

---

# ESTADISTICA GENERAL

- INTRODUCCIÓN
- CONCEPTOS BASICOS
- ORGANIZACIÓN DE DATOS

**Profesor del curso: Ing. Celso Gonzales**

---

# INTRODUCCION

---

## OBJETIVOS

- **Comprender qué es y porqué se estudia la estadística.**
- **Comprender los conceptos básicos de la estadística.**
- **Explicar lo que significan estadística descriptiva y estadística inferencial.**
- **Diferenciar entre una variable cualitativa y cuantitativa.**

# ¿Qué es estadística?

---

- **Recolecta**
- **Organiza**
- **Presenta**
- **Analiza**
- **interpreta**

# ¿Para qué sirve la estadística?

---

La **Estadística** se utiliza como **tecnología al servicio** de las ciencias.....

# TIPOS DE ESTADISTICA

---

- Estadística Descriptiva
- Estadística Inferencial

---

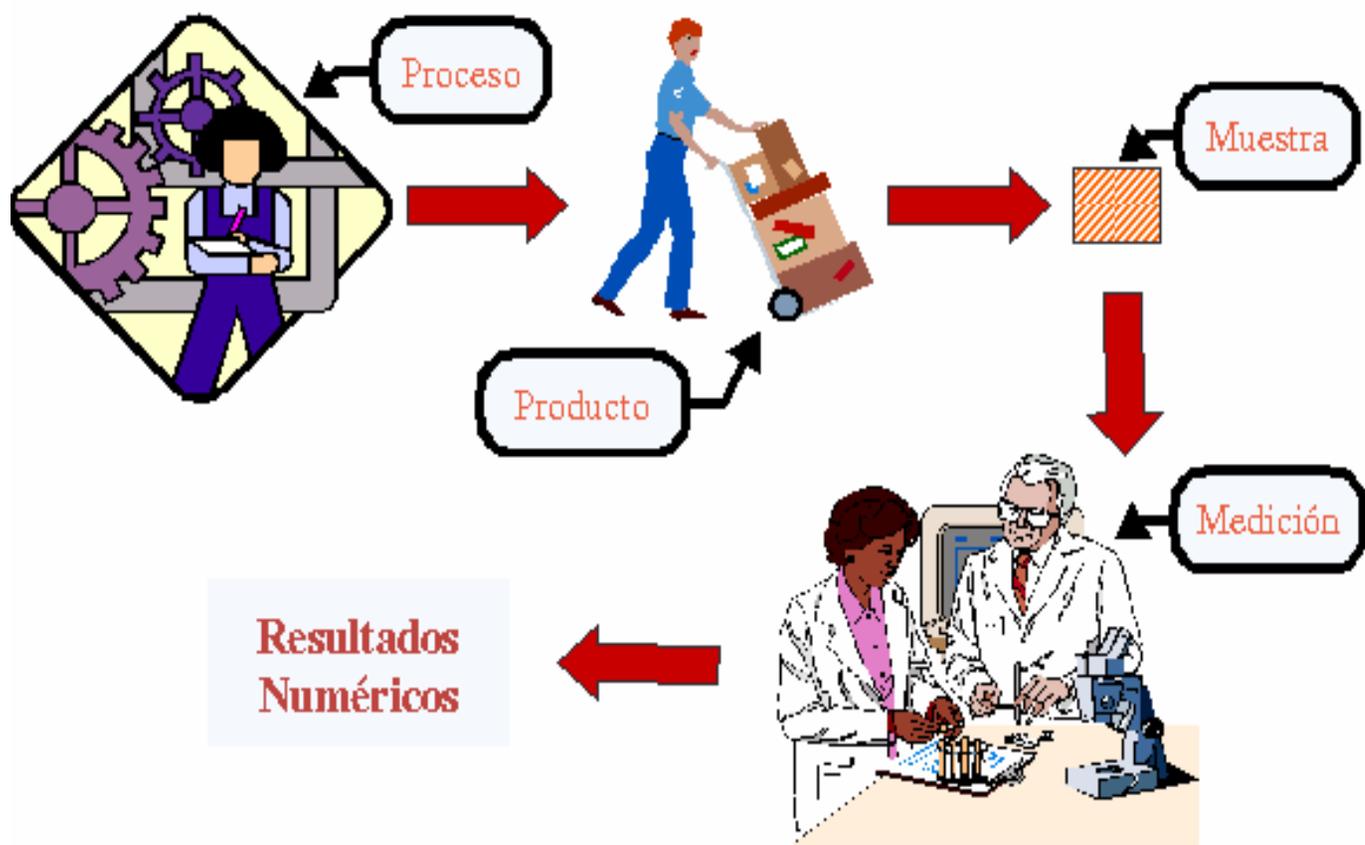
## RAMAS DE LA ESTADÍSTICA

### ESTADÍSTICA DESCRIPTIVA

SE OCUPA DE LA COLECCIÓN Y CLASIFICACIÓN DE INFORMACIÓN, DE SU RESUMEN EN CUADROS Y GRÁFICOS ADECUADOS QUE RESUMAN EN FORMA APROPIADA LA INFORMACIÓN CAPTADA.

