CAPITULO VIII

METODOS NO PARAMETRICOS PARA ANALISIS QUIMICO

CONTENIDO

8.0 METODOS ESTADÍSTICOS NO PARAMÉTRICOS

Los métodos estadísticos paramétricos suponen que los datos que se analizan siguen una distribución normal (Gaussiana). La validez de esta hipótesis se basa en el teorema central del límite, que postula que la distribución muestral de la media puede ser aproximadamente normal aunque la población de referencia tenga una distribución muy diferente. La aproximación mejora a medida que el tamaño de la muestra aumenta.

Con frecuencia se presentan a los químicos analistas situaciones donde no pueden asumirse los supuestos requeridos por desconocerse la distribución de la variable estudiada, o bien, por ser la muestra muy pequeña de manera que incluso el teorema central del límite sería de escasa relevancia. O bien, aunque se conozca la distribución de la variable y sea válido teóricamente hacer supuestos sobre la distribución y sus parámetros, no es razonable utilizar una prueba paramétrica ya que hay poca certeza de que se cumplan en este conjunto de datos. También puede ocurrir que la variable no sea continua por lo que no se cumplen las restricciones establecidas para las pruebas paramétricas.

En cualquiera de los casos anteriores hay que buscar técnicas alternativas que permitan darle solución a estas situaciones de forma eficiente. Surge así la necesidad de desarrollar una serie de técnicas estadísticas que tengan un mínimo de restricciones. A estas técnicas se les conoce como: Métodos no paramétricos.

Según WONNACOTT (1973) existen dos indicaciones para preferir las pruebas no paramétricas:

1. Cuando la prueba clásica correspondiente no es válida.
2. En aplicaciones en donde la prueba clásica es razonablemente válida, pero un estimador no paramétrico puede ser más eficiente.

Entre las ventajas del uso de métodos no paramétricos se encuentran las siguientes (PRIA, 2001, cap. 1):

1. Tienen mayor eficiencia que los métodos paramétricos en distribuciones asimétricas, o sea cuando hay valores atípicos o datos aberrantes.
2. Tienen validez en el sentido de que su nivel de confiabilidad es realmente el especificado en la mayoría de las pruebas.

3. Generalmente son de cómputo mucho más fácil que las técnicas de las estadísticas clásicas.
4. Son aplicables en situaciones donde los procedimientos clásicos no son aplicables.
5. Aún cuando se cumplan los requisitos para realizar una prueba paramétrica, si la muestra es pequeña la eficiencia relativa de la prueba no paramétrica es alta.

Entre las desventajas de la aplicación de los métodos no paramétricos se citan las siguientes (PRIA, 2001, cap. 1):

1. Fundamentalmente cuando las muestras son muy grandes las pruebas no paramétricas tienen una eficiencia relativa baja con relación a las paramétricas cuando se cumplen los supuestos.
2. Las hipótesis que se plantean en las pruebas no paramétricas son menos precisas, lo que hace que la interpretación de los resultados sea más ambigua.
3. Su aplicación en muestras grandes se hace muy laboriosa.
4. Para un problema particular pueden existir varias pruebas, por lo que en ocasiones se hace difícil seleccionar la mejor.

En general, este tipo de pruebas deben utilizarse cuando no se cumplan los supuestos para hacer una prueba paramétrica y cuando se duda sobre el cumplimiento de los supuestos en una muestra pequeña. La principal ventaja de las pruebas no paramétricas consiste en que pueden efectuarse inferencias exactas cuando las suposiciones fundamentales de los métodos estándar no pueden cumplirse en su totalidad; su principal desventaja radica en que exigen demasiada información y tienen una eficiencia menor cuando todas las suposiciones se satisfacen. Sin embargo, si se afirma que la eficiencia de cierto método no paramétrico es del 80%, se debe entender en realidad que es relativamente valioso, dado que la eficiencia del correspondiente método "estándar" es poco menos del 100% si todas las suposiciones no se cumplen con exactitud. En general es verdad que *cuanto menos se supone, tanto menos se puede inferir de un conjunto de datos,* pero también es cierto que *cuanto menos se supone, tanto más se amplía la aplicabilidad de un método* (FREUD y WALPOLE, 1990, cap. 16).
Es posible comparar la confiabilidad del uso de una prueba no paramétrica con respecto al uso de una prueba paramétrica homóloga a través del concepto de *potencia-eficiencia* de

una prueba, que en general, se refiere al aumento porcentual del tamaño de muestra que se necesita para obtener los mismos resultados con una prueba X, no tan poderosa como otra que es la prueba Y, que es la más potente que se conoce para resolver un problema específico cuando se cumplen los supuestos para efectuar la misma. La *potencia-eficiencia* de la prueba X puede calcularse de la forma siguiente (PRIA, 2001, cap. 1):

$$\textit{Potencia – Eficiencia de la prueba } X = \frac{N_Y}{N_x}\, x\ 100$$

En donde N_Y es el tamaño de muestra requerido para la prueba Y, y, N_X el tamaño de muestra requerido para la prueba X.

Existe una gran variedad de pruebas no paramétricas tanto para el análisis de variables cualitativas como cuantitativas. En el cuadro 8.1 se presenta un listado de dichas pruebas y de su aplicación.

En las secciones de la 8.1 a la 8.7 se describen con detalle las pruebas no paramétricas asociadas a variables de tipo cuantitativo de mayor aplicación en análisis químico.

Además de las pruebas expuestas en la tabla 8.1, existen pruebas no paramétricas para correlación y análisis de regresión, cuyas bases se presentan en el anexo A8.

8.1 PRUEBA DE LA MEDIANA

Esta prueba permite determinar si dos muestras independientes difieren con relación a sus medianas, o sea permite determinar si dos muestras independientes provienen de poblaciones con la misma mediana siempre que la variable esté al menos en escala ordinal (PRIA, 2001, cap. 1). Existe una generalización de esta prueba que permite la comparación de las medianas de tres o más muestras independientes que no será objeto de estudio en esta sección.

La hipótesis a probar es:

$$H_o:\ \tilde{X}_1 = \tilde{X}_2$$
$$H_1:\ \tilde{X}_1 \quad \tilde{X}_2$$

En donde $\tilde{X}_1$ y $\tilde{X}_2$ representan las medianas de las dos muestras que se están comparando.

CUADRO 8.1 DESCRIPCION DE PRUEBAS NO PARAMETRICAS

PRUEBA	TIPO DE VARIABLE	APLICACION
Prueba c^2 de independencia	Cualitativa	Esta prueba permite medir la significación de la asociación entre dos variables de clasificación.
Prueba c^2 de homogeneidad	Cualitativa	Cuando se tienen varias muestras y se desea determinar si son homogéneas con relación a la distribución en las mismas de una variable cualitativa.
Prueba Kolmogorov-Smirnov	Cuantitativa	Es una prueba no paramétrica que se utiliza para diferencias entre distribuciones acumuladas, es, pues, una prueba de bondad de ajuste.
Prueba de la mediana	Cuantitativa	Esta prueba permite determinar si dos muestras independientes difieren con relación a sus medianas, o sea permite determinar si dos muestras independientes provienen de poblaciones con la misma mediana.
Prueba de signos	Cuantitativa	Esta prueba permite la comparación de la mediana de una serie de datos con un valor especificado. También permite indicar la existencia de tendencias.
Prueba de rangos y signos de Wilcoxon	Cuantitativa	Permite probar la aleatoriedad de una secuencia de datos. También permite probar la simetría de una distribución. Otra aplicación de esta prueba es comparar la distribución de una serie de datos con un valor especificado.
Prueba de la suma de rangos de Wilcoxon	Cuantitativa	Constituye la base para el resto de pruebas que utilizan rangos y permite determinar si dos muestras proceden de la misma distribución, las muestras deben de ser del mismo tamaño y no necesariamente independientes.
Prueba de U-Man Whitney	Cuantitativa	Esta prueba se utiliza para resolver el mismo caso que resuelve la prueba de suma de rangos de Wilcoxon con muestras no necesariamente del mismo tamaño.
Prueba de Kruskal Wallis	Cuantitativa	Es una generalización de la prueba de suma de rangos de Wilcoxon, permitiendo comparar más de dos muestras con el propósito de conocer si proceden de la misma población o si hay diferencias entre las medidas de tendencia central de más de dos poblaciones.

