

# Unidad 1

**Obtención, Medición  
y Representación de Datos**

**Estadística E.S.O.**

# Objetivos

---

- ◇ Distinguir, localizar y manejar las **fuentes de información** estadística más usuales que proporcionan información útil.
- ◇ Definir, diferenciar y comprender el significado e interés de los **conceptos básicos** de la Estadística Descriptiva.
- ◇ Clasificar la **tipología de los datos** atendiendo a la escala de medida utilizada en su producción u observación.



# Objetivos

---

- ◇ *Ordenar y agrupar los datos correspondientes a una variable estadística en forma de una tabla y obtener su distribución de frecuencias.*
- ◇ *Realizar representaciones gráficas correctas de los datos acordes a su tipología.*



# Índice

---

## **1.- Definición y fuentes de información estadística**

## **2.- Conceptos básicos**

Dato, Población, Variable Estadística

## **3.- Clasificación de variables estadísticas**

## **4.- Tabulación de datos.**

Distribuciones de frecuencia

## **5.- Representaciones gráficas**

Diagramas de barras, diagramas de sectores, pictogramas, histogramas, polígonos de frecuencias, pirámides de población.

---



# 1.1 Definición

---

- ▶ La **Estadística** se ocupa de los **métodos** y **procedimientos** para recoger, clasificar, resumir, hallar regularidades y analizar los **datos**; así como de realizar **inferencias** a partir de ellos, con la finalidad de ayudar a la **toma de decisiones** y en su caso formular **predicciones**.
- ✓ Estadística **descriptiva**: describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.
- ✓ Estadística **inferencial**: efectúa estimaciones, toma de decisiones, formula predicciones apoyándose en cálculo de probabilidades y a partir de datos muestrales.



# 1.1 Fuentes de Información

---

◇ *En un estudio o investigación estadística son básicos los datos. Atendiendo a su procedencia o generación, las llamadas fuentes de información se clasifican en:*

▶ *Fuentes de Información **Primaria***

*Información que reúne o genera el investigador para alcanzar los objetivos del proyecto, análisis o estudio de un fenómeno*

▶ *Fuentes de Información **Secundaria***

*Recabada por empresas o agentes ajenos a la propia investigación*

*Suele ser información elaborada en otras investigaciones o recopilada y difundida por organismos oficiales*

---



# 1.1 Fuentes de Información

---

## Algunas Instituciones y Organismos

- ✓ *Instituto Nacional de Estadística (I.N.E.): [www.ine.es](http://www.ine.es)*
- ✓ *Centro de Investigaciones Sociológicas (C.I.S.): [www.cis.es](http://www.cis.es)*
- ✓ *Oficina Estadística de la Comunidad Europea(EUROSTAT): <http://europa.eu.int/comm/eurostat/>*

## Tipología de la Información

- ✓ *Información CENSAL*
  - ✓ *Información basada en ENCUESTAS*
  - ✓ *ESTUDIOS A MEDIDA*
- 



## 1.2 Conceptos básicos

---

- **Población**: Conjunto sobre el que se centra el objetivo de un análisis o investigación estadística. Esta compuesta por **unidades** a las que se denomina **elementos** (individuos, items..).
- **Dato**: Manifestación de un fenómeno de forma cualitativa o numérica. Cuantificación del mundo real en una representación cualitativa o numérica, comprensible para el individuo, y después representable para el ordenador.





## 1.2 Conceptos básicos

---

- **Muestra:** Subconjunto **representativo** de la población al que tenemos acceso y sobre el que realmente hacemos observaciones (mediciones).
- **Censo:** Observación de la/s característica/s en todos los elementos de una población.
- **Encuesta:** Método de medición de una o más variables en una muestra.



## 1.2 Conceptos básicos

---

- **Variable:** Característica observable que **varía entre los diferentes individuos** de una población. La información que disponemos de cada individuo es resumida en variables.

Ejemplos:

- ✓ Sexo (hombre / mujer)
- ✓ Salario bruto anual
- ✓ Tipo de contrato (duración indefinida / duración determinada)
- ✓ Edad (en años cumplidos)
- ✓ Edad (en grupos de edad)
- ✓ Nacionalidad
- ✓ Altura / Peso
- ✓ Régimen de tenencia de viviendas en España (Por compra, herencia etc)

Nota: Cada uno de los posibles valores de una variable se denomina **modalidad**.

---



# 1.3 Clasificación de variables

---

## Tipos de variables

### Cualitativas

Describen cualidades de un objeto o individuo (no se pueden hacer operaciones algebraicas con ellas)

**Nominales:** Si sus valores no se pueden ordenar

- Sexo, Grupo Sanguíneo, Religión, Nacionalidad, Fumar (Sí/No)

**Ordinales (semicuantitativas):** Si sus valores se pueden ordenar

- Mejoría a un tratamiento, Grado de satisfacción, Intensidad del dolor

### Cuantitativas o Numéricas

Si sus valores son numéricos (tiene sentido hacer operaciones algebraicas con ellos)

**Discretas:** Si toma valores enteros

- Número de hijos, Número de libros leídos, Num. de “cumpleaños”

**Continuas:** Si entre dos valores, son posibles infinitos valores intermedios.

- Altura, Salario bruto anual, Gasto farmacéutico, edad (años, meses, horas...)
- 



## 1.3 Clasificación de variables

---

### Tipos de variables

**Ejercicio:** Clasifica las siguientes variables..

“Paga semanal” (medida en €)

Nota de un examen (suspenso, aprobado, notable, sobresaliente, matrícula)

Número de grupos de E.S.O.

