

## Estadística

2º curso del Grado en Ciencias de la  
Actividad Física y el Deporte

---o0o---

### Introducción. Descripción de las Muestras: Estadística Descriptiva



**Bioestadística - Facultad de Medicina**

**Universidad de Granada (España)**

<http://www.ugr.es/~bioest>



Estadística descriptiva - 2

Resúmenes: 1

## ESTADÍSTICA

### *Necesidad de la Estadística*

Las Ciencias de la Actividad Física y el Deporte son experimentales y se basan en el método inductivo (extensión, al todo, de las conclusiones obtenidas en una parte). El único modo de validar tales inducciones es por el Método Estadístico. Las demás razones que siguen son reflejo de esta mayor razón:

- a) **La variabilidad biológica de los individuos objeto de estudio en las Ciencias de la Actividad Física y el Deporte** origina que sus datos sean impredecibles y que el modo de controlarlos sea a través del Método Estadístico.
- b) **La naturaleza cada vez más cuantitativa de las Ciencias de la Actividad Física y el Deporte** requiere del Método Estadístico para analizar y poner orden en los datos.
- c) **La investigación en el campo de las Ciencias de la Actividad Física y el Deporte** requiere de la Estadística en sus etapas de diseño, recopilación de datos y análisis de los resultados.
- d) **El volumen de la información que recibe el profesional del deporte** requiere de conocimientos estadísticos que le permitan leer crítica y comprensivamente los resultados científicos ajenos.
- e) **La naturaleza del trabajo en la Actividad Física y el Deporte** es en esencia de tipo probabilístico o estadístico, disciplinas que dan rigor y objetividad a los clásicos procesos subjetivos de diagnóstico, pronóstico y tratamiento.
- f) **La perspectiva comunitaria de las Ciencias del Deporte** requiere del uso de la Estadística para poder extrapolar las conclusiones desde la parte estudiada de la población a su globalidad.

#### **Objetivo del Curso:**

- Proporcionar los mínimos conocimientos necesarios para realizar d) en los experimentos más simples.
- Comprender los resultados estadísticos de los artículos científicos.



# ESTADÍSTICA

*Hoy día la Estadística no es solo una Ciencia Básica sino que representa una tecnología clave en el momento científico actual. Las ideas estadísticas básicas deben formar una parte imprescindible del equipo mental de la persona formada, y el Método Estadístico un instrumento de trabajo esencial para el científico.*

Sir Ronald Fisher (adaptado)

**Definición de Estadística:** La **Estadística** es el conjunto de métodos necesarios para recoger, clasificar, representar y resumir datos, así como para hacer **inferencias** (extraer consecuencias) científicas a partir de ellos.

Distinguimos:

Estadística descriptiva

(recogida, clasificación, representación y resumen de datos)

Inferencia Estadística

Teoría de la Estimación

*¿Cuál es el tiempo medio que tarda en hacer efecto un fármaco?*

Teoría de los Contrastes de Hipótesis

*¿El tratamiento A es igual de efectivo que el tratamiento B?*



# ESTADÍSTICA



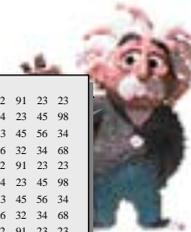
Nuestro interés es conocer qué es lo que ocurre en una **población** dada acerca de alguna cierta característica. Por ejemplo, *¿es efectivo el nuevo método de entrenamiento para un cierto deporte?*



Como es de suponer, no podemos observar a todos los individuos de esa población, de manera que trataremos de obtener una “**imagen representativa**” de la misma y estudiarla.

Las observaciones realizadas constituyen una **muestra**. Para poder tratar la información contenida en este conjunto de datos es preciso recurrir a las **técnicas estadísticas descriptivas**

12	91	23	23	12	91	23	23
34	23	45	98	34	23	45	98
23	45	56	34	23	45	56	34
56	32	34	68	56	32	34	68
12	91	23	23	12	91	23	23
34	23	45	98	34	23	45	98
23	45	56	34	23	45	56	34
56	32	34	68	56	32	34	68
12	91	23	23	12	91	23	23
34	23	45	98	34	23	45	98



La extensión de lo observado a la población vendrá después, mediante la **inferencia estadística**



## ESTADÍSTICA DESCRIPTIVA

### Consideraciones sobre los datos

#### Modalidad

Cada una de las maneras en las que se presenta un carácter.



#### Tipos de Datos

1. **Cualitativos:** aquellos que se refieren a una cualidad, no son expresables de manera rigurosa por un número.
  1. **Nominales:** las modalidades no son susceptibles de estar ordenadas (sexo, estado civil, grupo sanguíneo, etc). Si solo hay dos modalidades se habla de datos *Binarios* o *Dicotómicos*.
  2. **Ordinales:** cuando las modalidades son susceptibles de estar ordenadas (nivel de satisfacción, estado después del tratamiento,...)
2. **Cuantitativos:** aquellos que necesariamente requieren de un número para ser expresados de manera rigurosa.
  1. **Discretos:** aquellos datos que sólo pueden tomar "valores numéricos aislados" (número de hijos, número de visitas,... recuentos en general)
  2. **Continuos:** pueden tomar cualquier valor dentro de un intervalo, de modo que entre cualesquiera dos de ellos siempre existe otro valor posible (peso, estatura, valor hematocrito,... medidas en general)



## ESTADÍSTICA DESCRIPTIVA

### Resumen de la información

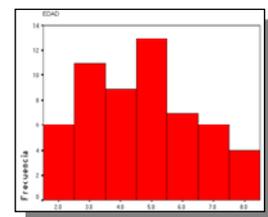
3 niveles:

- Tablas de frecuencias
- Representaciones gráficas
- Medidas descriptivas

Edad			
2	5	4	5
2	5	5	5
2	5	3	6
2	5	3	6
2	5	3	7
2	6	3	7
3	6	3	7
3	6	3	7
3	6	3	7
3	6	4	
4	5		
4	5		

**Datos**

EDAD			
Valores	Frecuencia	Porcentaje	Porcentaje acumulado
2.00	6	10.7	10.7
3.00	11	19.6	30.4
4.00	9	16.1	46.4
5.00	13	23.2	69.6
6.00	7	12.5	82.1
7.00	6	10.7	92.9
8.00	4	7.1	100.0
Total	56	100.0	



Desv. típ. = 1.75  
 Media = 4.7  
 N = 56.00

## ESTADÍSTICA DESCRIPTIVA

### Datos originales (tabla de casos x variables):

Tabla 2.1

**Grupo sanguíneo, número de hermanos y peso en (Kg.) de 500 alumnos varones de una Universidad**

Alumno n°	Grupo Sanguíneo	Número de hermanos	Peso
1	A	0	70,502
2	B	3	67,231
-	-	-	-
-	-	-	-
-	-	-	-
500	AB	2	71,676

## ESTADÍSTICA DESCRIPTIVA

### Tablas de frecuencias: variables nominales

Tabla 2.2

**Distribución del grupo sanguíneo de una muestra de 500 alumnos varones de una Universidad**

Grupo sanguíneo	f <sub>i</sub> (frecuencia absoluta)	h <sub>i</sub> (frecuencia relativa)	% (porcentaje)
<b>A</b>	150	0.300	30.00
<b>B</b>	75	0.150	15.00
<b>AB</b>	25	0.050	5.00
<b>0</b>	250	0.500	50.00
<b>Total</b>	500	1.000	100

Frecuencia absoluta (*recuento*):  $f_1 + f_2 + \dots + f_k = \sum_{i=1}^k f_i = n$

Frecuencia relativa (*porcentaje*):  $h_i = f_i/n$

$$h_1 + h_2 + \dots + h_k = \sum_{i=1}^k h_i = 1$$



## ESTADÍSTICA DESCRIPTIVA

### Tablas de frecuencias: variables cuantitativas discretas

Tabla 2.3

Distribución del número de hermanos (excluido él mismo) de una muestra de 500 alumnos varones de una Universidad

Número de hermanos	$f_i$	$h_i$	%	$F_i$	$H_i$	% acumulado
0	72	0.144	14.40	72	0.144	14.4
1	155	0.310	31.00	227	0.454	45.40
2	97	0.194	19.40	324	0.648	64.80
3	81	0.162	16.20	405	0.810	81.00
4	30	0.060	6.00	435	0.870	87.00
5	27	0.054	5.40	462	0.924	92.40
6	20	0.040	4.00	482	0.964	96.40
más de 6	18	0.036	3.60	500	1.000	100.00
<b>Total</b>	500	1.000	100			

Un 81% de los casos observados tienen 3 hermanos o menos, por tanto un 19% tienen 4 o más

A  
B 1  
B 2

## ESTADÍSTICA DESCRIPTIVA

### Tablas de frecuencias: variables cuantitativas continuas

Tabla 2.4

Distribución del peso (en Kg) de una muestra de 500 alumnos varones de una Universidad

Peso (Kg)	$f_i$	$h_i$	%	$F_i$	$H_i$	% acumulado
Menos de 45	1	0.002	0.20	1	0.002	0.2
[45 - 50)	3	0.006	0.60	4	0.008	0.8
[50 - 55)	12	0.024	2.40	16	0.032	3.2
[55 - 60)	75	0.150	15.00	91	0.182	18.2
[60 - 65)	103	0.206	20.60	194	0.388	38.8
[65 - 70)	155	0.310	31.00	349	0.698	69.8
[70 - 75)	101	0.202	20.20	450	0.900	90.0
[75 - 80)	29	0.058	5.80	479	0.958	95.8
[80 - 85)	11	0.022	2.20	490	0.980	98.0
[85 - 90)	8	0.016	1.60	498	0.996	99.6
90 o más	2	0.004	0.40	500	1.000	100.0
<b>Total</b>	500	1.000	100			

Un 71.8% de los alumnos de la muestra pesan entre 60 y 75 Kg



## ESTADÍSTICA DESCRIPTIVA

### Tablas de frecuencias

#### Condiciones de las tablas:

- 1º) Enunciado que las haga autosuficientes.
- 2º) En las columnas que sea necesario deben ir totales
- 3º) Unidades de medida.
- 4º) Igual número de decimales.
- 5º) Intervalos de clase de igual anchura.



## ESTADÍSTICA DESCRIPTIVA

### Representaciones gráficas

#### Principio general de toda representación gráfica:

La figura que representa a cada una de las modalidades de un carácter debe tener el área proporcional a la frecuencia de dicha modalidad.

