EVIDENCIAS EN PEDIATRÍA
Toma de decisiones clínicas basadas en las mejores pruebas científicas
www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Estadística descriptiva

Ochoa Sangrador C[1], Molina Arias M[2]
[1]Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.
[2]Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.

Correspondencia: Carlos Ochoa Sangrador, cochoas2@gmail.com

Palabras clave en español: estadística descriptiva.
Palabras clave en inglés: descriptive statistics.

Fecha de recepción: 13 de noviembre de 2018 • **Fecha de aceptación:** 5 de diciembre de 2018
Fecha de publicación del artículo: 12 de diciembre de 2018

Evid Pediatr. 2018;14:43.

CÓMO CITAR ESTE ARTÍCULO

Ochoa Sangrador C, Molina Arias M. Estadística descriptiva. Evid Pediatr. 2018;14:43.

Este artículo está disponible en: http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2018;14:43.
 • ISSN: 1885-7388

Estadística descriptiva

Ochoa Sangrador C[1], Molina Arias M[2]
[1]Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.
[2]Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.

Correspondencia: Carlos Ochoa Sangrador, cochoas2@gmail.com

Como hemos visto en el anterior artículo de esta serie de fundamentos, el primer paso del análisis estadístico es el cálculo de medidas descriptivas de la muestra de estudio. Podemos diferenciar varios grupos de medidas: de masa, de tendencia (o centralización) y de dispersión.

MEDIDAS DE MASA

Son medidas de masa el tamaño muestral (n), el sumatorio y las frecuencias absoluta y relativa.

- Tamaño muestral: el recuento del número de casos.
- Sumatorio (ΣX_i): suma aritmética del valor de una variable de todos los casos.
- Frecuencia absoluta: recuento del número de ocurrencias de cada valor de una variable.
- Frecuencia relativa: porcentaje respecto al total.

En la tabla 1 se puede ver la tabla de frecuencias de la variable número de hijo de 20 parejas, tal y como lo ofrecen la mayoría de los paquetes estadísticos.

MEDIDAS DE POSICIÓN O TENDENCIA O CENTRALIZACIÓN

Las principales medidas de tendencia son la media, la moda y la mediana. Cada una de ellas describe una característica de los datos que estamos analizando.

- **Media muestral:** si tenemos $X_1,\ldots,X_n$ datos se llama media muestral a la media aritmética de ellos.

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\Sigma_{i=1}^{n} X_i}{n} = \frac{\Sigma X_i}{n}$$

- **Moda muestral:** la moda es el valor que más se repite (puede no existir y si existe puede no ser única).
- **Mediana:** si ordenamos de menor a mayor los valores de una variable en una muestra (X_i), la mediana será el valor que esté en el medio o la media de los valores que estén en el medio (si la muestra es par):

$$\widetilde{X} \begin{cases} X_{\frac{n+1}{2}\ si\ n\ impar} \\ \frac{X_{n/2} + X_{n/2+1}}{2}\ si\ n\ par \end{cases} \{X_i\}\ ordenados$$

Veamos de forma gráfica cómo localizar la mediana. En una muestra de 20 pacientes se recogieron las siguientes estancias hospitalarias:

$$X = \{2, 20, 3, 4, 5, 2, 3, 6, 7, 4, 2, 1, 3, 4, 6, 8, 6, 5, 4, 3\}.$$

Si las ordenamos (figura 1), la posición central la ocupan dos 4. La media de ambos es 4. Esa es la mediana.

La medida más popular y empleada es la media; sin embargo, cuando los valores de una muestra no siguen una distribución normal o existen valores extremos en la distribución, no es una buena medida de tendencia. En estas circunstancias recomendamos utilizar la mediana. Si la media y la mediana son muy diferentes, es poco probable que el valor medio describa la tendencia de los datos (probablemente no tengan una distribución de Gauss o normal), por lo que tendremos que dar la mediana o ambos.

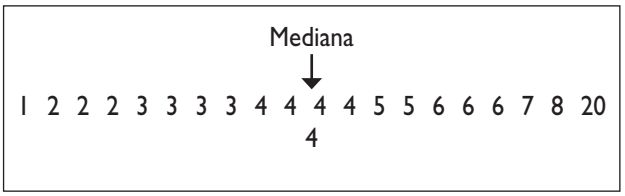
MEDIDAS DE DISPERSIÓN

Las medidas de tendencia no permiten describir los datos de una muestra, porque no informan de cuan alejados está cada

Tabla 1. Tablas de frecuencias de la variable número de hijos de 20 parejas.

X = {2, 20, 3, 4, 5, 2, 3, 6, 7, 4, 2, 1, 3, 4, 6, 8, 6, 5, 4, 3}	**Frecuencia absoluta**	**Frecuencia relativa**	**% acumulado**
1	9	45%	45%
2	6	30%	75%
3	2	10%	85%
4	1	5%	90%
5	2	10%	100 %
Total (tamaño muestral) = sumatorio = (X_i) = 41	20	100 %	

FIGURA 1. IDENTIFICACIÓN DE LA MEDIANA



uno de los valores respecto el valor central. Las principales medidas de dispersión son el rango, la varianza, la desviación típica, el coeficiente de variación y el rango intercuartílico.

Rango

Si ordenamos los valores de menor a mayor, es la diferencia entre los valores extremos (mínimo y máximo):
$\{X_i\}$ ordenados $X_n - X_i$ (máximo - mínimo).

