

# La Bioestadística y su papel en la investigación en salud

NELCY RODRÍGUEZ M.\*

---

## Resumen

La investigación en las áreas biológicas y muy especialmente en el área de la salud, es la base para la toma de decisiones que afectan a seres vivos. Este artículo hace una revisión del papel de la bioestadística como un área particular de la estadística en procesos de investigación, papel que resulta fundamental, dado que provee al investigador los elementos esenciales y suficientes para sustentar los resultados de sus observaciones y el análisis de las mismas.

*Palabras Clave:* Pregunta de investigación, medición, diseño, muestra, aleatorización, sesgo, validez, confiabilidad.

## Abstract

Decision making about living beings is supported by biological and health research. The role of biostatistics in research process is reviewed in this paper. Biostatistics is a key area in research since it can provide fundamental and enough elements which help the researcher to support his or her observations and analyses.

*Keywords:* Research question, measurement, design, sample, randomization, bias, validity, reliability.

## 1. Introducción

Con frecuencia, cuando un investigador está interesado en responder a una pregunta de su interés particular, aparecen a su vez otras de carácter

---

\*Profesora asistente, Departamento de Estadística, Universidad Nacional de Colombia.

metodológico como las siguientes: ¿Cómo medir lo que se quiere observar? ¿Cómo verificar las hipótesis del estudio si, para responder a la pregunta de interés, se necesita la evaluación de una o más de ellas? ¿Cuántos sujetos o elementos será necesario involucrar? ¿Cuántos grupos de sujetos será necesario determinar para contestar a la pregunta y qué proceso se deberá seguir en la asignación de esos sujetos a tales grupos?, ¿Cómo recolectar y controlar la calidad de la información necesaria para producir un resultado? Una vez obtenida esa información, ¿Cómo procesarla y analizarla?, y al final del estudio ¿Qué tanto se podrán extrapolar o generalizar los resultados que se encuentren y a quienes serán generalizables? Todos estos interrogantes están directamente relacionados con el quehacer de la estadística y su aplicación en investigación. Este artículo revisa algunos de los aspectos más importantes en los cuales el papel específico de la bioestadística es fundamental en un proceso de investigación, particularmente en el área de la salud y que, por tener un carácter decisivo en los resultados, deben ser cuidadosamente considerados y aplicados por el investigador.

## 2. Algunas definiciones

Para entender el quehacer de la estadística en la investigación, específicamente en el área de la salud, podemos revisar los siguientes dos enunciados: Para Elston & Johnson (1987) “la estadística tiene que ver con la colección, organización, presentación, análisis e interpretación de información que se exprese en forma numérica”. Según Echeverría (1982), “los métodos estadísticos juegan un papel importante en el proceso de investigación en cuanto reducen el nivel de incertidumbre a la hora de buscar resultados válidos y fiables”.

Si aceptamos que los dos enunciados anteriores describen de manera clara el papel de la estadística en un proceso de investigación, entonces el buen uso de las técnicas necesarias para solucionar su problema permitirá al investigador minimizar la ocurrencia de ciertos errores y evaluar aspectos tan importantes como la variabilidad en la ocurrencia de los eventos de su interés. Ahora bien, la bioestadística es el área de la estadística que sirve como herramienta de apoyo a las ciencias biológicas en sus procesos de investigación sobre problemas relacionados con los seres vivos. Es entonces la rama de la estadística puesta al servicio de las ciencias biológicas (Pagano & Kimberlee 2000). Por ejemplo, frente a un problema de investigación específico relacionado con la salud de las personas, nos podemos preguntar cómo algunos elementos de la bioestadística pueden ser de utilidad. En ese proceso, se tienen unos aspectos de carácter metodológico que se inician con la pregunta misma de investigación.