Debe señalarse que en el caso de que las distribuciones de ambas poblaciones sean simétricas, la mediana debe coincidir con la media aritmética, convirtiéndose ésta en una prueba de comparación de dos medias también (PRIA, 2001, cap. 1).

Como se está asumiendo que ambas muestras provienen de poblaciones con la misma mediana, se determina el valor de la mediana general, o sea se combinan los valores de ambos grupos para determinar el valor de la mediana.

Una vez determinado éste valor se determina en cada una de las muestras cuántas observaciones tienen valores mayores que la mediana general y cuántas tienen valores inferiores a la misma.

Según SIEGEL (1972) si hay valores que coincidan con los de la *mediana general* en las muestras estudiadas, se procederá de la siguiente forma:

1. Si la muestra es grande y un número pequeño de observaciones coincide con este valor pueden excluirse del análisis estas observaciones.
2. Si la muestra no es grande o si el número de observaciones que coincide con este valor es grande entonces pueden contarse estos valores junto con los valores menores que la mediana.

La información se resumirá en una tabla de contingencia de la forma siguiente (PRIA, 2001, cap. 1):

Valores/Mediana	**Muestra 1**	**Muestra 2**	**Total**
Observaciones > Mediana General	a	b	a+b
Observaciones ≤ Mediana General	c	d	c+d
Total	**a+c**	**b+d**	**N**

Para la prueba de hipótesis se necesita un estadígrafo con una distribución conocida para determinar si existen diferencias significativas entre las medianas de ambas poblaciones. Para ello se utilizará la expresión (PRIA, 2001, cap. 1):

$$\chi_o^2 = \frac{n\left(|ad - bc| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Este estadístico bajo el supuesto de que no existen diferencias entre las medianas de ambas poblaciones se distribuye χ^2 con un grado de libertad.

Si $\chi_o^2 > \chi_{tabulada}^2$ con un grado de libertad para un nivel de significación determinado, se rechaza la hipótesis nula.

Al usar el estadígrafo χ^2 hay que tener en cuenta que se deben cumplir las restricciones para la prueba χ^2. En este caso, como hay una tabla de contingencia de 2 x 2, debe cumplirse que todos los valores esperados sean mayores o iguales a 5, en caso de no cumplirse esta restricción se deberá utilizar la prueba de las probabilidades exactas de Fisher (PRIA, 2001, cap. 1).

Según SIEGEL y MOOD (1954) se ha demostrado que si la prueba de la mediana se aplica a datos que pueden analizarse adecuadamente por una prueba paramétrica más poderosa, la prueba t en éste caso, su potencia-eficiencia sería de cerca de un 95% cuando $n_1 + n_2 = 6$. Este valor iría disminuyendo a medida que aumenta el tamaño muestral llegando a tener una eficiencia asintótica eventual de 63%.

8.2 LA PRUEBA DE SIGNOS

La prueba de signos es uno de los métodos no paramétricos más simples. La prueba t supone que los datos se distribuyen normalmente. La prueba del signo prescinde de tal hipótesis y es mucho más fácil de realizar. Se puede utilizar de diferentes formas, la más simple se describe a continuación:

Para probar la hipótesis nula $\mu = \mu_o$ contra una alternativa apropiada, basándose en una muestra aleatoria de tamaño *n*, se reemplaza cada valor muestral mayor que μ_o por un signo más y cada valor muestral menor que μ_o por un signo menos (MILLER y FREUND, 1986, cap. 10).

Se ignoran por completo aquellos valores que son iguales a μ_o. Para contrastar si la preponderancia de signos menos, es significativa se utiliza la ley de la binomial acumulada. Esta ley establece que la probabilidad de que aparezcan *r* signos menos entre *n* signos está dada por (MILLER y MILLER, 1993, cap. 6):

$$p(r) = nC_r \rho^r q^{n-r}$$

donde:

nC_r: indica el número de combinaciones de *r* elementos de un total de *n* elementos.

p: es la probabilidad de que aparezca un signo menos en uno de los resultados.

q: es la probabilidad de que no aparezca un signo menos en uno de los resultados, es decir, ***q* = *1 – p*.**

Si la probabilidad experimental es menor que un nivel de significación α, la hipótesis nula debe rechazarse. Es decir, existe evidencia como para rechazar que los datos proceden de una población con $\mu = \mu_o$ (MILLER y MILLER, 1993, cap. 6).

La prueba de signos puede utilizarse también como una alternativa no paramétrica de la prueba t por parejas para comparar dos grupos de resultados para las mismas muestras. Así, si se examinan *n* muestras con cada uno de los dos métodos, A y B se puede contrastar si los dos proporcionan lecturas significativamente diferentes (MILLER y MILLER, 1993, cap. 6).

Calculando para cada muestra la diferencia en los resultados, es decir (MILLER y MILLER, 1993, cap. 6):

$$[(resultado\ obtenido\ por\ el\ método A) - (resultado\ obtenido\ por\ el\ método\ B)]$$

La hipótesis nula será que los dos métodos no proporcionan resultados significativamente diferentes. En la práctica esto significará que la probabilidad de obtener un signo más (o un signo menos) es 0.5. La probabilidad o frecuencia de un número signos positivos sigue una distribución binomial de parámetros *n, p = 0.5* (MILLER y MILLER, 1993, cap. 6).

El resumen de la prueba del signo para comparar dos poblaciones se presenta en el cuadro 8.2.

Cuando el número de pares en un experimento de diferencias por parejas es grande, se puede aproximar la distribución de probabilidad para x, el número de veces que x_A es mayor que x_B, por una distribución normal. Por lo tanto, con $p = P(X_A > X_B)$ se puede probar H_O: p = 0.5 utilizando la prueba para una proporción binomial. (Ver cuadro 3.14, capítulo 3).

Otro uso de la prueba de signos es indicar una tendencia. En esta prueba al primer resultado del primer grupo se le resta el primer resultado del segundo grupo, al segundo resultado del primer grupo se le resta el segundo resultado del segundo grupo y así sucesivamente. Se obtiene de esta forma una serie de signos asociados a las diferencias obtenidas e ignorando los ceros se calcula la probabilidad binomial de obtener *r* pares de signos iguales. La hipótesis nula de que no existe ninguna tendencia en los resultados, se rechaza si la probabilidad calculada es menor que el nivel de significación α de la prueba de dos colas (MILLER y MILLER, 1993, cap. 6).

CUADRO 8.2 RESUMEN DE LA PRUEBA DEL SIGNO APLICADA AL ANALISIS DE DOS MUESTRAS

1. *Hipótesis nula:* $H_0: P(X_A - X_B) = 0.5$
2. *Hipótesis alternativa*

Prueba de una cola	**Prueba de dos colas**
$H_1: P(X_A - X_B) > 0.5$ (o bien $P(X_A - X_B) < 0.5$)	$H_1: P(X_A - X_B) \neq 0.5$

3. *Estadístico de contraste:* ***x*** *= número de pares de observaciones para los cuales* X_A *excede a* X_B.

4. *Región de rechazo:*

Prueba de una cola	**Prueba de dos colas**
Para: $P(X_A - X_B) > 0.5$ Rechace H_0 para valores muy grandes de ***x***. *Para*: $P(X_A - X_B) < 0.5$ Rechace H_0 para valores muy pequeños de ***x***.	Rechace H_0 para valores muy grandes o muy pequeño de ***x***. Basta observar si la frecuencia observada de signos positivos se desvía significativamente de la esperada n/2, en cuyo caso se rechaza H_0.

