, la rata, la ameba, etc.). El conjunto total de observaciones será X_1, X_2, …, X_n. Este conjunto de observaciones procederá de una población de puntuaciones que siguen una función de distribución continua a la que denominaremos F. Pero si no conocemos ni podemos asumir la forma concreta de F no podemos utilizar los tests que se han visto en temas anteriores (tests paramétricos). Tendremos que utilizar, por consiguiente, tests no paramétricos. Esto quiere decir que debemos admitir que F puede adoptar cualquier forma (normal, exponencial, gamma, etc.) y, por consiguiente, el procedimiento estadístico que utilicemos no puede depender de la forma de F. Este procedimiento debe ser válido para cualquier forma que pueda adoptar F en la realidad.

Supongamos que estamos interesados en poner a prueba la hipótesis de que la mediana de F, a la que llamaremos η, es igual a un valor específico al que llamaremos m_0. Es decir, estamos poniendo a prueba el siguiente contraste:

$$H_0: \quad \eta = m_0$$
$$H_1: \quad \eta \neq m_0$$

Si recordamos el concepto de mediana y asumiendo que H_0 es cierta, esto significa que en el valor m_0 (un valor concreto de la variable medida X que queremos contrastar) aproximadamente la mitad de los datos de la muestra deben quedar por debajo del mismo y la mitad restante por encima, excepto variaciones por azar. En otras palabras, m_0 es el percentil 50 de la distribución (la mediana). Simbólicamente esto puede expresarse como $F(m_0) = 0.5$. Y esto sin importar el tipo de distribución concreta que pueda adoptar F en la realidad. Podríamos tener cualquiera de las situaciones que pueden verse en la Figura 3 en el sentido de que nuestros datos podrían provenir de una distribución Normal, Beta, Uniforme, etc.

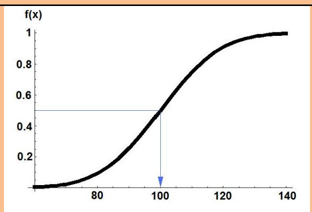
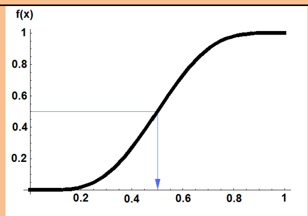
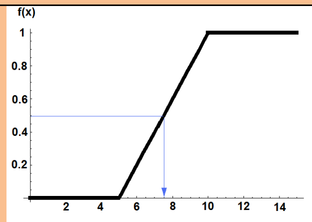
		
Normal(100, 15), Mediana 100 $F_{normal}(100)= 0.5$	Beta(α = 5, λ = 5) Mediana 0.5 $F_{Beta}(0.5)=0.5$	Uniforme(a=5, b=10) Mediana 7.5 $F_{Uniforme}(7.5)= 0.5$

Figura 3: Tres **funciones de distribución** diferentes (Normal con media 100 y desviación típica 15, Beta con $\alpha = 5$ y $\lambda = 5$ y una distribución uniforme entre 5 y 10, respectivamente) en las que se ha señalado explícitamente el valor de la mediana (m_0) para cada caso.

Como hemos dicho, es esencial darse cuenta de que si H_0 es cierta, cada observación será menor que m_0 con probabilidad 0.5 (y superior a m_0 con probabilidad 1- 0.5 = 0.5). Es decir, que si H_0 es cierta la probabilidad de que cada observación sea superior o inferior a m_0 será similar al experimento aleatorio de arrojar una moneda insesgada (no trucada) y determinar si cae cara o cruz. Ya se vio anteriormente que el modelo estadístico que subyace a un proceso dicotómico del tipo cara-cruz, repetido varias veces (*n* ensayos) se modela mediante una variable aleatoria de Bernouilli con parámetro *p* (probabilidad de éxito). En nuestro caso, el valor de *p* deberá ser 0.5 siempre y cuando H_0 sea cierta. En consecuencia, partiendo de los valores de X_i obtenidos empíricamente en el experimento u observación, podemos calcular otra variable aleatoria *I* de Bernouilli según la fórmula:

$$I_i = \begin{cases} 1 & si \quad X_i < m_0 \\ 0 & si \quad X_i \geq m_0 \end{cases}$$

Esto significa que cada valor concreto de X_i se transforma en un 1 si es menor que m_0 (el valor propuesto para la mediana en H_0) o en un 0 si es igual o superior a m_0. En consecuencia I_i se puede considerar una función indicatriz. Si H_0 es cierta, entonces la nueva variable aleatoria I_i será una v.a. de Bernouilli con p = 0,5. Ahora podemos sumar los valores de I_i (a esta suma solo aportarán valor efectiva los I_i que valgan 1 ya que el resto vale 0) para toda la muestra. Llamemos *v* al estadístico resultante de sumar todos los valores de I_i:

$$v = \sum_{i=1}^{n} I_i$$

Este valor *v* (que es igual al número de unos de I_i) se distribuye según la binominal con parámetros *p* y *n* (en donde p valdrá 0.5 si H_0 es cierta y *n* será el número de datos muestrales). Para calcular la probabilidad de obtener un valor igual o superior a *v* asumiendo que H_0 es cierta aplicaremos la binomial bilateral según:

$$2\ P(Binomial(n,\ 0.5) \leq v$$

En esta fórmula, $Binomial(n, 0.5)$ representa la distribución binomial con *n* ensayos y *p* = 0,5 (distribución asumida si H_0 es cierta). De esta binomial nos interesa la probabilidad de que sea menor o igual al estadístico *v* calculado en nuestra muestra, lo cual se expresa analíticamente como $P[Binomial(n, 0.5) \leq v]$. Como estamos trabajando bilateralmente, multiplicamos por dos esta probabilidad. En el caso de trabajar unilateralmente eliminaremos la multiplicación por dos y calcularemos la probabilidad oportuna considerando que la distribución binomial es simétrica.

A veces se utiliza como estadístico el número de signos positivos (o negativos) que se obtiene realizando la diferencia entre cada puntuación original y la mediana planteada en H_0, es decir, $X_i - m_0$. Es por ello que a este test también se le conoce como el test de los signos. Los valores resultado de este cálculo (signos positivos o negativos de $X_i - m_0$) son conceptualmente idénticos al cálculo de la variable aleatoria I_i ya que si X_i es inferior a m_0 entonces el signo es negativo y si X_i es superior a m_0, entonces el signo es positivo. Esto es lo mismo que determinar si X_i es inferior o superior a m_0 y asignar valores de acierto/fracaso o unos y ceros. Sin embargo, nuestra aproximación entronca mejor con la binominal que se estudió en cursos anteriores por lo que la consideramos preferible.

El test de los signos se suele utilizar para poner a prueba hipótesis sobre la mediana aunque podría ampliarse fácilmente para poner a prueba cualquier hipótesis sobre otro percentil razonando a partir de la binomial. Los supuestos para poder aplicar correctamente el estadístico de los signos son:

- Asumimos que la variable de interés es continua. Esto significa que teóricamente no deberían producirse diferencias cuando calculamos $X_i - m_0$. No obstante, como en la práctica se producen, el procedimiento usual es descartar aquellas observaciones que generen estas diferencias nulas y reducir *n* en consecuencia.
- La variable de interés viene medida, como mínimo, en una escala ordinal.
- La muestra se ha extraído aleatoriamente de una población con mediana desconocida.

A continuación presentamos de forma detallada un ejemplo de los cálculos a seguir.

Ejemplo 1:

Sabiendo que la distribución de los tiempos de reacción (TR) tiene una distribución asimétrica negativa, un investigador decide utilizar un test no paramétrico para contrastar la hipótesis de que la mediana en un experimento de TR simple es de 450 ms. Ha sometido a 8 participantes a un experimento de TR simple, calculando para cada uno de ellos la media de TR en 100 ensayos. Los datos obtenidos han sido {X_1= 430 , X_2= 480 , X_3= 510 , X_4= 500, X_5= 390 , X_6= 455, X_7= 440 , X_8= 470, X_9= 475, X_{10}= 465}. El investigador trabaja a un nivel de confianza del 95%.

El investigador desea utilizar un test no paramétrico debido a la asimetría de la distribución de los TR y no en base al nivel de medida de la variable dependiente, que en este caso sería de razón. Los resultados pueden verse en la Tabla 2.

Sujeto	X_i	I_i	(X_i-m_0) (X_i-450)	Signo de las diferencias entre (X_i-m_0)
1	X_1=430	1	-20	-
2	X_2=480	0	30	+

3	X_3=510	0	60	+
4	X_4=500	0	50	+
5	X_5=390	1	-60	-
6	X_6=455	0	5	+
7	X_7=440	1	-10	-
8	X_8=470	0	20	+
9	X_9=475	0	25	+
10	X_{10}=465	0	15	+
		v = 3		

Tabla 2: resultados y cálculos del Ejemplo 1.

En la Tabla 2 vemos en la segunda columna los datos directos (la media del TR simple en 100 ensayos) para los 10 participantes (primera columna). En la tercera columna vemos la función indicatriz, en donde 1 significa que la puntuación X_i es menor que la mediana postulada en H_0 y 0 significa que la puntuación es superior. El sumatorio de todos los valores de la función indicatriz es el estadístico *v*. Con este valor nos preguntamos por la probabilidad asociada al mismo bajo el supuesto de que H_0 es cierta. Para ello, veamos en primer lugar la función de probabilidad binomial con n=10 (el número de participantes) y p = 0.5 (la probabilidad de éxito si H_0 es cierta).

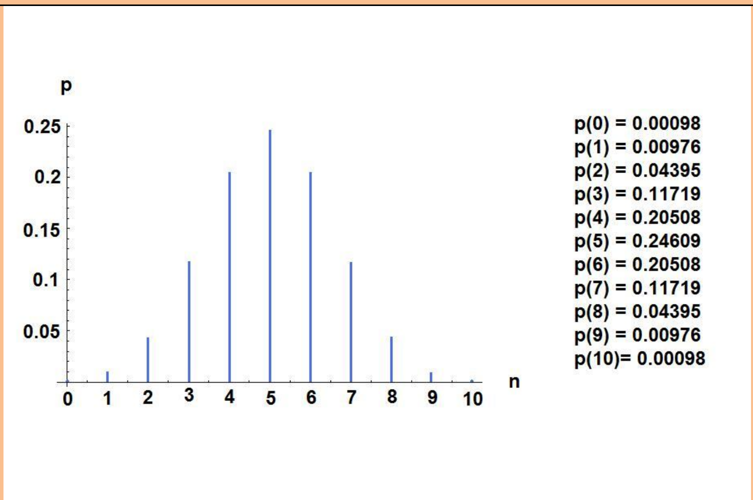


Figura 4: Función de probabilidad de una distribución binomial con n = 10 y p = 0.5.

