

¿Qué comparan las pruebas paramétricas?

José Rafael Cuauhtémoc Acoltzin-Vidal^a

He leído cuidadosamente el artículo de Rodolfo Rivas-Ruiz *et al.* titulado “Diferencias de medianas con *U* de Mann-Whitney” (Rev Med Inst Mex Seguro Soc. 2013;51[4]:414-9) y difiero con el concepto expresado, porque las llamadas pruebas paramétricas no comparan los valores obtenidos en las mediciones sino sus distribuciones.

La fórmula de la prueba *t* de Student para una muestra es $\mu - \bar{X}/SEM$ (*standard error of the mean*), que compara valores medios entre la variación esperada del resultado si se analizara otra muestra aleatoria de la misma población experimental.

La primera finalidad de la estadística es sintetizar los datos obtenidos. Para lograrlo, si los datos son numéricos se ordenan de menor a mayor y se cuentan las veces que se repiten cada uno, de tal forma que se acumulan en columnas para formar una gráfica llamada *histograma*. Se unen las cimas de las columnas para formar un polígono de frecuencias. Si la gráfica resultante tiene forma de campana simétrica, la distribución de los datos es normal porque refleja la de la población de referencia, cuyos datos se llaman parámetros, de ahí que se le denomine *curva paramétrica*.

El análisis de los datos de una curva normal requiere el cálculo de una medida de tendencia central llamada *media* (o valor medio), que se obtiene con la suma de todos los datos, cuyo resultado se divide entre el número de observaciones. También se calcula la desviación estándar, que se refiere a la dispersión de los datos, es decir, qué tan alejado está cada dato del valor medio. Como la curva de distribución es simétrica, todos los datos estarán dentro de tres desviaciones estándar por debajo de la media y tres desviaciones estándar por arriba de ella. Para el análisis se compararán los valores obtenidos, tal como se mostró con la fórmula de la prueba *t* de Student.

Si la distribución no adopta una forma de campana simétrica, se dice que la distribución es libre y puede ser de varios tipos: una campana asimétrica con joroba desviada (hacia uno u otro extremo), rectangular (cuando el número de observaciones es igual para cada valor) o fractal (cuando resulta absolutamente irregular). Casualmente, la mayoría de los estudios con muestras pequeñas tiene distribución libre. En ese caso, como la distribución es distinta que la paramétrica se recurre a las pruebas llamadas *no paramétricas*.

Las pruebas de distribución libre no comparan los datos sino la distribución de cada grupo y para lograrlo se reúnen los datos de los grupos que se comparan y se ordenan de menor a mayor. Si se repiten, no se apilan en columnas sino que se repiten en fila. Se identifican los de cada grupo con el color que se les designó o abajo se coloca una letra que represente al grupo al que pertenecen. En seguida se les asigna un rango para identificar el lugar que ocupan en la fila y se coloca encima el número correspondiente. Si un valor se repite, se dice que ha ocurrido una liga (la liga mayor representa la *moda* y puede ocasionar la deformidad de la curva) y como cada integrante de la liga tiene el mismo valor, se suma el rango

Recibido: 15/04/2013

Aceptado: 20/05/2013

correspondiente con el número de repeticiones y se asigna el resultado a cada dato, pero a los siguientes se les asigna el que les tocaría si cada dato hubiera recibido el que le tocaba en orden estricto. Se suman los rangos correspondientes a un grupo y los correspondientes al otro grupo, y cada resultado se integra a la fórmula de la prueba U que Rodolfo Rivas-Ruiz mostró en su publicación. Si el número de observaciones es el mismo en cada grupo, el desarrollo de ambas fórmulas será exactamente igual, con la diferencia de cada rango. Para ilustrarlo pongo en fila los datos aportados en el artículo original:

Rango:	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11
Datos:	5,	6,	7,	9,	10,	12,	15,	18,	24,	25,	25
Grupo:	a,	b,	a,	a,	b,	a,	b,	a,	b,	b,	b

A golpe de vista se aprecia que el color rojo (grupo a) se inclina hacia abajo (hacia el rango número 1) y que el color verde (grupo b) hacia arriba (hacia el número 11).

El complemento es la identificación gráfica de la dispersión de los datos que se obtiene al calcular los recorridos intercuartílicos, que dividen los grupos en cuatro partes iguales y permiten identificar los datos ubicados en 25, 50 (mediana), 75 y 100 %. La gráfica correspondiente se llama de *caja y línea* y permite comparar a golpe de vista las distribuciones. Si las gráficas están totalmente separadas se entenderá que los grupos son distintos, pero si se enciman serán homogéneos según la magnitud del traslape.

Actualmente, con la ayuda de los programas estadísticos para computadora, todos esos cálculos manuales se han vuelto improcedentes, lo que no invalida la conveniencia de conocerlos.

^aMédico cirujano, cardiólogo y maestro en Ciencias Médicas, profesor de estadística en el curso de Maestría en Ciencias Médicas, Universidad de Colima, Colima, México

Comunicación con: José Rafael Cuauhtémoc Acoltzin-Vidal
Correo electrónico: cuauhtemoc_acoltzin@uclm.mx
