

Objetivos

En esta quincena aprenderás a:

- Distinguir los distintos tipos de variables estadísticas.
- Agrupar en intervalos los datos de un estudio estadístico.
- Hacer la tabla estadística asociada a un conjunto de datos.
- Representar e interpretar gráficos estadísticos, y saber cuando es conveniente utilizar cada tipo.
- Calcular la media, la moda, la mediana y los cuartiles de un conjunto de datos.
- Que son y cómo se calculan los parámetros de dispersión: el rango o recorrido, la varianza y la desviación típica, el coeficiente de variación.

1. Hacer estadística	pág. 4
Necesidad	
Población y muestra	
Variables	
2. Recuento y gráficos	pág. 5
Recuento de datos	
Gráficos	
Agrupación de datos en intervalos	
3. Medidas de centralización	pág. 9
y posición	
Medida	
Moda	
Cuartiles y mediana	
Diagramas de caja y bigotes	
4. Medidas de dispersión	pág. 12
Rango y desviación media	
Desviación típica	
Coeficiente de variación	

Ejercicios para practicar

Para saber más

Resumen

Autoevaluación

Anexo

Antes de empezar

Estadística

La estadística, a nivel primario, es una actividad que todo el mundo hace desde muy pequeño. El mero hecho de contar y/o clasificar tus juguetes (tus coches, muñecas/os, canicas, videojuegos,...) ya es una actividad estadística.

Clasificar objetos

El destornillador a la caja de herramientas, los cubiertos al cajón de la cocina, los libros a la estantería, los videojuegos junto a la consola,...

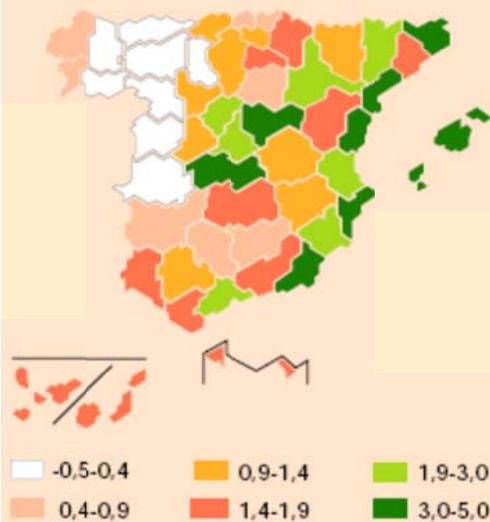


Competiciones escolares

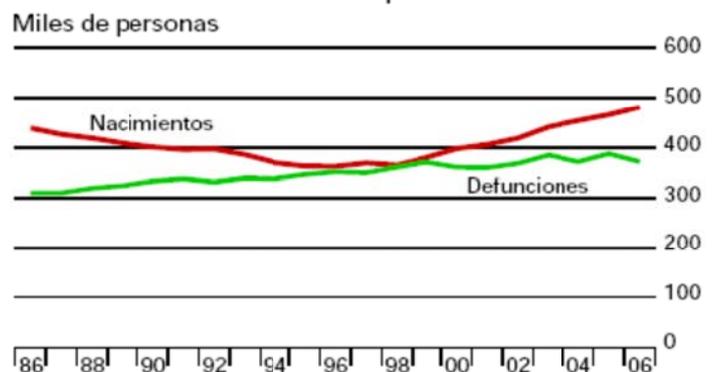


En tu Instituto habrá equipos y competiciones, y habrá que llevar un cómputo de resultados y una ordenación de equipos según una puntuaciones. Hasta es posible que haya un registro de varios años.

Crecimiento relativo de la población total
1 de enero de 2008 (%)



Crecimiento natural de la población



1. Hacer estadística

Necesidad

Al poner en práctica una medida social para saber su aceptación. ¿A cuántas personas puede ir dirigida?, ¿cuáles son los distintos niveles?. Frente a una iniciativa como esta, preguntar a toda la población puede agotar los recursos destinados a ella, una encuesta previa puede ahorrarnos algún que otro equívoco.

Población y muestra

Cuando se hace un estudio estadístico el investigador decide si analizará toda la población o una muestra elegida previamente.

Población es el conjunto de individuos, con alguna característica común, sobre el que se hace un estudio estadístico.

La **muestra** es un subconjunto de la población. Debe elegirse que sea representativa de toda la población en la característica estudiada.

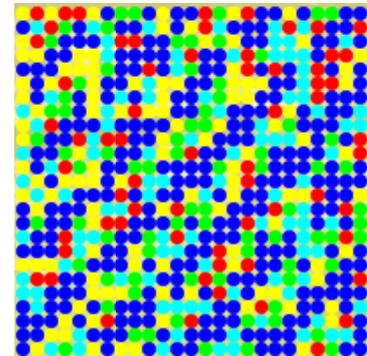


Atributos y Variables.

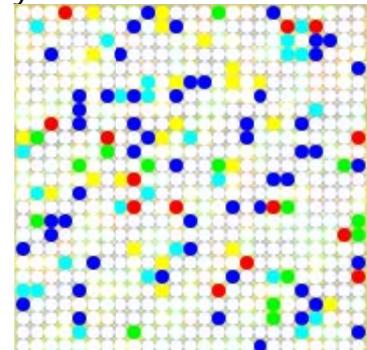
Cada una de las propiedades o características que podemos estudiar es una **variable estadística**. Dependiendo de los posibles valores que puedan tomar se clasifican en:

- **Variables cualitativas** o atributos. Los valores de la variable no son números sino cualidades, se expresan con palabras. El color, la forma, el sexo,...son ejemplos de variables cualitativas.
- **Variables cuantitativas**. Los datos se expresan numéricamente y pueden ser:
 - Discretas. Cada una de las variables solo puede tomar valores enteros (1, 2, 3...). El nº de hermanos, el nº ventanas de casa, el nº colegios de tu población,...
 - Continuas. Pueden tomar cualquier valor de un intervalo dado. Nuestro peso, altura, fuerza, no es posible medirlas con números enteros, la densidad del aire, la velocidad media de los fórmula 1 en una carrera,...

Población



y muestra



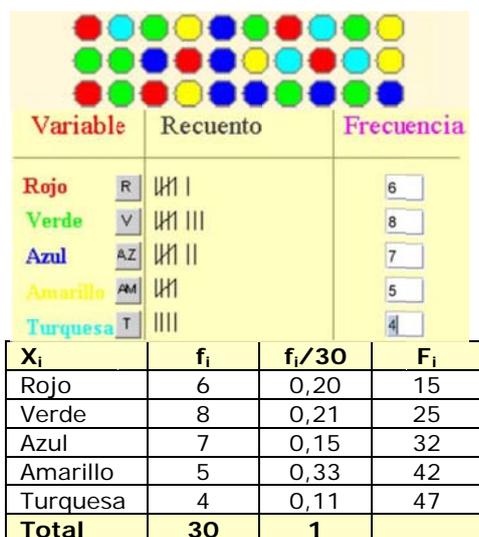
El color de los lápices, es una variable cualitativa



La altura, edad y peso, son variables cuantitativas.



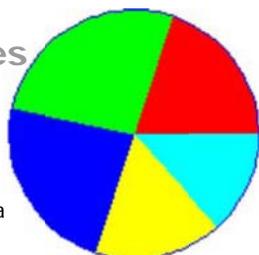
2. Recuento y gráficos



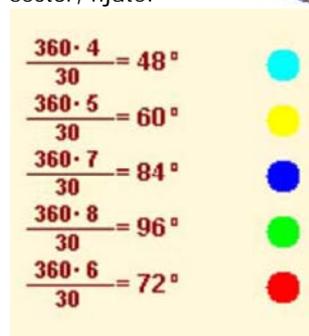
Es parte del proceso, después de recopilar los datos se procede a su recuento para expresarlos de forma ordenada y para que sea más fácil trabajar con ellos. Generalmente se elabora una tabla como se muestra a la izquierda donde puedes practicar.