Naturaleza de una finca (rústica / urbana )



## 1.3 Clasificación de variables

---

### Tipos de variables

**Ejercicio:** Clasifica las siguientes variables..

“Paga semanal” (medida en €) → **Cuantitativa Continua**

Nota de un examen → **Cualitativa Ordinal**

Número de grupos de E.S.O. → **Cuantitativa discreta**

Naturaleza de una finca (rústica / urbana ) → **Cualitativa nominal**



## 1.4 Tabulación de datos

---

### Tabulación

Proceso de ordenación y agrupamiento de un conjunto de datos

- ✓ Tamaño de la población:  $N$
- ✓ Valores de la variable o modalidades:  $x_i$
- ✓ Número de modalidades:  $k$
- ✓ Recorrido o campo de variación:  $\{x_1, \dots, x_k\}$



## 1.4 Tabulación de datos

---

Definimos:

$n_i$  frecuencia **absoluta**: Número de veces que se observa o repite el dato  $x_i$

$N_i$  frecuencia **absoluta acumulada**: Número de datos menores o iguales que  $x_i$

$f_i$  frecuencia **relativa**: Porcentaje de observaciones sobre el total  
 $f_i = n_i / N$

$F_i$  frecuencia **relativa acumulada**: Porcentaje por debajo o igual a  $x_i$   
 $F_i = N_i / N$

### Distribución de Frecuencias

Conjunto de valores que toma la variable y sus frecuencias asociadas

---



## 1.4 Tabulación de datos

---

### Distribución de Frecuencias

$X$	$n_i$	$f_i$	$N_i$	$F_i$
$x_1$	$n_1$	$f_1 = n_1/N$	$N_1 = n_1$	$F_1 = f_1$ ó $F_1 = N_1/N$
$x_2$	$n_2$	$f_2 = n_2/N$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$ ó $F_2 = N_2/N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k$	$n_k$	$f_k = n_k/N$	$N_k = n_1 + \dots + n_k = N$	$F_k = f_1 + \dots + f_k = 1$ ó $F_k = N_k/N = 1$
$\Sigma$	$N$	<b>1</b>		





## 1.4 Tabulación de datos

---

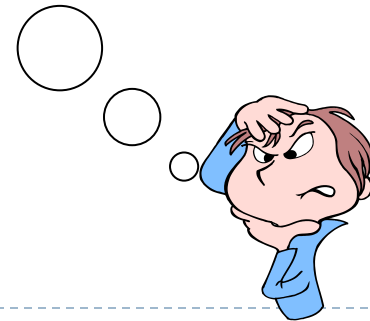
### Ejemplo:

En una encuesta realizada a 1.509 familias se les pregunta por el **número de hijos**

¿Cuántas familias tienen menos de 2 hijos?

¿Qué porcentaje de familias tiene 6 hijos o menos?

¿Qué cantidad de hijos es tal que al menos el 50% de las familias tiene una cantidad inferior o igual?



# 1.4 Tabulación de datos

$X$ Número de hijos	$n_i$ Frecuencias Absolutas	$f_i = n_i / N$ Frecuencias Relativas	$N_i$	$F_i$
0	419	419 / 1.509 = 27,8 %	419	0,278
1	255	16,9 %	419 + 255 = 674	0,278 + 0,169 = 44,7 %
2	375	24,9 %	1.049	69,5
3	215	14,2 %	1.264	83,8
4	127	8,4 %	1.391	92,2
5	54	3,6 %	1.445	95,8
6	24	1,6 %	1.469	97,3
7	23	1,5 %	1.492	98,9
8 ó más	17	1,1 %	<b>1.509</b>	100 %
$\Sigma$	<b>1.509</b>	100 %		

¿Cuántas familias tienen menos de 2 hijos?

¿Qué porcentaje de familias tiene 6 hijos o menos?

¿Qué cantidad de hijos es tal que al menos el 50% de las familias tiene una cantidad inferior o igual?



# 1.4 Tabulación de datos

$X$ Número de hijos	$n_i$ Frecuencias Absolutas	$f_i = n_i / N$ Frecuencias Relativas	$N_i$	$F_i$
0	419	419 / 1.509 = 27,8 %	419	0,278
1	255	16,9 %	419 + 255 = 674	0,278 + 0,169 = 44,7 %
2	375	24,9 %	1.049	69,5
3	215	14,2 %	1.264	83,8
4	127	8,4 %	1.391	92,2
5	54	3,6 %	1.445	95,8
6	24	1,6 %	1.469	97,3
7	23	1,5 %	1.492	98,9
8 ó más	17	1,1 %	1.509	100 %
$\Sigma$	1.509	100 %		

¿Cuántas familias tienen menos de 2 hijos?