La probabilidad de obtener un valor del estadístico de contraste *v* menor o igual al obtenido (recordemos que *v* = 3 y que este valor es idéntico al número de signos negativos en la quinta columna de la Tabla 2) viene dado por la suma de los niveles de probabilidad para *n* entre 0 y 3, es decir: $P(V \leq 3) = p(0) + p(1) + p(2) + p(3) = 0.17187$

Ahora bien, como estamos trabajando con una H_0 bilateral estamos asumiendo *a priori* que el estadístico *v* puede ser superior o inferior a la mediana. En el cálculo anterior sólo hemos contemplado las probabilidades de obtener valores de *v* iguales o inferiores a 3, pero tenemos que contemplar también que *v* pueda ser igual o superior a *n-v*, es decir, a 7. Considerando que la distribución binomial es simétrica es fácil ver que para calcular el nivel de probabilidad asociado con el estadístico v = 3 sólo tenemos que multiplicar la p unilateral obtenida por 2. El resultado es de 0.34375. Este valor de p se podría haber calculado también mediante la expresión:

$$\begin{aligned} P(V \leq v) + P(V \geq n - v) &= P(V \leq 3) + P(V \geq 10 - 3) \\ &= p(0) + p(1) + p(2) + p(3) + p(7) + p(8) + p(9) + p(10) = 0.34375 \end{aligned}$$

pero como hemos indicado, la simetría de la distribución binomial nos exime de este cálculo más engorroso. Además, utilizando las tablas de la binomial podemos extraer estas probabilidades directamente.

El último paso a realizar es comparar este valor de probabilidad empírico con el error tipo I que estemos dispuestos a cometer (α). Si utilizamos un α de 0.05 podemos observar que $\alpha < [P(V \leq v) + P(V \geq v)]$ ya que 0.05 < 0.34. Esto nos conduce a no rechazar H_0. Sólo en el caso de que α fuese mayor que la probabilidad de obtener un estadístico *v* igual o más extremo que el calculado rechazaríamos H_0.

La misma conclusión obtendríamos si utilizáramos la comparación entre $P(V \leq 3)$ y $\alpha / 2$. Observamos que la probabilidad de obtener un valor igual o más extremo que *v*=3 por la izquierda de la distribución es mayor que la mitad del error tipo I ($\alpha / 2$). Ambos procedimientos de comparación de las probabilidades empíricas y críticas conducen a no rechazar H_0: no existe evidencia para sospechar que la mediana de los TR sea diferente de 450 ms.

En el caso anterior hemos puesto a prueba el contraste bilateral ($H_0: \eta = m_0;\ H_1: \eta \neq m_0$) lo cual se representa gráficamente mediante un conjunto de valores del estadístico *v* (o el número de signos) compatibles con H_0 y otros valores a derecha e izquierda poco probables si H_0 es cierta. Si deseamos hacer un contraste unilateral izquierdo ($H_0: \eta \geq m_0;\ H_1: \eta < m_0$) estamos proponiendo que la mediana poblacional es superior a un valor concreto m_0. Por consiguiente si H_0 es cierta esto implica que el grueso de datos (X_i) serán superiores a m_0 y la diferencia $X_i - m_0$ será mayoritariamente positiva. A su vez, esto implica que se espera un valor muy bajo de *v* y de signos negativos. En este caso se calcula la probabilidad $P(V \geq n - v)$. Si tenemos información que nos conduzca a realizar un contraste unilateral derecho ($H_0: \eta \leq m_0;\ H_1: \eta > m_0$) el razonamiento anterior se invierte y se calculará $P(V \leq v)$.

Ya se vio en cursos anteriores que la binomial se aproxima a la normal conforme *n* aumenta. Por ello cuando el número de elementos muestrales es lo suficientemente elevado, puede utilizar el estadístico z como aproximación a la binomial mediante la fórmula:

$$Z = \frac{(n-v) - \frac{n}{2}}{\sqrt{n\frac{1}{2}\left(1-\frac{1}{2}\right)}} = \frac{(n-v) - \frac{n}{2}}{\sqrt{n/4}}$$

No obstante, no hay acuerdo preciso acerca del tamaño muestral que garantiza o recomienda la utilización de Z. Así, Daniel (1990) defiende que *n* debería ser igual o superior a 12. Sin embargo Siegel y Castellan (1988) recomiendan que la aproximación se utilice cuando n sea mayor que 25.

No obstante, esta aproximación normal a la binomial tiende a inflar el error tipo I con tamaños muestrales pequeños por lo que se recomienda realizar un ajuste para hacer más conservador este test. El ajuste consiste en sumar o restar una cantidad constante (0.5) al numerador del estadístico Z:

$$Z = \frac{(n-v) \pm 0.5 - \frac{n}{2}}{\sqrt{n\frac{1}{2}\left(1-\frac{1}{2}\right)}} = \frac{(n-v) \pm 0.5 - \frac{n}{2}}{\sqrt{n/4}}$$

Se sumará 0.5 cuando (*n*-*v*) sea inferior a *n*/2. Esto hace que el estadístico *n*-*v* se acerque al valor *n*/2 postulado por H_0, favoreciendo así a la hipótesis nula (por eso se denomina “ajuste conservador”). Por el contrario, cuando (*n*-*v*) sea superior a *n*/2 se restará 0.5 por la misma razón que antes: esta disminución del numerador del test Z favorecerá a H_0 en el sentido de que hará menos probable su rechazo.

El test de Wilcoxon para una muestra

Hemos visto que el test de los signos solo hace uso de la información presente en la muestra acerca de si cada valor muestral observado, X_i, es superior o inferior a la mediana (m_0). Esto se podía reflejar ya fuese en los signos de las diferencias $(X_i - m_0)$ o utilizando una función indicatriz. En cualquier caso, el test de los signos no utiliza la **magnitud** de esta diferencia entre X_i y m_0 en sus cálculos. Por tanto, es una información presente en los datos pero que no utiliza. Frank Wilcoxon (1945, 1949) desarrolló un tests que utiliza ambas informaciones: la dirección o signo de la diferencia y la magnitud de esta diferencia. Este test permite poner a prueba hipótesis sobre la mediana pero, además, pone a prueba la hipótesis de que la distribución sea simétrica (que el conjunto de datos superior a m_0 sea igual al conjunto de datos inferior a m_0).

Los supuestos que utiliza el test de Wilcoxon para una muestra son los siguientes:

- La muestra se ha seleccionado al azar de la población que representa.
- La variable dependiente obtenida de cada unidad de observación es de intervalo o de razón.
- La distribución poblacional subyacente es simétrica. En el caso de que este supuesto no se cumpla, se pude utilizar el test de signos binomial (Daniel, 1990).

Este test se recomienda especialmente como sustituto del t-test cuando se sospeche que el supuesto de normalidad se incumple de forma obvia. En este caso, y a semejanza de lo que hicimos en el test de los signos, se utiliza una función indicatriz ponderada por los rangos (*i*):

$$v = \sum_{i=1}^{n} i \cdot I_i$$

La hipótesis nula bilateral es similar a la del test de los signos: H_0: θ=m_0, H_1: $\theta \neq m_0$. Si H_0 es cierta y recordando que la mediana deja por debajo de sí a la mitad de los datos muestrales, esto se traduce en que la suma de los órdenes de las puntuaciones por encima de la mediana deberá ser aproximadamente igual a la suma de los órdenes de las puntuaciones por debajo de la mediana, es decir, ER+ ≈ ER-. El incumplimiento de esta desigualdad conducirá al rechazo de H_0.

Cálculo de los rangos promediados

Recordemos que el rango R (u orden) de una observación Z se obtiene disponiendo todas las observaciones en orden de la más pequeña a la mayor. A cada observación directa se le asocia su orden, es decir, a la observación más pequeña se le asocia el rango 1, a la siguiente observación el 2 y así sucesivamente. La mayor observación se le proporciona el rango *n*. Este procedimiento para asignar rangos a las observaciones directas asume que no hay empates, es decir, que no ha dos medidas exactamente iguales. Esto sólo podría suceder si las mediciones originales se realizasen con una precisión infinita, lo cual no es el caso en las situaciones reales. Las observaciones reales tienen una precisión finita y, por consiguiente, se pueden producir empates. Podemos pensar en dos estudiantes, Juan y Pedro, a los que se les ha sometido a un experimento de TR y podrían tener un TR medio de 453.5432…ms. y 453.5382…ms. respectivamente. Pero si el cronómetro que se utilice para su medición sólo puede precisar el milisegundo significa que obtendrá para ambos el mismo valor de 453 ms., aunque en realidad tienen diferente puntuación siempre y cuando pudiéramos medirla con una precisión inferior al milisegundo. Este argumento puede ampliarse a cualquier nivel de precisión que podamos alcanzar con un instrumento de medida real. Es por ello que en la realidad y debido a la precisión finita de nuestros instrumentos de medida tendremos empates. En estos casos los rangos iniciales (1, 2, 3, …, *n*) se reemplazan por los rangos promediados.

Ejemplo:

Supongamos que disponemos de los siguientes observaciones de TR: {450, 500, 732, 322, 450}. Si disponemos estas puntuaciones en orden y les asignamos a cada una el número de orden tendríamos:

Obsérvese cómo, al haber dos puntuaciones iguales a 450, inicialmente se le ha proporcionado a una de ellas el rango 2 y a la otra el rango 3 (no hay forma de identificar cual es cual, así que da igual cuál tenga el rango 2 y cuál el 3). Pero como no tiene sentido proporcionar a una única puntuación directa (450) dos rangos distintos (el 2 y el 3), se les proporciona un único rango (2.5) que es la media entre los rangos otorgados inicialmente. Y si tuviéramos tres puntuaciones empatadas, el procedimiento sería similar. Por ejemplo, asumamos que las puntuaciones originales hubiesen sido {322, 450, 450, 500, 450}

Puntuación directa	322	450	450	450	500
Rango u orden	1	2	3	4	5
Rango promediado	1	(2+3+4)/3=3	(2+3+4)/3=3	(2+3+4)/3=3	5

El procedimiento calcula primero la diferencia entre cada puntuación directa X_i y la mediana postulada en H_0. De esta forma obtendremos un conjunto de *n* puntuaciones diferencia $Y_i = X_i - m_0$. El signo de Y_i dependerá de que X_i sea mayor (signo positivo) o menor (signo negativo) que m_0. Para cada puntuación Y_i registraremos también su orden o rango utilizando los siguientes criterios:

- Ordenamos los valores absolutos de los Y_i, es decir, sin considerar el signo de Y_i.
- Eliminamos cualquier Y_i que valga 0 porque coincide con la m_0 lo que debería ser imposible según el supuesto de distribución subyacente F continua.
- Cuando hayan valores de Y_i que en término absoluto ($|Y_i|$) se encuentren empatados se proporcionará a cada valor empatado el promedio de los órdenes. A diferencia de otros tests que utilizan rangos, el test de Wilcoxon exige que se asigne el orden 1 a la puntuación con el valor absoluto más bajo y el rango mayor (n) a la puntuación diferencia con el valor absoluto más alto.
- A continuación se suman los valores absolutos de las puntuaciones diferencia que tienen signos positivos para producir ER+ y los valores absolutos de las puntuaciones diferencia que tienen signos negativos para producir ER-.

