

## PRINCIPIOS BÁSICOS SOBRE INFERENCIA ESTADÍSTICA

### PRINCIPLES FOR STATISTICAL INFERENCE

Valeria Camila Restrepo Arias MD\*, Mario Delgado Noguera MD, PhD\*\*

#### RESUMEN

*Para facilitar la interpretación y análisis de los resultados de una investigación, en el presente texto se considera la inferencia estadística y se describen tres principios para la selección de una prueba estadística: primero, sintetizar las diversas mediciones en un número adecuado de datos manejables, segundo, estimar e inferir a partir de las muestras extraídas de poblaciones y por último, ajustar los datos según la influencia de las variables de confusión.*

**PALABRAS CLAVE:** *Estadística, epidemiología, inferencia.*

#### ABSTRACT

*To attend the first purpose of medical research, the selection of appropriate statistical method, this review described the statistical inference and the principles for the selection of statistical test: first, synthesize measurements on a limited number of manageable data, second, estimates and inferences from samples taken from populations and finally, fit the data as the influence of confounding variables.*

**KEYWORDS:** *Observational research, epidemiology, research reporting*

Historia del artículo:

Fecha de recepción: 11/01/2015

Fecha de aceptación: 28/03/2015

\* Médica y cirujana, Universidad del Cauca.

\*\* Universidad del Cauca, profesor titular, Facultad Ciencias de la Salud, Departamento de Pediatría.

Correspondencia: Carrera 6 # 10N-142, Departamento de Pediatría, Hospital Universitario San José, tercer piso, Popayán, Colombia. Correo electrónico: mariodelg@gmail.com

Con el objetivo de facilitar el análisis e interpretación de los resultados de una investigación, se describen a continuación los principios básicos sobre la inferencia estadística. Generalmente la investigación se lleva a cabo en muestras de la población; en la muestra se trabaja con variables y se trata de establecer la probabilidad de relación entre ellas. La inferencia estadística, permite determinar con un grado de confianza si existe una diferencia real entre dos alternativas de tratamiento por ejemplo, o si la diferencia se explica por el azar.

Para empezar, consideramos que la estadística en la investigación en salud posee las siguientes finalidades: primero, elegir el método estadístico adecuado, segundo, analizar las mediciones de acuerdo al método elegido y tercero, hacer una interpretación adecuada. Para cumplir con la primera finalidad es necesario conocer los principios básicos para la selección de una prueba estadística, que son: 1) Sintetizar numerosas mediciones en un número limitado de datos manejables, 2) Estimar e inferir a partir de las muestras extraídas de poblaciones y 3) Ajustar los datos según la influencia de las variables de confusión (1).

Este manuscrito enfatiza en esos principios, por lo que se realizará una descripción detallada de cada uno de ellos.

## 1. SÍNTESIS DE LAS MEDIDAS

Este primer principio se refiere a resumir grandes cantidades de datos en un número adecuado y manejable de ellos. Para hacerlo, se debe tener claro el concepto de “muestra”, que significa que las mediciones realizadas en una investigación son parte de un grupo más numeroso de individuos (Población), y que tuvieron la misma probabilidad de haber sido incluidos en el mismo (1).

Para entender mejor el concepto es importante reconocer que las poblaciones pueden ser muy grandes y tener tantos miembros que es imposible estudiarlos a todos. Las poblaciones, por otra parte, pueden ser inestables, pues hay un flujo

constante de individuos que entran y salen imposibilitando también el análisis de cada sujeto, por eso se acude a un grupo de individuos que represente esa población y a ese grupo se le denomina muestra (2).

De la adecuada selección de la muestra depende que los resultados sean muy similares a los que se hubieran obtenido si el estudio se hubiese realizado en toda la población. El muestreo es el procedimiento destinado a producir una muestra (3). Existen varios tipos de muestreo, el aleatorio simple, por ejemplo, garantiza la equiprobabilidad de elección de cualquier elemento de la población; en otras palabras, todos los individuos de la población tienen la misma probabilidad de ser elegidos (2).

