The background features a close-up, slightly blurred view of a document. On the left, a line graph is visible with a blue line and a shaded area. A pen is positioned on the right side of the document, pointing towards the graph. The overall color palette is a cool, light blue.

UNIDAD 1. ESTADÍSTICA DESCRIPTIVA

Probabilidad y
Estadística

SUBTEMAS

1.1 Introducción, notación sumatoria.

1.1.1 Datos no agrupados.

1.1.2 Medidas de tendencia central.

1.1.3 Medidas de posición.

1.1.4 Medidas de dispersión.

1.1.5 Medidas de forma.

1.2 Datos agrupados.

1.2.1 Tabla de frecuencia.

1.2.2 Medidas de tendencia central y de posición.

1.2.3 Medidas de dispersión.

1.2.4 Medidas de asimetría y curtosis.

1.3 Representaciones gráficas.

1.3.1 Diagrama de Dispersión.

1.4 Teorema de Chebyshev.

DISTRIBUCIÓN DE FRECUENCIAS

- Una distribución de frecuencia es un resumen tabular de datos que muestra el número (frecuencia) de elementos en cada una de las diferentes clases disyuntas (que no se superponen).
 - Para realizar una tabla de distribución de frecuencias y sus gráficas se deben realizar los siguientes pasos:
 1. Obtención de datos
 2. Orden de datos
 3. Cálculo del rango
 4. Cálculo del número de clases
 5. Cálculo del intervalo de clases
 6. Cálculo de frecuencias
 7. Cálculo de límites, marcas de clase y amplitud
 8. Graficación
-

OBTENCIÓN Y ORDEN DE LOS DATOS

- Los datos cuantitativos obtenidos durante el experimento se deberán ordenar del menor al mayor de ellos. Esta lista se ocupará para todos los procesos estadísticos descriptivos.
 - Los datos cualitativos no operan de la misma manera.
-

DISTRIBUCIÓN DE FRECUENCIAS

- RANGO: El rango es el intervalo de valores entre el dato mayor y el menor.
 - *Rango = dato mayor – dato menor*
 - NÚMERO DE CLASES: Se calcula por medio de la fórmula:
 - $NC = 1 + \left(\log n * \frac{10}{3}\right)$ NOTA: El resultado se redondea al número entero más próximo.
 - INTERVALOS DE CLASES: Se obtiene por medio de la fórmula:
 - $IC = \frac{Rango}{NC}$ NOTA: El resultado se redondea al número entero más próximo si la variable es discreta.
-

DISTRIBUCIÓN DE FRECUENCIAS

- FRECUENCIA ABSOLUTA O DE CLASES: Es la frecuencia de la variable de estudio de acuerdo a los intervalos de clases. Se denomina como F_i .
- FRECUENCIA RELATIVA: Es la parte o proporción de los elementos que perteneces a cada clase. Se denomina como Fr y se calcula por medio de la fórmula:
- $Fr = \frac{F_i}{n}$



DISTRIBUCIÓN DE FRECUENCIAS

- FRECUENCIA RELATIVA PORCENTUAL: Es la frecuencia relativa multiplicada por 100. Se denomina como $Fr\%$ y se calcula por medio de la fórmula:
 - $Fr\% = \frac{Fi}{n} \times 100$
 - FRECUENCIA ACUMULADA Y FRECUENCIA RELATIVA ACUMULADA: Es la frecuencia es el resultado de sumar sucesivamente las frecuencias absolutas o relativas, desde el menor al mayor de sus valores. Se calcula por medio de las fórmulas:
 - $F_{Ac} = Fi \text{ de la clase} + Fi \text{ de la clase anterior}$
 - $Fr_{Ac} = Fr \text{ de la clase} + Fr \text{ de la clase anterior}$
-

LÍMITES REALES Y MARCA DE CLASES

- LÍMITE REAL INFERIOR: Es el límite real de la clase aquel que se calcula por medio de la fórmula:

- $$Lri = \frac{(\text{límite inferior de la clase} + \text{límite superior de la clase anterior})}{2}$$

- LÍMITE REAL SUPERIOR: Es el límite real de la clase aquel que se calcula por medio de la fórmula:

