

Taller de Matemática Computacional

TUDAI

FCEX - UNICEN

Estadística

- Describir Datos.
- Representación y Cálculo de medidas de Resumen.
- Toma de decisiones objetivas.



Estadística – Conceptos Básicos

Población: Conjunto de Datos Total. Acotados en un tiempo y espacio determinados, con alguna característica común, **observable o medible**.



A veces es impracticable examinar alguna característica en la población entera, por lo que se examina una parte, subconjunto de la población.

Estadística – Conceptos Básicos

Muestra: subconjunto de elementos de la población.

Los elementos tienen características, y son las variables que nos interesa evaluar, clasificar, medir, según corresponda.



Estadística – Conceptos Básicos

Variables: son las características que pueden cambiar de un elemento a otro del espacio muestral (edad, peso, volumen, etc.). Según su característica, las variables se clasifican en **cualitativas y cuantitativas**.

Cualitativas: no medible numéricamente. Pueden ser Nominales u Ordinales, dependiendo si puede ordenarse o no.

Ej. Nominales: nacionalidad, grupo sanguíneo, color de ojos.

Ej. Ordinales: grado de satisfacción, nivel educativo.

Cuantitativas: si sus valores son numéricos. Pueden ser Discretas o Continuas, dependiendo si toma valores enteros o reales.

Ej. Discretas: cantidad de alumnos, cantidad de hijos.

Ej. Continuas: altura, peso, ingresos mensuales.

Estadística – Conceptos Básicos

Ej.) Se desea realizar un estudio estadístico de la edad de los alumnos de TMC. Se registra la edad de los 10 primeros alumnos de la lista de inscriptos, entonces:

- Población: las edad de todos los alumnos de TMC.
- Muestra: la edad de los 10 primeros alumnos de la lista.
- Individuo: cada edad.
- Variable: la edad.

Estadística – Organización de los Datos

Consiste en representar los datos de manera general, y para ello se utilizan tablas de **distribución de frecuencias**.

Puede hacerse tanto con variables cualitativas y cuantitativas.

Frecuencia Absoluta f_i : #veces que se repite un valor x_i

Frecuencia Relativa: es el cociente f_i / N , donde N es el total de datos, toma valores entre 0 y 1. Multiplicada por 100 se obtiene la **frecuencia relativa porcentual**.

Frecuencia Acumulada Absoluta: #veces que el valor es menor a x_i .

Relativa: $\text{Frec.Ac.Abs.} / N$

Porcentual: $\text{Frec.Ac.Rel.} \times 100$

Representación de Distribuciones

Variables Cualitativas:

Tipo Familia	#	%
sin hijos	1168	53,5
con hijos	830	38,02
H solos c/hijos	27	1,24
M solas c/hijos	158	7,24

- sin hijos
- con hijos
- H solos c/hijos
- M solas c/hijos

Gráfico Sectores

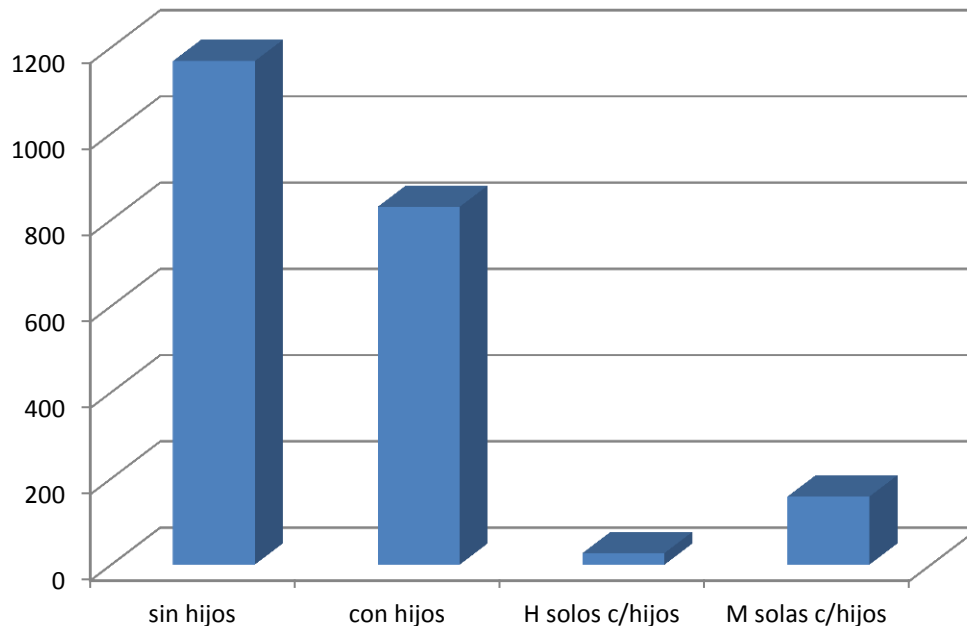
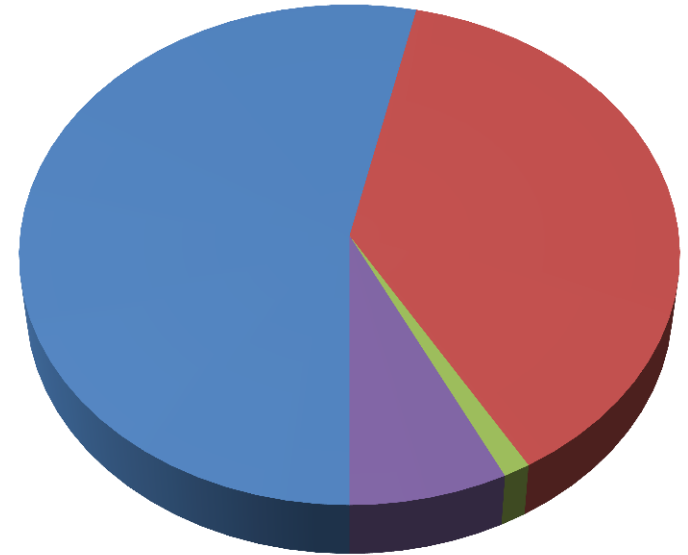