### ESTADÍSTICA INFERENCIAL

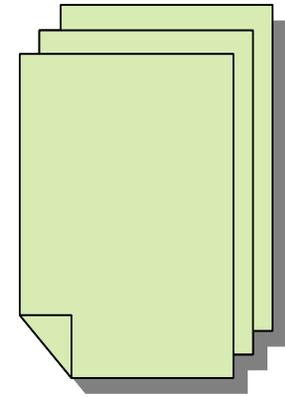
SE OCUPA DE LOS PROCESOS DE ESTIMACIÓN, ANÁLISIS Y PRUEBA DE HIPÓTESIS, CON EL PROPÓSITO DE LLEGAR A CONCLUSIONES QUE BRINDEN UNA ADECUADA BASE CIENTÍFICA PARA LA TOMA DE DECISIONES TOMANDO COMO BASE LA INFORMACIÓN CAPTADA POR LA MUESTRA.



## Pasos en un estudio estadístico

---

- Plantear **hipótesis** sobre una **población**
- Decidir qué datos recoger (diseño de experimentos)
  - Qué individuos pertenecerán al estudio (*muestras*)
  - Qué datos recoger de los mismos (*variables*)
- Recoger los datos (*muestreo*)
  - ¿Estratificado? ¿Sistemáticamente?
- **Describir** (resumir) los datos obtenidos
- Realizar una **inferencia** sobre la población
- Cuantificar la confianza en la inferencia
  - *Nivel de confianza*
  - *Margen de error:*



# CONCEPTOS BÁSICOS

---

- Población
- Muestra
- Individuo o unidad elemental
- Observación
- Variable
- Parámetro
- Estadístico

# POBLACIÓN Y MUESTRA

---

**Población** Conjunto de observaciones o datos obtenidos de una medición, conteo o cualidad de ciertos caracteres de los mismos.

## **Muestra**

- Deber ser “representativo”

# Unidad elemental

---

- Todo elemento que está afectado por la característica o factor que se desea estudiar.
- Ejemplo: Se desea hacer un estudio sobre el ingreso familiar de Lima Metropolitana  
Identifique: Población, una posible muestra, la característica en estudio y la unidad elemental.

# Observación

---

- Dato o registro realizado, producto de la apreciación de una característica en una unidad elemental.
- De una observación para el ejemplo anterior.

## EJEMPLO1

---

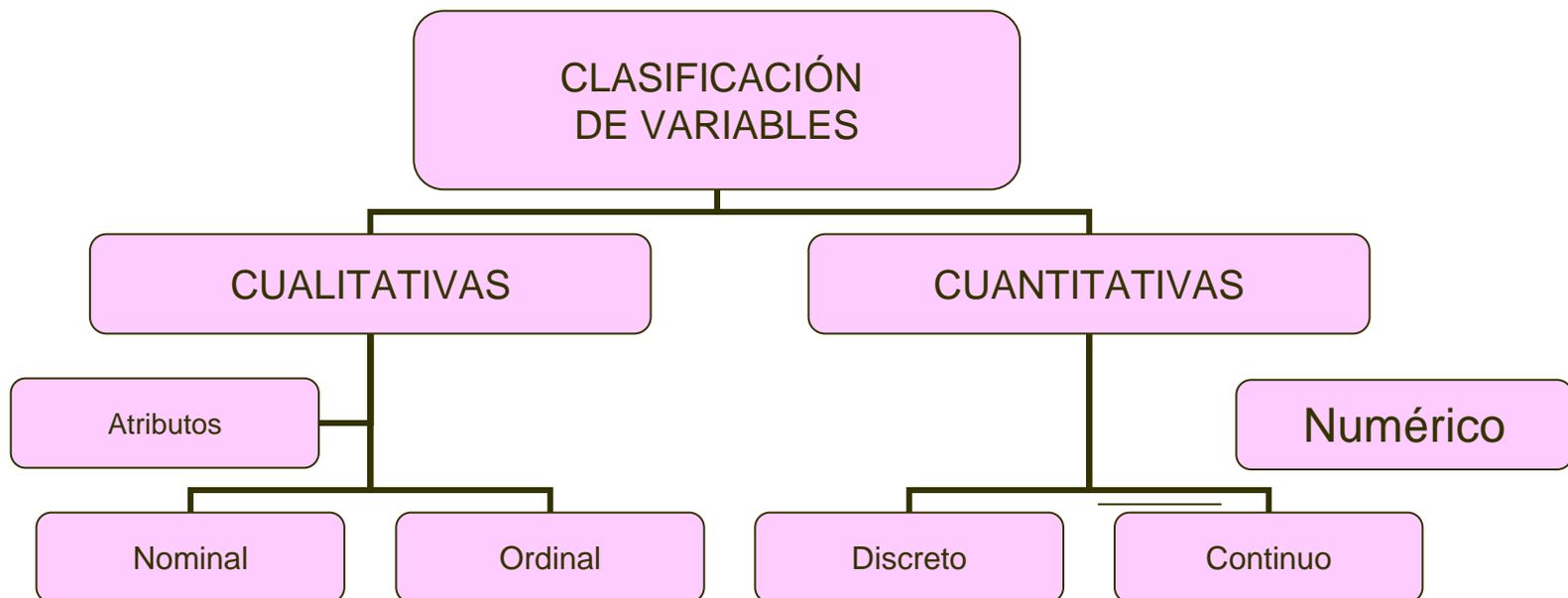
Se ha hecho un estudio para determinar la preferencia de una marca especial de detergente por parte de las amas de casa. Entre las 50 amas de casa entrevistadas, 30 dijeron que preferían esta marca.

- a. ¿Qué constituye la muestra?.
- b. ¿Qué constituye la población?.
- c. ¿Cuál es la proporción, dentro de la muestra, de las amas de casa que prefieren la marca del detergente?

# Variables

---

Son las características que se desea apreciar de los elementos de la población.



