

Cómo seleccionar una prueba estadística (Primera de dos partes)

(Choosing a statistical test. First of two parts)

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

RESUMEN

La elección de una prueba en estadística depende: 1) Del tipo de diseño que se usó en la investigación. 2) De la pregunta o preguntas que se hizo el autor. 3) Del interés particular de los autores. 4) De la distribución de los datos obtenidos en uno, dos o más grupos. 5) De la manera como se recopilaron los datos. 6) Del potencial de sesgo y confusión de los datos obtenidos. En esta contribución se examina cómo definir los tipos de datos, la comprobación de las hipótesis planteadas y cómo elegir la prueba estadística más adecuada para plantear una conclusión.

Palabras clave: Estadística paramétrica, estadística no paramétrica, variables, pruebas de hipótesis, pruebas estadísticas.

SUMMARY

The choice of a test in statistics depends on: 1) The type of research design. 2) Which is the question of research. 3) The outcome of interest. 4) Data distribution in one, two or more groups. 5) In the manner of collection of data. 6) The potential of bias and confusion. This article examines the types of data, hypothesis testing and how to choose the most appropriate statistical test.

Key words: Parametric statistics, non-parametric statistics, variables, hypothesis testing, statistical tests.

Los principales diseños en la investigación médica se pueden dividir en dos grupos: los estudios de intervención con pruebas clínicas o experimentales y los estudios observacionales; los que pueden ser descriptivos, como en este caso o serie de casos o una encuesta, un estudio transversal o uno de prevalencia. Los de tipo analítico son estudios transversales; unos con relación a casos clínicos o bajo control terapéutico y eventualmente estudios de cohorte.¹

La presentación de un caso clínico o de una serie de casos son a diseños más «débiles», basados en la descripción y observaciones, en casos particulares; sin embargo, éstos pueden ser la evidencia de una asociación particular que puede generar una hipótesis clínica, terapéutica o etiológica.²

Al revisar el diseño de un estudio planeado de manera apropiada para responder a las preguntas planteadas por el investigador, se debe analizar, en primer lugar, si los objetivos del estudio se definieron de manera correcta y si las hipótesis se formularon de manera clara;³ luego, el investigador debe considerar si el tamaño de la muestra en el estudio es suficiente para responder a la pregunta o preguntas planteadas⁴ y si el diseño ha sido el apropiado para responder a las preguntas hechas por él; teniendo siempre en cuenta las ventajas y desventajas de cada tipo de estudio.³

En ocasiones, cuando se hace una investigación, el investigador puede optar por formar dos grupos: uno de estudio y otro de controles; siempre que sean similares en cuanto a las características particulares y con excepción de la o las variables en estudio.

Otras veces, los investigadores siguen un diseño de casos y controles, o bien optan por un estudio de «cohorte», pero tal caso debe evitar sesgos de selección de casos; lo que acontece cuando los grupos seleccionados son escogidos de tal manera que puedan influir en los

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

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

resultados de la investigación por ser distintos ambos grupos.⁵

Por otra parte, los autores se deben preguntar: ¿Cuál es el significado de sus resultados?, de tal manera que si la magnitud de las diferencias o de la fuerza de asociación en el estudio son significativas, podrán saber si éstas serán útiles y trascendentales; sobre todo si los autores han cumplido con los criterios ligados a una relación de causa-efecto.⁶

SELECCIÓN DE UNA PRUEBA ESTADÍSTICA

La estadística, como parte de las matemáticas empleadas, es la secuencia de razonamientos para estudiar los fenómenos de la naturaleza, y considera que la inferencia estadística es la que permite obtener conclusiones en función de los resultados obtenidos en una muestra en estudio representativa; sin embargo, cabe señalar que hay dos tipos de inferencias: la que se conoce como prueba de hipótesis y la que concierne a la estimación de intervalos.⁷

Y en cuanto a las hipótesis en las pruebas estadísticas, la hipótesis de nulidad (H_0) plantea la ausencia de diferencias significativas, en tanto que la hipótesis alterna (H_1) afirma que hay diferencias en las variables de estudio. Es así como las pruebas de significancia estadística conducen a conocer la magnitud de las diferencias y la significancia de los resultados, considerando un nivel significativo cuando la $p < 0.05$, que se interpreta como rechazo de la hipótesis de nulidad y aceptación de la hipótesis alterna⁸ y de no existir diferencias estadísticas, la $p \geq 0.05$.

