

Obxectivos

Nesta quincena aprenderás a:

- Distinguir os distintos tipos de variables estatísticas.
- Agrupar en intervalos os datos dun estudio estatístico.
- Facer a táboa estatística asociada a un conxunto de datos.
- Representar e interpretar gráficos estatísticos, e saber cando é conveniente utilizar cada tipo.
- Calcular a media, a moda, a mediana e os cuartís dun conxunto de datos.
- Que son e como se calculan os parámetros de dispersión: o rango ou percorrido, a varianza e a desviación típica, o coeficiente de variación.

1.Facer estatística	páx. 4
Necesidade	
Poboación e mostra	
Variables	
2.Reconto de datos	páx. 5
Reconto de datos	
Gráficos	
Agrupación de datos en intervalos	
3.Medidas de centralización e posición	páx. 9
Medida	
Moda	
Cuartís e mediana	
Diagramas de caixa e bigotes	
4.Medidas de dispersión	páx. 12
Rango e desviación media	
Desviación típica	
Coeficiente de variación	

Exercicios para practicar

Para saber máis

Resumo

Autoavaliación

Anexo

Antes de empezar

Estatística

A estatística, a nivel primario, é unha actividade que todo o mundo fai dende moi pequeno. O mero feito de contar e/ou clasificar os teus xoguetes (os teus coches, bonecas/os, bólas, videoxogos ...) xa é unha actividade estatística.

Clasificar obxectos

O desparafusador á caixa de ferramentas, os cubertos ao caixón da cociña, os libros ao estante, os videoxogos xunto á consola ...

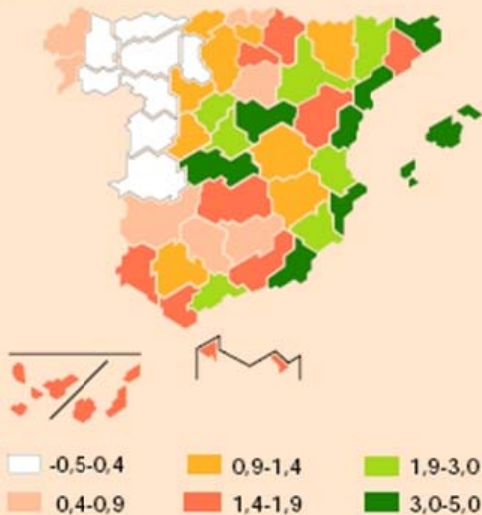


Competicións escolares

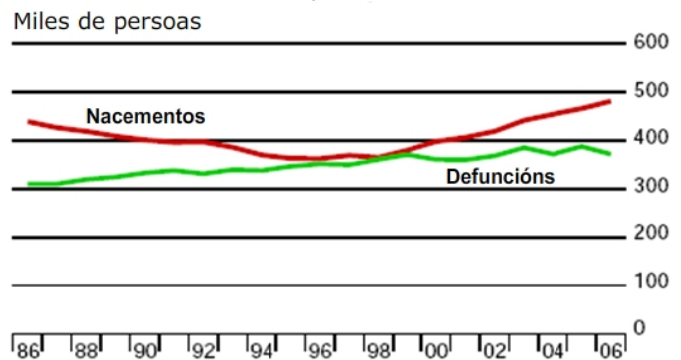


No teu instituto haberá equipos e competicións, e haberá que levar un cómputo de resultados e unha ordenación de equipos segundo unhas puntuacións. Ata é posible que haxa un rexistro de varios anos.

Crecedamento relativo da poboación total
1 de xaneiro de 2008 (%)



Crecedamento natural da poboación



1. Estatística descritiva

Necesidade

Ao poñer en práctica unha medida social para saber a súa aceptación, a cantas persoas pode ir dirixida?, cales son os distintos niveis? Fronte a unha iniciativa como esta, preguntar a toda a poboación se pode esgotar os recursos destinados a ela. Unha enquisa previa pode aforrarnos algún que outro equívoco.

Poboación e mostra

Cando se fai un estudo estatístico o investigador decide se analizará toda a poboación ou unha mostra elixida previamente.

Poboación é o conxunto de individuos, con algunha característica común, sobre o que se fai un estudo estatístico.

A **mostra** é un subconxunto da poboación. Debe elixirse que sexa representativa de toda a poboación na característica estudada.

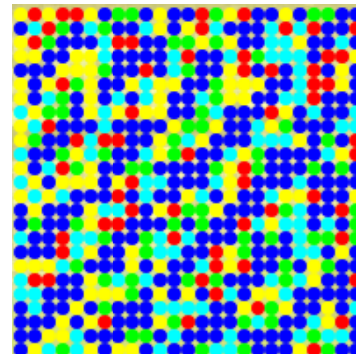


Atributos e Variables.

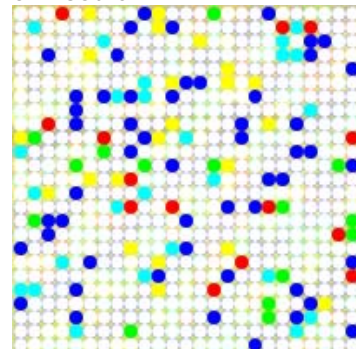
Cada unha das propiedades ou características que podemos estudar é unha **variable estatística**. Dependendo dos posibles valores que poidan tomar clasifícanse en:

- **Variables cualitativas** ou atributos. Os valores da variable non son números senón calidades, exprésanse con palabras. A cor, a forma, o sexo ... son exemplos de variables cualitativas.
- **Variables cuantitativas**. Os datos exprésanse numericamente e poden ser:
 - Discretas. Cada unha das variables só pode tomar valores enteiros (1, 2, 3 ...). O nº de irmáns, o nº ventás de casa, o nº colexios da túa poboación ...
 - Continuas. Poden tomar calquera valor dun intervalo dado: o noso peso, altura, forza. Non é posible medilas con números enteiros, a densidade do aire, a velocidade media dos fórmula 1 nunha carreira ...

Poboación



e mostra



A cor dos lapis, é unha variable cualitativa



A altura, idade e peso, son variables cuantitativas.



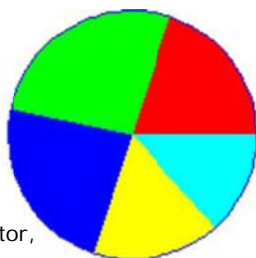
2. Reconto e gráficos



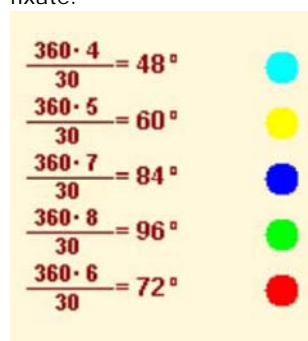
É parte do proceso. Despois de recompilar os datos, procédese ao seu reconto para expresalos de forma ordenada e para que sexa máis doado traballar con eles. Xeralmente elabórase unha táboa como se mostra á esquerda onde podes practicar.