Suposiciones: Las observaciones x_A *y* x_B *se seleccionan aleatoria e independientemente por pares. Cuando los datos de las observaciones de* x_A *son iguales a las de* x_B *se eliminan y con esto se reduce el número de pares n.*

8.3 PRUEBA DE RACHAS DE WALD – WOLFWITZ.

Las rachas son una secuencia de signos positivos seguida de una de signos negativos, y luego otra de signos positivos. En algunos casos podría tener interés no sólo investigar si las observaciones generan signos positivos o negativos, si no también si éstos aparecen en una secuencia aleatoria.

En el caso de ajuste de curvas, una secuencia no aleatoria de signos positivos y negativos conducirá a un número más pequeño de rachas que una aleatoria. El método de Wald – Wolfwitz prueba si el número de rachas es suficientemente pequeño para que se rechace la hipótesis nula de una distribución aleatoria de los signos (MILLER y MILLER, 1993, cap. 6).

El número de rachas en los datos experimentales se compara con los números de la tabla del anexo B8, de "prueba de rachas de Wald – Wolfwitz", que se presenta para un nivel de significación α. En esta tabla los valores de *N* representan el número de signos positivos y *M* es el número de signos negativos. Si el número experimental de rachas es más pequeño que el valor tabulado, entonces se puede rechazar la hipótesis nula.

En relación a la prueba de Wald-Wolfwitz, existen otros puntos de interés, por ejemplo, permite comparar la mediana de una serie de datos con un valor especificado, al crear una serie de signos a través de la comparación de cada valor de la serie con el valor especificado y asignando un signo negativo a los datos inferiores al valor especificado y un signo positivo a los superiores a dicho valor. Ignorando el número de ceros, se determina el número de rachas obtenidas y se compara su valor con el propuesto en la tabla del anexo B8.

También cabe resaltar que se puede encontrar números inusualmente grandes de rachas cortas, así como también números inusualmente pequeños de rachas largas. Así por ejemplo, como en el caso de 6 signos positivos y 6 signos negativos en el orden + - + - + - + - + - + -, se podría sospechar que existe una secuencia no aleatoria. La tabla muestra que, con $N = M = 6$, un total de 11 ó 12 rachas indica que la hipótesis nula de un orden aleatorio se debería rechazar y sospechar una cierta periodicidad en los datos (MILLER y MILLER, 1993, cap. 6).

8.4 PRUEBA DE RANGOS Y SINGOS DE WILCOXON

La importancia de la prueba de signos radica en los supuestos mínimos que se hacen sobre los datos experimentales. No se supone que la población de la cual se toma la muestra sea normal, ni incluso que sea simétrica. La única información a priori necesaria es el valor de la mediana. Una cierta desventaja de la prueba de signos es que no utiliza toda la información disponible. Sólo es necesario saber si una medición individual es más grande o más pequeña que la mediana, y la magnitud de esta desviación no se utiliza en absoluto (MILLER y MILLER, 1993, cap. 6).

En muchos casos un analista tendrá suficientes razones para creer que sus mediciones se distribuyen simétricamente pero no desea suponer que siguen una distribución normal. Este supuesto de datos simétricos, y la consecuencia de que la media y la mediana de la población sean iguales, permite desarrollar pruebas de significación más potentes. Wilcoxon contribuyó con importantes avances al respecto, y su prueba de rangos y signos tienen varias aplicaciones (MILLER y MILLER, 1993, cap. 6).

La comparación de la distribución de una serie de datos con un valor de referencia se realiza mediante la obtención de las diferencias entre cada dato de la muestra y el valor de referencia (conservando los signos). Los valores absolutos de estas diferencias se ordenan posteriormente de menor a mayor y a continuación se incorporan sus signos, los números entonces se jerarquizan, en este proceso se mantienen sus signos pero se les asignan números que indican su orden o rango.

Luego, asignando con X a la suma de los rangos positivos y con Y a la suma de rangos negativos, se selecciona la menor de estas cifras (X ó Y) y se toma como el estadístico de contraste. El teorema binomial dará la probabilidad de que aparezca este número. Si los datos provienen de una población con una mediana igual al valor especificado, se esperaría que la suma de rangos positivos y negativos sea aproximadamente igual.

La probabilidad de que aparezca una suma concreta está dada en la tabla que se presenta en el anexo C8. En esta prueba se rechaza la hipótesis nula si el valor tabulado es menor o igual que el valor experimental, es decir, situación opuesta de la observada en la mayoría de las pruebas de significación.

Una ventaja importante de la prueba de rangos y signos es que también se puede utilizar para datos por parejas, ya que las diferencias entre los datos de las dos series se pueden transformar en el tipo de datos como en el caso anterior. De esta manera se puede utilizar este método no paramétrico como una alternativa a la prueba t por parejas (MILLER y MILLER, 1993, cap. 6).

Si no hay diferencia entre las dos series de datos, entonces se esperaría que las diferencias entre los resultados para cada muestra, [(resultado muestra 1) – (resultado muestra 2)] deberán distribuirse en torno a cero. Cuando hay posiciones empatadas, el problema se resuelve asignando posiciones promedio a los valores empatados, con signos adecuados. Si la ordenación es correcta, al calcular la suma de todos los valores sin tomar en cuenta el signo, debe ser la misma que la suma de los ocho primeros números naturales (MILLER y MILLER, 1993, cap. 6).

Si la suma de los valores positivos es *X* y la de los valores negativos es *Y,* el estadístico de prueba será la menor de estas sumas. Este valor se compara con el valor obtenido de la tabla del anexo C8, rechazando la hipótesis nula si el valor del estadístico de prueba es menor que el especificado en dicha tabla.

Es de hacer notar que la prueba de rangos y signos es un método simple y de mucha utilidad, aunque su principal limitación radica en que no se puede aplicar a conjuntos de datos muy pequeños, para el caso de prueba de dos colas con un nivel de significación $p = 0.05$, n tiene que ser al menos 6 (MILLER y MILLER, 1993, cap. 6).

8.5 PRUEBA DE SUMA DE RANGOS DE WILCOXON

La prueba de rangos y signos descrita anteriormente es válida para el estudio de conjuntos individuales de mediciones y para conjuntos de datos por parejas que se pueden reducir con facilidad a conjuntos individuales. Sin embargo, en muchos casos es necesario comparar dos muestras independientes que no se puedan reducir a un conjunto único de datos, ya que pueden contener diferentes números de mediciones. Existen varias pruebas no paramétricas que permiten abordar estas situaciones (MILLER y MILLER, 1993, cap. 6).

Sean *A* y *B*, dos muestras con *M* y *N* observaciones respectivamente, donde se cumple que $M \geq N$, la prueba de suma de rangos de Wilcoxon para determinar si ambas muestras proceden de la misma distribución se describe a continuación (PRIA, 2001, cap. 3):

1. Se ordenan todas las observaciones de ambas muestras, como si fuera una sola muestra, en orden ascendente y se asignan los rangos a los valores ordenados.
2. Se identifican los valores que pertenecen a cada muestra.
3. Se determina el estadígrafo que en esta prueba es: $T_{o} =$ Suma de rangos de B = (Σ rangos B) (en donde B es la muestra más pequeña)
4. *La regla de decisión será:*

   a) Se plantea la Hipótesis que se adecue a la situación que se necesita resolver, y se aplica la regla de decisión de acuerdo a lo que se presenta en el cuadro 8.3:

CUADRO 8.3 APLICACIÓN DE LA REGLA DE DECISION PARA LA PRUEBA DE SUMA DE RANGOS DE WILCOXON

Hipótesis	Regla de decisión rechazar H_o si:	α más usados
$H_o : Me_B = Me_A$ $H_1 : Me_B \neq Me_A$	$T_o \leq T_I$ ó $T_o \geq T_S$	0.025 0.05
$H_o : Me_B \geq Me_A$ $H_1 : Me_B < Me_A$	$T_o \leq T_I$	0.05 0.01
$H_o : Me_B \leq Me_A$ $H_1 : Me_B > Me_A$	$T_o \geq T_S$	0.05 0.01

Donde T_I y T_S son los valores obtenidos en la tabla de Valores críticos para la estadística de prueba de la Suma de Rangos de Wilcoxon (ver anexo D8), considerando un tamaño de n_A y n_B y un nivel de significación dado. Esta tabla sirve para trabajar cuando el tamaño de la muestra llega hasta 25 la muestra menor y 50 la mayor. Me_B y Me_A representan los parámetros de tendencia central de las distribuciones de ambas muestras.