### 3. Aspectos metodológicos

#### 3.1. Pregunta de investigación

Una pregunta de investigación, definida por Hulley & Cummings (1988) como la “incertidumbre acerca de algo en una población o en un grupo, que el investigador quiere resolver haciendo mediciones en sus sujetos de estudio”, presenta dos elementos esenciales en los cuales la bioestadística tiene mucho que aportar: la definición de una población o de un subgrupo de ella y el establecimiento de mediciones en los sujetos de estudio. Con el primer elemento se hace referencia a la definición de un universo o población general, que con frecuencia es restringida por algunos criterios particulares para el problema considerado y que al restringirse, se convierte en lo que se conoce como “población de estudio” (Ardila & Rodríguez 2001); o de una muestra de tal población, para lo cual, las técnicas de diseño y selección de muestreo resultan pertinentes. Con el segundo elemento se referencia al establecimiento y definición de las variables. La definición y posterior operacionalización de dichas variables determinará, por un lado, el proceso de recolección de información y por otro, el plan de análisis para los datos recolectados. Una buena pregunta de investigación debe tener cinco características esenciales (Hulley & Cummings 1988), relacionadas de una u otra forma con la consideración de aspectos de la bioestadística, si la pregunta se relaciona con seres vivos. Estas características se refieren a la factibilidad, el interés, la novedad, la ética y la relevancia.

En términos de la factibilidad por ejemplo, una pregunta específica sobre un problema, puede ser susceptible de tener una respuesta, una vez evaluados aspectos como la población de referencia, la población de estudio, el número de sujetos a ser estudiados y la logística del proceso de recolección de la información. Cada uno de estos aspectos, tiene un alto grado de aplicación de metodología estadística (diseño y cálculo de tamaño de muestra, técnica de selección de sujetos al estudio, método de recolección de la información, procesos de control de calidad de los datos, etc.)

#### 3.2. Objetivos

Este aspecto permite determinar el tipo y el número de variables o de indicadores que serán medidos. Por ejemplo, supongamos que se pretende establecer la prevalencia de la desnutrición global en niños menores de cinco años de cierta ciudad. Al plantear este objetivo, es claro el establecimiento de una medida de frecuencia de tipo epidemiológico conocida como la tasa de prevalencia

y la medición de una variable como, por ejemplo, la desnutrición global, para cuya determinación, el investigador necesitará observar o medir a su vez, otras variables como el peso, la talla y la edad de los niños menores de cinco años que participen en el estudio.

### 3.3. Planeación

Aquí se identifican aquellos pasos necesarios para desarrollar el estudio y que permitan cumplir con los objetivos propuestos. Esta planeación incluye puntos tales como: la selección del tipo de diseño, el número de sujetos a estudiar, la medición de las variables identificadas en el planteamiento de los objetivos, el número de observaciones por sujeto, el número de grupos que será estudiado, la forma de asignación de los sujetos a esos grupos en el caso de estudios de tipo experimental, la forma de recolección de la información, la definición del procesamiento de ésta y el análisis de datos para obtener finalmente conclusiones válidas y confiables. Una incorrecta aplicación de uno o más de estos puntos, ha llevado a que más de una investigación en salud haya mostrado resultados carentes de validez. Revisemos entonces cómo la bioestadística tiene un papel definitivo en la aplicación de estos puntos.

#### 3.3.1. Selección del diseño del estudio

La selección del diseño indica automáticamente el uso del tipo de medida o medidas estadísticas pertinentes en la búsqueda de respuestas para la pregunta planteada. En investigación sobre problemas relacionados con la salud de las personas, existe toda una serie de estrategias que pueden ser aplicadas según el problema y los objetivos del estudio que se pretende llevar a cabo (Kleinbaum, Kupper & Morgenstern 1982). Por ejemplo, en el caso del problema sobre la desnutrición global, si el investigador decide que su diseño será observacional y para ello elige una encuesta de tipo transversal, entonces desde el punto de vista estadístico podrá recurrir a medidas de frecuencia simple y determinar además la tasa de prevalencia, tasa resultante de la frecuencia de niños con desnutrición global en la población considerada como expuesta a este evento, en un tiempo y lugar previamente determinados.