- Frecuencia **absoluta**, é o nº de veces que aparece un dato. Á de x_i chamarémola f_i .
- Frecuencia **relativa**, é o cociente entre a frecuencia absoluta e o nº total de datos.
- Frecuencia **acumulada** dun dato, é a suma das frecuencias absolutas dos valores que son menores ou iguais ca el, indicarémola con F_i . Tamén se poden calcular as frecuencias relativas acumuladas.

Diagrama de sectores



Para calcular os graos de cada sector, fíxate:



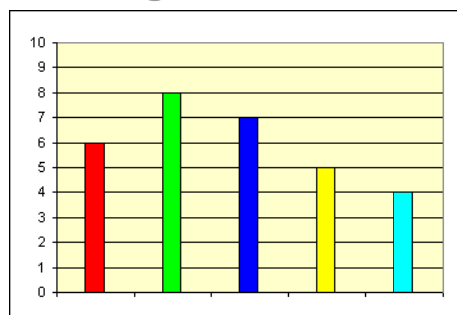
Diagramas de barras e de sectores

Os datos estatísticos adoitan representarse de forma gráfica, xa que desta forma podemos facernos unha idea da súa distribución dun só golpe de vista. En función do tipo de variable convén máis usar un tipo de gráfico ou outro.

- **Diagrama de sectores**, pode aplicarse a calquera tipo de variable, aínda que é o máis adecuado en variables cualitativas e para unha primeira toma de contacto cos valores dunha poboación. É un círculo dividido en sectores de ángulo proporcional á frecuencia de cada valor. A amplitude de cada sector obtense multiplicando a frecuencia relativa por 360° .

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

Diagrama de barras



- **Diagrama de barras**. Tamén pode aplicarse a calquera tipo de variable, aínda que se considera o idóneo para variables discretas. Cada valor correspóndese cunha barra de lonxitude proporcional á súa frecuencia.

EXERCICIOS resoltos

1. Cantas persoas supoñen unha mostra do 10% dunha poboación de 10.000 habitantes? E dunha de 6000 habitantes?.

Solución: a) $10.000 \cdot 10 / 100 = 1000$, b) $6000 \cdot 10 / 100 = 600$

2. Unha empresa de sondaxes estatísticas ten capacidade para entrevistar 1000 persoas por semana. Se dispón de 4 semanas a que porcentaxe dunha poboación de 100.000 habitantes pode entrevistar para obter unha mostra?.

Solución: En 4 semanas pode entrevistar 4000 persoas. 4000 de 100.000 equivale a 4 de 100. Así pois, o 4%.

3. Co fin de coñecer mellor a forma de viaxar dunha poboación prepararon unha enquisa. Algunhas das preguntas trataron sobre: N° de días de viaxe, diñeiro empregado, número de vultos, zonas xeográficas, medio de transporte, natureza da viaxe (negocios, turismo, familiar, saúde...) e n° de persoas. Clasifica estas variables estatísticas.

Solución:

V cualitativa: Zonas xeográficas, medio de transporte e natureza da viaxe.

V. cuantitativa discreta: N° de días, número de vultos e n° de persoas.

V. cuantitativa continua: Diñeiro empregado.

4. Fai un recuento dos seguintes datos:

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

Solución:

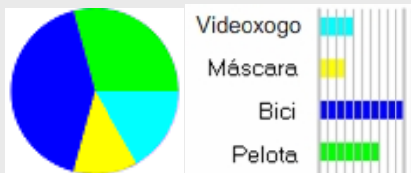
X_i	f_i
1	3
2	8
3	4
4	9

5. Fai un recuento dos seguintes datos, un gráfico de sectores e outro de barras. Indica o ángulo de cada sector.

Pelota, máscara, pelota, máscara, máscara, bici, máscara, bici, bici, máscara, máscara, máscara, máscara, videoxogo, máscara, pelota, videoxogo, pelota, videoxogo, pelota, pelota, videoxogo, pelota, máscara.

Solución:

X_i	f_i	graos
Videoxogo	4	60
Máscara	3	45
Bici	10	150
Pelota	7	105



Os 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 as marcas de cada clase como se corresponden coa media dos seus 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 agora 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, ou en discretas cando o número de datos distintos se fai case tan grande como o número de datos, e para poder estudalos, faise necesario agrupalos en **intervalos** ou **clases**, habitualmente da mesma amplitude e como mínimo catro.

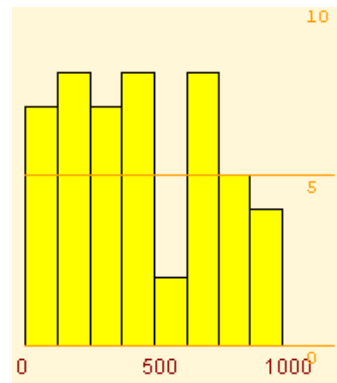
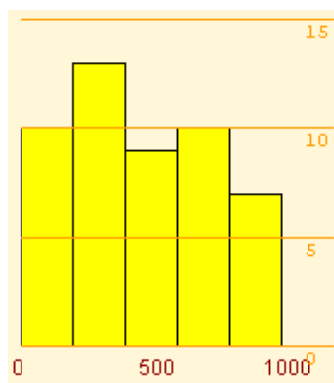
Por exemplo, nunha poboación hai case tantas alturas como individuos pero podemos agrupalos en baixos, medios e altos; tamén poderíamos facer baixos, medios-baixos, medios-altos e altos, ou clasificalos de 10 en 10 cm ou de 20 en 20 cm ...

- Para representar a todos os datos dun intervalo eliximos un valor, o punto medio do intervalo; chámase **marca de clase**.