Cuando el tamaño de la muestra menor excede las 25 observaciones puede trabajarse con la aproximación a la distribución normal (PRIA, 2001, cap. 3).

Otra forma de rechazar la hipótesis nula es cuando se está trabajando con un programa estadístico que calcule T_o y p que es la probabilidad de error de tipo I asociada a ese valor. En éste caso si el valor de p es menor que el α prefijado se rechaza la hipótesis nula.

Deben aclararse los aspectos siguientes (PRIA, 2001, cap. 3):

1. Si $n_A = n_B$ se seleccionará el estadígrafo Σ rangos de la muestra tomando en consideración las hipótesis alternativas, de la forma siguiente:

Si la Hipótesis Alternativa es:	Se seleccionará la muestra:
$H_1 : Me_B \neq Me_A$	Cualquiera de las dos tiene igual solución.
$H_1 : Me_B < Me_A$	T_o = la que tenga mayor suma de rangos.
$H_1 : Me_B > Me_A$	T_o = la que tenga menor suma de rangos.

2. Las hipótesis deben plantearse tomando como primer parámetro de referencia para contrastar el de la muestra más pequeña, en el caso de que sean de igual tamaño podrán plantearse de cualquier forma.

3. En caso de existir observaciones de igual valor se asignarán rangos promedios, por ejemplo si los tres primeros valores son:

VALORES	15	15	15
LUGARES	1	2	3
RANGOS	(1+2+3)/3 =**2**	(1+2+3)/3 =**2**	(1+2+3)/3 =**2**

4. Potencia - Eficiencia:

Según Siegel, si la prueba de la Suma de Rangos de Wilcoxon se aplica a datos que pueden analizarse adecuadamente por una prueba paramétrica más poderosa, la prueba *t* de Student en éste caso, su potencia-eficiencia sería de cerca de un 95.5% y se acerca a 95% para muestras de tamaño moderado, por lo que es una excelente alternativa ante la prueba *t*.

Debe destacarse que esta prueba es más potente que la prueba de la Mediana, pues esta última utiliza solamente la información de cómo están ubicadas las observaciones de cada muestra con relación al valor de la *mediana general*, en cambio la prueba de *suma de rangos de Wilcoxon* utiliza además, la información relativa a la ubicación de cada observación en las muestras, que se resume en el estadígrafo de la *suma de rangos*.

8.6 PRUEBA MANN-WHITNEY

Otra prueba que se utiliza para resolver el mismo caso que resuelve la prueba de la Suma de Rangos de Wilcoxon es la prueba de Mann-Whitney. El procedimiento seguido en ambas es muy parecido. Las hipótesis que se contrastan son las mismas y el estadígrafo utilizado se parece aunque por supuesto no es igual. A continuación se describe el procedimiento para el contraste de hipótesis mediante el uso de la prueba de Mann-Whitney:

Inicialmente se identifican ambas muestras A y B, con M y N observaciones respectivamente, donde se cumple que $M \geq N$. Todas las observaciones de ambas muestras se ordenan, como si fuera una sola muestra, en orden ascendente y se asignan los rangos a los valores ordenados. Posteriormente, se identifican los valores que pertenecen a cada muestra y se calculan las <u>*sumas de rangos*</u> de cada muestra y se define *S*, que es la <u>*suma de rangos*</u> de menor valor (PRIA, 2001, cap. 3).

El estadístico de contraste para esta prueba es (PRIA, 2001, cap. 3):

$$T_o = S - \frac{N(N+1)}{2}$$

La *regla de decisión* vendrá dada en función del planteamiento de la hipótesis que se adecue a la situación que se desea resolver, como se muestra en el cuadro 8.4 (PRIA, 2001, cap. 3):

CUADRO 8.4 APLICACIÓN DE LA REGLA DE DECISION PARA LA PRUEBA DE MANN-WHITNEY

Hipótesis	Regla de decisión rechazar H_O si:	a más usados
$H_o : Me_B = Me_A$ $H_1 : Me_B \neq Me_A$	$T_o \leq W_{\propto/2}$ o $T_o \geq W_{1-\propto/2}$	0.025 0.005
$H_o : Me_B \geq Me_A$ $H_1 : Me_B < Me_A$	$T_o \leq W_{\propto}$	0.05 0.01
$H_o : Me_B \leq Me_A$ $H_1 : Me_B > Me_A$	$T_o \geq W_{1-\propto}$ donde $W_{1-\propto} = NM - W_{\propto}$	0.05 0.01

Donde $W_{\propto}$ y $W_{\propto/2}$ son los valores críticos obtenidos en la tabla de valores críticos para la estadística de prueba de Mann-Whitney presentada en el anexo D8, en la que se consideran tamaños muestrales $n_{A,}$ n_B y un nivel de significación α. Esta tabla sirve para trabajar cuando los tamaños muestrales son de 20 para la muestra menor y 40 para la mayor.

Cuando el tamaño de la muestra menor excede las 20 observaciones puede trabajarse con la aproximación a la distribución normal (ver capítulo 3.0).

Otra forma de rechazar la hipótesis nula es cuando se está trabajando con un programa estadístico que calcule T_o y *p* que es la probabilidad de error de tipo I asociada a ese valor, en este caso se rechaza la hipótesis nula si *p* es menor al nivel de significación α tomado para la prueba.

8.7 PRUEBA DE KRUSKAL WALLIS

Cuando se presenta el problema de comparar más de dos muestras con el propósito de conocer si proceden de la misma población, o bien, comparar si existen diferencias entre las medidas de tendencia central de más de dos poblaciones y no se justifica la suposición de normalidad y de igualdad de varianzas, el químico analista podrá recurrir a un procedimiento alternativo al de la prueba *F* del análisis de la variancia y que no depende de esta suposición. Kruskal y Wallis (1952) desarrollaron un procedimiento como alternativa para dar solución a este problema, conocido como *la prueba de Kruskal – Wallis.*

La *prueba de Kruskal–Wallis* constituye una alternativa no paramétrica al análisis de varianza usual y se considera como una extensión del procedimiento de *suma de rangos de Wilcoxon* como se verá en su desarrollo (MILLER y FREUND, 1989, cap. 10).

La hipótesis nula para la prueba de Kruskal-Wallis es que no existe diferencia entre los tratamientos ($\mu_1 = \mu_2 = ... = \mu_a$), mientras que la hipótesis alternativa es que exista diferencia entre al menos un par de tratamientos (($\mu_i \neq \mu_j$).

Para realizar la prueba de Kruskal–Wallis los datos pueden agruparse como se presenta en la tabla a continuación (MONTGOMERY, 1991, cap. 4):

Repeticiones	**Tratamientos**			
	A	**B**	**C**	...
1	X_{1A}	X_{1B}	X_{1C}	...
2	X_{2A}	X_{2B}	X_{2C}	...
⋮	⋮	⋮	⋮	...
n	X_{nA}	X_{nB}	X_{nC}	...

Donde X_{ij} son las observaciones
i = 1, 2, ..., n
j = A, B, C, ...
N = $n_A + n_B + n_C$ ...

