11 Estatística

Obxectivos

Nesta quincena aprenderás a:

- Distinguir os conceptos de poboación e mostra.
- Diferenciar os tres tipos de variables estatísticas.
- Facer recontos e gráficos.
- Calcular e interpretar as medidas estatísticas de centralización máis importantes.
- Calcular as principais medidas de dispersión.
- Entender a importancia da elección da mostra para que sexa representativa.
- Utilizar e representar variables bidimensionais.
- Calcular o centro de gravidade, a covarianza, o coeficiente de correlación e a recta de regresión nunha distribución bidimensional.

Antes de empezar

Lembra

O curso pasado xa estudaches estatística e en numerosas ocasións fixeches estatística aínda que non te deras contas diso. Vexamos algúns exemplos.

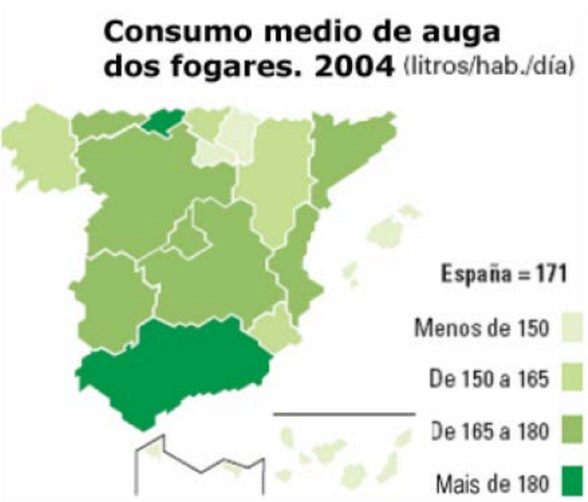
Nota media

Ao longo dun curso escolar terás moitas ocasións onde calcular este valor. Se unha nota depende de dous exames e nun tes un 4, intentarás sacar polo menos un 6 na outra.

Ao final do instituto, as medias do bacharelato e da proba selectividade. Comparacións coa media local ou nacional. As medias de corte para determinadas carreiras

Fútbol

O xogador que máis goles marcou, o porteiro que menos encaixou. A clasificación da liga. A mellor metade de liga. Os postos de competicións europeas, os de descenso, nº de veces internacional, nº de fases finais, minutos xogados, tiros a porta, faltas.



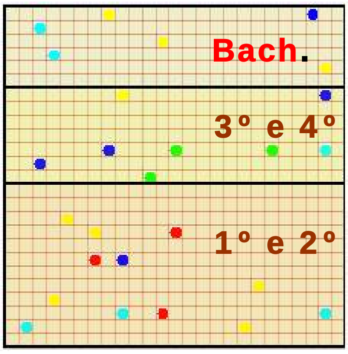
1. Estatística descritiva

Poboación e mostra.

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

Na práctica é frecuente ter que recorrer a unha mostra para inferir datos da poboación. A **mostra** é un subconxunto da poboación, seleccionada de modo que poña de manifesto as súas características, de aí que a propiedade máis importante das mostras é a súa representatividade.

O proceso seguido na extracción da mostra chámase **mostraxe**



Se cada cadriño representa cada un dos alumnos dun instituto ficticio e a cada cadriño se lle pregunta sobre a súa cor favorita, o total dos cadros é a poboación, 625 alumnos, e os 26 enquisados constitúen a mostra.

Variables estatísticas

A característica a estudar nunha poboación é a **variable estatística**.

As variables estatísticas poden ser esencialmente de dous tipos: **cualitativas e cuantitativas.**

As variables cualitativas son as que non aparecen en forma numérica senón como unha categoría ou atributo.

As variables cuantitativas son as que poden expresarse numericamente, e á súa vez poden ser:

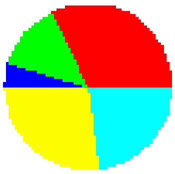
- ✓ Cuantitativas discretas, se só poden tomar un número finito de valores.
- ✓ Cuantitativas continuas cando poden tomar calquera valor dun intervalo.

- A cor dos ollos, o queixo preferido, o continente onde vives, son **variables estatísticas cualitativas**.
- O nº de ordenadores na casa, ou de televisores e o nº de habitantes por vivenda, por exemplo, son variables estatísticas **cuantitativas discretas.**
- O peso, a altura, a velocidade, a densidade, a presión, son **variables estatísticas cuantitativas continuas.**

Os datos:

Total **22**

Teñen este diagrama de sectores



Gráficos en variables cualitativas.

O diagrama de sectores é o mais indicado para este tipo de información. A porcentaxe de datos de cada valor nunha mostra correspóndese coa mesma porcentaxe de sector dun círculo. Así por exemplo, se os datos son A, A, A, A, A, B, B, B, C e C. As frecuencias son (A,5), (B,3) e (C,2), as porcentaxes serán (A,50%), (B,30%) e (C,20%) os que corresponde a un gráfico de sectores con (A, 180º), (B,108º) e (C, 72º).

$$\frac{\text{frecuencia}}{\text{nº total de datos}} = \frac{\text{graos do sector}}{360}$$

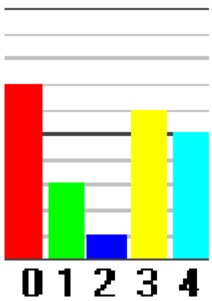
Gráficos en variables discretas.

Diagrama de barras. Abondará que observes un exemplo.

Aos datos,

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

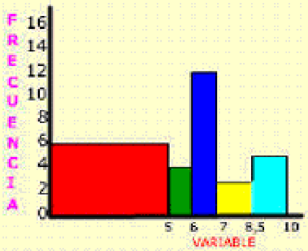
correspóndelles o gráfico da dereita.

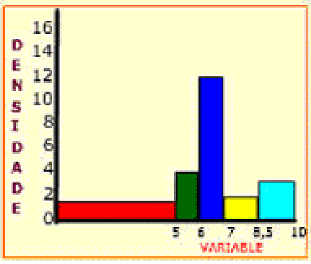


Intervalos	Reconto	fr.	Dens.
[0 5)	卌 I	6	1,2
[5 6)	IIII	4	4
[6 7)	卌 卌 II	12	12
[7 8,5)	III	3	2
[8,5 10)	卌	5	3,3

RECONTO DAS NOTAS EN 30 EXAMES

No diagrama de frecuencias a área maior corresponde á columna vermella que non é a de máis frecuencia





$$\text{Densidade} = \frac{\text{Frecuencia}}{\text{Lonxitude do intervalo}}$$

As **áreas** das barras-densidade resultan **proporcionais ás frecuencias** no intervalo.

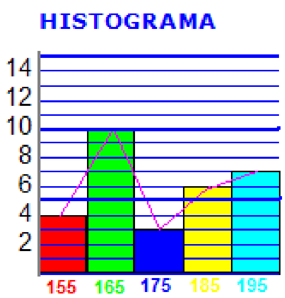
Gráficos en variables continuas.

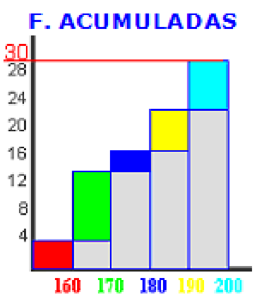
Histograma. Os datos represéntanse por rectángulos a base dos cales é a amplitude do intervalo representado e coa altura que nos indica a frecuencia absoluta, se todos os intervalos son da mesma amplitude. Se non é o caso, as alturas calcúlanse de maneira que as áreas sexan proporcionais ás frecuencias absolutas. Á esquerda tes un exemplo feito.

Polígono de frecuencias. Uniremos os centros da parte superior de todos os rectángulos para obtelo.

Tamén se adoita debuxar o histograma das **frecuencias acumuladas**, en cada dato acumúlase a frecuencia dos datos anteriores.

[150, 160]→4
[160, 170]→10
[170, 180]→3
[180, 190]→6
[190, 200]→7





EXERCICIOS resoltos

1. Clasifica as os seguintes exemplos de variables estatísticas: Lonxitude dun camión, Carga máxima, nº de rodas, nº de eixes, tipo de camión, marcas de pneumáticos, tipo de tapizaría, nº de portas, altura máxima.

Cualitativas: Tipo de camión, marcas de pneumáticos, tipo tapizaría

C. discretas: Nº de rodas, nº de eixes, nº de portas

C. continuas: Lonxitude dun camión, Carga máxima e altura máxima.

2. Calcula os graos que corresponden a cada valor nun gráfico de sectores feito a partir dos datos: R, R, V, V, V, V, V, A, A e A

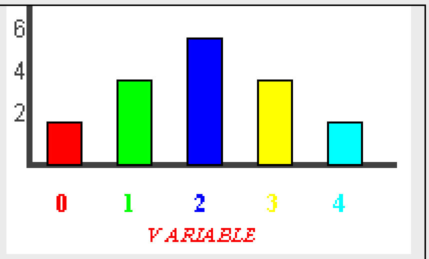
Facemos o reconto R→ 2, V→ 5 e A→3 e calculamos

$\frac{2}{10}=\frac{\text{Graos R}}{360}$, $\frac{5}{10}=\frac{\text{Graos V}}{360}$ e $\frac{3}{10}=\frac{\text{Graos A}}{360}$ e obtemos

Graos R =72, Graos V =180 e Graos A =108

3. Agrupa os datos seguintes e fai un diagrama de barras axeitado. Datos = { 0 1 0 2 3 4 1 2 2 1 2 2 3 4 3 2 1 3 }

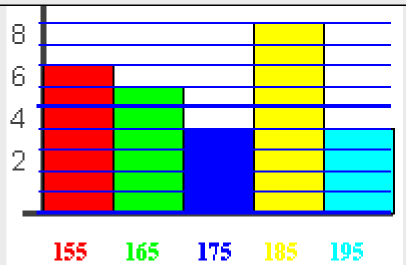
Marca	Frecuencia
0	2
1	4
2	6
3	4
4	2



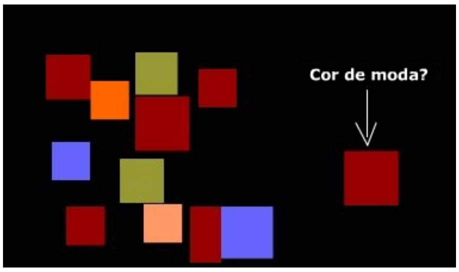
4. Clasifica os datos en intervalos e debuxa un histograma axeitado.