# TIPOS DE VARIABLES

---

- Cualitativas
  - **Nominales:** Si sus valores no se pueden ordenar
    - Sexo, Grupo Sanguíneo, Religión, Nacionalidad, Fumar (Sí/No)
  - **Ordinales:** Si sus valores se pueden ordenar
    - Grado de satisfacción, Intensidad del color
- Cuantitativas o Numéricas
  - **Discretas:** Si toma valores enteros
    - Número de hijos, Número de desempleados
  - **Continuas:** Si entre dos valores, son posibles infinitos valores intermedios.
    - Altura, ingreso mensual,

## EJEMPLO2

---

Clasificar cada una de las siguientes variables:

- a. Distancia diaria recorrida por cada estudiante para ir de su casa a la universidad.
- b. Tiempo que requiere un estudiante para responder a un examen.
- c. Llamadas que llegan a la central telefónica de la USB en un día.
- d. Preferencia por cierta marca de refresco.
- e. Sexo de las estudiantes que toman el curso de estadística en el semestre.
- f. Número de acciones vendidas en un día en la Bolsa de Valores.

# Parámetro

---

Es una función de todas las observaciones de una población ( $\theta$ )

MEDIDA USADA PARA DESCRIBIR UNA VARIABLE DE UNA POBLACION

El parámetro resume la información contenida en las observaciones que comprenden a una población, por lo que su **valor es único** y generalmente **desconocido** y por tato debe ser **ESTIMADO**.

# Parámetro

---

- La media  $\mu$

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N}$$

- La varianza  $\sigma^2$
- Proporción  $p$

# Estadístico o estimador

---

Es una función de los valores muestrales y que no depende de parámetro alguno. Se caracterizan porque pueden tomar valores diferentes de muestra a muestra.

$$\hat{\theta} = f(x_1, x_2, \dots, x_n)$$

Por ejemplo, la media muestral

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

# ORGANIZACIÓN DE DATOS

---

## OBJETIVOS

- Organizar los datos en una distribución de frecuencias
- Presentar una distribución de frecuencias en un histograma, un polígono de frecuencias y una ojiva.
- Elaborar e interpretar una representación de tallo y hoja.

---

# ORGANIZACIÓN DE DATOS

VARIABLE CUALITATIVA

- 
- Para representar gráficamente la distribución de frecuencias de una variable cualitativa se utilizan las ***barras*** y los ***sectores circulares***.

- Nota

Si trabajamos con variables nominales las categorías pueden ser colocadas en cualquier orden. En el caso de escala ordinal las categorías deberán ser colocadas en orden

---

## **EJEMPLO:**

En una planta embotelladora se registraron 54 accidentes y de acuerdo con la parte del cuerpo lesionada, dedos (D), ojos (O), brazo (B), y piernas (P); se registraron los datos (cuadro 1). Se pide organizar los datos.

---

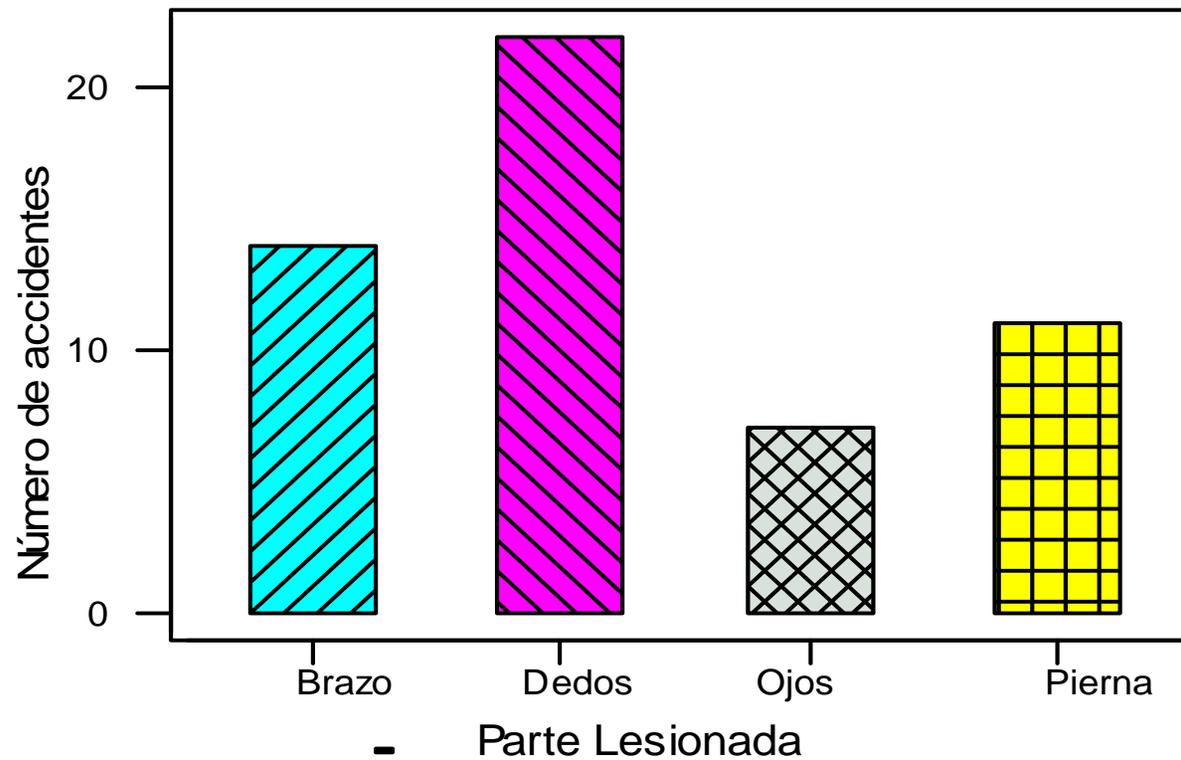
## CUADRO 1:

D	D	B	P	B	B
D	D	D	P	D	O
D	P	D	O	D	P
O	B	B	O	B	D
D	B	B	D	O	B
P	B	P	D	D	P
D	P	D	D	P	B
D	D	D	B	O	P
B	P	B	O	D	D

---

<b>PARTE LESIONADA</b>	<b>NÚMERO DE ACCIDENTES</b>
BRAZOS	14
DEDOS	22
OJO	7
PIERNA	11
Total	54

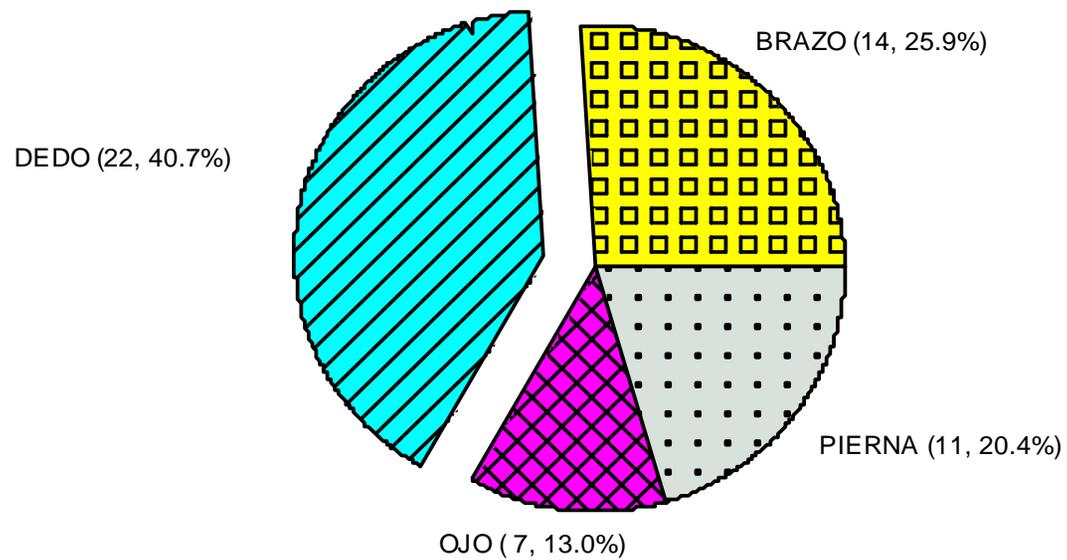
## Accidentes ocurridos en la planta según la parte lesionada



