

Investigar y publicar. 3. Cómo realizar un análisis estadístico de una investigación

Research and Publish. 3. How to Conduct a Statistical Analysis of Research

Gerardo Andrés Puentes-Leal.^{1*} 

ACCESO ABIERTO

Citación:

Puentes-Leal GA. Investigar y publicar. 3. Cómo realizar un análisis estadístico de una investigación. *Revista Colomb. Gastroenterol.* 2024;39(3):296-301. <https://doi.org/10.22516/25007440.1265>

¹ Médico Internista y Gastroenterólogo. Especialista en Epidemiología Clínica y Docencia Universitaria, Magister en Economía de la Salud, Centro Hospitalario Serena del Mar, Cartagena de Indias. Cartagena, Colombia.

*Correspondencia: Gerardo Andrés Puentes-Leal. gandrespl@yahoo.com.ar

Fecha recibido: 06/08/2024
Fecha aceptado: 20/08/2024



Resumen

Actualmente la práctica clínica debe integrar la medicina basada en la evidencia. Esto requiere un método científico riguroso que incluye identificar una incertidumbre del conocimiento (el problema), formular una pregunta, realizar una búsqueda sistemática de la evidencia disponible, seleccionar la mejor evidencia válida, interpretarla y aplicarla, y si no existe esta evidencia, el científico (el médico) debe generar ese conocimiento a través de un diseño de investigación. Una parte del proceso metodológico es el análisis estadístico, que a modo de ejemplo consiste en introducir los datos clínicos (la base de datos de un estudio) en una máquina estadística que tiene varias opciones de procesos matemáticos (calcular estadígrafos) para adquirir un resultado el cual debe ser validado, interpretado y analizado de modo que justifique su aplicabilidad real. Esta revisión consiste en contextualizar algunos conceptos básicos estadísticos y propone organizar de una manera sucinta unos pasos para realizar un análisis estadístico, el cual tiene como objetivo motivar a su profundización para generar una visión crítica a la hora de revisar un diseño de investigación.

Palabras clave

Análisis estadístico de datos, validez de los resultados, inferencia estadística, metodología.

Abstract

Currently, clinical practice must integrate evidence-based medicine. This requires a rigorous scientific method that includes identifying a knowledge uncertainty (the problem), formulating a question, conducting a systematic search of the available evidence, selecting the best valid evidence, interpreting and applying it, and if this evidence does not exist, the scientist (the physician) must generate that knowledge through a research design. A part of the methodological process is statistical analysis, which, for example, involves inputting clinical data (the study's database) into a statistical machine that has various mathematical processing options (calculating statistical metrics) to obtain a result that must be validated, interpreted, and analyzed to justify its real-world applicability. This review aims to contextualize some basic statistical concepts and proposes a succinctly organized set of steps for conducting a statistical analysis, which is intended to encourage further exploration to develop a critical perspective when reviewing a research design.

Keywords

Statistical Data Analysis, Validity of Results, Statistical Inference, Methodology.

INTRODUCCIÓN

La investigación clínica debe ser utilizada por los médicos como una herramienta que debe ayudar a disminuir la incertidumbre del conocimiento científico en busca de responder preguntas que pueden resolver problemas relacionados con las condiciones de salud de nuestros pacientes y poder tomar decisiones acertadas en nuestra práctica médica⁽¹⁾. Este proceso, que hace referencia a la evidencia científica sumado al conocimiento médico y a la experiencia clínica, es lo que condiciona la medicina basada en la evidencia⁽²⁾. La investigación requiere de un método científico y si se hace referencia a la investigación cuantitativa, el método consiste en transformar los datos clínicos en conocimiento médico⁽²⁾.

La organización y transformación de estos datos con fórmulas matemáticas es el análisis estadístico, pero la investigación puede arrojar resultados alejados de la realidad; es decir, errados (**Figura 1**). Uno de los problemas que se presentan en las investigaciones es el uso inadecuado de los estadísticos de prueba debido a un desconocimiento de la estadística y al uso inadecuado de una metodología de investigación^(3,4). El riesgo está en que los investigadores o los lectores desconozcan los errores posibles de toda investigación al mostrar resultados y conclusiones, lo que se denomina *falsos positivos* o *falsos negativos* (el error alfa y beta)⁽⁵⁾. La validez interna y externa de la investigación se da cuando se disminuye la probabilidad de tener un error alfa o beta y son aplicables los resultados a nuestra población. Una recomendación básica que se debe tener en cuenta cuando se leen o se realizan investigaciones es establecer la hipótesis que quiere responder la pregunta problema, y lo siguiente es cómo probar esa hipótesis planteada. Para esto se incluye en la investigación cuantitativa el análisis estadístico.