- Frecuencia **absoluta**, es el nº de veces que aparece un dato. A la de x_i la llamaremos f_i .
- Frecuencia **relativa**, es el cociente entre la frecuencia absoluta y el nº total de datos.
- Frecuencia **acumulada** de un dato, es la suma de las frecuencias absolutas de los valores que son menores o iguales que él, la indicaremos con F_i . También se pueden calcular las frecuencias relativas acumuladas.

Diagrama de sectores



Para calcular los grados de cada sector, fijate:



Diagramas de barras y de sectores

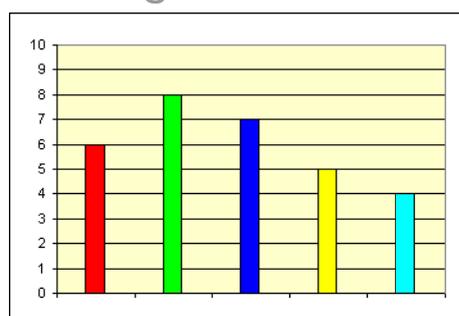
Los datos estadísticos suelen representarse de forma gráfica, ya que de esta forma podemos hacernos una idea de su distribución de un solo golpe de vista. En función del tipo de variable conviene más usar un tipo de gráfico u otro.

- **Diagrama de sectores**, puede aplicarse a cualquier tipo de variable, aunque es el más adecuado en variables cualitativas y para una primera toma de contacto con los valores de una población. Es un círculo dividido en sectores de ángulo proporcional a la frecuencia de cada valor.

La amplitud de cada sector se obtiene multiplicando la frecuencia relativa por 360° .

$$\frac{\text{frecuencia}}{\text{n}^\circ \text{ total de datos}} = \frac{\text{grados del sector}}{360}$$

Diagrama de barras



- **Diagrama de barras**. También puede aplicarse a cualquier tipo de variable, aunque se considera el idóneo para variables discretas. Cada valor se corresponde con una barra de longitud proporcional a su frecuencia.

EJERCICIOS resueltos

1. ¿Cuántas personas suponen una muestra del 10% de una población de 10.000 habitantes? ¿Y de una de 6000 habitantes?.

Solución: a) $10.000 \cdot 10 / 100 = 1000$,

b) $6000 \cdot 10 / 100 = 600$

2. Una empresa de sondeos estadísticos tiene capacidad para entrevistar a 1000 personas por semana. Si dispone de 4 semanas a qué porcentaje de una población de 100.000 habitantes puede entrevistar para obtener una muestra.

Solución: En 4 semanas puede entrevistar a 4000 personas. 4000 de 100.000 equivale a 4 de 100. Así pues el 4%.

3. Con el fin de conocer mejor la forma de viajar de una población han preparado una encuesta. Algunas de las preguntas trataron sobre: N° de días de viaje, dinero empleado, número de bultos, zonas geográficas, medio de transporte, naturaleza del viaje (negocios, turismo, familiar, salud...) y n° de personas. Clasifica estas variables estadísticas.

Solución:

V cualitativa: Zonas geográficas, medio de transporte y naturaleza del viaje.

V. cuantitativa discreta: N° de días, número de bultos y n° de personas.

V. cuantitativa continua: Dinero empleado.

4. Haz un recuento de los siguientes datos

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

Solución:

X_i	f_i
1	3
2	8
3	4
4	9

5. Haz un recuento de los siguientes datos, un gráfico de sectores y otro de barras. Indica el ángulo de cada sector.

Pelota, máscara, pelota, máscara, máscara, bici, máscara, bici, bici, máscara, máscara, máscara, máscara, videojuego, máscara, pelota, videojuego, pelota, videojuego, pelota, pelota, videojuego, pelota, máscara.

Solución:

X_i	f_i	grados
Videojuego	4	60
Máscara	3	45
Bici	10	150
Pelota	7	105



Los datos

55	491	42	465	653
829	798	254	155	427
153	533	945	878	230
690	652	476	110	87
14	751	47	211	341
737	473	452	352	246
499	109	694	308	933
326	209	729	651	397
161	329	975	848	823
240	640	319	526	

Agrupados en 5 intervalos
 Observa las marcas de cada clase como se corresponden con la media de sus extremos.

Intervalo	Marca	Frecuencia
[0 , 200)	100	10
[200 , 400)	300	13
[400 , 600)	500	9
[600 , 800)	700	10
[800 , 1000)	900	7

Agrupados ahora en 8 intervalos

Intervalo	Marca	Fr.
[0 , 125)	62,5	7
[125 , 250)	187,5	8
[250 , 375)	312,5	7
[375 , 500)	437,5	8
[500 , 625)	562,5	2
[625 , 750)	687,5	8
[750 , 875)	812,5	5
[875 , 1000)	937,5	4

Agrupación de datos en intervalos

En variables continuas, o en discretas cuando el número de datos distintos se hace casi tan grande como el número de datos, y para poder estudiarlos, se hace necesario agruparlos en **intervalos** o **clases**, habitualmente de la misma amplitud y como mínimo cuatro.

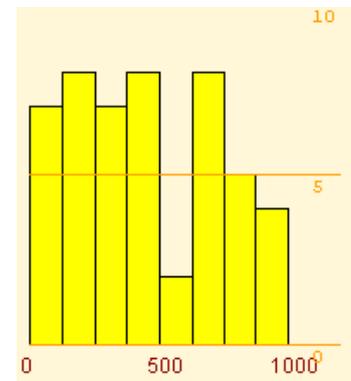
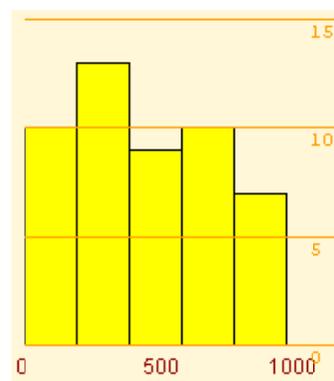
Por ejemplo, en una población hay casi tantas alturas como individuos pero podemos agruparlos en bajos, medios y altos; también podríamos hacer bajos, medios-bajos, medios-altos y altos, o clasificarlos de 10 en 10 cm, o de 20 en 20...

- Para representar a todos los datos de un intervalo elegimos un valor, el punto medio del intervalo, se llama **marca de clase**.

Histograma

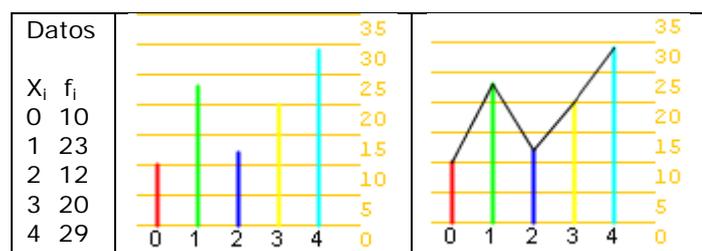
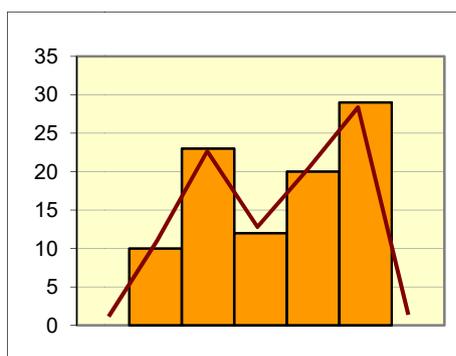
Cuando los datos vienen agrupados en intervalos se usa para representarlos gráficamente el **histograma**. Cada valor se representa con un rectángulo de anchura el intervalo correspondiente y con la altura proporcional a su frecuencia.

