

Conceptos prácticos en MBE y epidemiología clínica

Medwave 2013;13(10):e doi: 10.5867/medwave.2013.10.5826

Estadística para aterrizados: recomendaciones para describir sus datos

Statistics for the faint of heart – how to display your data

Autores: Ricardo Aris⁽¹⁾, Felipe Martínez^(1,2)

Filiación:

⁽¹⁾Centro de Investigaciones Biomédicas, Escuela de Medicina, Universidad de Valparaíso, Chile

⁽²⁾M.Sc. Programme in Evidence-Based Healthcare, University of Oxford, Inglaterra

E-mail: rjariss@gmail.com

Citación: Aris R, Martínez F. Statistics for the faint of heart – how to display your data. Medwave 2013;13(10):e5826 doi: 10.5867/medwave.2013.10.5826

Fecha de envío: 13/10/2013

Fecha de aceptación: 8/11/2013

Fecha de publicación: 11/11/2013

Origen: solicitado

Tipo de revisión: con revisión por tres pares externos, a doble ciego

Introducción

La estadística es una parte esencial de cualquier proyecto de investigación biomédica. Tradicionalmente se divide en dos ramas, la estadística descriptiva y la inferencial¹. El objetivo de la primera es simplemente mostrar y resumir los datos, mientras que la segunda busca establecer afirmaciones probabilísticas que permitan dar sentido a los hallazgos detectados¹⁻⁵. En este artículo discutiremos someramente aspectos esenciales respecto a cómo describir datos, sin pretender profundizar a nivel de textos de estadística clásicos.

Describiendo variables cualitativas

Se entiende por variables cualitativas a todas aquellas cuya variabilidad no pueda ser expresada por un número. Dentro de este grupo existen dos subtipos, las variables nominales que sólo pretenden distinguir valores (o categorías) distintos de otro; y las ordinales en las que fuera de distinguirse categorías existe un orden jerárquico dentro de la variable^{4,5}. Ejemplos de la primera son características como el color de ojos, nombres o género, mientras que en la segunda caen formas de expresar gravedad (leve, moderado y severo) o frecuencia (nunca, ocasionalmente, siempre).

Las variables cualitativas pueden describirse de varias maneras. La primera corresponde al simple conteo de eventos dentro de la muestra, vale decir, la frecuencia absoluta de presentación de la categoría. Las frecuencias relativas, como los porcentajes, expresan cuántas veces se expresó un valor determinado en relación al total. En general, se recomienda que los autores utilicen frecuencias absolutas en vez de relativas si sus tamaños muestrales son menores de 100 para evitar confusiones con el uso de porcentajes⁷. Existen además las llamadas

frecuencias acumuladas. La frecuencia absoluta acumulada es el número de veces que aparece en la muestra un número igual o inferior a la variable estudiada, mientras que la relativa acumulada representa al cociente entre la anterior cifra con el total de la muestra.

Alternativamente, pueden emplearse una serie de gráficos para expresar los distintos valores que pueden tomar las variables estudiadas. Los más importantes son:

Gráfico circular: en este gráfico las observaciones se dividen en un número de segmentos en un gráfico según las posibles categorías dentro de la variable. Su proporción en relación al gráfico total corresponde a su frecuencia relativa. Si bien son muy utilizados, no se recomienda su uso para expresar variables con más de dos categorías dada la dificultad cognitiva que supone la interpretación visual de áreas^{9,10}.

Gráfico de barras: los gráficos de barras expresan las frecuencias observadas en el eje de las abscisas (y). Permiten comparar directamente las alturas observadas por cada categoría en una variable dando una imagen rápida de las formas más frecuentes de presentación. A diferencia de un histograma (véase más adelante), las categorías no son expresadas en intervalos, sino que cada una representa un valor dentro de la muestra^{3,5} y debieran estar separadas por un pequeño espacio una de otra.

Describiendo variables cuantitativas

A diferencia de las anteriores, las variables cuantitativas sí pueden ser expresadas en base a números^{1,2}. Existen dos

subtipos: las continuas, en las que entre dos valores cualquiera de la variable existe un número infinito de valores, y las discretas en las que este número es finito. En general, las variables que se miden (estatura, peso, tiempo operatorio) son continuas, mientras que las que se cuentan (número de hijos) son discretas.