Es así como el método estadístico consiste en una secuencia de procedimientos relacionados con el manejo de los datos que proporciona una investigación y comprende la siguiente secuencia:

- 1) Recolección de datos (medición).
- 2) Recuento (cómputo).
- 3) Presentación.
- 4) Descripción.
- 5) Análisis.

En la fase inicial de la investigación se recoge información cualitativa y cuantitativa (variables), lo que se señala desde el diseño del estudio. La segunda etapa concierne a la revisión, clasificación y cómputo numérico de la información recogida. En la siguiente fase (presentación) se elaboran los cuadros y figuras para someter los datos numéricos a un análisis estadístico.

En la descripción, la información es resumida en medidas que permiten expresar de manera sintética las propiedades numéricas de los datos obtenidos; de tal

manera que las medidas son expresadas como medidas de resumen, para sintetizar los datos cualitativos en términos de razones, proporciones y tasas.

En lo que atañe a los datos cuantitativos, las medidas de resumen se sintetizan según que la distribución de los datos sea normal expresándolos en términos de promedio y desviación estándar; en cambio, cuando la distribución de los datos no es normal, la información numérica se expresa en términos de la moda y la amplitud, la mediana y los percentiles. Finalmente, al análisis de las fórmulas estadísticas y el uso de tablas que facilitan la comparación de las medidas de resumen mediante las pruebas estadísticas apropiadas, permiten saber a los investigadores si las variables, motivo de estudio, tienen diferencias significativas.

Cada prueba de análisis se emplea en función del diseño de investigación, que se halla seleccionado para comprobar si la consecuencia es verificable a partir de la hipótesis general de investigación.⁹

Es conveniente mencionar que se definen como variables: las cualidades, rasgos, atributos, características o propiedades de diferentes valores, magnitudes o intensidades valoradas en un grupo de elementos.

En cuanto al manejo estadístico de los datos numéricos, es importante conocer el grado de precisión con el que se obtuvieron las variables, ya que dependiendo de éstas se presenta la información en cuadros o gráficas.

Cabe reiterar que las variables pueden ser de carácter cualitativas y cuantitativas, de tal manera que cuando los datos se pueden observar, pero no son medidos, se trata de una variable cualitativa o categórica, por ejemplo, los colores, texturas, olores, sabores, apariencia, etc., y se clasifican en subgrupos en términos de:

- a) Dicotómicas o binarias, cuando hay sólo dos categorías posibles: «Sí» y «No»; por ejemplo: paciente con enfermedad y paciente sin enfermedad.
- b) Nominal o variables de «atributo» o «categóricas». La observación individual de una variable nominal suele ser considerada una palabra no un número y sigue un orden, por ejemplo: el género (los valores posibles son masculino o femenino), el grupo sanguíneo (los valores son A, B, AB y O).
- c) Ordinal, cuando es posible determinar un orden, como la clase social (alta, media o baja), el orden de nacimiento (primero, segundo, tercero, etc.), el grado escolar (de uno a sexto grado), etcétera.

Cuando se puede cuantificar una variable en enteros, como es el número de hijos, el número de empleados de una fábrica, etc., se le denomina variable cuantitativa discreta (discontinua o de intervalo); y cuando se expre-

sa en fracciones, por ejemplo: la estatura, el peso, etc., se les llama variables cuantitativas continuas (de cociente o de razón). También puede ser que los datos tengan una distribución estadística normal como los valores de sesgo de - 0.5 a + 0.5 y curtosis de 2 a 4.¹⁰

En cuanto a la clasificación de las variables estadísticas se dividen en: variables independientes (como sinónimos de antecedente, causa, origen, razón) y en variables dependientes, como consecuente, efecto, resultado o producto. Es conveniente mencionar que las variables independientes son el antecedente de la variable dependiente o efecto.