frecuencia familias con 0 hijos  
+  
frecuencia familias con 1 hijo

**674 familias**

**Es decir, el 44,7 % de las familias encuestadas**



# 1.4 Tabulación de datos

$X$ Número de hijos	$n_i$ Frecuencias Absolutas	$f_i = n_i / N$ Frecuencias Relativas	$N_i$	$F_i$
0	419	419 / 1.509 = 27,8 %	419	0,278
1	255	16,9 %	419 + 255 = 674	0,278 + 0,169 = 44,7 %
2	375	24,9 %	1.049	69,5
3	215	14,2 %	1.264	83,8
4	127	8,4 %	1.391	92,2
5	54	3,6 %	1.445	95,8
6	24	1,6 %	1.469	97,3
7	23	1,5 %	1.492	98,9
8 ó más	17	1,1 %	1.509	100 %
$\Sigma$	1.509	100 %		

¿Qué porcentaje de familias tiene 6 hijos o menos?

**97,3 %**



## 1.4 Tabulación de datos

$X$ Número de hijos	$n_i$ Frecuencias Absolutas	$f_i = n_i/N$ Frecuencias Relativas	$N_i$	$F_i$
0	419	419 / 1.509 = 27,8 %	419	0,278
1	255	16,9 %	419 + 255 = 674	0,278 + 0,169 = 44,7 %
2	375	24,9 %	1.049	69,5
3	215	14,2 %	1.264	83,8
4	127	8,4 %	1.391	92,2
5	54	3,6 %	1.445	95,8
6	24	1,6 %	1.469	97,3
7	23	1,5 %	1.492	98,9
8 ó más	17	1,1 %	<b>1.509</b>	100 %
$\Sigma$	<b>1.509</b>	100 %		

¿Qué cantidad de hijos es tal que al menos el 50% de las familias tiene una cantidad inferior o igual?

**2 hijos**



## 1.4 Tabulación de datos

---

Distinguiremos **distribución de frecuencias** para

- ✓ Variables cualitativas
- ✓ Variables cuantitativas discretas o no agrupadas
- ✓ Variables cuantitativas continuas o agrupadas



## 1.4 Tabulación de datos

### Distribución de frecuencias para **Variables Cualitativas**

$X$	$n_i$	$f_i$
$x_1$	$n_1$	$f_1 = n_1/N$
$x_2$	$n_2$	$f_2 = n_2/N$
$\vdots$	$\vdots$	$\vdots$
$x_k$	$n_k$	$f_k = n_k/N$
$\Sigma$	$N$	$1$

Ejemplo: Estadística de Nulidades, Separaciones y Divorcios  
Año 2008. Fuente INE

$X$ Tipo de ruptura	$n_i$ Matrimonios	$f_i = n_i/N$ Frecuencias Relativas	<del><math>N_i</math></del>	<del><math>F_i</math></del>
Divorcios	110.036	110.036 / 118.939 = 92,51 %		
Separaciones	8.761	7,37 %		
Nulidades	142	0,12 %		
$\Sigma$	<b>118.939</b>	100 %		

Carece de sentido obtener **frecuencias acumuladas**

## 1.4 Tabulación de datos

### Distribución de frecuencias para **Variables Cuantitativas Discretas** o no agrupadas

Partiendo de la hipótesis  $x_1 < x_2 < \dots < x_k$  se cuenta el número de veces que se repite cada dato ( $n_i$ ) de la que se deducen la frecuencias relativas, y las acumuladas.

$X$	$n_i$	$f_i$	$N_i$	$F_i$
$x_1$	$n_1$	$f_1 = n_1/N$	$N_1 = n_1$	$F_1 = f_1$ ó $F_1 = N_1/N$
$x_2$	$n_2$	$f_2 = n_2/N$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$ ó $F_2 = N_2/N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k$	$n_k$	$f_k = n_k/N$	$N_k = n_1 + \dots + n_k = N$	$F_k = f_1 + \dots + f_k = 1$ ó $F_k = N_k/N = 1$
$\Sigma$	$N$	<b>1</b>		



## 1.4 Tabulación de datos

---

### Distribución de frecuencias para **Variables Cuantitativas Discretas** o no agrupadas

Ejemplo: Distribución del Número de bibliotecas por número de puntos de servicio que dispone. Fuente Estadística de Bibliotecas 2006. INE.

$X$ Número puntos de servicio	$n_i$ Frecuencias Absolutas (Bibliotecas)	$f_i = n_i / N$ Frecuencias Relativas	$N_i$ Frecuencias Absolutas Acumuladas	$F_i$ Frecuencias Relativas Acumuladas
1	6.072	6.072 / 6.443 = 94,24 %	6.072	94,24 %
2	226	3,51 %	6.298	97,75 %
3	82	1,27 %	6.380	98,02 %
4	44	0,68 %	6.424	98,71 %
5	19	0,29 %	6.443	100 %
$\Sigma$	<b>6.443</b>	100 %		



## 1.4 Tabulación de datos

### Distribución de frecuencias para **Variables Continuas o agrupadas en intervalos**

Si el número de modalidades o valores de la variable es muy elevado, éstos se suelen agrupar en intervalos y se recoge la frecuencia de datos en cada intervalo

