Revista Mexicana de Pediatría

Volumen *Volume* 70 Número *Number* 2 Marzo-Abril *March-April* 2003

Artículo:

Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas

Revista Mexicana de Pediatría

Vol. 70, Núm. 2 • Mar.-Abr. 2003
pp 91-99

Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas

(Non parametric statistical tests synopsis. When are they used?)

Manuel Gómez-Gómez,* Cecilia Danglot-Banck,* Leopoldo Vega-Franco**

RESUMEN

Se describen las pruebas no paramétricas resaltando su fundamento y las indicaciones para su empleo cuando se trata de una sola muestra (Ji cuadrada, binomial, de rachas, Kolmogorov-Smirnov), de dos muestras con datos independientes (U de Mann-Whitney, Kolmogorov-Smirnov, Moses, o de las rachas de Wald-Wolfowitz), de dos muestras con datos pareados (T de Wilcoxon, del signo, McNemar), de varias muestras con datos independientes (H de Kruskal-Wallis, de la mediana) y de varias muestras con datos pareados (Ji cuadrada de Friedman, W de Kendall, Q de Cochran).

Palabras clave: Estadísticas no paramétricas, usos de la estadística, pruebas no paramétricas.

SUMMARY

A description of non parametric tests is done. Emphasis about its usefulness when it is studied one sample (chi square, binomial chi, of runs, Kolmogorov-Smirnov one sample test), two samples with independent data (Mann-Whitney, Kolmogorov-Smirnov of two samples, Moses or Wald-Wolfowitz), two samples with paired data (Wilcoxon, of the sign, McNemar), several samples with independent data (Kruskal-Wallis, of the median), or several samples with paired data (Friedman, Kendall, Cochran) it is done.

Key words: *Non parametric statistics, non parametric test, uses of statistical methods.*

INTRODUCCIÓN

Una manera de definir la estadística es considerándola una serie ordenada de métodos que se ocupan de la recolección, organización, presentación, análisis e interpretación de datos numéricos.[1] Se acostumbra dividirla, según el propósito que se persigue, en: descriptiva e inferencial. La estadística descriptiva se utiliza para describir la frecuencia y distribución de las características (o variables) del objeto en estudio, en tanto que la estadística inferencial se ocupa del proceso metódico para obtener conclusiones válidas de una muestra, con respecto a la población, de manera tal que se le pueda considerar representativa de ella. Para entender cuándo y cómo se usa una o la otra, es preciso entender y definir algunos conceptos básicos de la estadística.

CONCEPTOS GENERALES

A diferencia de la estadística paramétrica, en la que el investigador aspira encontrar en las características de la muestra que ha seleccionado, aquellas que distinguen a la población de donde ésta procede; hay dos formas de actuar: 1) estimar el valor de un parámetro a partir de la muestra, y 2) contrastar si su hipótesis es confirmada en la muestra, poniendo a prueba la hipótesis de las diferencias nulas (Ho), la que de no confirmarse se explica por la hipótesis alterna (H_1), que acepta que esas diferencias existen dentro de cierto margen de probabilidad: cuando son significativas (a nivel de una $p < 0.05$ o < 0.001) se rechaza la hipótesis nula y se acepta la hipótesis alterna.[2]

En estadística se definen como variables a los atributos, rasgos o propiedades de un grupo de elemen-

* Maestría en Ciencias (Epidemiología) Clínica, Centro Médico Nacional "La Raza" IMSS.

** Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México.

tos que toman diferentes valores, magnitudes o intensidades. En el proceso de medición de ellas se les asignan números o códigos de observación. La manera más aceptada para ordenar y cuantificar una variable, propuesta por Stevens,[3] es dividirlas en cualitativas (según su calidad o atributo) o cuantitativas (de acuerdo a la magnitud de su medición). Cuando la variable cualitativa no tiene punto de comparación como el color de los ojos (café, azul, verde, negro) se le denomina variable cualitativa nominal; cuando hay un determinado orden como clase social (alta, media, baja), duración de una enfermedad (aguda, subaguda, crónica), orden en la familia (primero, segundo, tercero, etc.) se le llama variable cualitativa ordinal. Cuando la variable cuantitativa sólo se puede medir en valores enteros: como el número de alumnos, el número de partos, el número de empleados, se le denomina variable cuantitativa discreta, discontinua o de intervalo, mientras que si la variable se puede expresar en fracciones, como peso al nacimiento (3,460 g) o estatura (51.3 cm), se les denominan variables cuantitativas continuas o de razón y puede ser que los datos tengan una distribución normal (sesgo de -0.5 a +0.5 y curtosis de 2 a 4).[4]

Cuadro 1. Valoración de las características de los datos.

1. Determinar el nivel de medida de la variable de interés.
2. Valorar la distribución de las variables.
   - Medidas de tendencia central para cada variable.
   - Sesgo y curtosis para cada variable.
   - Valoración visual de la distribución de los datos.
   - Examinar los diagramas de las probabilidades de la distribución.
   - Si se considera necesario transformar las variables.
   - Ver los resultados de la transformación.
3. Ver la homogeneidad de las varianzas.
4. Ver el tamaño de muestra total y de los subgrupos.
5. Determinar qué prueba estadística paramétrica o no paramétrica es la más adecuada.