Gráfico Barras

Representación de Distribuciones

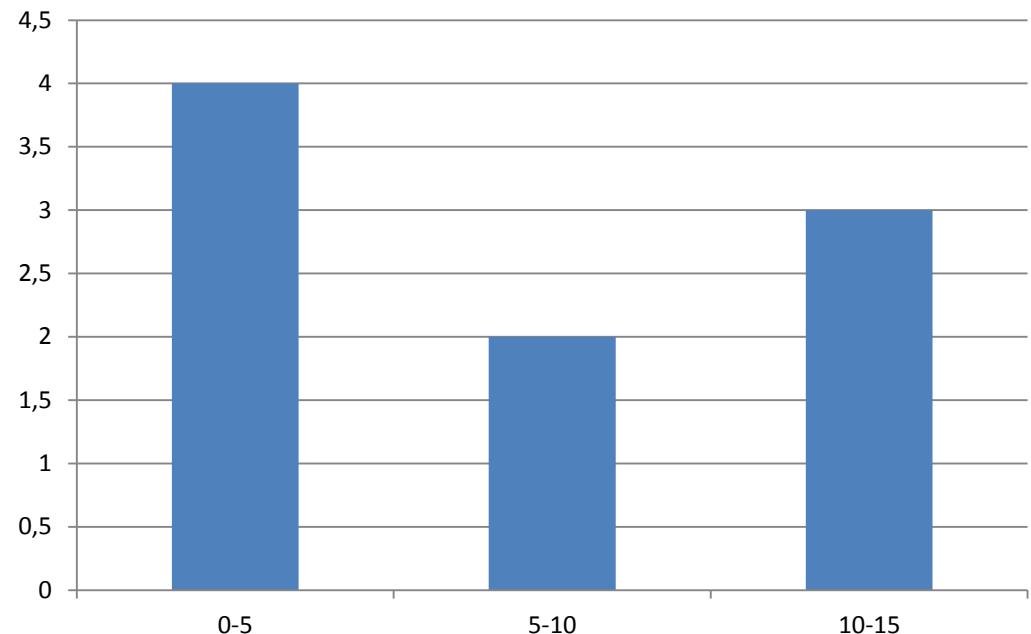
Variables Cuantitativas: Se utilizan **histogramas**. Se agrupan los valores en intervalos. Las barras deben cubrir todo el intervalo de valores.

Ej: % de alumnos > 30 años por carrera de la UNICEN

Carrera	%
Ing. Sistemas	13
Veterinaria	5,2
TUDAI	1,3
Cs. de la Educ.	14,9
Lic. En Física	3,5
Contador	2,8
Lic. En Admin.	2,3
Geografía	9,7
Ed. Inicial	11

$0 < x \leq 5$
 $5 < x \leq 10$
 $10 < x \leq 15$

Histograma de Frecuencias



Medidas de Resumen

Media, \bar{x} : se suman todos los valores y se divide por la cantidad de observaciones.

$$\bar{x} = \frac{\sum x_i}{N}$$

Mediana, M : punto medio de una distribución. La mitad de las observaciones tienen un valor menor y la otra mitad un valor mayor a M .

¿Cómo se hace?

1. Ordenar las observaciones de mínimo a máximo.
2. Si M es impar, la mediana es el valor central de la lista
3. Si M es par, se promedian los valores de las 2 posiciones centrales

Medidas de Resumen

Diferencias:

La **Media** es sensible a observaciones atípicas,
la **Mediana** no.