- $$Lrs = \frac{(\text{límite superior de la clase} + \text{límite inferior de la clase posterior})}{2}$$

- MARCA DE CLASE: Es el valor medio que se obtiene a través de la fórmula:

- $$Mc = \frac{Lri + Lrs}{2}$$

- AMPLITUD: Se obtiene a través de la fórmula:

- $$C = Lrs - Lri$$

EJERCICIOS



Tomados del libro *Probabilidad y Estadística para Ingeniería y Ciencias* de Walpole, Myers & Myers



CÁLCULOS ESTADÍSTICOS

DATOS NO AGRUPADOS



MEDIDAS DE TENDENCIA CENTRAL

- Media aritmética:

- $\bar{x} = \frac{\sum x_i}{n}$

- Mediana:

- *Si n es impar, la Md es el dato medio de la lista ordenada de datos.*
- *Si n es par, la Md es la \bar{x} de los dos datos medios de la lista ordenada de datos.*

- Moda

- *La Mo es el dato que más se repite.*
-

MEDIDAS DE VARIACIÓN

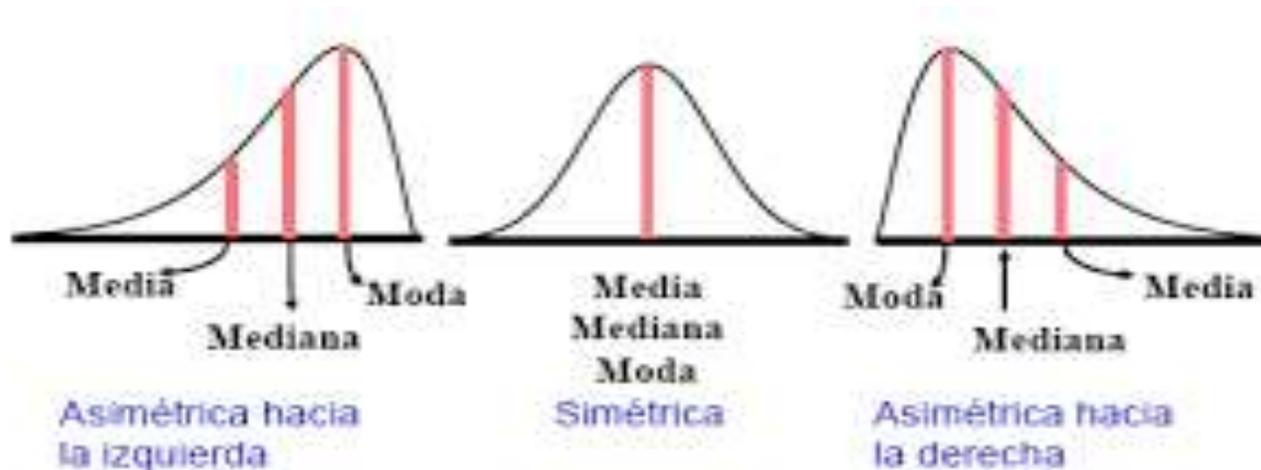
- Varianza: variación cuadrada de los datos con respecto a la media
 - $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$ varianza muestral
 - $s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$ varianza poblacional
 - Desviación estándar: variación de los datos con respecto a la media
 - $s = \sqrt{s^2}$
 - Coeficiente de variación: proporción de la desviación estándar respecto a la media
 - $CV = \frac{s}{\bar{x}}$
-

MEDIDAS DE DISTRIBUCIÓN

- Sesgo: indica la simetría de la distribución de una variable respecto a la media aritmética.