Otro estudio posterior puede plantear, por ejemplo, la evaluación de la asociación entre la desnutrición global de los niños menores de cinco años y algunos factores potenciales relacionados. Al estudiar este problema, la bioestadística proporcionará elementos esenciales como la posibilidad de construir y analizar tablas de contingencia, evaluar hipótesis sobre la asociación planteada y cal-

cular medidas de asociación que hablarán de la magnitud de la relación entre la desnutrición y los factores de interés, mediante el uso de análisis simples o recurriendo al modelamiento matemático para el caso de análisis ajustados (Preisser & Koch 1997).

### 3.3.2. Número de sujetos a estudiar

Con relación a este aspecto es posible, usando el ejemplo sobre la prevalencia de la desnutrición global, que el investigador no pueda estudiar toda la población expuesta a la desnutrición sino que deba extraer una muestra de ella. El muestreo de tipo probabilístico y las técnicas de selección de sujetos permitirán determinar el número de niños seleccionados y la forma en que se hará su selección. Este es uno de los puntos más cruciales en el diseño de investigaciones en salud y además una de las consultas hechas más frecuentemente a los bioestadísticos. Dependiendo de los objetivos planteados, de la definición de la variable más importante de estudio, del nivel de confiabilidad deseado, del error de muestreo que se está dispuesto a tolerar, del número de grupos estudiados, de la variabilidad esperada en las observaciones de los sujetos, de la potencia estadística esperada para evaluar hipótesis que quieren comparar grupos en términos de algún estimador específico, entre otros factores, el uso de fórmulas estadísticas acordes con la estrategia de diseño seleccionado permitirá tener el cálculo de un número mínimo de sujetos que deben ser involucrados en el estudio (Dennis 1989). Por ejemplo, si el interés es determinar si la tasa de prevalencia global en la población es aproximadamente del 10%, suponiendo una confiabilidad del 95% al hacer la estimación de esa tasa de prevalencia, un error del 5% y una diferencia mínima deseada del 4% y conociendo además que la población total del grupo de 2'800.000 niños, el número mínimo de niños a ser involucrados en el estudio será de 217. Este número es el resultado de la aplicación de una de las ecuaciones de tamaño de muestra, apropiada para el tipo de diseño y el objetivo que se pretende lograr. Esta ecuación está dada por la siguiente expresión (Ospina 2001).

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad (1)$$

donde

$$n_0 = \frac{Z_{\alpha/2}^2 p(1-p)}{e^2} \quad (2)$$

$N$  = Tamaño de la población,

$\alpha$  = Probabilidad de cometer el error tipo I,

$1 - \alpha/2$  = nivel de confianza,

$p$  = Proporción esperada en la población,

$e$  = Diferencia máxima esperada y

$Z_{\alpha/2}$  = percentil correspondiente en la distribución normal estándar para un nivel dado,  $\alpha$ .

La anterior fórmula permite obtener el tamaño de la muestra para estimar la proporción poblacional  $P$ , basada en una precisión establecida en unidades absolutas cuando se conoce el total poblacional.

### 3.3.3. Determinación de variables

Respecto a la definición del tipo y número de características o variables que serán medidas, será necesario revisar si la pregunta de investigación exige determinar una, dos o más variables de interés y además, la posible interrelación entre ellas. Esta consideración determinará el carácter univariado o multivariado del problema, desde el punto de vista estadístico. La definición de la escala de medición que se empleará en el proceso de medición y por tanto el tipo de datos que se obtendrá, es uno de los aspectos claves en la determinación de la técnica estadística a ser empleada en la fase de procesamiento y análisis de información. Si las variables definidas son todas cualitativas, nominales u ordinales, el tipo de datos que resultará de su observación, será diferente del dato que se obtiene cuando la medición se hace mediante el uso de una escala cuantitativa como la escala de intervalo o la de razón. En el ejemplo sobre la desnutrición global, si una de las variables de interés es el género de los niños bajo estudio, la medición de esta variable proporcionará un dato de tipo discreto que implicará luego obtener una distribución de frecuencias simples que refleje la distribución de los niños según el género. En el caso de otras características importantes del estudio como la edad, el peso y la talla, su observación permitirá obtener datos de tipo cuantitativo y continuo que posteriormente serán susceptibles de ser manejados mediante el uso de medidas de tendencia central y de variabilidad si la pregunta de investigación lo exige y el sistema de muestreo lo permite, se podrán aplicar técnicas estadísticas inferenciales para comparar grupos en cuanto a algunas de estas características (Milton 2001).