Desde sus inicios, las computadoras se han utilizado en el manejo de los datos y en ellas se puede hacer uso de las técnicas estadísticas, por lo que hay paquetes estadísticos entre los cuales el SPSS (Statistical Package for the Social Sciences)® es, quizá, el más usado, con más de tres décadas en el mercado.[5]

El procedimiento estadístico que se usará para el análisis depende de: 1) El tipo de medida de la variable a analizar; 2) La distribución que caracteriza a las mediciones de las variables, la homogeneidad de las varianzas en los grupos de ellas, el impacto de los residuos y el tamaño de la muestra; 3) El poder de la prueba que se usará, es decir, la capacidad de aceptar o rechazar, correctamente, la hipótesis nula.[6] En el *cuadro 1* se presenta una guía para la valoración de los datos estadísticos de carácter cuantitativo.

Cuando se pretende probar una hipótesis respecto a uno o más parámetros de una población que tiende a una distribución normal, las pruebas usadas son las de la estadística paramétrica, como la t de Student.[1] En el *cuadro 2* se presentan las características comunes a estas pruebas paramétricas.[7] Por lo contrario, si los procedimientos estadísticos no requieren plantear inferencias acerca de los parámetros de la población (su media y dispersión) se le conoce como no paramétricos, o de distribución libre (ya que no se hacen suposiciones acerca de la distribución de la población de donde procede la muestra. En el *cuadro 3* se presentan las características que son comunes a las pruebas de hipótesis no paramétricas.[6-10]

Con las pruebas no paramétricas se puede trabajar con muestras pequeñas de datos categóricos u ordinales, independientemente de la distribución de las muestras que se desea contrastar.[6] Moses[8] considera que las pruebas no paramétricas: 1) Son más fáciles de aplicar; 2) Son aplicables a los datos jerarquizados; 3) Se pueden usar

Cuadro 2. Características comunes de las pruebas paramétricas.

1. Independencia de las observaciones a excepción de datos pareados.
2. Las observaciones para la variable dependiente se han obtenido de manera aleatoria de una población con distribución normal.
3. La variable dependiente es medida al menos en una escala de intervalo.
4. Se recomienda un tamaño de muestra mínimo de 30 sujetos por grupo.
5. Los datos son obtenidos de poblaciones que tienen varianzas iguales (una varianza no debe ser el doble o mayor que la otra).
6. Habitualmente las hipótesis se hacen sobre valores numéricos, especialmente el promedio de una población (μ), como ejemplo:
   $Ho: \mu_1 = \mu_2$
   $H_1: \mu_1 \neq \mu_2$
7. Otros posibles requisitos: variable independiente nominal o de intervalo, homocedasticidad (para cada nivel de la variable independiente hay una variación similar de la variable dependiente) y casillas de igual tamaño.

Cuadro 3. Características comunes de las pruebas no paramétricas.

1. Independencia de las observaciones aleatorias a excepción de datos pareados.
2. Pocas asunciones con respecto a la distribución de la población.
3. La variable dependiente es medida en escala categórica.
4. El punto primario es el ordenamiento por rangos o por frecuencias.
5. Las hipótesis se hacen sobre rangos, mediana o frecuencias de los datos.
6. El tamaño de muestra requerido es menor (20 o <).

Cuadro 4. Características de las pruebas no paramétricas.

Ventajas

1. Determinación sencilla. Mediante fórmulas simples de combinación.
2. Fáciles de aplicar. Las operaciones matemáticas son la jerarquización, conteo, suma y resta.
3. Rápidas de aplicar. Cuando las muestras son pequeñas.
4. Campos de aplicación. A grupos mayores de poblaciones.
5. Menos susceptibles a la contravención de los supuestos. Ya que los supuestos son escasos y menos complicados.
6. Tipo de medición requerida. Se pueden utilizar con datos ordinales o nominales.
7. Tamaño de la muestra. Cuando la muestra es < 10 son sencillas, rápidas y sólo un poco menos eficaces. Conforme aumenta el tamaño de la muestra se hacen más laboriosas y tardadas, y menos efectivas.
8. Efectividad estadística. Cuando se satisfacen los supuestos de la prueba no paramétrica son igual de efectivas. Si se satisfacen los supuestos de una prueba paramétrica con muestras pequeñas son un poco menos efectivas y se vuelven menos eficaces a medida que aumenta el tamaño de muestra.

Desventajas

1. Si se puede utilizar una prueba paramétrica y se usa una no paramétrica hay una pérdida de información.
2. En muestras grandes las pruebas no paramétricas son muy laboriosas

Cuadro 5. Pruebas paramétricas y su alternativa no paramétrica.

Tipo de problema	Prueba paramétrica	Prueba no paramétrica
Medidas repetidas		Prueba del signo
2 periodos	t pareada	Wilcoxon
> 2 periodos	ANOVA	Friedman
Muestras independientes		Prueba de la mediana,
2 grupos	t independiente	U de Mann-Whitney
> 2 grupos	ANOVA de una vía	Prueba de la mediana, Kruskal-Wallis
Medidas de asociación	r de Pearson	rho de Spearman, tau de Kendall

Cuadro 6. Resumen de las pruebas estadísticas no paramétricas.