$(L_{i-1}, L_i]$ Intervalo	$X_i$ Marca de Clase	$n_i$	$f_i$	$N_i$	$F_i$
$(L_0, L_1]$	$x_1$	$n_1$	$f_1 = n_1/N$	$N_1 = n_1$	$F_1 = f_1$ ó $F_1 = N_1/N$
$(L_1, L_2]$	$x_2$	$n_2$	$f_2 = n_2/N$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$ ó $F_2 = N_2/N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(L_{k-1}, L_k]$	$x_k$	$n_k$	$f_k = n_k/N$	$N_k = n_1 + \dots + n_k = N$	$F_k = f_1 + \dots + f_k = 1$ ó $F_k = N_k/N = 1$
	$\Sigma$	N	1		

**Marca de clase**  
Valor medio de dicho intervalo  
 $(L_i + L_{i-1}) / 2$

**Amplitud del intervalo**  
 $a_i = L_i - L_{i-1}$

$x_1 < x_2 < \dots < x_k$   
se cuenta el número de datos comprendidos en cada intervalo

# 1.4 Tabulación de datos

## Distribución de frecuencias para Variables Continuas o agrupadas en intervalos

### Población por edad (grupos quinquenales)

Unidades: Personas

	Ambos sexos
0-4	2.397.418
05-09	2.250.648
10-14	2.134.301
15-19	2.335.412
20-24	2.780.855
25-29	3.593.846
30-34	4.181.737
35-39	3.980.413
40-44	3.794.228
45-49	3.430.898
50-54	2.973.498
55-59	2.587.260
60-64	2.440.606
65-69	1.954.495
70-74	1.874.577
75-79	1.737.803
80-84	1.224.908
85 y más	989.047

### Población en Edad Pediátrica

$(L_{i-1}; L_i]$ Intervalo	$X_i$ Marca de Clase	$n_i$	$f_i$	$N_i$	$F_i$
0 - 4	2	2.397.418	0,35 (35 %)	2.397.418	0,35 (35 %)
5 - 9	7	2.250.648	0,33 (33 %)	4.648.066	0,69 (69 %)
10 - 14	12	2.134.301	0,31 (31 %)	6.782.367	1
	$\Sigma$	6.782.367	1		

## 1.4 Tabulación de datos

### Distribución de frecuencias para **Variables Continuas o agrupadas en intervalos**

Estaturas de 50 niños. Fuente: Peña y Romo 1997.

$(L_{i-1}; L_i]$ Intervalo	$X_i = \frac{(L_{i-1} + L_i)}{2}$ Marca de Clase	$n_i$	$f_i$	$N_i$	$F_i$
[ 1,465 – 1,497 )	1,481	4	0,08 (8 %)	4	0,08 (8 %)
[ 1,497 - 1,529 )	1,513	4	0,08 (8 %)	8	0,16 (16 %)
[1,529 - 1,561 )	1,545	15	0,3 (30 %)	23	0,46 (46%)
[1,561 – 1,593 )	1,577	12	0,24 (24 %)	35	0,70 (70%)
[ 1,593 – 1,625 )	1,609	9	0,18 (18 %)	44	0,88 (88%)
[ 1,625 – 1,657 )	1,641	5	0,1 (10 %)	49	0,98 (98%)
[ 1,657 – 1,689 )	1,673	1	0,02 (2 %)	50	1
	$\Sigma$	50	1		

Sabemos que el niño más alto mide **1,689** y el niño más bajito mide **1,465**

Agrupamos nuestros datos en **7** intervalos con igual amplitud:

$$a_i = (\text{Max} - \text{Min}) / 7$$

$$\text{Ej: } (1,689 - 1,465) / 7 = 0,032$$

**Primer intervalo:**

$$1,465 + 0,032 = 1,497$$

**[1,465 , 1,497)**

# 1.5 Representaciones gráficas

---

## Tipos de representaciones Gráficas

Para Variables **Cualitativas** o Atributos

Cartogramas, Pictogramas

Diagramas de Sectores

Diagrama de Barras

Para Variables **Cuantitativas Discretas**

Diagrama de Barras

Para Variables **Cuantitativas Continuas o Agrupadas**

Histograma

Polígonos de Frecuencias

Otras representaciones gráficas

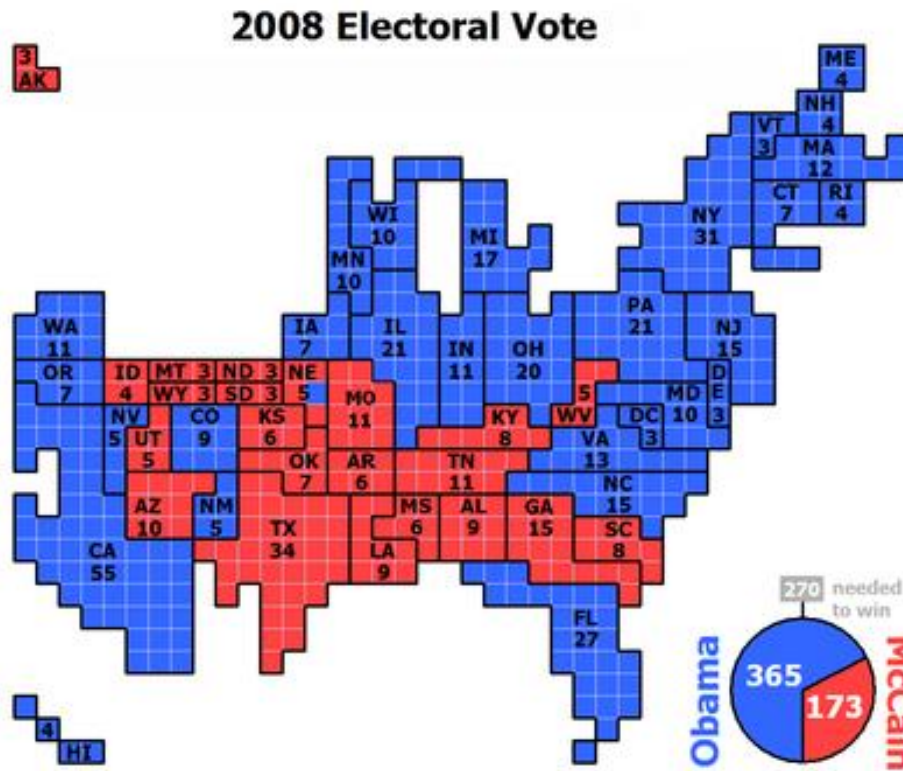
---



# 1.5 Representaciones gráficas

Tipos de representaciones Gráficas para variables cualitativas

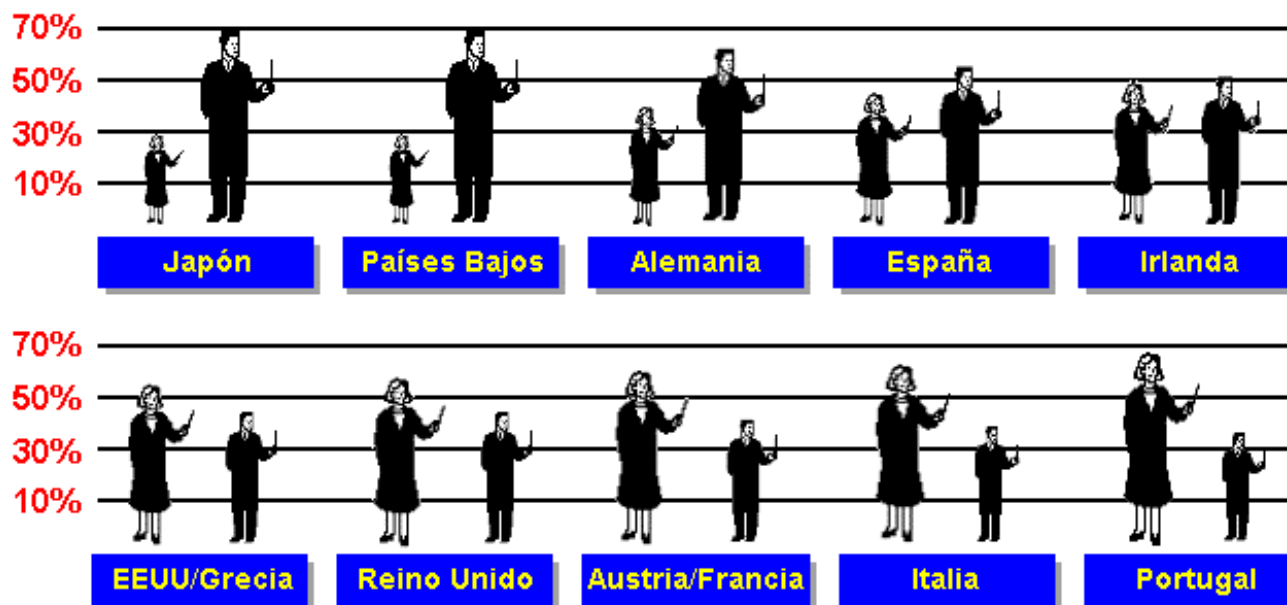
## Cartogramas



# 1.5 Representaciones gráficas

## Tipos de representaciones Gráficas para variables cualitativas

### Pictogramas



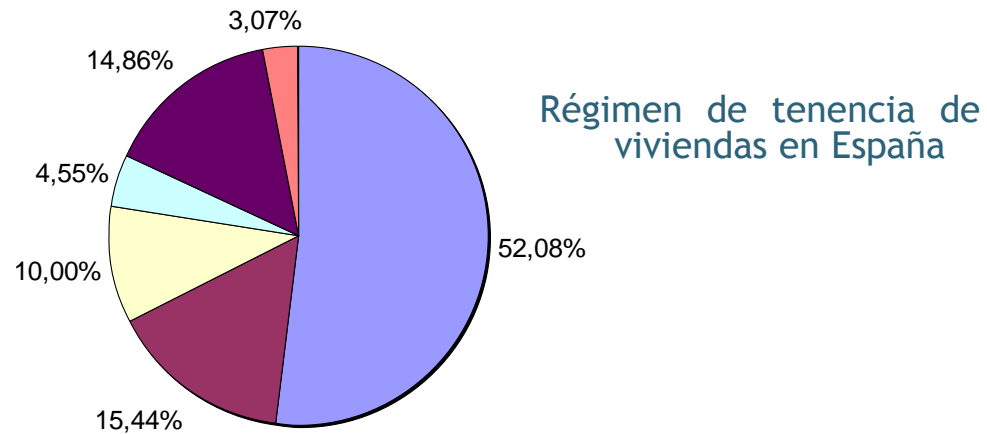
Porcentaje de profesoras y profesores en distintos países