180	197	154	181	189	162	152	162	167	190
189	160	166	197	187	194	152	181	173	154
177	184	186	174	177	159	158	189	160	150

[150, 160]	155→ 7
[160, 170]	165→ 6
[170, 180]	175→ 4
[180, 190]	185→ 9
[190, 200]	195→ 4



2. Medidas de centralización



Media, mediana e moda.

Un conxunto de N observacións, N números, pode que por si só non nos diga nada. En cambio, se ademais nos din que están situados ao redor dun ou varios valores centrais xa temos unha referencia que sintetiza a información.

Media. A suma dos N números dividida entre N. Por exemplo, para **3**, **4** e **5**, (**3** + **4** +**5**)/3 =12/3 =4; para **1**, **1**, **4**, **8**, **8** e **8**, (**1**·2+**4**+**8**·3)/6=5.

$$\text{Media} = \frac{x_1f_1 + x_2f_2 + \ldots + x_nf_n}{N}$$

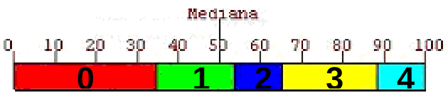
Moda. Se unha observación se repite máis que calquera outra, será considerada a moda deses datos. Por exemplo, se temos as observacións 6,7,8,6,7,6,8,6,9 e agrupamos os datos 6→4, 7→2, 8→2 e 9→1 vemos claramente que o valor 6 aparece mais que ningún outro. Neste caso a moda é 6.

No caso de variable continua, consideraremos por moda á marca do intervalo de maior frecuencia, cando isto aconteza. Tamén pode acontecer que haxa dúas modas ou que non haxa ningunha que destaque.

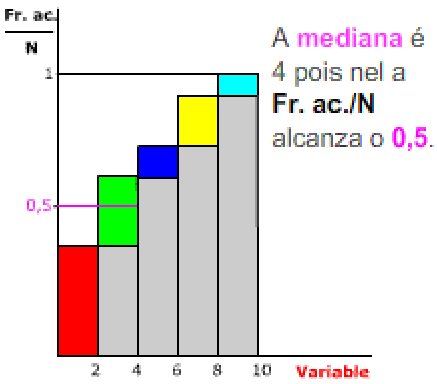
Mediana. O número tal que a metade das observacións son maiores que el e a outra metade menores.

En xeral, para poucos datos o mellor é proceder segundo o exemplo da esquerda, segundo sexa unha cantidade par ou impar.

Para cantidades maiores, haberá que agrupar os datos primeiro nunha táboa. E determinar segmentos de lonxitude proporcional á súa frecuencia, dispoñelos de forma lineal e marcar o centro como mostra o seguinte exemplo.



Neste outro gráfico vemos indicada a mediana nun diagrama de Frecuencias relativas acumuladas:



Por exemplo, se temos as observacións 6,7,8,6,7,6,8,6,9 e agrupamos os datos vemos claramente que o valor 6 aparece mais que ningún outro. Neste caso a **moda** é 6.

x → fr
6 → 4
7 → 2
8 → 2
9 → 1

Se ordenamos os datos, e dado que o nº de datos é impar xusto o 7 queda no centro.

6 6 6 6 7 7 8 8 1

Se os datos fosen 6,7,8,6,7,6,8,6,5 unha vez ordenados, e como hai unha cantidade par de datos, dous deles ocuparían o centro:

5 6 6 6 6 7 7 8 8 1

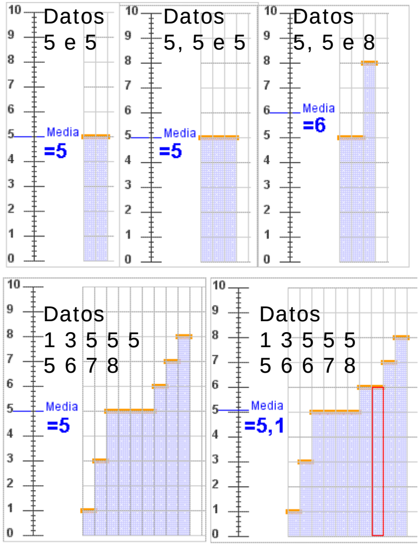
e a mediana será (6 +7)/2 =6.5

Media. Evolución ao engadir e/ou cambiar un dato

1 Para os datos 5 e 5 a media é 5. Se engadimos un 5 mantense en 5. Se engadimos un 8 a media pasa a ser 6. (Figura dereita).

2 Se temos 9 datos con media 5, necesitamos engadir un 6 para que a media pase a ser 5,1. Se temos 19 datos con media 5, necesitan un dato de valor 7 para que a media suba 5,1 (Figura dereita).

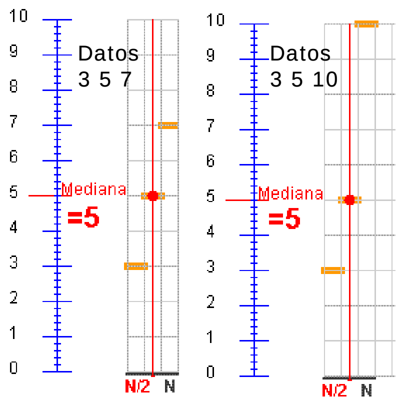
3 Para un conxunto de datos con media 5, se engadimos outro con media 5, por exemplo 6 e 4, o novo conxunto conserva a media.



Mediana. Evolución ao engadir e/ou cambiar un dato

1 A mediana, para os datos 2, 3 e 4 é Me=3. Se cambiamos o 4 por 5 ou por 6 ou por calquera outro valor maior segue sendo Me =3.

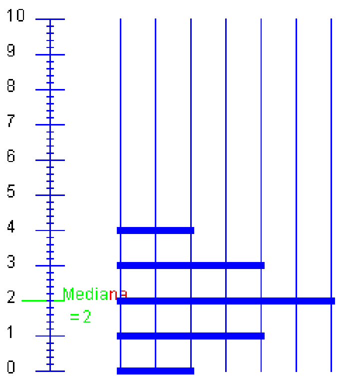
2 En cambio, se engadimos outro dato e temos 2, 3, 4 e 4, por exemplo, a Me =3,5. E se agora engadimos un quinto valor, un 4 ou un 5 ou un 6 ou calquera outro maior que 4, a mediana en 2,3, 4, 4 e ?? pasa a ser 4. Da igual que o valor ?? sexa 5, 10 ou 25**.**



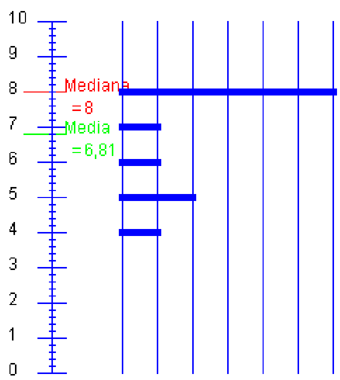
Para ver a mediana trázase unha vertical dende o eixe horizontal en N/2

Media e mediana comparadas

Para os datos 4 e 6 a media e a mediana coinciden en 5. Engadir un 8 ou un 11 da o mesmo para a mediana, que pasa a ser en ambos os dous casos 6. Non obstante a media cun 8 pasa a ser 6 e cun 11 pasa a ser 7. Os valores 8 e 11 considéranse observacións atípicas, están distanciados do resto de valores, tiran da media e non afectan á mediana.
Se os datos estivesen repartidos simetricamente respecto a un valor, ese valor sería á vez a media e a mediana. En cambio, se os valores a un lado da mediana están máis afastados dela que os do outro lado, a media desprázase cara a eses valores afastados que tiran dela. Hai unha asimetría.



Datos simétricos
Mediana igual á Media



Datos atípicos
Mediana distinta da Media

Por exemplo, se temos as observacións

1. 20, 24 e 28.