Si H_0 es cierta, es de esperar que los valores ER+ y ER- sean similares entre sí y no muy diferentes a $E[\sum_{i=1}^{n} i \cdot I_i]$ o valor esperado de la función indicatriz. Bajo el supuesto anterior entonces la media debe valer $\frac{n(n+1)}{4}$ y la varianza XXX.

En la Figura 5a podemos ver el conjunto de datos {3, 5, 7, 9.5, 12, 15, 22} representados por puntos sobre el eje X. Si quisiéramos poner a prueba la hipótesis de que la mediana vale 8 en la población de la que se han extraído estos datos (línea vertical) y calculáramos los rangos a partir de las observaciones directas con signo siempre obtendríamos que los valores inferiores a la mediana obtendrían los rangos inferiores (1, 2 y 3 en este caso) mientras que las observaciones superiores a la mediana siempre tendrían rangos elevados (4, 5, 6 y 7 en el ejemplo de la Figura 5). La suma de estos rangos no nos daría información acerca de H_0. Sin embargo, al calcular la distancia en valor absoluto (sin considerar el signo) observemos como los rangos se van alternando a derecha e izquierda. Ahora sí es razonable esperar que la suma de los rangos inferiores a m_0 sea aproximadamente igual a la suma de rangos superiores a m_0 (si H_0 es cierta).

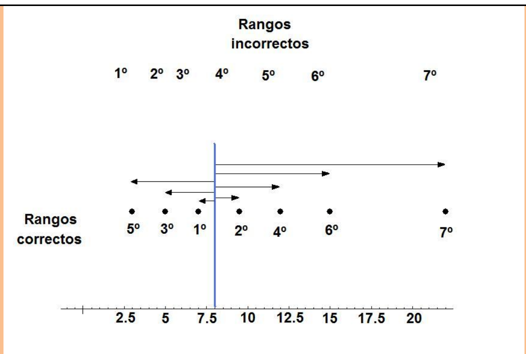


Figura 5a

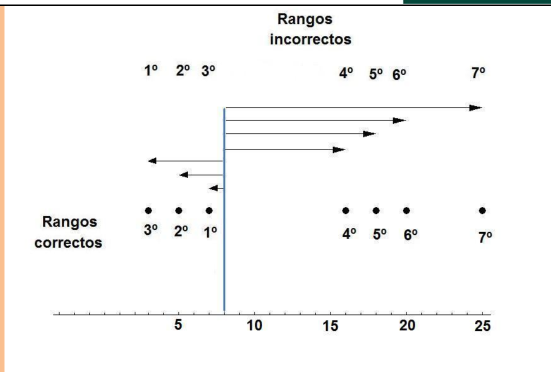


Figura 5b

Observemos en la Figura 5b que los datos han cambiado. Aunque la situación es cualitativamente similar a la Figura 5a (3 datos a izquierda de m_0 y 4 a la derecha) ahora los rangos correctos han cambiado totalmente porque al utilizar la distancia en valor absoluto de cada puntuación a m0 ahora los rangos inferiores quedan todos a mano izquierda de m_0 y los rangos superiores a mano derecha. Ahora el sumatorio de estos rangos produce valores totalmente diferentes para los rangos superiores e inferiores, lo cual es muy improbable según H_0.

Demostración de la igualdad $E[\sum_{i=1}^{n} i \cdot I_i] = \frac{n(n+1)}{4}$

Cuando H_0 es cierta, podemos calcular la media del estadístico ER- observando que los valores de Y_i deben situarse de forma simétrica en relación a 0. Esto significa que es igual de probable que Y_i sea positiva que negativa (para cualquier valor de *i*) debido a que la probabilidad de que I_i valga 1 o 0 es 0.5 (siempre asumiendo que H_0 es cierta). La media de la distribución vendrá dada por:

$$E[\sum_{i=1}^{n} i \cdot I_i] = \sum_{i=1}^{n} i \cdot I_i \cdot p_i$$

Bajo el supuesto de H_0 es cierta, todas las p_i son iguales a 0.5. Por consiguiente

$$E[\sum_{i=1}^{n} i \cdot I_i] = \sum_{i=1}^{n} i \cdot I_i \cdot \frac{1}{2}$$

Considerando que I_i valdrá 0 o 1 según la función indicatriz (y en el caso de que valga 0 eliminaría este término del sumatorio) podemos quedarnos con:

$$\sum_{i=1}^{n} i \cdot \frac{1}{2}$$

Podemos sacar la constante ½ fuera del sumatorio:

$\frac{1}{2}\sum_{i=1}^{n} i$ Y sabiendo que el sumatorio de los n primeros números naturales vale $\frac{n(n+1)}{2}$ entonces

$$\frac{1}{2}\sum_{i=1}^{n} i = \frac{1}{2}\frac{n(n+1)}{2} = \frac{n(n+1)}{4}$$

Nota: esta demostración no es materia de examen.

La pregunta a responder es si la diferencia entre ER+ y ER- es lo suficientemente elevada como para sospechar que H_0 no es cierta. Cuando estos valores sean equivalentes ambos tendrán un valor de n(n+1)/4. Por el contrario, si el valor de ER+ es significativamente mayor que el valor de ER- nos señala que existe una alta probabilidad de que la muestra se derive de una población con una mediana mayor que el valor hipotetizado en la población (y a la inversa).

La Tabla de la T de Wilcoxon se encuentra calculada en función de los valores inferiores que pueda adoptar cualquiera de las sumas de rangos. Por ello se adoptará el valor más pequeño de las dos sumas de rangos (ER+ y ER-) como el estadístico de Wilcoxon, al que llamaremos T. Para poder interpretar este valor es necesario calcular la probabilidad de obtener un valor igual o tan extremo como el obtenido. Si esta probabilidad (*p*) es inferior al nivel de significación α rechazaremos H_0. Existen diferentes formas de realizar este paso. La más directa sería utilizar un software que proporcionara el valor *p* exacto. Otra sería el cálculo manual del mismo (véase el ejemplo desarrollado en Ross -2004-) pero es bastante costoso cuando *n* se incrementa mínimamente. Por ello, lo más fácil es utilizar las tablas de valores críticos de Wilcoxon.

Para obtener la significación (rechazar H_0), el valor muestral T debe ser igual o menor que el valor tabulado para un nivel de significación especificado.

Ejemplo 2:

Una muestra aleatoria de estudiantes fueron entrevistados para determinar el número real de horas que pasaban estudiando para el examen final. Los resultados fueron

{25, 12, 10, 15, 20, 17, 30, 125, 3}

Si estudios anteriores han mostrado que la mediana del número de horas que se estudia es de 15 horas ¿apoya la evidencia esta observación? Trabaje a un α = 0.05.

El gráfico de estos datos aparece en la Figura 6.

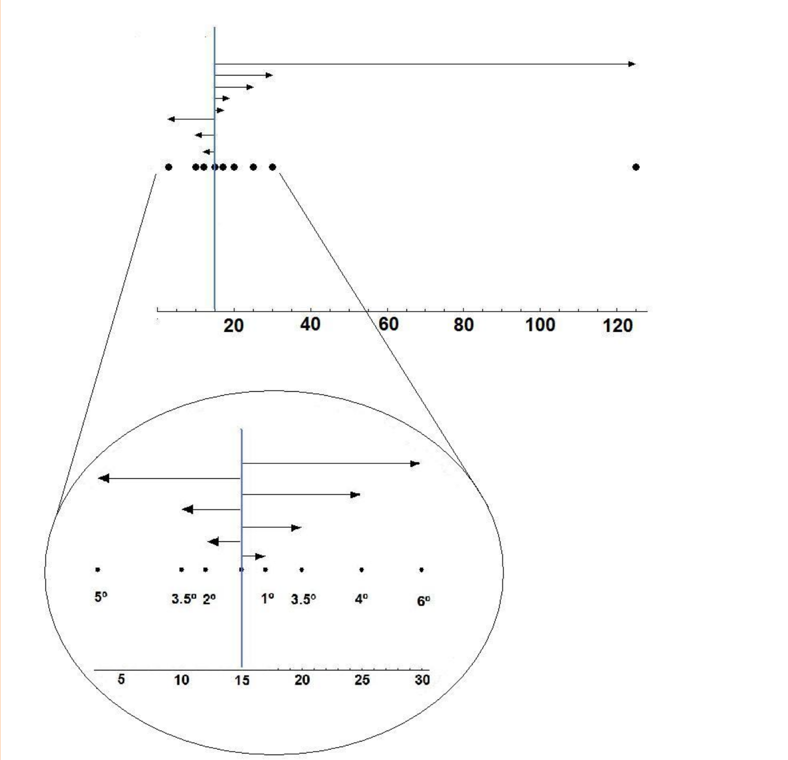


Figura 6: representación gráfica de los datos del ejemplo 2 con ampliación de la zona central para que puedan identificarse los rangos del grueso de datos.

Gráficamente puede verse varios aspectos interesantes de los datos. En primer lugar, tenemos un dato muy alejado del grueso (un *outlier*). Esta puede ser una de las razones por las que se ha adoptado por un test no paramétrico (si hubiéramos utilizado un contraste de la media, este valor extremo habría desplazado hacia la derecha la estimación de la media). En segundo lugar, observemos que tenemos un dato que coincide con la mediana (m_0=15) por lo que debemos eliminarlo pasando de tener 9 datos a 8 (n = 8). Además, tenemos dos datos que están a la misma distancia de m_0 (10 y 20) por lo que tendremos que utilizar el cálculo de los rangos ponderados para este par de datos. Además, tenemos solo 3 datos a la izquierda de m_0 y 5 a la derecha con los rangos alternando a derecha e izquierda. Intuitivamente no tenemos una base firme para sospechar que H_0 sea cierta o falsa. En otros casos, la simple gráfica nos puede dar indicaciones muy útiles.

Veamos analíticamente los cálculos a realizar (véase la Tabla 3) ordenando previamente los datos.