# 1.5 Representaciones gráficas

---

## Tipos de representaciones Gráficas para variables cualitativas

### Diagramas de Sectores



- Por compra (totalmente pagada)
- Por compra (con pagos pendientes)
- Por herencia o donación
- Facilitadas (por otras personas, empresas ú organismos)
- En alquiler
- Otras formas



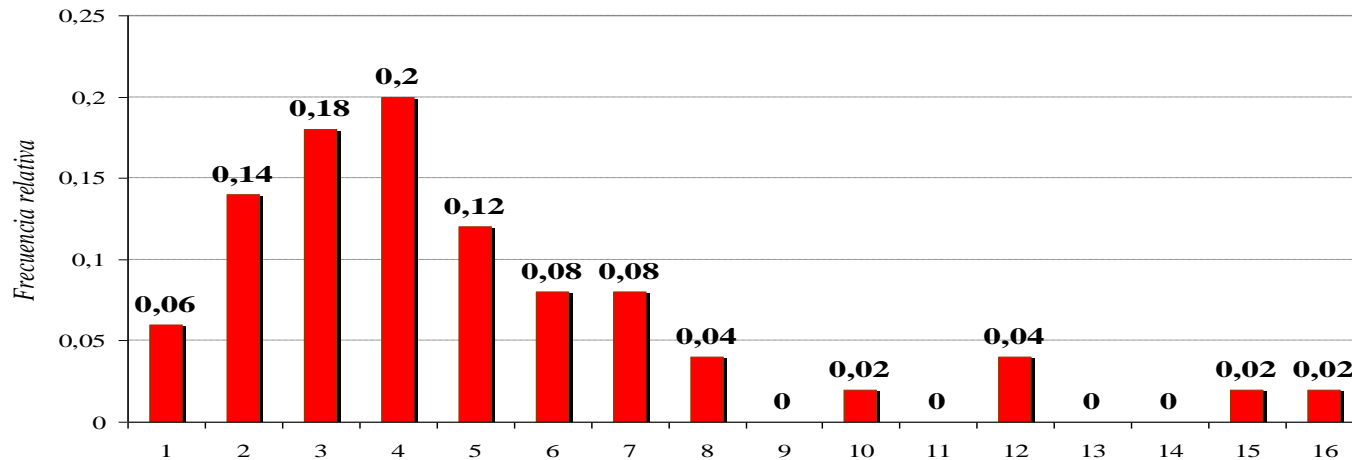


# 1.5 Representaciones gráficas

---

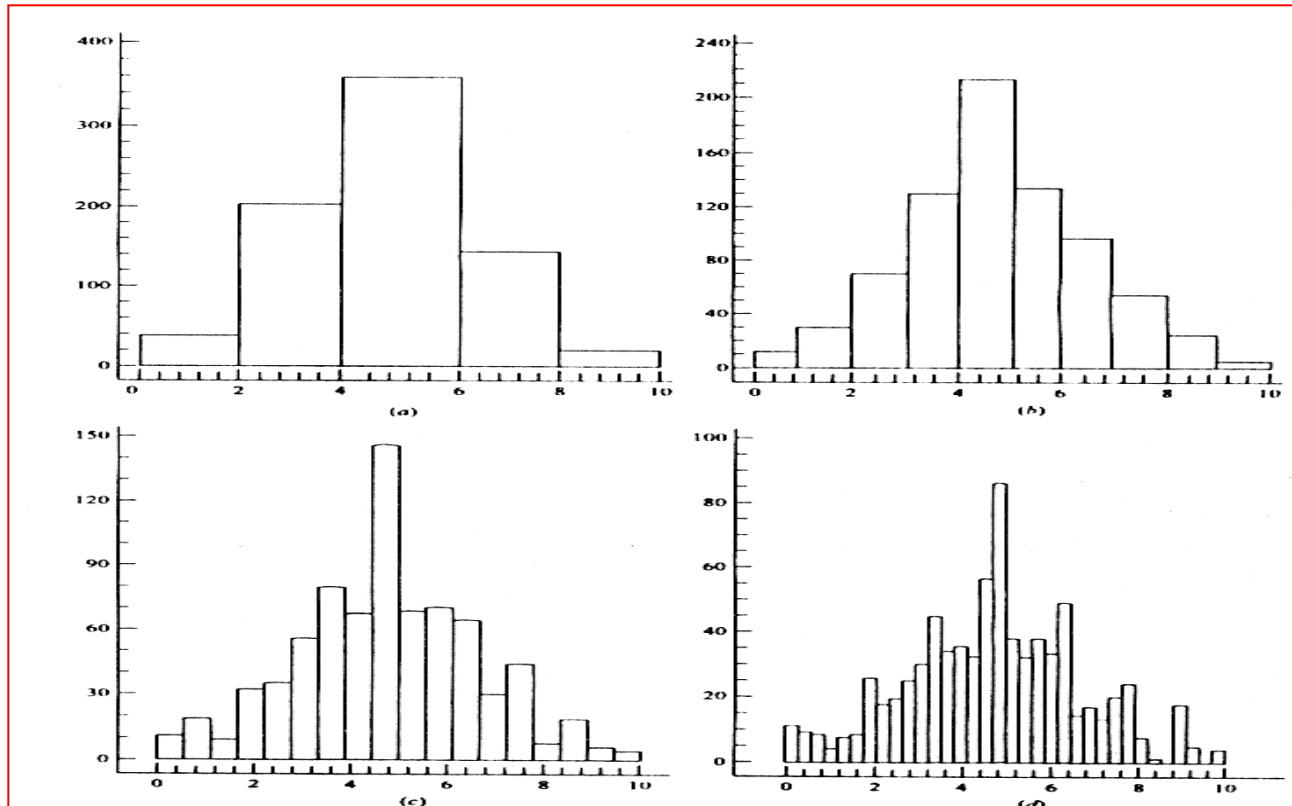
## Tipos de representaciones Gráficas

### Diagrama de Barras



# 1.5 Representaciones gráficas

## Tipos de representaciones Gráficas

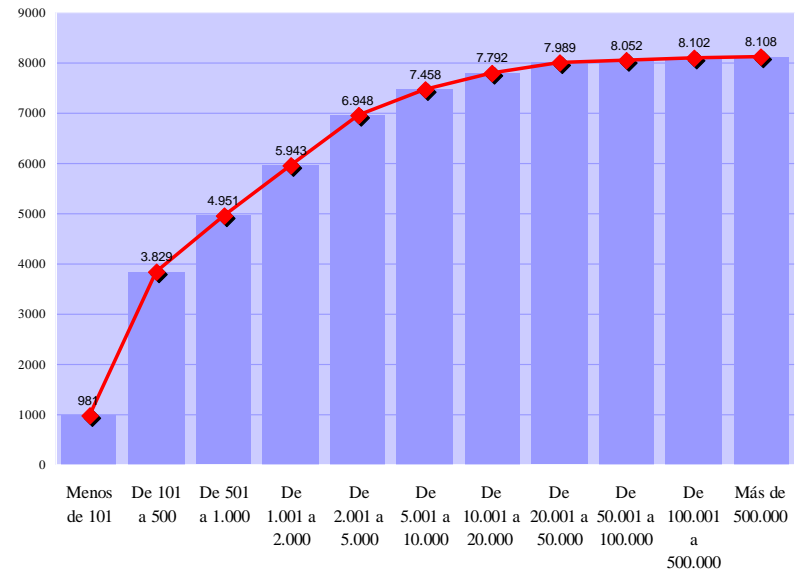
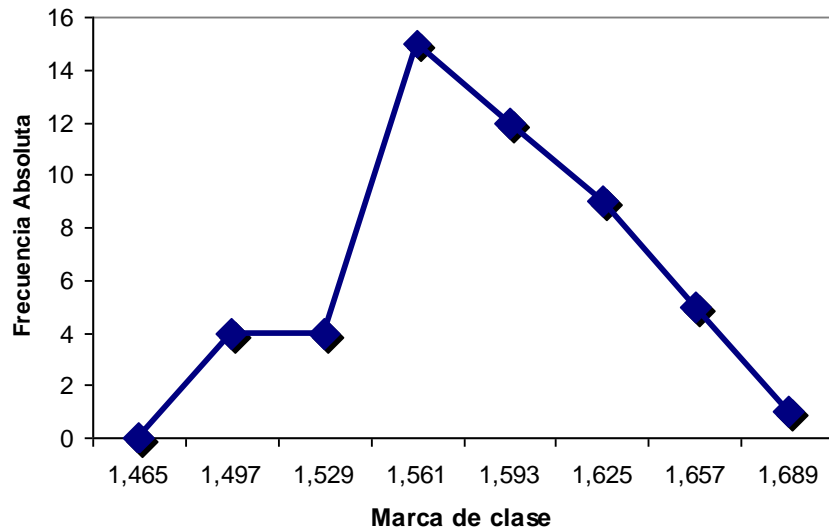


