



# Curso: Estadística

## Sesión 1: Estadística Descriptiva

Lic. Jose O. Henao M.<sup>1</sup>

<sup>1</sup>Economista / Investigador del Programa Internacional de Democracia Sociedad y Nuevas Economías (PIDESONE)

Universidad Surcolombiana  / Universidad de Buenos Aires 

Feb, 2020

# Contenido

## 1 Estadística descriptiva

- Distribución de frecuencias
- Medidas descriptivas de posición central
- Medidas descriptivas de variabilidad o dispersión

## 2 Ejercicios

- Ver documento PDF y/o archivo Excel

# Definición

## Definición 1

**Estadística Descriptiva** como un Método para ordenar, resumir o expresar en alguna forma las características de un conjunto de números y **Inferencia Estadística** campo que se ocupa de la formulación de generalizaciones, de predicciones y estimaciones de relaciones entre dos o más variables.<sup>a</sup>

---

<sup>a</sup>Introducción al análisis estadístico - Harnett, D. Murphy, J.

## Definición 2

**Estadística Descriptiva** como resúmenes de datos en forma de tabla, gráfica y números y, la **Inferencia Estadística** como el proceso de usar datos obtenidos de una muestra para efectuar estimaciones o probar hipótesis acerca de las características de una población.<sup>a</sup>

---

<sup>a</sup>Estadística para Administración y Economía - Anderson, D.; Sweeney, D y Williams, T.

# Definiciones

## Conceptos iniciales (Parte 1)

Algunos conceptos que puede observarse en la literatura sobre **Población Estadística**:

- **Conjunto** de todos los valores en consideración (Población o Universo.)
- **Conjunto** de datos cuantificables puede llamarse una población si ese conjunto de datos está constituido por todos los valores de interés.
- **Conjunto** de todos los valores pertinentes constituye una población.

Respecto a la **Muestra Estadística** se observa que:

- **Subconjunto** de una población

# Definiciones

## Conceptos iniciales (Parte 2)

Otros conceptos que puede observarse son:

- **Observaciones:** valores individuales pertenecientes a una muestra
- **Estadísticos:** características numéricas de las muestras (también se conocen como muestrales)
- **Tabla de frecuencias:** representación de la información y no de los datos en sí mismos. También puede definirse como un método de ordenación de datos.
- **Intervalo de clase:** longitud de cada categoría que compone una tabla de frecuencia. Su tamaño debe igual para todas las categorías definidas.

# Definiciones

## Conceptos iniciales (Parte 3)

Algunos consejos para definir los intervalos de clases:

- Es aconsejable que tenga **menos de 20 clases**
- Evitarse los intervalos con extremos abiertos. Ej. "menor que 65" o "mayor que 90".
- Ninguna observación puede estar en más de un intervalo
- punto medio de cada categoría representa los valores asignados a ella corresponde a "marca de clase".

Método de calculo para definir los intervalos o para calcular la amplitud de cada clase (**Formula de Sturges**):

$$C = \frac{R}{1+3.322 \log n}$$

Siendo: **R** el recorrido (Max-Min+1); **n** total de datos

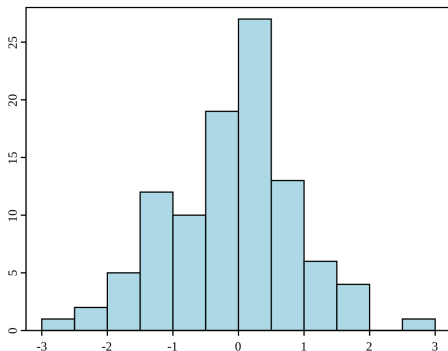
# Definiciones

## Conceptos iniciales (Parte 4)

- **Frecuencia relativa:** se obtiene dividiendo la frecuencia de cada clase entre el número total de observaciones

$$\text{Frecuencia Relativa} = \frac{\text{Frecuencia de clase}}{n}$$

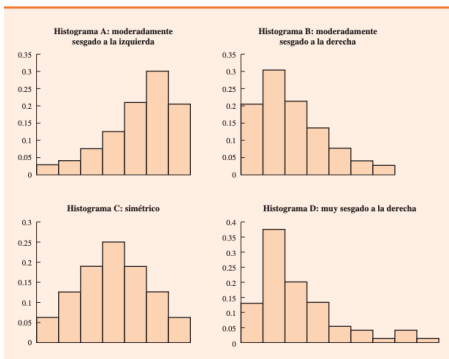
- **Histograma:** Gráfica que representa las calases en el eje horizontal y las frecuencias en el vertical (diagrama de barras).