#### Observaciones: Las representaciones gráficas

- Deben indicar las escalas y unidades de medida
- Deben explicarse por sí solas
- Deben contribuir a clarificar el material presentado

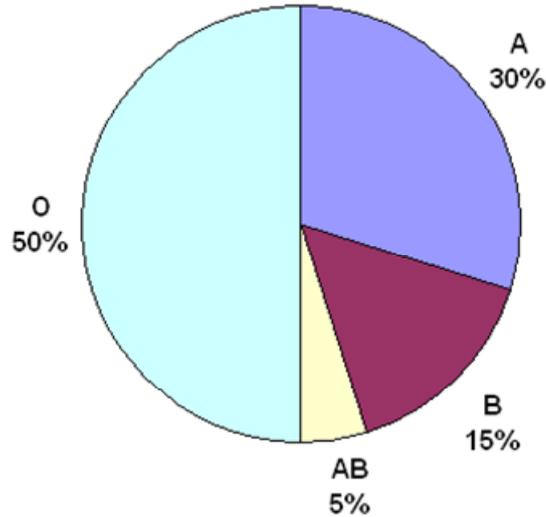


## ESTADÍSTICA DESCRIPTIVA

### Representaciones gráficas: variables nominales

#### Diagrama de sectores de la Tabla 2.2

Distribución del grupo sanguíneo de una muestra de 500 alumnos varones de una Universidad



$$\alpha_i^o = 360^o h_i$$



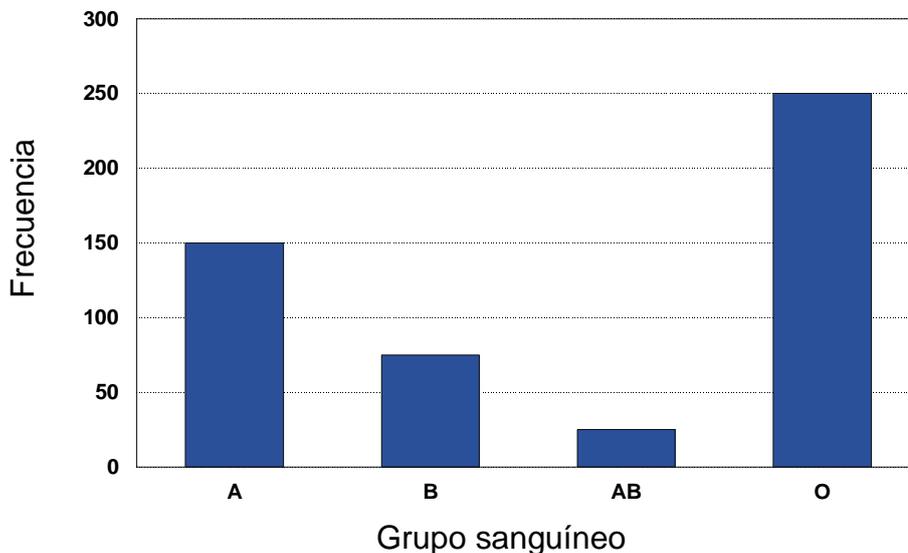



## ESTADÍSTICA DESCRIPTIVA

### Representaciones gráficas: variables nominales y ordinales

#### Diagrama de barras de la Tabla 2.2

Distribución del grupo sanguíneo de una muestra de 500 alumnos varones de una Universidad






## ESTADÍSTICA DESCRIPTIVA

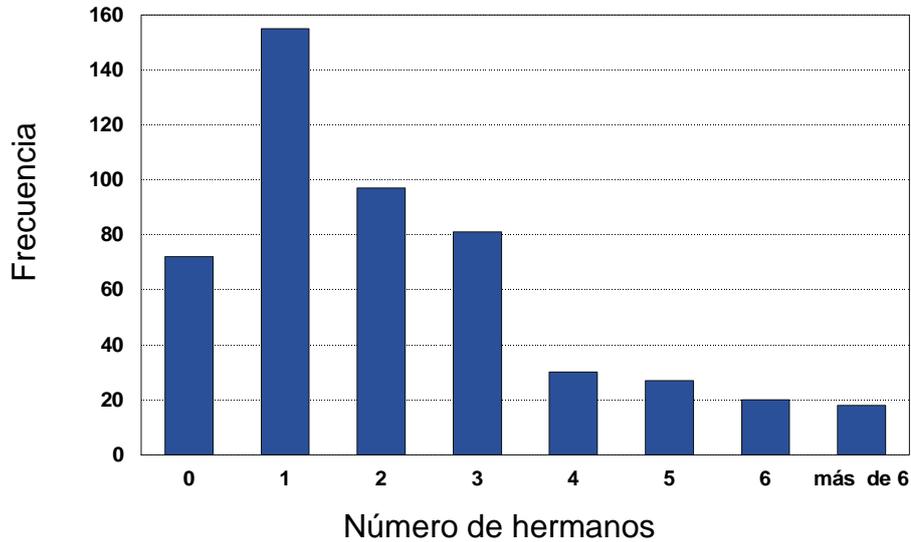
Estadística descriptiva - 15

Resúmenes: 2.3

### Representaciones gráficas: variables cuantitativas discretas

#### Diagrama de barras de la Tabla 2.3

Distribución del número de hermanos de una muestra de 500 alumnos varones de una Universidad