- $Sesgo = \frac{\bar{x} - Mo}{s}$ *Primer Coeficiente de Pearson*

- $Sesgo = \frac{3(\bar{x} - Md)}{s}$ *Segundo Coeficiente de Pearson*

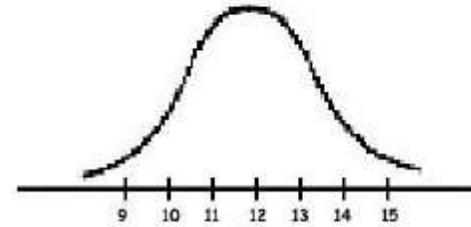


MEDIDAS DE DISTRIBUCIÓN

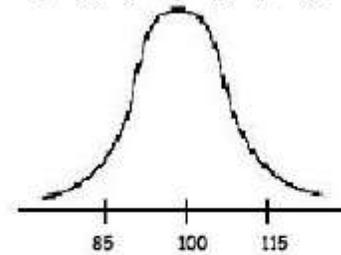
- Kurtosis

- $Kurtosis = \frac{\sum(x_i - \bar{x})^4}{s^4} - 3$

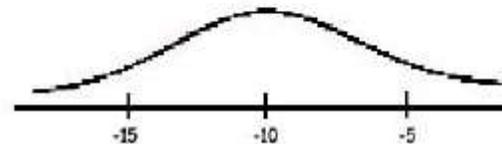
Kurtosis



Kurtosis = 0



Kurtosis > 0, *Leptokurtic*



Kurtosis < 0, *Platykurtic*

MEDIDAS DE LOCALIZACIÓN

- Cuartiles. Valores que dividen una muestra de datos en cuatro partes iguales.

- Localización del cuartil

- $LocQ_i = \frac{ni}{4}$

- NOTA: Si $LocQ_i$ no es un entero, entonces la posición i es el siguiente entero más grande.

Si $LocQ_i$ es un entero, entonces la posición i será $LocQ_i+0.5$ y así el valor del percentil es el promedio de las observaciones ordenadas $LocQ_i$ y $LocQ_i+1$

- Cálculo del cuartil

- $Q_i =$ será el dato en la posición localizada
-

MEDIDAS DE LOCALIZACIÓN

- Deciles. Valores que dividen una muestra de datos en diez partes iguales.

- Localización del decil

- $LocD_i = \frac{ni}{10}$

- NOTA: Si $LocD_i$ no es un entero, entonces la posición i es el siguiente entero más grande.

Si $LocD_i$ es un entero, entonces la posición i será $LocD_i+0.5$ y así el valor del percentil es el promedio de las observaciones ordenadas $LocD_i$ y $LocD_i+1$

- Cálculo del decil

- $D_i =$ será el dato en la posición localizada
-

MEDIDAS DE LOCALIZACIÓN

- Percentiles. Valores que dividen una muestra de datos en cien partes iguales.

- Localización del percentil

- $LocP_i = \frac{ni}{100}$

- NOTA: Si $LocP_i$ no es un entero, entonces la posición i es el siguiente entero más grande.

Si $LocP_i$ es un entero, entonces la posición i será $LocP_i+0.5$ y así el valor del percentil es el promedio de las observaciones ordenadas $LocP_i$ y $LocP_i+1$

