

Estadísticas: una breve descripción

Ryan Winters , Andrew Winters y Ronald G. Amedee

Ochsner J . Otoño de 2010; 10 (3): 213–216.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3096219/>

Resumen

El Consejo de Acreditación de Educación Médica para Graduados establece una serie de temas educativos obligatorios que deben abordarse en los programas de residencia y becas. Intentamos proporcionar una introducción sobre algunos de los conceptos estadísticos básicos importantes a tener en cuenta al examinar la literatura médica. No es esencial comprender el funcionamiento exacto y la metodología de cada prueba estadística encontrada, pero es necesario comprender conceptos seleccionados como pruebas paramétricas y no paramétricas, correlación y datos numéricos versus categóricos. Este conocimiento práctico le permitirá detectar irregularidades obvias en los análisis estadísticos que encuentre.

Palabras clave: educación para residentes de ACGME, matemáticas, estadística

INTRODUCCIÓN

El Consejo de Acreditación para la Educación Médica para Graduados establece una serie de temas educativos obligatorios que deben abordarse en los programas de residencia y becas. Uno de estos temas es la bioestadística, y para muchos alumnos es un tema confuso. Intentamos proporcionar una introducción sobre algunos de los conceptos estadísticos básicos importantes a tener en cuenta al examinar la literatura médica.

Ponemos mucha fe en los números. El concepto subyacente de la medicina basada en evidencia requiere “evidencia” de la eficacia o beneficio de un tratamiento, o en detrimento de un efecto secundario o complicación; Esta evidencia viene en forma numérica. Como nuestra profesión depende cada vez más de las matemáticas, debemos equiparnos con una comprensión básica de los métodos estadísticos para interpretar la literatura.

En su forma más básica, las estadísticas son estudios de muestras de población, y el objetivo es aplicar los resultados de estas muestras a poblaciones enteras. Este concepto aparentemente simple requiere varias consideraciones importantes, como el diseño del estudio, la selección de la muestra de estudio y la elección de la prueba estadística. Las aberraciones en cualquiera de estos arrojarán conclusiones erróneas sobre la importancia o aplicabilidad de los resultados. En palabras de Homer Simpson, *"la gente puede llegar a estadísticas para probar cualquier cosa ... el 14% de las personas lo saben"*. Si, al leer un artículo, la pregunta de investigación parece sencilla, la recopilación de datos parece sencilla, pero los autores tienen Utilizó una prueba estadística de la que nunca ha oído hablar, tenga cuidado con las conclusiones presentadas.

Definiciones

Media

La media es el promedio aritmético de todos los valores en un conjunto de datos. A menudo se usa para aproximar la tendencia central de un conjunto de datos, pero es muy susceptible a los valores atípicos. Por ejemplo, el ingreso promedio de Omaha, Nebraska, probablemente sería excesivamente alto debido a Warren Buffett.

Mediana

La mediana es el valor medio de un conjunto de datos cuando el conjunto se ordena cronológicamente. No está influenciado por valores atípicos, pero no da ninguna indicación de los valores reales de los números en un conjunto de datos; No es un promedio. Por lo tanto, es útil para conjuntos de números pequeños y muy sesgados.

Distribución

Cuando los valores de un conjunto de datos se trazan en un gráfico, la forma de la curva resultante define la distribución. La distribución *normal* (también llamada gaussiana) es la curva de campana clásica, y en esta distribución la media = mediana ([Figura](#)). Esto permite realizar varios supuestos matemáticos importantes, lo que permite el uso de pruebas

estadísticas que son muy sensibles. Una de las suposiciones de Fisher establece que cuanto mayor sea el tamaño de la muestra, más se acercará la distribución a una distribución normal. Esta es una razón matemática por la cual los estudios con muestras de gran tamaño son mejores.