Fuente: Elaborado por el Departamento de Personal.  
Julio del 2005

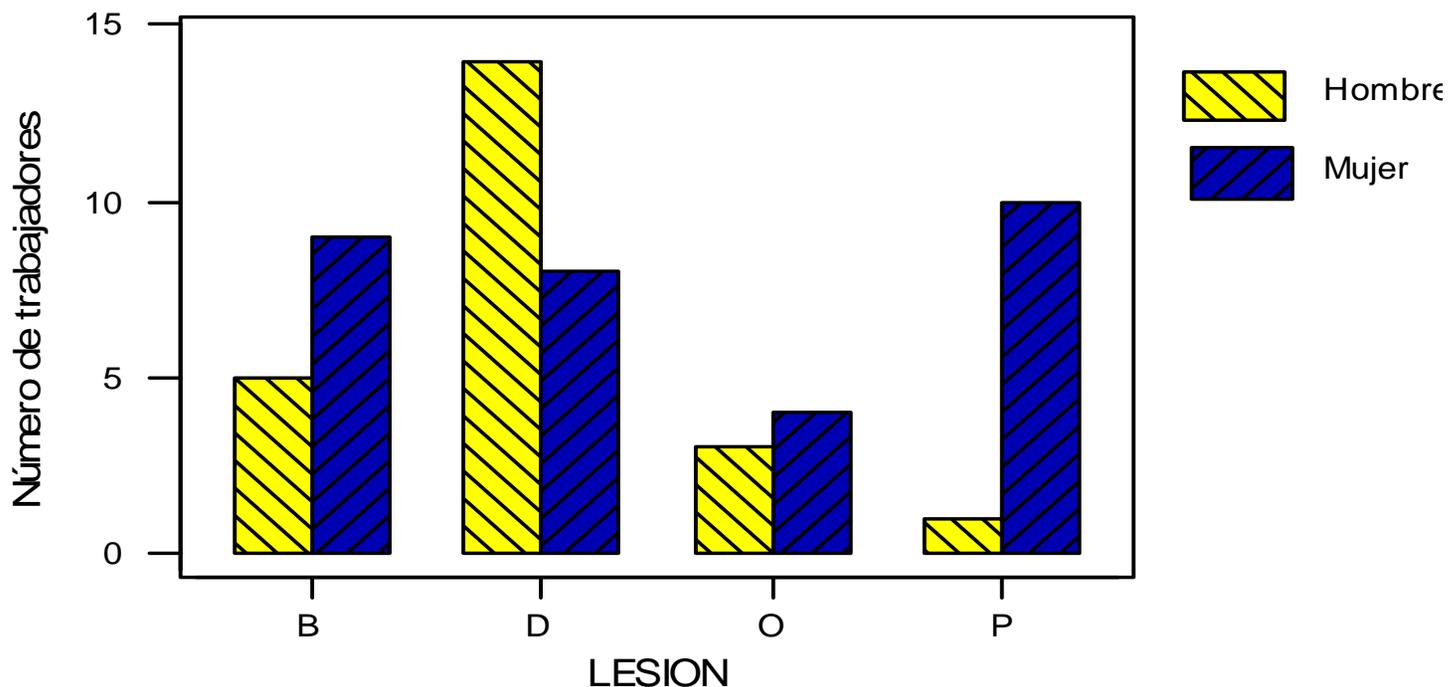
---

## ACCIDENTES OCURRIDOS EN LA PLANTA SEGÚN LA PARTE LESIONADA



Si además de registrar la parte lesionada se registra el sexo de los trabajadores se podría tener un gráfico de este tipo.

Registro de accidentes según el tipo de lesión y sexo



Fuente: Departamento Médico  
Julio del 2005

---

# ORGANIZACIÓN DE DATOS

VARIABLE CUANTITATIVA  
DISCRETA

## Variable Cuantitativa Discreta

(Pocos valores diferentes de la variable)

---

Variable	Frecuencia	Porcentaje
$X_1$	$f_1$	$(f_1/n) \times 100$
$X_2$	$f_2$	$(f_2/n) \times 100$
...	...	...
$X_n$	$f_n$	$(f_n/n) \times 100$
	$n$	1

---

- Ejemplo

Construir la distribución de frecuencias del número de trabajadores que se ausentaron en 20 días laborales:

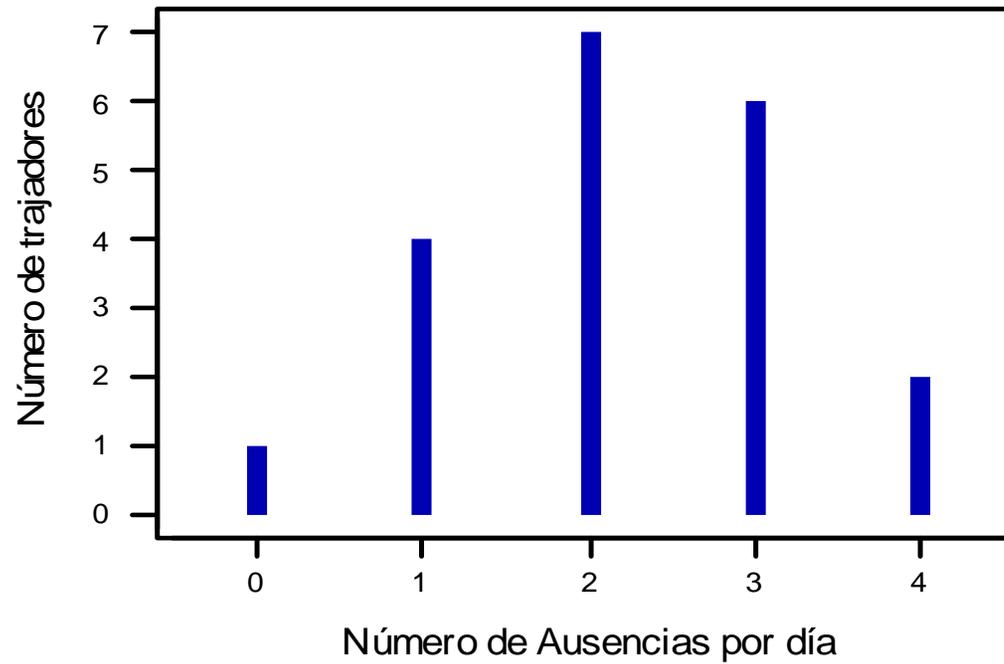
2	1	2	4	1
3	2	3	2	0
3	2	1	3	2
3	3	1	2	4

---

<b>Ausencias de Trabajadores</b>	<b>N° de días</b>	<b>Porcentajes</b>
0	1	$0.05 \times 100 = 5\%$
1	4	$0.20 \times 100 = 20\%$
2	7	$0.35 \times 100 = 35\%$
3	6	$0.30 \times 100 = 30\%$
4	2	$0.10 \times 100 = 10\%$
	20	100%

---

Distribución de Frecuencia del número de trabajadores que sea ausentaron por día.



# Tablas de frecuencias: Ejemplo

- ¿Cuántos individuos tienen menos de 2 hijos?
  - 674 individuos
- ¿Qué porcentaje de individuos tiene 6 hijos o menos?
  - 97,3%
- ¿Qué cantidad de hijos es tal que al menos el 50% de la población tiene una cantidad inferior o igual?
  - 2 hijos

Número de hijos

	Frec.	Porcent. (válido)	Porcent. acum.
0	419	27,8	27,8
1	255	16,9	44,7
2	375	24,9	69,5
3	215	14,2	83,8
4	127	8,4	92,2
5	54	3,6	95,8
6	24	1,6	97,3
7	23	1,5	98,9
Ocho+	17	1,1	100,0
Total	1509	100,0	

Diagrama de Venn: Se muestran tres círculos superpuestos que representan los grupos de hijos: 0 hijos (verde), 1 hijo (verde) y 2 hijos (rojo). El círculo rojo está completamente dentro del círculo verde, indicando que los individuos con 2 hijos o menos pertenecen al grupo de 6 hijos o menos.