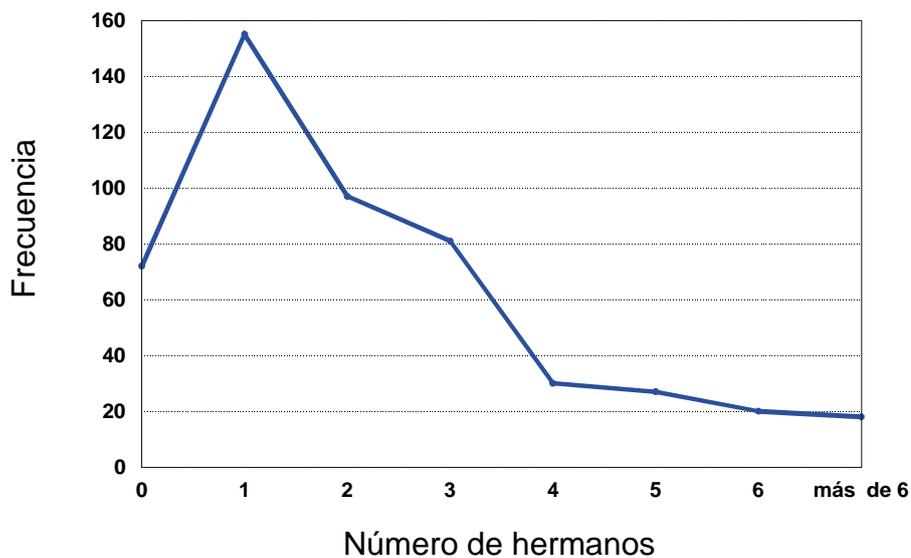
## ESTADÍSTICA DESCRIPTIVA

Estadística descriptiva - 16

Resúmenes: 2.3

#### Polígono de frecuencias de la Tabla 2.3

Distribución del número de hermanos de una muestra de 500 alumnos varones de una Universidad





