

UNIDAD 10 **DISTRIBUCIONES UNIDIMENSIONALES**

1. INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

La **estadística** es la parte de las matemáticas que se ocupa de *recoger, organizar y analizar* grandes cantidades de datos para estudiar las características o comportamiento de un colectivo.

Uno de sus principales objetivos es realizar una *síntesis* de la información recogida. Para ello se calcularán una serie de *valores numéricos o parámetros* que permitirán resumir la información que previamente se agrupará en una tabla **estadística de frecuencias**.

De esta manera podremos describir y comparar el comportamiento de diversas distribuciones, así como predecir, modelizar y tomar decisiones en el estudio de una población a partir de los datos suministrados por una muestra.

Se puede considerar como antecedentes de la estadística cualquier recuento que en sociedades primitivas se pudieran hacer sobre cabezas de ganado, bienes... Sin embargo, formalmente podemos considerar que los antecedentes de la estadística están registrados por los primeros censos como los llevados a cabo en China alrededor del año 2 200 a.C.

Posteriormente fue avanzando con el conocimiento humano desarrollándose como ciencia a partir de los siglos XV y XVI.

TERMINOLOGÍA BÁSICA:

Población: Conjunto de elementos que son objeto del estudio estadístico.

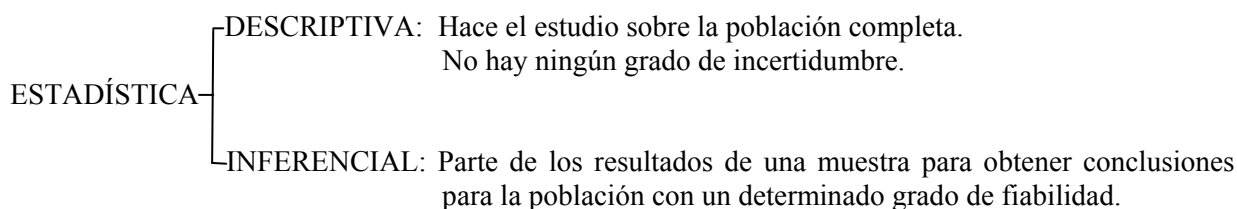
Individuo: Cada uno de los elementos de la población.

Muestra: Subconjunto de la población elegido para realizar un estudio estadístico.

Se toma una muestra cuando no es factible observar la característica objeto de estudio en todos los individuos de la población. Por ejemplo, si se quiere conocer la intención de voto en unas elecciones.

La muestra debe ser representativa de la población, es decir, las conclusiones obtenidas de su estudio se pueden extender a toda la población.

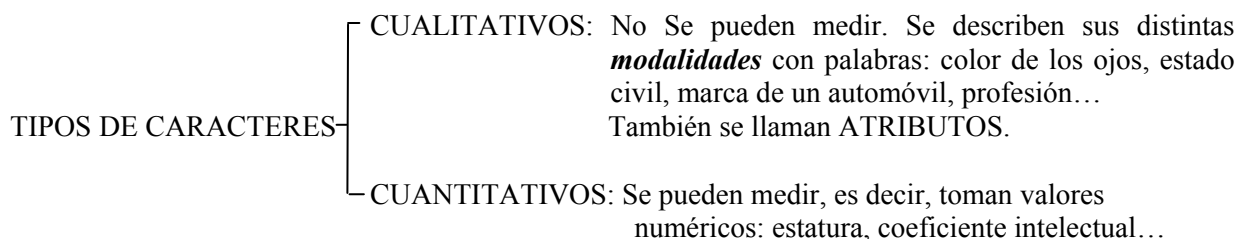
Distinguimos dos tipos de estadística:



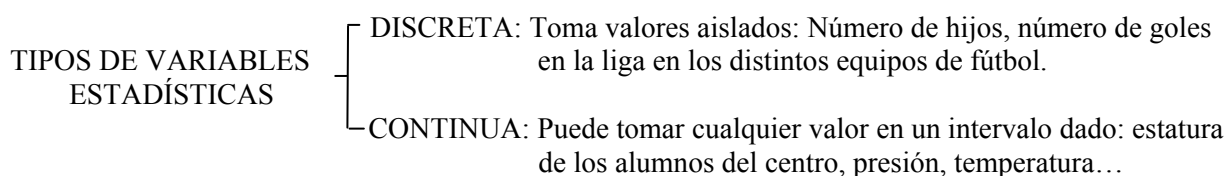
2. CARACTERES Y VARIABLES ESTADÍSTICAS

Carácter estadístico: Propiedad o característica de la población que se desea estudiar.

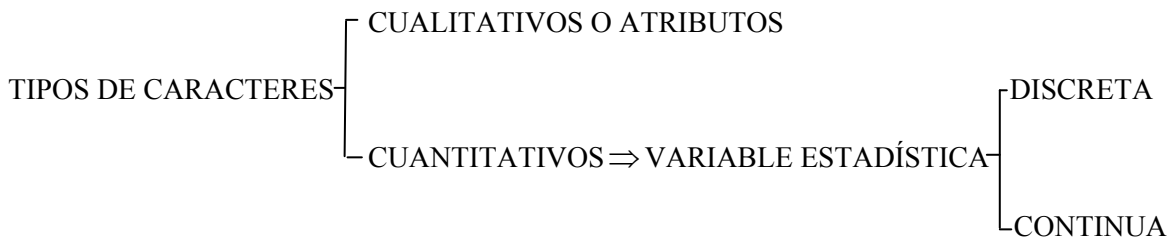
Distinguimos dos tipos de caracteres:



Variables estadísticas: Los caracteres cuantitativos pueden tomar distintos valores. Determinan, por tanto, una variable estadística.



Por tanto:



3. ORGANIZACIÓN DE DATOS. TABLAS ESTADÍSTICAS DE FRECUENCIAS

Frecuencia absoluta n_i de una modalidad o valor x_i : Número de veces que se repite ese valor o modalidad.

Frecuencia absoluta acumulada N_i : $N_i = n_1 + n_2 + \dots + n_i$, es decir, $N_i = \sum_{k=1}^i n_k$

Frecuencia relativa f_i de una modalidad o valor x_i : Es el cociente entre su frecuencia absoluta y el número total de datos N , es decir, $f_i = \frac{n_i}{N}$

Frecuencia relativa acumulada F_i : $F_i = f_1 + f_2 + \dots + f_i$, o bien, $F_i = \frac{N_i}{N}$

En toda tabla estadística de frecuencias se cumple: $\sum_{i=1}^n n_i = N$ $\sum_{i=1}^n f_i = 1$

3.1 TABLA PARA CARACTERES CUALITATIVOS

Ejemplo: Al observar el color de pelo de 110 alumnos de un instituto observamos que hay 16 con tono rubio, 24 morenos, 35 castaños, 10 pelirrojos y 25 con otras coloraciones.

Formamos la tabla estadística de frecuencias:

COLOR DE PELO	n_i	f_i
Rubio	16	$16/110 \approx 0.15$
Moreno	24	$24/110 \approx 0.22$
Castaño	35	$35/110 \approx 0.32$
Pelirrojo	10	$10/110 \approx 0.09$
Otros	25	$25/110 \approx 0.22$
	$N = 110$	1

3.2 TABLA PARA DATOS NO AGRUPADOS

Ejemplo: Un profesor tiene anotadas en su cuaderno las notas de los 30 alumnos/as de una clase:

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

Formamos la tabla estadística de frecuencias:

Notas (x_i)	n_i	N_i	f_i	F_i
0	2	2	$2/30 \approx 0.07$	$2/30 \approx 0.07$
1	3	5	$3/30 = 0.1$	$5/30 \approx 0.17$
2	1	6	$1/30 \approx 0.03$	$6/30 = 0.2$
3	1	7	$1/30 \approx 0.03$	$7/30 \approx 0.23$
4	1	8	$1/30 \approx 0.03$	$8/30 \approx 0.27$
5	3	11	$3/30 = 0.1$	$11/30 \approx 0.37$
6	2	13	$2/30 \approx 0.07$	$13/30 \approx 0.43$
7	5	18	$5/30 \approx 0.17$	$18/30 = 0.6$
8	7	25	$7/30 \approx 0.23$	$25/30 \approx 0.83$
9	5	30	$5/30 \approx 0.17$	$30/30 = 1$
	$N = 30$		1	

Esta tabla se usa con variables estadísticas discretas con una cantidad no excesiva de datos.

3.3 TABLA PARA DATOS AGRUPADOS EN INTERVALOS O CLASES

Si la variable estadística es continua, o bien es discreta pero con gran cantidad de datos, éstos deben agruparse en **intervalos** o **clases**.

Marca de clase: Valor medio de cada intervalo. Se obtiene haciendo la semisuma de los extremos del intervalo. Es el valor representativo de ese intervalo o clase.

Usualmente se toman todos los intervalos con la misma amplitud salvo que existan valores muy dispersos. En este último caso es conveniente obtener la **densidad de frecuencia** en cada clase que se define como la frecuencia por unidad de medida en cada intervalo:

$$d_i = \frac{n_i}{a_i} \quad a_i \rightarrow \text{Amplitud del intervalo } i\text{-ésimo}$$

En general, se considerarán los intervalos cerrados a la izquierda y abiertos a la derecha.