   Me =24

2. E para 20, 24, 28 e 30

   Me = (24 +28) / 2 =26

3. Para 20, 24, 28 e 100

   Me = (24 +28) / 2 =26

En cambio a media non se comporta da mesma forma para os mesmos datos

1 $\bar{X} = 24$

2 $\bar{X} = 25{,}5$

3 $\bar{X} = 43$

EXERCICIOS resoltos

5. Calcula a media en cada caso:

a) 4, 6, 8 Solucións: a) (4+6+8)/3 = 6

b) 4, 6, 8, 6 b) (4+6+8+6)= 24/4 = 6

c) 100, 120, 180, 200 c) (100+120+180+200)/4 = 150

6. Calcula a media en cada caso:

a

Marca	Fr
10	2
20	4
30	3
40	2

b

Marca	Fr
100	2
200	4
300	3
400	2

a) $\overline{X} = \dfrac{10\cdot 2 + 20\cdot 4 + 30\cdot 3 + 40\cdot 2}{11} = 24{,}54$

b) $\overline{X} = \dfrac{100\cdot 2 + 200\cdot 4 + 300\cdot 3 + 400\cdot 2}{11} = 245{,}45$

7. Determina a moda e a mediana

a) 5,6,6 c) 1,2,3,4,2 Solucións: a) Me=6, Mo=6 c) Me=2 Mo=2

b) 1,1,2,3 d) 3,2,3,2,2,2 b) Me=1,5 Mo=1 d) Me=2 Mo=2

8. Calcula a moda e a mediana en cada caso:

a

Marca	Fr
10	2
20	4
30	3
40	2

b

Marca	Fr
100	2
200	3
300	4
400	1

Solucións:

a) Me=20 Mo=4

b) Me=250 Mo=300

9. Medíronse as alturas en cm dun grupo de 30 persoas obténdose os datos seguintes:

Altura en cm	fi
(150,160]	7
(160,170]	9
(170,180]	10
(180,190]	3
(190,200]	1

Calcula a media, a moda e a mediana.

a) Completamos a táboa engadindo unha columna para x_i e outras dúas para $x_i \cdot fi$ e para as frecuencias acumuladas.

Altura en cm	x_i	f_i	$x_i \cdot fi$	F_i	
(150,160]	155	7	1085	7	
(160,170]	165	9	1485	16	← Me
(170,180]	175	10	1750	26	← Mo
(180,190]	185	3	555	29	
(190,200]	195	1	195	30	
	SUMA:	30	5070		

$\overline{x} = \dfrac{5070}{30} = 169$ Me $=165$ Mo $=175$

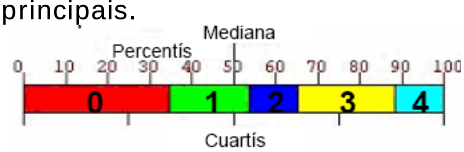
3. Medidas de posición

Cuartís e percentís

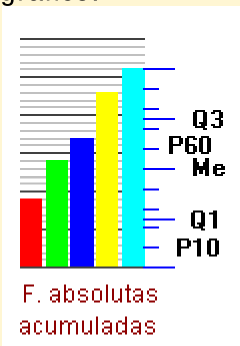
Dado un conxunto de datos numéricos correspondentes a un estudo estatístico, se os ordenamos de forma crecente e consideramos que estea no centro, estarémonos a fixar na **mediana**. É o primeiro que supera (ou iguala) ao 50% de valores, pero tamén podemos fixarnos noutras posicións:

- Se nos fixamos no primeiro valor que supera ao 25% ou ao 75%, estamos falando do **primeiro** e **terceiro cuartil, Q_1 e Q_3**.
- Para outros valores como o 10%, ou o 80% falamos de **percentís, P_{10} e P_{80}**.

Exemplo. Para a variable de valores 0, 1, 2, 3, 4, e frecuencias 0→9, 1→5, 2→3, 3→6, 4→3, debuxamos barras de lonxitude proporcional ás frecuencias e dividimos o total en partes iguais: en dúas partes para a mediana, catro para os cuartís e 10 para os percentís principais.



Tamén podemos facer un diagrama de frecuencias acumuladas e dividir en partes iguais como mostra o gráfico.

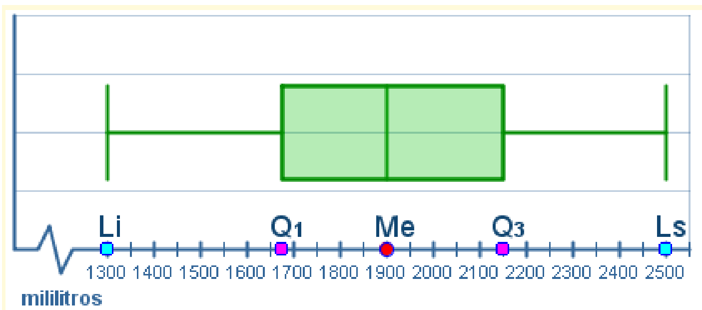


Diagramas de caixa e bigotes

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.

mín=1300 Q_1=1675 Me=1900 Q_3=2150 máx=2500



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

A táboa mostra o consumo diario de auga, en ml, dos 20 alumnos dunha clase.

Xoan	1650
Luis	1300
Alma	2400
Toño	2000
Rosa	2100
Lupe	1700
Paco	1900
Tere	1500
Iris	1900
Pepe	1850
Marco	2000
Lisa	2200
Xulio	2300
Maia	1600
Alex	1900
Beto	2500
Rita	2200
Marta	1650
Omar	2100
David	1750

Luís	1300	Min
Tere	1500	
Maia	1600	
Marta	1650	
Xoan	1650	Q1
Lupe	1700	
David	1750	
Pepe	1850	
Alex	1900	
Iris	1900	Me
Paco	1900	
Marco	2000	
Toño	2000	
Omar	2100	
Rosa	2100	Q3
Lisa	2200	
Rita	2200	
Xulio	2300	
Alma	2300	
Beto	2500	Max

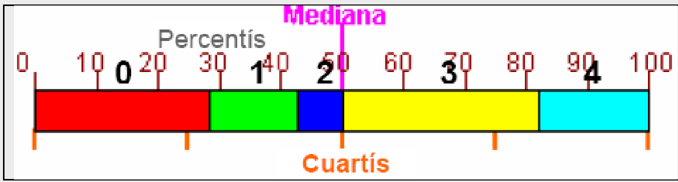
NOTA: A lonxitude dos bigotes non debe exceder unha vez e media a da caixa, se hai valores extremos que superan esa medida debúxanse como puntos illados.

EXERCICIOS resoltos

10. Calcula a mediana, cuartís primeiro e 3º, e o percentil 30º, 60º e 90º dos datos.

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

Facemos o reconto: 0→8, 1→4, 2→2, 3→9 e 4→5 e barras de lonxitude proporcional á frecuencia para cada valor. Ademais partimos a lonxitude total da barra en 2, 4 e 10 anacos para obter a mediana, cuartís e percentís, tal e como mostra a imaxe.

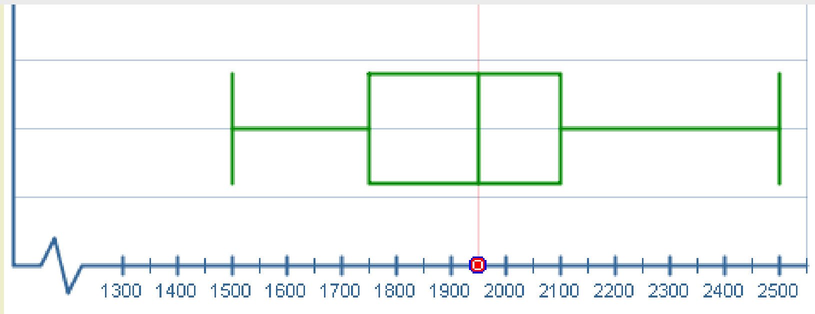


Vemos que a mediana está entre o azul e o amarelo, (3 +2)/2 =2.5, Q_1 no vermello, Q3 en amarelo.

$Q_1=0$ Me=2,5 $Q_3=3$

$P_{30}=1$ $P_{60}=3$ $P_{90}=4$

11. 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

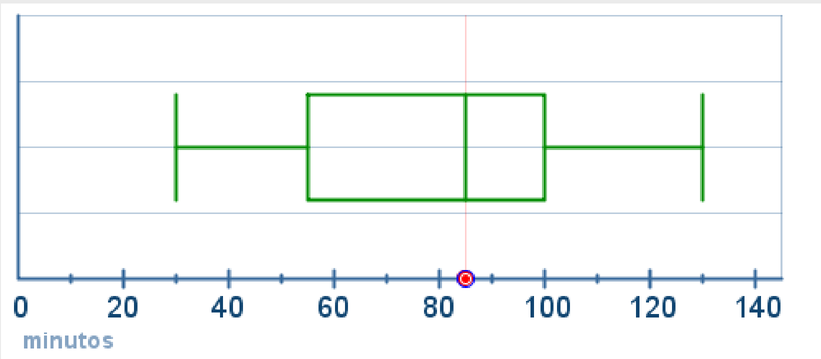
Q_1 = 1750

Me =1950

Q3 = 2100

Máximo =2500

12. 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

Q3 = 100

Máximo =130

a) A que porcentaxe da poboación fixera efecto ao cabo de 30 minutos?.
b) Ao cabo de cantos minutos fixera efecto ao 50% da poboación?.
c) Cantos minutos tardou en facer efecto ao 100% da poboación?
d) A que porcentaxe fixera efecto aos 55 minutos?.
e) 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, ¾ partes son o 75%

4. Medidas de dispersión.

Varianza, Desviación típica e rango

"A estatística é unha ciencia segundo a cal, se eu me como un polo e ti non te comes ningún, comemos como media medio polo cada un".

A estatística indicará que todos comen o mesmo cando as medidas de dispersión sexan todas nulas.

Rango. O intervalo definido polo menor e o maior dato. Tamén se chama rango á diferenza entre o maior e o menor dos datos.

Varianza. A media aritmética dos cadrados das diferenzas dos datos coa media.

$$\sigma^2 = \frac{\sum fi\cdot(Xi-\bar{X})^2}{n} \text{ que equivale a } \sigma^2 = \frac{\sum fi\cdot(Xi)^2}{n} - (\bar{X})^2$$

Desviación típica. A raíz cadrada positiva da varianza.