**Histogramas según tipo de intervalos**



# 1.5 Representaciones gráficas

## Tipos de representaciones Gráficas



Polígonos de frecuencias

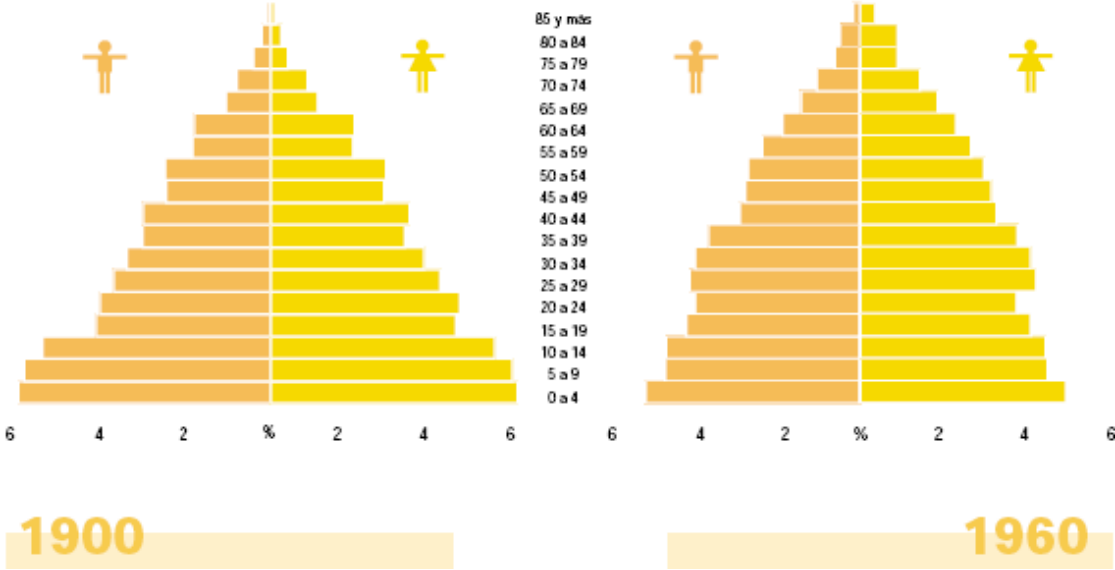
Polígonos de frecuencias acumuladas



# 1.5 Representaciones gráficas

## Otras representaciones Gráficas

Evolución de las pirámides de población



Pirámides de población



# ¿Qué es lo que hemos visto?

---

1.- Fuentes de información (primaria / secundaria)

2.- Conceptos básicos de Estadística

- ✓ Población
- ✓ Muestra
- ✓ Variable estadística

3.- Escalas de medidas

- ✓ Clasificación de variables estadísticas (Cualitativa / Cuantitativa)

4.- Distribución de frecuencias

- ✓ Frecuencias absolutas
- ✓ Frecuencias relativas
- ✓ Frecuencias acumuladas

5.- Representaciones gráficas



## Ejercicio

---

**8** Los alumnos que se presentaron al examen de Estadística I de primero de LADE en la primera convocatoria de Junio obtuvieron las calificaciones que se presentan en la tabla (redondeadas al entero más próximo):

7	3	2	4	5	1	8	6	1	5
3	2	4	9	8	1	0	2	4	1
2	5	6	5	4	7	1	3	0	5
8	6	3	4	0	10	2	5	7	4
0	2	1	5	6	4	3	5	2	3
9	7	3	4	3	5	7	4	6	5
6	1	0	5	7	8	5	2	3	10
4	6	2	1	1	2	6	7	4	5
4	7	3	6	5	0	2	8	2	7
8	5	2	7	1	4	6	3	5	6

Define tipo de variable

Tabula los datos

Establece intervalos si lo consideras necesario

---



# Ejercicio

**8** Los alumnos que se presentaron al examen de Estadística I de primero de LADE en la primera convocatoria de Junio obtuvieron las calificaciones que se presentan en la tabla (redondeadas al entero más próximo):

$X$ Nota	$n_i$ Frecuencias Absolutas Nº Alumnos	$f_i = \frac{n_i}{N}$ Frecuencias Relativas	$N_i$	$F_i$
0	6	6,00%	6	6,00%
1	10	10,00%	16	16,00%
2	13	13,00%	29	29,00%
3	11	11,00%	40	40,00%
4	13	13,00%	53	53,00%
5	16	16,00%	69	69,00%
6	11	11,00%	80	80,00%
7	10	10,00%	90	90,00%
8	6	6,00%	96	96,00%
9	2	2,00%	98	98,00%
10	2	2,00%	100	100,00%
$\Sigma$	100	100 %		

Variable Cuantitativa  
Discreta

El 53 % de los alumnos  
sacan menos de un 5 en el  
examen

## Ejercicio

**8** Los alumnos que se presentaron al examen de Estadística I de primero de LADE en la primera convocatoria de Junio obtuvieron las calificaciones que se presentan en la tabla (redondeadas al entero más próximo):

$X$ Nota	$n_i$ Alumnos	$f_i = n_i/N$ Frecuencias Relativas	$N_i$	$F_i$
Suspenso	53	53 %		<i>¿Tiene Sentido?</i>
Aprobado	27	27 %		
Notable	16	16 %		
Sobresaliente	4	4 %		
$\Sigma$	100	100 %		

**Variable Cualitativa Ordinal**

**El 53 % de los alumnos suspenden el examen**





## Ejercicio

**8** Los alumnos que se presentaron al examen de Estadística I de primero de LADE en la primera convocatoria de Junio obtuvieron las calificaciones que se presentan en la tabla (redondeadas al entero más próximo):

$(L_{i-1}; L_i]$ Intervalo	$X_i = \frac{(L_{i-1} + L_i)}{2}$ Marca de Clase	$a_i = (L_i - L_{i-1})$ Amplitud	$n_i$	$f_i$	$N_i$	$F_i$
0 - 2	1	3	29	29 %	29	29 %
3 - 4	3,5	2	24	24 %	53	53 %
5 - 6	5,5	2	27	27 %	80	80 %
7 - 8	7,5	2	16	16 %	96	96 %
9 - 10	9,5	2	4	4 %	100	100 %
	$\Sigma$		100	1		

**Variable Cuantitativa Agrupada en intervalos**

**El 29 % de los alumnos tienen una nota inferior al 3**