## ESTADÍSTICA DESCRIPTIVA

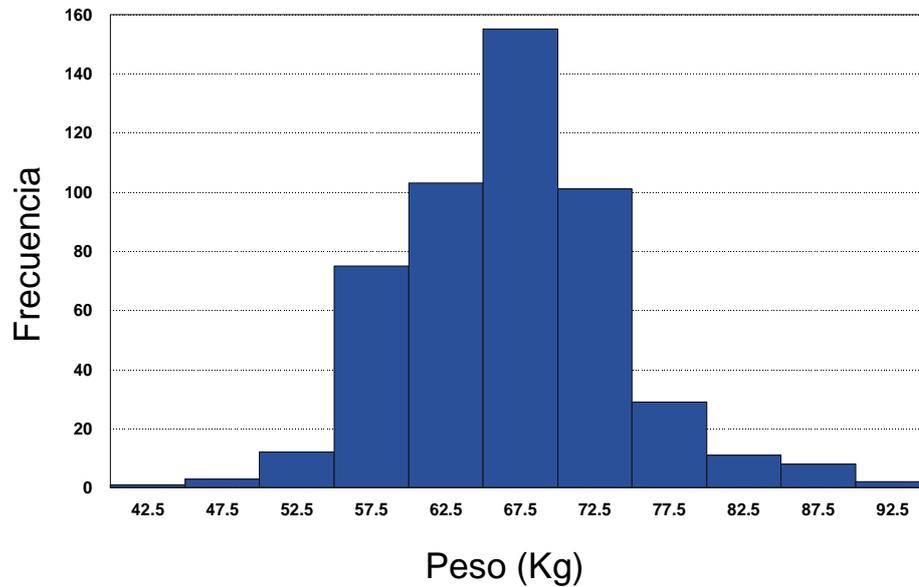
Estadística descriptiva - 17

Resúmenes: 2.3

### Representaciones gráficas: variables cuantitativas continuas

#### Histograma de la Tabla 2.4

Distribución del peso de una muestra de 500 alumnos varones de una Universidad



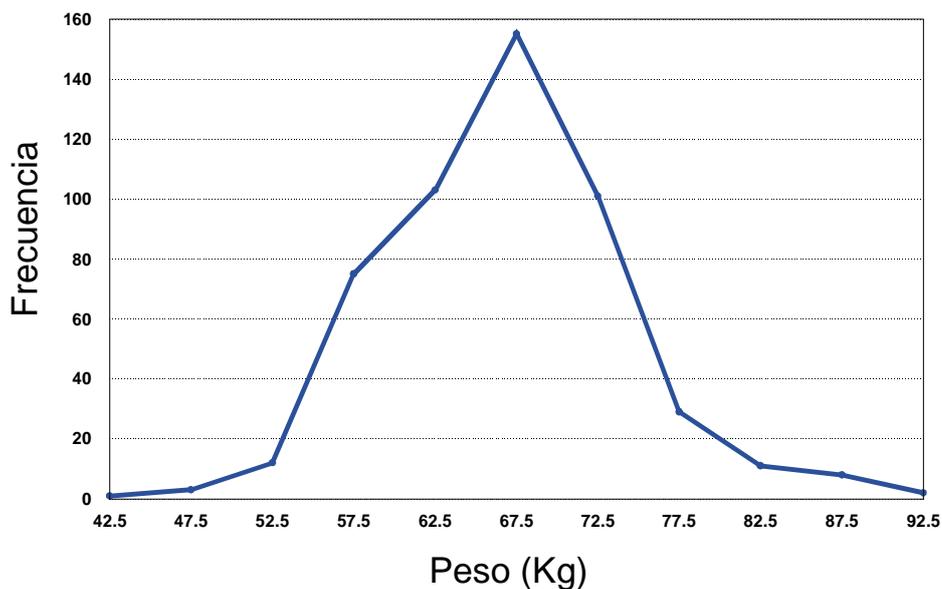
## ESTADÍSTICA DESCRIPTIVA

Estadística descriptiva -18

Resúmenes: 2.3

#### Polígono de frecuencias de la Tabla 2.4

Distribución del peso de una muestra de 500 alumnos varones de una Universidad





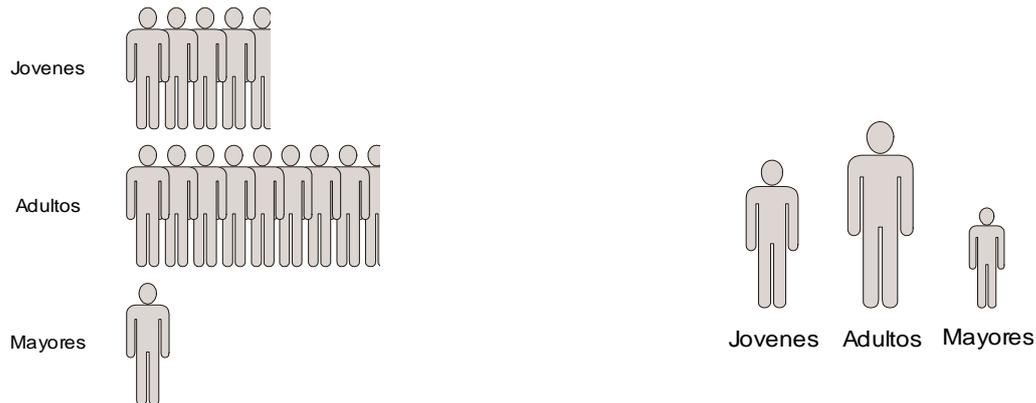
## ESTADÍSTICA DESCRIPTIVA

### Otras representaciones gráficas: pictogramas

Tabla 2.6

Pirámide de edad de una población de 31 millones de personas.

Edad (años)	0-19	20-64	Más de 64	Total
Porcentaje(%)	34,1	59,0	6,9	100,0



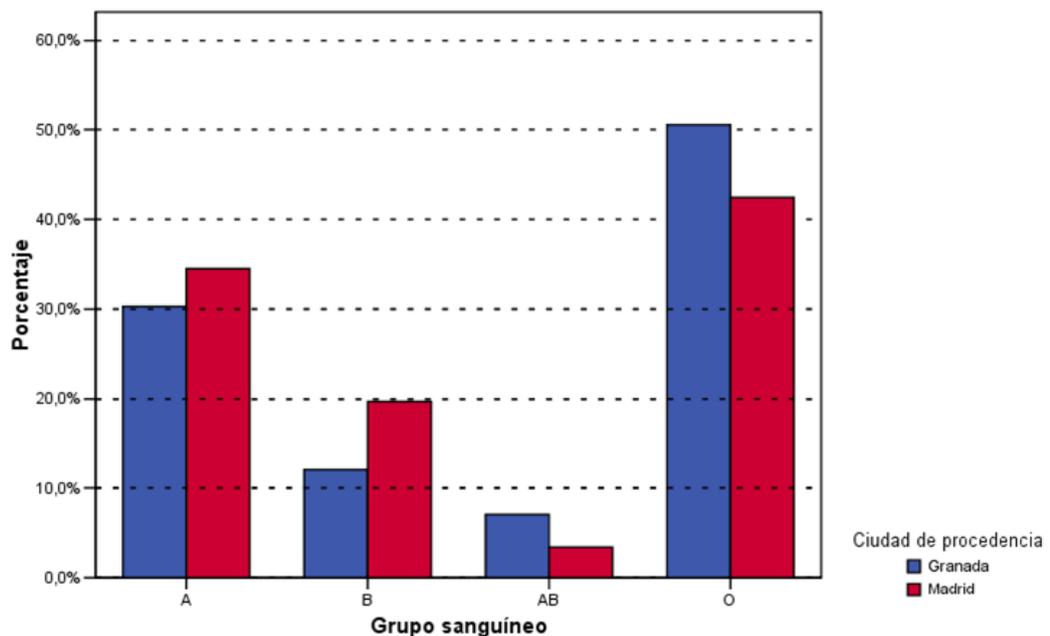
Pictograma de repetición

Pictograma de amplificación



## ESTADÍSTICA DESCRIPTIVA

### Gráficas comparativas: Diagramas de barras

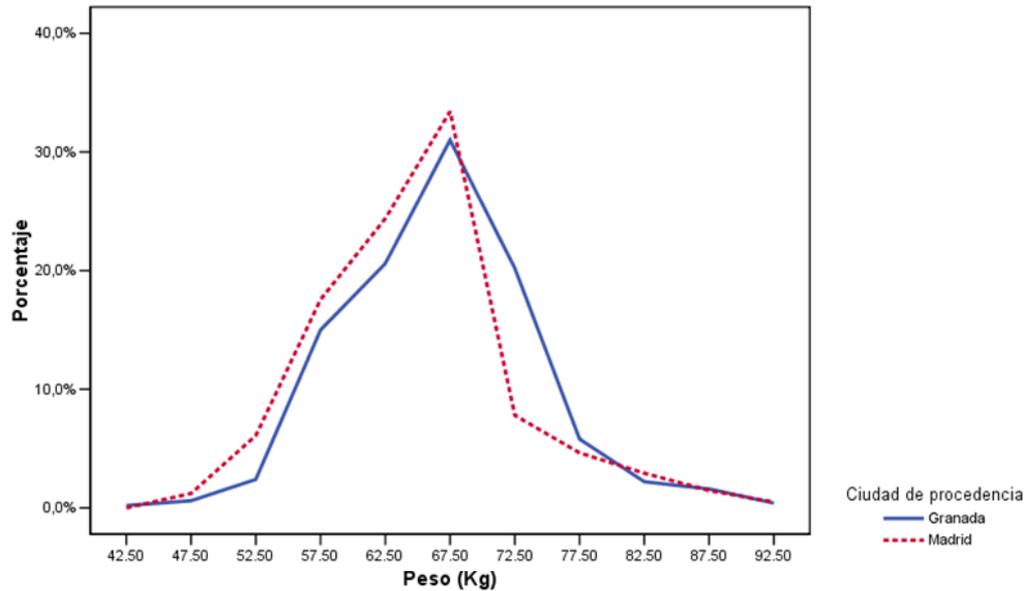
Distribución del grupo sanguíneo en dos muestras de universitarios de Granada ( $n_1=510$ ) y Madrid ( $n_2=520$ )