Las observaciones se organizan posteriormente, en orden ascendente y se asignan las posiciones (o rangos) con el rango 1 correspondiente a la observación más pequeña. En caso de empate (varias observaciones con el mismo rango o posición), se asigna el rango promedio a cada observación empatada (MONTGOMERY, 1991, cap. 4).

El estadístico de prueba de Kruskal-Wallis está dado por la expresión (MONTGOMERY, 1991, cap. 4):

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right]$$

En donde $R_{i.}$ es la suma de los rangos de las observaciones del i-ésimo tratamiento y n_i es el número de observaciones del i-ésimo tratamiento; N es el número total de observaciones y (MONTGOMERY, 1991, cap. 4)

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{ni} R_{ij}^2 - \frac{N(N-1)^2}{4}\right]$$

Debe notarse que S^2 es igual a la varianza de los rangos. Si no hay empate, S^2 = N(N+1)/12 y el estadístico de prueba se simplifica a (MONTGOMERY, 1991, cap. 4):

$$H = \frac{12}{N-1}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - 3(N+1)\right]$$

Si n_i es razonablemente grande, como sería el caso de $n_i \geq 5$, entonces H tiene una distribución aproximadamente χ^2_{a-1} (ver tabla del anexo C3) si la hipótesis nula es verdadera. Por lo tanto, si (MONTGOMERY, 1991, cap. 4)

$$H \geq \chi^2_{\alpha,a-1}$$

hay que rechazar la hipótesis nula.

Si H se calcula mediante un programa estadístico en una computadora se obtendrá el valor de H con la probabilidad exacta de error tipo I asociada a ese valor.

Debe señalarse que la potencia-eficiencia de esta prueba comparada con la prueba paramétrica más poderosa, la prueba *F*, considerando que se cumplen los supuestos para la misma es de un 95.5% (PRIA, 2001, cap. 3).

8.8 PROBLEMAS DE APLICACIÓN DE METODOS NO PARAMETRICOS

1. En relación al ejercicio 1 de la sección 3.5.2 (cap. 3), que se refiere a la determinación del tanto por ciento del níquel de una muestra particular de acero de referencia del NIST por un nuevo método espectrofotométrico, determinar mediante la prueba del signo para los niveles de significación del 0.01 y 0.05 si μ = 1.12% .

   Datos: 1.10, 1.08, 1.09, 1.12, 1.109

<u>*Solución:*</u>

a) Hipótesis nula: μ = 1.12%
Hipótesis alterna: $\mu \neq$ 1.12%

b) Nivel de significación: α = 0.01 y 0.05

c) Cálculo:
Reemplazando cada valor mayor que 1.12% por un signo positivo y cada valor menor que 1.12% por un signo menos, los cinco valores generados son:

- - - 0 -

d) Criterio: el criterio de decisión se basa en el número signos positivos (X) o en el número de signos negativos (Y). Utilizando el número de signos menos frecuente, se rechaza H_o si la probabilidad de obtener X ó Y o más signos es menor o igual a 0.01 ó 0.05 (dependiendo del nivel de significación dado a la prueba.
Así, para X = 0 que es el *número de signos positivos*, el cálculo de la probabilidad binomial para X $\geq$ 0 con parámetros *n = 4* y *p=0.5* es:

$$P\,(X \geq 0;\ n = 4\ ,\ p = 0.5\ ) = \binom{4}{0} 0.5^{0} (1 - 0.5)^{4-0} = 0.0625$$

Para un nivel de significación del 10%, puede rechazarse la hipótesis nula de que μ = 1.12%, dado que P(=0.0625) < 0.1. Mientras que si se usa un nivel de significación del 5% no existe evidencia como para rechazar H_o. Se observa, en el primer caso, que se coincide con los resultados obtenidos en la prueba paramétrica, mientras que en el segundo caso, no. La prueba *paramétrica* puede volverse más concluyente en la medida que se aumente el número de mediciones.

2. Los siguientes quince datos son mediciones del punto de ebullición de un compuesto de silicio en ºC: 166, 141, 136, 153, 170, 162, 155, 146, 183, 157, 148, 132, 160, 175, 150. Para un nivel de significación de 0.05, utilizar las pruebas del signo y de rangos y signos de Wilcoxon para probar la hipótesis de que μ = 158ºC.

<u>*Solución:*</u>

Prueba del signo

a) Hipótesis nula: μ = 158ºC
Hipótesis alterna: $\mu \neq$ 158ºC

b) Nivel de significación: α = 0.05

c) Cálculo:

Reemplazando cada valor mayor que 158ºC por un signo positivo y cada valor menor que 158ºC por un signo menos, los quince valores generados son:

+ - - - + + - - + - - - + + -

d) Criterio: el criterio de decisión se basa en el número signos positivos (X) o en el número de signos negativos (Y). Utilizando el número de signos menos frecuente, se rechaza H_o si la probabilidad de obtener X ó Y o más signos es menor o igual 0.05 (dependiendo del nivel de significación dado a la prueba).
El signo menos frecuente es el signo más (+) y su número es X = 6, el cálculo de la probabilidad binomial para X ≥ 6, con parámetros *n = 15* y *p = 0.5* es:

$$P\,(X \geq 6;\ n = 15,\ p = 0.5\,) = \sum_{X=0}^{6} \binom{15}{X} (0.5)^{X}\,(0.5)^{15-X} = 0.3036$$

Dado que 0.3036 es mayor que el nivel de significación α de 0.05, no existe evidencia como para rechazar que el punto de ebullición del compuesto de silicio sea de 158ºC.

Suma de rangos y signos de Wilcoxon

a) Hipótesis nula: μ = 158ºC
Hipótesis alterna: $\mu \neq$ 158ºC

b) Nivel de significación: α = 0.05

c) Cálculo:
Comparando cada dato obtenido con 158ºC, se obtiene:

8 -17 -22 -5 12 4 -3 -12 25 -1 -10 -26 2 17 -8

Ordenando los datos de menor a mayor sin tomar en cuenta los signos se tiene:

1 2 3 4 5 8 8 10 12 12 17 17 22 25 26

Colocando los rangos y los signos asociados a cada observación se tiene:

-1 2 -3 4 -5 6.5 -6.5 -8 9.5 -9.5 11.5 -11.5 -13 14 -15

d) Criterio: el criterio de decisión se basa en la suma de los rangos con signo positivo (X) o en la suma de los rangos con signo negativo (Y). Utilizando la suma menor, se rechaza H_o si al compararla con el valor obtenido de la tabla del anexo B8 se observa que es menor o igual que este.
En este problema:

X (suma de rangos con signo positivo) = 47.5
Y (suma de rangos con signo negativo) = 72.5

Por lo que el estadístico de contraste es X = 47.5, de la tabla del anexo B8, se obtiene el valor de 25 para $n = 15$ y $\alpha = 0.05$. Dado que el valor del estadístico de contraste es mayor que el valor leído en la tabla puede concluirse que no existe evidencia significativa al 5% como para afirmar que el punto de ebullición del compuesto de silicio difiere de 158ºC.
Como puede observarse este resultado, coincide con el generado a partir de la aplicación de la prueba del signo.

3. En relación al ejercicio 3 de la sección 3.2.5(cap. 3) que se refiere a la determinación de la homogeneidad de una muestra patrón de cloruros mediante análisis de porciones de material, tomadas en la superficie y en el fondo del contenedor, en el que los datos obtenidos fueron:

% de Cloruros en la superficie	% de cloruros en el fondo
26.32	26.28
26.33	26.25
26.38	26.38
26.39	

A partir de la prueba de la mediana, determinar si existe homogeneidad en el material a un nivel de significación α del 5%.