Varianza

La varianza es la media de las diferencias al cuadrado entre cada valor y la media. Se elevan al cuadrado para evitar que las diferencias negativas se anulen con las positivas. Se representa con s^2.

$$S^2 = \frac{\Sigma (X_i - \overline{X})^2}{n}$$

Cuasivarianza

La cuasivarianza es una fórmula de estimación corregida de la dispersión de los datos. Aunque la varianza describe fielmente la dispersión de los datos de la muestra, infraestima la dispersión de los datos en la población de la que procede la muestra si esta tiene pequeño tamaño muestral; por ello la fórmula se corrige disminuyendo su denominador. La varianza que se emplea en inferencia estadística es la cuasivarianza, también conocida como varianza muestral o estimada o simplemente varianza.

$$S^2 = \frac{\Sigma (X_i - \overline{X})^2}{n - 1}$$

Desviación típica muestral o estimada

Como en el cálculo de la varianza las distancias entre cada valor y la media se elevan al cuadrado la magnitud de la dispersión pierde sentido (por ejemplo, para la variable peso su varianza tiene una dimensión en kg^2). Por ello, recurrimos a redimensionar la dispersión haciendo la raíz cuadrada de la varianza. De ahí resulta la desviación típica, representada por *s*.

$$s = +\sqrt{s^2} = \sqrt{\frac{\Sigma (X_i - \overline{X})^2}{n - 1}}$$

Al igual que con la varianza, la fórmula no corregida o desviación típica poblacional:

$$s = +\sqrt{s^2} = \sqrt{\frac{\Sigma (X_i - \overline{X})^2}{n}}$$

Repasemos los pasos para el cálculo de la varianza y desviación típica. No recomendamos realizar los cálculos manualmente, ya que estas medidas son estimadas automáticamente por cualquier calculadora, hoja de cálculo o paquete estadístico. La intención de conocer sus pasos es para entender su significado:

- Anotar las observaciones y calcular la media.
- Calcular la diferencia entre cada valor observado y la media.
- Estas cifras se elevan al cuadrado y se suman, con lo que se obtiene la suma de los cuadrados.
- La suma de los cuadrados se divide por una cifra que es igual al número de observaciones menos uno, con lo que se obtiene la media de la suma de los cuadrados o varianza.
- Extrayendo la raíz cuadrada de la varianza, se obtiene la desviación estándar, que puede caer en cualquiera de los lados de la media.

En la tabla 2 podemos ver el cálculo de la varianza y desviación típica a partir de los datos de longitud de una serie de recién nacidos.

Coeficiente de variación

El coeficiente de variación expresa la dispersión de los datos como medida ajustada. Al dividir la desviación típica por la media, nos indica el porcentaje de dispersión con respecto a la media. Generalmente se expresa como tantos por ciento. Resulta útil para comparar el grado de dispersión de variables con distintas unidades de medida o rango.

$$CV = \frac{S}{\overline{X}} \text{ a veces; } CV = \frac{S}{\overline{X}} \times 100$$

Percentiles 25-75

Ordenando los X_i de menor a mayor el valor que deja a su izquierda el 25% de los casos es el percentil 25 y el que deja por arriba a un 25% de los casos el percentil 75. El **rango o recorrido intercuartílico** es el intervalo entre ambos percentiles (que, en ocasiones, pueden denominarse también como primer y tercer cuartil). En muestras con distribución no normal es la mejor alternativa a la desviación estándar como medida de dispersión.

Veamos de forma gráfica cómo localizar los percentiles 25 y 75. En una muestra de 20 pacientes se recogieron las siguientes estancias hospitalarias:

$$X = \{2, 20, 3, 4, 5, 2, 3, 6, 7, 4, 2, 1, 3, 4, 6, 8, 6, 5, 4, 3\}.$$

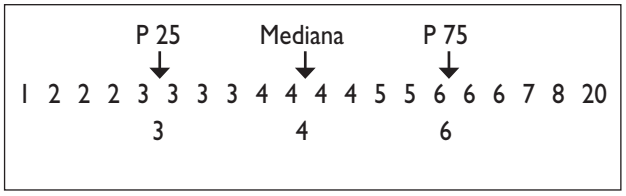
Tabla 2. Cálculo de la varianza y desviación típica a partir de los datos de longitud de una serie de recién nacidos

X_i (longitud RN)	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
48,5	-0,9	0,8
51,0	1,6	2,6
51,0	1,6	2,6
50,0	0,6	0,4
47,0	-2,4	5,8
47,5	-1,9	3,6
51,0	1,6	2,6
50,0	0,6	0,4
49,0	-0,4	0,2
50,0	0,6	0,4
51,0	1,6	2,6
51,0	1,6	2,6
48,0	-1,4	2,0
48,0	-1,4	2,0
50,0	0,6	0,4
49,0	-0,4	0,2
50,0	0,6	0,4
47,0	-2,4	5,8
51,0	1,6	2,6
48,0	-1,4	2,0
Sumatorio	988	39,3

Tamaño muestral: 20; media ($\overline{X}$): 49,4.
Varianza: 39,3/20=1,96; desviación típica: 1,43.

Si las ordenamos (figura 2), los valores que corresponden a los percentiles 25 y 75 son "3" y "6". El rango intercuartílico sería 6 – 3 = 3.

FIGURA 2. IDENTIFICACIÓN DE PERCENTILES 25 Y 75 (RANGO INTERCUARTÍLICO)



BIBLIOGRAFÍA

- Altman DG. Practical statistics for medical research. Londres: Chapman & Hall; 1991.
- Milton JS. Estadística para biología y ciencias de la Salud. México: McGraw-Hill; 2001.
- Norman GR, Streiner DL. Bioestadística. México: Mosby/Doyma Libros; 1996.
- Rosner B. Fundamentals of biostatistics. 7.ª edición. Boston: Brooks/Cole, Cengage Learning; 2011.