***i* (índice del participante)**	**X_i**	$\lvert X_i - m_0\rvert$	**Rango**	**Rango con signo**
1	3	$\lvert 3-15\rvert = 12$	6	-6

2	10	$\|10-15\|=5$	(3+4)/2=3.5	-3.5
3	12	$\|12-15\|=3$	2	-2
4	15	$\|15-15\|=0$	Eliminado	Eliminado
5	17	$\|17-15\|=2$	1	1
6	20	$\|20-15\|=5$	(3+4)/2=3.5	3.5
7	25	$\|25-15\|=10$	5	5
8	30	$\|30-15\|=15$	7	7
9	125	$\|125-15\|=110$	8	8

Tabla 3

Los estadísticos valen R- = 11.5 y R+= 24.5 (obsérvese que el signo negativo de R- se ha utilizado simplemente como señal para sumar estos valores pero con signo positivo, es decir, en vez de sumar (-6)+(-3.5)+(-2) sumamos 6+3.5+2 ya que el signo negativo solo es un indicador de que estos rangos van juntos en un único sumatorio). Utilizaremos como estadístico T el valor de R- ya que es el más pequeño de ambos. Luego T = 11.5. Ahora debemos acudir a las tablas de Wilcoxon para un contraste bilateral con $\alpha = 0.05$ para extraer los valores críticos. Para n = 8 en un contraste bilateral el valor crítico $T_c = 3$. Debido a que la H_0 solo puede rechazarse si $T \leq 3$, concluimos que no hay evidencia para cuestionar H_0.

Contrastes de hipótesis no paramétricas para dos muestras

Tanto el test de los signos como el test de Wilcoxon se han aplicado al caso de una única muestra. Los casos más interesantes surgen, sin embargo, cuando tenemos dos o más muestras ya que entonces podemos realizar comparaciones entre grupos o condiciones distintas. A continuación estudiamos dos tests, el test de Wilcoxon para dos muestras y el test WMM, que se pueden utilizar cuando tenemos dos muestras. La diferencia entre los mismos radica en que el test de Wilcoxon asume que las muestras están relacionadas mientras que en el test WMM estas deben ser independientes.

Dos muestras relacionadas: el test de Wilcoxon

Cuando se incumplan una o más de los supuestos del t-test para dos muestras relacionadas se podrá utilizar como procedimiento alternativo el test de Wilcoxon para dos muestras **relacionadas** debido a que sus supuestos son menos estrictos. Debido a su similitud con el test de Wilcoxon para una muestra, lo presentaremos al hilo de un ejemplo. Supongamos que un investigador está interesado en el efecto del ruido ambiental sobre la comprensión lectora. Si dispone de seis estudiantes podría asignar tres de ellos aleatoriamente a una condición sin ruido y los tres restantes a una condición con ruido (la variable independiente "condición" se manipula de manera intersujetos). Los resultados de este experimento podrían analizarse mediante el test WMM. Sin embargo, en esta situación existe otro diseño experimental más adecuado considerando los pocos participantes de que se dispone. En vez de dividir a los seis sujetos en dos grupos podría hacer que **cada** sujeto pasara por ambas condiciones, una con ruido y otra sin ruido. Las condiciones se contrabalancearán entre sujetos para evitar cualquier tipo de sesgo. Es decir, el primer sujeto puede pasar primero la condición sin ruido y luego la condición con ruido, el segundo sujeto lo hará a la inversa (primero con ruido y después sin ruido), el tercero lo hará como el primero y así sucesivamente. Obtendría de esta forma 6 puntuaciones para cada condición, dos puntuaciones por sujeto.

Las puntuaciones obtenidas en los dos tests de comprensión lectora para cada sujeto (a mayor puntuación, mejor comprensión) vienen reflejadas en las columnas segunda y tercera de la Tabla 4:

Sujeto	**Sin ruido** X_1	**Con ruido** X_2	**Diferencia** $D=X_1-X_2$	**Rango de** $\lvert D \rvert$	**Rango con signo**
1	76	60	16	4	4
2	89	49	40	5	5

3	65	75	-10	2	-2
4	95	91	4	1	1
5	98	22	76	6	6
6	35	50	-15	3	-3

$$\sum R_{-} = 5$$

$$\sum R_{+} = 16$$

Tabla 4

La cuestión es como analizar estos datos considerando que cada par de puntuaciones no son independientes ya que comparten una serie de variables (*v.g.*, de personalidad, fisiológicas, etc.). Si un sujeto es muy bueno en comprensión lectora, las puntuaciones tenderán a ser altas en ambas condiciones (con y sin ruido). Si un sujeto es malo en comprensión lectura, las puntuaciones tenderán a ser bajas en ambas condiciones. Esto significa que las puntuaciones no son independientes y debemos tratarlas como pares de observaciones con algo en común (en este caso el sujeto). Es por ello que en vez de basar nuestro análisis en las puntuaciones directas, el test W de Wilcoxon utiliza las **diferencias** entre cada par de puntuaciones observadas (cuarta columna de la tabla 4). Observamos en el ejemplo que la mayor parte de las diferencias son positivas. Como hemos realizado la diferencia en el sentido "sin ruido – con ruido", estas diferencias positivas parecen indicar que la comprensión lectora es superior en situaciones de silencio que en situaciones de ruido. Pero necesitamos establecer criterios precisos para llegar a esa conclusión ya que en este experimento hemos manipulado una única variable independiente (condición de ruido) pero hay otras variables que suponemos han actuado al azar (comprensión lectora previa, conocimientos, cansancio, etc.). Debemos evaluar si estos resultados se podrían haber producido por azar o si, por el contrario, esto es muy improbable y en este caso concluiríamos que la condición ha tenido efecto sobre la comprensión lectora.

Para calcular el estadístico W de Wilcoxon para dos muestras dependientes necesitamos utilizar los rangos de las diferencias **absolutas** de las puntuaciones (utilizando los rangos promediados en el caso de que hubiesen empates y eliminando cualquier diferencia nula disminuyendo correspondientemente el valor de *n*).

Cuando la hipótesis nula de igualdad de efectos del tratamiento es cierta se espera que la mediana de las puntuaciones diferencia valga 0, es decir, que la mitad de las puntuaciones diferencia sean negativas y la mitad restante sean positivas, siendo el valor 0 el P_{50} de estas puntuaciones diferencia ($H_0: \theta_D = 0$). Otra manera de formular la hipótesis es afirmar que si H_0 es cierta, la distribución de las diferencias será simétrica con el centro de simetría θ_D en cero. Si, por el contrario, los dos tratamientos tienen diferentes efectos, las diferencias tendrán una mediana distinta de 0 ($H_1: \theta_D \neq 0$). Al calcular las diferencias entre cada par de puntuaciones hemos convertido nuestro problema en una hipótesis de que la mediana vale 0 ($\theta_D = 0$) en contraposición a que sea superior ($\theta_D > 0$) o inferior ($\theta_D < 0$) a cero (en un contraste bilateral). En consecuencia podemos utilizar bien el test de los signos o el test de Wilcoxon de una muestra para evaluar esta hipótesis sin más supuestos sobre la distribución de las diferencias. Utilizaremos el test de Wilcoxon por ser más completo que el test de los signos.

Para realizarlo las diferencias **en valor absoluto** se ordenan desde 1 hasta *n* (columna quinta de la Tabla 4). A continuación se asigna a cada orden el signo de la diferencia original que aparecía en la columna cuarta de la Tabla 4 (el resultado puede verse en la columna sexta de la Tabla 3). Por último se calcula la suma de los rangos con signo positivo ($\sum R_+ = 16$) y con signo negativo ($\sum R_- = 5$).

Si H_0 es cierta (la mediana de las diferencias poblacionales es cero) esperaríamos encontrar entre los rangos altos tantos signos positivos como signos negativos (lo mismo sucedería entre los rangos bajos). Es por ello que la suma de rangos con signo positivo y con signo negativo debería ser similar. Sin embargo, si observamos un alejamiento suficientemente elevado de esta expectativa en los datos se dudaría de la veracidad de H_0.

De la misma que hicimos en el test de Wilcoxon para una muestra, el estadístico W de Wilcoxon para dos muestras dependientes se calcula como el valor más pequeño de los dos sumatorios ($\sum R_-$ o $\sum R_+$). En nuestro caso W = 5 (el mínimo de 16 y 5). Trabajando a un α de 0.05 extraemos de la Tabla de valores críticos de Wilcoxon con *n* = 6 y en un contraste bilateral un valor de $T_{crítico} = 0$. Debido a que $W > T_{crítico}$ no podemos rechazar H_0.

En el caso de que tengamos más de 30 sujetos, no podríamos utilizar la Tabla del estadístico de Wilcoxon para obtener los valores críticos. En este caso y de forma similar a como hicimos en el test de Wilcoxon para una muestra, la Ley de los Grandes Números nos permite utilizar una aproximación para tamaños muestrales superiores a 20 utilizando la distribución normal tipificada z:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Dos muestras relacionadas: test U de Mann-Whitney-Wilcoxon (MWW)

Desarrollado independientemente por Wilcoxon (1945) y por Mann y Whitney (1947) es un test equivalente al test de Wilcoxon para muestras **no relacionadas** o independientes. En él se pone a prueba la hipótesis acerca de si las medianas de las poblaciones a partir de las que se ha extraído cada muestra son diferentes.

Los supuestos necesarios para utilizar este tests son los siguientes:

- La variable original observada es una variable aleatoria continua. Este supuesto se incumple frecuentemente empleándose este test para variables dependientes que representan variables aleatorias discretas.
- La escala de medida de la variable dependiente empleada es, al menos, de nivel ordinal.
- Las distribuciones subyacentes a partir de las que se han extraído las muestras tienen la misma forma (aunque esta no tiene porqué ser normal). El supuesto de igual forma (Maxwell y Delaney, 1990) conlleva también el supuesto de homocedasticidad. No obstante, este test no se muestra tan afectado por la violación del supuesto de homocedasticidad como lo es el test paramétrico correspondiente (*t*-test para dos muestras independientes).