### 3.3.4. Número de grupos de estudio

Referente al número de grupos a estudiar y al tamaño de éstos, definidos desde la pregunta de investigación, la bioestadística proporcionará técnicas que pueden ser aplicadas a una, dos o más poblaciones o muestras, técnicas tanto de tipo paramétrico como no paramétrico. Así por ejemplo, puede ser de interés, comparar el promedio del ingreso familiar entre aquellos niños que

presentan un problema de desnutrición global y aquellos que no lo presentan con relación además, a un grupo específico de edad. Si los niños del estudio han sido seleccionados mediante una muestra probabilística, el tamaño de ésta es suficiente y se cumplen ciertos supuestos estadísticos como el de normalidad, será lícito emplear una técnica de tipo paramétrico como la de *Student* para comparación de muestras independientes. Si el interés es comparar un medicamento nuevo frente a dos ya conocidos para la disminución de la tensión arterial, automáticamente se están definiendo tres grupos de estudio para la aplicación de un diseño experimental. Dado que la medición de la tensión arterial de los individuos bajo estudio se hace en milímetros de Mercurio (mmHg), una de las técnicas estadísticas a usar, luego de aplicar estos tratamientos y obtener los resultados, puede ser la comparación de los promedios en las mediciones de tensión arterial de los tres grupos, a través de un análisis de varianza si los supuestos para la aplicación de esta técnica, se satisfacen (Dowdy & Wearden 1991). Supongamos además que se encuentran diferencias en los promedios de las tensiones arteriales entre los tres grupos de tratamiento. Si un objetivo del estudio es determinar cuál de los tratamientos es significativamente diferente de los otros dos con respecto al promedio de tensión arterial, entonces puede ser factible, el uso pruebas de comparación múltiple. Cabe anotar que el uso de estas técnicas debería, en lo posible, estar determinado desde la planeación misma del estudio, debido a consideraciones que deben hacerse sobre la modificación del error tipo I al evaluar las hipótesis de diferencias potenciales entre los grupos de tratamiento (Dowdy & Wearden 1991, Shott 1990).

### 3.3.5. Asignación de sujetos a los grupos de estudio

Respecto a la forma de asignación de sujetos a uno de los grupos de estudio en aquellos casos en que la estrategia de diseño sea de tipo experimental, el investigador y, especialmente el bioestadístico, pueden recurrir a métodos eminentemente aleatorios, que permitirán idealmente, tener grupos comparables en cuanto a algunas variables que podrían de otra forma sesgar la estimación de parámetros de interés. El proceso de aleatorización de individuos a grupos de intervención, permite remover el sesgo del investigador en esa asignación y garantiza además que las pruebas estadísticas tendrán niveles de significancia válidos (Pocock 1983).

En la tarea de aleatorizar sujetos a diferentes grupos o tratamientos, se puede optar por alguno de los métodos conocidos como son la aleatorización simple, por bloques, sistemática, adaptativa o la estratificación. Como se mencionó antes, la correcta aplicación de estos métodos conllevará a la reducción en sesgos de asignación de sujetos (Friedman, Furberg & Demets 1983, Spilker

1984).

A manera de ejemplo, supongamos que en el caso de los medicamentos para reducir los niveles de tensión arterial, el interés es comparar no tres sino dos de estos medicamentos. Para reducir sesgos en la asignación de individuos, se decide hacer una asignación por bloques. Esta alternativa, asegura un número igual de individuos en cada grupo, en la secuencia de su asignación de acuerdo al tamaño del bloque. Si se definieran bloques de tamaño  $k$ , el número de sujetos asignados a cada tratamiento es proporcional dentro del bloque  $k$ -ésimo. Por ejemplo, si el bloque es de tamaño 4, entonces para cada bloque de 4 individuos, se tendrá 2 de ellos asignados a un tratamiento y 2 al otro. Esto quiere decir que habrá 6 posibilidades para aleatorizar los sujetos en bloques balanceados de tamaño 4: Si  $A$  representa uno de los tratamientos y  $B$  el otro, esas posibilidades son:  $ABAB$ ,  $AABB$ ,  $ABBA$ ,  $BABA$ ,  $BBAA$  y  $BAAB$ .