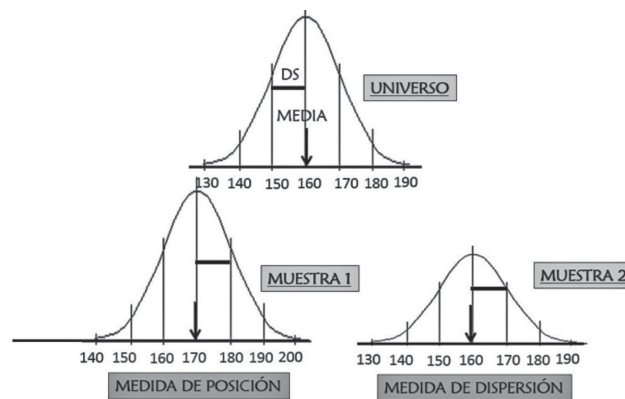
Así mismo, se debe entender el concepto sobre la “distribución poblacional”, que es la manera como se grafican las frecuencias de ciertas variables dentro de una población. Sin embargo, a veces es difícil entender o transmitir la información de una gráfica por lo que los métodos estadísticos son una medida sintética de esa distribución.

Cada distribución de una población posee un número limitado de valores sintéticos que la describen de manera concreta y completa y se denominan Parámetros, son ejemplo en el caso de la distribución gaussiana o normal, la media y la desviación estándar, medidas de posición o medidas de tendencia central, y dispersión respectivamente (1).

Para explicar a que se hacen referencia las medidas de dispersión y posición, se toma como ejemplo hipotético la distribución poblacional de la talla de los estudiantes de medicina, donde la media es 160 cm y la desviación estándar es 10cm.

En la Figura 1 se puede observar tres situaciones: (a) Si se tomara una muestra de esa población donde la media es 170 cm en lugar de 160 cm, la posición de su centro de gravedad se mueve, pero la distribución no cambia. (b) Si se altera la dispersión, es decir la distancia de los datos respecto a la media, se modifica la forma del gráfico, pero no su posición (c).

**Figura 1.** Medidas de posición y dispersión de la distribución poblacional hipotética de la talla de los estudiantes de medicina.



## 2. ESTIMACIÓN E INFERENCIA

Debido a que en muy raras ocasiones es posible realizar todas las mediciones de una población, acudimos a la estimación de esos parámetros a través de las mediciones en una muestra extraída de la misma. Las muestras se comportan de una manera previsible y se distribuyen en un conjunto de valores esperados. Las estimaciones muestrales de un parámetro poblacional se denominan estadísticos o estimadores (1-3).

A través de la inferencia estadística podemos saber qué tan cerca del valor real de los parámetros poblacionales se encuentran las estimaciones muestrales. Sin embargo, aunque nunca será posible conocer el valor real de esos parámetros, lo que sí podemos saber es cuánto se espera que los estimadores se acerquen a un valor real hipotético (1).

Dicho de otra manera, la inferencia estadística usa datos de una muestra para hacer comparaciones y sacar conclusiones sobre la población que representa (2). Por ejemplo, se quiere conocer la mortalidad en los pacientes sometidos a dos tratamientos diferentes para el cáncer de mama. Para ello, se asignan de manera aleatoria dos tipos de tratamientos a los pacientes de un hospital durante un tiempo determinado y luego se compara la mortalidad en ambos grupos. Al hacer la comparación, se puede determinar si alguna de las opciones aportó un beneficio o un riesgo significativos. Este resultado se puede inferir desde la muestra a la población general cumpliendo con ciertas condiciones como el control del error sistemático.

Así mismo, existen dos consideraciones que permiten investigar la influencia del azar en las estimaciones muestrales, las pruebas de significación estadística o de hipótesis y la estimación por intervalo de confianza.

Para entender el concepto de prueba de significación estadística o de hipótesis se debe considerar que entre dos factores, siempre habrá dos posibilidades de relación: o se asocian, o no se asocian. Se dice que se asocian cuando aparecen a la vez con más frecuencia de lo esperado por el azar (1). Cuando se encuentran diferencias entre los grupos de la muestra, los investigadores calculan la probabilidad (azar) de observar esas diferencias suponiendo, como hipótesis nula, que no existieran diferencias reales en la población de la que fueron seleccionados.