<u>*Solución:*</u>

a) *Hipótesis nula:* hay homogeneidad en el contenido de cloruros de la muestra patrón, es decir: $Me_{fondo} = Me_{superficie}$

b) *Hipótesis alterna:* no hay homogeneidad en el contenido de cloruros de la muestra patrón, es decir: $Me_{fondo} \neq Me_{superficie}$

c) *Nivel de significación(α):* 0.05

d) *Procedimiento de prueba:*

- El primer paso consiste en ordenar la serie de datos en orden ascendente como si procedieran de la misma muestra, para los datos del ejercicio se tiene:

26.25 26.28 26.32 26.33 26.38 26.38 26.39

- La mediana de la serie anterior es: *26.33*
- El resumen de la información del problema para la prueba de la mediana se presenta en la tabla a continuación:

Valores / mediana	Muestra 1	Muestra 2	Total
Obs. > mediana general	2(a)	2(b)	4
Obs. ≤ mediana general	2(c)	1(d)	3
Totales	4	3	7

- El cálculo del estadístico para el contraste es el siguiente:

$$\chi_o^2 = \frac{n\left(\left|ad - bc\right| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\chi_o^2 = \frac{7\left(\left|2x1 - 2x2\right| - \frac{7}{2}\right)^2}{4x\,3x\,4x\,3} = \frac{15.75}{144} = 0.109$$

e) *Criterio de decisión: si* $\chi_o^2 \geq \chi_{tabulado}^2$ *con un grado de libertad y para un* α *=0.05 (ver anexo C3, tabla de* c^2*) se rechaza la hipótesis nula. Dado que 0.109 < 3.84 (ver anexo C3), no existe evidencia significativa al 5% como para rechazar la hipótesis nula. Por lo que se concluye que no existe evidencia a un nivel de significación de 0.05 de que las medianas de las distribuciones de los valores de cloruro varien entre la superficie y el fondo del contenedor. Como puede observarse este resultado coincide con el obtenido mediante la prueba paramétrica.*

4. En el desarrollo de un nuevo método para la determinación de niveles de alcohol en la sangre, se analizó cinco veces una muestra de sangre, con los resultados siguientes: 64.5, 66.0, 63.9, 65.1 y 64.0 mg/100 ml. El método de análisis estándar aplicado a la misma muestra proporciona los resultados 66.2, 65.8, 66.3, 65.6 mg/100ml. Utilizando las pruebas de U-Mann Whitney y la de suma de rangos de Wilcoxon, probar si los métodos difieren significativamente.

Solución:

Prueba de U-Mann Whitney

a) *Hipótesis nula:* $Me_{\text{método estándar}} = Me_{\text{método nuevo}}$

b) Hipótesis alterna: $Me_{\text{método estándar}} \neq Me_{\text{método nuevo}}$

c) Nivel de significación(α): 0.05

d) Procedimiento de prueba:

- El primer paso consiste en asignar los rangos a los datos de ambas muestras como si todos fueran una sola muestra, de la siguiente forma:

Método nuevo	Rango	Método estándar	Rango
65.5	4	66.2	8
66.0	7	65.8	6
63.9	1	66.3	9
65.1	3	65.6	5
64.0	2		

- La suma de rangos de cada muestra es:

Muestra del nuevo método = ***17***
Muestra del método estándar = ***28***

- El estadístico viene dado por tanto como:

$$T_o = S - \frac{N(N+1)}{2} = 17 - \frac{5(5+1)}{2} = 2$$

Para *S = 17* (menor de las sumas de rangos) y *N = 5* (número de datos para la muestra con la menor suma de rangos).
De esta forma $T_o = 2$

e) Criterio de decisión:

El valor de $W_{\alpha/2,n1,n2}$ obtenido de la tabla del anexo D8 es:

$$W_{0.05,\,4,\,5} = 1$$

Dado que $T_o = 2$ > $W_{0.05,\,4,\,5} = 1$, *no existe evidencia a un nivel de significación del 5% como para rechazar la hipótesis nula, o, en otras palabras, se concluye que no existe evidencia significativa al 5% como para rechazar que los dos métodos producen resultados equivalentes.*

<u>Prueba de Suma de Rangos de Wilcoxon:</u>

La variante de esta prueba con respecto a la prueba de U-Mann Whitney se encuentra únicamente en el estadístico de contraste empleado, pues en la prueba de Suma de rangos de Wilcoxon, se calculan dos estadísticos que vienen dados como:

$$T_1 = S_1 - \frac{N_1(N_1+1)}{2} = 17 - \frac{5(5+1)}{2} = 2$$
$$T_2 = S_2 - \frac{N_2(N_2+1)}{2} = 28 - \frac{4(4+1)}{2} = 18$$

Criterio de decisión:

El valor de $W_{\alpha/2,n1,n2}$ obtenido de la tabla del anexo D8 es:

$$W_{0.05,\ 4,\ 5} = 1$$

Dado que la menor de estas sumas es 2 y supera al valor tabulado de 1, no existe evidencia significativa al 5% como para rechazar H_o.

5. En relación al ejercicio 6 de la sección 3.5.2, de la comparación de los resultados obtenidos en la aplicación de un nuevo método de determinación de Indio en Zinc y sus aleaciones y el valor reportado en el patrón certificado. Estudiar con un nivel de significación del 5%, mediante la prueba de rangos y signos de Wilcoxon, si existe diferencia entre las mediciones con el nuevo método y el valor certificado del patrón. La tabla de datos se presenta a continuación:

Patrón	Observación	Certificado
CRM322	2	3
CRM323	6	5
CRM324	15	16
CRM325	44	46
CRM352	2	3
CRM353	1	3
CRM354	7	10
CRM357	6	3
CRM358	12	7
CRM359	16	16
CRM360	29	30

<u>*Solución:*</u>

a) *Hipótesis nula:* $\mu_d = 0$
b) *Hipótesis alterna:* $\mu_d \neq 0$

c) Nivel de significación(μ): 0.05

d) Procedimiento de prueba:

En la tabla a continuación se presentan en la primera columna las diferencias calculadas para cada par de datos, en la segunda columna se presentan estas diferencias ordenadas en orden ascendente sin tomar en cuenta el signo y en la columna tres se presenta el rango asignado a cada diferencia:

d_i = obs- Cert.	orden	rango
-1	0	1
1	-1	-4
-1	-1	-4
-2	-1	-4
-1	-1	-4
-2	1	4
-3	-2	-7. 5
3	-2	-7.5
5	-3	-9.5
0	3	9.5
-1	5	11

La suma de los rangos con signo positivo y negativo son respectivamente:

X (suma de rangos con signo +) = 25.50

Y (suma de rangos con signo -) = 40.50

Por lo tanto el estadístico de prueba es: X = 25.5

e) Criterio de decisión:

Al comparar el estadístico de prueba con el valor obtenido de la tabla del anexo C8, para n = 11, se observa que X = 25.5 > 10, por lo tanto no existe evidencia de que la mediana (media) de la diferencia sea distinta de cero, es decir, no existe evidencia significativa al 5% de la diferencia entre las mediciones de la concentración del patrón por el nuevo método y valor certificado del mismo.

6. Considere el ejercicio 1 de la sección 4.1.10 (cap. 4), en el que se quiere comparar el trabajo de cuatro analistas de un laboratorio en el ensayo de determinación del % de alcohol metílico en muestras de un producto químico, mediante la técnica de cromatografía líquida de alta resolución (HPLC). Los analistas reportaron los resultados siguientes:

Analista	% de alcohol metílico		
1	84.99	84.02	84.38
2	85.15	85.13	84.88
3	84.72	84.48	85.16
4	84.20	84.10	84.55

Mediante la prueba no paramétrica de Kruskall-Wallis, determinar si el trabajo de los analistas difiere significativamente. Use un nivel de significación de 0.05.