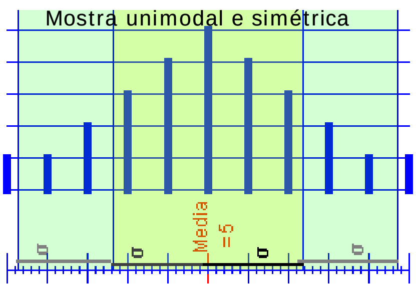
$$\sigma = \sqrt{\frac{\sum fi\cdot(Xi-\bar{X})^2}{n}} \quad \text{o } \sigma = \sqrt{\frac{\sum fi\cdot Xi^2}{n} - \bar{X}^2}$$

Medir a dispersión

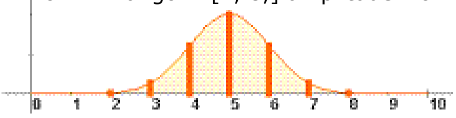
Ese é o obxectivo destas medidas. Por exemplo, os datos A= {20, 20}, B={15, 20, 20, 25} teñen a mesma media, moda e mediana. En todos os casos igual a 20. Non obstante, podes comprobar que en ningunha das tres medidas de dispersión definidas arriba coinciden.

Media e desviación típica.

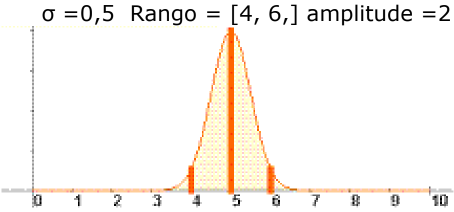
Para mostras unimodais (unha soa moda) e case simétricas, arredor da media podemos considerar un intervalo que conteña a maioría dos datos. Por exemplo, para unha mostra con media 100 e desviación típica 10, a maior parte dos datos estarán entre 90 e 110, aproximadamente o 68%; entre 80 e 120 estará o 95% aproximadamente. E case todos entre 70 e 130. Hai unha forma de distribución de datos chamada **normal** que cumpre co anterior, e dun xeito ou doutro, de todas as poboacións grandes se poden extraer datos que se axustan a ela. En cursos superiores verás a importancia destas distribucións.



σ =1 Rango = [2, 8,] amplitude =6



σ =0,5 Rango = [4, 6,] amplitude =2



En ambos os dous gráficos a media, mediana e moda valen 5

Na práctica adóitase usar a fórmula reducida para o cálculo da desviación típica.

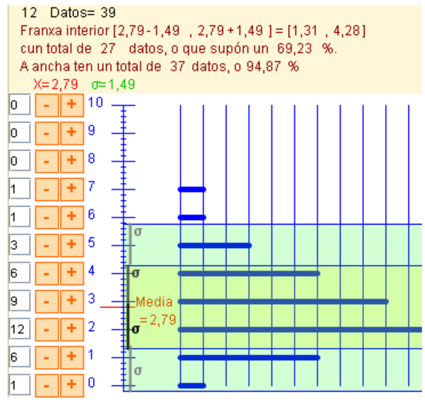
$$\sigma = \sqrt{\frac{\sum fi\cdot Xi^2}{n} - \bar{X}^2}$$

Así, para

Marca	Fr
4	3
5	3
6	2

Tense que a media $\bar{X} = 4,85$ e

$$\sigma = \sqrt{\frac{3\cdot4^2 + 3\cdot5^2 + 2\cdot6^2}{8} - 4,85^2}$$



EXERCICIOS resoltos

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

a) $\overline{X} = \dfrac{250+200}{2} = 225 \quad \sigma = \sqrt{\dfrac{(250-225)^2+(200-225)^2}{2}} = \sqrt{\dfrac{25^2+25^2}{2}} = 25$

b) $\overline{X} = \dfrac{175+275}{2} = 225 \quad \sigma = \sqrt{\dfrac{(175-225)^2+(275-225)^2}{2}} = \sqrt{\dfrac{50^2+50^2}{2}} = 50$

c) $\overline{X} = \dfrac{250+250}{2} = 250 \quad \sigma = \sqrt{\dfrac{(250-250)^2+(250-250)^2}{2}} = \sqrt{\dfrac{0^2+0^2}{2}} = 25$

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

a) $\overline{X} = \dfrac{7+5+3+2+4+5}{6} = \dfrac{26}{6} = 4,33$

$$\sigma = \sqrt{\frac{7^2+5^2+3^2+2^2+4^2+5^2}{6} - 4,33^2} = \sqrt{\frac{128}{6} - 18,75} = 1,59$$

b) $\overline{X} = \dfrac{20+25+20+22+21}{5} = \dfrac{108}{5} = 21,6$

$$\sigma = \sqrt{\frac{20^2+25^2+20^2+22^2+21^2}{5} - 21,6^2} = \sqrt{\frac{2350}{5} - 466,56} = 1,85$$

(Nota. - Observa a fórmula utilizada para a desviación)

15. Organiza os datos seguintes en intervalos de 10 cm dende 150 a 200. Amplía a táboa con dúas columnas, unha para o produto das marcas coas frecuencias e outra para o produto das frecuencias cos cadrados das diferenzas coa media. Calcula a media e a desviación típica.

174	158	150	185	186	178	166	185	199
183	175	173	175	164	173	178	179	164
176	159	190	173	189	163	156	169	

	x_i	f_i	$x_i f_i$	$f_i \cdot (x_i - \overline{X})^2$
[150,160)	155	5	775	1733,65
[160,170)	165	5	825	371,58
[170,180)	175	10	1750	19,02
[180,190)	185	7	1295	906,42
[190,200)	195	2	390	914,14
Total		29	5035	3944,82

Cos datos da táboa é mais doado, e tense:

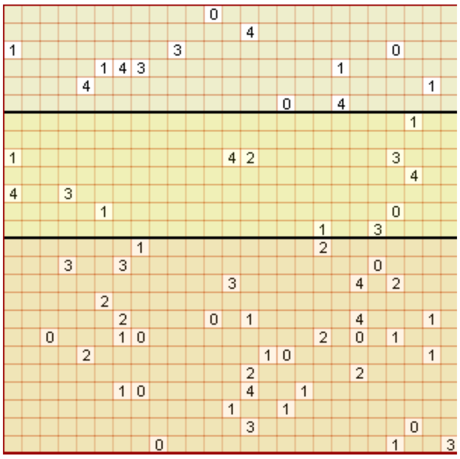
Media e **Desviación típica**

$\overline{X} = \dfrac{5035}{29} = 173,62 \quad \sigma = \sqrt{\dfrac{3944,82}{29}} = 11,66$

5. Representatividade

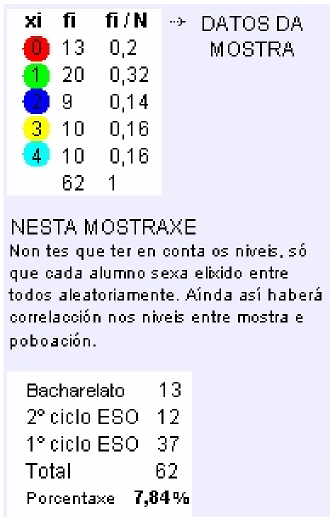
Mostraxe aleatoria

A característica máis importante dunha mostra é a súa **representatividade** respecto ao estudo estatístico que se estea a facer. Se a mostra non é representativa diremos que está **sesgada**.



O proceso mediante o cal se elixe unha mostra chámase **mostraxe**, e para que nos proporcione unha mostra representativa debe ser aleatorio. Unha mostraxe é **aleatoria** cando os individuos da mostra se elixen ao chou, de forma que todos teñen a mesma probabilidade de ser elixidos.

Exemplo: Chamadas telefónicas voluntarias. Estas enquisas teñen varias fontes de sesgo. Hai familias que non teñen teléfono, o custo da chamada non todo o mundo está disposto a asumilo. Pero sobre todo, o factor de resposta voluntaria, os enquisados autoselecciónanse. Adoitan contestar aqueles cunha forte opinión negativa sobre o tema. O anoxo anímalles a participar.



xi	fi	fi / N
0	13	0,2
1	20	0,32
2	9	0,14
3	10	0,16
4	10	0,16
	62	1

DATOS DA MOSTRA

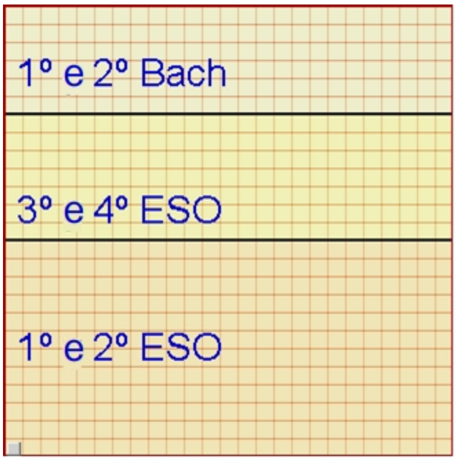
NESTA MOSTRAXE

Non tes que ter en conta os niveis, só que cada alumno sexa elixido entre todos aleatoriamente. Aínda así haberá correlacción nos niveis entre mostra e poboación.

