

Cómo seleccionar una prueba estadística (Segunda parte)

(Choosing a statistical test. [Second part])

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

La prueba t de Student, en su uso más común, es la prueba de hipótesis acerca de la media de una sola población y también valoración de si las medias de dos grupos son estadísticamente diferentes entre sí, lo que tiene los siguientes supuestos: 1) las observaciones en cada grupo siguen una distribución normal; 2) la desviación estándar en ambas muestras es igual, y 3) independencia. El valor de las observaciones en un grupo no proporciona información alguna sobre el otro grupo.

Es probablemente la prueba estadística más ampliamente usada y conocida.²² Es simple, sencilla, fácil de usar y adaptable a una amplia gama de situaciones.

El análisis de varianza define las diferencias en el número de hospitalizaciones entre tres o más grupos; por ejemplo, derechohabientes en el ISSSTE, IMSS y Seguro Popular; medio socioeconómico alto, medio y bajo.²³

La prueba t de Student pareada. Ésta se usa para comparar las diferencias con relación a los promedios respecto de sí mismos, o variaciones, relacionados con el tiempo o circunstancias diferentes. Se emplea, por ejemplo, para saber si hay diferencia en el número de hospitalizaciones según el empleo de un nuevo medicamento contra la displasia broncopulmonar en pacientes, en una clínica ambulatoria, donde cada paciente sirve como su propio control.

La hipótesis nula es que no existan diferencias en el número de admisiones con el uso del nuevo medicamento. Por otra parte, considera la variación de una prueba en un solo grupo (por ejemplo, antes y después).²⁴

La prueba estadística conocida como chi cuadrada (χ^2) es una prueba cualitativa para saber si las diferencias

entre las frecuencias observadas y las esperadas son o no significativas en uno o más grupos categóricos. La hipótesis nula en esta prueba plantea que no existen diferencias entre las frecuencias esperadas y las observadas, es así que la χ^2 se emplea también para conocer la bondad de ajuste de las observaciones y resume si la discrepancia entre las frecuencias observadas y las esperadas es o no significativa; en cambio, la χ^2 de independencia se usa para conocer la relación entre dos variables de una muestra, y para saber si los dos factores están o no relacionados. La χ^2 usa variables nominales (categóricas) o datos ordinales, en lugar de medias y varianzas.^{3,7,8,25}

En cuanto a la bondad de ajuste de la χ^2 , ésta parte de la suposición planteada en términos de que si se tira una moneda al aire 100 veces, los resultados esperados son que 50 veces sea águila y 50 sol; si a un lado de esta suposición se obtienen 47 águilas y 53 soles, tal divergencia ocurre porque ¿la moneda tiene un sesgo o es por casualidad? La *hipótesis nula* para tal experimento plantea que las frecuencias observadas son cercanas a las frecuencias esperadas. La *hipótesis alterna* es que las frecuencias no sean próximas a lo esperado, así la fórmula:

$\chi^2 = \sum [(O-E)^2/E]$, en donde O es la frecuencia observada y E la frecuencia esperada; en tal caso, para águila: $\chi^2 = (47-50)^2/50 = 0.18$ y para sol: $\chi^2 = (53-50)^2/50 = 0.18$, por lo que la suma de estas «categorías» es $0.18 + 0.18 = 0.36$.

El significado de la bondad de ajuste se conoce calculando los grados de libertad y usando tabla de distribución de χ^2 : si la chi cuadrada calculada es mayor que el de la tabla, se rechaza la hipótesis nula, concluyendo que las predicciones eran incorrectas (en este experimento, los grados de libertad serían $[2 - 1] = 1$). El valor crítico para una χ^2 para este ejemplo es $\alpha = 0.05$, y $v = 1$ es 3.84, o sea que es mayor de la χ^2 estimada 0.36, lo que no rechaza la hipótesis nula y, por lo tanto, el sorteo fue justo.²⁶

* Pediatra-Neonatólogo, Maestría en Ciencias con Énfasis en Epidemiología, UNAM.

