

Precurso de Estadística

1

Repasar las medidas de tendencia central (media, mediana y moda)

2

Repasar cómo la media, mediana y moda son afectadas por la asimetría

3

Repasar las medidas de dispersión



- **Medidas de Tendencia Central:** proveer información descriptiva sobre el valor numérico que es considerado el más usual para una variable cuantitativa:
 - Media**
 - Mediana**
 - Moda**
- **Asimetría** en la distribución de datos. Efecto de la media, mediana y moda.
- **Medidas de Variabilidad:**
 - Rango**
 - Varianza**
 - Desvío Estándar**

La **media** se define de la siguiente manera:

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

Por ejemplo, para la muestra 8, 5 y -1, su media es:

$$\bar{x} = \frac{8 + 5 + (-1)}{3} = 4$$

La **mediana** puede pensarse de manera simple como el valor del "medio" de una lista ordenada de datos (o el valor que separa la primera mitad y la segunda mitad de una distribución).

Para una lista ordenada la mediana es calculada de diferente manera dependiendo de la cantidad de elementos de la misma:

- **Impar:**

[1, 2, 3, 5, **7**, 8, 9, 10, 15]

#elementos: 9

La mediana es el valor de la posición 5 (la posición del "medio")

Mediana = 7

- **Par:**

[-5, -1, 0, **1**, **2**, 3, 8, 20]

#elementos: 8

La mediana es la media de los valores en las dos posiciones centrales

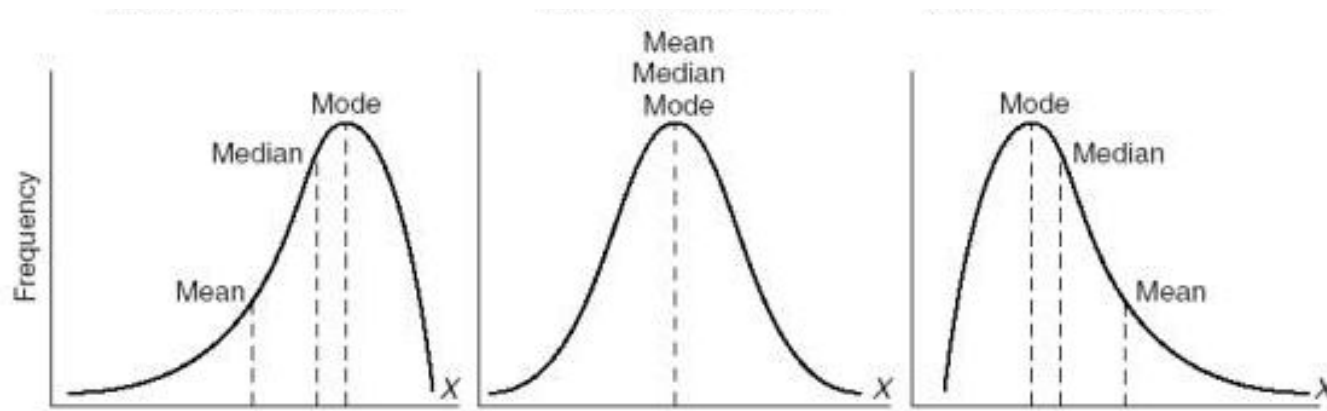
Mediana = $(1+2)/2 = 1.5$

La **moda** es el valor que aparece con mayor frecuencia o más veces en la distribución.

Por ejemplo, la moda de $[0,1,1,2,2,2,2,3,3,4,4,4,5]$ es 2.

La moda no es necesariamente única. Puede ocurrir que haya dos valores diferentes que sean los más frecuentes. Por ejemplo, para $[10, 13, 13, 20, 20]$, tanto 13 como 20 son la moda.

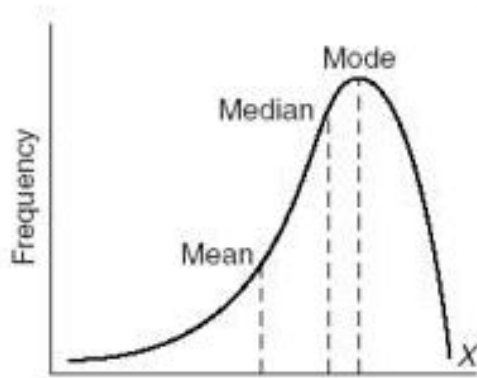
Nos referimos a la **asimetría** en cuanto a la distribución de los datos¹:



- Una distribución con **asimetría a derecha** significa que la cola del lado derecho es más larga que la de la izquierda (gráfico a la derecha)
- De la misma manera, una distribución con **asimetría a izquierda**, significa que la cola de la izquierda es más larga que la de la derecha (gráfico a izquierda).
- Por último, una **distribución simétrica** no presenta este fenómeno dado que sus colas son de igual longitud al ser simétrica.