Indicador de 50%: Una línea horizontal roja con una etiqueta "≥50%" apunta a la fila correspondiente a 2 hijos, indicando que el 50% de la población tiene 2 hijos o menos.

---

# MEDIDAS DE VARIABILIDAD

- Rango
- Rango intercuartil
- Variancia
- Desviación estándar
- Coeficiente de variabilidad

---

# ORGANIZACIÓN DE DATOS

## **VARIABLE CUANTITATIVA CONTINUA**

---

En la oficina de un diario, el tiempo que se tardan en imprimir la primera plana fue registrado durante 50 días. A continuación se transcriben los datos, aproximados a décimas de minuto:

20,8	22,8	21,9	22,0	20,7	20,9	25,0	22,2	22,8	20,1
25,3	20,7	22,5	21,2	23,8	23,3	20,9	22,9	23,5	19,5
23,7	20,3	23,6	19,0	25,1	25,0	19,5	24,1	24,2	21,8
21,3	21,5	23,1	19,9	24,2	24,1	19,8	23,9	22,8	23,9
19,7	24,2	23,8	20,7	23,8	24,3	21,1	20,9	21,6	22,7

- 
- a. Construya con los datos una tabla de distribución de frecuencia, usando la regla de sturges.
  - b. Construya un polígono de frecuencias.
  - c. Construya una ojiva.
  - d. Por medio de la ojiva estime que porcentaje de las veces la primera plana del periódico puede imprimirse en menos de 24 minutos.

---

## PROCEDIMIENTO

- Calcular el rango (R) o recorrido, el cual se define de la siguiente manera:

$$R = \text{Obs. máxima} - \text{Obs. M\u00ednima}$$

$$R = 25.3 - 19.0 = 6.3$$

- Determine el n\u00famero de intervalos (k)

$$K = 1 + 3.3 \log (n) \text{ (F\u00f3rmula de Sturges)}$$

$$K = 1 + 3.3 \log (50) = 6.607 = 7 \text{ (entero)}$$

(Redondeo simple)

Determinar del tamaño del intervalo TIC  
(Redondeado por exceso)

Se redondea según el número de  
decimales que se tiene en la muestra

$$TIC = \frac{R}{K}$$
$$TIC = \frac{6.3}{7} = 0.9$$

- Elabore la tabla de frecuencia según la información anterior donde:

$f_i$  = Frecuencia absoluta

$F_i$  = Frecuencia acumulada absoluta

$fr_i$  = Frecuencia relativa =  $f_i/n$

$Fr_i$  = Frecuencia acumulada relativa

$$\sum_{i=1}^k f_i = n$$

$$\sum_{i=1}^k fr_i = 1$$

$$F_i = \sum_{j=1}^i f_j$$

$$Fr_i = \sum_{j=1}^i fr_j = F_i / n$$

Marca de clase

$$X'_i = \frac{LI_i + LS_i}{2}$$

## Distribución de frecuencias del tiempo que tardan en imprimir la primera plana de un diario

---

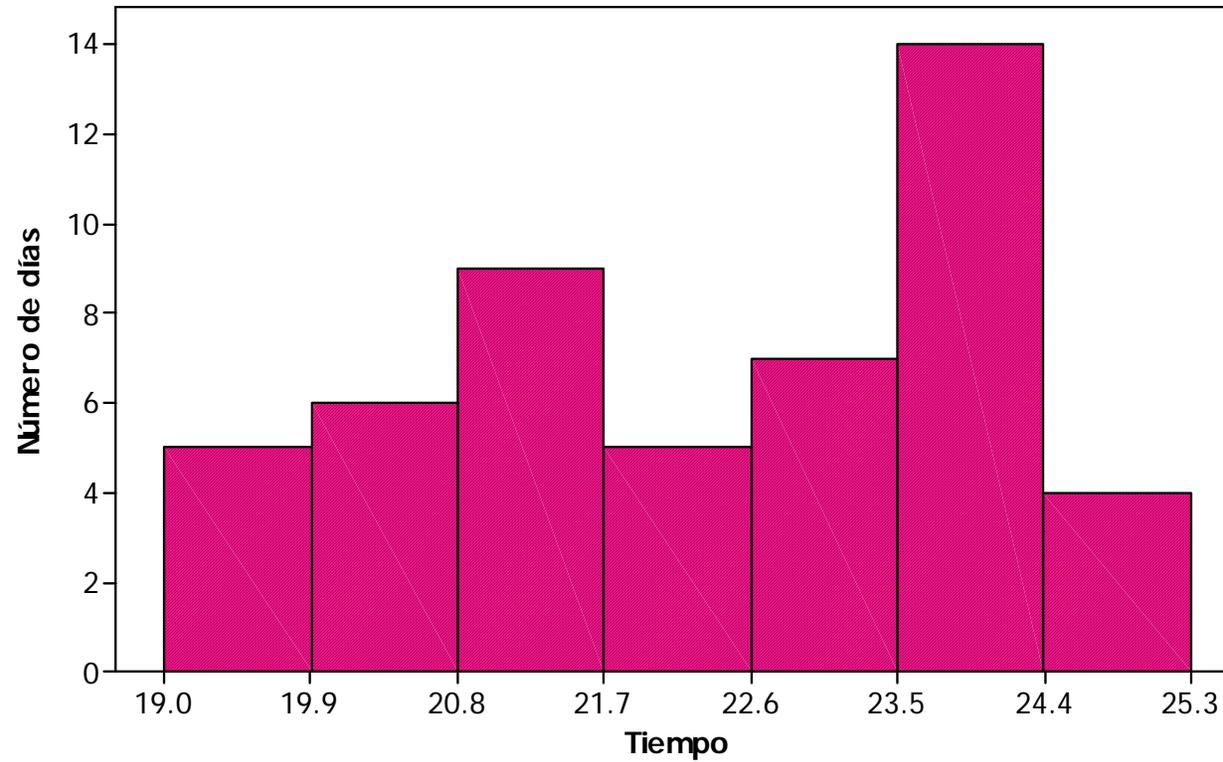
<b>Tiempo (minutos)</b>	<b>fi</b>	<b>fri</b>	<b>Fi</b>	<b>Fri</b>	<b><math>X'_i</math></b>
[19.0-19.9>	5	0.10	5	0.10	19.45
[19.9-20.8>	6	0.12	11	0.22	20.35
[20.8-21.7>	9	0.18	20	0.40	21.25
[21.7-22.6>	5	0.10	25	0.50	22.15
[22.6-23.5>	7	0.14	32	0.64	23.05
[23.5-24.4>	14	0.28	46	0.92	23.95
[24.4-25.3]	4	0.08	50	1.00	24.85
	50	1			