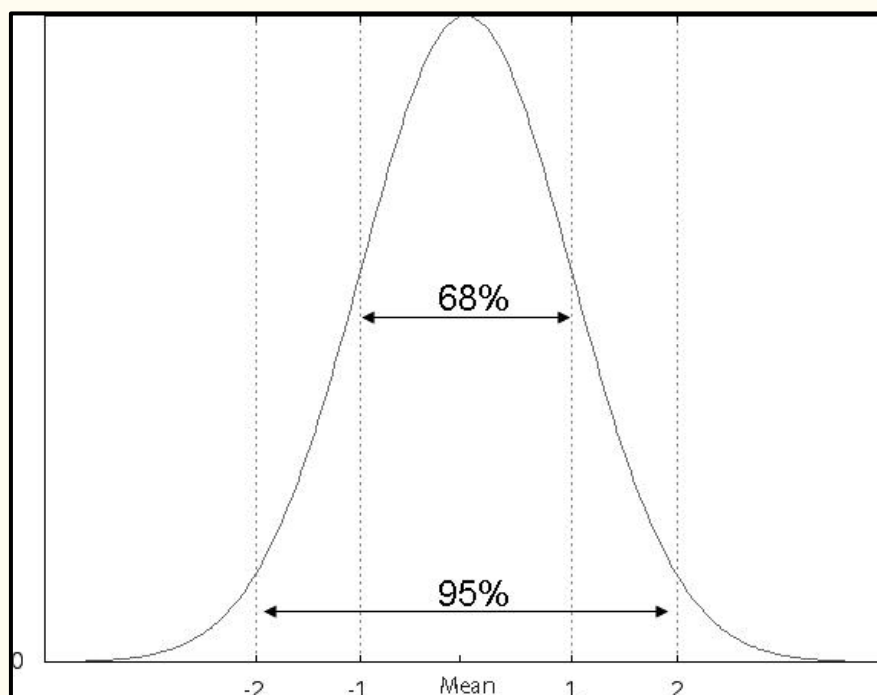


Figura 1. Distribución normal (gaussiana). Media = 0 en este ejemplo. El 1 SD a cada lado de la media abarca ~68% de todos los valores bajo la curva, mientras que 2 SD abarca el 95% de todos los valores bajo la curva.

Desviación estándar

La desviación estándar (σ o SD) indica qué tan lejos de la media se encuentra un valor individual. Para un conjunto de datos normalmente distribuido, aproximadamente el 68% de los valores estarán dentro de 1 SD de la media, y aproximadamente el 95% de los valores estarán dentro de 2 SD ([Figura](#)).

Valor P

El valor P representa la probabilidad de que el resultado observado sea el resultado del azar. Arbitrariamente en las comunidades científicas y médicas, $P < .05$ ha sido elegido como el punto de corte para "estadísticamente significativo". Esto significa que hay una posibilidad menor al 5% (1 probabilidad en 20) de que el resultado observado sea solo por casualidad. Un valor de P fuera del rango significativo podría indicar una de dos cosas. O no hay una diferencia "real" entre los 2 conjuntos de datos, o el tamaño de la muestra era demasiado pequeño para detectar una diferencia entre los 2 conjuntos, incluso si existe. Sin embargo, no se puede determinar por el valor P cuál de estas posibilidades tiene la culpa.

Intervalo de confianza

Un intervalo de confianza (IC) es un rango de valores dentro del cual es bastante seguro que se encuentra el verdadero valor. Esto se basa en la idea de que si repitiera exactamente el mismo estudio en grupos aleatorios de sujetos varias veces, no obtendría exactamente los mismos resultados cada vez. Tendrías un rango. A los efectos de interpretar la literatura médica, se habla con frecuencia del IC del 95%. Lo que esto representa matemáticamente, para un resultado estadístico dado, es que podemos estar 95% seguros de que el valor verdadero se encuentra dentro del rango denotado por el IC (es decir, dentro de 2 SD). Cuanto más estrecho sea el rango del IC, más cercano será cualquier valor observado al valor verdadero, y más preciso será el resultado. Los intervalos de confianza pueden ser números positivos o negativos, y esto no tiene implicaciones. Sin embargo, si el IC del 95% incluye cero y valores de importancia práctica, Valor de P . 1 Esto se aplica a las estadísticas inferenciales y es ligeramente diferente para las razones de probabilidades, que no se analizan en esta revisión. Matemáticamente, cero significa que no hay diferencia entre la muestra y el verdadero valor de la población. Por lo tanto, no se ha excluido la posibilidad de ninguna diferencia entre los 2 grupos, a pesar de $P < .05$. También es fundamental recordar que tales cálculos solo sirven para determinar la validez matemática de los resultados; no determinan la utilidad clínica de dichos resultados. Determinar dicha utilidad es un tema fascinante en sí mismo y está mucho más allá de esta breve descripción de las estadísticas.