Los histogramas para los datos del margen agrupados en cinco y ocho intervalos:



Polígono de frecuencias.

Lo creamos al unir los extremos superiores de las barras de los histogramas o de los diagramas de barras.

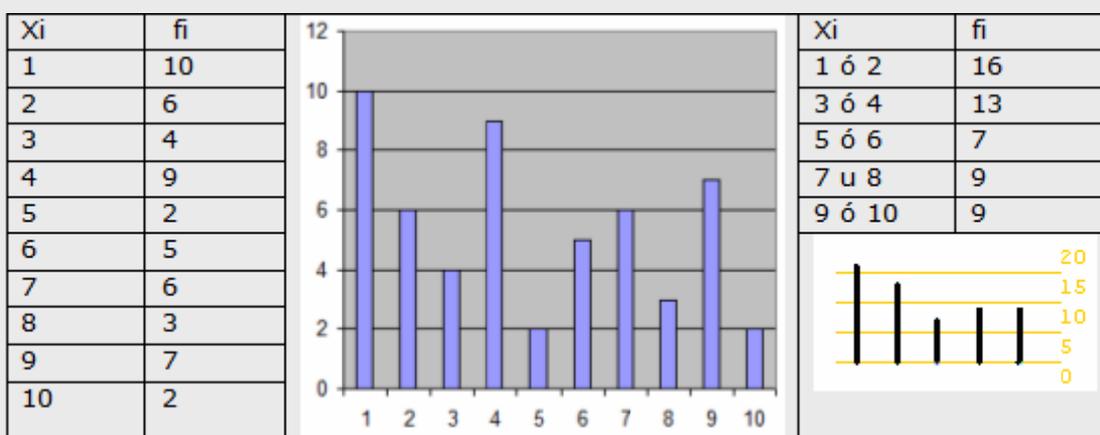


EJERCICIOS resueltos

6. Agrupa los siguientes datos en 10 grupos. Agrupa los mismos datos, ahora, en 5 grupos y haz un gráfico para cada agrupación.

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

Solución:

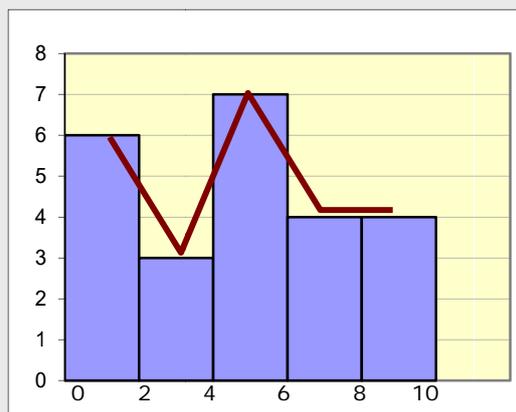


7. Agrupa los datos siguientes en 5 intervalos de igual amplitud, haz un gráfico y un polígono de frecuencias.

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

Solución:

	f_i
[0,2)	6
[2,4)	3
[4,6)	7
[6,8)	4
[8,10)	4



Media

Ejemplo 1

10, 12, 10, 14 y 13

$$\bar{x} = \frac{10+12+10+14+13}{5} = \frac{59}{5} = 11.8$$

Ejemplo 2

X_i	f_i	$X_i \cdot f_i$
5	4	20
10	6	60
15	7	105
20	9	180
25	4	100
30	6	180
36	645	

La media.
 $\bar{x} = \frac{645}{36} = 17,91$

Moda

Ejemplo 1

X_i	f_i	Moda
5	8	5
10	5	
15	1	
20	8	20
25	0	
30	3	

Ejemplo 2

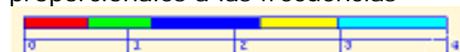
X_i	f_i	Moda
100	5	
200	5	
300	1	
400	3	
500	2	
600	9	600

Medidas de posición

Ampliamos la tabla con la columna de frecuencias acumuladas y porcentajes de estas

x_i	f_i	F_i	%
1	10	10	15,152
2	10	20	30,303
3	17	37	56,061
4	12	49	74,242
5	17	66	100

y/o una barra con longitudes proporcionales a las frecuencias



y podremos saber la mediana y cuartiles

$$Q_1=2, Q_2=Me=3 \text{ y } Q_3=5$$

3. Medidas de centralización y posición

La media

Todos los alumnos saben que con un 6 y un 4 tienen de media 5. Pues la media en estadística no es otra cosa que eso, solo que, habitualmente, con más datos.

Para calcular la media si son pocos los datos, se suman todos y se divide entre el número total. Si son muchos, los tendremos agrupados, entonces se suman los productos de cada dato por su frecuencia absoluta y se divide esta suma por el número total de datos. Se indica con \bar{x} .

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum x_i f_i}{N}$$

La moda

¿Quién no ha oído alguna vez: "Está de moda ir a...", "Se lleva este tipo de pantalón, está de moda", o "Se ha puesto de moda el grupo"... y todo el mundo entiende que hay una buena cantidad de personas en esas opciones.

Así pues, el valor que más frecuencia tenga será "el de moda", aunque puede ocurrir que haya más de uno.

- La **moda**, **Mo**, de una distribución estadística es el valor de la variable que más se repite, el de mayor frecuencia absoluta.

La mediana y los cuartiles

La mediana y los cuartiles, como la media aritmética, sólo se pueden calcular cuando la variable es cuantitativa.

- La **mediana**, **Me**, es el valor que ocupa la posición central una vez ordenados los datos en orden creciente, es decir el valor que es mayor que el 50% y menor que el otro 50%.

La mediana divide la distribución en dos partes con igual nº de datos, si la dividimos en cuatro partes obtenemos los **cuartiles**, 1º, 2º y 3º, que se indican respectivamente **Q₁**, **Q₂** y **Q₃**.

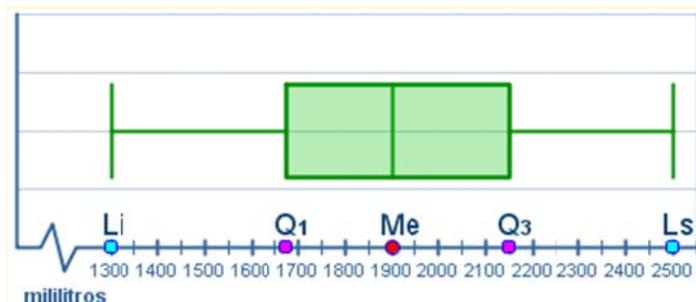
Ordenados los datos, **el primer cuartil**, es mayor que el 25% de estos; **el tercer cuartil**, mayor que el 75%, y el segundo coincide con la mediana.

Diagramas de caja y bigotes

A partir del valor de la mediana y los cuartiles se pueden representar las distribuciones estadísticas mediante los llamados "diagramas de caja y bigotes".

Veamos como se construye con los datos de la tabla de la derecha. Una vez ordenados los datos, se calculan los valores mínimo y máximo, los cuartiles y la mediana.

$$\text{mín}=1300 \quad Q_1=1675 \quad \text{Me}=1900 \quad Q_3=2150 \quad \text{máx}=2500$$



Se sitúan estos valores sobre el eje de abscisas y se dibuja la "caja" desde el primer al tercer cuartil (el recorrido *intercuartílico*), y los "bigotes" como indica la figura.

La tabla muestra el consumo diario de agua, en ml, de los 20 alumnos de una clase.