Histograma

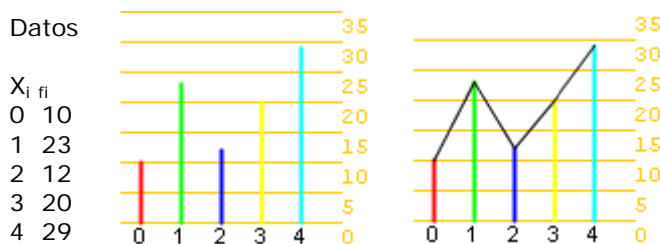
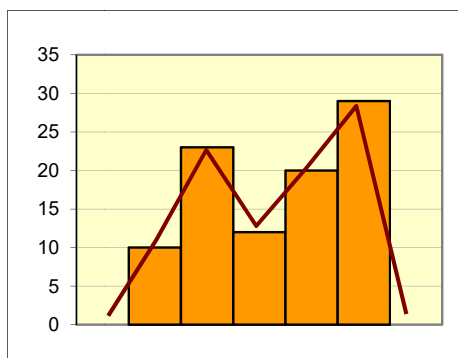
Cando os datos veñen agrupados en intervalos úsase para representalos graficamente, o **histograma**. Cada valor represéntase cun rectángulo de anchura ao intervalo correspondente e coa altura proporcional á súa frecuencia.

Os histogramas para os datos da marxe agrupados en cinco e oito intervalos:



Polígono de frecuencias.

Creámolo ao unir os extremos superiores das barras dos histogramas ou dos diagramas de barras.

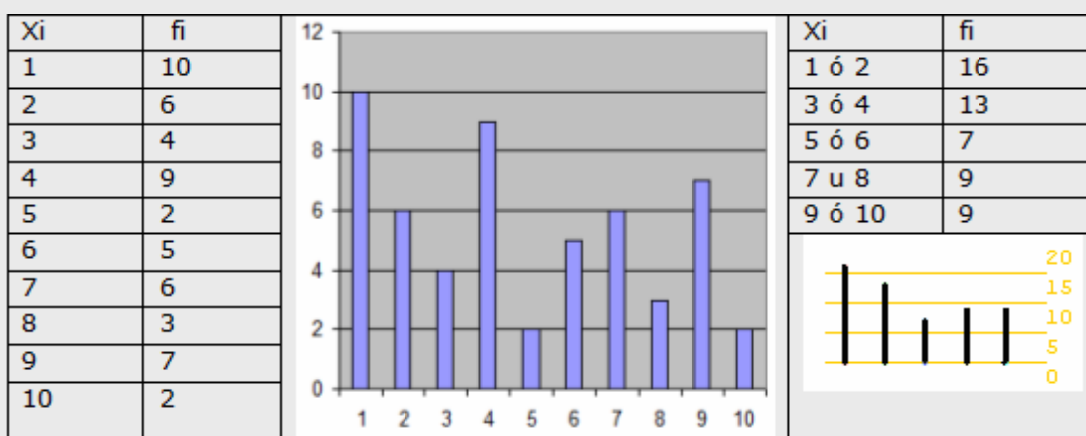


EXERCICIOS resoltos

6. Agrupa os seguintes datos en 10 grupos. Agrupa os mesmos datos, agora, en 5 grupos e realiza 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:

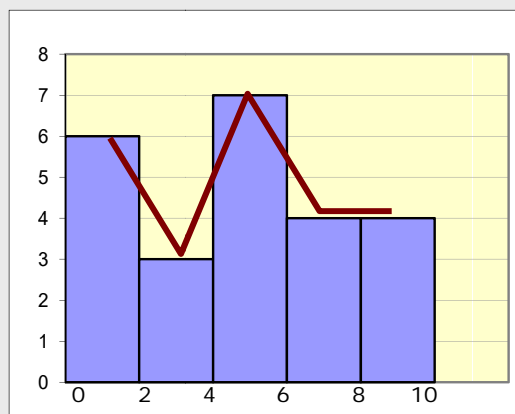


7. Agrupa os datos seguintes en 5 intervalos de igual amplitude. Realiza un gráfico e un polígono de frecuencias.

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

Exemplo 1

10, 12, 10, 14 y 13

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

Exemplo 2

Xi	fi	Xi · fi
5	4	20
10	6	60
15	7	105
20	9	180
25	4	100
30	6	180
36	645	

A media:
 $\bar{X} = \frac{645}{36} = 17,91$

Moda

Exemplo 1

Xi	fi	Moda
5	8	5
10	5	
15	1	
20	8	20
25	0	
30	3	

Exemplo 2

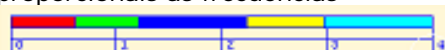
Xi	fi	Moda
100	5	
200	5	
300	1	
400	3	
500	2	
600	9	600

Medidas de posición

Ampliamos a táboa coa columna de frecuencias acumuladas e porcentaxes destas

xi	fi	Fi	%
1	10	10	15,152
2	10	20	30,303
3	17	37	56,061
4	12	49	74,242
5	17	66	100

e/ou unha barra con lonxitudes proporcionais ás frecuencias



e poderemos saber a mediana e cuartís

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

3. Medidas de centralización e posición

A media

Todos os alumnos saben que cun 6 e un 4 teñen de media 5. Pois a media en estatística non é outra cousa que iso, só que, habitualmente, con máis datos.

Para calcular a media se son poucos os datos, súmanse todos e divídese entre o número total. Se son moitos, terémolos agrupados; entón súmanse os produtos de cada dato pola súa frecuencia absoluta e divídese esta suma polo número total de datos. Indícase con \bar{X} .

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

A moda

¿Quen non oíu algunha vez: "Está de moda ir a...", "Leva este tipo de pantalón, está de moda", ou "Púxose de moda o grupo"... e todo o mundo entende que hai unha boa cantidade de persoas nesas opcións.

Así pois, o valor que máis frecuencia teña será "o de moda", aínda que pode acontecer que haxa máis dun.

- A **moda**, **Mo**, dunha distribución estatística é o valor da variable que máis se repite, o de maior frecuencia absoluta.

A mediana e os cuartís

A mediana e os cuartís, como a media aritmética, só se poden calcular cando a variable é cuantitativa.

- A **mediana**, **Me**, é o valor que ocupa a posición central unha vez ordenados os datos en orde crecente; é dicir, o valor que é maior que o 50% e menor que o outro 50%.

A mediana divide a distribución en dúas partes con igual nº de datos, se a dividimos en catro partes obtemos, os **cuartís**, 1º, 2º e 3º, que se indican respectivamente **Q₁**, **Q₂** e **Q₃**.

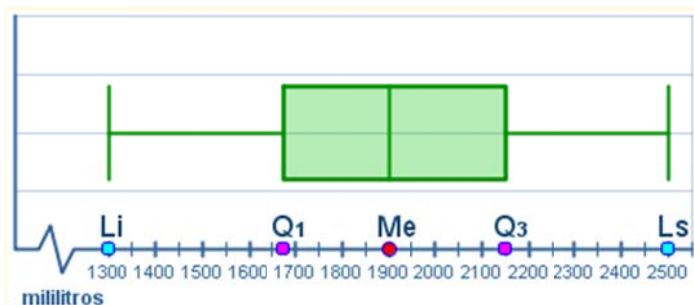
Ordenados os datos, o **primeiro cuartil**, é maior que o 25% destes; o **terceiro cuartil**, maior que o 75%, e o segundo coincide coa mediana.