## ESTADÍSTICA DESCRIPTIVA

### Gráficas comparativas: polígonos de frecuencias

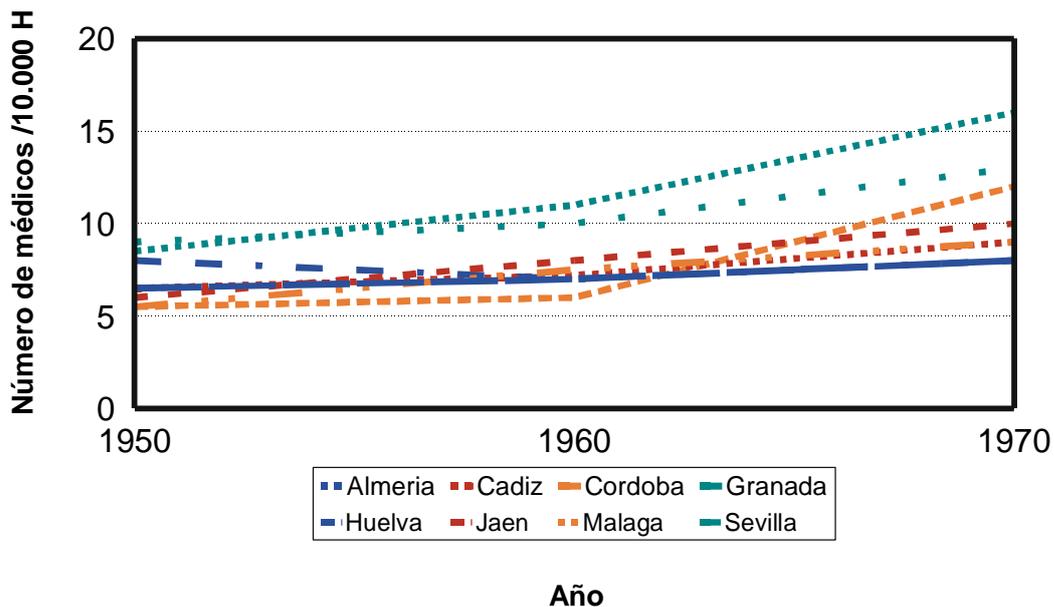
Distribución del peso en dos muestras de universitarios de Granada ( $n_1=500$ ) y Madrid ( $n_2=525$ )



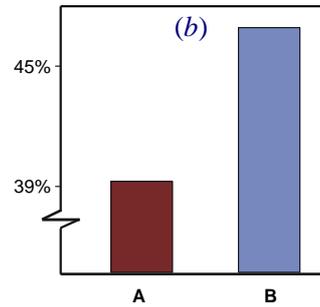
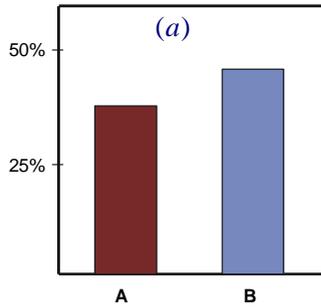
## ESTADÍSTICA DESCRIPTIVA

### Gráficas confusas

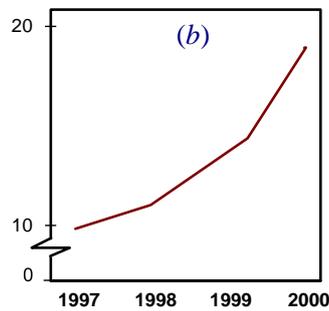
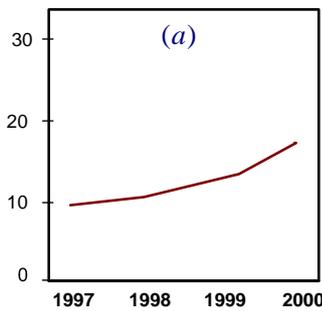
Número de médicos por 10.000 habitantes en tres décadas de este siglo en las 8 provincias andaluzas



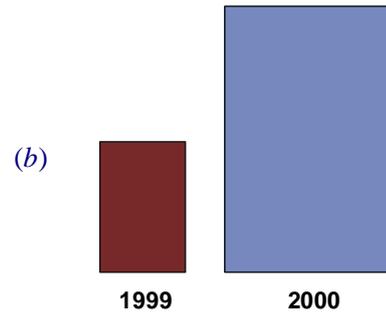
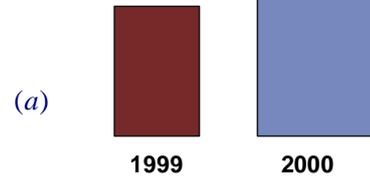
## Gráficas tendencias



1.- Porcentajes de curación de cierta enfermedad con dos fármacos A y B



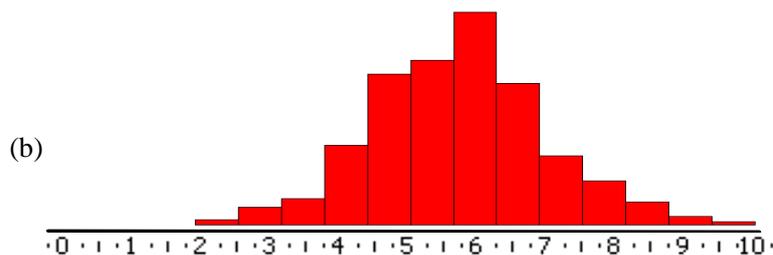
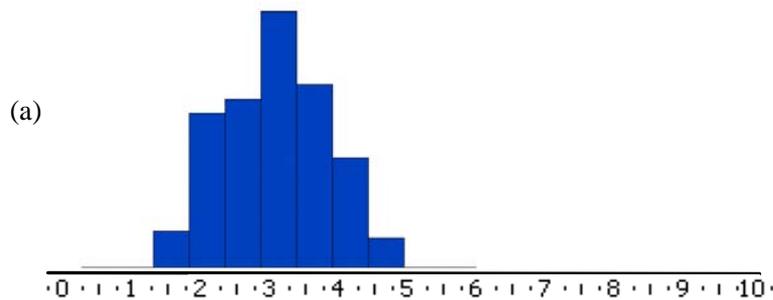
2.- Número de entradas (en miles) en un Servicio de Urgencias