Juan	1650	Luis	1300	Mín
Luis	1300	Tere	1500	
Alma	2400	Maya	1600	
Toño	2000	Marta	1650	
Rosa	2100	Juan	1650	Q ₁
Lupe	1700	Lupe	1700	
Paco	1900	David	1750	
Tere	1500	Pepe	1850	
Iris	1900	Alex	1900	Me
Pepe	1850	Iris	1900	
Marco	2000	Paco	1900	
Lisa	2200	Marco	2000	
Julio	2300	Toño	2000	
Maya	1600	Omar	2100	
Alex	1900	Rosa	2100	Q ₃
Beto	2500	Lisa	2200	
Rita	2200	Rita	2200	
Marta	1650	Julio	2300	
Omar	2100	Alma	2300	
David	1750	Beto	2500	Máx

NOTA: La longitud de los bigotes no debe exceder una vez y media la de la caja, si hay valores extremos que superan esa medida se dibujan como puntos aislados.

EJERCICIOS resueltos

8. Calcula la media en cada caso:

- a) 4, 6, 8
b) 4, 6, 8, 6
c) 100, 120, 180, 200

Soluciones: a) $(4+6+8)/3 = 6$
b) $(4+6+8+6)/4 = 24/4 = 6$
c) $(100+120+180+200)/4 = 150$

9. Calcula la media de los siguientes datos

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

Solución:

Sumamos todo y dividimos entre 24, el número de datos. $\bar{X} = 2,29$

10. Calcula la media de los siguientes datos

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

Solución: $\bar{X} = 2,19$

11. Determina la moda para los datos

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

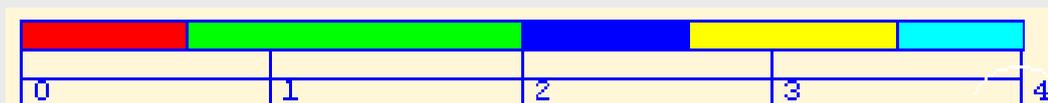
Solución:

Mo = 1, es el valor que mas veces aparece, 8 en total.

12. Calcula la mediana, el primer y el segundo cuartil de los datos del ejercicio anterior.

Solución:

Hacemos el recuento, 0→4, 1→8, 2→4, 3→5 y 4→3, dibujamos barras de colores de longitudes proporcionales a las frecuencias, valdrían por ejemplo de 4mm, 8mm, 4mm, 5mm y 3mm. Dividimos toda la barra en 4 partes y nos fijamos el color en el que queda la división.

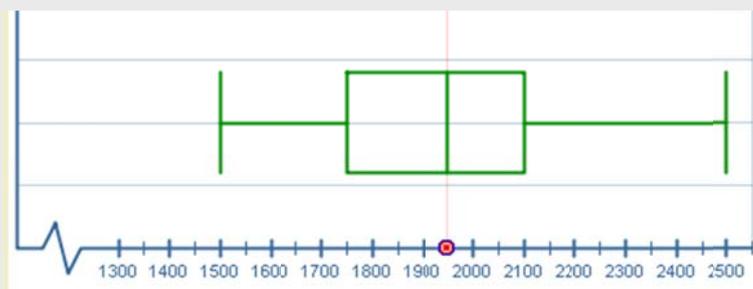


Mediana = 1,5,
Q1=1 y Q3=3

Ó bien construimos la tabla y observamos donde quedan en la última columna los valores 25%, 50% y 75%.

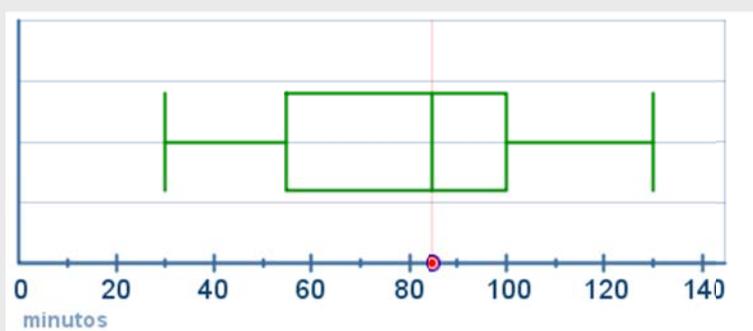
X_i	f_i	F_i	%
0	4	4	16,667
1	8	12	50
2	4	16	66,667
3	5	21	87,5
4	3	24	100

13. Analiza el siguiente diagrama de caja y bigotes y calcula, a partir de él, los valores máximo y mínimo, la mediana y los cuartiles.



Mínimo = 1500
 $Q_1 = 1750$
 $Me = 1950$
 $Q_3 = 2100$
 Máximo = 2500

14. Analiza el siguiente diagrama de caja y bigotes. Muestra los minutos que tarda en hacer efecto un medicamento en una población. Interpreta la información que presenta y responde a las preguntas.



Mínimo = 30
 $Q_1 = 55$
 $Me = 85$
 $Q_3 = 100$
 Máximo = 130

- ¿A qué porcentaje de la población había hecho efecto al cabo de 30 minutos?.
 - Al cabo de cuántos minutos había hecho efecto al 50 % de la población?.
 - Cuántos minutos tardó en hacer efecto al 100% de la población?
 - A qué porcentaje había hecho efecto a los 55 minutos?.
- ¿Cuánto tardó en hacer efecto a las tres cuartas partes de la población?

RESPUESTAS: a) Al 0%, 30 es el valor mínimo. b) a los 85 minutos (la mediana)
 c) 130 minutos (valor máximo) d) 55 es el primer cuartil, al 25%
 e) 100 minutos, $\frac{3}{4}$ partes son el 75%

4. Medidas de dispersión.

Rango y Desviación media

Las medidas de **dispersión** indican si los datos están más o menos agrupados respecto de las medidas de centralización.

- **Rango** o recorrido, es la diferencia entre el mayor y el menor valor de la variable, indica la longitud del intervalo en el que se hallan todos los datos.

Aunque el rango da una información importante, resulta más interesante calcular cuánto se desvían en promedio los datos de la media.

- **Desviación media**, es la media de los valores absolutos de las diferencias entre la media y los diferentes datos.

Varianza y desviación típica

Es otra forma de medir si los datos están o no próximos a la media y es la más utilizada.

- La **varianza** es la media de los cuadrados de las desviaciones.
- La **desviación típica** es la raíz cuadrada positiva de la varianza. Para designarla emplearemos la letra griega "sigma" σ .

$$\sigma = \sqrt{\frac{\sum f_i (X_i - \bar{X})^2}{n}} \quad \text{o} \quad \sigma = \sqrt{\frac{\sum f_i X_i^2}{n} - \bar{X}^2}$$

Es importante que entiendas el significado de estas medidas, cuanto mayores sean más dispersos estarán los datos.

Los intervalos alrededor de la media de amplitud 2 o 4 veces la desviación típica, tienen mucha importancia en estadística por el porcentaje de datos que hay en ellos.

Coefficiente de variación

Es el cociente entre la desviación típica y la media, se utiliza para comparar las dispersiones de datos de distinta media.

Por ejemplo, para los datos 4 y 6, el $CV=1/5=0,2$ y para 101 y 99 es $CV=1/100=0.01$. En ambos casos la desviación típica es la misma, pero en relación a la media es mas importante en el primero.