El test MWW se utiliza frecuentemente para reducir o eliminar el impacto de los valores atípicos (*outliers*) ya que utiliza los valores ordinales de los datos y no las observaciones originales. Las hipótesis que se ponen a prueba pueden ser bilaterales o unilaterales (dependiendo de qué mediana se considere más elevada en H_0, la del grupo 1 o la del grupo 2):

Bilateral	Unilateral izquierda	Unilateral derecha
$H_0: M_1 = M_2$ $H_1: M_1 \neq M_2$	$H_0: M_1 \geq M_2$ $H_1: M_1 < M_2$	$H_0: M_1 \leq M_2$ $H_1: M_1 > M_2$

Para calcular el estadístico del test MWW debemos combinar las dos muestras y transformamos las observaciones directas en rangos (promediados si hiciera falta). A continuación sumamos los rangos pertenecientes a cada muestra por separado, obteniendo dos sumatorios S_1 y S_2:

$$S_1 = \sum_{i=1}^{n_1} R_i, \qquad S_2 = \sum_{i=1}^{n_2} R_i$$

La lógica del estadístico MWW es que si la mediana de la población de la que se ha extraído la primera muestra es inferior a la mediana de la segunda población (de la que se ha extraído la segunda muestra) deberíamos esperar que la suma de los órdenes pertenecientes a la primera muestra fuese inferior a la suma de los órdenes pertenecientes a la segunda muestra, es decir, que $S_1 < S_2$ (si la inversa fuera cierta, esperaríamos encontrar $S_1 > S_2$). Recordando que la suma de los rangos de ambas muestras debe proporcionar:

$\sum_{i=1}^{n_1} R_i + \sum_{j=1}^{n_2} R_j = \sum_{k=1}^{N} R_k = 1 + 2 + 3 + ... + N = \frac{1}{2}N(N+1)$ se sigue que $S_1 + S_2 = \frac{1}{2}N(N+1)$. Esta fórmula nos permite calcular S_1 a partir de S_2 (o a la inversa). Si consideramos las muestras individualmente, podemos razonar que el valor teórico que debería alcanzar la suma de rangos sólo depende del número de datos de cada muestra y valdría $\frac{1}{2}n_1(n_1+1)$ para la primera muestra y $\frac{1}{2}n_2(n_2+1)$ para la segunda muestra. Si restamos la suma de rangos S_i realmente obtenida para cada muestra en el ordenamiento global (es decir, en el ordenamiento de los datos sin considerar la muestra) del valor teórico de este sumatorio considerando cada muestra individual, obtendremos los estadísticos U_1 y U_2 que nos servirán para realizar la inferencia:

$$U_1 = n_1 \cdot n_2 + \frac{1}{2}n_1(n_1+1) - S_1$$

$$U_2 = n_1 \cdot n_2 + \frac{1}{2}n_2(n_2+1) - S_2$$

Tanto U_1 como U_2 deben ser valores positivos y, como verificación de la corrección de los cálculos realizados se debe verificar también que $n_1 \cdot n_2 = U_1 + U_2$. Si realizáramos simplemente la diferencia entre $\frac{1}{2}n_i(n_i+1)$ y S_i podríamos obtener valores negativos. Para evitar esta incomodidad en el cálculo se les suma a ambos estadísticos (U_1 y U_2) una constante que depende del número de datos total existente ($n_i \cdot n_j$). La Tabla del apéndice se encuentra tabulada con esta consideración

Como estadístico U elegimos el valor inferior de U_1 o U_2. La Tabla U de Mann-Whitney-Wilcoxon del apéndice expresa los valores críticos U en función del número de sujetos de cada grupo, del nivel de confianza y del tipo de contraste (bilateral o unilateral). La significación se alcanza si el valor U es igual o menor que el valor crítico extraído de la tabla al nivel de significación especificado.

Cuando n_1 o n_2 son superiores a 20 podemos utilizar el Teorema del Límite Central para demostrar que el siguiente estadístico se distribuye según una *z*:

$$z = \frac{U_i - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}}$$

El término ${n_1 \cdot n_2}/{2}$ representa el valor medio esperado de U si H_0 es cierta, es decir, si realmente los dos grupos tienen idéntica mediana. El denominador representa el error típico de la distribución muestral esperada del estadístico U_i. En la aproximación normal se ha puesto U_i ya que, en este caso, podemos utilizar U_1 o U_2 para realizar el contraste.

Dos casos especiales de la aproximación de la normal se producen si el valor *z* obtenida está cerca de los valores críticos tabulados (corrección por continuidad) y/o si se producen empates (corrección por empates). En el primer caso y debido a que el estadístico U es discontinuo ya que es la suma de rangos que, en definitiva, son números naturales mientras que la distribución muestral del estadístico z es continua, Siegel y Castellan (1988) emplean lo que ellos llaman "corrección por continuidad" de la aproximación normal al test MWW. Esta corrección consiste en modificar ligeramente la fórmula anterior:

$$z = \frac{\left|U_i - \frac{n_1 \cdot n_2}{2}\right| - 0.5}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}}$$

La corrección por empates se produce en el caso de que hayamos calculado rangos promediados. En este caso se introduce una corrección en el denominador:

$$z = \frac{U_i - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 \cdot n_2 \left[\sum_{i=1}^{s} (t_i^3 - t_i)\right]}{12(n_1 + n_2)(n_1 + n_2 - 1)}}}$$

Siendo *s* el número de grupos empatados y *t* el número de datos empatados en cada grupo.

Ejemplo 4:

Un equipo de psicólogos evolutivos ha estudiado como adquieren los niños la capacidad para empatizar con otros. Para este fin han desarrollado un tests para medir la empatía en niños pequeños. Los estudios preliminares han mostrado que los chicos son más lentos en desarrollar esta capacidad y, además, la desarrollan en menor cuantía que las chicas. En uno de estos estudios se eligió a un grupo de niños al azar y las puntuaciones obtenidas fueron las siguientes:

Chicos = {25, 6, 1, 11, 18, 20, 16, 17, 5, 29, 17, 20, 6, 13, 16}; $n_1 = 14$.

Chicas = {25, 19, 12, 23, 10, 27, 3, 20, 19, 22, 10, 17, 19, 15, 18, 27, 28 29, 11, 20};

$n_2 = 19$.

Determine si estos resultados apoyan la afirmación de que los niños tienen menor empatía que las chicas (α = 0.05).

En primer lugar debemos observar que la hipótesis alternativa está formulada en términos direccionales: los niños tienen menor empatía que las chicas, es decir: $H_1 : \eta_o < \eta_a$ y por tanto la hipótesis nula es $H_0 : \eta_o \geq \eta_a$. Los cálculos a realizar se presentan en forma tabular (véase Tabla 5). La primera fila representa los datos originales (el fondo de gris de esta primera fila y de la tercera representa un dato de la muestra de chicos; un fondo blanco es un dato de la muestra de chicas), la segunda fila representa los rangos sin ajustar (el fondo en gris en estas filas representa rangos que hay que promediar):

Empatía	1	3	5	6	6	10	10	11	12	13	15
Rango sin ajustar	1	2	3	4	5	6	7	8	9	10	11
Rango ajustado	1	2	3	4.5	4.5	6.5	6.5	8	9	10	11

Empatía	16	16	17	17	17	18	18	19	19	20	20
Rango sin ajustar	12	13	14	15	16	17	18	19	20	21	22
Rango ajustado	12.5	12.5	15	15	15	17.5	17.5	19.5	19.5	22.5	22.5

Empatía	20	20	22	23	25	25	27	27	28	29	29
Rango sin ajustar	23	24	25	26	27	28	29	30	31	32	33
Rango ajustado	22.5	22.5	25	26	27.5	27.5	29.5	29.5	31	32.5	32.5

Tabla 5

El sumatorio de los rangos para el grupo de chicos produce

S_1=1+3+4.5+4.5+10+12.5+12.5+15+15+17.5+22.5+22.5+27.5+32.5 = 200.5

Y para el grupo de chicas

S_2=2+6.5+6.5+8+9+11+15+17.5+19.5+19.5+22.5+22.5+25+26+27.5+29.5+29.5+31+32.5 = 360.5

Calculamos los estadísticos U_i:

$$U_1 = n_1 \cdot n_2 + \frac{1}{2} n_1 (n_1 + 1) - S_1 = 14 \cdot 19 + \frac{1}{2} 14(15) - 200.5 = 170.5$$

$$U_2 = n_1 \cdot n_2 + \frac{1}{2} n_2 (n_2 + 1) - S_2 = 14 \cdot 19 + \frac{1}{2} \cdot 19 \cdot 20 - 360.5 = 95.5$$

Verificamos los cálculos

$$n_1 \cdot n_2 = U_1 + U_2 \qquad \Rightarrow \qquad 14 \cdot 19 = 170.5 + 95.5 = 266$$

El mínimo de U_1 y U_2 es 95.5. El valor crítico según la tabla de WMM en un contraste unilateral con α = 0.05 con n_1= 14 y n_2 = 19 es 87. Como 95.5 > 87 no podemos rechazar H_0. No obstante, obsérvese que el valor U se encuentra relativamente cercano a 87. El investigador haría bien en incrementar el tamaño muestral para incrementar la potencia o en revisar otros trabajos similares al suyo haciendo un meta-análisis de todas las investigaciones que hayan tratado el mismo tema.

Aunque no tenemos suficientes sujetos podemos realizar este mismo contraste utilizando z lo realizaremos a modo de ejemplo.

$$z = \frac{\left| U_i - \frac{n_1 \cdot n_2}{2} \right| - 0.5}{\sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}}} = \frac{\left| 95.5 - \frac{14 \cdot 19}{2} \right| - 0.5}{\sqrt{\frac{14 \cdot 19 (14 + 19 + 1)}{12}}} = 1.35$$

Según las tablas de la curva normal, para obtener la significación estadística con un α = 0.05 unilateral deberíamos alcanzar una z crítica de 1.65, tal y como se representa en la Figura 7.

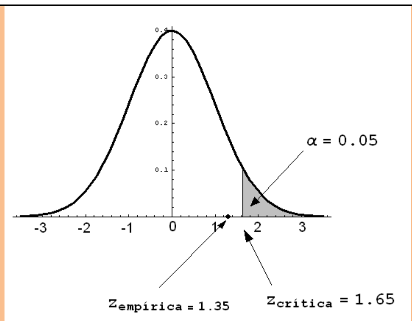


Figura 7: representación gráfica de los resultados del ejemplo 4.

Pero, como además, se han producido empates, podemos utilizar la puntuación z con corrección por empates. Para lograrlo debemos conocer el número de empates (s) y el número de datos empatados en cada uno de los empates. Disponemos de 8 grupos de empates, luego s = 8. El primer grupo de empates se corresponde con el valor 6 que se repite 2 veces ($t_1 = 2$), el segundo corresponde al valor 10 que se repite 2 veces ($t_2 = 2$), el tercero corresponde al valor 16 que se repite 2 veces ($t_3 = 2$), el cuarto corresponde al valor 17 que se repite 2 veces ($t_4 = 2$), el quinto se corresponde al valor 20 que se repite 4 veces ($t_5 = 4$), el sexto corresponde al valor 25 que se repite 2 veces ($t_6 = 2$), el séptimo corresponde al valor 27 que se repite 2 veces ($t_7 = 2$) y el octavo que se corresponde al valor 29 que se repite 2 veces ($t_8 = 2$). El valor z con esta corrección vale:

$$z = \frac{\left|95.5 - \frac{14 \cdot 19}{2}\right|}{\sqrt{\frac{14 \cdot 19(14+19+1)}{12} - \frac{14 \cdot 19\left[7(2^3-2)+(4^3-4)\right]}{12(14+19)(14+19-1)}}} = 1.37$$

Vemos que se acerca al valor criterial pero no lo alcanza. En consecuencia, los tres tests producen los mismos resultados: no hay evidencia suficiente para rechazar H_0.

