



**Universidad Maimónides**

---

# Estadística Aplicada

---

Universidad Maimónides  
2016

Clase 2. Medidas de Tendencia Central y Dispersión

Pedro Elosegui

# Medidas de Descripción del Conjunto de Datos

- Estadísticas sumarias que permiten describir las características de un conjunto de datos
- Las características de una población se denominan Parámetros
- Las características de una muestra se denominan estadísticos
- Resumen de la Información
  - Tendencia Central
  - Dispersión
  - Simetría
  - Sesgo
  - Kurtosis

# Medidas de Tendencia Central

- Estadísticas sumarias que permiten describir las características de un conjunto de datos
- Las medidas de tendencia central se refieren a la posición de una distribución de frecuencia de los datos
- Las medidas típicas corresponden a la Media, la Mediana y la Moda
- La más común se refiere a la Media Aritmética o Promedio

# Medidas de Tendencia Central

- La media aritmética de una población se identifica como el parámetro
- La media aritmética de una muestra se identifica como el estadístico

$$\mu = \frac{\sum x}{N}$$

- En ambos casos corresponde a la suma de los valores de las observaciones (de la población o de la muestra) dividido el número de elementos en la muestra

$$\bar{X} = \frac{\sum x}{n}$$

# Medidas de Tendencia Central

- La media aritmética es simple de calcular y describe la tendencia central de los datos bajo estudio
- Permite comparar los datos
- Es sensible a la presencia de datos extremos
- Es en general, la medida sumaria de tendencia central más utilizada

# Medidas de Tendencia Central

## - La **Mediana**

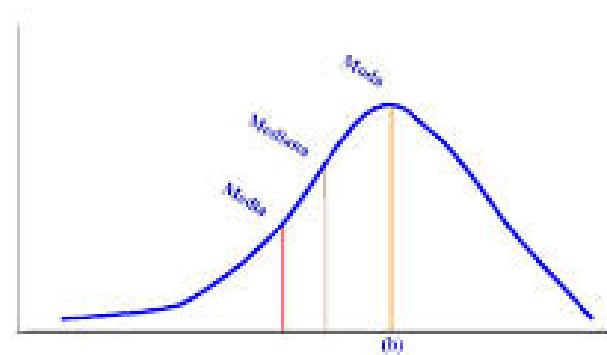
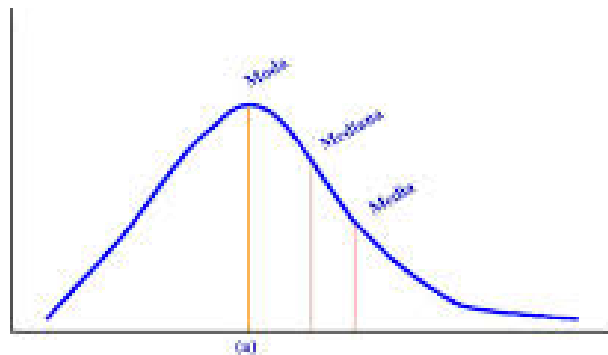
- Valor que toma la observación central del conjunto ordenado de datos
- Valor ubicado en la mitad del conjunto ordenado tal que existe el mismo número de observaciones antes que después de ese número
- La mitad de los elementos están por debajo y la otra mitad por encima de tal valor
- Si el número de datos es impar existe una única observación
- Si el número de datos es par se promedian las dos observaciones centrales
- No es sensible a la presencia de datos extremos

# Medidas de Tendencia Central

- La **Moda**
- Valor que se repite con más frecuencia en el conjunto de datos
- No es sensible a la presencia de datos extremos
- Puede repetirse con frecuencia más de un dato
- El conjunto de datos puede ser bimodal o multimodal
- Corresponde al techo de la distribución de frecuencia en el gráfico de campana

# Medidas de Tendencia Central

- Si la distribución tiene sesgo negativo
  - Se observa que la Media es menor que la Mediana que es menor que la Moda
- Si la distribución tiene sesgo positivo
  - Se observa que la Media es mayor que la Mediana que es mayor que la Moda





# Medidas de Dispersión

- Se refiere a la distribución de las observaciones en torno a la tendencia central
- Indica en una estadística sumaria que tan concentrada o dispersa se encuentran los datos entre si
- La variabilidad de los datos como desvío de la tendencia central

# Medidas de Dispersión

- Una forma de caracterizar la dispersión se denomina Rango
- Rango = Valor de la observación más grande – Valor de la observación más pequeña
  - Mide solamente valores extremos
  - Se ignora la varianza del resto de la población
- Rango Interfractil
- En una distribución de frecuencias el fractil indica el % de los datos que corresponden o son menores a dicha proporción
- La mediana es el fractil 0,5

# Medidas de Dispersión

- Ejemplo de fractiles
  - Cuartiles (1/4)
  - Deciles (1/10)
  - Percentiles (1/100)
- Rango Intecuartile =  $Q3 - Q1$

# Medidas de Dispersión

- La Varianza de la Población
- Es la medida más usual de dispersión para una población  $\sigma^2$
- Se expresa en unidades al cuadrado por eso suele hablarse del desvío estándar  $\sigma$

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N} = \frac{\sum x^2}{N} - \mu^2$$

- Teorema de Chebyshev
  - El 68% de los valores de la población se ubican dentro del 1 desvío de la media
  - El 95% dentro de 2 desvíos de la media
  - El 99 % dentro de 3 desvíos de la media

# Medidas de Dispersión

- La Varianza de la Muestra
- Es la medida más usual de dispersión para una muestra y se indica como  $S^2$

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

- El desvío estándar corresponde a s
- Fórmula ( se divide por n-1)

# Medidas de Dispersión

- Cómo puede conocerse si un determinado desvío estándar es importante respecto al valor de la media?
- Puede ser importante considerar un valor relativo entre los desvíos y el valor medio....
- Para obtener la Dispersión Relativa se utiliza el Coeficiente de Variación

$$CV = \frac{\sigma}{\mu}(100)$$

# Normalización

- En ocasiones las observaciones se normalizan
- Se calcula la diferencia entre cada observación y la media y se lo divide por el desvío estándar

$$Z = \frac{\bar{X} - \mu}{\sigma}$$

-Esta normalización, veremos, se vuelve relevante más adelante....