Variable dependiente	Una muestra (Bondad de ajuste)	Muestras relacionadas dos muestras	Muestras relacionadas > 2 muestras	Muestras independientes dos muestras	Muestras independientes > 2 muestras	Pruebas de asociación
Nominal	Binomial χ^2;	McNemar	Q de Cochran	Fisher, χ^2 para 2 muestras independientes	χ^2; Mantel-Haenszel	Coef. phi, Coef. Cramér, Kappa
Ordinal/intervalo	Kolmogorov-Smirnov de 1 muestra, 2 muestras	Del signo, Wilcoxon	Friedman	Mediana, U de Mann-Whitney	Mediana, Kruskal-Wallis	Punto biserial, Rho de Spearman, tau de Kendall

cuando dos series de observaciones provienen de distintas poblaciones; 4) Son la única alternativa cuando el tamaño de muestra es pequeño; y, 4) Son útiles a un nivel de significancia previamente especificado. En el *cuadro 4* se pueden ver las características más importantes de las pruebas no paramétricas.

En términos generales, se puede considerar que aunque la potencia de las pruebas estadísticas paramétricas es mayor que la que ofrecen las pruebas no paramétricas, ya que con ellas es la probabilidad de rechazar la hipótesis nula cuando ésta realmente es falsa (error de tipo II: 1-β), es conveniente comentar que el adecuado tamaño de muestra es un requisito indispensable para aumentar la eficacia de una prueba: a medida que aumenta el tamaño de muestra, disminuye la posibilidad de cometer el error de tipo II.[9] En el *cuadro 5* aparecen algunas de las pruebas paramétricas más usadas y sus alternativas no paramétricas.[6,9,10] y en el *cuadro 6* se resumen las indicaciones de las pruebas estadísticas no paramétricas.

PRUEBAS CON UNA SOLA MUESTRA

Ji cuadrada

Esta prueba de hipótesis se usa para comparar la posible diferencia entre las frecuencias observadas en la distribución de una variable con respecto a las esperadas, en razón de una determinada hipótesis.[4,11-16] Por ejemplo: al comparar los resultados obtenidos con una nueva técnica quirúrgica usada en 255 individuos intervenidos en comparación con la técnica utilizada ordinariamente.

	Invalidez total	Invalidez parcial	Funcionamiento normal	Mejoría funcional
Nueva	31 casos	45 casos	73 casos	106 casos
Tradicional	25.10%	30.87%	29.22%	14.81%

Pasos a seguir

Primero: Planteamiento de hipótesis estadísticas

Ho: Fo = Fe. Las frecuencias observadas son iguales a las frecuencias esperadas

Ha: Fo ≠ Fe. Las frecuencias observadas difieren de las frecuencias esperadas

Segundo: Disposición de ambas distribuciones de frecuencias. Para obtener la distribución de frecuencias esperada (teórica) se aplican los porcentajes de los resultados de la técnica quirúrgica tradicional al total de pacientes.

Distribución de frecuencias observadas y esperadas

	Invalidez total	Invalidez parcial	Funcionamiento normal	Mejoría funcional
Frecuencias observadas (Fo)	31 casos	45 casos	73 casos	106 casos
Frecuencias esperadas (Fe)	25.10% de 255 = 64 casos	30.87% de 255= 79 casos	29.22% de 255 = 74 casos	14.81% de 255 = 38 casos

Tercero: Cálculo del valor de χ^2 mediante la fórmula:

$$\chi^2 = \Sigma \frac{(\text{o-e})^2}{\text{e}}$$

En donde : o = frecuencia observada en una modalidad
e = frecuencia esperada en la misma modalidad

Resultados de las diferencias

	Invalidez total	Invalidez parcial	Funcionamiento normal	Mejoría funcional
(Fo)	31 casos	45 casos	73 casos	106 casos
(Fe)	64 casos	79 casos	74 casos	38 casos
$\frac{(o-e)^2}{e}$	$\frac{(31-64)^2}{64}$ = 17.02	$\frac{(45-79)^2}{79}$ = 14.63	$\frac{(73-74)^2}{74}$ = 0.01	$\frac{(106-38)^2}{38}$ = 121.68

χ^2 = 17.02 + 14.63 + 0.01 + 121.68 = 153.34

Cuarto: Comparación de la χ^2 calculada con el valor crítico que aparece en el cuadro de χ^2 y conclusión respecto a las hipótesis planteadas.

Extracto del cuadro de valores críticos de χ^2

	Niveles de significancia	
Grados de libertad	0.05	0.01
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
..		
100	124.34	135.85

Los grados de libertad se refieren, en esta prueba, al número de modalidades menos una. Como fue de cuatro modalidades (columnas) en el renglón a considerar, son 3 grados de libertad (número de columnas menos uno); así, en la tabla de χ^2 al cruzar renglón de los grados de libertad con las columnas de los niveles de significancia al 0.05 se obtiene un valor crítico de $p = 0.05$ de 7.81 y al 0.01 es de 11.34. En vista de que el valor calculado de la ji cuadrada rebasa, en ambos casos, los valores críticos de las tablas al nivel de 5 % y 1 %, se puede rechazar la hipótesis nula (Ho: Fo = Fe) con una $p < 0.01$.