Notemos además que a medida que el tamaño del bloque aumenta, el número de combinaciones posibles para crear bloques balanceados, también aumentará. Usualmente en ensayos clínicos, se prefiere el método de aleatorización por bloques al método de aleatorización simple, especialmente porque si el estudio no incluye el número total de individuos inicialmente planeado, aún se tendrá un número igual o aproximadamente igual en cada grupo de tratamiento si se ha hecho una asignación usando bloques balanceados (Spilker 1984).

Con relación al proceso de asignación de los sujetos a los grupos, es importante en estudios experimentales, que quien suministre el tratamiento o los tratamientos a los sujetos, no conozca la identificación de éstos (por ejemplo la droga o la dosis), para mantener lo que se conoce como el proceso de “enmascaramiento” en la aplicación de tratamientos. Cuando el bioestadístico esté involucrado en la fase de análisis de datos, es aconsejable que tampoco él conozca la identificación de tales tratamientos.

### 3.3.6. Recolección de información

La logística de recolección de los datos para el estudio, que deberá ser claramente definida en la fase de planeación, implica entre otras cosas, la determinación del tipo de instrumento que será empleado para la obtención de información como por ejemplo, un cuestionario estandarizado y previamente probado. Si este es el caso, tal cuestionario deberá presentar las variables de acuerdo a las escalas de medición seleccionadas, precodificado en lo posible para facilitar y controlar la calidad de la información y libre de posibles sesgos de medición cuya ocurrencia pueda comprometer su validez. En el estudio

sobre desnutrición global, si lo que se emplea para la medición de variables es un instrumento como una balanza o un tallímetro, será necesario revisar cuidadosamente la calibración de tales instrumentos en la toma de las mediciones respectivas para que los datos obtenidos estén libres del sesgo debido a problemas en estos instrumentos (falta de calibración por ejemplo). Igualmente, quienes tomen las mediciones o hagan las lecturas correspondientes sobre estas variables, deberán tener criterios definidos o estandarizados, en la forma como se deben hacer tales procesos para evitar el sesgo o error sistemático de medición debido al observador. Vale la pena hacer énfasis en que la ocurrencia de los sesgos mencionados y otros que pueden ocurrir tanto por parte del observador como del observado, no podrán ser corregidos con aplicación de técnicas estadísticas. En estudios experimentales la logística del proceso de recolección de información, implica tener una buena asignación que puede ser enmascarada en forma simple (quien recibe el tratamiento no sabe la identificación de éste) o doble (cuando ni quien administra el tratamiento ni quien lo recibe, conocen la identificación del mismo). Esto ayuda a controlar el sesgo de asignación de sujetos que si se presenta, repercute de nuevo en la validez de los resultados.

### **3.3.7. Procesamiento de información**

La fase de procesamiento de datos implica una revisión inicial de posibles inconsistencias en la información obtenida, una verificación de éstas y en lo posible, una corrección de ellas. Esto implica, que se debe tener definido desde la etapa de la planeación, un tiempo suficientemente amplio para esta fase de un estudio. Nuevamente aquí, la bioestadística tendrá un papel importante en el aseguramiento de una información suficientemente válida y susceptible de ser analizada.

### **3.3.8. Análisis de resultados**

La fase de análisis de datos, debe corresponder a un plan determinado de antemano. El plan de análisis de la información deberá ser concordante con los objetivos del estudio, la estrategia de diseño seleccionada, corresponder al tipo y número de variables definidas y efectivamente medidas, a la relación establecida entre tales variables, a la estrategia de muestreo y técnicas de selección aplicadas, al número de grupos definidos y estudiados, a la unidad de observación y unidad de análisis definidas previamente, al número de observaciones hechas por sujeto y a la forma de recolección de los datos, entre otros aspectos (Rodríguez & Ruíz 2001).