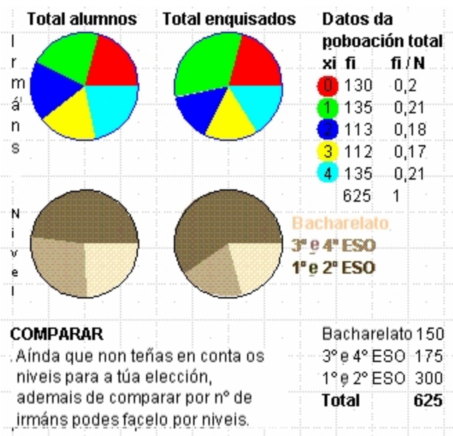
Bacharelato	13
2° ciclo ESO	12
1° ciclo ESO	37
Total	62
Porcentaxe	7,84 %

Exemplo

Na imaxe tes 625 cadros que representan os alumnos dun instituto ficticio, quérese estudar o "número de irmáns" e para iso elixiuse unha mostra aleatoria como podes ver a dereita.



Faino así: Decide primeiro o tamaño da mostra, por exemplo 62 alumnos, ordenados os alumnos elíxese un deles ao chou (podes simulalo elixindo un cadriño cos ollos pechados), a partir deste conta e sinala cada 10 cadriños (625/62≈10), cando chegues ao final da lista (cadrado) segue dende o principio. Este tipo de mostraxe aleatoria chámase **sistemática**.



Mostraxe estratificada

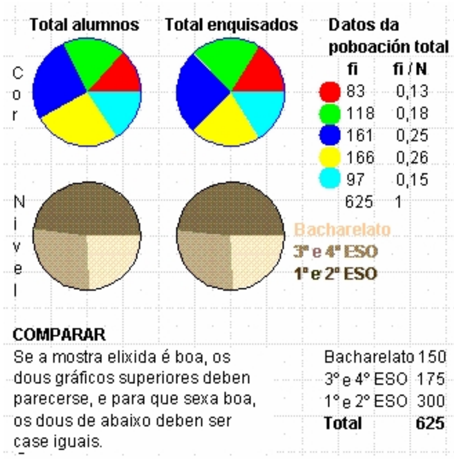
En ocasións cando a poboación obxecto de estudo, pertence a distintos grupos ou estratos convén elixir a mostra de forma que todos eles queden representados.

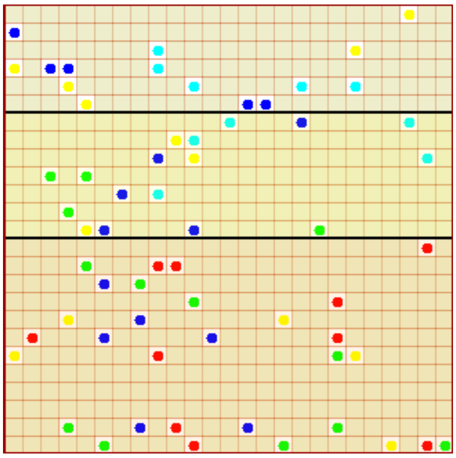
Este tipo de mostraxe, escollendo unha repartición proporcional aos estratos, chámase **estratificada**.

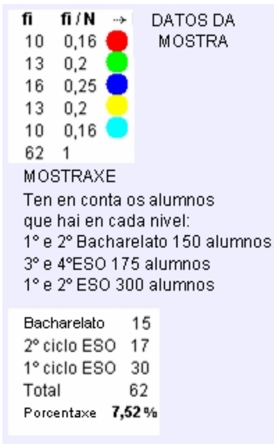
Por exemplo, se queremos estudar o poder adquisitivo dunha poboación, e solo eliximos individuos dunha determinada zona, ou principalmente dunha determinada zona, a mostra con toda seguridade non será representativa. A mostra hase de elixir tomando mostras de individuos proporcionais á poboación de cada zona. Se hai tres zonas con 12.000, 18.000 e 20.000 habitantes, a mostra deberá ter un 24% da primeira zona, 36% da segunda e 40% da última.

A continuación sobre a poboación do instituto ficticio anterior fíxose unha enquisa sobre a cor preferida e neste caso decidiuse facer estratificada. De cada nivel seleccionouse aleatoriamente un número de individuos proporcional ao número de compoñentes.

Debaixo vemos a mostra aleatoria que se elixiu e o resultado da enquisa. Os últimos diagramas de sectores comparan a realidade cos resultados da enquisa.







EXERCICIOS resoltos

16. Unha grande empresa ten traballadores en catro áreas. Operarios, representantes, administración e dirección. As condicións de traballo son bastantes diferentes en cada área, polo que o grao de satisfacción non é igual en cada unha delas. Para descubrilo, se hai 1000, 500, 300 e 200 traballadores nas áreas de operarios, representantes, administrativos e directivos, cantos hai que seleccionar de cada área para unha mostra de tamaño?

a) 200 b) 100 c) 300

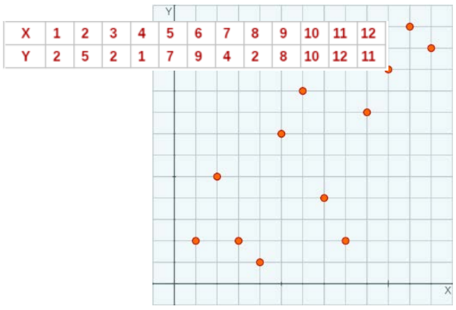
a) Dun total de 2000 empregados, as porcentaxes para operarios, repartidores, administrativos e directivos son do 50%, 25%, 15% e 10%. O cal fai que a mostra tome 100 operarios, 50 repartidores, 30 administrativos e 20 directivos.
b) 50, 25, 15 e 10.
c) 150, 75, 45 e 30

6. Estatística bidimensional

Distribucións bidimensionais

Unha distribución bidimensional é aquela na que interveñen dúas variables X e Y que poden ou non estar relacionadas.

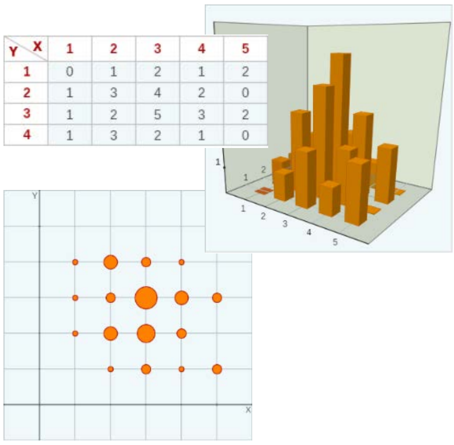
Podemos representar conxuntamente as dúas variables nun **diagrama de dispersión** ou **nube de puntos**, simplemente asociando un punto do plano a cada par (x_i, y_i).



X	1	2	3	4	5	6	7	8	9	10	11	12
Y	2	5	2	1	7	9	4	2	8	10	12	11

Cando hai moitos datos e os pares de valores se repiten, acudimos a unha **táboa de continxencia**, como a da dereita. Neste caso a representación gráfica faise mediante un gráfico tridimensional chamado prismograma, ou máis polo miúdo poñendo puntos de tamaño proporcional á frecuencia.

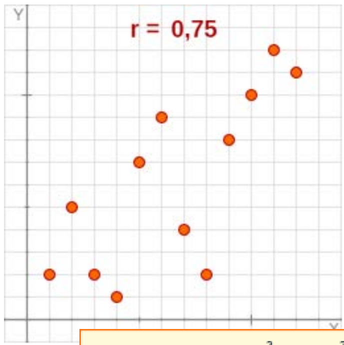
Y \ X	1	2	3	4	5
1	0	1	2	1	2
2	1	3	4	2	0
3	1	2	5	3	2
4	1	3	2	1	0



Correlación lineal

O obxectivo dun estudo bidimensional é observar se hai algún tipo de relación entre as dúas variables. Esta relación, que chamaremos **correlación**, pódese apreciar vendo se a nube de puntos se achega ou non á gráfica dunha función, no noso caso a unha recta, por iso falaremos de correlación lineal.

Canto máis se achegue a nube de puntos a unha recta máis forte será a correlación lineal, ademáis será positiva ou directa se a recta é crecente (se crece X crece Y) e negativa ou inversa en caso contrario (se crece X decrece Y ou viceversa).



Para cuantificar esta relación empregaremos un parámetro **r**, o **coeficiente de correlación lineal**, que toma valores entre -1 e 1. Canto máis se achege a valer 1 ou -1 máis forte será a correlación.

- Se **r=1** ou **r=-1**, hai dependencia funcional, os puntos están sobre unha recta.
- Se **0,5<r<1** consideraremos que a correlación é forte e directa ou forte e inversa se **-1<r<-0,5**.
- Se **r=0** ou moi próximo a 0, non hai correlación lineal entre as dúas variables.

Dispoñemos os datos en columnas e calculamos a **media** e a **desviación típica** das dúas distribucións X e Y. Para calcular r temos que calcular antes outro parámetro, a **covarianza**, definida como:

$$\sigma_{XY} = \frac{\sum(x_i - \bar{x})\cdot(y_i - \bar{y})}{n} = \frac{\sum x_i \cdot y_i}{n} - \bar{x}\cdot\bar{y}$$

E o **coeficiente de correlación lineal**:

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \qquad (-1 \le r \le 1)$$

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	2	1	4	2
2	5	4	25	10
3	2	9	4	6
4	1	16	1	4
5	7	25	49	35
6	9	36	81	54
7	4	49	16	28
8	2	64	4	16
9	8	81	64	72
10	10	100	100	100
11	12	121	144	132
12	11	144	121	132
78	**73**	**650**	**613**	**591**

$\bar{x} = \frac{78}{12} = \mathbf{6{,}5}$ $\quad \sigma_x = \sqrt{\frac{650}{12} - 6{,}5^2} = \mathbf{3{,}45}$

$\bar{y} = \frac{73}{12} = \mathbf{6{,}08}$ $\quad \sigma_y = \sqrt{\frac{613}{12} - 6{,}08^2} = \mathbf{3{,}75}$

$$\sigma_{xy} = \frac{591}{12} - 6{,}5 \cdot 6{,}08 = \mathbf{9{,}71}$$

$$r = \frac{9{,}71}{3{,}45 \cdot 3{,}75} = \mathbf{0{,}75}$$

Rectas de regresión

Cando se aprecia un certo grao de correlación entre as dúas variables dunha distribución bidimensional, búscase a recta que mellor se axusta á nube de puntos.