Ejemplo: A los 100 empleados de una empresa de piezas de precisión se les ha realizado una prueba de habilidad manual. En una escala de 0 a 100 se han obtenido las siguientes puntuaciones:

27 66 32 55 46 37 75 81 18 33 47 74 37 52 47 66 80 87 37 29
 46 15 29 90 76 67 23 35 94 23 25 56 73 78 17 28 76 58 45 36
 55 60 17 56 23 82 64 50 51 45 37 65 62 26 69 36 54 42 40 54
 27 62 28 65 46 92 36 33 23 66 18 82 47 49 59 45 73 43 47 83
 78 65 39 36 53 91 38 35 68 78 91 23 34 43 55 56 74 56 62 38

Formamos la tabla estadística de frecuencias:

Habil. Manual	x_i	n_i	N_i	f_i	F_i
[15, 25)	20	10	10	10/100 ≈ 0.10	10/100 = 0.10
[25, 35)	30	12	22	12/100 ≈ 0.12	22/100 = 0.22
[35, 45)	40	17	39	17/100 ≈ 0.17	39/100 = 0.39
[45, 55)	50	17	56	17/100 ≈ 0.17	56/100 = 0.56
[55, 65)	60	14	70	14/100 ≈ 0.14	70/100 = 0.70
[65, 75)	70	13	83	13/100 ≈ 0.13	83/100 = 0.83
[75, 85)	80	11	94	11/100 ≈ 0.11	94/100 = 0.94
[85, 95)	90	6	100	6/100 ≈ 0.06	100/100 = 1
		$N = 100$		1	

4. REPRESENTACIONES GRÁFICAS

Permiten comunicar mediante un gráfico los resultados de un estudio estadístico.

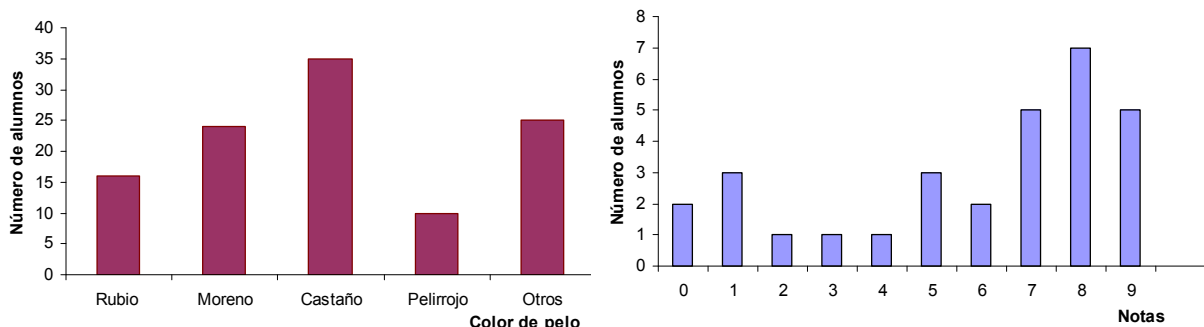
4.1 DIAGRAMA DE BARRAS

Se usa con caracteres cualitativos (atributos), o cuantitativos discretos.

En el eje horizontal se ponen las distintas modalidades del carácter en el caso de los atributos, o bien los valores x_i que toma la variable estadística discreta.

A continuación se escriben las frecuencias absolutas en el eje vertical y se levanta una barra o rectángulo correspondiente a cada modalidad o valor x_i de altura su correspondiente frecuencia n_i .

El diagrama de barras de los ejemplos de los apartados 3.1 y 3.2 sería:

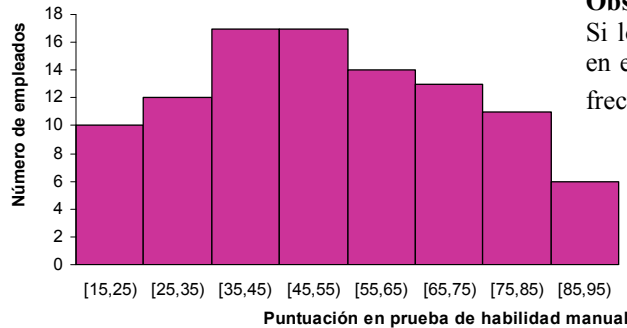


4.2 HISTOGRAMA

Se utiliza con caracteres cuantitativos agrupados en intervalos o clases.

En cada intervalo o clase se forma un rectángulo que tiene por base la amplitud del intervalo y como altura su n_i correspondiente. Por tanto los rectángulos están unidos entre sí.

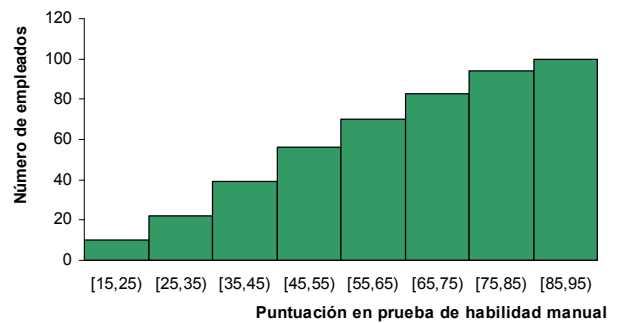
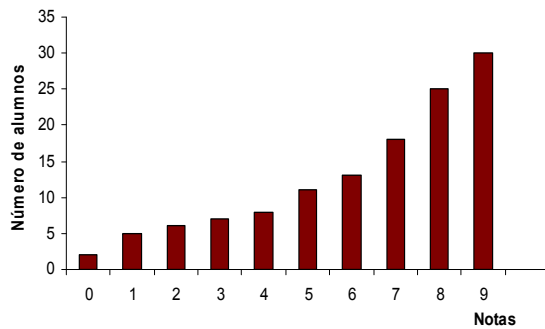
El histograma del ejemplo del apartado 3.3 sería:



Observación:

Si los intervalos presentan distinta amplitud, en el eje vertical se usarán las densidades de frecuencia d_i en lugar de n_i .

En el caso de variables cuantitativas, si en lugar de usar en el eje vertical la frecuencia absoluta n_i se usa la frecuencia absoluta acumulada N_i , se obtiene el diagrama de barras acumulado o el histograma de frecuencias acumulado según corresponda. En los ejemplos anteriores quedan así:

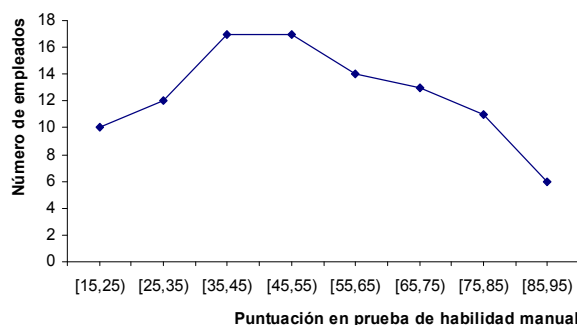
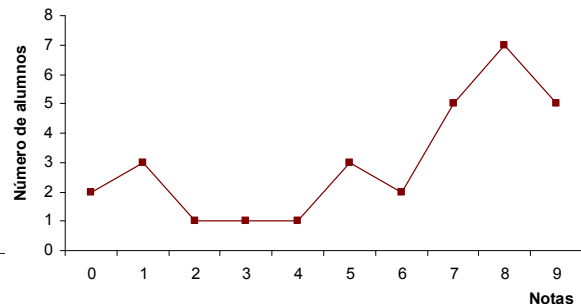
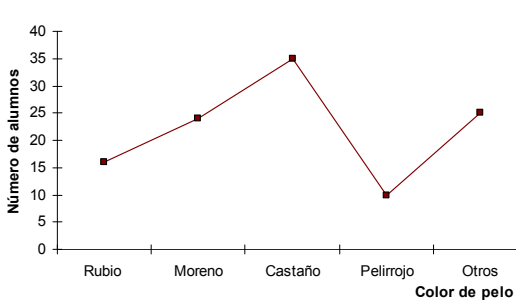


4.3 POLÍGONO DE FRECUENCIAS

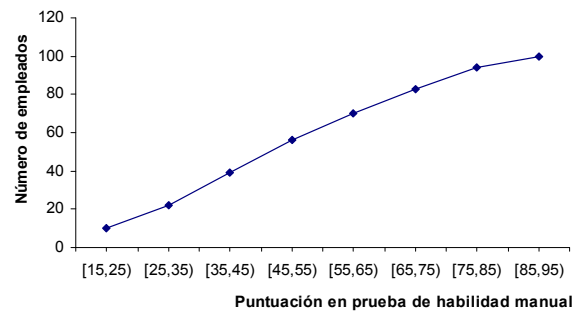
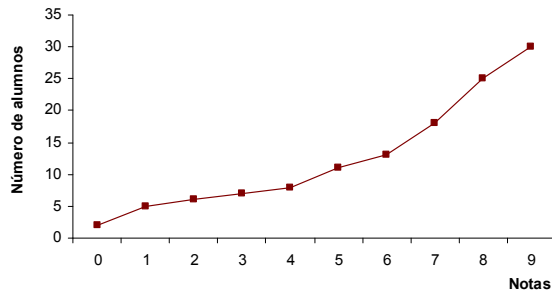
Se usa con todo tipo de caracteres.