Cuando hay una asociación entre dos variables y ésta es real, corresponde a colocar en la celda «a» de la tabla de contingencia de 2 x 2, si no hay asociación entre las variables, esto corresponde a la realidad y se coloca en la celda «d». En ambos casos no se ha cometido error. Si se dice que hay una asociación entre dos variables y no es real se coloca en la celda «b», entonces se comete el error de tipo I, que se representa con el valor $\alpha = 0.05$, y es conocido como el valor crítico de p. En cambio, cuando se dice que la asociación entre las variables es estadísticamente significativa, que hay menos del 5% de posibilidades de error tipo I, lo que en general tiene que ver con el tamaño de la muestra. De no haber una asociación y ésta en realidad existe, corresponde a la celda «c» y se comete un error tipo II, representado con un valor $\beta = 0.20$. Su contraparte ($1-\beta$) es el poder de la muestra, que constituye la sensibilidad mínima requerida para probar alguna diferencia (*Cuadro 1*).¹¹

El procedimiento estadístico aplicable al problema de la significancia estadística de una prueba, va a depender de:

1. El diseño estadístico seleccionado para la investigación.
2. La distribución de casos en uno, dos o más grupos.
3. El tipo de medida o variable a analizar.
4. La manera en que se distribuyan las variables, la homogeneidad de las varianzas en los grupos, el impacto de los residuos y el tamaño de la muestra.
5. El poder de la prueba, es decir, la capacidad para aceptar o rechazar la hipótesis nula.¹²

Cuadro 1. La realidad y sus posibles apreciaciones.

Sí	No
Realidad	Error tipo I
Poder	$\alpha = 0.05$
a	b
c	d
Error tipo II	Confianza
$\beta = 0.20$	Realidad

Las pruebas de significancia estadística son métodos estadísticos que permiten contrastar las hipótesis para valorar los efectos del azar, de acuerdo con los resultados de una investigación.¹³

En los ensayos clínicos y el estudio de cohorte, el riesgo relativo es la medida básica de la fuerza de una asociación, como cuando se mide la magnitud de la asociación al contrastar la incidencia de una enfermedad entre personas expuestas a un factor de riesgo, respecto a las personas no expuestas a éste. La fórmula para saberlo es $(a/a + b)/(c/c + d)$.

En el estudio de casos y controles se emplea la razón de momios o razón de productos cruzados (*Odds Ratio* de la literatura sajona) cuya fórmula es ad/bc , y en el estudio transversal la razón de momios de prevalencia informa de la fuerza de la asociación, y ambas razones tienen una interpretación similar al riesgo relativo.¹⁴

El valor de p (probabilidad asociada con la hipótesis nula de la prueba) es la probabilidad de que un resultado sea debido al azar; entre menor sea la diferencia encontrada entre dos o más grupos hay una mayor significancia estadística (y se considera significativa $p < 0.05$ en adelante); el valor de p tiene distinto significado de acuerdo con la prueba de que se trata, por ejemplo: la t de Student y la prueba Z ponen a prueba la hipótesis de que la diferencia entre dos promedios se debe al azar; la p de una correlación indica la probabilidad de que la pendiente de esa recta sea igual a la pendiente de una recta igual a 0, y no indica qué tan importante es esa correlación, mientras que la p de la prueba exacta de Fisher indica cuál es la probabilidad de que la distribución observada se deba al azar.¹⁵

El intervalo de confianza de 95% indica la imprecisión de las estimaciones puntuales; esto es si se tomó un número infinito de muestras del mismo tamaño que la que se estudió, ¿entre cuál de estos valores (alto y bajo) se encontraría el 95% de las mediciones de esta diferencia o correlación? Si la magnitud de los valores de los dos extremos son menores a la unidad, se considera un efecto protector; y si la magnitud de estos valores de los dos extremos son mayores a la unidad, el resultado es significativo. Si la magnitud de los valores de uno de los extremos es menor a la unidad y el otro es mayor a la unidad, el resultado es no significativo.¹⁶

Los intervalos de confianza y las pruebas de hipótesis llevan a las mismas conclusiones.

