

Revista Chilena de Endocrinología y Diabetes

[Edición Actual](#)[Ediciones Anteriores](#)[Comité Editorial](#)[Instrucciones a los Autores](#)[Directorio Soched](#)[Contacto](#)[Buscar](#)Rev. chil. endocrinol. diabetes 2008; 1 (2) [Volver a Índice](#)

Rincón de la Bioestadística

Representación de datos: Introducción a la Estadística Descriptiva

Gabriel Cavada Ch¹¹División de Bioestadística, Escuela de Salud Pública, Universidad de Chile.²Departamento de Salud Pública, Universidad de Los Andes.

Se llama estadística descriptiva, al conjunto de técnicas que permiten ordenar, resumir y representar la información recolectada. Como su nombre lo indica, la estadística descriptiva sólo pretende describir cuantitativamente el fenómeno, sin proyectar, aún, sus resultados a la universalidad del mismo.

Una vez obtenida la información que se desea analizar es necesario ordenarla y realizar una exploración de la misma para poder, finalmente, presentar esta información mediante tablas o gráficos que permitan una primera impresión de la muestra en análisis; empero, todo esto dependerá del tipo y de la medida de cada variable o atributo de interés; así, de manera general, las variables se pueden clasificar en cualitativas y cuantitativas.

Las variables cualitativas son aquellas que representan una cualidad o atributo que clasificará a cada unidad de observación dentro de una categoría, por lo que se las conoce también como variables categóricas. Estas pueden ser: 1) Ordinales, si dichas categorías siguen un orden (ej. grado de anemia, estadios de tumores, escala de Apgar, etc.); 2) Nominales, cuando las categorías no siguen un determinado orden (ej. género, estado civil, tipo de diabetes). Si la variable cuenta con sólo dos categorías, como es el caso del género, ésta recibe el nombre de variable Dicotómica.

Las variables cuantitativas son aquellas que pueden cuantificarse o medirse numéricamente y pueden ser de dos tipos: 1) Continuas, si permiten tomar cualquier valor dentro de un espectro determinado (ej. edad, talla, concentración de hemoglobina plasmática, etc.) y 2) Discretas, cuando no admiten valores intermedios dentro del espectro, es decir, sólo consideran valores enteros (ej. número de hijos, número de embarazos, etc.)

En el caso de las variables cualitativas, el objetivo radica en conocer la frecuencia absoluta y/o relativa de casos por cada categoría de la variable, es decir la moda (Tabla 1). La representación gráfica más adecuada para este tipo de variables será con gráficos de barras y de sectores. Ej: Se desea conocer las características clínicas de pacientes ingresados a un servicio de medicina interna por insuficiencia cardíaca (IC). Se cuenta con la variable "etiología de IC" medida en escala nominal y cuya distribución de frecuencias se muestra en la Tabla 1, y donde las figuras 1a y 1b son su representación gráfica. Nótese que las frecuencias relativas pueden ser presentadas también como porcentajes en el gráfico de barras, sin embargo, el gráfico de sectores provee una ventaja visual mayor para variables medidas en escala nominal.

Tabla 1 Frecuencia absoluta y porcentual de etiologías de insuficiencia cardíaca en pacientes internados en un servicio de medicina interna en el período 1995-1998.

Etiología	Frecuencia absoluta	Frecuencia porcentual
Cardiopatía isquémica	47	1.54 %
HTA	1 628	53.38 %
Fibrilación auricular	742	24.33 %
Valvulopatías	633	20.75 %
	3 050	100.00 %

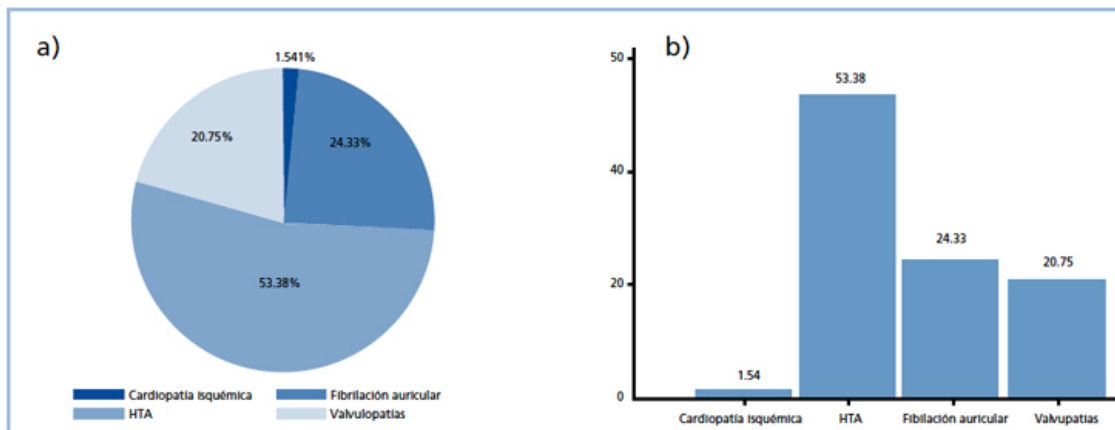


Figura 1. Distribución de las etiologías de insuficiencia cardíaca.