Desviación media

Ejemplo

X_i	f_i	$X_i \cdot f_i$	$ \bar{X} - X_i \cdot f_i$
5	2	10	29,44
10	1	10	9,72
15	3	45	14,16
20	4	80	1,11
25	6	150	31,66
30	2	60	20,55
	18	355	106,66

$$DM = \frac{106,66}{18} = 5,92$$

Desviación típica

Ejemplo

X_i	f_i	$X_i \cdot f_i$	$(X_i - \bar{X})^2 \cdot f_i$	$f_i \cdot X_i^2$
5	5	25	1164,81	125
10	1	10	105,33	100
15	1	15	27,7	225
20	2	40	0,13	800
25	1	25	22,43	625
30	9	270	853,25	8100
	19	385	2173,68	9975

$$\sigma = \sqrt{\frac{2173,68}{19}} = 10,69$$

O bien, como $\bar{x} = 20,26$

$$\sigma = \sqrt{\frac{9975}{19} - 20,26^2} = 10,69$$

Coefficiente de variación

$$\bar{x}=12 \quad \sigma=7 \quad CV= \boxed{58,33} \%$$

$$\bar{x}=110 \quad \sigma=7 \quad CV= \boxed{6,36} \%$$

$$\bar{x}=1148,25 \quad \sigma=7 \quad CV= \boxed{0,60} \%$$

Utilizar la Calculadora científica

MODELO 1

Modo estadístico

Primero se ha de elegir el modo estadístico. En muchas calculadoras se hace pulsando:

[MODE][.]

Datos desordenados

A continuación hay que introducir los datos, por ejemplo para 2, 3, 4, 3 teclareamos:

[2][M+] [3][M+] [4][M+] [3][M+]

Y para hacer los cálculos:

- Para la media **[SHIFT][x̄]**
- Para la desviación típica **[SHIFT][σ_n]**

También se puede sumar todos, o los cuadrados, o contar el n° de datos introducidos, pulsando respectivamente:

[SHIFT][Σx] [SHIFT][Σx²] [SHIFT][n]

Datos en una tabla

x_i f_i Se introducen los datos según la
2 4 secuencia: **[2][x][4][M+]**
3 3 **[3][x][3][M+]**
4 5 **[4][x][5][M+]**

Y ahora ya se pueden realizar los cálculos como antes.

Nota: Hay muchos modelos de calculadoras, pero afortunadamente, todas son bastante parecidas, trata de averiguar el funcionamiento de la tuya si no coinciden con éstas, consulta el manual o pregunta a tu profesor.

MODELO 2

Modo estadístico e introducción de datos

Elegimos el modo estadístico (mode stat 1-VAR) y nos aparece una columna donde introducir datos, uno tras otro, no importa que vayan desordenados. Si tuviéramos una tabla con frecuencias tendríamos que activar las frecuencias (Setup frequency on) y rellenar las columnas. Después del último dato pulsar AC.

Cálculos

Pulsando SHIFT STAT nos aparece un menú, **1:type, 2:Data, 3>Edit, 4:Sum, 5:Var 6:MinMax**. Con la opción **5:Var** accederemos a calcular la media, desviación típica y cantidad de datos. Con la opción **4:sum** las sumas que habitualmente necesitamos. Con la opción **6:MinMax** el mínimo y el máximo. Y con la opción **2:Data** podremos modificar los datos introducidos.



EJERCICIOS resueltos

13. Calcula el rango y la desviación media de los datos:

8 8 6 10 9 6 7 8 9 7
 7 6 6 7 9 5 5 7 10 7

x_i	f_i	$x_i \cdot f_i$	$ x_i - \bar{x} \cdot f_i$
5	2	10	4,8
6	4	24	5,6
7	6	42	2,4
8	3	24	1,8
9	3	27	4,8
10	2	20	5,2
20	147	24,6	

Solución:

El rango oscila entre 8 y 12 con una amplitud de 4. Hacemos el recuento.

La media:

Calculamos la desviación de cada dato respecto a la media, en valor absoluto. La media de las desviaciones:

14. Calcula la desviación media de los datos tabulados siguientes:

x_i	f_i	$x_i \cdot f_i$	$ x_i - \bar{x} \cdot f_i$	
[0,200)	100	7	700	2831,82
[200,400)	300	8	2400	1636,36
[400,600)	500	13	6500	59,09
[600,800)	700	9	6300	1759,09
[800,1000)	900	7	6300	2768,18
Total:	44	22200	9054,55	

Solución:

Calculamos la media:

Completamos la última columna:

Para practicar



- ¿Cuántas personas suponen una muestra del 5% de una población de 20.000 habitantes? ¿Y de una de 1000 habitantes?.
- De una población de 30000 individuos se ha estudiado varias características en 150 individuos. ¿Qué porcentaje del total ha sido estudiado?
- Un veterinario estudia las siguientes características en una muestra de animales de una granja tipo de animal, peso, color de los ojos, temperatura corporal, número de compañeros y metros cuadrados por animal.
- Haz un recuento de los siguientes datos, un gráfico de sectores y otro de barras. Indica el ángulo de cada sector.

a	b	c	a	c	c
d	c	d	b	d	a
d	a	b	b	c	c
a	a	b	a	b	d
- Haz un recuento de los siguientes datos y un diagrama de barras con polígono de frecuencias

3	3	1	1	3	2
3	3	2	1	3	2
2	3	1	1	4	3
2	2	4	4	3	3
- Agrupar los siguientes datos en 10 grupos. Agrupa los mismos datos, ahora, en 5 grupos.

3	6	5	9	2	6
2	2	7	9	4	6
2	5	9	9	1	0
2	5	3	6	7	8
6	4	3	6	7	9
10	10	9	1	6	8
6	2	3	9	6	5
6	6	5	7	6	6
10	1	3	4	4	4
- Calcula la media en cada caso:
 - 14,16, 18
 - 24, 26, 28, 26
 - 1000, 1200, 1800, 2000
- Calcula la media de los siguientes datos

3	3	1	1	3	2
3	3	2	1	3	2
2	3	1	1	4	3
2	2	4	4	3	3
- Calcula la media de los siguientes datos

10	1,5	18	20	16	1
9,5	5,50	15,5	6,5	4,5	4
8,5	7,5	1,5	15	13	0
20	12,5	7,5	4,5	14,5	9
- Determina la moda para los datos

3	3	1	1	3	2
3	3	2	1	3	2
2	3	1	1	4	3
2	2	4	4	3	3
- Calcula la mediana, el primer y el segundo cuartil de los datos del ejercicio anterior.
- Calcula de desviación media en cada caso:
 - 14, 16, 18
 - 34, 36, 38, 36
 - 1000, 1200, 1800, 2000
- Calcula el rango y la desviación media de los datos:

23	8	21	24	20	9
33	20	11	36	13	1
40	25	30	12	18	5
40	27	16	26	9	7
- Calcula la desviación media de los datos tabulados siguientes:

Intervalo	Marca = X_i	Fr	$F_i \cdot \bar{X} - X_i $
[0 , 200)	100	1	450
[200 , 400)	300	3	750
[400 , 600)	500	3	150
[600 , 800)	700	2	300
[800 , 1000)	900	3	1050

15. Calcula la media y la desviación típica en
 a) 2000, 2500
 b) 1750, 2750
 c) 2500, 2500

16. Calcula la media y la desviación típica de los datos:

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

17. Calcula el coeficiente de variación de los datos del ejercicio anterior.

18. Calcula la media y la desviación típica de los datos:

25	29	40	9	32	4
15	35	26	24	16	2
11	16	37	10	30	2
35	17	8	40	38	5

19. Calcula el coeficiente de variación de los datos del ejercicio anterior.

20. Calcula la media y la desviación típica de los datos agrupados siguientes:

X_i	f_i
5	7
10	0
15	2
20	2
25	4
30	2

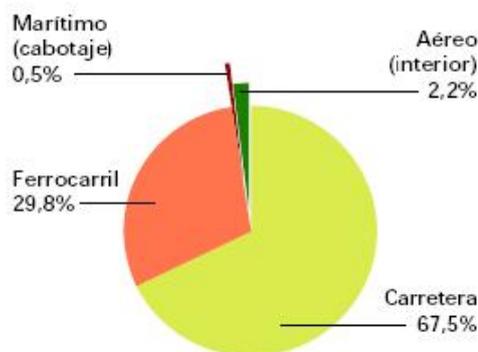
21. Haz los cálculos para un millón de habitantes en cada comunidad.