Contraste de hipótesi no paramétrica para más de dos muestras

El último caso que vamos a tratar se aplica cuando disponemos de tres o más grupos independientes.

Tests de Kruskal-Wallis para más dos muestras independientes

Este test es una ampliación del test de Mann-Whitney para más de dos muestras independientes (denotaremos el número de tratamientos a comparar mediante *k*, siendo *k* necesariamente mayor que 2 ya que en caso contrario estaríamos en el test MWW). Por tanto, es un test que pone a prueba las medianas muestrales con datos medidos a nivel ordinal (o transformados a nivel ordinal a partir de formatos superiores, es decir, de intervalo o de razón). Su correspondencia paramétrica sería el Análisis de Varianza (Anova) de un único factor intersujetos. Los supuestos necesarios para poder aplicar adecuadamente este test son que:

- Las muestras han sido seleccionadas al azar de la población que representan.
- Las *k* muestras son independientes entre sí.
- La variable dependiente es una variable continua aunque posteriormente se transforma a una escala ordinal. No obstante, en la práctica este test también se emplea con variables aleatorias discretas.
- La función de densidad de probabilidad poblacional adoptada por cada distribución puede ser de cualquier tipo. Este supuesto implica homogeneidad de varianza.

Un aspecto positivo del test de Kruskal-Wallis es que al trabajar a partir de datos ordinales, disminuye o se elimina el impacto que pueden tener los valores extremos (*outliers*). En su aspecto negativo, el test de Kruskal-Wallis proporciona un test inferencial menos potente que el Anova de un único factor entre sujetos.

Debido a que es un test derivado del test MWW, con el que tiene muchas similitudes, lo presentaremos directamente mediante un ejemplo. Supongamos que un psicólogo está interesado en estudiar la efectividad de dos medicamentos analgésicos y de un placebo. El tiempo (en horas) transcurrido desde la administración de la píldora hasta que el paciente vuelve a quejarse de dolor se presenta en la Tabla 6:

Medicamento A	**Medicamento B**	**Placebo**
2.7	1.0	2.2
1.3	4.1	0.2
1.6	3.7	1.0
4.2	2.5	1.9
3.3	0.5	3.4

Tabla 6

El investigador desea poner a prueba la hipótesis de que el tipo de medicamento es irrelevante con respecto a la disminución del dolor que provoca. Como hicimos en el caso de dos muestras independientes, reemplazamos cada observación directa por sus rangos u órdenes. Así, reemplazamos la más pequeña de las 15 observaciones por 1, la siguiente más pequeña por 2, y así sucesivamente. Si existiesen empates (como es el caso en el ejemplo) utilizaremos el cálculo de rangos promediados. Los órdenes así como sus sumas para cada tratamiento son (el matiz de gris de la fila Tiempo y Rango ajustado representa el grupo, siendo el blanco el grupo Placebo, el gris claro el medicamento B y el gris oscuro el medicamento A).

Tiempo	0.2	0.5	1	1	1.3	1.6	1.9	2.2	2.5	2.7	3.3	3.4	3.7	4.1	4.2
Rango sin ajustar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rango ajustado	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15

El sumatorio de rangos para cada grupo es:

$$\sum R_A = 5+6+10+11+15 = 47$$
$$\sum R_B = 2+3.5+9+12+14 = 41.5$$
$$\sum R_P = 1+3.5+7+8+12 = 31.5$$

La suma de todos los rangos en un conjunto de 15 datos es 1+2+3+…+15 = 15 (15+1)/2= 120. Si asumimos que los rangos individuales se han distribuido aleatoriamente entre las tres condiciones, en promedio deberían valer 120/3= 40 si H_0 es cierta. Excepto las variaciones aleatorias producidas por el muestreo, esperamos que en cada grupo exista una suma de rangos igual a 40. Una medida de la magnitud de las diferencias entre el valor esperado de la suma de rangos y los valores realmente encontrados viene dado por el sumatorio de las diferencias al cuadrado:

$$\sum_{i=1}^{k}(R_i - \overline{R})^2 = (47-40)^2 + (41.5-40)^2 + (31.5-40)^2 = 123.5$$

Aunque esta medida es una buena indicación del grado de alejamiento de los datos con respecto a H_0, el criterio utilizado es otro: el estadístico H de Kruskal-Wallis:

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

siendo N el total de sujetos, n_i el número de sujetos en cada uno de los *k* grupos y R_i el sumatorio de los rangos para cada grupo. Aplicado a nuestro ejemplo, produce:

$$H = \frac{12}{15(15+1)}\left[\frac{47^2}{5} + \frac{41.5^2}{5} + \frac{31.5^2}{5}\right] - 3(15+1) = 1.235$$

Observemos que H es, forzosamente, igual o superior a 0. La estadística demuestra que H se distribuye según la distribución χ^2 con *k*-1 grados de libertad. Buscamos en la Tabla de χ^2 con 3-1 = 2 grados de libertad la probabilidad de obtener un valor de H igual o mayor a 1.235. Aunque no podemos obtener un valor preciso, vemos que el valor de χ^2 más aproximado es 1.35, el cual deja por encima de sí el 0.5. De aquí deducimos que el estadístico H = 1.235 deja por encima de sí un valor un poco inferior a 0.5. Este valor es demasiado elevado si trabajamos a un α de 0.05 o de 0.01 (los niveles de α más utilizados en Psicología). En consecuencia los datos nos conducen a no poder rechazar H_0. Los niveles críticos que nos indica la Tabla χ^2 de son de 5.99 para un α del 0.05 y de 9.21 para un α del 0.01. Como nuestro valor de H es inferior a ambos, no podemos rechazar H_0.

El test H de Kruskal-Wallis en este ejemplo sugiere que los diferentes tratamientos producen la misma reducción del dolor. Lo que debemos plantearnos es la situación en que nos encontraríamos si hubiéramos rechazado H_0. En este caso el investigador sabría que hay al menos dos condiciones que difieren entre sí pero no sabe cuáles son. Es muy probable que la pregunta que deseara responder es si los tratamientos A y B difieren entre sí y, a su vez, si estos tratamientos difieren con respecto al placebo. Cuando, después de un test H significativo, se nos plateen este tipo de preguntas significa que debemos acudir al método de las comparaciones múltiples. Mientras que el test H nos ha permitido rechazar H_0 de forma global, el test de las comparaciones múltiples nos permitirá determinar entre qué grupos se ha producido la diferencia.

Para determinar la diferencia mínima en las medias de los rangos de cada par de condiciones que nos permite clasificarlos como estadísticamente diferentes a un nivel específico de significación debemos utilizar una la ecuación:

$$MV_{KW} = z_{adj}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

El valor z_{adj} se extrae de la tabla de la curva normal que deja por encima de sí el $\alpha/(2\cdot c)$ del área de la distribución, siendo c el número de comparaciones a realizar en un contraste bilateral. Si se realizan comparaciones unilaterales el valor de z_{adj} se calcula como α/c. Si declararán como significativas las diferencias de medias de rangos en valor absoluto que superen MV_{KW}, es decir, si

$$\left|\overline{R}_i - \overline{R}_j\right| > MV_{KW}$$

Expondremos estos cálculos mediante un ejemplo en donde las diferencias sean tan notables que se rechace H_0.

Ejemplo 5

Un psicólogo está interesado en estudiar el efecto que la deprivación del sueño tiene sobre la ansiedad. Un grupo de sujetos se asigna al azar a uno de tres tratamientos: A) sin interrupción del sueño; B) a los participantes se les despierta dos veces durante la noche mientras están en sueño profundo según el EEG y C) a los participantes se les despierta cuatro veces durante la noche mientras están en sueño profundo también según el EEG. El procedimiento se repitió durante cinco noches y cada día se les pasaba un test para determinar el nivel de ansiedad. Puntuaciones altas indican alta ansiedad. Las puntuaciones registradas fueron {7, 7, 3, 6, 5, 8} para el grupo A, {10, 9, 11, 10, 7, 6, 8} para el grupo B y {15, 11, 12, 9, 10} para el grupo C.

A continuación presentamos el cálculos de los rangos (como hay empates, hay que realizar el cálculo ajustado). El matiz de gris de la variable ansiedad indica el grupo (blanco grupo A, gris claro el grupo B y gris oscuro el grupo C). :

Ansiedad	3	5	6	6	7	7	7	8	8	9	9	10	10	10	11	11	12	15
Rango	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Rango ajustado	1	2	3.5	3.5	6	6	6	8.5	8.5	10.5	10.5	13	13	13	15.5	15.5	17	18

Los rangos para cada grupo valen:

$$R_A = 1 + 2 + 3.5 + 6 + 6 + 8.5 = 27$$
$$R_B = 3.5 + 6 + 8.5 + 10.5 + 13 + 13 + 15.5 = 70$$
$$R_C = 10.5 + 13 + 15.5 + 17 + 18 = 74$$

El cálculo del estadístico H proporciona un valor de 10.253

$$H = \frac{12}{18(18+1)}\left[\frac{27^2}{6} + \frac{70^2}{7} + \frac{74^2}{5}\right] - 3(18+1) = 10.253$$

Podemos observar en la Tabla de χ^2 para un α de 0.05 un valor crítico de 5.99 (y de 9.21 para un α de 0.01). Como H = 10.253 > 5.99 podemos rechazar H_0 y afirmamos que hay al menos dos grupos que difieren significativamente entre sí. Como hemos rechazado H_0 es necesario realizar los contrastes *a posteriori* para determinar entre qué grupos se encuentra esta diferencia. Con tal fin, calcularemos primero las diferencias en valor absoluto entre las medias de los rangos:

$$\overline{R}_A = \frac{\sum R_A}{n_A} = \frac{27}{6} = 4.5 \qquad \overline{R}_B = \frac{\sum R_B}{n_B} = \frac{70}{7} = 10 \qquad \overline{R}_C = \frac{\sum R_C}{n_C} = \frac{74}{5} = 14.8$$

$$\left|\overline{R}_A - \overline{R}_B\right| = |4.5 - 10| = 5.5$$
$$\left|\overline{R}_A - \overline{R}_C\right| = |4.5 - 14.8| = 10.3$$
$$\left|\overline{R}_B - \overline{R}_C\right| = |10 - 14.8| = 4.8$$

A continuación calculamos el valor crítico MVKW considerando que queremos realizar todas las comparaciones posibles entre nuestros tres grupos ($k(k-1)/2$) y con un α bilateral del 0.05:

$$\alpha/(2\cdot c) = 0.05/2\cdot 3 = 0.0083$$

En las tablas de la curva normal encontramos el valor de z que deja por encima de sí el 0.0083 del área de la curva normal. Este valor es 2.39 correspondiente a un área inferior acumulada de 0.9916, es decir, 1-0.0083.