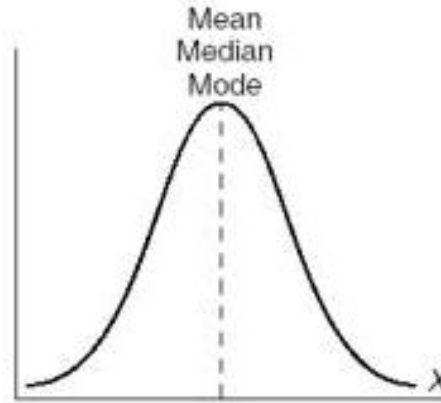
¹. Escribiremos hablando de asimetría en el contexto de distribuciones unimodales

La media, mediana y moda son afectadas por la asimetría:



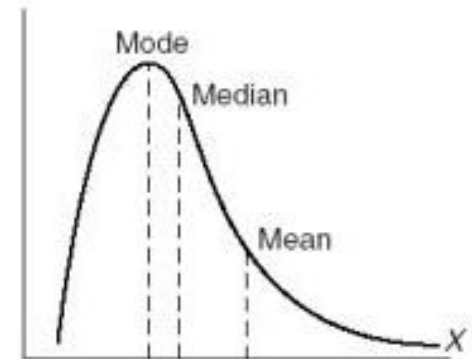
Asimetría a izquierda

$\text{Media} < \text{Mediana} < \text{Moda}$



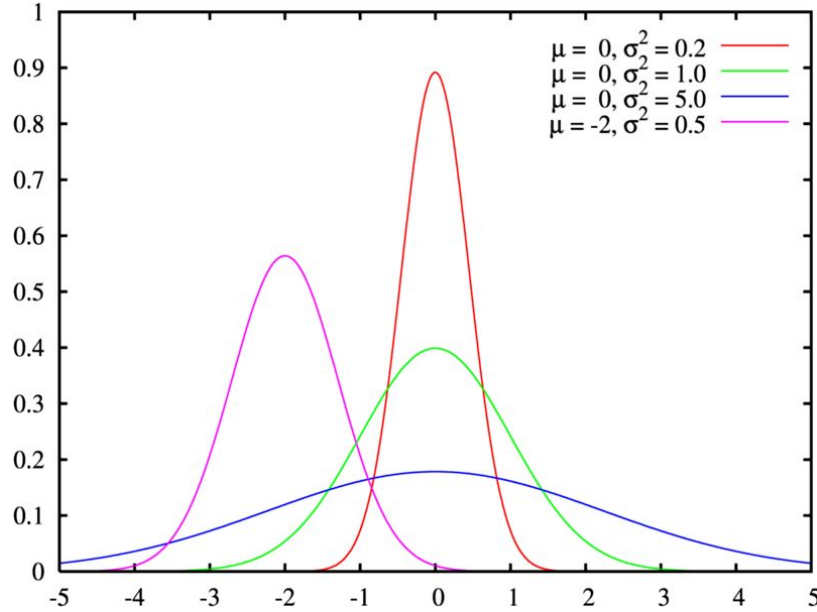
Simetría

$\text{Media} = \text{Mediana} = \text{Moda}$

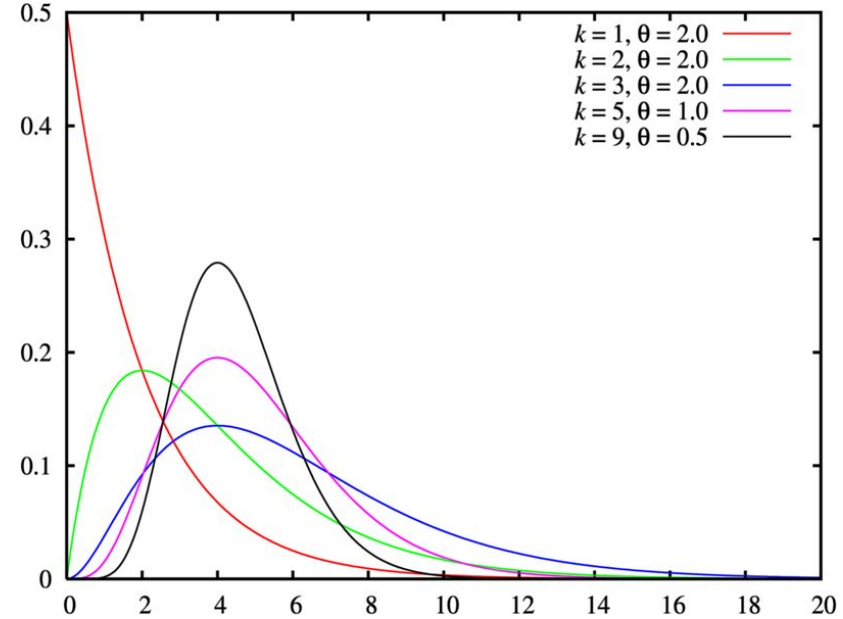


Asimetría a derecha

$\text{Moda} < \text{Mediana} < \text{Media}$



La distribución normal es un ejemplo de distribución **simétrica**



La distribución gamma es un ejemplo de distribución **asimétrica**

Las medidas de variabilidad indican cómo los datos están esparcidos. Nos vamos a focalizar en:

- **Rango**
- **Varianza**
- **Desvío estándar**

El **rango** es la diferencia entre el valor más bajo y más alto de la distribución.

La **varianza** es un valor numérico utilizado para describir cuánto varían los números de una distribución respecto a su media.

La varianza puede ser calculada como:

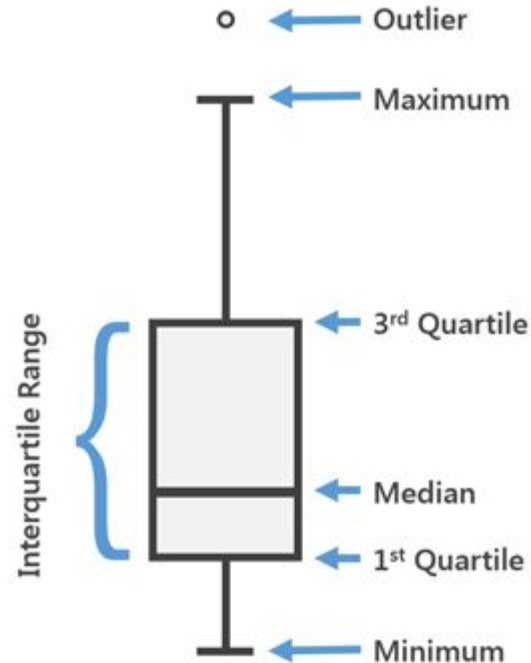
$$s^2 = \frac{\sum (X - \bar{X})^2}{N}$$

Esto es el **promedio de la diferencia elevada al cuadrado entre cada valor y la media**.

El **desvío estándar** es la raíz cuadrada de la varianza.

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

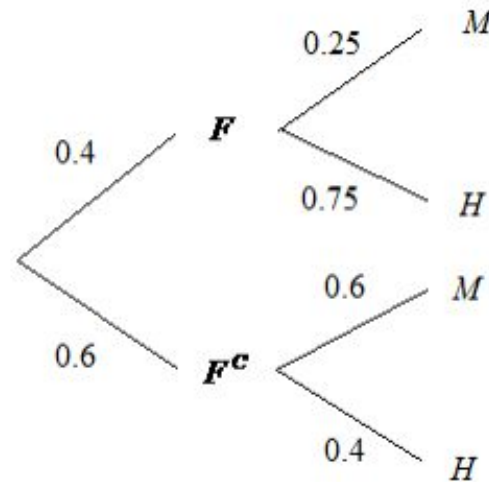
- El desvío es una medida de la dispersión de los datos
- NO ES la desviación promedio con respecto de la media. Como los desvíos están elevados al cuadrado los desvíos muy grandes cuentan más que proporcionalmente.



Notas:

- El máximo es $1,5 \text{ el IR} + 3^{\text{rd}} \text{ Q.}$
- El mínimo es $1^{\text{st}} \text{ Q} - 1,5 \text{ IR}$
- El primer cuartil equivale al 25 percentil. Es decir, es el valor que deja al 25% inferior a la izquierda.

Un médico ha observado que el 40% de sus pacientes fuma y de estos, el 75% son hombres. Entre los que no fuman, el 60% son mujeres. Calcula la probabilidad de:



1. Un paciente no fumador sea hombre.
2. Un paciente sea hombre y fume
3. Un paciente sea mujer.