Tasa de criminalidad. 2006
 Infracciones penales por 1.000 hab.

Tasas más altas

Illes Balears	78,8
Comunidad de Madrid	70,8
Comunitat Valenciana	67,5
Ceuta	67,4
Cataluña	65,3

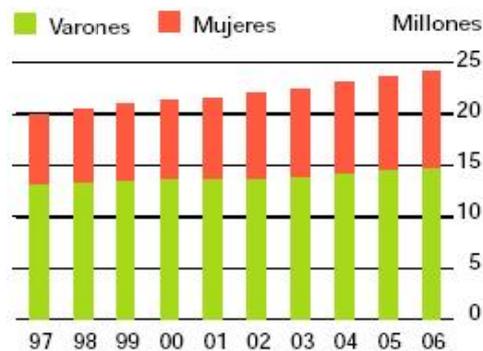
22. De cada millón de viajeros, ¿cuántos corresponden a cada sector?
Viajeros que utilizan transporte interurbano. 2007



Fuentes: INE, RENFE, FEVE, D. Gral. de Aviación Civil y D. Gral. de Puertos y Costas

23. ¿Cuántos conductores había en el año 2002? ¿Cuántos eran hombres y cuántas mujeres?

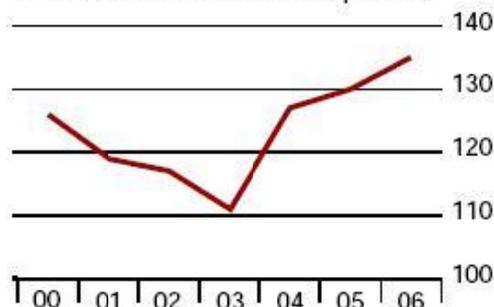
Censo de conductores



Fuente: Dirección General de Tráfico

24. ¿Entre qué años aumentaron más los detenidos por infracciones penales?

Tasa de detenidos
 Por cada 1.000 infracciones penales



Fuente: Ministerio del Interior

Para saber más



España en cifras

El Instituto Nacional de Estadística publica periódicamente datos como estos. Consulta la web. (www.ine.es)



Viviendas que disponen de acceso a internet. 2007



Uso de TIC en hogares. 2007

	Hogares con acceso a internet (%)	Personas que compraron por internet (%)
Países Bajos	83	55
Suecia	79	53
Dinamarca	78	56
Luxemburgo	75	47
Alemania	71	52
Finlandia	69	48
Reino Unido	67	53
Bélgica	60	21
Austria	60	36
Eslovenia	58	16
Irlanda	57	33
UE-27	54	29
Estonia	53	9
Letonia	51	11
Francia	49	:
Eslovaquia	46	16
España	45	18
Lituania	44	6
Italia	43	10
Polonia	41	16
Portugal	40	9
Chipre	39	10
Hungría	38	11
República Checa	35	17
Grecia	25	8
Rumania	22	3
Bulgaria	19	3
Malta	:	:

: Dato no disponible

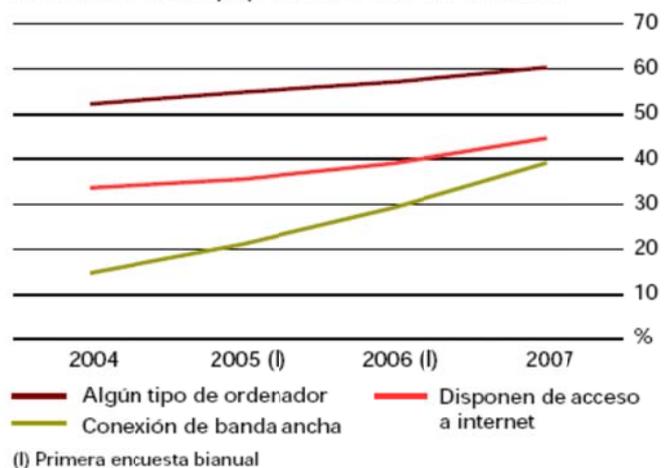
Fuente: Eurostat

Equipamiento tecnológico de los hogares

Según datos de la Encuesta sobre Equipamiento y Uso de las Tecnologías de Información y Comunicación en los Hogares, en 2007 el 99% de los hogares dispone de teléfono (ya sea fijo o móvil), el 99,5% dispone de televisión y el 22,8% tiene recepción de televisión digital terrestre (TDT). En una de cada cinco viviendas con televisión, alguna de ellas es de pantalla plana (plasma, LCD). El DVD se halla presente en tres de cada cuatro hogares, en detrimento del vídeo.

Un 60,4% dispone de algún tipo de ordenador.

Evolución del equipamiento TIC en viviendas



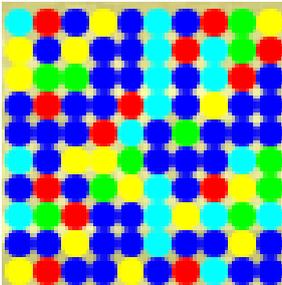
La banda ancha gana adeptos

6,5 millones de viviendas familiares tienen acceso a internet (un 44,6% del total). El 39% de los hogares disponen de conexión mediante banda ancha (ADSL, red de cable,...), lo que supone un incremento de 10 puntos respecto al año 2006. El mayor porcentaje de viviendas con acceso a internet se da en Comunidad de Madrid (56,6%) y Cataluña (51,3%). La diferencia en puntos porcentuales entre sexos respecto a los principales indicadores de uso de TIC se reduce en 2007, tanto en el uso de ordenador como en los usuarios frecuentes; sin embargo, aumenta en el acceso a internet. El 60,5% de los varones han usado internet alguna vez frente a un 54,1% de las mujeres. En la Unión Europea, los países que disponen de indicadores TIC superiores a la media se encuentran, principalmente, en el norte de Europa. Por debajo de la media, se sitúan los países del área mediterránea o los que son de reciente ingreso.

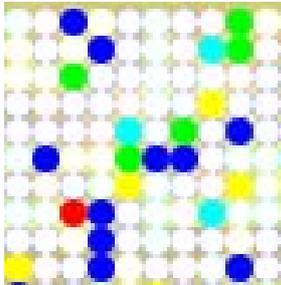


Recuerda lo más importante

Población



Muestra



Nº de hermanos: 4 3 2 3 1 2 0 2 0 1 2 3 1 2 4
0 1 1 4 1 1 4 0 4 2 0 4 1

Recuento de datos:

x_i	f_i	$x_i \cdot f_i$	$f_i \cdot (x_i - \bar{x})^2$
0	5	0	0
1	8	8	6,37
2	6	12	0,06
3	3	9	3,67
4	6	24	26,64
Total	28	53	54,67

Gráficos de **sectores y barras**

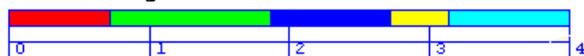


Media y moda

$$\text{Media} = \bar{X} = \frac{53}{28} = 1.89$$

$$\text{Moda} = M_o = 1$$

Cuartiles y mediana



$$Me=2, Q1=1, Q3=3$$

Rango.

De 0 a 4, de amplitud 4

Desviación típica

$$\sigma = \sqrt{\frac{54.67}{28}} = 1.39$$

Coefficiente de variación

$$CV = 1,39 / 1,89 = 0,73 = 73\%$$

Variables estadísticas:

- Cualitativa, color preferido;
- Cuantitativa discreta, nº de hermanos
- cuantitativa continua, altura.