Conclusión. Existen diferencias en la frecuencia de pacientes ubicados en las diversas modalidades de los resultados obtenidos con la técnica quirúrgica nueva respecto a las frecuencias que se encontrarían en las mismas modalidades si se hubiera aplicado la técnica quirúrgica usual, al menos en el caso de estos 255 pacientes intervenidos.

Prueba binomial

La prueba binomial compara las frecuencias observadas en cada una de las dos categorías de una variable dicotómica con respecto a las frecuencias esperadas bajo una distribución binomial que tiene un parámetro de probabilidad específico que, por defecto, para ambas categorías es 0.5. Para cambiar las probabilidades se puede ingresar una proporción de la prueba para el primer grupo por lo que la probabilidad para el segundo será 1 menos la probabilidad especificada para el primero. La prueba está basada en la distribución binomial, que permite estimar que la probabilidad en una muestra de sujetos que puedan proceder de una población binomial cuyo valor de p y q (donde q es la probabilidad contraria) son similares a los de la población de donde se obtuvo la muestra. Se asume que: 1) Las observaciones son seleccionadas al azar, son independientes y se obtienen de una sola muestra; 2) Los datos son de dos categorías distintas, que se les ha asignado un valor de 1 y 0. Esto quiere decir que si la variable no es dicotómica se deben colapsar los datos en dos categorías mutuamente excluyentes; y, 3) Se debe de especificar la probabilidad de ocurrencia de un evento en la población dada. Esta proporción teórica puede venir de registros públicos, censos o investigaciones previas. La prueba binomial está indicada cuando la variable a ser examinada es dicotómica, es especialmente útil en casos de tamaño de muestra pequeños, que no se cumplen los requisitos de la bondad de ajuste de la Ji cuadrada.

Pasos a seguir

Primero: Planteamiento de hipótesis estadísticas

Ho: $p = po$ Las frecuencias observadas son iguales a las frecuencias esperadas

Ha: $p \neq po$ Las frecuencias observadas difieren de las frecuencias esperadas

Segundo. Conocer el número total de casos observados (N).

Tercero. Conocer la frecuencia de las ocurrencias en cada una de las categorías

Cuarto. Se habla de valores binomiales, con una N de 2-30, k de 0-30 y p desde 0.01 a 0.50.

Quinto. Si la probabilidad asociada con el valor observado de valores aún más extremos, es igual o menor al de alfa se rechaza la hipótesis nula.[12-16]

Alternativa. Debido a que se utilizan sólo datos categóricos no hay opción. Si la variable de la prueba no es dicotómica, por lo que se requiere considerar más de dos categorías, se deberá usar la Ji cuadrada para bondad de ajuste.

Prueba de las rachas

La prueba de las rachas mide hasta qué punto en una variable dicotómica la observación de uno de sus atributos puede influir en las siguientes observaciones; es decir, si el orden de ocurrencia en la observación de uno de los atributos de una variable dicotómica ha sido por azar. [12-16] Una racha es una secuencia de observaciones de un mismo atributo o cualidad. Una serie de datos en los que hay muchas o pocas rachas, hacen pensar que éstas no han ocurrido por azar.

Alternativa. Para probar que dos muestras vienen de poblaciones con las mismas distribuciones, se emplea la prueba de rachas sugerida por Wald-Wolfowitz.

Prueba de Kolmogorov-Smirnov Para una muestra

La prueba se usa para definir si el grado de ajuste de los datos a una distribución teórica: que puede ser con tendencia a la normal, a la de Poisson o exponencial. La prueba Z de Kolmogorov-Smirnov (K-S), se computa a partir de la diferencia mayor (en valor absoluto) entre la distribución acumulada de una muestra (observada) y la distribución teórica. La bondad de ajuste de la muestra permite suponer de manera razonable, que las observaciones pudieran corresponder a la distribución específica.

La contribución de Kolmogorov[17] corresponde al problema relacionado con una sola muestra, mientras que la de Smirnov[18] se ocupa de responder al problema respecto a dos muestras, tratando de probar la hipótesis de igualdad entre las poblaciones de origen de una con respecto a la de la otra. La prueba de K-S no precisa que las observaciones sean agrupadas (como es el caso de la Ji cuadrada). Se usa en cualquier muestra de cualquier tamaño, mientras que la Ji cuadrada requiere muestras con un tamaño mínimo. Esta prueba no se debe usar cuando los parámetros tienen que ser estimados a partir de la población y es útil, especialmente cuando se conoce la estructura en que subyace la distribución de la variable en estudio. Es más poderosa que la Ji cuadrada, especialmente cuando el tamaño de la muestra es pequeño y el nivel de medición de la variable es ordinal. Se considera más poderosa que la Ji cuadrada y que la prueba binomial; requiere que la variable dependiente sea una variable cuantitativa continua.