A recta de regresión de Y sobre X pasa polo punto $(\overline{x}, \overline{y})$, centro de gravidade da nube de puntos. A sua ecuación é:

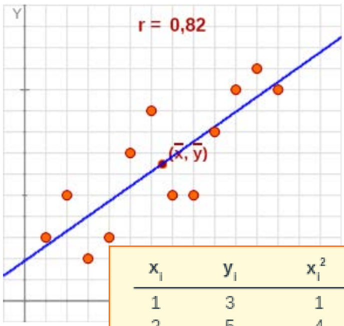
$$\boldsymbol{y = \overline{y} + \frac{\sigma_{XY}}{\sigma_X^2}(x - \overline{x})}$$

A pendente $\frac{\sigma_{XY}}{\sigma_X^2}$ é o **coeficiente de regresión**.

Esta recta de regresión serve para estimar o valor que tomará la variable Y para un determinado valor de X. O valor desta estimación será tanto máis fiable canto:

- Máis se achegue o coeficiente de correlación a 1 ou a -1.
- O valor quede dentro do rango de valores de X e máis cerca esté do centro de gravidade.

Se se quere estimar valores de X a partires dos de Y hai que utilizar outra recta análoga, a recta de regresión de X sobre Y.



x_i	y_i	x_i^2	$x_i \cdot y_i$
1	3	1	3
2	5	4	10
3	2	9	6
4	3	16	12
5	7	25	35
6	9	36	54
7	5	49	35
8	5	64	40
9	8	81	72
10	10	100	100
11	11	121	121
12	10	144	120
78	**78**	**650**	**608**
$\overline{x} = 6{,}5$	$\overline{y} = 6{,}5$	$\sigma_X = 3{,}45$	$\sigma_{XY} = 8{,}42$

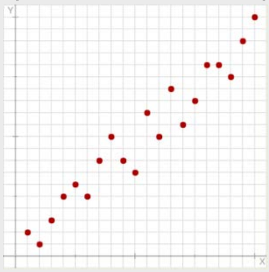
Recta de regresión de Y sobre X

$$y = 6{,}5 + \frac{8{,}42}{3{,}45^2}(x - 6{,}5)$$

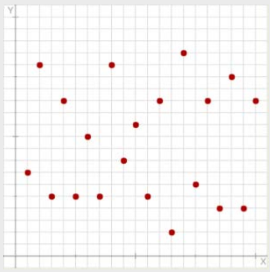
$$\mathbf{y = 0{,}7x + 1{,}9}$$

EXERCICIOS resoltos

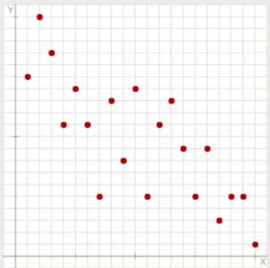
17. En vista da nube de puntos indica se cres que a correlación é moi forte, forte (directa ou inversa), débil ou moi débil.



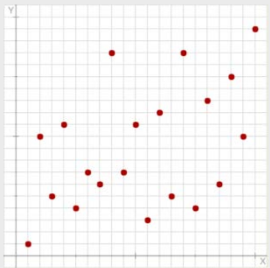
Moi forte e directa



Moi débil



Forte e inversa

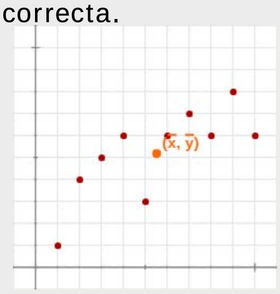


Débil

18. Unha das cuatro ecuacións corresponde á recta de regresión de Y sobre X da nube de puntos. Indica a correcta.

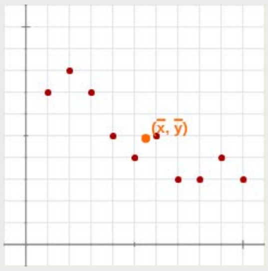
1. y = 0,7x + 2,4
2. y = -0,7x + 10,2
3. y = 0,7x + 10,2
4. y= -0,7x + 2,4

Sol: (b)



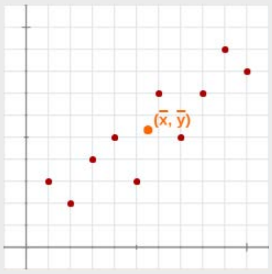
1. y = -0,5x + 2,5
2. y = 0,5x + 8
3. y = -0,5x + 8
4. y= 0,5x + 2,5

Sol: (d)



1. y = -0,5x + 2,2
2. y = 0,5x + 2,2
3. y = -0,5x + 7,7
4. y= 0,5x + 7,7

Sol: (c)



1. y = -0,7x + 1,4
2. y = 0,7x + 9,2
3. y = -0,7x + 9,2
4. y= 0,7x + 1,4

Sol: (d)

Algúns dos exercicios propostos a continuación están elaborados a partir desta publicación de INE. Podes ver artigos similares en
http://www.ine.es/prodyser/pubfolletos.htm

4/2007

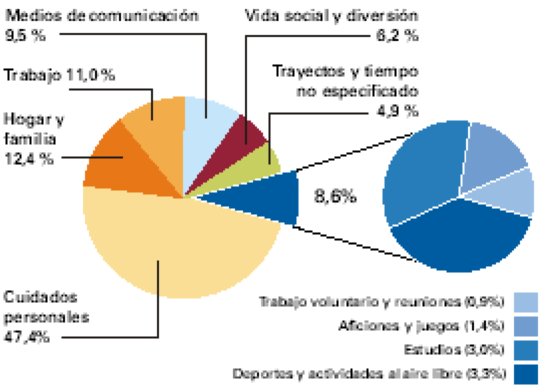
cifras INE

Boletín Informativo del Instituto Nacional de Estadística

Encuesta de Empleo del Tiempo

Qué hacemos y durante cuánto tiempo

Distribución del tiempo por actividades



NOTA: Los informantes de 10 y más años han anotado las actividades realizadas en un día concreto (de lunes a domingo) elegido al azar. El tiempo así estimado se refiere a un 'día promedio' obtenido al concentrar todas las actividades de todos los informantes en un solo día. Los datos que aquí se presentan se refieren a toda la población investigada, salvo que se indique expresamente lo contrario.

Más información en: **www.ine.es**

DEPÓSITO LEGAL: M-13947-2001 ISSN: 1579-2277 NIPO: 605-07-006-1

El Instituto Nacional de Estadística (INE) presenta en esta publicación algunos de los principales resultados de la **Encuesta de Empleo del Tiempo**, **primera y única encuesta de ámbito nacional sobre la utilización del tiempo**. Se realizó en España entre los años 2002 y 2003 de manera armonizada con las de otros países europeos, siguiendo las recomendaciones de la Oficina Estadística de la Unión Europea (Eurostat). Entre los años 1998 y 2004 otros países de la Unión llevaron a cabo investigaciones similares.

La encuesta facilita información, entre otras cosas, del **porcentaje de personas que realizan una determinada actividad en el transcurso del día y la duración media diaria dedicada a esa actividad por dichas personas**. Esta información primaria nos permite analizar con rigor la dimensión del trabajo no remunerado realizado por los hogares, la distribución de las responsabilidades familiares en el hogar, la participación de la población en actividades culturales y de ocio, etc. Por otra parte, la información recogida también permite comparar **datos nacionales** de uso del tiempo **en relación con los demás países europeos** que han realizado la encuesta.

Como principales resultados, cabe destacar el dato de que **las tareas domésticas y el cuidado de niños y ancianos son tareas eminentemente femeninas, ya que el 93% de las mujeres las realizan, frente al 70% de los varones**. En el contexto europeo, es de señalar **la primera posición de España en tiempo dedicado a caminar y pasear; pero también el último lugar por lo que se refiere a tiempo dedicado a la lectura**.

El INE quiere aprovechar esta ocasión para expresar su **agradecimiento a los cerca de 24.000 hogares de la muestra**, y pone a su disposición los resultados obtenidos.

Fuentes estadísticas utilizadas:
Procedentes del INE: Encuesta de Empleo del Tiempo. La información internacional procede de **Eurostat**.

Para practicar

1. Clasifica as seguintes variables: a) Peso, b) densidade, c) nº de plantas dos edificios, d) Tipo de fachada dos edificios, e) nº de ventás, f)metros de fachada, g) nº de habitantes por edificio, h)tipo de porta principal.

2. Escribe tres variables cualitativas que teñan que ver con embarcacións.

3. Escribe tres variables cuantitativas discretas que teñan que ver con avións.

4. Escribe tres variables cuantitativas continuas que teñan que ver con trens.

5. Se as frecuencias para R, V, A e T son R→3, V→2, A→4 e T→1. Cantos graos lle corresponde a cada letra nun gráfico de sectores?

