

# Estadística Descriptiva

## Estadísticos

Edimer David Jaramillo - Bioestadística 1

Agosto de 2018

- 1 **Medidas de resumen para una variable**
- 2 **Medidas de tendencia central**
- 3 **Medidas de dispersión**
- 4 **Medidas de posición**
- 5 **Medidas de distribución o forma**

# Medidas de resumen para una variable

- Las medidas de resumen permiten caracterizar la información que aportan los  $n$  datos de la muestra considerada.
- Los **estadísticos** o **valores típicos** pretenden caracterizar o describir la muestra, indicando el valor de la variable hacia el cual tienden a agruparse los datos.

# Medidas de tendencia central

- Las medidas de centralización son las más importantes, aunque no basta con ellas para resumir la información.

- **Media aritmética ( $\bar{x}$ ):** el promedio es la suma de los valores  $x_i$  de las observaciones dividido por el tamaño  $n$  de la muestra.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- La media es susceptible a valores extremos.
- Cuando se tienen variables discretas, es posible que el resultado de  $\bar{x}$  no pertenezca al conjunto de valores posibles (**ej.** el número de hermanos).

- **Mediana ( $Me$ ):** la mediana es el valor central en el sentido del orden. La mediana deja por encima y por debajo de ella, el 50% de los datos.
  - No se afecta por valores extremos.
  - En distribuciones simétricas, la media es aproximadamente igual a la mediana.
  - En distribuciones con sesgo a la derecha (sesgo positivo), la media tiende a ser mayor que la mediana.
  - En distribuciones con sesgo a la izquierda (sesgo negativo) la media tiende a ser menor que la mediana.



- **Moda ( $M_o$ ):** la moda es el valor que ocurre con mayor frecuencia, es decir, el dato que más se repite.
  - Unimodal: una sola moda
  - Bimodal: dos modas
  - Multimodal: más de dos modas
- **Media ponderada ( $M_p$ ):** es una medida donde todas las observaciones no poseen el mismo grado de importancia.

# Medidas de dispersión

Las medidas de dispersión expresan el grado de homogeneidad o heterogeneidad de los datos en torno a un valor central.

- **Rango:** es la diferencia entre el valor máximo y el mínimo.

$$\text{Máximo} - \text{Mínimo}$$

- **Varianza( $S^2$ ):** proporciona una medida de dispersión relativa al tamaño muestral de las distintas observaciones respecto a la media aritmética  $\bar{x}$ . Se define como la suma de los cuadrados de las diferencias entre cada valor de la muestra y su media.

- Para la población:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$$

- Para la muestra:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

- **Desviación estándar ( $S$ ):** se define como la raíz cuadrada de la varianza. Tiene la propiedad de estar definida en las mismas unidades de la variable analizada.

- Para la población:

$$\sigma = \sqrt{\sigma^2}$$

- Para la muestra:

$$S = \sqrt{S^2}$$

- **Coefficiente de variación (CV):** se define como el cociente entre la desviación típica y la media aritmética, expresado en porcentaje.

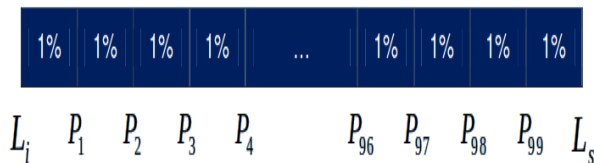
$$CV = \frac{S}{\bar{x}} \times 100$$

# Medidas de posición

# Percentiles

Los percentiles, cuantiles y deciles indican la posición relativa de un dato respecto al resto de la muestra.

- **Percentiles:** dividen la muestra en 100 partes iguales. Son los valores que dejan determinado porcentaje de observaciones por debajo de ellos. Por ejemplo, el percentil 10, es el valor superior al 10% de los datos, pero inferior al 90% restante.



Ubicación del percentil  $k$ :

$$P_k = \frac{k(n+1)}{100}$$



- **Deciles:** dividen la muestra en 10 partes iguales.