Alternativa. No hay opción paramétrica. Una alternativa no paramétrica es la prueba de bondad de ajuste de Ji cuadrada.[12-16]

PRUEBAS DE DOS MUESTRAS INDEPENDIENTES

Las pruebas de dos muestras independientes comparan dos grupos de casos con una variable. Hay disponibles cuatro pruebas para ver si las dos muestras independientes (grupos) vienen de la misma población y son la U de Mann-Whitney,[19] la Z de Kolmogorov-Smirnov, las reacciones extremas de Moses[20] y la prueba de rachas de Wald-Wolfowitz.[12-16]

La U de Mann-Whitney es la más popular de las pruebas para el estudio de dos muestras independientes. Es equivalente a la prueba de suma de rangos de Wilcoxon y a la prueba de dos grupos de Kruskal-Wallis. Es la alternativa no paramétrica a la comparación de dos promedios independientes a través de la t de Student. Se utiliza cuando se desea efectuar la comparación de dos grupos en quienes se les ha medido una variable cuantitativa continua que no tiene una distribución normal o cuando la variable es de tipo cuantitativa discreta. Tiene tres asunciones: 1) La variable independiente es dicotómica y la escala de medición de la variable dependiente es al menos ordinal; 2) Los datos son de muestras aleatorias de observaciones independientes de dos grupos independientes, por lo que no hay observaciones repetidas; 3) La distribución de la población de la variable dependiente para los dos grupos independientes comparte una forma similar no especificada, aunque con una posible diferencia en las medidas de tendencia central. Las observaciones de ambos grupos se combinan y acomodan, con el rango promedio en el caso de pares. El número de pares debe ser pequeño en relación al número total de observaciones. Si las poblaciones son idénticas en situación, los rangos deben mezclarse al azar entre las dos muestras. Se calcula el número de veces que una cuenta del grupo 1 precede una cuenta del grupo 2 y el número de veces que una cuenta del grupo 2 precede una cuenta del grupo 1. La U de Mann-Whitney es el número más pequeño de estos dos números.

Alternativas. La alternativa paramétrica es la t de Student para muestras independientes, que es más poderosa que la U de Mann-Whitney cuando se llenan todas las asunciones, mientras que si los datos no se distribuyen normalmente, el tamaño de muestra es pequeño, los grupos son de diferente tamaño, la U de Mann-Whitney es más poderosa, sobre todo cuando las colas de la distribución son grandes y hay la presencia de residuales. Una alternativa no paramétrica que puede ser utilizada, sobre todo si las colas de la distribución no son similares es la prueba de la mediana.

La prueba Z de Kolmogorov-Smirnov está basada en la diferencia absoluta máxima entre la función de distribución acumulada observada para ambas muestras. Cuando esta diferencia es significativamente grande, las dos distribuciones son consideradas diferentes.

La prueba de las reacciones extremas de Moses[20] asume que la variable experimental afecta algunos sujetos en una dirección y otros sujetos en la dirección opuesta. Se prueba las reacciones extremas comparadas a un grupo de control. Esta prueba se enfoca en la distribución del grupo de control y es una medida de cuantos valores extremos del grupo experimental influencian la distribución cuando se combinan con el grupo de control.

La prueba de rachas de Wald-Wolfowitz es una alternativa no paramétrica para contrastar si dos muestras con datos independientes proceden de poblaciones con la misma distribución. Combina y acomoda las observaciones de ambos grupos.

Si las dos muestras son de la misma población, los dos grupos deben distribuirse al azar a lo largo de la clasificación jerárquica. Si hay pocas rachas habla de que se tratan de grupos diferentes mientras que, si hay muchas rachas no hay diferencias significativas en la distribución de los dos grupos.

La prueba Z de Kolmogorov-Smirnov y la prueba de rachas de Wald-Wolfowitz son pruebas más generales que detectan diferencias en la localización y formas de las distribuciones.

PRUEBAS PARA DOS O MÁS MUESTRAS INDEPENDIENTES

Las pruebas para muestras independientes comparan las variables de dos o más series de casos; permiten suponer que las muestras provienen de la misma población. Las más conocidas son la de Kruskal-Wallis,[21-23] la de la mediana,[24] y la de Jonckherrere-Terpstra.[12-16]

La prueba de Kruskal-Wallis o de H es una extensión de la de U de Mann-Whitney; de cierta manera es el equivalente no paramétrico del análisis de varianza de una vía y permite conocer si hay diferencias en las distribuciones de la variable en estudio en las poblaciones. Su aplicación asume: 1) Que los datos provienen de un grupo aleatorio de observaciones; 2) Que la variable dependiente es, al menos, ordinal; 3) Que la variable independiente es nominal, con más de dos niveles; 4) Que las observaciones son independientes dentro de cada grupo y entre los grupos; 5) Que no hay medidas repetidas o categorías de respuestas múltiples; y, 5) que es similar la forma en que la distribución de la variable dependiente dentro de cada uno de los grupos, excepto por la posible diferencia de las medidas de tendencia central en al menos uno de estos grupos. Se utiliza cuando la variable independiente tiene más de dos grupos y la variable dependiente es cuantitativa continua.