Es importante mencionar que el primer paso a seguir es hacer un análisis exploratorio de los datos numéricos para conocer el grado de exactitud, y al mismo tiempo saber si no hay error en la recolección y registro de los datos recabados.¹⁷

Por otra parte, el análisis exploratorio de los datos numéricos corresponde a la estadística descriptiva, la que incluye el cálculo de las medidas de tendencia central: media y mediana, las medidas de dispersión: la desviación estándar y los rangos intercuartiles, según la naturaleza de los datos.

La generalidad de las pruebas estadísticas parten de la suposición de que los datos de una muestra tienen una distribución normal. Estas pruebas son conocidas como pruebas paramétricas como la t de Student y el análisis de varianza (ANOVA).

La llamada curva normal en estadística es un polígono de frecuencias en forma de campana, para el cual se han calculado las áreas de la curva en función de los valores obtenidos del eje horizontal (abscisa). Es en la abscisa donde se encuentran los valores que corresponden a los cuantitativo continuos, genéricamente denominados valores «z», cuyas magnitudes en teoría, van de izquierda a derecha, de menos infinito ($-\infty$) hasta más infinito ($+\infty$).

El promedio de todos los valores zeta de la abscisa equivale a 0, ya que la mitad son positivos y la otra mitad son negativos, y corresponde a la parte más alta de la curva, la que es simétrica alrededor del promedio, es decir, hay una mitad izquierda que es reflejo de la mitad derecha.

En la abscisa hay segmentos unitarios de igual longitud y de magnitud 1, de tal manera que los segmentos a la izquierda del promedio tienen un signo negativo y los segmentos de la derecha tienen un signo positivo. En tales segmentos, la desviación estándar puede dividirse en fracciones infinitamente pequeñas y continuas; la curva es asintótica, es decir, que los extremos en teoría nunca tocan a la abscisa.

Por costumbre se grafica la distancia de tres segmentos a la izquierda y a la derecha del promedio, y toda el área bajo la curva vale 1 (el área a la izquierda del

promedio vale 0.5 y el área a la derecha vale 0.5); sin embargo, el área que se encuentra sobre el segmento de la abscisa, que va desde el promedio hasta el valor z de $+1$ vale 0.3413; por simetría, el área que se encuentra sobre el segmento que va desde el promedio hasta el valor z de -1 de la abscisa también vale 0.3413.

En cuanto al área que se encuentra sobre el segmento de la abscisa, que va más allá del valor z de $+1$ vale 0.1587; por simetría, el área que se encuentra sobre el segmento que va más allá ($-\infty$) del valor z de -1 de la abscisa, también vale 0.1587.

Para cualquier segmento de la abscisa, y aun para fracciones del segmento, se encuentran calculadas las áreas correspondiente en las tablas de valores z .

Así, en el intervalo de una desviación estándar $[\mu - \sigma, \mu + \sigma]$ se comprende, aproximadamente el 68.26% de la distribución, y en el intervalo de dos desviaciones estándar $[\mu - 2\sigma, \mu + 2\sigma]$ se encuentra el 95.44% del área de la curva. Por otra parte, en el intervalo de tres desviaciones estándar $[\mu - 3\sigma, \mu + 3\sigma]$ está comprendida 99.74% de la distribución.