Como tenemos un número de sujetos distinto en cada grupo debemos calcular un valor distinto de MV_{KW} para cada comparación:

$$MV_{KW} = 2.39\sqrt{\frac{18(18+1)}{12}\left(\frac{1}{6}+\frac{1}{7}\right)} = 7.09 > 5.5 = \left|\bar{R}_A - \bar{R}_B\right| \quad (no \quad significativa)$$

$$MV_{KW} = 2.39\sqrt{\frac{18(18+1)}{12}\left(\frac{1}{6}+\frac{1}{5}\right)} = 7.72 < 10.3 = \left|\bar{R}_A - \bar{R}_C\right| \quad (significativa)$$

$$MV_{KW} = 2.39\sqrt{\frac{18(18+1)}{12}\left(\frac{1}{7}+\frac{1}{5}\right)} = 7.47 > 4.8 = \left|\bar{R}_B - \bar{R}_C\right| \quad (no \quad significativa)$$

Observamos que sólo la comparación entre los grupos A y C (sin deprivación de sueño y con 4 horas de deprivación) resultó significativa.

Tabla de la distribución binomial acumulativa

						θ				
n	*x*	.05	.10	.15	.20	.25	.30	.35	.40	.45
1	0	.95	.90	.85	.80	.75	.70	.65	.60	.55
	1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025
	1	.9975	.9900	.9775	.9600	.9375	.9100	.8775	.8400	.7975
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664
	1	.9928	.9720	.9392	.8960	.8438	.7840	.7182	.6480	.5748
	2	.9999	.9990	.9966	.9920	.9844	.9730	.9561	.9360	.9089
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915
	1	.9860	.9477	.8905	.8192	.7383	.6517	.5630	.4752	.3910
	2	.9995	.9963	.9880	.9728	.9492	.9163	.8735	.8208	.7585
	3	1.000	.9999	.9995	.9984	.9961	.9919	.9850	.9744	.9590
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503
	1	.9774	.9185	.8352	.7373	.6328	.5282	.4284	.3370	.2562
	2	.9988	.9914	.9734	.9421	.8965	.8369	.7648	.6826	.5931
	3	1.000	.9995	.9978	.9933	.9844	.9692	.9460	.9130	.8688
	4	1.000	1.000	.9999	.9997	.9990	.9976	.9947	.9898	.9815
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277
	1	.9672	.8857	.7765	.6554	.5339	.4202	.3191	.2333	.1636
	2	.9978	.9842	.9527	.9011	.8306	.7443	.6471	.5443	.4415
	3	.9999	.9987	.9941	.9830	.9624	.9295	.8826	.8208	.7447
	4	1.000	.9999	.9996	.9984	.9954	.9891	.9777	.9590	.9308
	5	1.000	1.000	1.000	.9999	.9998	.9993	.9982	.9959	.9917
	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152
	1	.9556	.8503	.7166	.5767	.4449	.3294	.2338	.1586	.1024
	2	.9962	.9743	.9262	.8520	.7564	.6471	.5323	.4199	.3164
	3	.9998	.9973	.9879	.9667	.9294	.8740	.8002	.7102	.6083
	4	1.000	.9998	.9988	.9953	.9871	.9712	.9444	.9037	.8471
	5	1.000	1.000	.9999	.9996	.9987	.9962	.9910	.9812	.9643
	6	1.000	1.000	1.000	1.000	.9999	.9998	.9994	.9984	.9963
	7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

n	x	θ .50	.55	.60	.65	.70	.75	.80	.85	.90	.95
1	0	.500	.450	.400	.350	.300	.250	.200	.150	.100	.050
	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0	.2500	.2025	.1600	.1225	.0900	.0625	.0400	.0225	.0100	.0025
	1	.7500	.6975	.6400	.5775	.5100	.4375	.3600	.2775	.1900	.0975
	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0	.1250	.0911	.0640	.0429	.0270	.0156	.0080	.0034	.0010	.0001
	1	.5000	.4252	.3520	.2818	.2160	.1562	.1040	.0608	.0280	.0072
	2	.8750	.8336	.7840	.7254	.6570	.5781	.4880	.3859	.2710	.1426
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0	.0625	.0410	.0256	.0150	.0081	.0039	.0016	.0005	.0001	.0000
	1	.3125	.2415	.1792	.1265	.0837	.0508	.0272	.0120	.0037	.0005
	2	.6875	.6090	.5248	.4370	.3483	.2617	.1808	.1095	.0523	.0140
	3	.9375	.9085	.8704	.8215	.7599	.6836	.5904	.4780	.3439	.1855
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0	.0312	.0185	.0102	.0053	.0024	.0010	.0003	.0001	.0000	.0000
	1	.1875	.1312	.0870	.0540	.0308	.0156	.0067	.0022	.0005	.0000
	2	.5000	.4069	.3174	.2352	.1631	.1035	.0579	.0266	.0086	.0012
	3	.8125	.7438	.6630	.5716	.4718	.3672	.2627	.1648	.0815	.0226
	4	.9688	.9497	.9222	.8840	.8319	.7627	.6723	.5563	.4095	.2262
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0	.0156	.0083	.0041	.0018	.0007	.0002	.0001	.0000	.0000	.0000
	1	.1094	.0692	.0410	.0223	.0109	.0046	.0016	.0004	.0004	.0000
	2	.3438	.2553	.1792	.1174	.0705	.0376	.0170	.0059	.0013	.0001
	3	.6562	.5585	.4557	.3529	.2557	.1694	.0989	.0473	.0158	.0022
	4	.8906	.8364	.7667	.6809	.5798	.4661	.3446	.2235	.1143	.0328
	5	.9844	.9723	.9533	.9246	.8824	.8220	.7379	.6229	..4686	.2649
	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	0	.0078	.0037	.0016	.0006	.0002	.0001	.0000	.0000	.0000	.0000
	1	.0625	.0357	.0188	.0090	.0038	.0013	.0004	.0001	.0000	.0000
	2	.2266	.1529	.0963	.0556	.0288	.0129	.0047	.0012	.0002	.0000
	3	.5000	.3917	.2898	.1998	.1260	.0706	.0333	.0121	.0027	.0002
	4	.7734	.6836	.5801	.4677	.3529	.2436	.1480	.0738	.0257	.0038
	5	.9375	.8976	.8414	.7662	.6706	.5551	.4233	.2834	.1497	.0444
	6	.9922	.9848	.9720	.9510	.9176	.8665	.7903	.6794	.5217	.3017
	7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45
						θ				
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084
	1	..9428	.8131	.6572	.5033	.3671	.2553	.1691	.1064	.0632
	2	.9942	.9619	.8948	.7969	.6785	.5518	.4278	.3154	.2201
	3	.9996	.9950	.9786	.9437	.8862	.8059	.7064	.5941	.4770
	4	1.000	.9996	.9971	.9896	.9727	.9420	.8939	.8263	.7396
	5	1.000	1.000	.9998	.9988	.9958	.9887	.9747	.9502	.9115
	6	1.000	1.000	1.000	.9999	.9996	.9987	.9964	.9915	.9819
	7	1.000	1.000	1.000	1.000	1.000	.9999	.9998	.9993	.9983
	8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046
	1	.9288	.7748	.5995	.4362	.3003	.1960	.1211	.0705	.0385
	2	.9916	.9470	.8591	.7382	.6007	.4628	.3373	.2318	.1495
	3	.9994	.9917	.9661	.9144	.8343	.7297	.6089	.4826	.3614
	4	1.000	.9991	.9944	.9804	.9511	.9012	.8283	.7334	.6214
	5	1.000	.9999	.9994	.9969	.9900	.9747	.9464	.9006	.8342
	6	1.000	1.000	1.000	.9997	.9987	.9957	.9888	.9750	.9502
	7	1.000	1.000	1.000	1.000	.9999	.9996	.9986	.9962	.9909
	8	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997	.9992
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025
	1	.9139	.7361	.5443	.3758	.2440	.1493	.0860	.0464	.0233
	2	.9885	.9298	.8202	.6778	.5256	.3828	.2616	.1673	.0996
	3	.9990	.9872	.9500	.8791	.7759	.6496	.5138	.3823	.2660
	4	.9999	.9984	.9901	.9672	.9219	.8497	.7515	.6331	.5044
	5	1.000	.9999	.9986	.9936	.9803	.9527	.9051	.8338	.7384
	6	1.000	1.000	.9999	.9991	.9965	.9894	.9740	.9452	.8980
	7	1.000	1.000	1.000	.9999	.9996	.9984	.9952	.9877	.9726
	8	1.000	1.000	1.000	1.000	1.000	.9999	.9995	.9983	.9955
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014
	1	.8981	.6974	.4922	.3221	.1971	.1130	.0606	.0302	.0139
	2	.9848	.9104	.7788	.6174	.4552	.3127	.2001	.1189	.0652
	3	.9984	.9815	.9306	.8389	.7133	.5696	.4256	.2963	.1911
	4	.9999	.9972	.9841	.9496	.8854	.7897	.6683	.5328	.3971
	5	1.000	.9997	.9973	.9883	.9657	.9218	.8513	.7535	.6331
	6	1.000	1.000	.9997	.9980	.9924	.9784	.9499	.9006	.8262
	7	1.000	1.000	1.000	.9998	.9988	.9957	.9878	.9707	.9390
	8	1.000	1.000	1.000	1.000	.9999	.9994	.9980	.9941	.9852
	9	1.000	1.000	1.000	1.000	1.000	1.000	.9998	.9993	.9978
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9998
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