Es una línea poligonal que une los extremos de las barras que forman el diagrama de barras o los puntos medios de la parte superior de los rectángulos del histograma.

Los polígonos de frecuencias de los tres ejemplos anteriores son:



En el caso de variables cuantitativas, si en lugar de usar en el eje vertical la frecuencia absoluta n_i se usa la frecuencia absoluta acumulada N_i , se obtienen los polígonos de frecuencias acumulados. En los ejemplos anteriores son:



4.4 DIAGRAMA DE SECTORES

Se utiliza, en la mayoría de los casos, con caracteres cualitativos.

Se divide el círculo en sectores circulares de modo que su área sea proporcional a la frecuencia absoluta n_i de la modalidad considerada del carácter estudiado.

El diagrama de sectores del ejemplo 3.1 sería:

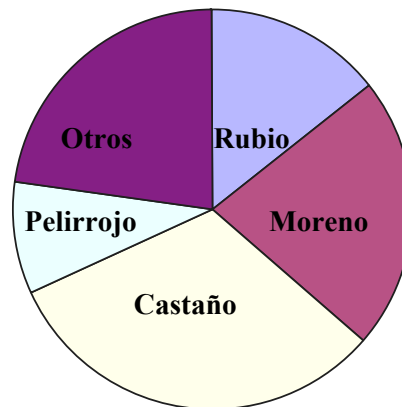
$$\text{Rubio: } \frac{16}{110} = \frac{\alpha}{360^\circ} \Rightarrow \alpha \approx 52.36^\circ$$

$$\text{Moreno: } \frac{24}{110} = \frac{\alpha}{360^\circ} \Rightarrow \alpha \approx 78.55^\circ$$

$$\text{Castaño: } \frac{35}{110} = \frac{\alpha}{360^\circ} \Rightarrow \alpha \approx 114.55^\circ$$

$$\text{Pelirrojo: } \frac{10}{110} = \frac{\alpha}{360^\circ} \Rightarrow \alpha \approx 32.73^\circ$$

$$\text{Otros: } \frac{25}{110} = \frac{\alpha}{360^\circ} \Rightarrow \alpha \approx 81.82^\circ$$



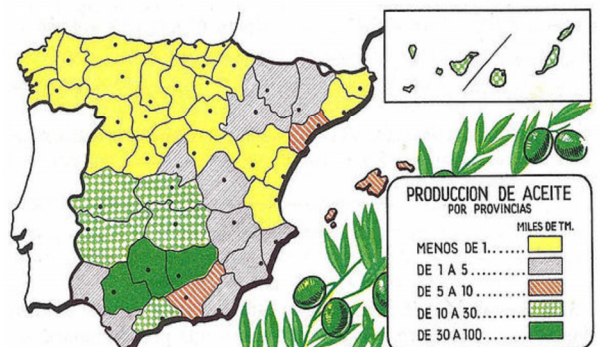
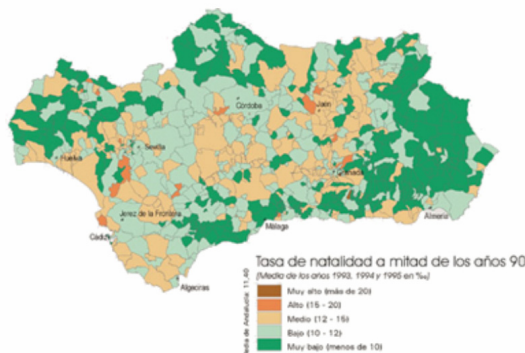
4.5 PICTOGRAMA

Son dibujos alusivos a la distribución que se utilizan para comparar datos no agrupados en intervalos o clases. La frecuencia de las distintas modalidades se pondrá de relieve bien por el tamaño de estas figuras o bien repitiendo el dibujo un número de veces proporcional a la frecuencia absoluta de cada una de las modalidades representadas. Su principal problema es que tienen muy poca precisión.



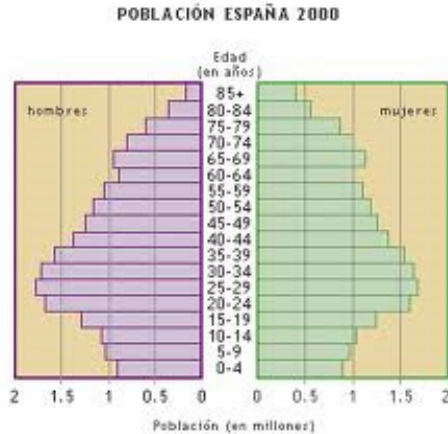
4.6 CARTOGRAMA

Consiste en representar sobre un mapa cualquier tipo de dato relacionado con un área geográfica.



4.7 PIRÁMIDES DE POBLACIÓN

Son dos histogramas horizontales que comparten el eje vertical que contiene los intervalos de clase. En uno de ellos, se representan los datos correspondientes a la distribución por edades del sexo masculino y en el otro, los del sexo femenino.



5. MEDIDAS DE POSICIÓN

Tienen como objetivo resumir, en pocos valores numéricos, la información contenida en las tablas estadísticas de frecuencias.

Estos valores representativos del conjunto de datos reciben el nombre de **parámetros** o **medidas de posición**.

5.1 MEDIDAS DE TENDENCIA CENTRAL

Aportan valores numéricos que se interpretan como resumen global de la distribución.

- **MEDIA ARITMÉTICA**

Media ponderada:

$$\bar{x}_p = \frac{\sum_{i=1}^n x_i \cdot p_i}{\sum_{i=1}^n p_i}$$

$p_i \rightarrow$ Pesos

Valor que se obtiene al dividir la suma de todos los valores de la variable, multiplicados por sus frecuencias absolutas, entre el número total de éstos.

$$\bar{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_n \cdot n_n}{N}$$

de modo abreviado

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot n_i}{N}$$

(También $\bar{x} = \sum_{i=1}^n x_i \cdot f_i$)

Media geométrica:

$$G = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_n^{n_n}}$$

Cálculo: tomar *log* en ambos miembros.

Si la distribución está agrupada en intervalos o clases, x_i será la marca de clase de cada intervalo.

Cálculo práctico de la media aritmética:

Media armónica:

$$H = \frac{N}{\sum_{i=1}^n \frac{1}{x_i} \cdot n_i}$$

Caso 1: Datos no agrupados

Ejemplo: El número de hijos de las familias de 40 alumnos de 1º de Bachillerato viene dado por la tabla:

Nº hijos (x_i)	n_i	$x_i \cdot n_i$
1	7	7
2	14	28
3	9	27
4	8	32
5	2	10
	$N = 40$	104

$$\bar{x} = \frac{\sum_{i=1}^5 x_i \cdot n_i}{N} = \frac{104}{40} = 2.6$$

Es decir, el número medio de hijos por familia es de 2.6

Caso 2: Datos agrupados en intervalos o clases

Según el estadístico **Lidny Yule**, un parámetro de centralización o de dispersión debe verificar:

- Definirse de modo objetivo.
- Usar todas las observaciones.
- Tener un significado concreto.
- Ser sencillo de calcular.
- Prestarse al cálculo algebraico.
- Ser poco sensible a fluctuaciones muestrales.

Ejemplo: La duración en horas de 30 bombillas viene dada por la tabla:

Número horas	x_i	n_i	$x_i \cdot n_i$
[310, 420)	365	1	365
[420, 530)	475	9	4275
[530, 640)	585	11	6435
[640, 750)	695	5	3475
[750, 860)	805	3	2415
[860, 970)	915	1	915
		$N = 30$	17880

$$\bar{x} = \frac{\sum_{i=1}^6 x_i \cdot n_i}{N} = \frac{17880}{30} = 596$$

Es decir, la duración media de cada bombilla es de 596 horas.

• **MODA**

Es el valor de la variable estadística, o la modalidad en el caso de atributos, con mayor frecuencia absoluta.

La moda puede no ser única, es decir, la distribución puede tener dos modas, tres..., en cuyo caso reciben en nombre de bimodal, trimodal...

Estas distribuciones se llaman plurimodales.

Cálculo práctico de la moda:

Caso 1: Datos no agrupados o atributos

Ejemplo 1: Si observamos la tabla de frecuencias del punto 3.1:

COLOR DE PELO	n_i
Rubio	16
Moreno	24
Castaño	35
Pelirrojo	10
Otros	25
$N = 110$	

$Mo = \text{Castaño}$

Es decir, la moda de la distribución es tener color de pelo castaño.