Diagramas de caixa e bigotes

A partir do valor da mediana e os cuartís pódense representar as distribucións estatísticas mediante os chamados "diagramas de caixa e bigotes".

Vexamos como se constrúe cos datos da táboa da dereita. Unha vez ordenados os datos, calcúlanse os valores mínimo e máximo, os cuartís e a mediana.

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



Sitúanse estes valores sobre o eixe de abscisas e débúxase a "caixa" dende o primeiro ao terceiro cuartil (o percorrido *intercuartilico*), e os "bigotes" como indica a figura.

A táboa mostra o consumo diario de auga, en ml, dos 20 alumnos dunha 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	
Pepe	1850	Iris	1900	Me
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: A lonxitude dos bigotes non debe exceder unha vez e media a da caixa, se hai valores extremos que superan esa medida débúxanse como puntos illados.

EXERCICIOS resoltos

8. Calcula a media en cada caso:

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

Solucións:

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

9. Calcula a media dos seguintes 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 e dividimos entre 24, o número de datos. $\bar{X} = 2,29$

10. Calcula a media dos seguintes 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 a moda para os 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

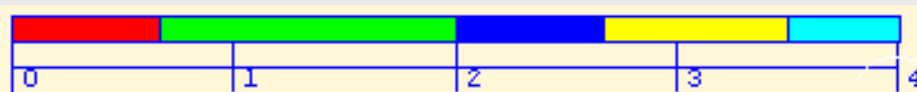
Solución:

Mo = 1 é o valor que mais veces aparece, 8 en total.

12. Calcula a mediana, o primeiro e o segundo cuartil dos datos do exercicio anterior.

Solución:

Facemos o reconto, $0 \rightarrow 4$, $1 \rightarrow 8$, $2 \rightarrow 4$, $3 \rightarrow 5$ e $4 \rightarrow 3$, debuxamos barras de cores de lonxitudes proporcionais ás frecuencias, valerían por exemplo de 4 mm, 8 mm, 4 mm, 5 mm e 3 mm. Dividimos toda a barra en 4 partes e fixémonos a cor na que queda a división.

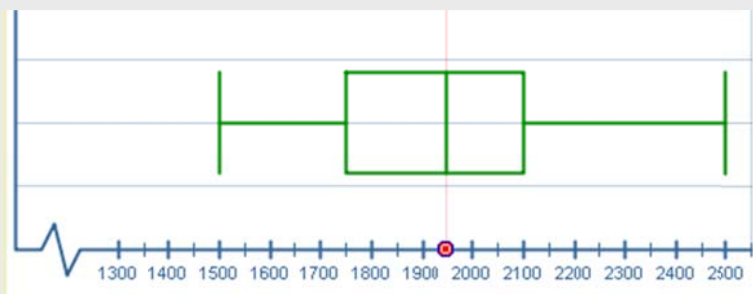


Mediana = 1,5, $Q_1=1$ e $Q_3=3$

Ou ben construímos a táboa e observamos onde quedan na última columna os valores 25%, 50% e 75%.

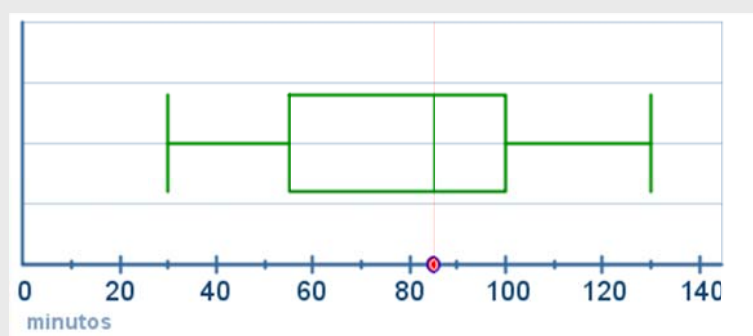
X_i	f_i	F_i	%
0	4	4	16,667
1	8	12	50
2	4	16	66,667
3	5	21	87,5
4	3	24	100

13. Analiza o seguinte diagrama de caixa e bigotes e calcula, a partir del, os valores máximo e mínimo, a mediana e os cuartís.



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

14. Analiza o seguinte diagrama de caixa e bigotes. Mostra os minutos que tarda en facer efecto un medicamento nunha poboación. Interpreta a información que presenta e responde ás preguntas.



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

- A que porcentaxe da poboación fixera efecto ao cabo de 30 minutos?.
- Ao cabo de cantos minutos fixera efecto ao 50% da poboación?.
- Cantos minutos tardou en facer efecto ao 100% da poboación?
- A que porcentaxe fixera efecto aos 55 minutos?.
- Canto tardou en facer efecto ás tres cuartas partes da poboación?

RESPOSTAS: a) Ao 0%, 30 é o valor mínimo. b) aos 85 minutos (a mediana)
 c) 130 minutos (valor máximo) d) 55 é o primeiro cuartil, ao 25%
 e) 100 minutos, $\frac{3}{4}$ partes son o 75%

4. Medidas de dispersión.

Rango e Desviación media

As medidas de **dispersión** indican se os datos están máis ou menos agrupados respecto das medidas de centralización.

- **Rango** ou percorrido, é a diferenza entre o maior e o menor valor da variable, indica a lonxitude do intervalo no que se achan todos os datos.

Aínda que o rango dá unha información importante, resulta máis interesante calcular canto se desvían en media os datos da media.

- **Desviación media**, é a media dos valores absolutos das diferenzas entre a media e os diferentes datos.

Desviación media

Exemplo

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

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

Varianza e desviación típica

É outra forma de medir se os datos están ou non próximos á media e é a máis utilizada.

- A **varianza** é a media dos cadrados das desviacións.
- A **desviación típica** é a raíz cadrada positiva da varianza. Para designala empregaremos a letra grega "sigma" σ .

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

É importante que entendas o significado destas medidas; canto maiores sexan, máis dispersos estarán os datos.

Os intervalos arredor da media de amplitude 2 ou 4 veces a desviación típica, teñen moita importancia en estatística pola porcentaxe de datos que hai neles.

Desviación típica

Exemplo

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

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

Ou ben como $\bar{x} = 20,26$

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

Coeficiente de variación

É o cociente entre a desviación típica e a media; utilízase para comparar as dispersións de datos de distinta media.