TIPOS DE DATOS

Los números se asignan a muchas cosas diferentes en los anales de investigación, pero se dividen en general en 2 categorías, cada una de ellas susceptible de diferentes pruebas estadísticas.

Datos numéricos

El número en sí tiene relevancia. Ejemplos de datos numéricos son cosas que se miden directamente como un número, como los recuentos de CD4, los niveles de lipoproteínas de baja densidad, el contenido de alcohol en la sangre y el tamaño del tumor.

Datos categóricos

Los números asignados a valores no numéricos de interés representan datos categóricos. Ejemplos de datos categóricos incluyen la ciudad de residencia o el sexo de los participantes del estudio.

TIPOS DE PRUEBAS ESTADÍSTICAS

Hay 2 amplias categorías de pruebas estadísticas.

Pruebas paramétricas

Las pruebas paramétricas suponen que los datos de muestra provienen de un conjunto con una distribución particular, típicamente de una distribución normal. Generalmente, esto requiere un gran tamaño de muestra. Como se conoce la distribución (matemáticamente, porque se conoce la forma de la curva), las pruebas paramétricas pueden examinar diferencias absolutas entre valores individuales en una muestra y son más potentes. Pueden identificar diferencias más pequeñas que las pruebas no paramétricas y deben usarse siempre que sea posible.

Pruebas no paramétricas

Las pruebas no paramétricas no hacen tales suposiciones sobre la distribución de datos de origen y, por lo tanto, deben ignorar los valores absolutos de los puntos de datos y centrarse en las propiedades ordinales (por ejemplo, cuál es la más pequeña, la más común). Es más difícil demostrar significación estadística con una prueba no paramétrica (es decir, la diferencia entre los 2 grupos debe ser mayor) que con una prueba paramétrica.

Para cada prueba estadística paramétrica hay un análogo no paramétrico que se utilizará cuando no se puedan cumplir las condiciones para el análisis paramétrico ([Tabla 1](#)).

Tabla 1. Para cada prueba estadística paramétrica hay una prueba no paramétrica análoga

Parametric	Nonparametric
Chi-square test ^a	Fisher exact test
Paired Student <i>t</i> test	Wilcoxon signed rank test
Unpaired Student <i>t</i> test	Mann-Whitney <i>U</i> test
ANOVA by sum of squares	ANOVA by rank
Pearson product moment coefficient	Spearman rank correlation coefficient

ANOVA: analysis of variance.

^a Chi-square is a nonparametric test. Some authors propose thinking of it as parametric, as it works with the sample distribution, mathematically speaking.

PRUEBAS

Chi-cuadrado

Matemáticamente hablando, la prueba de chi-cuadrado (χ^2) es una prueba no paramétrica. Prácticamente, requiere grandes tamaños de muestra y no debe usarse cuando los números son menores que 20. La prueba de ji al cuadrado se usa con datos categóricos, y se deben usar números de conteo reales, no porcentajes o medias. Prueba la distribución de 2 o más conjuntos de datos independientes en comparación con una distribución teórica. Cuanto más parecidas son las distribuciones, más relacionadas están determinadas, y mayor es el valor de chi-cuadrado. Un valor de $\chi^2 = 0$ implica que no hay relación entre las muestras.

Prueba exacta de Fisher

Análogamente a la prueba de ji al cuadrado, la prueba exacta de Fisher es una prueba no paramétrica para datos categóricos, pero se puede usar en situaciones en las que la prueba de ji al cuadrado no puede, como con tamaños de muestra pequeños. Esta prueba se utiliza al comparar 2 conjuntos de datos que crean una tabla de contingencia ([Tabla 2](#)) y prueba la asociación (contingencia) entre los 2 criterios. Esto se observa cuando cada conjunto de datos tiene una respuesta "sí / no", como células tumorales presentes o ausentes, hemocultivos positivos o negativos, muestras de cáncer de mama que son positivas o negativas para el receptor de estrógeno / receptor de progesterona, y así sucesivamente.