Ejemplo 2: El número de hijos de las familias de 40 alumnos de 1º de Bachillerato viene dado por la tabla:

Nº hijos (x_i)	n_i
1	7
2	14
3	9
4	8
5	2
$N = 40$	

$Mo = 2$

Es decir, la moda de la distribución es tener dos hijos por familia.

Caso 2: Datos agrupados en intervalos o clases

Intervalo o clase modal (I_{Mo}): la que tiene mayor frecuencia absoluta.

Como una primera aproximación puede tomarse como MODA la marca de clase de la clase modal, pero si se desea una mayor precisión se obtendrá con la expresión:

$$Mo = L_i + a_i \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

Siendo:

L_i → Límite inferior de la clase modal

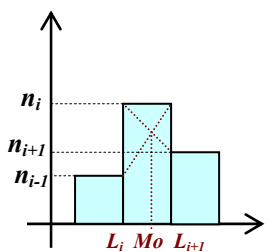
a_i → Amplitud de la clase modal

n_i → Frecuencia absoluta clase modal

Si los intervalos no tienen la misma amplitud se usará d_i en lugar de n_i .

Ejemplo: La duración en horas de 30 bombillas viene dada por la tabla:

Método gráfico de cálculo de la moda (semejanza)



En distribuciones agrupadas

Número horas	x_i	n_i
[310, 420)	365	1
[420, 530)	475	9
[530, 640)	585	11
[640, 750)	695	5
[750, 860)	805	3
[860, 970)	915	1
		$N = 30$

Intervalo Modal: $I_{Mo} = [530, 640)$

$$Mo = 530 + 110 \frac{11 - 9}{(11 - 9) + (11 - 5)} = 557.5$$

⇒ $Mo = 557.5$ horas

Es decir, la moda de la distribución es que una bombilla dure 557.5 horas.

Fíjate, una aproximación menos precisa hubiese sido tomar la marca de clase $Mo = 585$.

• **MEDIANA**

Es el valor de la variable estadística que divide a la población en dos partes iguales, es decir, el 50% de las observaciones serán inferiores a dicho valor y el otro 50% serán superiores.

Es, por tanto, el valor que ocupa el lugar **central** en un conjunto ordenado de datos.

Cálculo práctico de la mediana:

En cualquier caso, lo primero que se hará es calcular $\frac{N}{2}$.

Caso 1: Datos no agrupados

Observamos el primer valor de las frecuencias absolutas acumuladas N_i igual o superior a $\frac{N}{2}$ y se obtiene la mediana Me como se describe en los dos ejemplos siguientes:

Ejemplo 1:

x_i	n_i	N_i
1	7	7
2	14	21
3	9	30
4	8	38
5	2	40
$N = 40$		

$$\frac{N}{2} = \frac{40}{2} = 20$$

$Me = 2$

Ejemplo 2:

x_i	n_i	N_i
1	7	7
2	13	20
3	10	30
4	7	37
5	3	40
$N = 40$		

$$\frac{N}{2} = \frac{40}{2} = 20$$

$$Me = \frac{2+3}{2} = 2.5$$

$Me = 2.5$

Fíjate:

1º) Se calcula $\frac{N}{2}$ y observamos la tabla en N_i (frecuencias absolutas acumuladas).

2º) Si no coincide con ningún valor de N_i ⇒ Me es el valor del primer valor x_i tal que $\frac{N}{2} < N_i$.

3º) Si $\frac{N}{2} = N_i$ entonces $Me = \frac{x_i + x_{i+1}}{2}$.

Observación:

Si el número de datos es pequeño no hace falta obtener la tabla estadística de frecuencias:

Número impar de datos:

20, 23, 25, 20, 29, 26, 25, 25, 23

⇓

20, 20, 23, 23, 25, 25, 25, 26, 29

⇓

$$Me = 25$$

Número par de datos:

2, 3, 8, 3, 7, 8, 5, 4, 3, 2

⇓

2, 2, 3, 3, 3, 4, 5, 7, 8, 8

⇓

$$Me = \frac{3+4}{2} = 3.5$$

Caso 2: Datos agrupados en intervalos o clases

Una vez calculado $\frac{N}{2}$ se determina el **Intervalo mediano o clase mediana** I_{Me} , que es el que va a contener a la mediana.

Aunque se podría tomar su marca de clase como una primera aproximación a la mediana, ésta se puede obtener con mayor precisión a partir de la expresión:

$$Me = L_i + a_i \frac{\frac{N}{2} - N_{i-1}}{n_i}$$

Siendo:

L_i → Límite inferior de la clase mediana

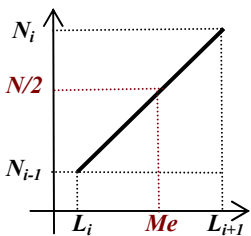
a_i → Amplitud de la clase mediana

n_i → Frecuencia absoluta clase mediana

N_i → Frecuencia absoluta acumulada clase mediana

Ejemplo: Una estación meteorológica registró 88 días de lluvia el pasado año, según se muestra en la tabla siguiente:

Método gráfico de cálculo de la mediana (semejanza)



En distribuciones agrupadas

Litros/ m ²	x_i	n_i	N_i
[0, 5)	2.5	3	3
[5, 10)	7.5	7	10
[10, 15)	12.5	19	29
[15, 20)	17.5	23	<u>52</u>
[20, 25)	22.5	18	70
[25, 30)	27.5	12	82
[30, 35)	32.5	6	88

$N = 88$

Obtención del intervalo mediano:

$$\frac{N}{2} = \frac{88}{2} = 44 \Rightarrow I_{Me} = [15, 20)$$

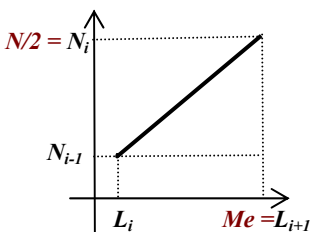
Por tanto:

$$Me = 15 + 5 \frac{44 - 29}{23} \approx 18.26 \Rightarrow$$

$$\Rightarrow \boxed{Me \approx 18.26} \text{ litros/m}^2$$

Ejemplo: Obtén la mediana de este conjunto de datos agrupados por intervalos:

Método gráfico de cálculo de la mediana en este caso



En distribuciones agrupadas

$[L_i, L_{i+1})$	x_i	n_i	N_i
[0, 1)	0.5	1	1
[1, 3)	2	6	7
[3, 6)	4.5	7	<u>14</u>
[6, 10)	8	12	26
[10, 12)	11	2	28

$N = 28$

Obtención del intervalo mediano:

$$\frac{N}{2} = \frac{28}{2} = 14 \Rightarrow I_{Me} = [3, 6)$$

En este caso $\Rightarrow \boxed{Me = 6}$

Fíjate, si aplicamos la fórmula:

$$Me = 3 + 3 \frac{14 - 7}{7} = 6 \Rightarrow \boxed{Me = 6}$$

5.2 MEDIDAS DE TENDENCIA NO CENTRAL (CUANTILES)

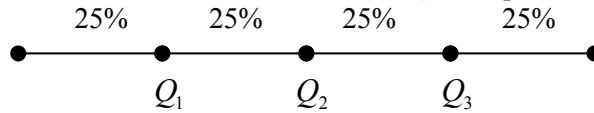
Los cuantiles son medidas de posición ya que van a suministrar valores entre los cuales se encuentra parte de la distribución.

Van a dividir a la distribución en partes iguales y van a recibir distintos nombres, según el número de partes en que se divida la distribución.

• **CUARTILES**

Son 3 valores de la variable, Q_1, Q_2, Q_3 , que dividen a la población en 4 partes iguales.

Entre dos cuartiles consecutivos se encuentra 1/4 (25%) de la población.



Es evidente que $Q_2 = Me$.

Fíjate:

Q_1 deja por debajo el 25% de los datos.

Q_2 deja por debajo el 50% de los datos.

Q_3 deja por debajo el 75% de los datos.

Cálculo práctico de cuartiles:

El procedimiento es análogo al del cálculo de la mediana, tanto en el caso de datos no agrupados como en el caso de datos agrupados en intervalos o clases.

Caso 1: Datos no agrupados

Observamos el primer valor de las frecuencias absolutas acumuladas N_i igual o inferior a $\frac{N}{4}$, $\frac{2N}{4} = \frac{N}{2}$ o bien $\frac{3N}{4}$ según se desee obtener Q_1, Q_2 o Q_3 respectivamente, y se procede de la misma forma que con la mediana.

Caso 2: Datos agrupados en intervalos o clases

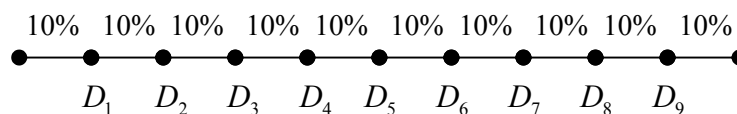
Una vez calculado $\frac{N}{4}$, $\frac{2N}{4} = \frac{N}{2}$ o bien $\frac{3N}{4}$, según se desee obtener Q_1, Q_2 o Q_3 respectivamente, se determina el intervalo en el que se va a encontrar el cuartil correspondiente. Aunque se podrían tomar sus marcas de clase como una primera aproximación de los cuartiles, éstos se pueden obtener con mayor precisión a partir de las expresiones:

$$Q_1 = L_i + a_i \frac{\frac{N}{4} - N_{i-1}}{n_i} \qquad Q_2 = Me \qquad Q_3 = L_i + a_i \frac{\frac{3N}{4} - N_{i-1}}{n_i}$$

• **DECILES**

Son 9 valores de la variable, D_1, D_2, \dots, D_9 , que dividen la población en 10 partes iguales.