La descripción numérica de estas variables se realiza fundamentalmente mediante dos herramientas estadísticas: las medidas de tendencia central y las de dispersión. Las medidas de tendencia central buscan mostrar al centro de los datos, un valor que caracterizará a la mayoría de la muestra. Las más utilizadas son el promedio (o media, suma de todos los valores dividido por el total), la mediana (el valor céntrico de los datos ordenados de menor a mayor) y la moda (valor más repetido). Las medidas de dispersión dan una medida de la heterogeneidad de los datos; grandes medidas de dispersión denotan mayor variabilidad en los datos cuantitativos. Medidas de dispersión de uso común incluyen al rango (valor mínimo y máximo), el rango intercuartil (valores correspondientes al percentil 25 y 75) y la desviación estándar (promedio de las distancias observadas en los valores al contrastarlos con la media)^{2,8,11}.

Todo dato cuantitativo debiera ser descrito con una medida de tendencia central y una de dispersión¹. La elección de cuál utilizar depende parcialmente de las características de distribución de los datos. Si la variable se distribuye en forma similar a la normal, es recomendable utilizar medias y desviaciones estándar para describirlas. En caso contrario, las medianas y rangos intercuartiles pueden representar una mejor opción¹.

Al igual que con las variables cualitativas, métodos gráficos⁵ pueden emplearse en la descripción de variables cuantitativas:

Gráfico de cajas y bigotes: estos gráficos pueden dividirse en dos secciones. La caja corresponde a la sección central y representa a la mayoría de los datos. Al centro está expresada la mediana (o p50) con una línea horizontal. El límite superior de la caja es el p75 y el inferior el p25, lo que corresponde al rango intercuartil. Los bigotes en cambio muestran la variabilidad fuera de la caja, pero dentro de un límite que corresponde al 1,5 veces el recorrido intercuartil (diferencia entre el p75 y p25). Si existen valores más allá de esta frontera, éstos se dibujan como puntos externos al bigote y reciben el nombre de valores extremos o *outliers* (Figura 1).

La Figura 1 muestra la mediana de edad de la muestra estudiada que fue de 67 años. El p25 fue 58 años y el p75 fue 73 años. Por tanto, el rango intercuartil es 58 a 73 y el recorrido intercuartil, 15 años (73-58=15). Los bigotes muestran el resto de la variabilidad de la muestra, con un máximo de 1,5 veces el valor del recorrido intercuartil aplicado a los límites de la caja. Por tanto, el límite superior del bigote será 95 años {73+(1,5x15)} y el inferior 36 años {58-(1,5x15)}. Existieron valores que

estuvieron más allá de estos límites dentro de la muestra (*outliers*), denotados con puntos.

Histogramas: los histogramas muestran la distribución de los valores de una variable cuantitativa. En estos gráficos la superficie de cada barra es proporcional a la frecuencia de los valores representados, los que son clasificados en un número de intervalos. De esta manera, estos gráficos pueden detectar asimetrías en los datos observados (o sesgos de una curva) o bien la presencia de otros valores que contienen gran parte de la información de la muestra (distribuciones bimodales), información que puede ser relevante al interpretar los resultados. Si bien son ideales para determinar una distribución de datos, la información descriptiva que aportan es algo menor en relación a los gráficos de cajas y bigotes, por lo que son menos utilizados para estos fines que los anteriores (Figura 2).

La Figura 2 es el histograma de los mismos datos de la Figura 1, pero en su lugar entrega información respecto a la distribución de los mismos entre los pacientes estudiados. Se aprecia que la distribución sigue una forma de campana y tiene una discreta asimetría favoreciendo a pacientes más jóvenes (sesgo negativo).

Recursos adicionales

Hemos revisado someramente técnicas simples para describir datos en trabajos científicos. Mayor información respecto a estos y otros métodos puede encontrarse en la lista de referencias a continuación.