Por exemplo, para os datos 4 e 6, o $CV = 1/5 = 0,2$ e para 101 e 99 é $CV = 1/100 = 0,01$. En ambos os dous casos a desviación típica é a mesma, pero en relación á media é mais importante no primeiro.

Coeficiente 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 a Calculadora científica

MODELO 1

Modo estatístico

Primeiro hase de elixir o modo estatístico. En moitas calculadoras faise pulsando:

[MODE][.]

Datos desordenados

A continuación hai que introducir os datos, por exemplo para 2, 3, 4, 3 teclearemos:

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

E para facer os cálculos:

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

Tamén se pode sumar todos, ou os cadrados, ou contar o n° de datos introducidos, premendo respectivamente:

[SHIFT][Σx] [SHIFT][Σx²] [SHIFT][n]

Datos nunha táboa

x_i f_i Se introducen os datos segundo a
2 4 secuencia: **[2][x][4][M+]**
3 3 **[3][x][3][M+]**
4 5 **[4][x][5][M+]**

E agora xa se poden realizar os cálculos como antes.

Nota: Hai moitos modelos de calculadoras, pero afortunadamente, todas son bastante parecidas. Trata de descubrir por ti mesmo o funcionamento da túa, se non coinciden con estas, consulta o manual ou pregunta ao teu profesor.

MODELO 2

Modo estatístico e introdución de datos

Eliximos o modo estatístico (mode stat 1-VAR) e aparécenos unha columna onde introducir datos, un tras outro, non importa que vaian desordenados. Se tivésemos unha táboa con frecuencias teriamos que activar as frecuencias (Setup frequency on) e encher as columnas. Despois do último dato pulsar AC.

Cálculos

Pulsando SHIFT STAT aparécenos un menú, **1:type**, **2:Data**, **3>Edit**, **4:Sum**, **5:Var**, **6:MinMax**. Coa opción **5:Var** accederemos a calcular a media, desviación típica e cantidade de datos. Coa opción **4:sum** as sumas que habitualmente necesitamos. Coa opción **6:MinMax** o mínimo e o máximo. E coa opción **2:Data** poderemos modificar os datos introducidos.



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	

	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	

EXERCICIOS resoltos

17. Calcula a media e a desviación típica en
 a) 200, 250 b) 175, 275

Solución:

$$\begin{aligned} \text{a) } \bar{X} &= \frac{250 + 200}{2} = 225 & \sigma &= \sqrt{\frac{(250 - 225)^2 + (200 - 225)^2}{2}} = \sqrt{\frac{25^2 + 25^2}{2}} = 25 \\ \text{b) } \bar{X} &= \frac{175 + 275}{2} = 225 & \sigma &= \sqrt{\frac{(175 - 225)^2 + (275 - 225)^2}{2}} = \sqrt{\frac{50^2 + 50^2}{2}} = 50 \end{aligned}$$

18. Calcula a media e a desviación típica en:
 a) 7, 5, 3, 2, 4, 5 b) 20, 25, 20, 22, 21

$$\begin{aligned} \text{a) } \bar{X} &= \frac{7 + 5 + 3 + 2 + 4 + 5}{6} = \frac{26}{6} = 4,33 \\ \sigma &= \sqrt{\frac{7^2 + 5^2 + 3^2 + 2^2 + 4^2 + 5^2}{6} - 4,33^2} = \sqrt{\frac{128}{6} - 18,75} = 1,59 \\ \text{b) } \bar{X} &= \frac{20 + 25 + 20 + 22 + 21}{5} = \frac{108}{5} = 21,6 \\ \sigma &= \sqrt{\frac{20^2 + 25^2 + 20^2 + 22^2 + 21^2}{5} - 21,6^2} = \sqrt{\frac{2350}{5} - 466,56} = 1,85 \end{aligned}$$

19. Cal das dúas distribucións anteriores presenta maior dispersión?

Solución:

As desviacións típicas son moi similares 1,59 e 1,85

Para comparar calculamos o coeficiente de variación:

$$\text{a) } CV = (1,59/4,33) \cdot 100 = 36,72\% \quad \text{b) } CV = (1,85/21,6) \cdot 100 = 8,56\%$$

Agora pódese apreciar que a dispersión é moito maior na distribución a).

20. Calcula a media e a desviación típica dos datos agrupados seguintes:

X_i	5	10	15	20	25	30
f_i	9	2	3	5	9	4

Solución:

X_i	f_i	$X_i \cdot f_i$	$(X_i - \bar{X})^2 \cdot f_i$	$f_i \cdot X_i^2$
5	9	45	1371,31	225
10	2	20	107,86	200
15	3	45	16,47	675
20	5	100	35,27	2000
25	9	225	527,56	5625
30	4	120	640,72	3600
	32	555	2699,21	12325

Calculamos previamente a media:

$$\bar{x} = \frac{555}{32} = 17,34$$

$$\text{1ª Forma de calculala } \sigma = \sqrt{\frac{2699,21}{32}} = 9,18$$

$$\text{2ª Forma } \bar{x} = 17,34 \quad \sigma = \sqrt{\frac{12325}{32} - 17,34^2} = 9,18$$

21. Cal é o coeficiente de variación da distribución anterior?

Solución: $CV = \frac{9,18}{17,34} \cdot 100 = 52,94\%$

Para practicar



- Cantas persoas supoñen unha mostra do 5% dunha poboación de 20.000 habitantes? E dunha de 1000 habitantes?.
- Dunha poboación de 30000 individuos estudouse varias características en 150 individuos. Que porcentaxe do total foi estudada?
- Un veterinario estuda as seguintes características nunha mostra de animais dunha granxa tipo de animal, peso, cor dos ollos, temperatura corporal, número de compañeiros e metros cadrados por animal. Indica de que tipo é cada unha destas variables estatísticas.
- Fai un recuento dos seguintes datos, un gráfico de sectores e outro de barras. Indica o 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
- Fai un recuento dos seguintes datos e 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 os seguintes datos en 10 grupos. Agrupa os mesmos datos, agora, 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 a media en cada caso:
 - 14, 16, 18
 - 24, 26, 28, 26
 - 1000, 1200, 1800, 2000
- Calcula a media dos seguintes 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 a media dos seguintes 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 a moda para os 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 a mediana, o primeiro e o segundo cuartil dos datos do exercicio anterior.
- Calcula a desviación media en cada caso:
 - 14, 16, 18
 - 34, 36, 38, 36
 - 1000, 1200, 1800, 2000
- Calcula o rango e a desviación media dos 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 a desviación media dos datos tabulados seguintes:

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 a media e a desviación típica en:

- a) 2000, 2500
- b) 1750, 2750
- c) 2500, 2500

16. Calcula a media e a desviación típica dos 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 o coeficiente de variación dos datos do exercicio anterior.