** Pediatra-Gastroenterólogo, Editor de la Revista Mexicana de Pediatría.

Es pertinente mencionar que en la prueba de χ^2 aplicada a tablas de contingencia de 2 x 2 no se calculan las frecuencias esperadas:

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

En general, cuando la frecuencia esperada es menor de dos en alguna de las casillas, o éstas tienen un 20% o más de cifras menores a cinco, no se debe usar la χ^2 , y en las tablas de 2 x 2 se recomienda hacer la llamada corrección de continuidad (de Yates): que consiste en restar la mitad de las diferencias entre las frecuencias observadas y las esperadas en el numerador de la χ^2 , antes de elevar al cuadrado, por lo que el valor de la χ^2 disminuye.²⁷

La prueba exacta de Fisher es, en cierta forma, semejante a la χ^2 y se usa al analizar las tablas de contingencia de dos filas y dos columnas, cuando los números en cualquier casilla de la tabla de contingencia, son < 5 .²⁸

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

Se hace la comparación de medias con la prueba de la suma de rangos de Wilcoxon, la cual usa la magnitud de las diferencias entre las mediciones respecto a un supuesto parámetro de ubicación; cabe agregar que esta prueba tiene los siguientes supuestos: 1) que la muestra sea aleatoria; 2) que la variable sea continua; 3) que la población se distribuya de manera simétrica alrededor de la media, y 4) que la escala de medición sea al menos de intervalos; en esencia, esta prueba revela si las medianas son diferentes.²⁹

La prueba de la mediana es para probar la hipótesis nula de que dos muestras independientes fueron extraídas de poblaciones con medianas iguales.⁹

Cuando se desea comparar dos grupos en los que se ha medido una variable cuantitativa continua, cuya distribución no es semejante a la curva normal, o bien cuando la variable en estudio es cuantitativa y discreta, la prueba U de Mann-Whitney es una buena alternativa para comparar dos promedios independientes usando la *t* de Student; esto con los siguientes supuestos: 1) que las dos muestras de tamaño, n y m , usadas para el análisis hayan sido extraídas de manera independiente y en forma aleatoria, de sus respectivas poblaciones; 2) que la escala de medición sea por lo menos ordinal; 3) que la variable de interés sea continua, y 4) si las poblaciones son diferentes, que éstas varíen solamente respecto de sus medianas.³⁰

En la comparación de dos grupos pareados usando sus proporciones, cuando la variable es nominal, la prueba indicada es la de McNemar, que se puede usar cuando se comparan dos procedimientos para medir la misma característica o cuando se contrastan las opiniones de dos expertos.³¹

La comparación de tres o más distribuciones independientes usando la prueba de Kruskal-Wallis se usa cuando la variable no tiene una distribución semejante a la curva normal o es de tipo discreto. Esta prueba supone que hay un diseño con una variable independiente o de índole cualitativa nominal con más de dos modalidades: las que definen a los grupos por pertenecer a conjuntos diferentes y se denominan grupos independientes. Las conclusiones con esta prueba se basan en la comparación de un valor H calculado *versus* un valor H crítico, de acuerdo con si los grupos son uno grande ($n > 5$) y el otro pequeño ($K = 3, 4$ o 5 y $n < 5$).³²

La prueba de Cochran evalúa si tres o más conjuntos con frecuencias o proporciones iguales difieren significativamente entre ellos. La igualdad se fundamenta en las características relevantes de los sujetos, o bien en el hecho de que los mismos son usados en diferentes condiciones. Esta prueba es particularmente aplicable a datos de tipo categórico (en escala nominal) o bien en observaciones ordinales (o de intervalo) dicotómicas.³³ El análisis bilateral de la varianza por jerarquías de Friedman se usa siempre y cuando los datos sean medidos al menos en una escala ordinal y puedan ordenarse en una clasificación bilateral.³⁴

Si se pretende saber si existe una relación entre dos variables, se habla de una correlación, y ésta se calcula estimando su coeficiente de correlación (r), la que puede ser positiva o negativa. Cabe mencionar que la correlación positiva ocurre cuando las puntuaciones altas de una variable se asocian con puntajes altos de la otra, en tanto que una correlación negativa es cuando las puntuaciones altas de una variable están asociadas con puntuaciones bajas en el otro extremo.²⁹

Es pertinente mencionar que el análisis de correlación se usa para cuantificar la magnitud o fuerza de la relación entre variables dependientes (por ejemplo, el número de admisiones) y las variables independientes (como el número de dosis de medicamentos, la edad, visita al Servicio de Urgencias, etc.). Es pertinente mencionar que el valor de r se encuentra siempre entre -1 y $+1$. Un valor r cerca de $+1$ indica una relación lineal positiva fuerte, en comparación con un valor cercano a 1 , lo que indica una fuerte relación lineal negativa.