3.- Consumo de tranquilizantes en los años indicados

## Síntesis de datos: medidas descriptivas

- **Medidas de posición:** describen la localización de la muestra
- **Medidas de dispersión:** describen cuánto de variables o dispersos son los datos



## ESTADÍSTICA DESCRIPTIVA

### Medidas descriptivas: Medidas de posición

**1.- Moda:** es el valor de la variable que tiene mayor frecuencia (puede no ser única).

- Ejemplos: [tabla 2.2](#), [tabla 2.3](#)

**2.- Mediana:** es el valor de la variable que divide a la muestra ordenada en dos partes iguales (es decir, deja tanto por debajo como por encima el 50% de las observaciones).

$$Me = x_{\left(\frac{n+1}{2}\right)}$$

- Ejemplos:

(a)  $\{3, 6, 7, 10, 15\} \rightarrow Me = x_{(3)} = 7$

(b)  $\{3, 6, 10, 15\} \rightarrow Me = x_{\left(\frac{4+1}{2}\right)} = \frac{6+10}{2} = 8$

(c) En la [tabla 2.3](#):  $Me = x_{((500+1)/2)} = x_{(250.5)} = 2$

## ESTADÍSTICA DESCRIPTIVA

### Medidas descriptivas: Medidas de posición

**3.- Media aritmética:**

a) Con datos no agrupados:  $x_1, \dots, x_n \rightarrow \bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$

b) Con datos agrupados:

$x$	$x_1$	...	$x_k$
$f$	$f_1$	...	$f_k$

$$\rightarrow \bar{x} = \frac{x_1 f_1 + \dots + x_k f_k}{f_1 + \dots + f_k} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

Ejemplo:

- con los datos de la tabla 2.3:  $\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n} = \frac{(0)(72) + (1)(155) + \dots + (7)(18)}{72 + 155 + \dots + 18} = \frac{1093}{500} = 2.19$

**4.- Media ponderada:**

$x$	$x_1$	...	$x_k$
$w$	$w_1$	...	$w_k$

$$\rightarrow \bar{x}_p = \frac{x_1 w_1 + \dots + x_k w_k}{w_1 + \dots + w_k} = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}$$

Ejemplo:  $\begin{array}{c|ccc} x & 5 & 7 & 9 \\ \hline w & 3 & 3 & 5 \end{array} \rightarrow \bar{x}_p = 81/11 = 7.36 \quad (\bar{x} = 7; \bar{x}_p \neq \bar{x})$

## ESTADÍSTICA DESCRIPTIVA

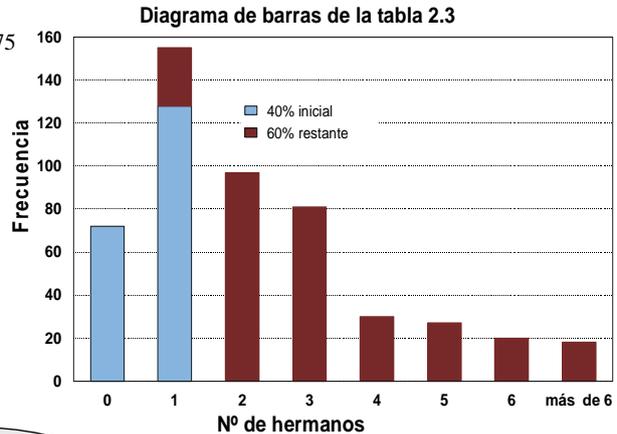
### Medidas descriptivas: Medidas de posición

**5.- Percentiles:** el **percentil  $\alpha$** ,  $P_\alpha$ , es el valor de la variable que divide a la muestra ordenada en dos partes, dejando por debajo el  $\alpha\%$  de las observaciones y por encima el  $(1 - \alpha)\%$ . Se habla entonces de percentil 1 ( $P_1$ ), ..., percentil 99 ( $P_{99}$ ). Casos particulares:

- Cuartiles:  $Q_1=P_{25}$ ,  $Q_2=P_{50}=Me$ ,  $Q_3=P_{75}$
- Deciles:  $D_1=P_{10}$ , ...,  $D_9=P_{90}$

- Ejemplo:

Percentil 40 de la [tabla 2.3](#)

$$501 \times 0.4 = 200.4; \quad P_{40} = \frac{x_{(200)} + x_{(201)}}{2}$$


Posición:	1	...	72	73	...	200	200...	227	228	...	
Valor:	0	...	0	1	...	1	1	...	1	2	...
						↑	↑				
						$x_{(200)}$	$x_{(201)}$			$\Rightarrow P_{40} = 200$	



## ESTADÍSTICA DESCRIPTIVA

### Medidas descriptivas: medidas de dispersión

**1.- Rango:**  $R = x_{\max} - x_{\min}$

Medida 'pobre', solo tiene en cuenta a dos observaciones de la muestra

**2.- Varianza:**

Dados  $n$  valores (no agrupados):  $x_1, \dots, x_n$   $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$