Entre dos deciles consecutivos se encuentra 1/10 (10%) de la población.



Cálculo práctico de deciles:

El procedimiento es análogo al del cálculo de los cuartiles, tanto en el caso de datos no agrupados como en el caso de datos agrupados en intervalos o clases.

Caso 1: Datos no agrupados

Observamos el primer valor de las frecuencias absolutas acumuladas N_i igual o inferior a $\frac{N}{10}, \frac{2N}{10}, \dots, \frac{9N}{10}$ según se desee obtener D_1, D_2, \dots, D_9 respectivamente, y se procede de la misma forma que con los cuartiles.

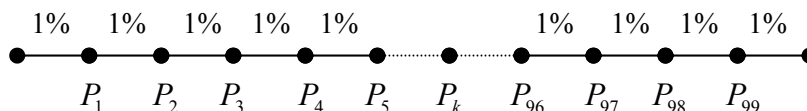
Caso 2: Datos agrupados en intervalos o clases

Una vez calculado $\frac{N}{10}, \frac{2N}{10}, \dots, \frac{9N}{10}$, según se desee obtener D_1, D_2, \dots, D_9 respectivamente, se determina el intervalo en el que se va a encontrar el decil correspondiente. Aunque se podrían tomar sus marcas de clase como una primera aproximación de los deciles, éstos se pueden obtener con mayor precisión a partir de la expresión:

$$D_k = L_i + a_i \frac{\frac{kN}{10} - N_{i-1}}{n_i} \quad k \in \{1, 2, \dots, 9\}$$

• **PERCENTILES O CENTILES**

Son 99 valores de la variable, P_1, P_2, \dots, P_{99} , que dividen la población en 100 partes iguales. Entre dos percentiles consecutivos se encuentra 1/100 (1%) de la población.



Cálculo práctico de percentiles:

El procedimiento es análogo al del cálculo de los cuartiles y deciles, tanto en el caso de datos no agrupados como en el caso de datos agrupados en intervalos o clases.

Caso 1: Datos no agrupados

Observamos el primer valor de las frecuencias absolutas acumuladas N_i igual o inferior a $\frac{N}{100}, \frac{2N}{100}, \dots, \frac{99N}{100}$ según se desee obtener P_1, P_2, \dots, P_{99} respectivamente, y se procede de la misma forma que con los cuartiles y deciles.

Ejemplo: En una empresa se lleva a cabo un estudio sobre la calidad de su producción. La distribución siguiente informa sobre el número de piezas defectuosas encontradas en 100 cajas examinadas con 50 unidades cada una de ellas.

Nº piezas defectuosas (x_i)	Nº cajas (n_i)	Ni
0	6	6
1	9	15
2	10	25
3	11	36
4	14	50
5	16	66
6	16	82
7	9	91
8	4	95
9	3	98
10	2	100
	$N = 100$	

- a) Obtén el valor de los cuartiles.
- b) Calcula los deciles 3 y 8.
- c) Obtén los percentiles 27 y 84.

$$Q_1 \Rightarrow \frac{N}{4} = \frac{100}{4} = 25 \Rightarrow \boxed{Q_1 = 2.5}$$

$$Q_2 \Rightarrow \frac{N}{2} = \frac{100}{2} = 50 \Rightarrow \boxed{Q_2 = Me = 4.5}$$

$$Q_3 \Rightarrow \frac{3N}{4} = \frac{300}{4} = 75 \Rightarrow \boxed{Q_3 = 6}$$

$$D_3 \Rightarrow \frac{3N}{10} = \frac{300}{10} = 30 \Rightarrow \boxed{D_3 = 3}$$

$$D_8 \Rightarrow \frac{8N}{10} = \frac{800}{10} = 80 \Rightarrow \boxed{D_8 = 6}$$

$$P_{27} \Rightarrow \frac{27N}{100} = \frac{2700}{100} = 27 \Rightarrow \boxed{P_{27} = 3}$$

$$P_{84} \Rightarrow \frac{84N}{100} = \frac{8400}{100} = 84 \Rightarrow \boxed{P_{84} = 7}$$

Caso 2: Datos agrupados en intervalos o clases

Una vez calculado $\frac{N}{100}, \frac{2N}{100}, \dots, \frac{99N}{100}$ según se desee obtener P_1, P_2, \dots, P_{99} respectivamente, se determina el intervalo en el que se va a encontrar el percentil correspondiente. Aunque se podrían tomar sus marcas de clase como una primera aproximación de los percentiles, éstos se pueden obtener con mayor precisión a partir de la expresión:

$$P_k = L_i + a_i \frac{\frac{kN}{100} - N_{i-1}}{n_i} \quad k \in \{1, 2, \dots, 99\}$$

Ejemplo: En el ejemplo de la estación meteorológica obtener el valor de los cuartiles, del decil 7 y del percentil 34.

Litros/ m ²	x_i	n_i	N_i
[0, 5)	2.5	3	3
[5, 10)	7.5	7	10
[10, 15)	12.5	19	29
[15, 20)	17.5	23	52
[20, 25)	22.5	18	70
[25, 30)	27.5	12	82
[30, 35)	32.5	6	88
$N = 88$			

En primer lugar, hay que obtener el intervalo o clase en el que se encuentra cada uno de los cuantiles que vamos a calcular.

$$Q_1 \Rightarrow \frac{N}{4} = \frac{88}{4} = 22 \Rightarrow I_{Q_1} = [10, 15) \Rightarrow Q_1 = 10 + 5 \frac{22 - 10}{19} \approx 13.16 \Rightarrow \boxed{Q_1 \approx 13.16}$$

$$Q_2 \Rightarrow \frac{N}{2} = \frac{88}{2} = 44 \Rightarrow I_{Q_2} = [15, 20) \Rightarrow Q_2 = 15 + 5 \frac{44 - 29}{23} = 18.26 \Rightarrow \boxed{Q_2 = Me \approx 18.26}$$

$$Q_3 \Rightarrow \frac{3N}{4} = \frac{264}{4} = 66 \Rightarrow I_{Q_3} = [20, 25) \Rightarrow Q_3 = 20 + 5 \frac{66 - 52}{18} \approx 23.89 \Rightarrow \boxed{Q_3 \approx 23.89}$$

$$D_7 \Rightarrow \frac{7N}{10} = \frac{616}{10} = 61.6 \Rightarrow I_{D_7} = [20, 25) \Rightarrow D_7 = 20 + 5 \frac{61.6 - 52}{18} \approx 22.67 \Rightarrow \boxed{D_7 \approx 22.67}$$

$$P_{34} \Rightarrow \frac{34N}{100} = \frac{2992}{100} = 29.92 \Rightarrow I_{P_{34}} = [15, 20) \Rightarrow P_{34} = 15 + 5 \frac{29.92 - 29}{23} = 15.2 \Rightarrow \boxed{P_{34} = 15.2}$$

6. MEDIDAS DE DISPERSIÓN

Permiten conocer el grado de agrupamiento de los datos en torno a las medidas de centralización. Miden, por tanto, la representatividad de las medidas de centralización involucradas en su definición, es decir, cuanto más pequeño sea el valor de la medida de dispersión calculada mayor grado de agrupamiento de los valores de la variable respecto a la medida de centralización considerada (generalmente la media aritmética) y más representativa será ésta como síntesis del comportamiento de la distribución.

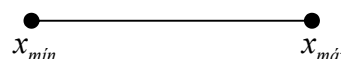
6.1 MEDIDAS DE DISPERSIÓN ABSOLUTAS

Se utilizan en estudios individualizados de distribuciones.

- **RECORRIDO O RANGO**

Se define como la diferencia entre el mayor y el menor valor de la variable:

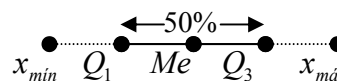
$$R_E = x_{\max} - x_{\min}$$



- **RECORRIDO O RANGO INTERCUARTÍLICO**

Se define como la diferencia entre el tercer y el primer cuartil:

$$R_I = Q_3 - Q_1$$



Nos proporciona la longitud del intervalo en el que se encuentran el 50% de las observaciones centrales.

- **DESVIACIÓN MEDIA RESPECTO A LA MEDIA**

Se define como:

$$D_{\bar{x}} = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot n_i}{N}$$

- **VARIANZA**

Se define como:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot n_i}{N}$$

o bien

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 \cdot n_i}{N} - \bar{x}^2$$

(Compruébalo)

También se representa con S^2 .