La hipótesis nula es, por ejemplo, cuando no hay ninguna relación entre el número de admisiones y el número de dosis de medicamentos, donde la correlación

es de $p < 0.05$, indica que el valor nulo de no relación es rechazado, aceptando la hipótesis alterna de que sí existe una relación. Es conveniente mencionar que el coeficiente de correlación de Pearson (paramétrico) se emplea si se puede asumir que dos variables continuas se distribuyen normalmente;³⁵ de lo contrario, se elige el coeficiente de correlación de Spearman³⁶

La regresión lineal se usa para predecir los cambios en una variable independiente (en caso de una regresión lineal simple) o varias (en caso de una regresión lineal múltiple) que pueden afectar el valor de la variable dependiente.

Para el análisis de la regresión lineal es necesario considerar que la variable dependiente sea continua y con-

Cuadro 2. Flujograma para elegir la prueba estadística adecuada.

Grupos	Objetivos	Distribución normal	Distribución no normal	Binomial
Uno	Variables cualitativas Dos categorías (tablas de 2 x 2)	RM = Casos y controles RMP = Estudio transversal RR = Cohorte y ensayo clínico		
	Diferentes resultados Diferentes categorías de exposición		> 30 casos = χ^2 < 30 casos = Exacta de Fisher	
	Variables cuantitativas Dos categorías, una variable numérica	Prueba t de Student de una muestra	U de Mann-Witney	
	Una variable directa y una variable indirecta	Regresión lineal simple r de Pearson r^2 = Coeficiente de correlación	r de Spearman	
	Una variable directa y dos o más variables indirectas	r de Pearson r^2 = Coeficiente de correlación Regresión lineal múltiple		
	Comparar un grupo con un valor hipotético	t de Student de una muestra	Wilcoxon	χ^2 o prueba binomial
Dos	No pareados	Prueba de t no pareada	U de Mann-Withney	Prueba exacta de Fisher (χ^2 para muestras grandes)
	Pareados	Prueba de t pareada	Wilcoxon	McNemar
Tres o más	No pareados	ANOVA unidireccional	Kruskal-Wallis ANOVA en bloques	Prueba de χ^2
	Pareados	Medidas repetidas ANOVA	Friedman	Q de Cochran
	Cuantificar la asociación entre dos variables	Pearson	r de Spearman	Coeficiente de contingencias
	Predecir el valor de otra variable medida	Regresión lineal o no lineal	Regresión no paramétrica	Regresión logística simple
	Predecir el valor de varias medias o de variables binomiales	Regresión lineal o no lineal múltiple		Regresión logística múltiple

tenga una distribución normal, en tanto que la relación entre una variable dependiente e independiente es lineal, y si tal suposición no se cumple, los resultados pueden ser incorrectos,^{2,9} mientras que para una variable independiente, no es necesario que se distribuya normalmente; sin embargo, para el análisis de correlación, ambas deben distribuirse de manera normal.

En cuanto a la regresión lineal, ésta se usa para expresar, en una ecuación, la relación que hay entre las variables dependiente e independiente; de tal manera que si estamos interesados en saber del número de visitas a un Servicio de Urgencias (variable independiente conocida como variable x), con relación al número de admisiones al hospital (variable dependiente conocida como variable y), el análisis de la regresión lineal se usa para saber su ecuación. De esta forma se estima la ecuación de la línea recta representada como $y = a + bx$, donde los coeficientes a y b son el intercepto de la línea estimada en el eje, ya que y es la pendiente.^{2,9}

En lo que atañe a la regresión logística, ésta es una variante de la regresión lineal que se usa para definir la relación que hay entre dos o más variables, cuando la variable dependiente es dicotómica y las independientes son datos categóricos o continuos;³⁷ a este respecto, en el *cuadro 2* se pueden ver las pruebas estadísticas aplicadas según el tipo de variable. Es así como se puede concluir que cuando se emplean variables cuantitativas continuas, la media aritmética y la desviación estándar de las muestras con una distribución normal, se dice que existe homogeneidad de varianzas, es decir, con varianzas similares, con un tamaño de muestra suficiente (mayor a 30 casos); en tal caso se usan las pruebas estadísticas paramétricas; pero cuando no se cumplen estos requisitos, si la distribución es anormal y el tamaño de muestra es menor a 11 casos, lo que está indicado es emplear pruebas no paramétricas o de libre distribución.