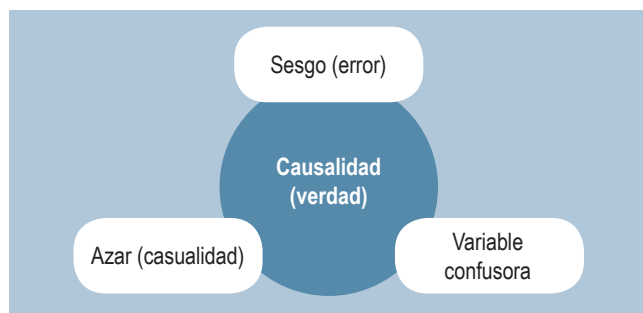


Figura 1. Posibles resultados de un estudio de investigación. Imagen propiedad del autor.

Cada vez más los médicos en la actualidad están obligados a justificar la conducta médica basada en la evidencia científica, y se ve la necesidad de generar investigación clínica, pero no todos son epidemiólogos y no tienen un conocimiento estructurado de estadística, lo que clasifica a los médicos en

dos grupos: los que basan sus decisiones médicas en guías de manejo, artículos y protocolos, y los que tienen un conocimiento básico de la interpretación de la epidemiología clínica y pueden criticar de una manera constructiva la metodología y los resultados de esas guías, protocolos y artículos científicos. Por eso, esta revisión propone repasar unos mínimos que ayuden a criticar de una manera responsable toda la información que aportan los estudios de investigación, y unas bases para poder generar análisis estadísticos para los propios proyectos de investigación.

CONCEPTOS BÁSICOS DESDE LO SIMPLE PARA COMPRENDER EL PROCESO DE UN ANÁLISIS ESTADÍSTICO

Prueba de hipótesis^(6,7)

Un problema requiere de una solución, para esto se plantea una pregunta científica que dé respuesta a la posible solución. La pregunta se plantea con base en una proposición formulada de acuerdo con el conocimiento científico previo y teniendo en cuenta datos. Esa proposición es lo que se llama una *hipótesis científica* y se basa en hechos y conocimiento científico. Existe una hipótesis alterna y una nula, que confirma o refuta la respuesta a la pregunta. La comprobación de la hipótesis es la metodología científica en la que se usan los datos mediante la observación o experimentación.

Estadígrafos

Es una fórmula o cálculo aplicado a los datos de una muestra que da un valor y que tiene una interpretación para describir los datos de esa muestra. Hay estadígrafos descriptivos de tendencia central (media, mediana, moda) y de tendencia de dispersión (desviación estándar, rango y varianza). También hay estadígrafos de asociación que buscan comparar dos grupos de datos entre poblaciones y, de acuerdo con el diseño de investigación, estos pueden ser riesgo relativo (RR), razón de probabilidades (*odds ratio*; OR), cociente de riesgo (HR), entre otros. Es importante aclarar que los estadígrafos se calculan según la naturaleza de la variable (por ejemplo, cualitativa nominal frente a cuantitativa de razón).

El azar⁽⁸⁾

Es un resultado no predecible que se debe a la casualidad. En estadística, el resultado de un estadígrafo no es reproducible ni debido a un algoritmo matemático; en otras palabras, el resultado es aleatorio debido a la suerte. El método científico busca disminuir un resultado al azar a través del

establecimiento de probabilidades de un fenómeno al cruzar variables.