		θ									
n	*x*	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
8	0	.0039	.0017	.0007	.0002	.0001	.0000	.0000	.0000	.0000	.0000
	1	.0352	.0181	.0085	.0036	.0013	.0004	.0001	.0000	.0000	.0000
	2	.1445	.0885	.0498	.0253	.0113	.0042	.0012	.0002	.0000	.0000
	3	.3633	.2604	.1737	.1061	.0580	.0273	.0104	.0029	.0004	.0000
	4	.6367	.5230	.4059	.2936	.1941	.1138	.0563	.0214	.0050	.0004
	5	.8555	.7799	.6846	.5722	.4482	.3215	.2031	.1052	.0381	.0058
	6	.9648	.9368	.8936	.8309	.7447	.6329	.4967	.3428	.1869	.0572
	7	.9961	.9916	.9832	.9681	.9424	.8999	.8322	.7275	.5695	.3366
	8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0	.0020	.0008	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0195	.0091	.0038	.0014	.0004	.0001	.0000	.0000	.0000	.0000
	2	.0898	.0498	.0250	.0112	.0043	.0013	.0003	.0000	.0000	.0000
	3	.2539	.1658	.0994	.0536	.0253	.0100	.0031	.0006	.0001	.0000
	4	.5000	.3786	.2666	.1717	.0988	.0489	.0196	.0056	.0009	.0000
	5	.7461	.6386	.5174	.3911	.2703	.1657	.0856	.0339	.0083	.0006
	6	.9102	.8505	.7682	.6627	.5372	.3993	.2618	.1409	.0530	.0084
	7	.9805	.9615	.9295	.8789	.8040	.6997	.5638	.4005	.2252	.0712
	8	.9980	.9954	.9899	.9793	.9596	.9249	.8658	.7684	.6126	.3698
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	.0010	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0107	.0045	.0017	.0005	.0001	.0000	.0000	.0000	.0000	.0000
	2	.0547	.0274	.0123	.0048	.0016	.0004	.0001	.0000	.0000	.0000
	3	.1719	.1020	.0548	.0260	.0106	.0035	.0009	.0001	.0000	.0000
	4	.3770	.2616	.1662	.0949	.0473	.0197	.0064	.0014	.0001	.0000
	5	.6230	.4956	.3669	.2485	.1503	.0781	.0328	.0099	.0016	.0001
	6	.8281	.7340	.6177	.4862	.3504	.2241	.1209	.0500	.0128	.0010
	7	.9453	.9004	.8327	.7384	.6172	.4744	.3222	.1798	.0702	.0115
	8	.9893	.9767	.9536	.9140	.8507	.7560	.6242	.4557	.2639	.0861
	9	.9990	.9975	.9940	.9865	.9718	.9437	.8926	.8031	.6513	.4013
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11	0	.0005	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0059	.0022	.0007	.0002	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0327	.0148	.0059	.0020	.0006	.0001	.0000	.0000	.0000	.0000
	3	.1133	.0610	.0293	.0122	.0043	.0012	.0002	.0000	.0000	.0000
	4	.2744	.1738	.0994	.0501	.0216	.0076	.0020	.0003	.0000	.0000
	5	.5000	.3669	.2465	.1487	.0782	.0343	.0117	.0027	.0003	.0000
	6	.7256	.6029	.4672	.3317	.2103	.1146	.0504	.0159	.0028	.0001
	7	.8867	.8089	.7037	.5744	.4304	.2867	.1611	.0694	.0185	.0016
	8	.9673	.9348	.8811	.7999	.6873	.5448	.3826	.2212	.0896	.0152
	9	.9941	.9861	.9698	.9394	.8870	.8029	.6779	.5078	.3026	.1019

10	.9995	.9986	.9964	.9912	.9802	.9578	.9141	.8327	.6862	.4312
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

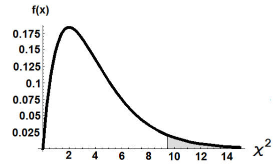
		θ								
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008
	1	.8816	.6590	.4435	.2749	.1584	.0850	.0424	.0196	.0083
	2	.9804	.8891	.7358	.5583	.3907	.2528	.1513	.0834	.0421
	3	.9978	.9744	.9078	.7946	.6488	.4925	.3467	.2253	.1345
	4	.9998	.9957	.9761	.9274	.8424	.7237	.5833	.4382	.3044
	5	1.000	.9995	.9954	.9806	.9456	.8822	.7873	.6652	.5269
	6	1.000	.9999	.9993	.9961	.9857	.9614	.9154	.8418	.7393
	7	1.000	1.000	.9999	.9994	.9972	.9905	.9745	.9427	.8883
	8	1.000	1.000	1.000	.9999	.9996	.9983	.9944	.9847	.9644
	9	1.000	1.000	1.000	1.000	1.000	.9998	.9992	.9972	.9921
	10	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997	.9989
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004
	1	.8646	.6213	.3983	.2336	.1267	.0637	.0296	.0126	.0049
	2	.9755	.8661	.7296	.5017	.3326	.2025	.1132	.0579	.0269
	3	.9969	.9658	.9033	.7473	.5843	.4206	.2783	.1686	.0929
	4	.9997	.9935	.9740	.9009	.7940	.6543	.5005	.3530	.2279
	5	1.000	.9991	.9947	.9700	.9198	.8346	.7159	.5744	.4268
	6	1.000	.9999	.9987	.9930	.9757	.9376	.8705	.7712	.6437
	7	1.000	1.000	.9998	.9988	.9944	.9818	.9538	.9023	.8212
	8	1.000	1.000	1.000	.9998	.9990	.9960	.9874	.9679	.9302
	9	1.000	1.000	1.000	1.000	.9999	.9993	.9975	.9922	.9797
	10	1.000	1.000	1.000	1.000	1.000	.9999	.9997	.9987	.9959
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9995
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002
	1	.8470	.5846	.3567	.1979	.1010	.0475	.0205	.0081	.0029
	2	.9699	.8416	.6479	.4481	.2811	.1608	.0839	.0398	.0170
	3	.9958	.9559	.8535	.6982	.5213	.3552	.2205	.1243	.0632
	4	.9996	.9908	.9533	.8702	.7415	.5842	.4227	.2793	.1672
	5	1.000	.9985	.9885	.9561	.8883	.7805	.6405	.4859	.3373
	6	1.000	.9998	.9978	.9884	.9617	.9067	.8164	.6925	.5461
	7	1.000	1.000	.9997	.9976	.9897	.9685	.9247	.8499	.7414
	8	1.000	1.000	1.000	.9996	.9978	.9917	.9757	.9417	.8811
	9	1.000	1.000	1.000	1.000	.9997	.9983	.9940	.9825	.9574
	10	1.000	1.000	1.000	1.000	1.000	.9998	.9989	.9961	.9886
	11	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9994	.9978
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.9999	.9997
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tabla de valores críticos para la prueba de Wilcoxon para una muestra o para dos muestras relacionadas

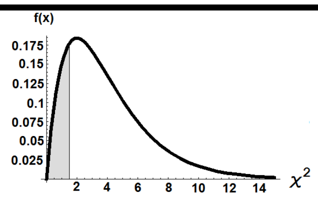
	Unilateral					Unilateral			
	.05	.025	.01	.005		.05	.025	.01	.005
	Bilateral					Bilateral			
	.10	.05	.02	.01		.10	.05	.02	.01
n					n				
5	0	-	-	-	28	130	116	101	91
6	2	0	-	-	29	140	126	110	100
7	3	2	0	-	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

Tabla C (χ^2)



Probabilidades superiores

Grados de libertad	**.990**	**.975**	**.950**	**.900**	**.750**	**.500**	**.250**	**.100**	**.050**	**.025**	**.010**
1		.001	.004	.016	.102	.455	1.32	2.71	3.84	5.02	6.63
2	.020	.051	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21
3	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3
4	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3
5	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1
6	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8
7	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5
8	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1
9	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7
10	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2
11	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7
12	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2
13	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7
14	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1
15	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6
16	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0
17	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4
18	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8
19	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2
20	8.26	9.39	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6
21	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9
22	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3
23	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6
24	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0
25	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3
26	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6
27	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0
28	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3
29	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6
30	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9
….											
40	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7
60	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4
	.01	.025	.050	.100	.250	.500	.750	.900	.950	.975	.990



Probabilidades inferiores

Tabla de valores críticos para el test U de Mann-Whitney-Wilcoxon

Contraste bilateral al 0.05

n_2	n_1																			
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
1																				
2								0	0	0	0	1	1	1	1	1	2	2	2	2
3					0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4				0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5			0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6			1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7			1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8		0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9		0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10		0	3	5	8	11	1	17	20	23	26	29	33	36	39	42	45	48	52	55
11		0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12		1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13		1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14		1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15		1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16		1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17		2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18		2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19		2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20		2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Contraste unilateral al 0.05

n_2	n_1																			
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
1																			0	0
2					0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3			0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4			0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5		0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6		0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7		0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8		1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9		1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10		1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11		1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12		2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13		2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14		2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92

15		3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16		3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17		3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18		4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

Contraste bilateral al 0.01

n_2	n_1																			
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
1																				
2																			0	0
3									0	0	0	1	1	1	2	2	2	2	3	3
4						0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5					0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6				0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7				0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8				1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9			0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10			0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11			0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12			1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13			1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14			1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15			2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16			2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17			2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18			2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19		0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20		0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

Contraste unilateral al 0.01

n_2	n_1																			
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
1																				
2													0	0	0	0	0	0	1	1
3							0	0	1	1	1	2	2	2	3	3	4	4	4	5
4					0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
5				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6				1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
7			0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28
8			0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34
9			1	3	5	7	9	11	14	16	18	21	23	26	28	31	33	36	38	40
10			1	3	6	8	11	13	16	19	22	24	27	30	33	36	38	41	44	47
11			1	4	7	9	12	15	18	22	25	28	31	34	37	41	44	47	50	53
12			2	5	8	11	14	17	21	24	28	31	35	38	42	46	49	53	56	60

13	0	2	5	9	12	16	20	23	27	31	35	39	43	47	51	55	59	63	67
14	0	2	6	10	13	17	22	26	30	34	38	43	47	51	56	60	65	69	73
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56	61	66	70	75	80
16	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66	71	76	82	87
17	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77	82	88	93
18	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88	94	100
19	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	107
20	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114

Bibliografía

Daniel, W. W. (1990). Applied Nonparametric Statistics. The Duxbury Advanced Series in Statistics and Decision Sciences, Boston.

Hyvtirinen, A., Karhunen, J. y Oja, E. (2001). Independent Component Analysis. John Wiley & Sons, New Cork.

Nemirovski, A. (2000). Topics in non-parametric statistics. En M. Emery, A. Nemirovski y D. Voiculescu (Eds.), Lectures on Probability Theory and Statistics. Springer, New York.

Ross, S. M. (2006). Introduction to probability and statistics for engineers and Scientists. New York: Wiley.

Siegel, S. y Castellan, N. J. (1988). Nonparametric Statistics for the behavioral sciences (2nd ed.). New York: McGraw-Hill Book Company.

Wasserman, L. (2006). All of Nonparametric Statistics. Springer.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80-83.