Cuando una variable cuantitativa medida en escala discreta toma pocos valores, su representación numérica y gráfica puede hacerse de la misma manera que una variable categórica (ej. número de partos, número de recidivas, etc.), donde el valor que más se repite, “moda”, representaría el número de partos o recidivas más frecuente en la muestra estudiada. Sin embargo, si la variable está medida en escala continua o discreta con valores altos (ej. índice de masa corporal, concentración de hemoglobina o la edad como variable discreta), la representación gráfica más adecuada de este tipo de variables lo constituirán los histogramas, gráficos de cajas y polígonos de frecuencia; aún así, la medida de resumen que represente toda la información dependerá de la simetría de la distribución de los valores de estas variables.

Así, los gráficos de cajas e histogramas, permiten explorar visualmente la forma de la distribución de la variable cuantitativa. En la Figura 2 se despliega la distribución de la talla de niños escolares de un estudio poblacional de sobrepeso, donde la mayor concentración de los valores se encuentra agrupado en la parte central, por lo tanto, la medida de resumen más adecuada corresponde a las medidas de tendencia central y de dispersión como son el promedio y su desviación estándar respectivamente; dada esta simetría, la mediana, la moda y el promedio deberían ser aproximadamente iguales.

Del mismo ejemplo anterior, la mayor concentración de las observaciones se agrupa en los valores menores de la variable peso, lo que contribuye a la asimetría positiva de la distribución de la variable; por tal razón, la medida de resumen adecuada que representa toda la información que contiene la variable peso, será la mediana con su respectiva dispersión intercuartílica (Figura 3).

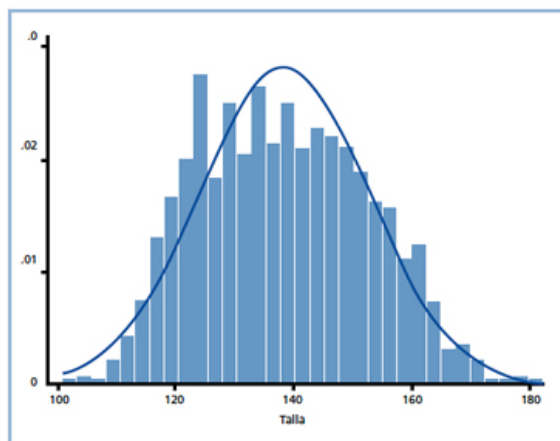


Figura 2. Histograma simétrico.

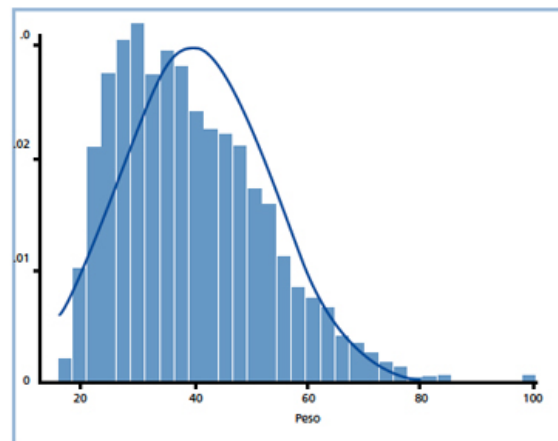


Figura 3. Histograma con asimetría a la derecha.

Otra alternativa para representar una variable continua, la constituyen los gráficos de cajas o “bigotes”, los que además de permitir la exploración de la forma de la distribución, permiten darse cuenta de la dispersión de los datos como de la presencia de valores extremos “outliers”. El límite superior e inferior de la caja representa el percentil 75 y 25 de la distribución de la variable; dentro de la caja, representada por una línea horizontal, se encuentra la mediana la que a su vez representa el percentil 50 de la distribución; a su vez, el ancho de la caja nos da una idea de la variabilidad de los datos, a mayor ancho mayor dispersión. Los bigotes del gráfico representan el valor mínimo y máximo no aberrante dentro del rango, ya que entre estos dos límites se concentran el 95% de la información. Así, todos aquellos puntos más allá de los límites del bigote serán posibles valores extremos u aberrante que merecen un trato específico (Figura 4).

Varios de los métodos de análisis estadístico exigen el supuesto de normalidad de las variables cuantitativas con escala continua; es decir, que la distribución de sus valores se aproxime a una distribución Gaussiana (Figura 2); la fase de exploración de las variables resulta imprescindible para poder evaluar la calidad de los datos antes ser analizados y presentados.

La información presentada solo pretende dar una idea general sobre la estadística descriptiva; invitamos al lector a profundizar sobre el tema en próximos artículos.

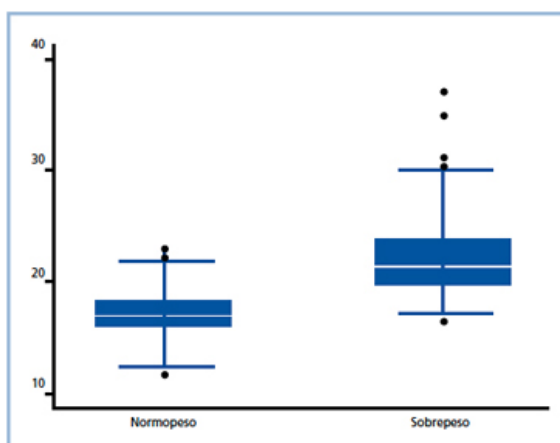


Figura 4. Histograma simétrico.