6. Fai unha táboa e un gráfico de sectores dos datos: R R A R A R V N V R N

7. Fai unha táboa e un gráfico de barras cos datos:
   3 3 4 5 4 5 3 2 1 2 3 4 5 4 5 4 3 3 4 4

8. Agrupa os datos seguintes en intervalos e fai un histograma.

195	194	194	182	168	179	191	154	177	189
184	187	155	167	177	187	161	171	190	162
190	152	166	180	156	186	184	167	184	162

9. Calcula a media en cada caso:
   a) 4, 6, 8
   b) 4, 6, 8, 6
   c) 100, 120, 180, 200

10. Calcula a media en cada caso:

a)

Marca	Fr
1	3
2	5
3	3
4	2

b)

Marca	Fr
1000	3
2000	5
3000	3
4000	2

11. Determina a moda e a mediana
   a) 50,60,60
   b) 12,12,22,32
   c) 10,20,30,40,20
   d) 35,25,35,25,25,25

12. Calcula a moda e a mediana en cada caso:

a)

Marca	Fr
100	5
200	4
300	6
400	3

b)

Marca	Fr
100	2
200	7
300	9
400	2

13. Cal ou cales dos datos seguintes se pode considerar unha observación atípica en cada unha das dúas series?
   a) 4 5 6 5 7 8 4 5 8 7 5 12 6 7 6 5 4
   b) 8 9 1 9 8 9 7 9 6 7 8

14. Calcula a mediana, o primeiro e o terceiro cuartil e o percentil 90 de:
   1 1 4 3 3 4 2 2 5 3 1 2 1 2 2 4 2 2 4 3 1

15. Calcula a mediana, primeiro e terceiro cuartil e o percentil 20 de
   3 1 1 1 4 1 5 3 1 3 3 4 5 5 4 4 2 1 4 4

16. Calcula a media e a desviación típica en cada un dos seguintes casos:
   a) 100 e 100
   b) 99 e 101
   c) 110 e 90
   d) 120 e 80

17. Completa a táboa cos datos:

190	151	193	187	158	175	165	158	184	172
197	161	157	157	183	180	150	161	182	169
162	177	160	155	188	157	189	167	186	157

Intervalo	Marca x_i	Frec. f_i	$f_i \cdot x_i$	$f_i(\bar{X}-x_i)^2$
[150,160)	155			
[160,170)	165			
[170,180)	175			
[180,190)	185			
[190,200)	195			

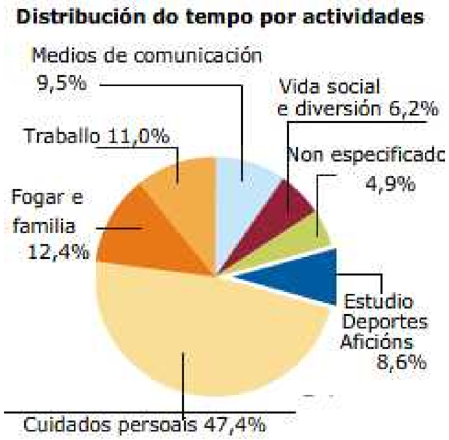
18. Determina a media e a desviación típica, dos datos da táboa anterior.

19. Determina os intervalos $\left(\bar{X}-\sigma, \bar{X}+\sigma\right)$ e $\left(\bar{X}-2\sigma, \bar{X}+2\sigma\right)$ e o número de elementos que hai en cada un.

Marca	Fr
0	5
1	4
2	7
3	3
4	2

20. Observa os seguintes gráficos e responde ás preguntas de cada un

a)

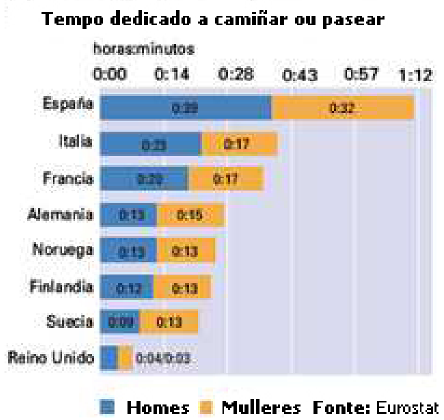


a1. Cal é a variable estudada? e a frecuencia?

a2. A que grupo de actividades dedicamos máis tempo os españois?

a3. Calcula canto tempo dedicamos ao fogar e a familia cantos graos ocupa este sector no diagrama?

b)



b1. En que países pasean máis as mulleres que os homes?

b2. Calcula o tempo medio que se dedica en cada país a pasear.

b3. Que país está no percentil 50?

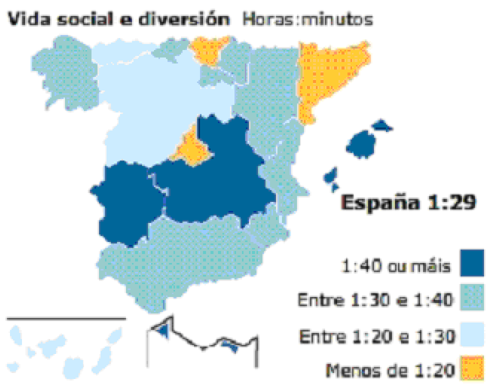
c)

c1. Cres que durmir se contou como actividade de coidado persoal?

c2. Ás 15:00 hai un máximo local na gráfica a que se debe?

c3. Á hora da comida o 38% das persoas dedícase ao coidado persoal. Significa isto que un 62% das persoas non come?

d)



d1. Cales son as comunidades nas que se dedica menos tempo á vida social e á diversión?

d2. Canto tempo dedican á diversión ou á vida social a maior parte das comunidades?

d3. Cal é o tempo medio que se dedica en España a esta actividade?

21. As cualificacións de 8 alumnos en Lingua e Inglés foron:

Lingua (X)	4	4	4	5	7	8	9	9
Inglés (Y)	3	5	6	5	8	9	10	9

Debuxa a nube de puntos e calcula o coeficiente de correlación lineal.

22. As horas semanais que adican 10 persoas a facer deporte e ver TV son:

Dep (X)	1	3	5	6	7	8	9	10	11	12
TV (Y)	14	14	13	10	8	9	4	8	5	5

Debuxa a nube de puntos e calcula o coeficiente de correlación lineal.

23. Dunha distribución bidimensional conocemos $\overline{x} = 8, \overline{y} = 7, \sigma_X = 1{,}5,\ \sigma_Y = 2{,}7$ e $\sigma_{XY} = 3{,}28$. Calcula o coeficiente de correlación lineal, a recta de regresión de Y sobre X e o valor estimado de y para x = 8.

24. Dunha distribución bidimensional conocemos $\overline{x} = 8, \overline{y} = 5, \sigma_X = 1{,}9,\ \sigma_Y = 2{,}5$ e $r = 0{,}83$. Calcula a recta de regresión de Y sobre X e o valor estimado de y para x = 10. És fiable esta predición?.

Para saber máis

A profesión de enfermaría.

Florence Nightingale (1820-1910), coñecida por ser a fundadora da profesión de enfermaría. Durante a guerra de Crimea decatouse de que a causa principal das mortes de feridos en combate era a falta de medidas sanitarias. Ao aplicalas, a taxa de mortalidade pasou dun 42,7% a un 2,2%. Grazas a un uso eficaz dos datos conseguiu modificar o sistema de atención sanitaria á súa volta a Gran Bretaña. Cambiou o sistema de rexistro de datos e foi unha das primeiras persoas en utilizar os gráficos estatísticos para representar os datos dunha forma sinxela de forma que ata os parlamentarios e xenerais puidesen entender.

Para Florence, os datos non eran algo abstracto, eran unha forma de poder salvar vidas humanas.

O pai da estatística.

Sir Ronald A. Fisher (1890-1962) está considerado o pai de estatística. Os escritos de Fisher axudaron a organizar a estatística como campo de estudo preciso os métodos do cal se aplican a problemas prácticos de moitas disciplinas. Como case todos os pioneiros na estatística, os seus traballos naceron da necesidade de resolver problemas prácticos.

Inferencia estatística

A estatística desenvolvida neste tema é o que se coñece como estatística descritiva, nela recóllese información e fanse cálculos que describen como están repartidos. Poñamos o caso que unha mostra elixida ao chou nos da unha media. A verdadeira media está próxima á da mostra? Se considero un intervalo arredor da media mostral, a verdadeira con qué probabilidade estará ou non nel? Destas preguntas e outras encárgase a inferencia estatística.

Principais campos de aplicación da estatística

A estatística aplícase en moitos campos como en **Industria e empresas**. Para o control de calidade na produción en cadea, para a análise de mercados, para o estudo de prezo de venda ao público dos artigos fabricados, en xestión financeira, ...

Na parte dereita cítanse algunhas outras das súas aplicacións.

Algúns campos de aplicación da estatística

Administración pública

A través das Delegacións territoriais e provinciais, recóllense datos para analizalos e sometelos a procesos estatísticos. Desta forma coñécense datos referidos a nacementos, defuncións, matrimonios, prezos, salarios, traballo, ensino, sanidade, ... Todos estes datos adóitanse publicar polo INE.

Economía.

Neste campo é imprescindible, sobre todo en macro-magnitudes.

Psicoloxía.

A maior parte dos traballos científicos en psicoloxía experimental teñen como principal ferramenta de traballo a estatística.

Medicina.