- Es una medida de dispersión que tiene en cuenta a todas las observaciones
- Se expresa en las unidades de la variable al cuadrado
- Obsérvese que  $\sum (x - \bar{x}) = 0$

Cálculo práctico:

$$s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right)$$

## ESTADÍSTICA DESCRIPTIVA

### Medidas descriptivas: medidas de dispersión

- Caso con datos agrupados:

$x$	$x_1$	...	$x_k$
$f$	$f_1$	...	$f_k$

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

- Ejemplos:

(a)  $x_i : 4, 6, 8, 10 \rightarrow \bar{x} = 7; s^2 = \frac{1}{3} \left( (4^2 + 6^2 + 8^2 + 10^2) - \frac{(4+6+8+10)^2}{4} \right) = 6.667$

(b)

$x_i$	4	6	8	10
$f_i$	1	4	3	2

$n=10$

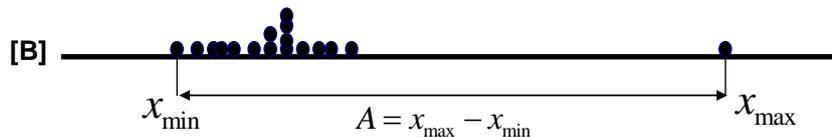
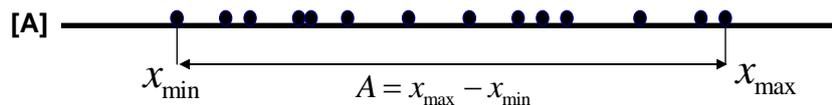
$$s^2 = \frac{1}{10-1} \left( ((1)(4^2) + (4)(6^2) + (3)(8^2) + (2)(10^2)) - \frac{((1)(4) + (4)(6) + (3)(8) + (2)(10))^2}{10} \right) = 3.733$$

### 3.-Desviación típica: $s = \sqrt{s^2}$

- Es una medida de dispersión que tiene en cuenta a todas las observaciones
- Se expresa en las mismas unidades que la variable

## ESTADÍSTICA DESCRIPTIVA

### Obtención de la varianza y desviación típica



**¿En cuál de estas dos distribuciones hay mas dispersión?**

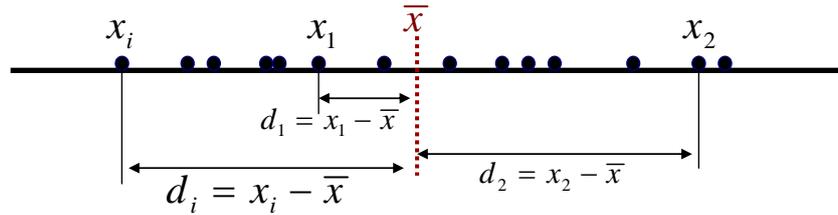
La amplitud es la misma en las dos, sin embargo es obvio que en B los datos están mas agrupados, son mas **homogéneos**.

El problema del rango es que solo considera a las dos observaciones mas extremas; no tiene en cuenta al resto

Es necesario encontrar una medida que refleje la **heterogeneidad, o dispersión**, de los datos pero considerándolos a todos, y no solo a los dos extremos



## ESTADÍSTICA DESCRIPTIVA



Podemos tomar una medida de referencia, por ejemplo la media aritmética, y calcular la distancia de cada observación a dicha referencia. Una posible medida de dispersión es la media de estas distancias:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{\sum (x_i - \bar{x})}{n}$$

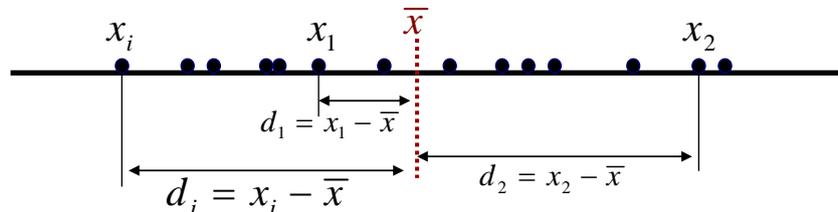
El problema es que  $\sum (x_i - \bar{x}) = 0$  siempre (¿por qué?)

Soluciones para evitar las distancias negativas:

1. Tomar valores absolutos ← ¡problemático! (aunque parezca lo contrario)
2. Elevar las distancias al cuadrado: da lugar a la **varianza**:

$$s^2 = \frac{\sum d_i^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n}$$

## ESTADÍSTICA DESCRIPTIVA



Por razones que veremos después conviene corregir este promedio tomando en el denominador no  $n$  sino  $n-1$ . En adelante siempre lo haremos así, de forma que la varianza resulta

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

El problema de la varianza como medida de dispersión es que sus unidades son las de la variable pero al cuadrado. Por ejemplo si  $x$  son centímetros (unidad de longitud) entonces la varianza son  $\text{cm}^2$  (unidad de superficie). Para expresar la dispersión en las mismas unidades de la variable se considera entonces su raíz cuadrada, que es la **desviación típica** o **estándar**

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

## ESTADÍSTICA DESCRIPTIVA

### Medidas descriptivas: medidas de dispersión

**4.-Coeficiente de variación:**  $CV = \frac{s}{\bar{x}} 100\%$

- Se utiliza para comparar métodos de medida
- Es una medida que tiene en cuenta a todas las observaciones
- Es adimensional y no depende de las unidades de medida
- Se expresa en %
- Cuanto menor es el CV, mayor es la precisión del método
- Ejemplo: CV=10% indica que por cada 100 unidades de medida, el método comete un error, medido en desviaciones típicas, de 10