Referencias

22. Lowry R. Concepts and applications of inferential statistics. t-test for two independent samples. In: <http://faculty.vassar.edu/lowry/ch11pt1.html> (acceso_ el 9 11 2012).
23. Moses LE. Statistical concepts fundamental to investigations. *NEJM* 1985; 312: 890-7.
24. Gosset WS. The probable error of mean. *Biometrika* 1908; 6: 1-25.
25. Gómez-Gómez M, Danglot-Banck C, Velásquez-Jones L. Bases para la revisión crítica de artículos médicos. *Rev Mex Pediatr* 2001; 68: 152-159.
26. Reynaga-Obregón J. *Estadística básica en ciencias de la salud*. México: Distribuidora y Editora Mexicana; 1999. pp. 37-52.
27. Daniel WW. *Bioestadística. Bases para el análisis de las ciencias de la salud*. 4ª ed. México: Editorial Limusa Wiley; 2002.
28. Mejía-Aranguré JM, Fajardo-Gutiérrez A, Gómez-Delgado A, Cuevas-Urióstegui ML, Hernández-Hernández DM, Garduño-Espinosa J et al. El tamaño de muestra: un enfoque práctico en la investigación clínica pediátrica. *Bol Med Hosp Infant Mex* 1995; 52: 381-391.
29. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32: 51-63.
30. Álvarez-Martínez H, Pérez-Campos E. Causalidad en medicina. *Gac Med Mex* 2004; 140: 467-472.
31. Armitage P, Berry G. *Estadística para la investigación biomédica*. 3ª ed. Madrid: Harcourt Brace; 1997. pp. 145-64.
32. Dawson-Saunders B, Trapp RG. *Bioestadística médica*. 3ª ed. México: El Manual Moderno; 2002. p. 335-52.
33. Reynaga-Obregón J. *Análisis estadístico en ciencias de la salud*. México: Distribuidora y Editora Mexicana; 2001.
34. Stevens SS. On the theory of scales of measurement. *Science* 1946; 103: 677-680.
35. Hernández-Ávila M, Garrido-Latorre F, López-Moreno S. Diseño de estudios epidemiológicos. *Salud Pública Mex* 2000; 42: 144-154.
36. Velásquez-Jones L, Gómez-Gómez M. Valoración de un artículo biomédico. En: Martínez y Martínez R, editor. *Cómo escribir un texto en ciencias de la salud. Anatomía de un libro*. 2ª ed. México: Manuel Moderno; 2002. p. 93-104.
37. Soyemi K. Choosing the right statistical test. *Pediatr Rev* 2012; 33(5): e38-44.

Correspondencia:
Dr. Manuel Gómez Gómez
Parque Zoquiapan Núm. 25,
Col. Lomas del Parque, 53398,
Naucalpan, Estado de México.
Teléfono: 55 76 56 06
E-mail: mangomez@prodigy.net.mx

Diferencia por litro de leche entre vacas mexicanas y francesas en 1932

Leche de vaca	Proteínas (g)	Grasas (g)	Lactosa (g)
México	22	34	45
Francia	36	40	47

«Como se verá, la leche de vaca de nuestro país es inferior a la leche de vaca de Francia, porque en un litro de leche existen 14 gramos menos de proteínas, 6 gramos menos de grasas y 2 gramos menos de lactosa; y si valoramos las calorías tenemos que un litro de leche de vaca en Francia produce

712.3 calorías

y un litro de leche de vaca en México produce

590.0 calorías

restando se obtiene:

122.3 calorías

¿será que una de ellas estaba «bautizada»?

Es decir, que un litro de leche de vaca en Francia posee 122.3 calorías más que el mismo volumen de leche de vaca en México; por otra parte, la diferencia por onza es la siguiente:

Una onza de leche de vaca en México, 16 calorías, y una onza de leche de vaca en Francia, 20.5 calorías.

Resultando la mexicana en 4.5 calorías inferior a la onza de leche de vaca de Francia.

Me propuse hacer el estudio comparativo anterior, porque es fundamental conocer esto para administrar el alimento adecuado a las necesidades del niño mexicano. Como veremos posteriormente, y lo tendré muy presente al comentar el método de alimentación que aún en los actuales momentos se está empleando para hacer el cálculo del alimento del niño en nuestra patria, si tenerse en cuenta la pobreza alimenticia de la leche de nuestras vacas, como continuaré demostrándolo teórica y prácticamente en los capítulos siguientes».

[Nota extraída de una de las páginas de la *Rev Mex de Pediatría*, en el volumen 2 de 1932].

www.medigraphic.org.mx