En calquera estudo experimental destas áreas Existe unha materia específica chamada Bioestatística para cubrir eses estudos experimentais. En Xenética e Antropometría atopamos dous dos campos de maior aplicación.

Lembra o máis importante

Poboación. Alumnos dun instituto ficticio.

Mostra. Alumnos enquisados

Variables estatísticas: Cualitativa, cor preferida; Cuantitativa discreta, nº de irmáns e cuantitativa continua, altura.

Consideremos as dúas mostras seguintes:

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

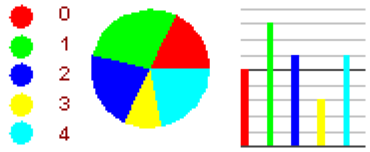
Altura: 182 172 157 194 150 166 163 196 167 199 172 185 172 168 173 160 162 173 161 192 156 164 173 180 193 172

Reconto de datos:

x_i	f
0	5
1	8
2	6
3	3
4	6
	28

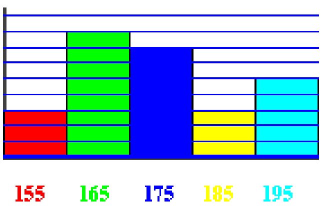
Intervalo	x_i	fi
[150,160)	155	3
[160,170)	165	8
[170,180)	175	7
[180,190)	185	3
[190,200)	195	5
	Total	26

Gráficos de **sectores** e **barras**
Nº de irmáns



Altura.

Histograma



Media e **moda** e **desviación típica**

	xi	fi	xi·fi	$fi\cdot(xi-\bar{x})^2$
	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

Media:

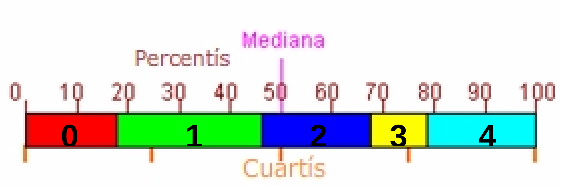
$$\bar{x} = \frac{53}{28} = 1{,}89$$

Moda = 1

Desviación típica:

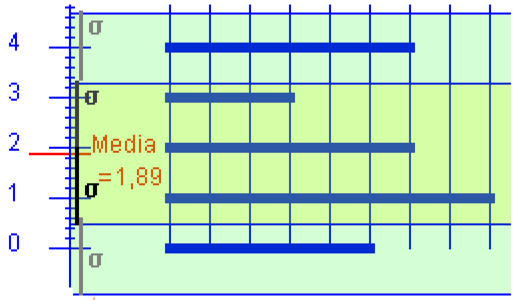
$$\sigma = \sqrt{\frac{54{,}67}{28}} = 1{,}39$$

Cuartil, mediana, percentil



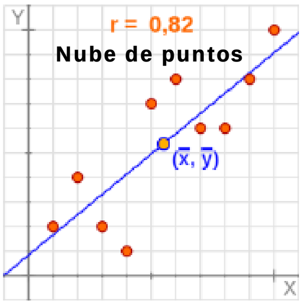
Me=2, Q1=1, Q3=3,
P20=1, P60=2, P90=4
Percorrido. De 0 a 4, de amplitude 4

Media e desviación No noso exemplo, 17 de 28 datos non se afastan da media máis da desviación típica, son o 60,7%, e o 100% non se afastan da media mais de dúas veces a desviación.



Representatividade
Unha mostra é representativa da poboación cando nela podemos atopar as mesmas proporcións das características de estudo que no conxunto da poboación.

Distribucións bidimensionais



x_i	y_i	x_i^2	y_i^2	$x_i\cdot y_i$
1	2	1	4	2
2	4	4	16	8
3	2	9	4	6
4	1	16	1	4
5	7	25	49	35
6	8	36	64	48
7	6	49	36	42
8	6	64	36	48
9	8	81	64	72
10	10	100	100	100
55	54	385	374	365

Covarianza: $\sigma_{XY} = \dfrac{\Sigma x_i\cdot y_i}{n} - \bar{x}\cdot\bar{y}$

Coeficiente de correlación lineal $r = \dfrac{\sigma_{XY}}{\sigma_X\cdot\sigma_Y}$

$x = 5{,}5$ $\sigma_X = 2{,}87$

$\sigma_{XY} = 6{,}8$

$y = 5{,}4$ $\sigma_Y = 2{,}87$

Recta de regresión
de Y sobre X

$$y = \bar{y} + \frac{\sigma_{XY}}{\sigma_X^2}(x - \bar{x})$$

$$r = \frac{6{,}8}{2{,}87\cdot 2{,}87} = 0{,}82$$

$$-1 \le r \le 1$$

$$y = 0{,}82x + 0{,}87$$

Autoavaliación

1. Cantos graos corresponden ao valor de frecuencia 3?

Xi	Fi
0	1
1	2
2	3
3	8
4	2

2. A mediana da distribución anterior é?

3. Cal é a moda?

Xi	Fi
3	1
4	3
5	4
6	2

4. Cal é a porcentaxe da mostra que corresponde ás dúas primeiras marcas ?

Xi	Fi
0	1
1	2
2	3
3	8
4	2

5. Cal é o percentil máis pequeno que deixa por debaixo os valores menores a 3?

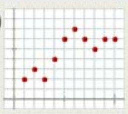
Xi	Fi
0	1
1	2
2	3
3	8
4	2

6. Cal é a media?

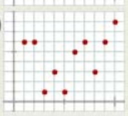
Xi	Fi
155	2
165	2
175	8
185	1
195	5

7. Cal é a desviación típica dos datos anteriores?

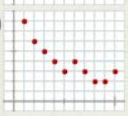
A)



B)



C)



8. Asocia cada nube de puntos (A, B, C) co seu coeficiente de correlación:

1) 0.32   2) 0.79   3) -0,88

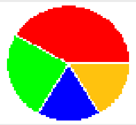
x_i	y_i	$x_i \cdot y_i$
1	1	1
2	2	4
3	3	9
4	5	20
5	4	20
6	5	30
7	5	35
8	5	40

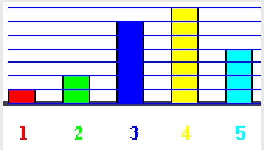
9. Na distribución da esquerda calcula a covarianza.

10. O centro de gravidade dunha distribución bidimensional é (4,5 , 3,75) e a pendiente da recta de regresión de Y sobre X é 0,57. Estima o valor de y para x = 7

Solucións dos exercicios para practicar

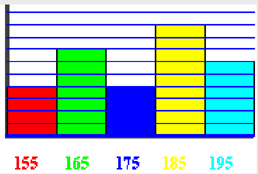
1. Cualitativas: d) h)
   Cuantitativas discretas c) e) g)
   C. continuas: a) b) f)
2. Propulsión, Carga, Tipo de travesía
3. Nº de pasaxeiros, nº rodas, nº ventás
4. Velocidade máxima, carga máxima, potencia.
5. R→108º, V→72º, A→144º, T→36º
6. R→5, A→3, V→2, N→2





7. 1→1, 2→2, 3→6, 4→ 7, 5→4
8.

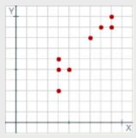
Intervalo	x_i	f_i
[150,160)	155	4
[160,170)	165	7
[170,180)	175	4
[180,190)	185	9
[190,200)	195	6



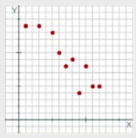
9. a) 6 b) 6 c) 150
10. a) 2.3 b) 2307
11. a)Mo=60, Me=60 b)Mo=12, Me=17
    c)Mo=20, Me=20
    d)Mo=25 Me=25
12. a) Mo=300, Me=250 b) Mo=300, Me=300
13. a) 12 b) 1
14. Me= 2, Q1=2, Q3=3, P90=4
15. Me=3, Q1=1, Q3=4 e P20=1
16. A media é 100 nos 4, e a desviación 0, 1, 10 e 20.
17.

Intervalo	Marca x_i	Frec. f_i	$f_i \cdot x_i$	$f_i(X-x_i)^2$
[150,160)	155	9	1395	2401
[160,170)	165	7	1155	280,77
[170,180)	175	3	525	40,33
[180,190)	185	8	1480	1494,22
[190,200)	195	3	585	1680.33
		30	5140	5896,66

18. $\bar{x} = 171,3$ $\sigma \approx 14,02$
19. En (0.42, 2.9) hai 11,
    e en (–0.88, 4.14) todos
20. a1)variable:actividades. Fr:porcentaxe de tempo diario que se dedica a cada actividade
    a2)coidados persoais
    a3) 2 h 58 m 34 s 44,64grados
    b1) Alemaña, Suecia e Finlandia
    b2) E35,5 I20, F18,5 A14 N13 F12,5 S11 R3,5 en minutos
    b3) Francia
    c1) Si. c2) Comida e Sesta
    c3) Non, o pico ocupa dúas horas e algúns comen en media hora
    d1) País Vasco, Cataluña e Madrid
    d2) entre 1:30 e 1:40 horas:minutos
    d3) 1:29
21. r = 0,93



22. r = -0,91



23. r = 0,81
    y = 1,46x - 1,66
    x = 8, y' = 7
24. y = 1,09x - 3,73
    x = 10, y' = 7,18 Bastante fiable

Solucións AUTOAVALIACIÓN

1. Sol 67,5º	6. Sol 177,78
2. Sol 3	7. Sol 12,83
3. Sol 5	8. 1-B 2-A 3-C
4. Sol 18,75%	9. Sol 3
5. Sol 37	10. Sol 5,18