El objetivo de estas pruebas estadísticas es entonces calcular el valor p, que es una representación de la probabilidad, con un rango que va de 0 (probabilidad nula de que ocurra un acontecimiento) a 100 (el acontecimiento ocurrirá siempre) (2).

La segunda consideración es el intervalo de confianza o estimación por intervalo de confianza. Cuando nos referimos a él, podemos decir que se tiene un nivel de confianza determinado de que el parámetro poblacional está incluido en un rango de valores. De la misma manera, se refiere a la precisión con que se ha realizado la estimación de un parámetro poblacional, de modo que, entre más amplio sea el rango de valores del intervalo, más posibilidades habrá de que lo incluya; habitualmente se aceptan intervalos del 95%, de tal manera que si se repitiera la investigación en 100 oportunidades, existe una probabilidad del 95% de que el parámetro poblacional estará incluido dentro del intervalo o rango de valores (1, 4).

Para explicar el concepto se supone que se quiere conocer el nivel de glicemia de los neonatos macrosómicos (Población), de

modo que se obtiene un grupo aleatorio de 50 niños (Muestra) y se supone que estos datos toman una distribución normal, con una media de 70 mg/dl y una desviación estándar de 10 mg/dl; para estimar el intervalo de confianza, la fórmula es: Estimado +/- Margen de error o error estándar. Se supone que al llevar a cabo el cálculo de la fórmula resultó un intervalo de confianza de 70 +/- 5 mg/dl (Rango de 65 a 75 mg/dl), con un nivel de confianza del 95%; la interpretación del intervalo es que, en esa muestra de 50 niños macrosómicos, existe una probabilidad del 95% de incluir el parámetro poblacional de la glicemia. Si se repite el estudio en 100 ocasiones, el 95% de las veces el parámetro poblacional se encontrará en ese rango.

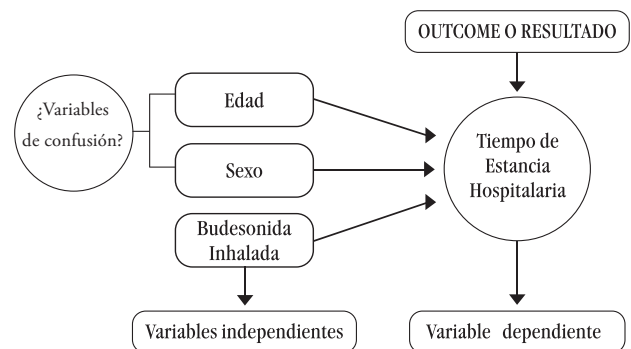
### 3. AJUSTE DE LOS DATOS SEGÚN LAS VARIABLES DE CONFUSIÓN

En todo estudio es necesario averiguar cuál es la variable dependiente, la existencia de variables independientes, y si las hay, cuántas son y qué tipo de datos las constituyen. Con esta información, es posible proceder a la selección adecuada de la técnica estadística.

Se recuerda que una variable es la característica que se quiere medir en un estudio, y que la variable dependiente es la de interés principal o desenlace final. De ella se desea realizar una estimación o contrastar en una prueba de hipótesis. Las variables independientes determinan las características que se desean medir o con las que se contrasta la hipótesis y pueden ser una o varias (1, 5).

Para ilustrar esta distinción entre el tipo de variables, se toma como ejemplo un ensayo clínico en el que se desea estimar como resultado el "Tiempo de estancia hospitalaria" de pacientes con crisis asmática a los que se les administró una nueva terapia, la budesonida inhalada, versus el tratamiento habitual (6). Se incluyeron variables como edad y sexo. En este caso, el desenlace principal es el tiempo de estancia hospitalaria, que será la variable dependiente. La variable independiente será la administración del corticoide inhalado (budesonida). La edad y el sexo serán también variables independientes que pueden influir en el resultado esperado a la terapia nueva que se está probando (Figura 2).