10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

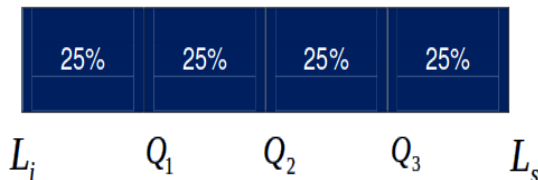
$L_i$     $D_1$     $D_2$     $D_3$     $D_4$     $D_5$     $D_6$     $D_7$     $D_8$     $D_9$     $L_s$

Ubicación del decil  $k$ :

$$D_k = \frac{k(n+1)}{10}$$

# Cuartiles

- **Cuartiles:** dividen la muestra en cuatro partes iguales.
  - Cuartil 1: 25% de las observaciones
  - Cuartil 2: 50% de las observaciones
  - Cuartil 3: 75% de las observaciones

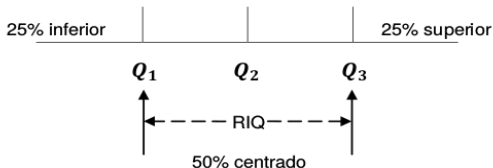


Ubicación del cuartil  $k$ :

$$Q_k = \frac{k(n+1)}{4}$$

# Rango intercuartílico

- **Rango intercuartil:** el rango intercuartílico (RIQ) se obtiene como la diferencia entre los cuartiles 3 ( $Q_3$ ) y 1 ( $Q_1$ ). La mitad de las observaciones se clasifican dentro de este rango.



$$RIQ = Q_3 - Q_1$$

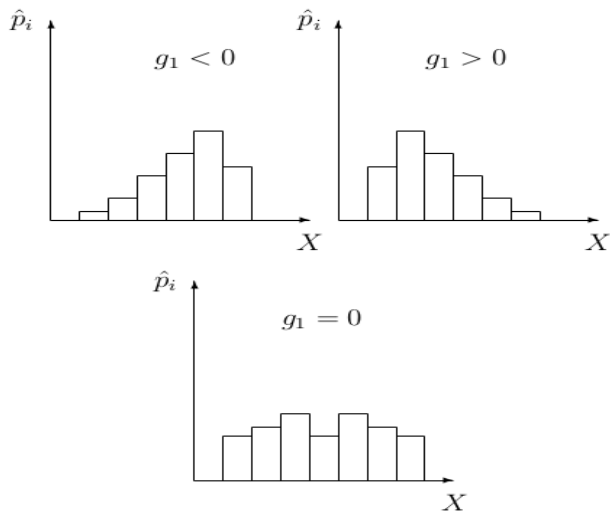
# Medidas de distribución o forma

- **Coficiente de asimetría ( $g_1$ ):** indica el grado de simetría de la distribución de datos. Puede ser:
  - Distribución simétrica si  $g_1 = 0$
  - Distribución asimétrica negativa si  $g_1 < 0$
  - Distribución asimétrica positiva si  $g_1 > 0$

Fórmula del coeficiente de asimetría:

$$g_1 = \left[ \frac{n}{(n-1)(n-2)} \right] \left[ \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{S^3} \right]$$

# Coefficiente de asimetría



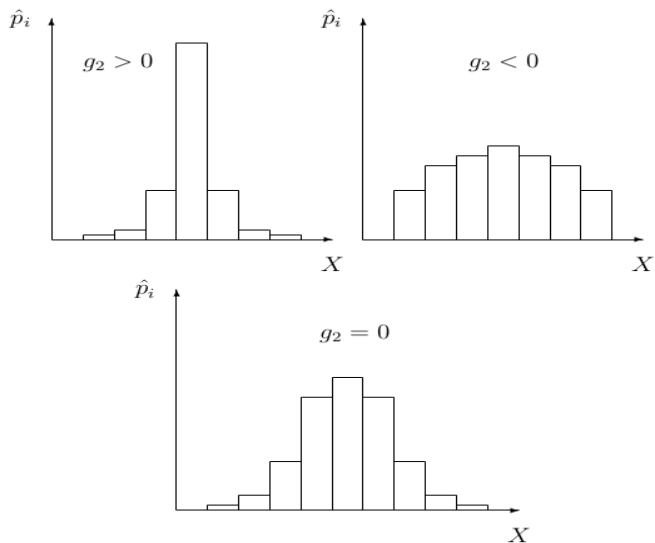
# Coefficiente de curtosis

- **Coefficiente de curtosis ( $g_2$ ):** evalúa el grado de concentración de los datos alrededor de la media y de los extremos (o colas). Puede ser:
  - Distribución mesocúrtica si  $g_2 = 0$
  - Distribución leptocúrtica si  $g_2 > 0$
  - Distribución platicúrtica si  $g_2 < 0$

Fórmula del coeficiente de curtosis:

$$g_2 = \left[ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \right] \left[ \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{S^4} \right] - \left[ \frac{3(n-1)^2}{(n-2)(n-3)} \right]$$

# Coefficiente de curtosis





GRACIAS!