La no cuidadosa aplicación de estos diferentes pero muy relacionados puntos

puede comprometer no sólo la validez interna y la confiabilidad de los resultados sino que además puede generar problemas éticos tanto para el investigador principal de un estudio como para el bioestadístico, miembros de un equipo multidisciplinario de investigación.

### 3.3.9. Generalización de resultados

Establecer claramente la población general como de la población de estudio, ésta última definida a partir de criterios de inclusión y exclusión previamente determinados así como un diseño de muestra apropiado según la estrategia seleccionada, permitirá al investigador determinar aquellos individuos a quienes serán extrapolables o generalizables los resultados obtenidos de su investigación. De nuevo en esta tarea, resulta de gran importancia el trabajo conjunto entre el investigador y el bioestadístico.

## 4. Conclusiones

Con base en esta revisión, podemos decir entonces que el papel de la bioestadística en un proceso de investigación especialmente en salud, no se limita a aspectos específicos. Desde la formulación de la pregunta, hasta el análisis de resultados, esta área del conocimiento proporciona excelentes herramientas que ayudan en la búsqueda de resultados válidos y confiables, los cuales a su vez, constituyen la base para la toma de decisiones que pueden modificar la salud tanto de individuos que son sometidos a estudio como de aquellos a quienes se extrapolan o generalizan los resultados.

## Bibliografía

- Ardila, M. & Rodríguez, N. (2001), *Población y Muestreo*, Vol. Investigación clínica: Epidemiología clínica aplicada, CEJA, Bogotá, chapter 3.
- Dennis, R. (1989), 'Cómo estimar el tamaño de la muestra en investigaciones con humanos', *Acta Médica Colombiana* **14**(2).
- Dowdy, S. & Wearden, S. (1991), *Statistics for Research. Second Edition*, John Wiley and Sons, N.Y.
- Echeverría, B. (1982), *Estadística aplicada a las Ciencias Humanas*, Daimon, Madrid.

- Elston, R. & Johnson, W. (1987), *Essentials of Biostatistics*, Davis Co., Philadelphia.
- Friedman, L. M., Furberg, C. D. & Demets, D. L. (1983), *Fundamentals of Clinical Trials, Chapter 5. Second Edition*, PSG Inc., Boston.
- Hulley, S. & Cummings, S. (1988), *Designing Clinical Research. An Epidemiologic Approach*, Williams and Wilkins, London.
- Kleinbaum, D., Kupper, L. & Morgenstern, H. (1982), *Epidemiologic Research. Principles and Quantitative Methods. Caps. 3,4.*, Van Nostrand Reinhold Co., New York.
- Milton, J. S. (2001), *Estadística para Biología y Ciencias de la Salud*, 3a. edición edn, McGraw- Hill, Madrid (España).
- Ospina, D. (2001), *Introducción al Muestreo*, Facultad de Ciencias. Universidad Nacional, Bogotá.
- Pagano, M. & Kimberlee, G. (2000), *Principles of Biostatistics*, second edition edn, Duxbury, USA.
- Pocock, S. J. (1983), *Clinical Trials. A practical Approach. Chap. 5.*, John Wiley and Sons, Chichester.
- Preisser, J. S. & Koch, G. C. (1997), 'Categorical data analysis', *Public Health. Annu. Rev. Public Health* **18**, 51–82.
- Rodríguez, N. & Ruíz, J. G. (2001), *Plan de Análisis. Cap 25 en: Investigación Clínica: Epidemiología Clínica Aplicada*, CEJA, Bogotá.
- Shott, S. (1990), *Statistics for Health Professionals*, W.B. Saunders Co., Philadelphia.
- Spilker, B. (1984), *Guide to Clinical Trials. Chapter 10*, Raven Press, USA.