Alternativas. La alternativa paramétrica es el análisis de varianza de una vía, en la que se asume la normalidad de la distribución dentro de cada nivel de la variable dependiente y la igualdad de las varianzas entre los niveles de la variable independiente. Las alternativas no paramétricas son la prueba de la mediana, la Ji cuadrada de Mantel-Haenszel y la Ji cuadrada para varias muestras independientes.

La prueba de la mediana está indicada cuando la variable independiente es categórica y la variable dependiente tiene, al menos, un nivel de medida de tipo ordinal, aunque ésta habitualmente es cuantitativa continua, y se desea investigar diferencias entre dos o más grupos con relación a su mediana, sea porque no cumplen las condiciones de normalidad para usar el promedio como medida de tendencia central o porque la variable es cuantitativa discreta.

Se define como mediana al valor que en una serie ordenada de datos deja por debajo de ella a la mitad de los valores y la otra mitad por arriba de ella. Responde a la cuestión de que si dos o más grupos proceden de poblaciones que tienen distribuciones similares. Es especialmente útil cuando los valores exactos de resultados extremos son truncados por abajo o por arriba de cierto punto de corte. También está indicada cuando no hay simetría en la forma de la U de Mann-Whitney. La prueba es directa, fácil de aplicar y es particularmente útil cuando no se conocen los valores exactos de todos los resultados, en especial en los valores extremos. La limitación es que esta prueba considera únicamente dos posibilidades: por arriba o por debajo de la mediana, y no se toma en cuenta el tamaño de la diferencia entre los resultados observados respecto a la mediana, por lo que es menos, es de menor potencia que la U de Mann-Whitney y la H de Kruskal-Wallis.

Alternativas. Hay dos alternativas paramétricas que son: la t de Student, cuando la variable independiente es dicotómica y, cuando la variable independiente tiene más de dos niveles, el análisis de varianza de una vía. Las alternativas no paramétricas son la U de Mann-Whitney y la prueba H de Kruskal-Wallis, las que generalmente se prefieren cuando se conoce el rango exacto de valores de la variable dependiente, ya que se toma en cuenta el tamaño de las diferencias entre los resultados observados y la gran mediana.

Cuando, *a priori*, hay un ordenamiento natural (ascendente o descendente) de las poblaciones, la prueba de Jonckheere-Terpstra es más poderosa.

PRUEBAS DE DOS MUESTRAS DEPENDIENTES

Las pruebas para dos muestras dependientes compara en ellas las distribuciones de dos variables que se asume están relacionadas. Para seleccionar la prueba es preciso conocer el tipo de datos que se tienen. Si los datos son continuos se usa la prueba del signo[1,25] o la prueba de rangos signados de Wilcoxon,[26] pero si los datos son binarios se usa la prueba de McNemar.[27] La prueba del signo[1,2,12-16] es una prueba simple, versátil y fácil de aplicar; puede ser usada para saber si una variable tiende a ser mayor que otra. También es útil para probar la tendencia que siguen una serie de variables ordinales positivas o para una valoración rápida de un estudio exploratorio. La desventaja es que no toma en cuenta la magnitud de la diferencia entre dos variables pareadas: computa las diferencias entre las dos variables para todos los casos y clasifica la diferencia como positiva, negativa o empate. Si las dos variables tienen una distribución similar, el número de diferencias positivas y negativas no diferirá significativamente.

Alternativas. La alternativa paramétrica es la t de Student pareada, aunque a la prueba del signo se considera una eficiencia de 95% al compararla con la t de Student, por lo que esta prueba es particularmente útil cuando el tamaño de las muestras es pequeño o cuando no se cumplen los requisitos que exige una prueba paramétrica, como son que las variables sean nominales o que las distribuciones estén sesgadas. Cuando las variables son, al menos, ordinales, una alternativa no paramétrica es la prueba de rangos signados de Wilcoxon, que permite una mejor valoración de las diferencias cuantitativas entre los pares de observaciones.

Rangos signados de Wilcoxon[1,2,12-16] es una prueba flexible que se puede utilizar en distintas situaciones, con muestras de diferente tamaño y con pocas restricciones. Lo único que se requiere es que la variable sea continua y que sean observaciones pareadas, es decir, que sean sujetos de una misma muestra con medidas pre y posprueba, o bien sujetos que hayan sido pareados bajo criterios bien definidos. Contiene varias asunciones críticas:1) Que los datos sean observaciones pareadas, de una muestra seleccionada al azar u obtenida por pares, o bien mediante sujetos considerados como sus propios controles; 2) Que los datos que se van a analizar sean continuos, o al menos ordinales, dentro y entre las observaciones pareadas; y, 3) Que haya simetría en los resultados de las diferencias con la mediana verdadera de la población.

Para efectuar esta prueba se calculan las diferencias entre los pares de datos y se registran los valores absolutos entre ellas. Luego, los valores absolutos de las diferencias entre las dos variables se ordenan del valor menor al mayor y para finalizar, a cada rango se le da un signo positivo o negativo, dependiendo del signo de la diferencia original. Los signos positivos y los negativos se suman separadamente y se obtienen los promedios. Los pares que no tienen cambio alguno se retiran del análisis. Se usa el valor de Z para probar la hipótesis nula de la no diferencia entre los pares. Si la hipótesis nula es cierta, la suma de los rangos positivos debe ser similar a los rangos negativos. Como la prueba de los rangos signados de Wilcoxon incorpora más información acerca de los datos, es más poderosa que la prueba del signo.