La significancia estadística⁽⁶⁾

Cuando se comparan dos grupos de datos y se obtiene un número que busca su relación o diferenciación (un estadígrafo, por ejemplo, un OR), se aplica una fórmula matemática (estadística) que evalúe si el resultado es por el azar. Si se considera estadísticamente significativo, se asume que hay una probabilidad baja de que el resultado sea por el azar (nunca se puede garantizar en un 100 % la probabilidad de no casualidad en los resultados). Se busca que ese riesgo de azar sea entre el 1 %, 5 % y 10 %, por eso se usa en algunos casos el valor $p < 0,05$ (menor que 5 %). Igualmente, al tener resultados de comparación entre dos grupos se calculan fórmulas matemáticas (estadísticas) que respondan si el estadígrafo obtenido es un valor que tenga una diferencia real y estadística “diferencia estadísticamente significativa”. Es de anotar que esto no siempre significa que esta diferencia sea grande, importante o clínicamente significativa.

Sesgo

Al recolectar, procesar, medir y aplicar fórmulas matemáticas a los datos, buscar asociación entre variables o causalidad se pretende mostrar que los valores representan el fenómeno verdadero; es decir, que son reales. Un sesgo es una desviación en el desarrollo de recolectar, procesar y medir los datos, lo cual va a dar resultados falsos alejados de la realidad. Existen sesgos en todo el proceso de creación de bases de datos y análisis de datos. Antes de realizar análisis estadísticos se deben medir sesgos, cuantificarlos, mitigarlos y predecirlos. Reconocer los sesgos hace parte de la discusión y análisis de resultados para que el lector pueda evaluar si acepta los resultados y los aplica a su práctica clínica dado que considera que los sesgos no influyen en la probabilidad de errar al obtener un resultado o, por el contrario, rechaza la información obtenida por existir un riesgo de obtener resultados errados en la práctica clínica. La calidad de una investigación es indirectamente proporcional a los sesgos de la misma.

Confusión

Cuando se analiza la asociación entre dos variables A y B, asociación que es reproducible y consistente, pero existe una tercera variable C que explica esta asociación, y al controlarla, es decir, retirarla o emparejar entre dos grupos, ya no hay una asociación entre A y B. La variable C es la variable de confusión.

Causalidad

Cuando se confirma que una variable causa un efecto en otra variable y no se debe al azar, está libre de sesgos y no hay una variable de confusión.

Error tipo I (alfa)

La mejor manera de entender este error es cuando la investigación arroja un falso positivo. La conclusión arroja que la hipótesis alterna se confirmó, se acepta. Por ejemplo, la variable A causa B. El factor C previene el desenlace D, entre otros. Al reproducir la metodología del diseño no se obtiene este mismo resultado y, al confirmar la razón, se evidencian sesgos, azar, variables de confusión, entre otros. Se usa la significancia estadística para disminuir este error. “Es tan grave como privar de la libertad al inocente del delito por sentenciarlo como culpable”.

Error tipo II (beta)

Este es el falso negativo. La conclusión del estudio dice que se acepta la hipótesis nula, es decir que no hay causalidad o asociación entre las variables que se quiere probar, pero en la realidad sí existe esta asociación o diferencia. “Es tan grave como dejar en libertad al culpable de un delito”.

Estadística descriptiva

Describe variables de datos de una población. Describe toda la población y sintetiza los datos en gráficas o estadígrafos de tendencia central, dispersión (si son variables cuantitativas) o frecuencias (si son variables nominales).

Estadística inferencial

Cuando no se pueden describir todos los datos de una población, se toman muestras aleatorias (tamaños de muestras), se analizan y se busca representar (inferir) estos resultados a toda la población o extrapolar a otras poblaciones con similares características. Para extrapolar los resultados de un estudio se requiere el uso de fórmulas estadísticas que busquen probabilidades de representación de resultados.

ANÁLISIS ESTADÍSTICO

Es una metodología matemática y estadística que se utiliza para organizar, describir, analizar e interpretar de una manera válida los datos, estableciendo probabilidades para extrapolarlos; es decir, poder utilizarlos para tomar decisiones.

Tipos de análisis estadísticos

En la **Figura 2** se muestran los tipos de análisis.

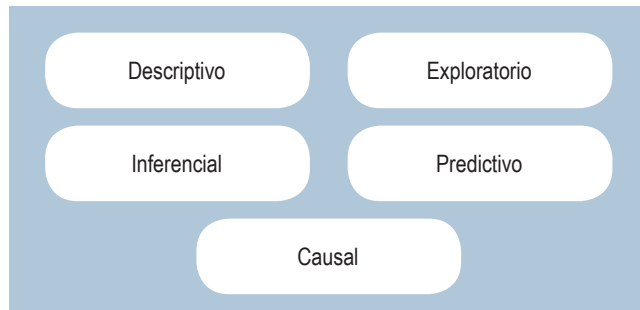


Figura 2. Tipos de análisis estadísticos. Imagen propiedad del autor.