La **Media** está asociada al valor total,
la **Mediana** no. $\bar{x}.N = \text{total}$

Moda: es el valor que se repite más veces. Puede haber más de una. Si son 2 se dice **bimodal**. Cuando los datos están agrupados en intervalos se considera el intervalo de mayor frecuencia

Medidas de Resumen

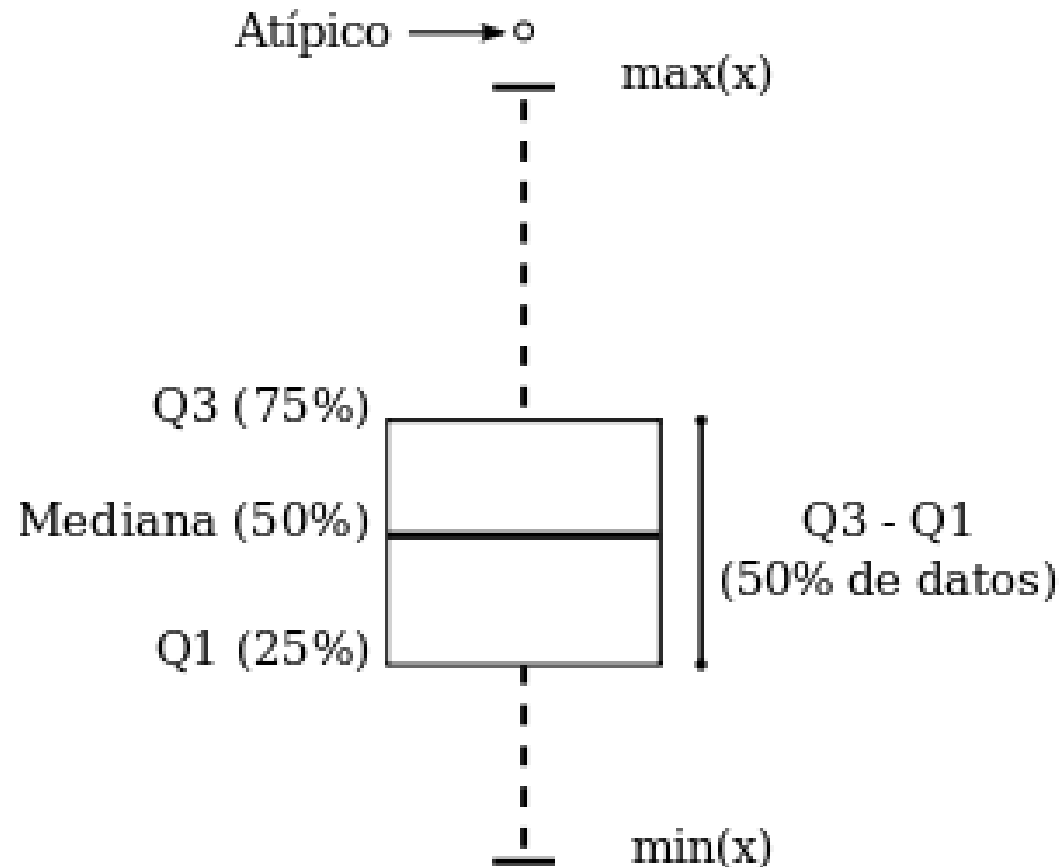
Los **Cuartiles** son medidas que separan a los datos en partes iguales. Así como la mediana los divide en **dos** partes iguales, los cuartiles en **cuatro** partes iguales. Los **deciles** en **diez** y los **percentiles** en **100**. Se calculan de manera análoga a la mediana.

El **Box-Plot** es una herramienta gráfica útil para describir la distribución de los datos. Refleja la **forma** de la distribución, acerca de su **simetría** y **sesgo**, sus cuartiles, y detectando la presencia de valores extremos y/o atípicos.

Medidas de Resumen

El Box-Plot consiste en construir una caja, cuyos límites son el Q_1 y el Q_3 , luego se ubican la media y la mediana.

Se considera a las líneas extremas que unen la caja al máximo y mínimo (o 5% y 95%, para contener en la representación al 90% central de los datos).



Medidas de Dispersión - Varianza

Una medida de dispersión es una medida de cuan alejados están los datos del centro de la distribución.

Varianza: es el promedio de los cuadrados de las desviaciones respecto de la media.

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

Si los datos fueron organizados en intervalos de clase en la expresión se utiliza la marca de clase de cada intervalo en lugar de x_i .

Medidas de Dispersión - Desvío Estándar

Se llama **Desvío Estándar** a la raíz cuadrada de la varianza. Suele ser más útil ya que tiene las mismas unidades que los datos.

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