- Cálculo del Percentil

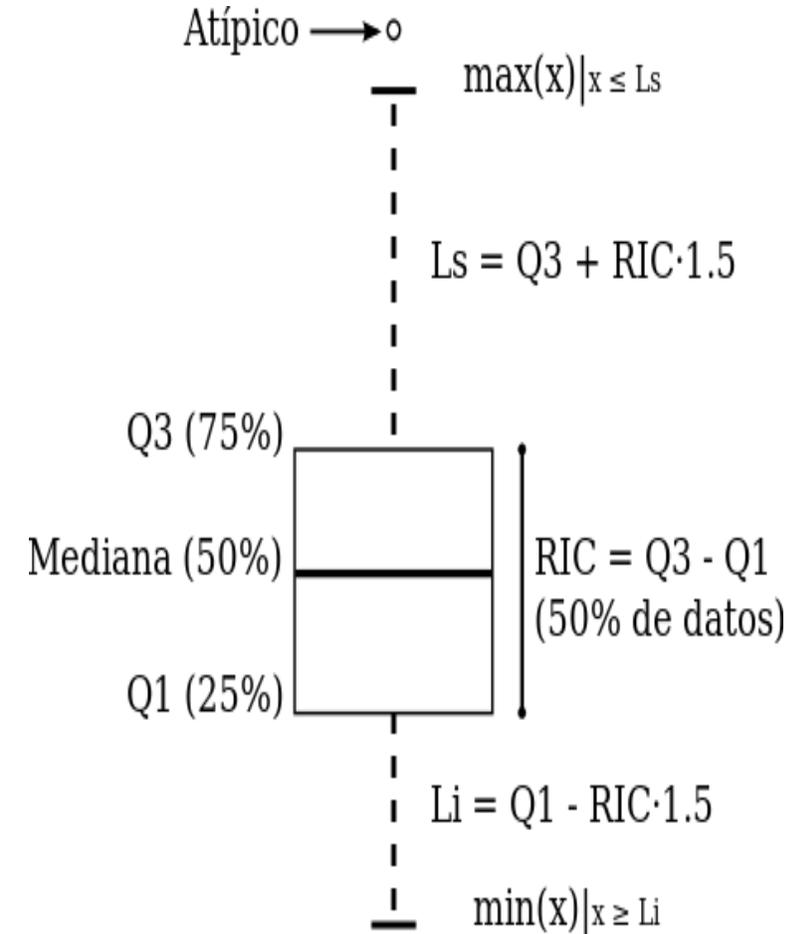
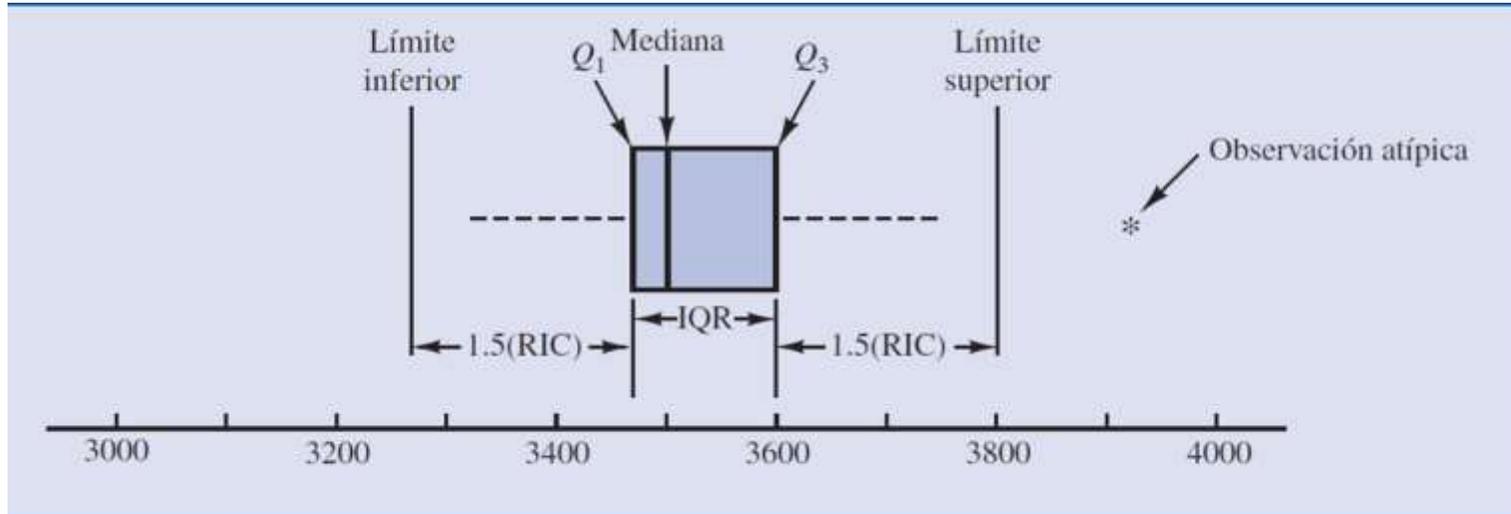
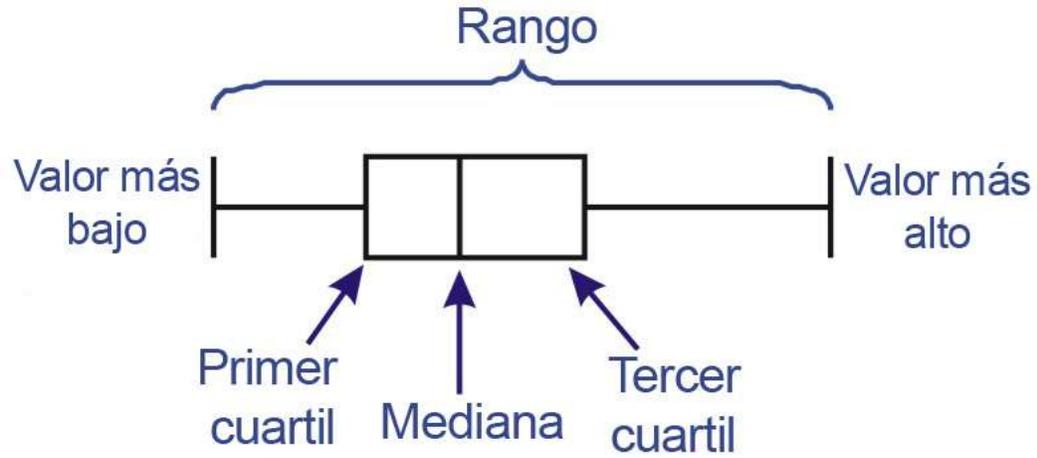
- $P_i =$ será el dato en la posición localizada
-