Altura: 172 162 147 184 140 156
153 186 157 189 162 175 162 158 163
150 152 163 151 182 146 154 163 170
183 162 176 167 168 165

Intervalo	x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
[140,150)	145	3	435	1200
[150,160)	155	8	1240	800
[160,170)	165	10	1650	0
[170,180)	175	4	700	400
[180,190)	185	5	925	2000
Total		30	4950	4400

Media:

$$\bar{x} = \frac{4950}{30} = 165$$

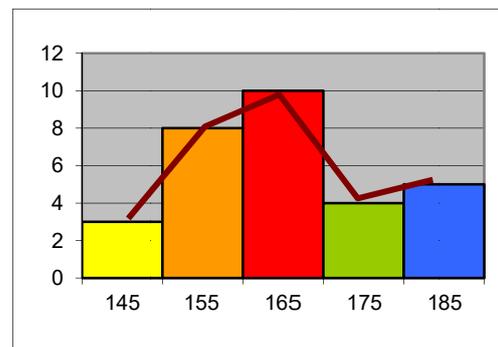
Varianza y Desviación típica:

$$\sigma^2 = \frac{4400}{30} = 146,67 \quad \sigma = \sqrt{146,67} = 12,11$$

Coefficiente de variación:

$$CV = 12,11/165 = 0,073 = 7,3\%$$

Histograma y polígono de frecuencias:



Autoevaluación



1. Haz un recuento de los datos siguientes

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

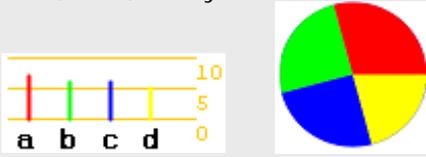
2. Haz un gráfico de barras para los datos anteriores.

3. Calcula la media de los datos dados por la tabla

X_i	f_i
1	11
2	3
3	5
4	5

4. Calcula la mediana de los datos anteriores.
5. Calcula el primer cuartil de los datos del ejercicio 3
6. Calcula el tercer cuartil de los datos del ejercicio 3
7. Calcula el rango de los datos del ejercicio 3
8. Calcula la desviación media de los datos anteriores
9. Calcula la desviación típica de los datos del ejercicio 3
10. Calcula el coeficiente de variación para los datos del ejercicio 3

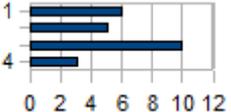
Soluciones de los ejercicios para practicar

1. a) 400 b) 50
2. 0,5%
3. V- cualitativa: tipo animal, color de ojos.
V. cuantitativa discreta: número de compañeros.
V. cuantitativa continua: peso, temperatura y metros cuadrados.
4. $a \rightarrow 7$, $b \rightarrow 6$, $c \rightarrow 6$ y $d \rightarrow 5$

5. $0 \rightarrow 0$, $1 \rightarrow 5$, $2 \rightarrow 6$, $3 \rightarrow 10$, $4 \rightarrow 3$

6. En 10: $1 \rightarrow 3$, $2 \rightarrow 6$, $3 \rightarrow 5$, $4 \rightarrow 5$, $5 \rightarrow 5$, $6 \rightarrow 13$, $7 \rightarrow 4$, $8 \rightarrow 2$, $9 \rightarrow 7$ y $10 \rightarrow 4$.
En 5: $(1 \text{ ó } 2) \rightarrow 9$, $(3 \text{ ó } 4) \rightarrow 10$, $(5 \text{ ó } 6) \rightarrow 18$, $(7 \text{ u } 8) \rightarrow 6$ y $(9 \text{ ó } 10) \rightarrow 11$
7. a) $(14+16+18)/3 = 16$
b) $(24+26+28+26)/4 = 26$
c) $(1000+1200+1800+2000)/4 = 1500$
8. $\bar{X} = 2,46$
9. $\bar{X} = 9,77$
10. $Mo = 3$, es el valor que mas veces aparece, 10 en total.
11. $1 \rightarrow 5$, $2 \rightarrow 6$, $3 \rightarrow 10$ y $4 \rightarrow 3$.
Mediana = 3, $Q1=2$ y $Q3=3$
12. a) $(2+0+2)/3 = 1,3333\dots$
b) $(2+0+2+0)/4 = 1$
c) $(500+300+300+500)/4 = 1600/4 = 400$
13. El rango oscila entre 1 y 40 con una amplitud de 39. $\bar{X} = 20,58$, $DM=8,92$

$$DM = \frac{2700}{12} = 225$$
- 14.
15. a) $\bar{X} = 2250$ $\sigma = 250$
b) $\bar{X} = 2250$ $\sigma = 500$
c) $\bar{X} = 2500$ $\sigma = 0$
16. Media = 2,58 D. típica = 1,21
17. $CV = 1,21/2,58 = 0,46$
18. $\bar{X} = 21,08$ y $\sigma = 12,98$
19. $CV = 0,62$
20. $\bar{X} = 15,58$ y $\sigma = 9,68$
21. 78800, 70800, 67500, 67400 y 65300
22. Carretera 67500, aéreo 22000, marítimo 5000 y ferrocarril 298000
23. Aproximadamente 22.000.000, de los que hombres son 14.000.000 y mujeres 8.000.000

Soluciones AUTOEVALUACIÓN

- | | | | | |
|-------|---|---|----|---|
| x_i | 1 | 2 | 3 | 4 |
| f_i | 6 | 5 | 10 | 3 |
- 
- 1.
 - 2.
 3. $\bar{x} = 2,17$
 4. $Me = 2$
 5. $Q1 = 1$
 6. $Q3 = 3$
 7. rango=3
 8. $DM = 1,11$
 9. $\sigma = 1,24$
 10. $Cv = 0,571$

Utilizar la Hoja de Cálculo

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

Introducción de datos y primeros cálculos

En general, las siguientes opciones están disponibles en todas las hojas de cálculo, en particular en la de Openoffice y en Excel.

✓ **Abre** una hoja de cálculo e introduce los datos 3, 4, 5, 3, 4, 6, 3, 3, 4 en la columna A, en A1, A2,... estos valores serán los datos sobre los que haremos los primeros cálculos.

Los datos están en el área "A1:A9" de la hoja de cálculo.

Si los datos estuvieran en 3 filas x 3 columnas, empezando por A1, sería "A1:C3".

Y por ejemplo, si con el ratón selecciono las celdas de la imagen adjunta, el área que cubro es "A2:C3".

	A	B	C
1	3		
2	4		
3	5		

Hay funciones para todas y cada uno los cálculos estadísticos que hemos estudiado y muchas más que puedes investigar por tu cuenta. Las funciones hacen cálculos con los números del área que se les proporciona.

Veamos como se harían los cálculos para los datos que hemos introducido. Haremos los cálculos en la columna C, la B la usaremos para indicar el nombre del cálculo que hay a continuación.

	A	B	C
1	3	Moda	3
2	4	Máximo	6
3	5	Mínimo	3
4	3	Rango	3
5	4	Suma	35
6	6	Media	3,888888889
7	3	Mediana	4
8	3	Cuartil 1º	3
9	4	Cuartil 3º	4
10		Desviación típica	1,054092553
11		Desviación media	0,790123457
12			

Moda: En la celda C1 introduce la fórmula =**moda(A1:A9)**.