- 
- Gráficos para datos cuantitativos Agrupados en intervalos de clase
    - HISTOGRAMA DE FRECUENCIA
    - POLIGONO DE FRECUENCIA
    - OJIVA
  - Gráfico de Tallos y hojas
  - Gráfico para datos cuantitativos seriados en el tiempo

# HISTOGRAMA DE FRECUENCIA

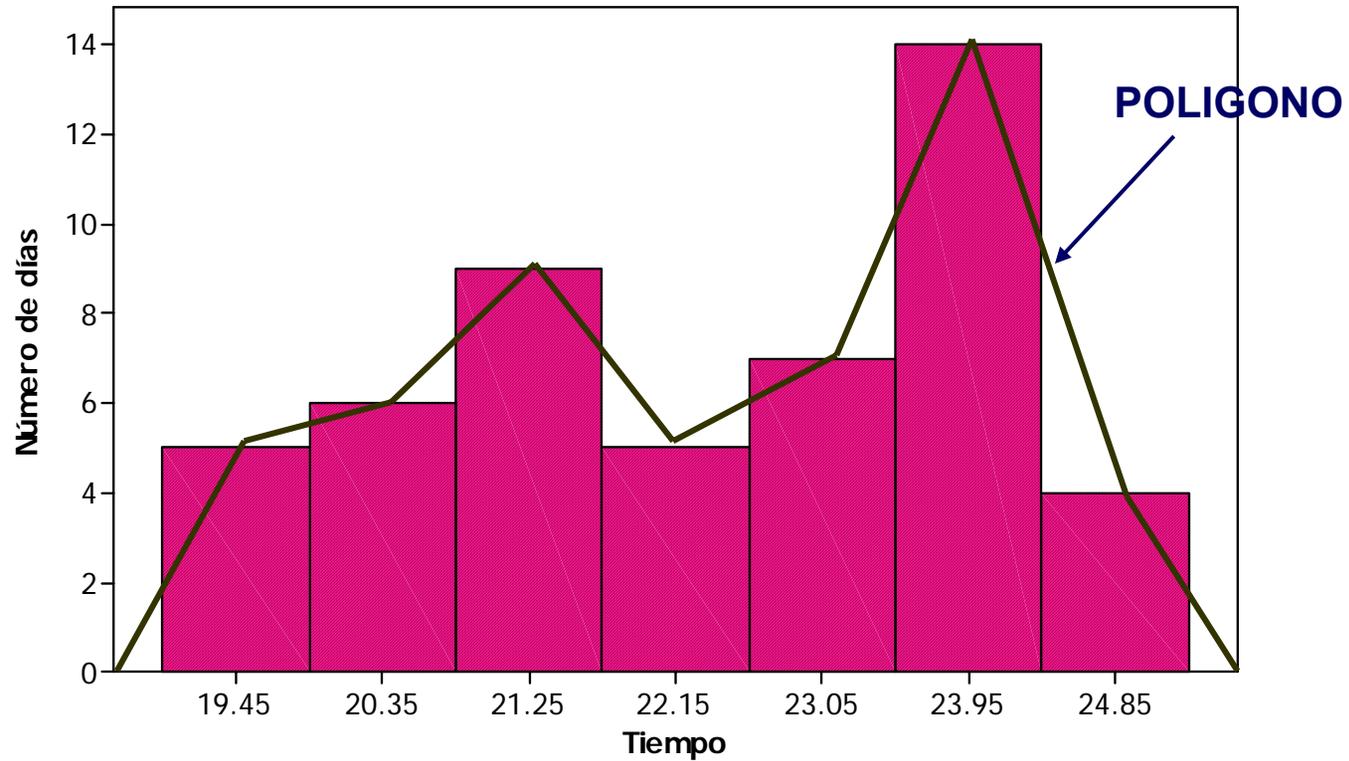
---

Distribución del tiempo que se tardan en imprimir la primera plana del diario



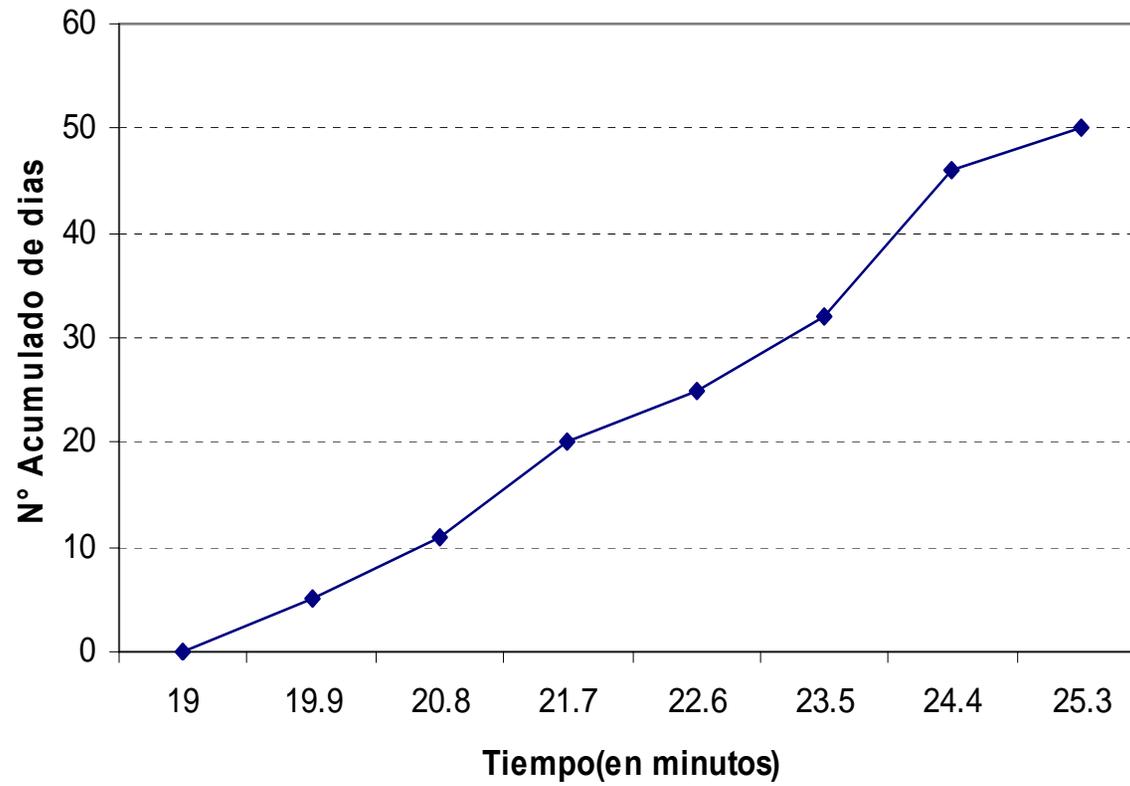
# POLIGONO DE FRECUENCIAS

Distribución del tiempo que se tardan en imprimir la primera plana del diario



# OJIVA

OJIVA



## EJERCICIO

---

Para evaluar la viabilidad de un proyecto de reforestación de una zona sometidas a estrés turístico, para el que se ha solicitado una subvención pública, se analizó la composición en mg/cm<sup>3</sup> de desechos orgánicos del territorio. Los datos obtenidos fueron:

Composición (mg/cm <sup>3</sup> )				
8.2	12.9	15.3	18.8	20.8
9.2	12.9	15.8	19.7	21
9.4	14	15.9	20.3	21.4
10.8	14.3	16.9	20.3	22.2
10.8	14.4	18.2	20.3	22.5
11.1	15.1	18.4	20.4	23.2
11.9	15.1	18.6	20.7	25.8

- 
- A. Construir la tabla de frecuencias y representar la información gráficamente.**
  - B. A partir de la tabla de frecuencias interprete:  $f_3$  y  $1-Fr_2$**
  - C. Entre qué niveles de composición se encuentra aproximadamente el 51,4% de las observaciones con menores niveles registrados?.**

---