Notas

Declaración de conflictos de intereses

Los autores han completado el formulario de declaración de conflictos de intereses del ICMJE traducido al castellano por *Medwave*, y declaran no haber recibido financiamiento para la realización del artículo; no tener relaciones financieras con organizaciones que podrían tener intereses en el artículo publicado, en los últimos tres años; y no tener otras relaciones o actividades que podrían influir sobre el artículo publicado. Los formularios pueden ser solicitados contactando al autor responsable.

Referencias

1. De Muth JE. Preparing for the first meeting with a statistician. *American journal of health-system pharmacy* : AJHP : official journal of the American Society of Health-System Pharmacists. 2008 Dec 15;65(24):2358-66. | [PubMed](#) |
2. Altman DG. *Practical Statistics for Medical Research*. London: Chapman & Hall CRC Press, 1997.
3. De Muth JE. Overview of biostatistics used in clinical research. *American journal of health-system pharmacy* 2009;66(1):70-81. | [CrossRef](#) | [PubMed](#) |
4. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet*. 2002 Jan 12;359(9301):145-9. | [CrossRef](#) | [PubMed](#) |
5. Spriestersbach A, Röhrig B, Prel du J-B, Gerhold-Ay A, Blettner M. Descriptive statistics: the specification of statistical measures and their presentation in tables

- and graphs. Part 7 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009 Sep;106(36):578-83. | [PubMed](#) | [PMC](#) |
6. Bowers D, House A, Owens D. Understanding Clinical Papers. NY: John Wiley & Sons, 2006.
 7. Feinstein AR. Principles of Medical Statistics. London: Chapman & Hall CRC Press, 2002.
 8. Sonnad SS. Describing data: statistical and graphical methods. Radiology. 2002 Dec;225(3):622-8. | [CrossRef](#) | [PubMed](#) |
 9. Tuft ER. The Visual Display of Quantitative Information. Connecticut: Graphics Press, 2011.
 10. Yau N. Visualize This: The FlowingData Guide to Design, Visualization and Statistics. Indianapolis: John Wiley & Sons, 2011.
 11. Curran-Everett D. Explorations in statistics: standard deviations and standard errors. Adv Physiol Educ. 2008 Sep;32(3):203-8. | [CrossRef](#) | [PubMed](#) |
 12. Gaddis GM, Gaddis ML. Introduction to biostatistics: Part 2, Descriptive statistics. Ann Emerg Med. 1990 Mar;19(3):309-15. | [CrossRef](#) | [PubMed](#) |