Máximo: En C2 escribe =**máx(A1:A9)**

Mínimo: Introduce =**min(A1:A9)**

Rango: =**C2-C3**

Suma: =**suma(A1:A9)**

Media: =**promedio(área)** Recuerda lo que es el **área** de los datos, **A1:A9**

Mediana: =**mediana(área)**

Cuartil 1º. =**cuartil(área;1)**

Cuartil 3º. =**cuartil(área;3)**

Desviación media: =**desprom(área)**

Desviación típica: =**desvest(área)**

Ejercicio

Introduce datos en un área mas grande que la anterior, al menos de 20 celdas y haz todos los cálculos de dos formas, con la hoja de cálculo y como lo has estudiado. Si todos los resultados son correctos esta hoja te vale para cualquier conjunto de datos de ese tamaño.

ANEXO

Valores aleatorios.

Si en 20 celdas escribimos **=aleatorio()** (podemos escribirlo en una y copiarlo en las otras) tendremos 20 datos entre 0 y 1 aleatorios. Si queremos que sean números comprendidos entre 0 y 20 escribiremos **=20*aleatorio()** Y si queremos que no lleven decimales escribiremos **=entero(20*aleatorio())** De esta forma nos inventamos 20 datos para hacer cálculos estadísticos.

Contar

Si tenemos introducidos una cantidad grande de datos. Con la función **=contar(área)** nos da la cantidad de números del área (imagina que hay celdas vacías y no es fácil contar las que están rellenas de números).

Si lo que tenemos son datos cualitativos (textos) usaremos **=contara(área)** para que nos cuente las celdas ocupadas del área indicada.

Agrupar datos.

- En **variable discreta**. Si en un área de la hoja de cálculo tenemos 20 números enteros entre 0 y 4, por ejemplo. ¿Cómo podemos contar cuántos de ellos son un 4? Escribiremos en una celda en blanco **=contar.si(área;4)**. Para contar el resto usaremos en la celda que nos convenga **=contar.si(área;0)**, **=contar.si(área;1)**, **=contar.si(área;2)**, ...

14	Datos	Marca	Frecuencia
15	1	4 X1=0	5
16	0	0 X2=1	2
17	3	0 X3=2	5
18	2	3 X4=3	3
19	2	3 X5=4	5
20	4	4	Suma
21	2	1	20
22	4	0	
23	0	4	
24	2	2	

Por ejemplo. En la imagen de la izquierda hay 20 datos generados aleatoriamente entre 0 y 4. En la columna Marca indico que valores voy a contar en la siguiente columna, y en la columna frecuencia uso la fórmula '**=contar.si(área;valor)**' escribiendo: **=contar.si(área;0)**, **=contar.si(área;1)**, **=contar.si(área;2)**, ...

Y finalmente compruebo sumando las frecuencias que se han contado todos los datos (en la celda D21 escribo **=suma(d15:d19)**).

Ejercicio. Genera 20 datos aleatorios entre 0 y 4 y haz una tabla de frecuencias usando la función contar.si Calcula todos los parámetros estadísticos estudiados (rango, máximo, mínimo, media,...).

- En **variable continua**. Si tenemos 100 alturas entre 150 y 200 centímetros en un área de una hoja de cálculo y queremos saber cuántas están en el intervalo [150,160) usaremos **=contar.si(área;">=150")-contar.si(área;">=160")** ; para el intervalo [160,170) usaremos **=contar.si(área;">=160")-contar.si(área;">=170")** y así para cada uno de los intervalos.

- En **variable cualitativa**. Si lo que tenemos es 100 colores y queremos saber cuántos rojos hay se escribe: **=contar.si(área;"rojo")**

Clasificar

Según el tipo de variable que empleemos, usaremos la fórmula 'contar.si' sobre el área de datos tantas veces como sea necesario para clasificar todos los datos y nos pueda ser de utilidad. Así, para colores usaremos **=contar.si(área;"rojo")**, **=contar.si(área;"verde")**, ... tantas veces como colores tengamos, de forma ordenada, en una columna por ejemplo.

Ejemplo

Vamos a inventarnos el peso de 50 personas con valores entre 50 y 110. Usaremos la fórmula **=aleatorio()*60+50**, si queremos que sean valores enteros se pone **=entero(aleatorio()*50+50)** y la copiaremos en el área **A1:E10**.

Ahora hay que clasificar los datos. Como todos están entre 50 y 110 los intervalos tendrán que reflejarlo. Vamos a realizar dos tablas distintas, una con valores de 20 en 20 y otra con valores de 10 en 10

Para primera, copiaremos la siguiente tabla en un lugar vacío de la hoja de cálculo

Intervalo, marca	Frecuencia
[50,70] 60	=contar.si(A1:E10;">=50")-contar.si(A1:E10;">=70")
[70,90] 80	=contar.si(A1:E10;">=70")-contar.si(A1:E10;">=90")
[90,110] 100	=contar.si(A1:E10;">=90")-contar.si(A1:E10;">=110")

Para la segunda, copiaremos esta tabla.

Intervalo, marca	Frecuencia
[50,60] 55	=contar.si(A1:E10;">=50")-contar.si(A1:E10;">=60")
[60,70] 65	=contar.si(A1:E10;">=60")-contar.si(A1:E10;">=70")
[70,80] 75	=contar.si(A1:E10;">=70")-contar.si(A1:E10;">=80")
[80,90] 85	=contar.si(A1:E10;">=80")-contar.si(A1:E10;">=90")
[90,100] 95	=contar.si(A1:E10;">=90")-contar.si(A1:E10;">=100")
[100,110] 105	=contar.si(A1:E10;">=100")-contar.si(A1:E10;">=110")

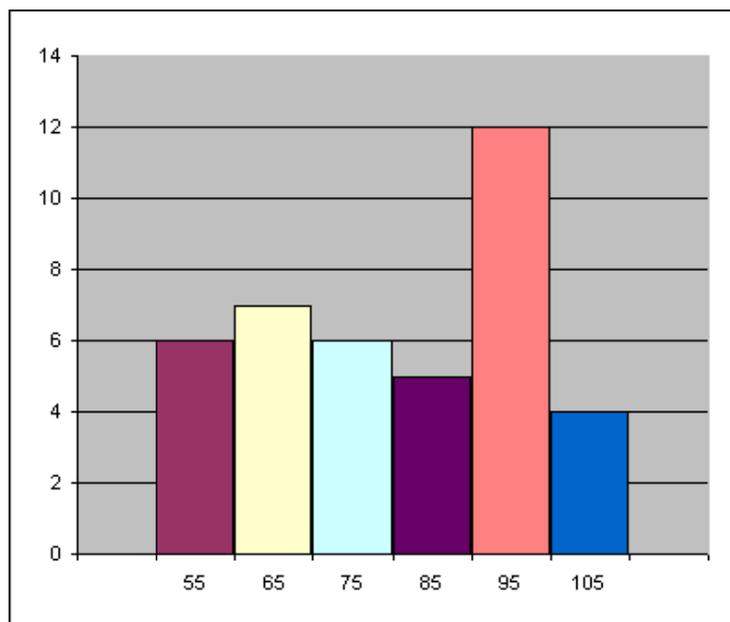
Ejercicio

Genera 100 números de 0 a 80 y haz dos agrupaciones en intervalos distintas. Por ejemplo, una de 10 en 10 y otra de 20 en 20.

Gráficos

Sencillo y sin etiquetas. Basta con seleccionar el área de las frecuencias de los datos y crear el gráfico pulsando en gráficos del menú insertar. Después podemos elegir el tipo de gráfico que queremos, y el aspecto de este entre una gran variedad de opciones que incorporan todas las hojas de cálculo.

Con etiquetas: Lo más sencillo es utilizar la primera fila o primera columna para las etiquetas. En las opciones del gráfico habrá que activar y/o desactivar la que corresponda con la forma de presentar los datos en la hoja.



Ejercicio

Haz dos gráficos diferentes, uno para cada una de las distribuciones del ejercicio anterior.