Tabla 2. Tabla de contingencia hipotética que compara el estado del receptor de estrógeno / receptor de progesterona (ER / PR) con metástasis ^a

Status	ER/PR+	ER/PR—	Totals
Metastasis+	a	b	a + b
Metastasis—	c	d	c + d
Totals	a + c	b + d	a + b + c + d

^a To form a contingency table, each category must have a yes/no or positive/negative (+/-) answer.

Prueba t de Student

La prueba *t de Student* es probablemente la prueba más utilizada para la significación estadística y es una prueba paramétrica adecuada para datos numéricos o categóricos. La prueba compara las medias de 2 conjuntos de datos para determinar si son iguales; si lo son, entonces no existe diferencia entre los conjuntos. Existe como prueba *emparejada* y *no emparejada* . Una prueba emparejada significa que se midió lo mismo en cada sujeto dos veces. Por ejemplo, midió la frecuencia cardíaca de cada sujeto, administró un betabloqueante, luego midió la frecuencia cardíaca de cada persona nuevamente y desea comparar la diferencia antes y después de la administración del medicamento. Si esto no se hizo, use una prueba no emparejada.

Prueba de rango firmado de Wilcoxon y prueba U de Mann-Whitney

La prueba de rango con signo de Wilcoxon y la prueba *U de Mann-Whitney* son análogos no paramétricos de las pruebas *t* pareadas y no pareadas , respectivamente, en muchas situaciones. Hay escenarios específicos en los que se usan otras pruebas como análogos no paramétricos de la prueba *t de Student* , como cuando se examina el tiempo de supervivencia o cuando se examinan datos categóricos, que están fuera del alcance de esta revisión. Si se encuentran estos escenarios, es aconsejable examinar detenidamente las referencias enumeradas en el estudio para determinar el motivo de la elección de la prueba utilizada. Estas pruebas analizan si la mediana de los 2 conjuntos de datos es igual o si los conjuntos de muestras se extraen de la misma población. Como con todas las pruebas no paramétricas, tienen menos potencia que la contraparte paramétrica (*Student t* prueba) pero se puede usar con muestras pequeñas o datos distribuidos de manera no normal.

Análisis de variación

El análisis de varianza (prueba ANOVA o F) es una generalización de la prueba *t de Student* (o prueba *U de Wilcoxon* o Mann-Whitney) cuando se comparan 3 o más conjuntos de datos. Existen análisis de varianza tanto paramétricos como no paramétricos denominados ANOVA por suma de cuadrados o ANOVA por rango, respectivamente.

Coefficiente de correlación de momento y producto de Pearson

Tal vez el análisis estadístico de correlación de momento del producto de Pearson es el más mal utilizado y mal entendido de todos los análisis estadísticos. Fundamentalmente, la correlación es una desviación de la independencia de 2 o más variables. Esto puede ser en forma de cualquier relación, positiva o negativa. La correlación no es igual a causalidad, ni implica causalidad; simplemente registra el hecho de que 2 o más variables no son completamente independientes entre sí. Coeficiente de Pearson (*r*) es una prueba paramétrica que se puede usar con datos numéricos o categóricos, pero solo tiene sentido en circunstancias muy selectas. Para que sea una medida válida, se deben cumplir los 4 criterios siguientes. (1) Los datos deben distribuirse normalmente. (2) Los 2 conjuntos de datos deben ser independientes entre sí. Un valor no debe variar automáticamente con el otro. Por ejemplo, la cantidad de bebidas consumidas y el nivel de alcohol en la sangre no son independientes entre sí; debe beber alcohol para cambiar su nivel de alcohol en la sangre. (3) Solo se debe realizar un solo par de mediciones en cada tema. (4) Cada valor *r* calculado debe ir acompañado de un valor *P* y / o CI. [2](#)

Coeficiente de correlación de rango de Spearman

La prueba del coeficiente de correlación de rango de Spearman (r_s o ρ) es la contraparte no paramétrica del coeficiente de Pearson y es una buena opción cuando no se pueden cumplir todos los criterios para calcular un valor r significativo y se están examinando datos numéricos. Como prueba no paramétrica, el grado de alejamiento de la independencia tendrá que ser mayor para alcanzar significación usando esta prueba que con la prueba de Pearson. También es más laborioso calcularlo, aunque en la era moderna del software estadístico esto es en gran medida irrelevante.