Figuras

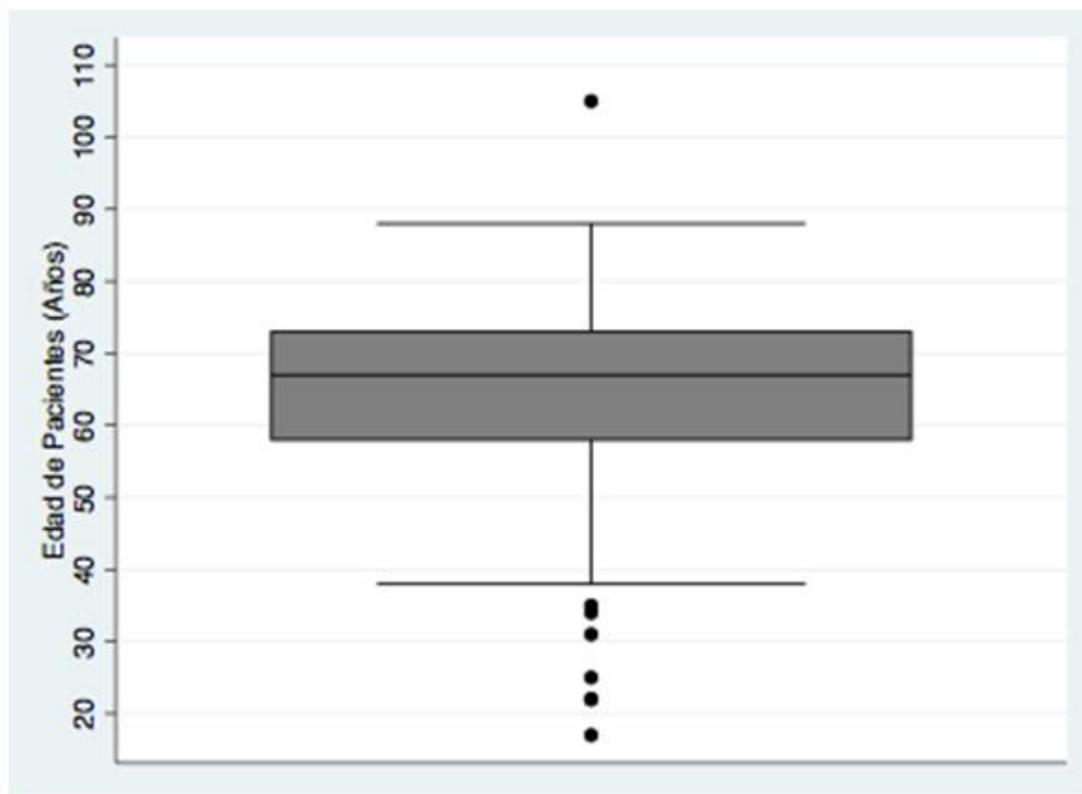


Figura 1. Gráfico de cajas: edades de pacientes ingresados a una Unidad de Cuidados Intensivos Cardiovasculares.

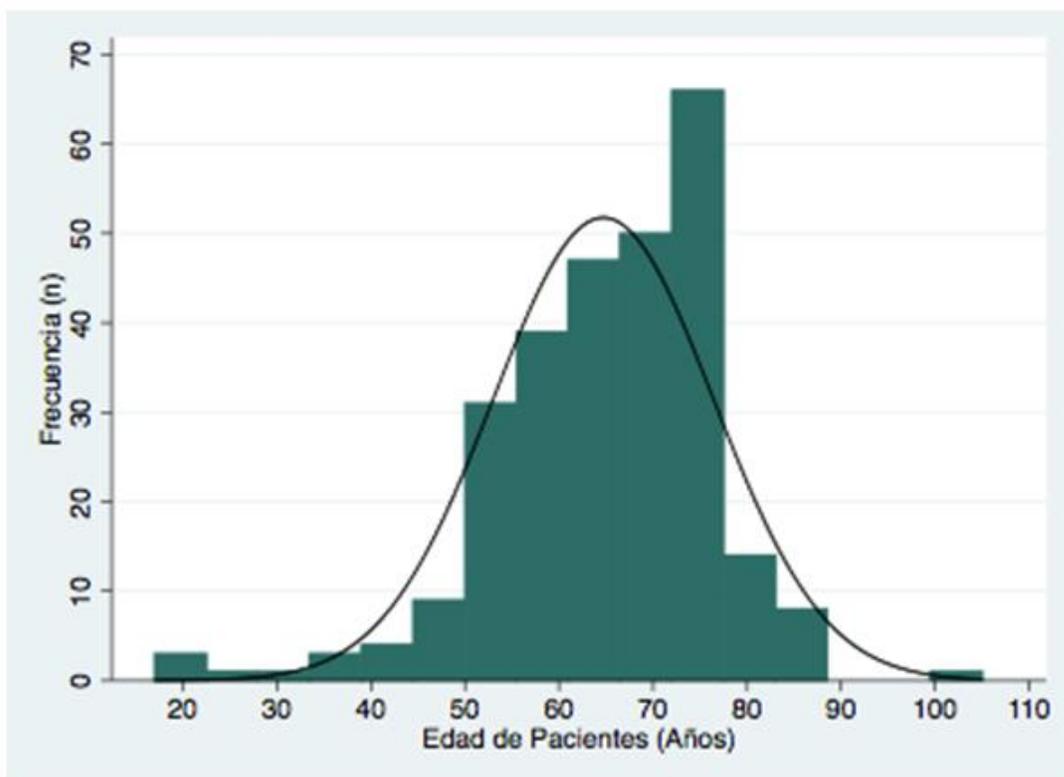


Figura 2. Histograma de edades entre pacientes ingresados a una Unidad de Cuidados Intensivos Cardiovasculares.

Correspondencia a:

Hontaneda 2664
 Edificio Dr. Bruno Günther Schaffeld
 Valparaíso
 Chile



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-No Comercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.