DIAGRAMA DE CAJA

- Una gráfica de este tipo consiste en una **caja** rectangular, donde los lados más largos muestran el **recorrido intercuartílico**. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero (recordemos que el segundo cuartil coincide con la mediana).

Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman **bigotes**. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente.

DIAGRAMA DE CAJA



EJERCICIOS



Tomados del libro *Probabilidad y Estadística para Ingeniería y Ciencias* de Walpole, Myers & Myers

DATOS AGRUPADOS

Estado Conyugal	Total	Hombre	Mujer
Soltero	163.717	40,1%	31,6%
Casado	132.744	28,6%	29,6%
Unido	107.638	23,3%	23,9%
Separado	25.377	4,4%	6,8%
Viudo	17.349	1,9%	5,7%
Divorciado	9.669	1,8%	2,5%

Fuente: Tomado de la página de Instituto Nacional de Estadísticas y Censos

MEDIDAS DE TENDENCIA CENTRAL

- Media aritmética:

- $\bar{x} = \frac{\sum F_i x_i}{n}$

- *Nota: la MC serán las x_i*
-

MEDIDAS DE TENDENCIA CENTRAL

- Mediana:
- Localización de la mediana (en la columna de la frecuencia acumulada)
- $LocMd = \frac{n}{2}$
- Cálculo de la mediana
- $Md = Lri + \left(\frac{LocMd - FAc\ anterior}{Fi} \right) * C$

MEDIDAS DE TENDENCIA CENTRAL

- Moda:
 - Localización de la moda (en la columna de la frecuencia acumulada)
 - *LocMo = la F_i de mayor valor*
 - Cálculo de la moda
 - *$Md = Lri + \left(\frac{A_1}{A_1+A_2}\right) * C$*
 - Donde:
 - *$A_1 = F_i$ de mayor valor – F_i anterior*
 - *$A_2 = F_i$ de mayor valor – F_i posterior*
-

MEDIDAS DE VARIACIÓN

- Varianza: variación cuadrada de los datos con respecto a la media

- $s^2 = \frac{\sum F_i(x_i - \bar{x})^2}{n-1}$ varianza muestral

- $S^2 = \frac{\sum F_i(x_i - \bar{x})^2}{n}$ varianza poblacional

- Desviación estándar: variación de los datos con respecto a la media

- $s = \sqrt{s^2}$

- Coeficiente de variación: proporción de la desviación estándar respecto a la media