Composición	Número de observaciones	Fi	fri	Fri
[8.2 - 11.2>	6	6	0.17142857	0.17142857
[11.2 - 14.2>	4	10	0.11428571	0.28571429
[14.2 - 17.2>	8	18	0.22857143	0.51428571
[17.2 - 20.2>	5	23	0.14285714	0.65714286
[20.2 - 23.2>	10	33	0.28571429	0.94285714
[23.6 - 26.2>	2	35	0.05714286	1

## EJERCICIO

---

Para los artículos anteriores se tienen las siguientes distribuciones de los precios (en dólares) en la primera semana de agosto y en la primera semana de setiembre del 2005.

CANTIDAD DE ARTICULOS		
PUNTOS MEDIOS	AGOSTO	SETIEMBRE
61	0	3
64	1	4
67	2	6
70	4	4
73	6	2
76	4	1
79	3	0

- 
- a) Grafique las dos distribuciones anteriores en un sólo gráfico donde se aprecien el comportamiento de las mismas y calcule el promedio y la desviación estándar de la distribución de agosto
  - b) Realice un comentario según lo observado en el gráfico anterior

# DIAGRAMA DE TALLOS Y HOJAS

Los datos tienen un decimal

tallo

**Stem-and-Leaf Display: Tiempo**  
**Stem-and-leaf of Tiempo N = 50**  
**Leaf Unit = 0.10**

hojas

1	19	0	Terminación de dígitos del 0 al 4
6	19	55789	Terminación de dígitos del 5 al 9
8	20	13	
15	20	7778999	
18	21	123	
22	21	5689	
24	22	02	
(6)	22	578889	
20	23	13	
18	23	56788899	
10	24	112223	
4	24		
4	25	0013	

15 datos menores e iguales que 20.9

6 datos entre 22.5 y 22.9 y uno de ellos es la mediana

10 datos mayores e iguales que 24.1