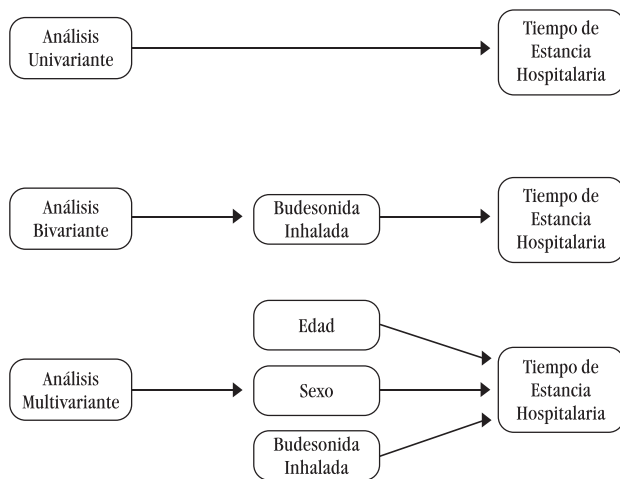
Figura 2: Variables: Dependiente e independientes.



Teniendo en cuenta los anteriores conceptos, se puede realizar la elección del tipo de análisis estadístico. Para ello, en el ejemplo anterior, si solo se desea estimar la estancia hospitalaria de los pacientes con crisis asmática, se elige, en este caso, un análisis de tipo univariado. Si se lleva a cabo un ensayo clínico para probar la eficacia de una intervención, se requerirá un análisis bivariado si solo se realizan las mediciones intervención y tiempo de estancia hospitalaria; pero si además se miden otras variables como edad y sexo de manera conjunta, se debe recurrir al análisis multivariado para determinar la influencia de estas dos características en la respuesta a la budesonida inhalada. Por otra parte, se debe determinar si las anteriores variables son de confusión, pues de una manera sistemática pueden influir en dicha asociación y causar distorsión de los resultados (3), ya que es posible que sean la edad o el sexo del paciente las variables que determinen la reducción de la estancia hospitalaria y no la administración de budesonida.

De lo expuesto, podemos resumir que un análisis univariado es aquel que estima una medición cuando hay una sola variable; el bivariado, cuando se desea observar la asociación entre una variable dependiente y una independiente y el multivariado, cuando existe una variable dependiente y más de una independiente (Figura 3).

**Figura 3:** Pruebas estadísticas, según tipo y cantidad de variables.



Finalmente se puede decir que la inferencia estadística es un proceso lógico que permite contestar a preguntas sobre las relaciones entre las variables o diferencias entre grupos. De manera que la inferencia es uno de los procesos que ayudan a predecir lo que pasará en el futuro en un evento en salud al elegir con cierta precisión entre diferentes alternativas.

## REFERENCIAS

1. Riegelman R, Hirsch R. Cómo estudiar un estudio y probar una prueba: Lectura crítica de la literatura médica. 2ª. Ed. Washington, D.C: Editorial Organización Panamericana de la Salud; 1992. p. 173-233.

2. Dawson G. Interpretación fácil de la bioestadística. La conexión entre la evidencia y las decisiones médicas. 1ª. Ed. Barcelona: Editorial Elsevier España, S.L; 2009. p. 2-82.
3. Guerrero R, González C, Medina E. Epidemiología. 1ª Ed. Washington, D.C, Addison Wesley Iberoamericana Editorial; 1986. p. 52-61, 139-146, 152-160.
4. Calvache J, Delgado M. Estimaciones puntuales e Intervalos de confianza: Resumen de los Resultados de una Investigación. Revista Facultad Ciencias de la Salud. Universidad del Cauca. 2007;9:57-59.
5. Shoemaker R. El primer acercamiento a la bioestadística: Clasificación de las variables. Revista Facultad Ciencias de la Salud. Universidad del Cauca. 2011;13: 24-27.
6. Hasan C, Zulfikar A, Harmanci K, Kocak M, Kuras Y. The Addition of Inhaled Budesonide to Standard Therapy Shortens the Length of Stay in Hospital for Asthmatic Preschool Children: A Randomized, Double-Blind, Placebo-Controlled Trial. Int Arch Allergy Immunol 2015;166:297-303.