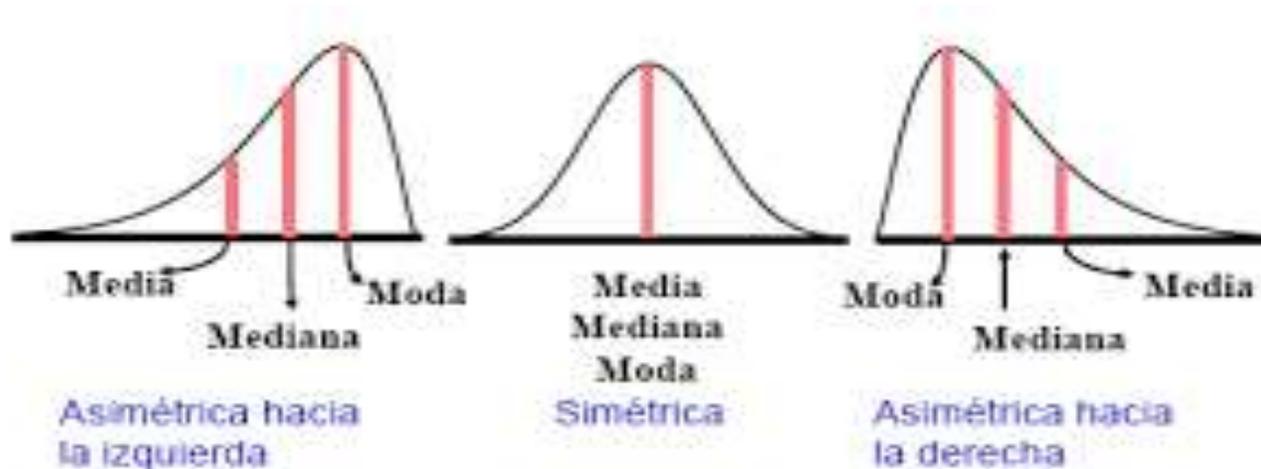
- $CV = \frac{s}{\bar{x}}$

MEDIDAS DE DISTRIBUCIÓN

- Sesgo: indica la simetría de la distribución de una variable respecto a la media aritmética.

- $Sesgo = \frac{\bar{x} - Mo}{s}$ *Primer Coeficiente de Pearson*

- $Sesgo = \frac{3(\bar{x} - Md)}{s}$ *Segundo Coeficiente de Pearson*

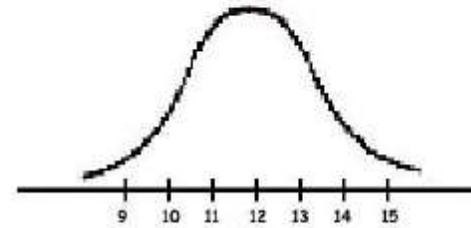


MEDIDAS DE DISTRIBUCIÓN

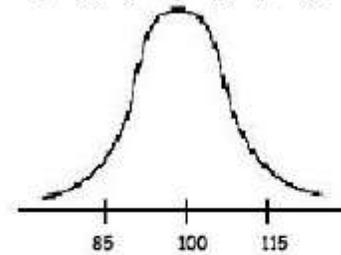
- Kurtosis

- $Kurtosis = \frac{\sum F_i(x_i - \bar{x})^4}{s^4} - 3$

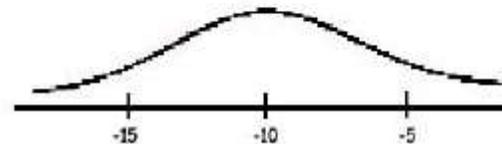
Kurtosis



Kurtosis = 0



Kurtosis > 0, *Leptokurtic*



Kurtosis < 0, *Platykurtic*

MEDIDAS DE LOCALIZACIÓN

- Cuartiles. Valores que dividen una muestra de datos en cuatro partes iguales.
- Localización del cuartil
- $LocQ_i = \frac{ni}{4}$
- Cálculo del cuartil
- $Q_i = Lri + \left(\frac{LocQ_i - FAc anterior}{Fi} \right) * C$



MEDIDAS DE LOCALIZACIÓN

- Deciles. Valores que dividen una muestra de datos en diez partes iguales.
- Localización del decil
- $LocD_i = \frac{ni}{10}$
- Cálculo del decil
- $D_i = Lri + \left(\frac{LocD_i - FAc\ anterior}{Fi} \right) * C$



MEDIDAS DE LOCALIZACIÓN

- Percentiles. Valores que dividen una muestra de datos en cien partes iguales.
- Localización del percentil
- $LocP_i = \frac{ni}{100}$
- Cálculo del Percentil
- $P_i = Lr_i + \left(\frac{LocP_i - FAc\ anterior}{Fi} \right) * C$



EJERCICIOS



Tomados del libro *Estadística para Administración y Economía* de Anderson, Sweeney & Williams