- **DESVIACIÓN TÍPICA**

Se define como la raíz cuadrada de la varianza:

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

Es la medida de dispersión más utilizada. Nos proporciona la dispersión de los datos respecto a la media aritmética en su misma dimensión. También se representa con S .

6.2 MEDIDAS DE DISPERSIÓN RELATIVAS

Permiten comparar distintos grados de dispersión de distribuciones heterogéneas.

Se va a estudiar únicamente el coeficiente de variación (de Pearson).

COEFICIENTE DE VARIACIÓN

Se define como:

$$C_v = \frac{\sigma}{\bar{x}} \text{ siempre que } \bar{x} \neq 0. \text{ Se puede dar en porcentaje}$$

Permite comparar dos poblaciones o muestras para determinar en cual de ellas es más representativa su media.

Cuanto menor sea el valor obtenido del coeficiente de variación más concentrados estarán los datos alrededor de la media y por tanto, más representativa será ésta.

Ejemplo: Calcula las medidas de dispersión anteriores en la siguiente distribución de frecuencias:

Otras medidas de dispersión relativas:

- Recorrido relativo

$$R_r = \frac{R_E}{\bar{x}}$$

- Rec. semi-intercuartílico

$$R_{SI} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

- Coeficiente de apertura

$$C_A = \frac{x_n}{x_1}$$

x_i	n_i	N_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$	$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot n_i$
1	7	7	7	7	1.6	11.2
2	14	21	28	56	0.6	8.4
3	9	30	27	81	0.4	3.6
4	8	38	32	128	1.4	11.2
5	2	40	10	50	2.4	4.8
$N = 40$			104	322		39.2

Recorrido: $R_E = x_{máx} - x_{mín} = 5 - 1 = 4$

Recorrido intercuartílico: $R_I = Q_3 - Q_1 = 3.5 - 2 = 1.5$

ya que $\frac{N}{4} = \frac{40}{4} = 10 \Rightarrow Q_1 = 2$; $\frac{3N}{4} = \frac{120}{4} = 30 \Rightarrow Q_3 = \frac{3+4}{2} = 3.5$

Desviación media respecto a la media:

$$D_{\bar{x}} = \frac{\sum_{i=1}^5 |x_i - \bar{x}| \cdot n_i}{N} = \frac{39.2}{40} = 0.98 \quad \text{ya que} \quad \bar{x} = \frac{\sum_{i=1}^5 x_i \cdot n_i}{N} = \frac{104}{40} = 2.6$$

Varianza: $\sigma^2 = \frac{\sum_{i=1}^5 x_i^2 \cdot n_i}{N} - \bar{x}^2 = \frac{322}{40} - 2.6^2 = 8.05 - 6.76 = 1.29$

Desviación típica: $\sigma = \sqrt{\sigma^2} = \sqrt{1.29} \approx 1.14$

Coeficiente de variación: $C_v = \frac{\sigma}{\bar{x}} = \frac{\sqrt{1.29}}{2.6} \approx 0.44 \Rightarrow 44\%$

Si la distribución se agrupa en intervalos se hará igual pero tomando las marcas de clase.

Ejemplo: En unas caballerizas los pesos de los caballos se distribuyen con una media de $\bar{x} = 400$ kg y una desviación típica $\sigma = 30$ kg. En una tienda de mascotas el peso de los perros tienen una media de $\bar{x} = 22$ kg y una desviación típica $\sigma = 12$ kg.

Compara la dispersión en términos absolutos y relativos de los pesos de estos animales.

Caballerizas: $\left. \begin{matrix} \bar{x} = 400 \text{ kg} \\ \sigma = 30 \text{ kg} \end{matrix} \right\} \Rightarrow C_v = \frac{\sigma}{\bar{x}} = \frac{30}{400} = 0.075 \Rightarrow 7.5\%$

Tienda de mascotas: $\left. \begin{matrix} \bar{x} = 22 \text{ kg} \\ \sigma = 12 \text{ kg} \end{matrix} \right\} \Rightarrow C_v = \frac{\sigma}{\bar{x}} = \frac{12}{22} \approx 0.545 \Rightarrow 54.5\%$

La desviación típica de los pesos de los caballos es mayor que la de los perros. Sin embargo, la diferencia no es muy notable en el caso de los equinos pero sí lo es en el de los canes, al contrario de lo que puede parecer a primera vista si observamos los datos absolutos. Los caballos tienen pesos similares, mientras que los perros son muy heterogéneos.

El cálculo del coeficiente de variación nos permite confirmar nuestras conclusiones anteriores y además cuantifica esta situación: el peso de los caballos varía en un 7.5% mientras que el de los chuchos lo hace en un 54.5%.

7. DIAGRAMA DE TALLOS Y HOJAS. DIAGRAMA DE CAJAS Y BIGOTES

7.1 DIAGRAMA DE TALLOS Y HOJAS

Es una forma rápida de resumir los datos de una distribución estadística y mantener a la vista todas las observaciones individuales.

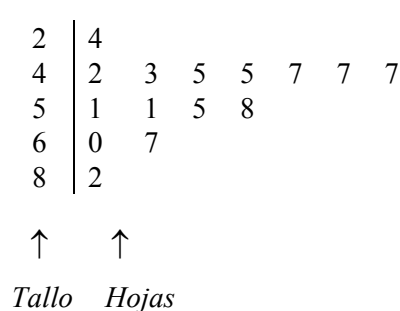
Ejemplo: El número de corbatas que se han vendido en unos grandes almacenes en 15 días son:

45 47 24 42 43 51 58 60 45 47 47 67 82 51 55

Se ordenan de menor a mayor:

24 42 43 45 45 47 47 47 51 51 55 58 60 67 82

Se forma el *diagrama de tallos y hojas*:



Número corbatas (x_i)	n_i	N_i
24	1	1
42	1	2
43	1	3
45	2	5
47	3	8
51	2	10
55	1	11
58	1	12
60	1	13
67	1	14
82	1	15
$N = 15$		

Podemos obtener la tabla estadística de frecuencias a partir del diagrama de tallos y hojas:

También se puede formar primero el diagrama de tallos y hojas y, posteriormente, ordenarlos.

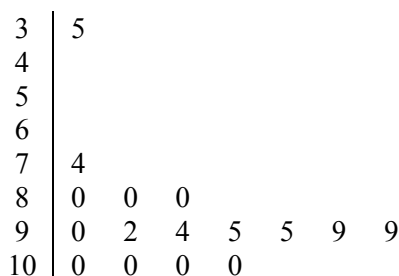
Ejemplo: Los resultados obtenidos en Matemáticas II de la Prueba de Acceso a la Universidad en un determinado centro en el curso 2015/2016 fueron:

10 8 9.5 8 9.9 10 9.9 9 10 9.4 7.4 9.5 9.2 8 3.5 10

Se ordenan de menor a mayor:

3.5 7.4 8 8 8 9 9.2 9.4 9.5 9.5 9.9 9.9 10 10 10 10

Formamos el *diagrama de tallos y hojas* y la tabla estadística de frecuencias:



Puntuación (x_i)	Nº alumnos (n_i)	N_i
[0, 1)	0	0
[1, 2)	0	0
[2, 3)	0	0
[3, 4)	1	1
[4, 5)	0	1
[5, 6)	0	1
[6, 7)	0	1
[7, 8)	1	2
[8, 9)	3	5
[9, 10]	11	16
$N = 16$		

En este caso se han agrupado los datos de la tabla en intervalos.

7.2 DIAGRAMA DE CAJAS Y BIGOTES (Box and Whisker)

Se usa como representación gráfica de variables cuantitativas.

Sintetiza la información de una distribución dividiéndola en cuatro partes con igual número de datos y describe gráficamente tanto la dispersión como la simetría de la distribución.

Los cuartiles marcan los puntos de división. Las *cajas* encierran los datos comprendidos entre cuartiles y los *bigotes* los datos que están por debajo y por encima del primer y tercer cuartil respectivamente.

Ejemplo: Obtener el diagrama de cajas y bigotes de la distribución:

x_i	n_i	N_i
2	3	3
4	5	8
6	7	15
7	4	19
9	2	21
$N = 21$		

Se calculan $Q_1, Me, Q_3, x_{min}, x_{max}$

$$x_{min} = 2; x_{max} = 9$$

$$Q_1; \frac{N}{4} = \frac{21}{4} = 5.25 \Rightarrow Q_1 = 4$$

$$Q_2 = Me; \frac{N}{2} = \frac{21}{2} = 10.5 \Rightarrow Q_2 = 6$$

$$Q_3; \frac{3N}{4} = \frac{63}{4} = 15.75 \Rightarrow Q_3 = 7$$

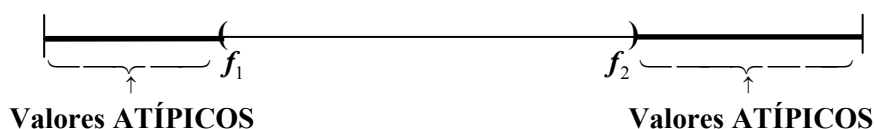
Se calcula el rango intercuartílico:

$$R_I = Q_3 - Q_1 = 7 - 4 = 3$$

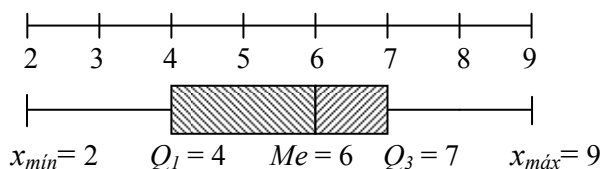
y se obtienen los posibles *valores atípicos*. Para ello se calculan los límites internos del conjunto de datos:

$$\left. \begin{aligned} f_1 &= Q_1 - 1.5R_I = 4 - 1.5 \cdot 3 = -0.5 \\ f_2 &= Q_3 + 1.5R_I = 7 + 1.5 \cdot 3 = 11.5 \end{aligned} \right\} \Rightarrow (f_1, f_2) = (-0.5, 11.5)$$

Nos indican que todos los valores que tome la variable estadística y que sean inferiores a f_1 y superiores a f_2 , es decir exteriores al intervalo (f_1, f_2) , se considerarán atípicos (anómalos).