Alternativas. La alternativa paramétrica es la t de Student para muestras pareadas, o relacionadas. Las alternativas no paramétricas son la prueba del signo y la prueba binomial.

Prueba de McNemar. Es especialmente útil cuando se tiene un diseño pre y posprueba, en el que el sujeto sirve como su propio control y la variable dependiente es dicotómica.[27] Se usa cuando hay una situación en la que las medidas de cada sujeto se repiten, por lo que la respuesta de cada uno de ellos se obtiene dos veces: una vez antes y la otra después de que ocurre un evento específico: examina la extensión del cambio de la variable dicotómica antes y después del evento. Si la frecuencia de la respuesta en una dirección es mayor de lo esperado por el azar, se rechaza la hipótesis nula (de que no hay cambio alguno). Tiene cuatro presunciones críticas: 1) Que la variable dicotómica que se va a medir tenga valores asignados para cada nivel (ej: 0 y 1), con el mismo valor en los dos periodos; 2) Que los datos representen frecuencias, no valores; 3) Que las medidas dicotómicas sean observaciones pareadas, de la misma selección aleatoria de sujetos o de sus pares; 4) Que los niveles de la variable dicotómica sean mutuamente excluyentes, lo que significa que un sujeto sólo puede asignarse a un nivel de la variable dicotómica que va a ser examinada en todo el tiempo.

Para efectuar la prueba lo primero es colocar los datos en una tabla de 2 x 2, en la que numéricamente se representen los cambios de cada individuo antes y después de la intervención. Si los datos son categóricos se usa la prueba de homogeneidad marginal; ésta es una extensión de la prueba de McNemar de la respuesta binaria y considera una respuesta multinomial; prueba los cambios en las respuestas que se obtiene y usa la distribución Ji cuadrada. Es útil para reconocer cambios de la respuesta debido a la intervención experimental en diseños antes y después.

Alternativas. No hay alternativa paramétrica. Cuando hay más de dos periodos de colección de datos (ej: preprueba, posprueba y seguimiento) se recomienda la Q de Cochran y si los datos son continuos y ordenados adecuadamente, la alternativa no paramétrica es la prueba del signo o la de Wilcoxon.

PRUEBAS DE VARIAS MUESTRAS RELACIONADAS

Las pruebas para varias muestras relacionadas comparan las distribuciones de dos o más variables. Hay tres pruebas disponibles para comparar las distribuciones de varias muestras relacionadas.

Prueba de Friedman. Es una extensión de la prueba de Wilcoxon para incluir datos registrados en más de dos periodos de tiempo o grupos de tres o más sujetos pareados, con un sujeto de cada grupo que ha sido asignado aleatoriamente a una de las tres o más condiciones.[28,29] La prueba examina los rangos de los datos generados en cada periodo de tiempo para determinar si las variables comparten la misma distribución continua de su origen. Es especialmente útil cuando la variable dependiente es continua pero su distribución se encuentra sesgada.

Alternativas. La contraparte paramétrica es el análisis de varianza intrasujetos, cuando ésta es medida de manera repetida. Se compara con la prueba de F del análisis de varianza y se considera que tiene un poder del 64% cuando son dos series ($k = 2$), de 80% cuando $k = 5$ y llega a ser de 87% cuando $k = 10$.

Prueba W de Kendall. En cierta forma es una normalización de la estadística de Friedman.[30] Se interpreta como el coeficiente de concordancia, que es una medida de acuerdo entre los rangos. Cada caso es una base o rango, y cada variable se considera un artículo o persona a juzgar. Para cada variable se computa la suma de cada línea. Su valor final está comprendido entre 0 (ningún acuerdo) y 1 (acuerdo completo). Tiene las mismas indicaciones que la prueba de Friedman, aunque su uso en investigación ha sido, principalmente, para conocer la concordancia entre rangos, más que para probar que hay una diferencia entre las medianas.

Q de Cochran. Esta prueba es idéntica a la prueba de Friedman, pero se aplica cuando todas las respuestas son binarias.[31-33] Es una extensión de la prueba de McNemar ante la situación de k-muestras. La Q de Cochran prueba la hipótesis de que varias variables dicotómicas que están relacionadas entre sí, tienen el mismo promedio. En observaciones múltiples las variables son medidas en el mismo individuo o en individuos pareados. Tiene la ventaja de examinar cambios en las variables categóricas.

Alternativas. No tiene equivalente paramétrico. Si los datos son continuos se prefiere la prueba de Friedman, en especial cuando el tamaño de muestra es pequeño (< 16) y los datos son ordenados.