Solución:

f) *Hipótesis nula:* los cuatro analistas trabajan de forma equivalente, es decir: $Me_1 = Me_2 = Me_3 = Me_4$

g) *Hipótesis alterna:* al menos dos analistas trabajan de forma diferente:

h) *Nivel de significación(* α *):* 0.05

i) *Procedimiento de prueba:*

- El primer paso consiste en ordenar la serie de datos en orden ascendente como si procedieran de la misma muestra, luego se establecen los rangos. La asignación de rangos se presenta en la siguiente tabla:

Muestra	Analista 1	Analista 2	Analista 3	Analista 4
1	9	11	7	3
2	1	10	5	2
3	4	8	12	6
$\sum R_{ij} =$	**14**	**29**	**24**	**11**

- El cálculo del estadístico para el contraste es el siguiente:

Como no hay empates, el estadístico de prueba será:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{a} \frac{R_{i.}^2}{n_i} - 3(N+1)$$

$$H = \frac{12}{12x13}\left(\frac{14^2}{3} + \frac{29^2}{3} + \frac{24^2}{3} + \frac{11^2}{3}\right) - 3x13 = 5.46$$

Criterio de decisión: si H > $\chi^2_{\alpha,a-1}$ *se rechaza la hipótesis nula. En este caso* $\chi^2_{0.05,3} = 7.81$ (de tabla del anexo B3). *Dado que 5.46 < 7.81, no existe evidencia significativa al 5% como para rechazar la hipótesis nula, por lo que se concluye que no existe evidencia a un nivel de significación de 0.05, como para afirmar que los analistas trabajan de forma diferente.*

REFERENCIAS BIBLIOGRAFICAS

- **Freund, J.E. y Walpole, R.E. (1990).** Estadística matemática con aplicaciones. Cuarta edición. México: PRENTICE HALL HISPANOAMERICANA S.A.

- **Mendenhall, W. (1990).** Estadística para administradores. Segunda edición. México: Grupo Editorial Iberoamérica.

- **Miller, I. y Freund, J.E. (1986).** Probabilidad y estadística para ingenieros. Tercera edición. México: PRENTICE HALL HISPANOAMERICANA S.A.

- **Miller, J. C. y Miller, J. N. (1993).** Estadística para química analítica. Segunda edición. Wilmington, Delaware, E.U.A: Addison Wesley Iberoamericana, S.A.

- **Pérez, C. (2001).** Técnicas estadísticas con SPSS. España: Pearson Education.

- **Pria Barros, M. C. (2001).** Métodos no paramétricos. [En línea]. Cuba: Universidad de La Habana. Disponible en Word Wide Web: ⟨http://www. Vcl.sld.cy/75 cm/facmedic/ webosalud/materiales/mnoparam. html⟩ [citado 10 de enero 2003]

- **Walpole, R. E. y Myers, R. H. (1992).** Probabilidad y estadística. Cuarta edición. México: Mc GRAW-HILL.

- **Wayne, W. D. (1992).** Bioestadística. Base para el análisis de las ciencias de la salud. Cuarta edición. México: Limusa.

ANEXO A8: METODOS NO PARAMETRICOS DE CORRELACION, REGRESION LINEAL SIMPLE Y BONDAD DE AJUSTE
(PRIA, 2001, cap. 4)

1. COEFICIENTE DE CORRELACION DE SPEARMAN

En muchas ocasiones es importante conocer si dos variables están relacionadas y si lo están, evaluar cuál es la intensidad de dicha relación.
En las estadísticas clásicas la medida de correlación usada es el coeficiente de Pearson que se estima por el método de mínimos cuadrados que requiere:

1. Que las variables estén medidas en una escala continua
2. Si se quiere medir su significación estadística es necesario que estas variables tengan una distribución normal bivariada.

Cuando estos supuestos no se cumplen pueden utilizarse medidas de correlación no paramétricas, que permiten además establecer una correlación existente, determinar su significación estadística.
Existen varios coeficientes de correlación no paramétricos, el que se expone aquí es el coeficiente de correlación de Spearman, que fue la primera estadística de rangos que se desarrolló y que requiere que ambas variables estén medidas al menos en una escala ordinal de forma que las unidades experimentales puedan colocarse en dos series y que las observaciones se hayan extraído aleatoriamente de una población.
Sean *n* observaciones de dos variables aleatorias *x* e *y*, medidas ambas al menos, en escala ordinal de forma simultánea en una misma unidad de observación. Los datos así obtenidos pueden presentarse de la forma siguiente:

Elementos	**Variable**	
	X	**Y**
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
⋮	⋮	
n	x_n	y_n

Cada elemento queda representado por un par de valores de la variables *x* e *y*, es decir por el punto (x_i, y_i), que posteriormente puede ser llevado a un gráfico de dispersión que se utiliza para explorar la relación lineal entre las variables.
A cada valor de la variable *x* e *y* se asignan R_{1j} , R_{2j} respectivamente, convirtiéndose la información en una matriz de rangos como la siguiente:

Individuos	**Variable x**	**Rango de x**	**Variable y**	**Rango de y**	$d_{ij} = R_{1j} - R_{2j}$
1	x_1	R_{11}	y_1	R_{21}	$R_{11} - R_{21}$
2	X_2	R_{12}	y_2	R_{22}	$R_{12} - R_{22}$
3	X_3	R_{13}	y_3	R_{23}	$R_{13} - R_{23}$
⋮	⋮	⋮	⋮	⋮	⋮
N	x_n	R_{1n}	y_n	R_{2n}	$R_{1n} - R_{2n}$

d_{ij} mide las discrepancias que hay entre los rangos de *x* e *y* para cada individuo.
En el caso de existir observaciones de igual valor deben asignarse rangos promedios, por ejemplo si los 3 primeros valores son:

Valores	**15**	**15**	**15**
Lugares	1	2	3
Rangos	(1+2+3)/3=2	(1+2+3)/3=2	(1+2+3)/3=2

La fórmula del coeficiente de correlación de Rangos de Spearman se deduce a partir de la del coeficiente de correlación de Pearson, sustituyendo los valores de las observaciones de cada variable por la de sus rangos, llegándose a obtener la siguiente expresión:

$$R_s = 1 - \frac{6\sum_{i=1}^{n} di^2}{n^3 - n}$$

De esta forma R_s brinda una estimación del grado de asociación entre las variables *x* e *y*.
Debe recordarse que cuando se analiza un coeficiente de correlación se deben considerar dos aspectos:

1. Su magnitud: oscila entre 0 y 1, valores cercanos a 0 significan que no hay relación lineal entre ambas variables, mientras que los valores cercanos a 1 significan que la relación lineal entre las variables es muy intensa.

2. Su sentido. Está dado por el signo + o -, el signo – significa que existe una relación inversamente proporcional; es decir, que a valores altos de una variable corresponden valores bajos de la otra. El signo + significa que existe una relación directamente proporcional entre las variables o sea que a valores altos de una se corresponden valores altos de la otra y viceversa.

Las hipótesis a plantear para resolver la situación descrita son las siguientes:

Hipótesis	**Rechazar Ho Si**	α **más usados**
$H_0 : R_s = 0$ $H_1 : R_s \neq 0$	$-R_s \leq -R_{tabulada}$ o $R_s \geq R_{rabulada}$	0.025 0.005
$H_0 : R_s \geq 0$ $H_1 : R_s < 0$	$-R_s \leq -R_{tabulada}$	0.05 0.01
$H_0 : R_s \leq 0$ $H_1 : R_s > 0$	$R_s \geq R_{tabulada}$	0.05 0.01

Donde $-R_{tabulada}$ y $R_{tabulada}$ son los valores críticos obtenidos en la tabla de *valores críticos para el coeficiente de correlación de rangos de Spearman* R_s, ver tabla F.8.1 al final del anexo, para tamaños de muestra que oscilen entre 4 y 30 y un nivel de significación dado.
Cuando el tamaño de la muestra excede las 30 observaciones puede trabajarse entonces con la aproximación a la distribución normal:

$Z = R_s\sqrt{n-1}$ que se distribuye Normal (0, 1)

Otra forma de rechazar la hipótesis nula es cuando se está trabajando con un programa estadístico entonces se calcula R_s y p que es la probabilidad de error de tipo I asociada a ese valor. En este caso si el valor de p es menor que el α prefijado se rechaza la hipótesis nula.
Según Siegel, si la prueba de correlación de rangos de Spearman se aplica a datos que pueden analizarse adecuadamente por una prueba paramétrica más poderosa, como la prueba del coeficiente de correlación de Pearson en éste caso, su potencia eficiencia sería de cerca de un 91%.