Estas propiedades son de gran utilidad para establecer los intervalos de confianza. El hecho de que prácticamente la totalidad de la distribución esté a tres desviaciones estándar de la media, significa los límites de las tablas de la curva normal estándar (*Figura 1*).¹⁸

Es pertinente mencionar que en la valoración de las características de los datos se debe seguir la siguiente secuencia:

- 1) Conocer el nivel de medida de la variable de interés.
- 2) Saber la distribución de las variables.
 - a) Las medidas de tendencia central para cada variable.
 - b) El sesgo y curtosis para cada variable.
 - c) La apreciación visual de la distribución de los datos.

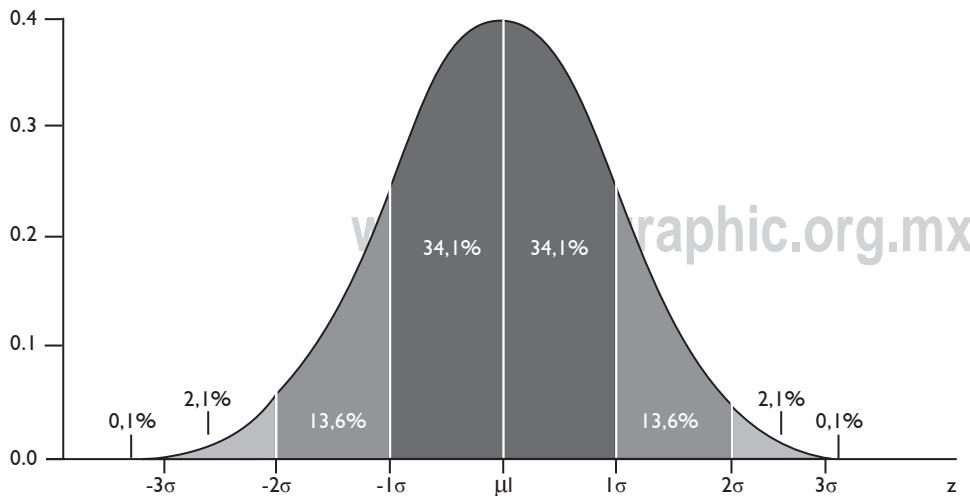


Figura 1.

Principales propiedades de la distribución normal (Campana de Gauss). μ = promedio; σ = desviación estándar.

- d) Examinar los diagramas y las posibilidades de distribución de los datos.
 - e) Si se considera necesario, transformar las variables.
 - f) Ver los resultados de la transformación.
- 3) Ver la homogeneidad de las varianzas.
- 4) Ver el tamaño de la muestra total y de los subgrupos.
- 5) Decidir qué prueba estadística paramétrica o no paramétrica es la más adecuada.

Si el procedimiento estadístico no precisa plantear inferencias de los parámetros de la población (media y dispersión) se le conoce como no paramétrica o de distribución libre, ya que no se hacen suposiciones acerca de la distribución de la población de donde procede la muestra, debido a que en éstas se puede trabajar con muestras pequeñas de datos categóricos u ordinales, independientemente de la distribución de muestras con las que se desea contrastar.¹⁹

Las principales ventajas de las pruebas no paramétricas son:

- 1) Sencillas, mediante fórmulas simples.
- 2) Fáciles de aplicar, con operaciones de jerarquización, conteo, suma y resta.
- 3) Pequeñas, son fáciles de usar.
- 4) Se aplican a los grupos mayores de poblaciones.
- 5) Son menos susceptibles a la contravención de los supuestos, ya que son escasos y sencillos.
- 6) Se pueden usar con datos ordinales o nominales.
- 7) Cuando la muestra es menor de 10 son sencillas, rápidas y sólo poco menos eficaces; conforme aumenta el tamaño de la muestra se hacen laboriosas, tardadas y menos efectivas.
- 8) Cuando se cumplen los supuestos de una prueba no paramétrica son igual de efectivas.