CONCLUSIONES

Cuando se usan variables cuantitativas continuas y la media aritmética y desviación estándar de las muestras tienden a tener una distribución normal, con varianzas similares (homogeneidad), y el tamaño de las muestras es suficiente (mayor a 30 casos) se deben utilizar las pruebas estadísticas paramétricas. En caso de que no se cumplan estos requisitos, y sobre todo cuando la normalidad de las distribuciones de la variable en estudio esté en duda y el tamaño de la muestra sea menor a once casos, el empleo de las pruebas no paramétricas está indicado.

Cuando una o varias muestras es menor a 11 casos, la potencia estadística de las pruebas paramétricas y no paramétricas es similar; a medida que aumenta el tamaño de las muestras las pruebas paramétricas aumentan su potencia, por lo que las pruebas no paramétricas están indicadas cuando la muestra sea menor de once o bien cuando hay una muestra mayor pero no se cumplen los requisitos de aplicabilidad de las pruebas paramétricas.

Referencias

1. Daniel WW. *Biostatistics*. A foundation for analysis in the health sciences. 7th ed. New York: John Wiley and Sons Inc, 1999: 658-736.
2. Gómez-Gómez M, Danglot-Banck C, Velásquez-Jones L. Bases para la revisión crítica de artículos médicos. *Rev Mex Pediatr* 2001; 69: 152-9.
3. Stevens SS. On the theory of scales of measurement. *Science* 1946; 103: 677-80.
4. Reynaga-Obregón J. Estadística básica en ciencias de la salud. México: *DEMSA*, 2001.
5. Norusis MJ. *SPSS 10.0.1 for Windows*. Chicago: SPSS Inc, 1999.
6. Pett MA. *Nonparametric statistics for health care research*. Thousand Oaks, Cal: Sage Publications Inc, 1997.
7. Bradley JV. *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall, 1968.
8. Moses LE. Non-parametric statistics for psychological research. *Psychol Bull* 1952; 49: 122-43.
9. Downie NM, Heath RW. *Métodos estadísticos aplicados*. 5ª ed. México: Harla, 1986: 251-65.
10. Siegel S, Castellan NJ. *Estadística no paramétrica aplicada a las ciencias de la conducta*. 4ª ed. México: Editorial Trillas, 1995: 151-7.
11. Pearson ES. The choice of statistical test illustrated on the interpretation of data in a 2 x 2 table. *Biometrika* 1947; 34: 139-67.
12. Ferran-Aranaz M. SPSS para Windows. *Análisis estadístico*. Madrid: Osborne McGraw-Hill, 2001.
13. Armitage P, Berry G. *Estadística para la investigación biomédica*. 3ª ed. Madrid: Harcourt Brace, 1997: 424-43.
14. Visauta-Vinacua B. Análisis estadístico con SPSS para Windows. *Estadística básica*. Madrid: McGraw-Hill, 1997: 238-74.
15. Álvarez-Cáceres R. *Estadística multivariante y no paramétrica con SPSS*. Aplicación a las ciencias de la salud. Madrid: Díaz de Santos, 1996.
16. Pérez-López C. *Técnicas estadísticas y SPSS*. Madrid: Prentice Hall, 2001.
17. Kolmogorov AN. Sulla determinazione empirical di una legge di distribuzione. *Giornale Inst Ital Altuari* 1933; 4: 83-91.
18. Smirnov NV. Estimate of deviation between empirical distribution functions in two independent samples. *Bull Moscow University* 1939; 2: 3-16.
19. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947; 18: 50-60.
20. Moses LE. Nonparametrical statistics for psychological research. *Psychol Bull* 1952; 49: 122-43.
21. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952; 47: 583-621.
22. Kruskal WH. A nonparametric test for the several sample problem. *Ann Mat Stat* 1941; 12:461-3.
23. Kruskal WH. Ordinal measures of association. *J Am Stat Assoc* 1958; 1958; 53: 814-61.
24. Reynaga-Obregón J, Gómez-Gómez M. Análisis estadístico en ciencias de la salud. México: *UNAM*, 2002: 125-8.
25. Clayton D, Hills M. *Statistical models in epidemiology*. Oxford: Oxford University Press, 1996: 246-7.
26. Wilcoxon F. Individual comparison by ranking methods. *Biometrika* 1945; 1: 80-3.
27. McNemar Q. *Psychological statistic*. 4th ed. New York: Wiley, 1969.
28. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc* 1937; 32: 675-701.
29. Friedman M. A comparison of alternative test of significance for the problem of rankings. *Ann Mat Stat* 1940; 11: 86-92.
30. Kendall MG. *Rank correlations methods*. 2th ed. New York: Hafner, 1955.
31. Dawson-Saunders B, Trapp RG. *Bioestadística médica*. 3ª ed. México: Manual Moderno, 2002: 184-5.
32. Cochran WG. The χ^2 test of goodness of fit. *Ann Mat Stat* 1952; 23: 315-45.
33. Cochran WG. Some methods for strengthening the common χ^2 tests. *Biometrics* 1954; 10: 417-51.

Correspondencia:
Dr. Manuel Gómez Gómez
Parque Zoquiapan 25,
Col. del Parque, CP 53398,
Naucalpan de Juárez,
Estado de México
Teléfono: 55 76 56 06
Correo electrónico:
mangomez38@hotmail.com