2. METODOS DE REGRESION NO PARAMETRICOS

En muchas ocasiones existe interés en realizar aproximaciones no paramétricas cuando se quiere ajustar una línea recta a un conjunto de puntos. Theil desarrolló dos métodos: el método completo de Theil y el método incompleto de Theil, denominado de esta forma para distinguirlo del otro método y por ser quizá el más simple, el cual se expone a continuación. Theil supone que un conjunto de puntos (x_1, y_1), (x_2, y_2), etc., va a ser ajustado por una recta de la forma y = bx + a.
Para los cálculos Theil desarrolla los siguientes pasos:

1. Disponer los puntos en orden de *x* creciente. Si el número de puntos *x*, es impar, el punto medio, es decir, el valor mediano de *x* se borra, el cálculo simple exige un número par de puntos.
2. Para cualquier par de puntos (x_i, y_i), (x_j, y_j) donde $x_j < x_i$, la pendiente, b_{ij} , de la línea que une los puntos es: $b_{ij} = \dfrac{y_i - y_i}{x_j - x_i}$.

   Se calculan las pendientes b_{ij} para el par de puntos (x_1, y_1) y el inmediatamente posterior al valor mediano de la *x*, para (x_2, y_2) y el segundo punto después del valor mediano de la *x*, y así sucesivamente hasta que se calcula la pendiente para la línea que une el punto inmediatamente anterior a la mediana de *x* con el último punto.

3. Las estimaciones de la pendiente, se disponen en orden creciente y su valor mediano, es la pendiente estimada de la línea recta.

4. Con este valor de *b*, los valores a_i se estiman para cada punto con la ayuda de la ecuación y = bx + a.

5. Las estimaciones de a se ordenan en forma creciente y el valor mediano se elige con la mejor estimación.

3. LA PRUEBA DE BONDAD DEL AJUSTE DE KOLMOGOROV

La bondad del ajuste surge cuando se necesita probar la procedencia de una muestra de observaciones de una distribución completa, como puede ser la distribución normal. La prueba Chi-cuadrada es utilizada para este propósito cuando los datos se presentan como

frecuencias, aunque la prueba normalmente no se utiliza para menos de cincuenta observaciones, y es difícil usarla con datos continuos y ais lados.
El método de Kolmogorov tiene dos aplicaciones frecuentes la *prueba de aleatoriedad y la prueba de normalidad de una distribución*.
Al utilizar la prueba de bondad del ajuste de Kolmogorov se hace una comparación entre alguna función de distribución acumulada y teórica $(F_{T(x)})$, y la función de distribución acumulada de una muestra, $F_{s(x)}$ el estadístico a utilizar es:

$$D = Máx\left|F_{s(x)} - F_{T(x)}\right| .$$

La hipótesis nula se rechaza al nivel de significación α si el valor calculado *D* excede al valor mostrado en la tabla A8-1 al final del anexo, para 1 - α .

TABLA A8-1 VALORES CRITICOS DE D PARA LA PRUEBA DE KOLMOGOROV SMIRNOFF
(MILLER y FREUND, 1986, pág. 542)

Tamaño muestral n	$D_{.10}$	$D_{.05}$	$D_{.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29

* Adaptada de F. J. Massey, Jr., "The Kolgomorov-Smirnov test for goodness of fit," *J. Amer. Statist Ass.*, Vol. 46 (1951), p. 70, con autorización del autor y del editor.

ANEXO B8: VALORES CRITICOS PARA LA PRUEBA DE RACHAS DE WALD WOLFOWITZ

(MILLER y MILLER, 1993, pag. 202)

N	M	A P = 0.05, el número de rachas es significativo si es:	
		menor que	mayor que
2	12-20	3	NA
3	6-14	3	NA
3	15-20	4	NA
4	5-6	3	8
4	7	3	NA
4	8-15	4	NA
4	16-20	5	NA
5	5	3	9
5	6	4	9
5	7-8	4	10
5	9-12	4	NA
5	13-18	5	NA
6	6	4	10
6	7-8	4	11
6	9-12	5	12
6	13-18	6	NA
7	7	4	12
7	8	5	12
7	9	5	13
7	10-12	6	13
8	8	5	13
8	9	6	13
8	10-11	6	14
8	12-15	7	15

ANEXO C8: VALORES CRITICOS DE LA PRUEBA DE RANGOS Y SIGNOS DE WILCOXON

(OSTLE, 1977, pag. 601)

	Nivel de significancia para prueba lateral		
	.025	.01	.005
	Nivel de significancia para prueba bilateral		
n	.05	.02	.01
6	0	—	—
7	2	0	—
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68
*			

***** Para n > 25, T es aproximadamente estándar normal con media n(n + 1)/4
y varianza n(n + 1)(2n + 1)/24

ANEXO D8: VALORES CRITICOS PARA LA PRUEBA DE SUMA DE RANGOS DE WILCOXON - MANN-WITHNEY

(JURAN y GRYNA, 1993, pág. A11-29)

n_2	α para prueba bilateral	α para prueba unilateral	n_1 (muestra más pequeña)											
			1	2	3	4	5	6	7	8	9	10	11	12
3	0.20	0.10		3	7									
	0.10	0.05			6									
	0.05	0.025												
	0.01	0.005												
4	0.20	0.10		3	7	13								
	0.10	0.05			6	11								
	0.05	0.025				10								
	0.01	0.005												
5	0.20	0.10		4	8	14	20							
	0.10	0.05		3	7	12	19							
	0.05	0.025			6	11	17							
	0.01	0.005					15							
6	0.20	0.10		4	9	15	22	30						
	0.10	0.05		3	8	13	20	28						
	0.05	0.025			7	12	18	26						
	0.01	0.005				10	16	23						
7	0.20	0.10		4	10	16	23	32	41					
	0.10	0.05		3	8	14	21	29	39					
	0.05	0.025			7	13	20	27	36					
	0.01	0.005				10	16	24	32					
8	0.20	0.10		5	11	17	25	34	44	55				
	0.10	0.05		4	9	15	23	31	41	51				
	0.05	0.025		3	8	14	21	29	38	49				
	0.01	0.005				11	17	25	34	43				
9	0.20	0.10	1	5	11	19	27	36	46	58	70			
	0.10	0.05		4	9	16	24	33	43	54	66			
	0.05	0.025		3	8	14	22	31	40	51	62			
	0.01	0.005			6	11	18	26	35	45	56			
10	0.20	0.10	1	6	12	20	28	38	49	60	73	87		
	0.10	0.05		4	10	17	26	35	45	56	69	82		
	0.05	0.025		3	9	15	23	32	42	53	65	78		
	0.01	0.005			6	12	19	27	37	47	58	71		
11	0.20	0.10	1	6	13	21	30	40	51	63	76	91	106	
	0.10	0.05		4	11	18	27	37	47	59	72	86	100	
	0.05	0.025		3	9	16	24	34	44	55	68	81	96	
	0.01	0.005			6	12	20	28	38	49	61	73	87	
12	0.20	0.10	1	7	14	22	32	42	54	66	80	94	110	127
	0.10	0.05		5	11	19	28	38	49	62	75	89	104	120
	0.05	0.025		4	10	17	26	35	46	58	71	84	99	115
	0.01	0.005			7	13	21	30	40	51	63	76	90	105