Si se satisfacen los supuestos de una prueba paramétrica con una muestra pequeña, es sólo poco menos efectiva y son menos sensibles a medida que aumenta el tamaño de la muestra. Si se puede usar una prueba paramétrica y se emplea una no paramétrica, hay una pérdida de información. Vale la pena hacer énfasis en que en muestras grandes las pruebas no paramétricas son muy laboriosas.²⁰

Puede el número de niveles de una variable categórica influir sobre el estadístico de prueba que se puede usar, por ejemplo: si se quiere comparar si el género afecta la duración del tiempo de acción de un medicamento. El sexo es una variable cualitativa que tiene dos posibles valores: mujer y hombre.

La frecuencia de administración de un medicamento es una variable dependiente continua y el investigador encuentra que tiene una distribución normal; en este

ejemplo, una prueba paramétrica como la t de Student de dos muestras es una prueba estadística correcta. Si la variable independiente categórica tiene más de dos valores de contraste se debe aplicar la ANOVA unidireccional.²¹

Referencias

1. Gómez GM, Danglot BC, Velásquez JL. Bases para la revisión crítica de artículos médicos. *Rev Mex Pediatr* 2001; 68:152-159.
2. Reynaga OJ. *Estadística básica en ciencias de la salud*. México: Distribuidora y Editora Mexicana; 1999; 37-52.
3. Daniel WW. *Bioestadística. Bases para el análisis de las ciencias de la salud*. México: Editorial Limusa Wiley; 2002.
4. Mejía AJM, Fajardo GA, Gómez DA, Cuevas UML, Hernández HDM, Garduño Ej et al. El tamaño de muestra: un enfoque práctico en la investigación clínica pediátrica. *Bol Med Hosp Infant Mex* 1995; 52: 381-91.
5. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32: 51-63.
6. Alvarez MH, Pérez CE. Causalidad en medicina. *Gac Med Mex* 2004; 140: 467-472.
7. Armitage P, Berry G. *Estadística para la investigación biomédica*. 3^a ed. Madrid: Harcourt Brace; 1997. p. 145-164.
8. Dawson SB, Trapp RG. *Bioestadística médica. El Manual Moderno* 2002; 3: 335-352.
9. Reynaga OJ. *Análisis estadístico en ciencias de la salud*. México: Distribuidora y Editora Mexicana; 2001.
10. Stevens SS. On the theory of scales of measurement. *Science* 1946; 103: 677-680.
11. Hernández AM, Garrido LF, López MS. Diseño de estudios epidemiológicos. *Salud Pública Méx* 2000; 42: 144-154.
12. Velásquez JL, Gómez GM. Valoración de un artículo biomédico. En: Martínez y Martínez R, editor. *Cómo escribir un texto en ciencias de la salud. Anatomía de un libro*. 2^a ed. México: El Manual Moderno; 2002: 93-104.
13. Soyemi K. Choosing the right statistical test. *Pediatr Rev* 2012; 33(5): 38-44.
14. Connor KA. Quantifying associations: understanding Relative Risks and Odds Ratios. *Pediatric Rev* 2012; 33: 473-474.
15. Wacher RN. Utilidad del valor de p y del intervalo de confianza. *Rev Med IMSS (México)* 1993; 31: 143-144.
16. Granhill FA, Bowden DC. Linear segment confidence bands for sample linear models. *J Am Stat Assoc* 1967; 62: 403-408.
17. Windish DM, Diener WM. A clinician-educator's roadmap to choosing and interpreting statistical test. *J Gen Intern Med* 2006; 21(6): 656-660.
18. Galton F. Co-relations and their measurement, chiefly from anthropometric data. *Proc Roy Soc* 1888; 45: 135-145.
19. Gómez GM, Danglot BC, Vega FL. Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. *Rev Mex Pediatr* 2003; 70: 91-99.
20. Bailar JC, Mosteller F. La información estadística que deben proporcionar los artículos publicados en revistas médicas. *Bol Of Sanit Panam* 1990; 108: 317-332.
21. Gómez GM, Danglot BC, Huerta ASG, García de la TG. El estudio de casos y controles: su diseño, análisis e interpretación en investigación clínica. *Rev Mex Pediatr* 2003; 70: 257-263.

Correspondencia:

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