# Definiciones

## Conceptos iniciales (Parte 5)

Podemos afirmar sobre el histograma que existe niveles de sesgo como: 1) moderadamente sesgado a la izquierda, 2) moderadamente sesgado a la derecha, 3) simétrico y 4) muy sesgado a la derecha:

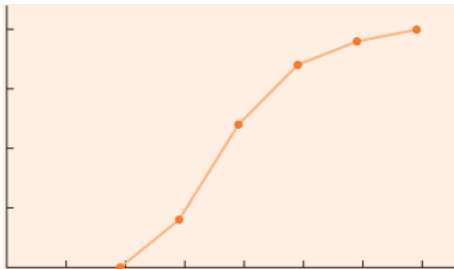




# Definiciones

## Conceptos iniciales (Parte 6)

- **Frecuencia Relativa:** representa la suma de las frecuencias desde la clase inferior hasta la superior. Su gráfica se conoce como ojiva, la cual es suavizada mediante una línea que conecta los puntos en los límites de clase (marcas de clase).



# Medidas de Resumen

## Conceptos Parte (1)

Las medidas descriptivas que resumen los datos se pueden observar dos principalmente: 1) las de posición central de los datos y 2) la variabilidad o dispersión de las observaciones.

las medidas de resumen que se refieren a una población completa, se llaman parámetros y cuando corresponde a una muestra se llaman estadísticas.

Las medidas de resumen más comunes:

**Tab. 1.** Medidas de resumen

Posición central	Variabilidad
Media aritmética	Desviación estándar
Mediana	Varianza
Moda	Recorrido
media geométrica	Recorrido intercuartílico

# Medidas de Resumen

## Posición Central (Parte 1)

- **La moda:** Valor que aparece el mayor número de veces ó valor que tiene la mayor frecuencia. En el caso de datos agrupados es la marca de clase del intervalo al cual le corresponde la mayor frecuencia.
- **La mediana:** es el valor central en un conjunto de números ordenados según su magnitud. Si **N es impar**, puede determinarse el valor central contando  $(N+1)/2$ . El número resultante divide a los datos en los dos grupos deseados, y representa así a la mediana. Si **N es par**, hay dos valores centrales, y se divide la mediana como el número que está a mitad de camino entre estos dos valores.

# Medidas de Resumen

## Posición Central (Parte 2)

Finalmente tenemos a **La media**, también conocida como media aritmética o promedio. Es la suma de todos los valores considerados, dividida por el número total de valores del conjunto. Se simboliza  $\mu$  para la media de una población.

### Media poblacional

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} * \sum_{i=1}^N x_i$$

### Media poblacional para una distribución de frecuencias

$$\mu = \frac{1}{N} * \sum_{i=1}^N x_i f_i$$

O la siguiente formula:

$$\mu = \sum_{i=1}^N x_i \frac{f_i}{N}$$

# Medidas de Resumen

## Medidas de dispersión (Parte 1)

- Recorrido: valor absoluto de la diferencia entre los valores máximo y mínimo.
- Desviación respecto a la media: Es la resta de cada valor del conjunto la media de los valores originales. Cada resultado se llama desviación

$$\sum_{i=1}^N (x_i - \mu) = 0$$

La suma de estas desviaciones son siempre igual a cero.

# Medidas de Resumen

## Medidas de dispersión (Parte 2)

### Desviación estándar y varianza

- Varianza: Promedio de los cuadrados de las desviaciones

#### Varianza poblacional

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

Pasos a seguir:

- 1) se calculan las desviaciones
- 2) se elevan al cuadrado
- 3) suman los resultados
- 4) divide la suma entre N

## Varianza poblacional para una distribución de frecuencias

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 f_i$$

O la siguiente formula:

$$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 \left( \frac{f_i}{N} \right)$$

## Formula abreviada para calcular la varianza

$$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 \frac{f_i}{N} - \mu^2$$

- Desviación estándar poblacional: Es la raíz cuadrada de la varianza, se denota con  $\sigma$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

# Medidas de resumen

## Otras medidas descriptivas (Parte 1)

### Percentiles, deciles y cuartiles

Conocidas como medidas de posición estos nos permite posicionarnos en un punto particular de interés.

- Cuartiles (4 partes iguales)  $\frac{n}{4}$
- Deciles (10 partes iguales)  $\frac{n}{10}$
- Percentiles (100 partes iguales)  $\frac{n}{100}$

El quinto decil y el segundo cuartil son equivalentes a la mediana



# Medidas de resumen

## Otras medidas descriptivas (Parte 2)

### Cálculo de los cuartiles

- Rango del primer cuartil:

$$\frac{1}{4}N + \frac{1}{4}$$

- Rango del segundo cuartil:

$$\frac{1}{2}N + \frac{1}{2} = \textit{Mediana}$$

- Rango del tercer cuartil:

$$\frac{3}{4}N + \frac{1}{4}$$



# Ejercicios

## Ejercicios de este capítulo

Hacer click en la imagen y esta lo llevará a mi página web:



# Bibliografía I

-  Harnett, D. & Murphy, J.,  
*Introducción al análisis estadístico.*  
Cap. 1.
-  Anderson, D. Sweeney, D & Williams, T.,  
*Estadística para Administración y Economía.*  
Cap. 3.