En este caso **no hay valores atípicos** por lo que el diagrama de cajas y bigotes queda así:



Valores atípicos son aquellos que se alejan significativamente del resto de las observaciones de una distribución de frecuencias. Puede ocurrir por diversas causas: error en la toma de datos, suceso extraordinario... Lo representaremos con \circ .

En nuestro ejemplo la variable no toma ningún valor atípico por lo que los extremos de los bigotes van de la caja a los valores x_{min} y x_{max} . En caso contrario, los bigotes finalizarán en el menor y/o mayor valor observados respectivamente del intervalo (f_1, f_2) .

Un valor atípico suele ser **leve (moderado)**, pero si excede de $Q_1 - 3R_I$ o $Q_3 + 3R_I$, llamados límites externos, se considerará un **valor atípico extremo** y lo representaremos con $*$.

Veamos un ejemplo en el que aparecen valores atípicos:

Ejemplo: Obtener el diagrama de cajas y bigotes de la distribución:

x_i	n_i	N_i
45	1	1
51	2	3
57	1	4
58	2	6
61	1	7
65	1	8
67	2	10
71	2	12
73	1	13
84	1	14
98	1	15
$N = 15$		

Se calculan $Q_1, Me, Q_3, x_{min}, x_{max}$

$$x_{min} = 45;$$

$$x_{max} = 98$$

$$Q_1; \frac{N}{4} = \frac{15}{4} = 3.75 \Rightarrow Q_1 = 57$$

$$Q_2 = Me; \frac{N}{2} = \frac{15}{2} = 7.5 \Rightarrow Q_2 = 65$$

$$Q_3; \frac{3N}{4} = \frac{3 \cdot 15}{4} = 11.25 \Rightarrow Q_3 = 71$$

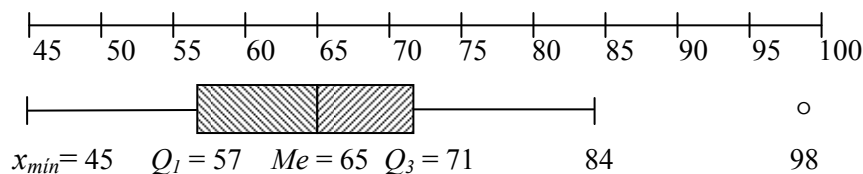
$$\text{Además: } R_I = Q_3 - Q_1 = 71 - 57 = 14$$

Se obtienen los posibles valores atípicos:

$$\left. \begin{aligned} f_1 &= Q_1 - 1.5R_I = 57 - 1.5 \cdot 14 = 36 \\ f_2 &= Q_3 + 1.5R_I = 71 + 1.5 \cdot 14 = 92 \end{aligned} \right\} \Rightarrow (f_1, f_2) = (36, 92)$$

Como $98 \notin (f_1, f_2) = (36, 92) \Rightarrow 98$ es un valor atípico

Por tanto, el diagrama de cajas y bigotes queda así:



El valor atípico obtenido es leve. Para comprobar que no es un valor atípico extremo obtenemos:

$$\left. \begin{aligned} Q_1 - 3R_I &= 57 - 3 \cdot 14 = 15 \\ Q_3 + 3R_I &= 71 + 3 \cdot 14 = 113 \end{aligned} \right\} \Rightarrow (15, 113)$$

Como $98 \in (15, 113) \Rightarrow 98$ es un valor atípico leve.

Fíjate:

Hubiese bastado con calcular $Q_3 + 3R_I = 71 + 3 \cdot 14 = 113$ y comprobar que $98 < 113$.

Recuerda que en el caso de haber sido un valor atípico extremo se hubiera representado con un asterisco *.

8. CARACTERÍSTICAS DE FORMA

Tienen como objetivo describir numéricamente la forma de una distribución sin tener que llevar a cabo su representación gráfica. Las medidas de forma van a ser unos coeficientes adimensionales que van a cuantificar diversas propiedades de una distribución de frecuencias. En particular vamos a evaluar dos propiedades que van a ser la simetría de una distribución y el grado de concentración de frecuencias en la zona central de una distribución (apuntamiento o curtosis).

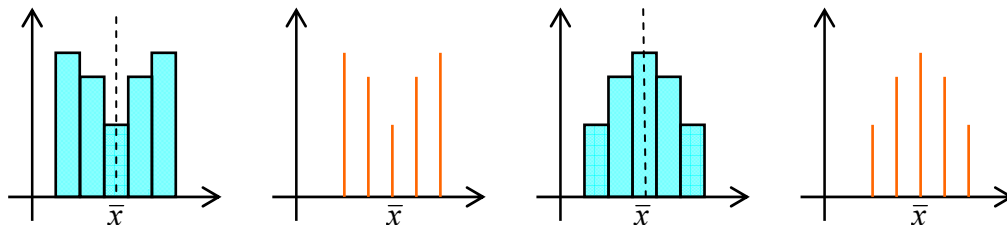
8.1 MEDIDAS DE SIMETRÍA

Tienen por objetivo detectar si una distribución es simétrica o no, y en caso de no serlo cuantificarán el grado de asimetría detectando también si ésta es a la derecha o positiva, o a la izquierda también llamada negativa.

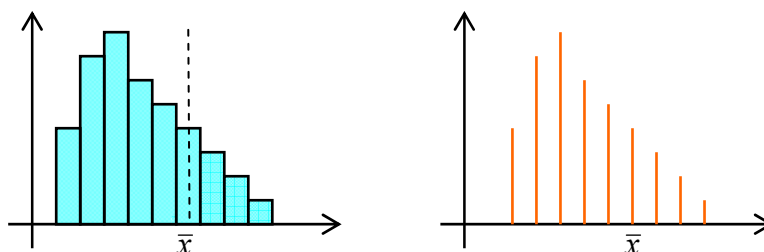
En general, la simetría se va a considerar respecto a una medida de tendencia central que generalmente va a ser la media aritmética.

Dada una distribución de frecuencias, si trazamos una perpendicular al eje de abscisas por su media aritmética, decimos que es:

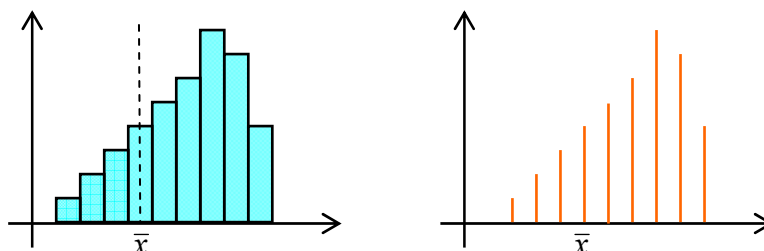
- **Simétrica (o insesgada)** respecto a esta recta como eje de simetría, si existe el mismo número de valores a ambos lados de la media aritmética, si dos a dos de los valores situados a ambos lados son equidistantes y además valores equidistantes presentan la misma frecuencia.



- **Asimétrica a la derecha o positiva (o sesgada a la derecha)** si su media aritmética está desplazada a la derecha.



- **Asimétrica a la izquierda o negativa (o sesgada a la izquierda)** si su media aritmética está desplazada a la izquierda.



Coefficiente de asimetría de Fisher:

$$\gamma_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 \cdot n_i}{N \cdot \sigma^3}$$

Interpretación:

- Si $\gamma_1 = 0$ la distribución es simétrica (insesgada).
- Si $\gamma_1 > 0$ la distribución es asimétrica (sesgada) a la derecha.
- Si $\gamma_1 < 0$ la distribución es asimétrica (sesgada) a la izquierda.