Descriptivo

Se presentan los resultados sintetizados de una base de datos de una muestra en tablas, gráficas y cuadros, y puede plantear hipótesis. Estos resultados no se pueden generalizar a toda la población; sin embargo, se usan en la mayoría de los diseños de investigación y son la base de los estudios observacionales descriptivos.

Exploratorio

Va más allá de lo descriptivo, busca cruzar variables y ver asociaciones (correlaciones) entre observaciones. Busca controlar algunas variables en subgrupos de análisis de la base de datos. Se pueden establecer características demográficas que diferencian dos grupos. Se usan los diseños de investigación observacionales analíticos.

Inferencial

Los resultados descriptivos de una base de datos se someten a pruebas estadísticas que buscan evaluar si se pueden extrapolar a otras poblaciones similares o generales. Igualmente, se analiza la probabilidad de relaciones entre variables de la muestra evaluada y poder confirmar asociaciones y causalidades. Para este análisis se requiere un tamaño muestral significativo y un control de sesgos, además de someter los resultados a descartar probabilidades del azar. Se usan en los diseños de investigación observacionales analíticos y experimentales, así como en las revisiones sistemáticas.

Predictivo

Los resultados analíticos e inferenciales de un diseño de investigación son sometidos a modelos matemáticos que pueden estimar tendencias futuras, comportamientos o probabilidades de resultados al modificar variables. Se usa en diseños de supervivencia y modelos económicos.

Causal

En este análisis se busca confirmar que una variable, además de correlacionarse con otra, es la causa de ese efecto y no debido a una tercera variable de confusión o por al azar. Existen diferentes modelos para establecer causalidad, y el más recomendado es el modelo de Bradford-Hill⁽⁹⁾, que propone el cumplimiento de los siguientes criterios: fuerza de asociación, consistencia, especificidad, temporalidad, gradiente biológico, plausibilidad biológica, coherencia, evidencia experimental, analogía, reversibilidad y juicio crítico. Se usan en los diseños de investigación experimentales y en las revisiones sistemáticas.

Propuesta para realizar un análisis estadístico para estudios cuantitativos

Los pasos para realizar un análisis estadístico en estudios cuantitativos son los siguientes (**Figura 3**):

1. Establecer la pregunta de investigación y el problema: ¿qué quiere investigar el autor? Proponga la hipótesis nula y alterna⁽¹⁾.
2. Revisar la base de datos: explore los datos, las características de las variables (cualitativas o cuantitativas) y organice los datos por variables. Dar una descripción clara y realista de los datos que se tienen. Revisar el diseño de investigación⁽⁴⁾.
3. Evalúe la confiabilidad y validez de la recolección de datos (medir sesgos de información y medición). Evaluar cómo se recolectaron y se midieron, y determinar posibles errores en el registro de los mismos, qué tan cerca a la realidad están de lo que se quiere medir. Analizar cómo se relacionan los datos con los sujetos del estudio⁽¹⁰⁾.
4. Seleccionar el programa estadístico que sea más práctico y confiable según el tipo de análisis estadístico. Existen diferentes programas desde Excel, SPSS, R, Stata, OpenEpi, Epidat, SPSS, entre otros.
5. Realizar análisis estadístico descriptivo a cada variable, por lo general representado en la tabla demográfica de un artículo:
 - distribución de frecuencias (para variables cualitativas);
 - medidas de tendencia central para variables cuantitativas (media, moda y mediana);
 - medidas de variabilidad para variables cuantitativas (rango, desviación estándar y varianza).
6. Realizar análisis estadístico inferencial teniendo en cuenta las hipótesis planteadas, el tamaño de la muestra y su distribución. Este análisis se realiza en los casos en que se quiere mostrar resultados que pueden ser usados por los lectores para tomar decisiones en su práctica

clínica. Diseñar un modelo que considere y describa la relación entre los datos y los sujetos del estudio⁽¹¹⁾:

- **Análisis paramétrico:** se usa cuando se plantean hipótesis con variables cuantitativas de intervalo o razón. Existen diferentes pruebas estadísticas como la prueba t de Student, regresión lineal, coeficientes de correlación, pruebas de diferencia de proporciones como el error estándar Z, chi cuadrado, prueba exacta de Fisher, análisis de varianza y análisis de covarianza).
 - **Análisis no paramétrico:** se usa cuando se plantean hipótesis con variables cualitativas nominales u ordinales. Existen diferentes pruebas estadísticas como la prueba de chi cuadrado de Pearson, chi cuadrado, prueba binomial, prueba de Fisher, coeficientes de Spearman y Kendall, coeficientes para tabulaciones cruzadas, Wilcoxon, entre otras)⁽¹²⁾.
 - **Análisis multivariado:** cuando se quiere evaluar la probabilidad del efecto de más de una variable sobre un resultado. Existen diferentes pruebas estadísticas el modelo log-lineal, regresión lineal múltiple, correlación canónica, ANOVA, MANOVA, ANCOVA y MANCOVA, entre otras)⁽¹³⁾.
7. **Evaluar el modelo para determinar su validez:** se debe considerar que en la base de datos estén todas las variables necesarias para resolver la pregunta, que el método de medición de la variable haya sido el más adecuado y que los estadígrafos o medidas operativas seleccionadas para calcular diferencias o asociaciones sean las que describen mejor el fenómeno a investigar. Al método se le puede aplicar una base de datos diferente, pero

con rangos similares de parámetros, y si los resultados son consistentes y reproducibles, se puede considerar válido el modelo⁽¹⁰⁾.

8. **Interpretar los resultados con el conocimiento científico de base y plantear nuevas hipótesis:** es importante tener en cuenta las limitaciones de los datos y las técnicas utilizadas, así como considerar cualquier otra información relevante que pueda influir en las conclusiones^(2,6,10).
9. **Graficar y presentar resultados:** la selección de la gráfica o algoritmos de resultados no es un proceso menor y tiene como objetivo mostrar los resultados de una manera clara y precisa⁽¹⁰⁾.

CONCLUSIONES

Para que los resultados de un estudio tengan validez interna, se debe realizar un adecuado análisis estadístico, y para esto es fundamental alinear la pregunta de investigación, el diseño de investigación y la base de datos que recolecta las variables a analizar. Cada variable debe interpretarse con cuidado y se debe aplicar la estadística adecuada para su reprocesamiento e interpretación. Siempre se debe considerar qué tan alejadas a la realidad representan los resultados de esta investigación, qué tanto pueden estar influenciadas por sesgos, por variables de confusión o por el azar. Un buen análisis estadístico disminuye la probabilidad de error. El mejor amigo de un médico investigador es un bioestadístico que procese la parte matemática de la base de datos sin perder el concepto clínico de todo el diseño de investigación.