18. Calcula a media e a desviación típica dos 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 o coeficiente de variación dos datos do exercicio anterior.

20. Calcula a media e a desviación típica dos datos agrupados seguintes:

Xi	fi
5	7
10	0
15	2
20	2
25	4
30	2

21. Fai os cálculos para un millón de habitantes en cada comunidade.

Taxa de criminalidade. 2006

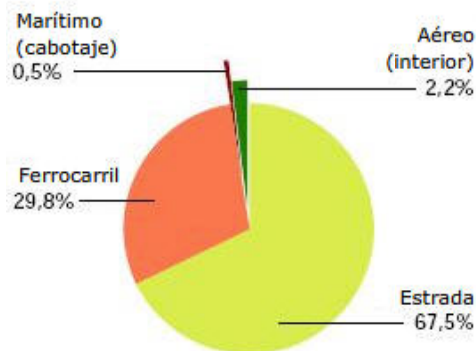
Infraccións penais por 1.000 hab.

Taxas máis 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 viaxeiros, cantos corresponden a cada sector?

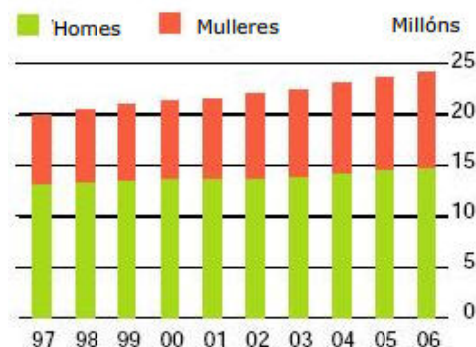
Viaxeiros que utilizan transporte interurbano. 2007



Fontes: INE, RENFE, FEVE, D. Xeral de Aviación Civil e D. Xeral de Portos e Costas

23. Cantos condutores había no ano 2002? Cantos eran homes e cantas mulleres?

Censo de condutores

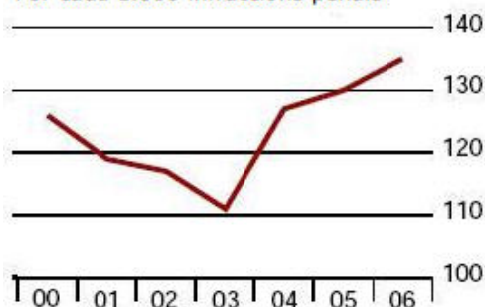


Fonte: Dirección Xeral de Tráfico

24. Entre que anos aumentaron máis os detidos por infraccións penais?

Taxa de detidos

Por cada 1.000 infraccións penais



Fonte: Ministerio do Interior

Para saber máis

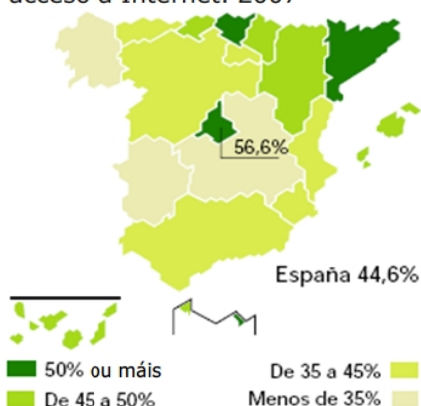


España en cifras

O Instituto Nacional de Estatística publica periodicamente datos como estes. Consulta a web. (www.ine.es)



Vivendas que dispoñen de acceso a Internet. 2007



Uso de TIC en fogares. 2007

	Fogares con acceso a Internet (%)	Persoas que compraron por Internet (%)
Paises Bajos	83	55
Suecia	79	53
Dinamarca	78	56
Luxemburgo	75	47
Alemania	71	52
Finlandia	69	48
Reino Unido	67	53
Belgica	60	21
Austria	60	36
Eslovenia	58	16
Irlandia	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 non dispoñible

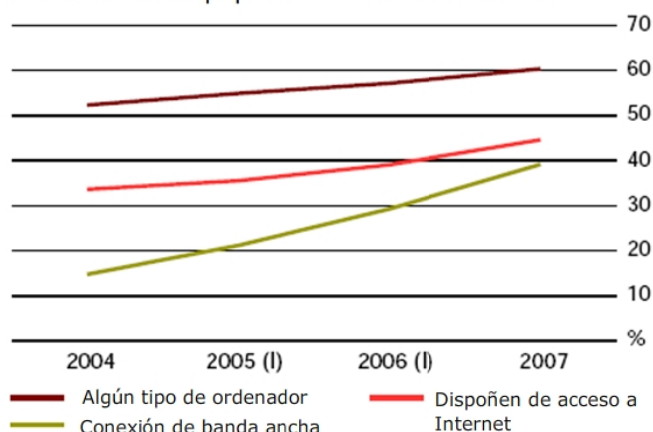
Fonte: Eurostat

Equipamento tecnolóxico dos fogares

Segundo datos da Enquisa sobre Equipamento e Uso das Tecnoloxías de Información e Comunicación nos Fogares, en 2007 o 99% dos fogares dispón de teléfono (xa sexa fixo ou móbil), o 99,5% dispón de televisión e o 22,8% ten recepción de televisión dixital terrestre (TDT). Nunha de cada cinco vivendas con televisión, algunha delas é de pantalla plana (plasma, LCD). O DVD áchase presente en tres de cada catro fogares, en detrimento do vídeo.

Un 60,4% dispón dalgún tipo de ordenador.

Evolución do equipamento TIC en vivendas



(I) Primeira enquisa bianual

A banda ancha gaña adeptos

6,5 millóns de vivendas familiares teñen acceso a internet (un 44,6% do total). O 39% dos fogares dispoñen de conexión mediante banda ancha (ADSL, rede de cable...), o que supón un incremento de 10 puntos respecto ao ano 2006.

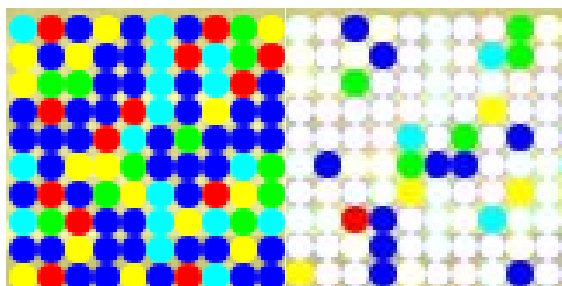
A maior porcentaxe de vivendas con acceso a internet dáse en Comunidade de Madrid (56,6%) e Cataluña (51,3%). A diferenza en puntos porcentuais entre sexos respecto aos principais indicadores de uso de TIC redúcese en 2007, tanto no uso de ordenador coma nos usuarios frecuentes; non obstante, aumenta no acceso a internet. O 60,5% dos varóns usaron internet algunha vez fronte a un 54,1% das mulleres. Na Unión Europea, os países que dispoñen de indicadores TIC superiores á media atópanse, principalmente, no norte de Europa. Por debaixo da media, sitúanse os países da área mediterránea ou os que son de recente ingreso.