Coefficiente de asimetría de Pearson:

Menos general y usado que el de Fisher, se podrá utilizar en distribuciones unimodales, campaniformes y no muy asimétricas en las que sabemos que:

- Si es asimétrica a la derecha $\rightarrow Mo \leq Me \leq \bar{x}$
- Si es asimétrica a la izquierda $\rightarrow \bar{x} \leq Me \leq Mo$
- Si es simétrica $\rightarrow \bar{x} = Me = Mo$

Teniendo en cuenta lo anterior, se define el **coeficiente de asimetría de Pearson** como:

$$C_p = \frac{\bar{x} - Mo}{\sigma}$$

Interpretación:

- Si $C_p = 0$ la distribución es simétrica (insesgada).
- Si $C_p > 0$ la distribución es asimétrica (sesgada) a la derecha.
- Si $C_p < 0$ la distribución es asimétrica (sesgada) a la izquierda.

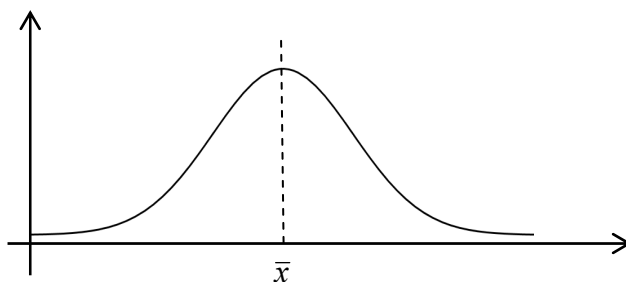
Observación: Si no se conoce la moda de la distribución, en la práctica se puede usar la expresión:

$$C_p = \frac{\bar{x} - Mo}{\sigma} \approx \frac{3(\bar{x} - Me)}{\sigma}$$

8.2 MEDIDA DE CURTOSIS

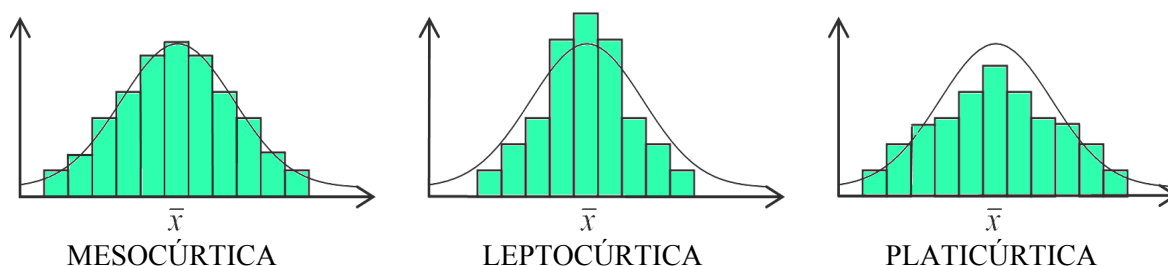
Solo tendrá sentido en distribuciones que sean unimodales, campaniformes y simétricas o ligeramente asimétricas. Tienen como objetivo evaluar la concentración de frecuencias que presenta una distribución en su zona central.

Para ello, se considera como modelo la Curva Normal, también llamada Campana de Gauss.



Se caracteriza por ser simétrica en torno a su media, luego ésta coincidirá con la moda que es única, y campaniforme.

Si la distribución presenta el mismo grado, mayor o menor de concentración de frecuencias en su zona central que la curva normal, recibe el nombre de mesocúrtica, leptocúrtica o platicúrtica respectivamente.



Se define el **coeficiente de curtosis o apuntamiento** como:

$$\gamma_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4 \cdot n_i}{N \cdot \sigma^4} - 3$$

Interpretación:

- Si $\gamma_2 = 0$ la distribución es mesocúrtica (igual de apuntada que la normal).
- Si $\gamma_2 > 0$ la distribución es leptocúrtica (más apuntada que la normal).
- Si $\gamma_2 < 0$ la distribución es platicúrtica (menos apuntada que la normal).

Nota: Se toma como referencia la distribución Normal ya que tiene un coeficiente de curtosis $\gamma_2 = 0$.

Ejemplo: Obtén el coeficiente de asimetría de Fisher y el coeficiente de curtosis de esta tabla de datos:

x_i	n_i	N_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^3 \cdot n_i$	$(x_i - \bar{x})^4 \cdot n_i$
1	7	7	7	7	-1.725	-35.93067	61.98041
2	9	16	18	36	-0.725	-3.42970	2.48654
3	14	30	42	126	0.275	0.29116	0.08007
4	8	38	32	128	1.275	16.58138	21.14125
5	2	40	10	50	2.275	23.54909	53.57419
$N = 40$			109	347		1.06126	139.26246

$$\text{Media aritmética: } \bar{x} = \frac{\sum_{i=1}^5 x_i \cdot n_i}{N} = \frac{109}{40} = 2.725$$

$$\text{Varianza: } \sigma^2 = \frac{\sum_{i=1}^5 x_i^2 \cdot n_i}{N} - \bar{x}^2 = \frac{347}{40} - 2.725^2 = 8.675 - 7.4375 = 1.2375 \approx 1.25$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} \approx 1.12$$

Coeficiente de asimetría de Fisher:

$$\gamma_1 = \frac{\sum_{i=1}^5 (x_i - \bar{x})^3 \cdot n_i}{N \cdot \sigma^3} = \frac{1.06126}{40 \cdot 1.12^3} \approx 0.0189 > 0 \Rightarrow \text{La distribución es asimétrica a la derecha}$$

Coeficiente de curtosis:

$$\gamma_2 = \frac{\sum_{i=1}^5 (x_i - \bar{x})^4 \cdot n_i}{N \cdot \sigma^4} - 3 = \frac{139.26246}{40 \cdot 1.12^4} - 3 \approx -0.7874 < 0 \Rightarrow \text{La distribución es platicúrtica}$$

9. APÉNDICE

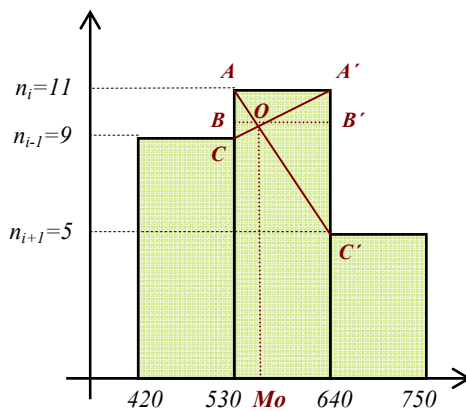
9.1 MÉTODO GRÁFICO DE CÁLCULO DE LA MODA PARA DATOS AGRUPADOS EN INTERVALOS O CLASES (SEMEJANZA)

Ejemplo: La duración en horas de 30 bombillas viene dada por la tabla:

Número horas	x_i	n_i
[310, 420)	365	1
[420, 530)	475	9
[530, 640)	585	<u>11</u>
[640, 750)	695	5
[750, 860)	805	3
[860, 970)	915	1
		$N = 30$

Intervalo Modal: $I_{Mo} = [530, 640)$

Utilizando el histograma de frecuencias no acumuladas:



Por semejanza de triángulos:

$$\hat{AOC} \sim \hat{A'OC'} \Rightarrow \frac{AC}{OB} = \frac{A'C'}{OB'} = \frac{AC + A'C'}{OB + OB'}$$

Por tanto:

$$\frac{11 - 9}{Mo - 530} = \frac{(11 - 9) + (11 - 5)}{640 - 530} \Rightarrow \frac{2}{Mo - 530} = \frac{8}{110} \Rightarrow Mo = 557.5 \text{ horas}$$

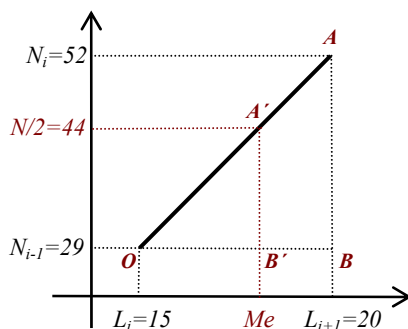
9.2 MÉTODO GRÁFICO DE CÁLCULO DE LA MEDIANA PARA DATOS AGRUPADOS EN INTERVALOS O CLASES (SEMEJANZA)

Ejemplo: Una estación meteorológica registró 88 días de lluvia el pasado año, según se muestra en la tabla siguiente:

Litros/ m ²	x_i	n_i	N_i
[0, 5)	2.5	3	3
[5, 10)	7.5	7	10
[10, 15)	12.5	19	29
[15, 20)	17.5	23	<u>52</u>
[20, 25)	22.5	18	70
[25, 30)	27.5	12	82
[30, 35)	32.5	6	88
		$N = 88$	

Obtención del intervalo mediano:

$$\frac{N}{2} = \frac{88}{2} = 44 \Rightarrow I_{Me} = [15, 20)$$



Por semejanza de triángulos

$$\hat{AOB} \sim \hat{A'OB'} \Rightarrow \frac{AB}{A'B'} = \frac{OB}{OB'}$$

Por tanto:

$$\frac{52 - 29}{44 - 29} = \frac{20 - 15}{Me - 15} \Rightarrow \frac{23}{15} = \frac{5}{Me - 15} \Rightarrow Me \approx 18,26 \text{ litros/m}^2$$

Observación: Este procedimiento puede usarse de forma análoga para el cálculo de cuantiles (cuartiles, deciles o percentiles).