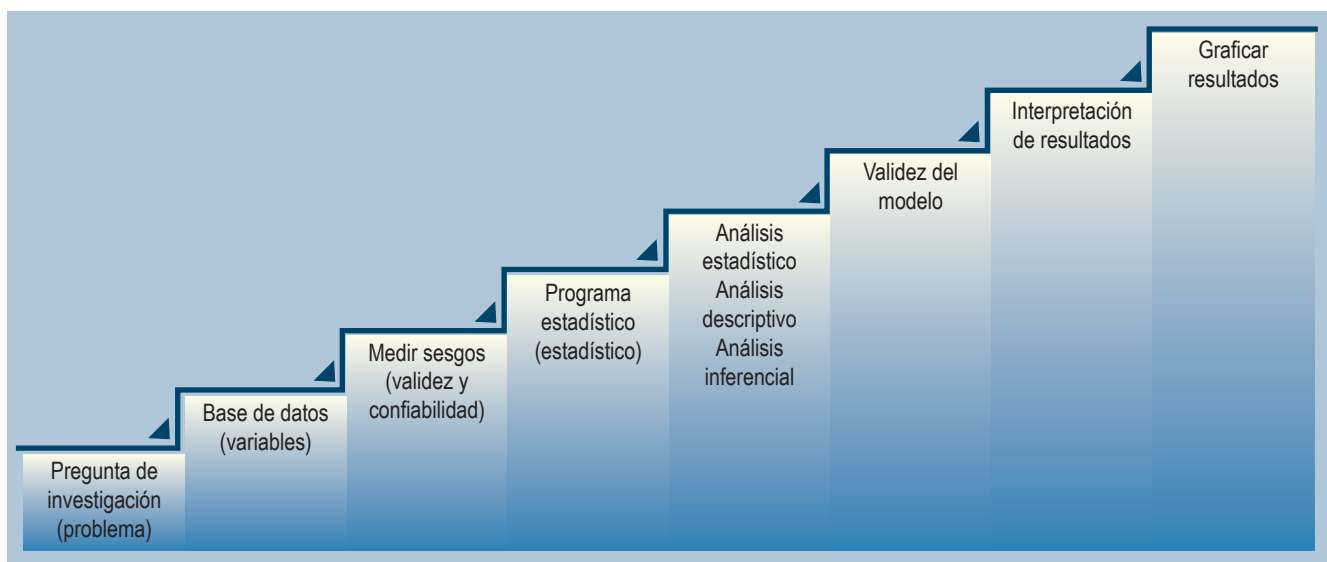


Figura 3. Propuesta para realizar un análisis estadístico para estudios cuantitativos

REFERENCIAS

1. Albis-Feliz R. Investigar y publicar. 1. Cómo formular una pregunta de investigación. *Rev Colomb Gastroenterol*. 2024;39(1):59-61.
<https://doi.org/10.22516/25007440.1174>
2. Vega-de Céniga, M, Allegue-Allegue N, Bellmunt-Montoya S, López-Espada C, Riera-Vázquez R, Solanich-Valldaura T, et al. Medicina basada en la evidencia: concepto y aplicación. *Angiología*. 2009;61(1):29-34.
[https://doi.org/10.1016/S0003-3170\(09\)11004-0](https://doi.org/10.1016/S0003-3170(09)11004-0)
3. Ramírez Ríos A, Polack Peña AM. Estadística inferencial. Elección de una prueba estadística no paramétrica en investigación científica. *Horizonte de la Ciencia*. 2020;10(19):191-208.
<https://doi.org/10.26490/uncp.horizonteciencia.2020.19.597>
4. Rey-Rubiano M. Investigar y publicar. 2. Cómo responder la pregunta “diseño del estudio”. *Rev Colomb Gastroenterol*. 2024;39(2):176-178.
<https://doi.org/10.22516/25007440.1189>
5. Illigens BMW, Lopes F, Fregni F, Brunoni A. Parametric Statistical Tests. En: Fregni F, Illigens B (editores). *Critical Thinking in Clinical Research*. Nueva York: Oxford University Press; 2018.
<https://doi.org/10.1093/med/9780199324491.003.0009>
6. Suemoto CK, Lee C, Fregni F. Basics of Statistics. En: Fregni F, Illigens B (editores). *Critical Thinking in clinical research*. Nueva York: Oxford University Press; 2018.
<https://doi.org/10.1093/med/9780199324491.003.0008>
7. Sánchez M, Marín G, Quintero I. La importancia de la prueba de hipótesis. *Semilla Científica*. 2024;5:211-216.
<https://doi.org/10.37594/sc.v1i5.1381>
8. Jiménez Ríos E. Los diccionarios de la Real Academia Española. En: *Lexicografía hispánica/The Routledge Handbook of Spanish Lexicography*. Routledge; 2024. p. 392-406.
<https://doi.org/10.4324/9780429244353-29>
9. Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med*. 2015;108(1):32-7.
<https://doi.org/10.1177/0141076814562718>
10. Hernández-Sampieri R, Mendoza C. *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw Hill; 2018.
11. Dagnino J. Análisis de proporciones. *Rev Chil Anest*. 2014;43(2):134-138.
<https://doi.org/10.25237/revchilanestv43n02.12>
12. Fregni F, Lopes F, Matsubayashi SR. Non-Parametric Statistical Tests. En: Fregni F, Illigens B (editores). *Critical Thinking in clinical research*. Nueva York: Oxford University Press; 2018.
<https://doi.org/10.1093/med/9780199324491.003.0010>
13. Sagaró Del Campo NM, Zamora Matamoros L. Técnicas estadísticas multivariadas para el estudio de la causalidad en medicina. *Rev Ciencias Médicas*. 2020;24(2):287-300.