Análisis de regresión

La regresión cuantifica la relación numérica entre 2 variables que están correlacionadas. Esencialmente, crea una ecuación en la que si se conoce una variable, se puede estimar la otra. Este proceso, llamado *extrapolación*, debe hacerse con cautela, e igualmente interpretarse con cautela. Las numerosas razones para esto están más allá del alcance de esta introducción.

DISCUSIÓN

Al examinar la sección de estadísticas de un artículo, es útil tener un enfoque sistemático. No es esencial comprender el funcionamiento exacto y la metodología de cada prueba estadística encontrada, pero es necesario comprender conceptos seleccionados como pruebas paramétricas y no paramétricas, correlación y datos numéricos versus categóricos. Este conocimiento práctico le permitirá detectar irregularidades obvias en los análisis estadísticos que encuentre. Greenhalgh [2](#), [3](#) propone hacer las siguientes preguntas, entre otras, como primer paso de cualquier artículo: (1) ¿Se evaluó la comparabilidad de los 2 grupos al inicio? (2) ¿La prueba elegida refleja el tipo de datos presentados (paramétrico vs no paramétrico, categórico vs numérico)? (3) ¿Se han analizado los datos de acuerdo con el protocolo original del estudio? (4) Si se utilizó una prueba oscura (esencialmente cualquier prueba no mencionada en esta revisión), ¿se proporcionó una explicación y una referencia?

Para leer más, una excelente serie de artículos de la *Canadian Medical Association Journal* [4](#) - [7](#) profundiza en mayores profundidades y explora algunos conceptos adicionales de estadística. Un poco de conocimiento básico sólido y un enfoque sistemático son invaluableles en el uso de la literatura médica y ayudan a que la idea de "estadísticas" sea un poco menos desalentadora.

Referencias

1. Blume J., Peipert JF Lo que su estadístico nunca le contó sobre los valores P. J Am Assoc Gynecol Laparosc. 2003; 10 ((4)): 439–444. [[PubMed](#)] [[Google Scholar](#)]
2. Greenhalgh T. Cómo leer un documento: estadísticas para el no estadístico, I: diferentes tipos de datos necesitan diferentes pruebas estadísticas [aparece una errata en *BMJ*. 1997; 315 (7109): 675] *BMJ*. 1997; 315 : 364–366. [[Artículo gratuito de PMC](#)] [[PubMed](#)] [[Google Scholar](#)]
3. Greenhalgh T. Cómo leer un documento: estadísticas para el no estadístico, II: relaciones "significativas" y sus dificultades. *BMJ*. 1997; 315 : 422-425. [[Artículo gratuito de PMC](#)] [[PubMed](#)] [[Google Scholar](#)]
4. Guyatt G., Jaeschke R., Heddle N. y col. Estadísticas básicas para médicos: 1. prueba de hipótesis. *CMAJ*. 1995; 152 ((1)): 27–32. [[Artículo gratuito de PMC](#)] [[PubMed](#)] [[Google Scholar](#)]
5. Guyatt G., Jaeschke R., Heddle N. y col. Estadísticas básicas para médicos: 2. interpretación de los resultados del estudio: intervalos de confianza. *CMAJ*. 1995; 152 ((2)): 169-173. [[Artículo gratuito de PMC](#)] [[PubMed](#)] [[Google Scholar](#)]
6. Guyatt G., Jaeschke R., Heddle N. y col. Estadísticas básicas para médicos: 3. evaluación de los efectos del tratamiento: medidas de asociación. *CMAJ*. 1995; 152 ((3)): 351–357. [[Artículo gratuito de PMC](#)] [[PubMed](#)] [[Google Scholar](#)]
7. Guyatt G., Jaeschke R., Heddle N. y col. Estadísticas básicas para médicos: 4. correlación y regresión. *CMAJ*. 1995; 152 ((4)): 497–504. [[Artículo gratuito de PMC](#)] [[PubMed](#)] [[Google Scholar](#)]