**Lembra
o máis importante**

Poboación

Mostra

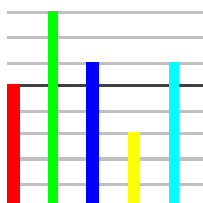


Nº de irmáns: 4 3 2 3 1 2 0 2 0 1 2 3 1 2 4 0 1
1 4 1 1 4 0 4 2 0 4 1

Reconto 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 e barras



Media e moda

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

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

Cuartís e mediana



$$M_e = 2, Q_1 = 1, Q_3 = 3$$

Rango.

De 0 a 4, de amplitude 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, cor preferida.
- Cuantitativa discreta, nº de irmáns.
- 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:

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

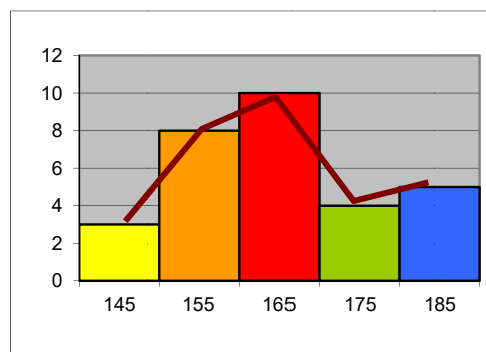
Varianza e Desviación típica:

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

Coefficiente de variación:

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

Histograma e polígono de frecuencias:



Autoavaliación



1. Realiza un recuento dos datos seguintes.

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

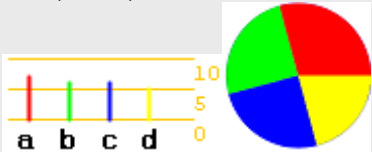
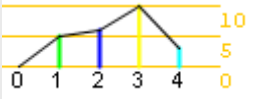
2. Realiza un gráfico de barras para os datos anteriores.

3. Calcula a media dos datos dados pola táboa.

X_i	f_i
1	11
2	3
3	5
4	5

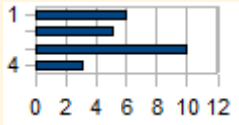
4. Calcula a mediana dos datos anteriores.
5. Calcula o primeiro cuartil dos datos do exercicio 3.
6. Calcula o terceiro cuartil dos datos do exercicio 3.
7. Calcula o rango dos datos do exercicio 3.
8. Calcula a desviación media dos datos anteriores.
9. Calcula a desviación típica dos datos do exercicio 3.
10. Calcula o coeficiente de variación para os datos do exercicio 3.

Soluções dos exercícios para praticar

- a) 400 b) 50
- 0,5%
- V. qualitativa: tipo animal, cor de olhos.
V. quantitativa discreta: número de companheiros.
V. quantitativa contínua: peso, temperatura e metros cadrados.
- a→7, b→6, c→6 e d→5

- 0→0, 1→5, 2→6, 3→10, 4→3

- En 10: 1→3, 2→6, 3→5, 4→5, 5→5, 6→13, 7→4, 8→2, 9→7 y 10→4 .
En 5: (1 ou 2)→ 9, (3 ou 4)→ 10, (5 ou 6)→ 18, (7 ou 8)→ 6 e (9 ou 10)→ 11
- a) $(14+16+18)/3 = 16$
b) $(24+26+28+26)/4 = 104/4 = 26$
c) $(1000+1200+1800+2000)/4 = 1500$
- $\bar{X} = 2,46$
- $\bar{X} = 9,77$
- $M_o = 3$, é o valor que mais vezes aparece, 10 em total.
- 1→5, 2→6, 3→10 e 4→3.
Mediana = 3, $Q_1 = 2$ e $Q_3 = 3$
- a) $(2+0+2)/3 = 1,3333\dots$
b) $(2+0+2+0)/4 = 1$
c) $(500+300+300+500)/4 = 1600/4 = 400$
- O rango oscila entre 1 e 40 cunha amplitude de 39. $\bar{X} = 20,58$, $DM = 8,92$

$$DM = \frac{2700}{12} = 225$$
- a) $\bar{X} = 2250$ $\sigma = 250$
b) $\bar{X} = 2250$ $\sigma = 500$
c) $\bar{X} = 2500$ $\sigma = 0$
- Media = 2,58 D. típica = 1,21
- $CV = 1,21/2,58 = 0,46$
- $\bar{X} = 21,08$ e $\sigma = 12,98$
- $CV = 0,62$
- $\bar{X} = 15,58$ e $\sigma = 9,68$
- 78800, 70800, 67500, 67400 y 65300
- Estrada 67500, aéreo 22000, marítimo 5000 e ferrocarril 298000
- Aproximadamente 22.000.000, dos que homes son 14.000.000 e mulleres 8.000.000

Soluções AUTOAVALIAÇÃO

- | | | | | |
|-------|---|---|----|---|
| x_i | 1 | 2 | 3 | 4 |
| f_i | 6 | 5 | 10 | 3 |
- 
 -
 - $\bar{x} = 2,17$
 - Me = 2
 - $Q_1 = 1$
 - $Q_3 = 3$
 - rango = 3
 - DM = 1,11
 - $\sigma = 1,24$
 - Cv = 0,571

Utilizar a Folla de Cálculo

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

Introdución de datos e primeiros cálculos

En xeral, as seguintes opcións están dispoñibles en todas as follas de cálculo, en particular na de Openoffice e en Excel.

✓ **Abre** unha folla de cálculo e introduce os datos 3, 4, 5, 3, 4, 6, 3, 3, 4 na columna A, en A1, A2, ... estes valores serán os datos sobre os que faremos os primeiros cálculos.

Os datos están na área "A1:A9" da folla de cálculo.

Se os datos estivesen en 3 filas x 3 columnas, empezando por A1, sería "A1:C3".

E por exemplo, se co rato selecciono as celas da imaxe adxunta, a área que cubro é "A2:C3".

	A	B	C
1	3		
2	4		
3	5		

Hai funcións para todas e cada un dos cálculos estatísticos que estudamos, e moitas máis que podes investigar pola túa conta. As funcións fan cálculos cos números da área que se lles proporciona.

Vexamos como se farían os cálculos para os datos que introducimos. Faremos os cálculos na columna C, a B usáremola para indicar o nome do cálculo que hai 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: Na cela C1 introduce a 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)** Recorda o que é a **área** dos datos, **A1:A9**

Mediana: =**mediana(área)**

Cuartil 1º. =**cuartil(área;1)**

Cuartil 3º. =**cuartil(área;3)**

Desviación media: =**desprom(área)**

Desviación típica: =**desvest(área)**

Exercicio

Introduce datos nunha área mais grande que a anterior, ao menos de 20 celas e realiza todos os cálculos de dúas formas, coa folla de cálculo e como o estudaches. Se todos os resultados son correctos esta folla váleche para calquera conxunto de datos dese tamaño.

ANEXO

Valores aleatorios.

Se en 20 celas escribimos **=aleatorio()** (podemos escribilo nunha e copialo nas outras), teremos 20 datos entre 0 e 1 aleatorios. Se queremos que sexan números comprendidos entre 0 e 20 escribiremos **=20*aleatorio()** E se queremos que non leven decimais, escribiremos **=entero(20*aleatorio())** Desta forma inventamos 20 datos para facer cálculos estatísticos.

Contar

Se temos introducidos unha cantidade grande de datos. Coa función **=contar(área)** dános a cantidade de números da área (imaxina que hai celas baleiras e non é doado contar as que están recheas de números).

Se o que temos é datos cualitativos (textos) usaremos **=contara(área)** para que nos conte as celas ocupadas da área indicada.

Agrupar datos.

- En **variable discreta**. Se nunha área da folia de cálculo temos 20 números enteiros entre 0 e 4, por exemplo. Como podemos contar cantos deles son un 4? Escribiremos nunha cela en branco **=contar.si(área;4)**. Para contarmos o resto usaremos na cela que nos conveña **=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 exemplo. Na imaxe da esquerda hai 20 datos xerados aleatoriamente entre 0 e 4. Na columna Marca indico que valores vou contar na seguinte columna, e na columna frecuencia uso a fórmula **'=contar.si(área;valor)'** escribindo:
=contar.si(área;0), **=contar.si(área;1)**,
=contar.si(área;2), ...

E finalmente comprobo, sumando as frecuencias que se contaron todos os datos (na cela D21 escribo **=suma(d15:d19)**).

Exercicio. Xera 20 datos aleatorios entre 0 e 4 e fai unha táboa de frecuencias usando a función **contar.si**. Calcula todos os parámetros estatísticos estudados (rango, máximo, mínimo, media ...).

- En **variable continua**. Se temos 100 alturas entre 150 e 200 centímetros nunha área dunha folia de cálculo e queremos saber cuántas están no intervalo [150,160) usaremos **=contar.si(área;">=150 ")-contar.si(área;"<160")**; para o intervalo [160,170) usaremos **=contar.si(área;">=160 ")-contar.si(área;"<170")** e así para cada un dos intervalos.

- En **variable cualitativa**. Se o que temos é 100 cores e queremos saber cuántas vermellas hai, escribese: **=contar.si(área;"vermello")**

Clasificar

Segundo o tipo de variable que empreguemos, usaremos a fórmula 'contar.si' sobre a área de datos tantas veces como sexa necesario para clasificar todos os datos e nos poida ser de utilidade. Así, para cores usaremos **=contar.si(área;"vermella")**, **=contar.si(área;"verde")**,... tantas veces como cores teñamos, de forma ordenada, nunha columna por exemplo.

Exemplo

Imos inventar o peso de 50 persoas con valores entre 50 e 110. Usaremos a fórmula $=\text{aleatorio}() * 60 + 50$, se queremos que sexan valores enteiros pon $=\text{entero}(\text{aleatorio}() * 50 + 50)$ e copiarémola na área **A1:E10**.

Agora hai que clasificar os datos. Como todos están entre 50 e 110 os intervalos terán que reflectilo. Imos realizar dúas táboas distintas, unha con valores de 20 en 20 e outra con valores de 10 en 10.

Para a primeira, copiaremos a seguinte táboa nun lugar baleiro da folia 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 a segunda, copiaremos esta táboa.

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")

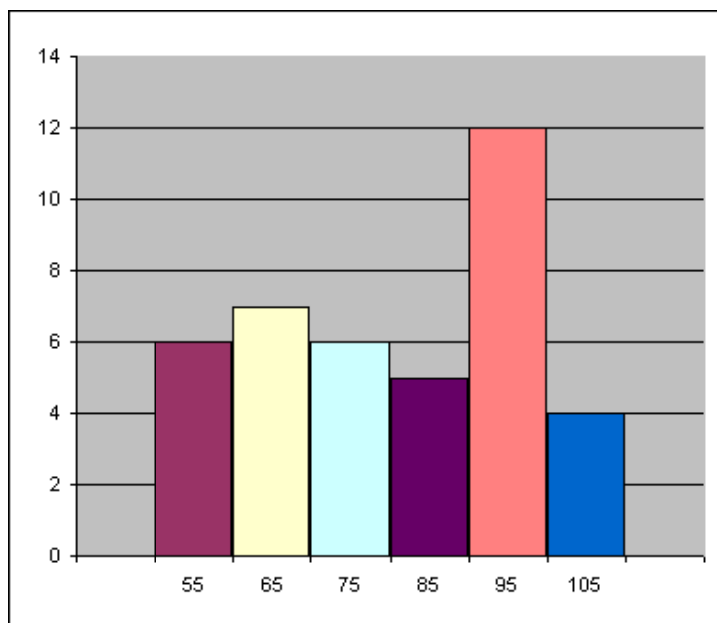
Exercicio

Xera 100 números de 0 a 80 e realiza dúas agrupacións en intervalos distintos. Por exemplo, unha de 10 en 10 e outra de 20 en 20.

Gráficos

Sinxelo e sen etiquetas. Abonda con seleccionar a área das frecuencias dos datos e crear o gráfico pulsando en gráficos do menú inserir. Despois podemos elixir o tipo de gráfico que queremos e o aspecto deste entre unha grande variedade de opcións que incorporan todas as follas de cálculo.

Con etiquetas: O máis sinxelo é utilizar a primeira fila ou primeira columna para as etiquetas. Nas opcións do gráfico haberá que activar e/ou desactivar a que corresponda coa forma de presentar os datos na folia.



Exercicio

Fai dous gráficos diferentes, un para cada unha das distribucións do exercicio anterior.