# 13 Estatística

Neste torneo de xadrez participan 80 xogadores. A finalidade do torneo é clasificar o 10% dos participantes para a fase final. A táboa resume os resultados finais:

| PUNTUACIÓN | 6 | 5,5 | 5 | 4,5 | 4 | ... |
|---|---|---|---|---|---|---|
| N.° DE XOG. | 4 | 3 | 3 | 8 | 7 | ... |

**1** Tendo en conta a táboa da esquerda, cal é a menor puntuación coa que se clasificou un xogador?

**2** Dos 80 xogadores, hai 16 co título de Mestre Provincial (M.P.). As súas puntuacións no torneo danse na táboa seguinte:

| PUNTOS | 6 | 5,5 | 5 | 4,5 | 4 |
|---|---|---|---|---|---|
| N.° DE M.P. | 4 | 2 | 2 | 6 | 2 |

Cal foi a puntuación media destes xogadores (M.P.)?

1. Solucións a estes problemas.

# 1 Conceptos básicos

■ **POBOACIÓN**

É o conxunto de todos os elementos cuxo coñecemento nos interesa e que serán obxecto do noso estudo.

■ **MOSTRA**

É un subconxunto extraído da poboación, cuxo estudo serve para inferir características de toda a poboación.

■ **INDIVIDUO**

É cada un dos elementos que forman a poboación ou a mostra.

■ **CARACTERES E VARIABLES**

**Caracteres** son os aspectos que desexamos estudar nos individuos dunha poboación. Cada carácter pode tomar distintos valores ou modalidades.

Unha **variable estatística** percorre todos os valores dun certo carácter.

As variables estatísticas poden ser:

- **Cuantitativas** se toman valores numéricos.
  - — **Discretas:** só toman valores illados.
  - — **Continuas:** poden tomar calquera valor dun intervalo.
- **Cualitativas** tómanse valores non numéricos.

Por exemplo, os 17 942 estudantes dunha universidade forman unha **poboación.**

Cada un deles é un **individuo.**

Pódense analizar múltiples **caracteres:** *facultades* á que pertence, *sexo*, *número de anos* que estivo matriculado, *idade*, *estatura*, *tipo de lectura preferida*, …

As **variables** correspondentes son dos seguintes tipos:

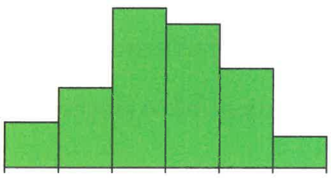
*facultade*, *sexo* e *tipo de lectura* → cualitativas

*número de anos matriculado* → cuantitativa discreta

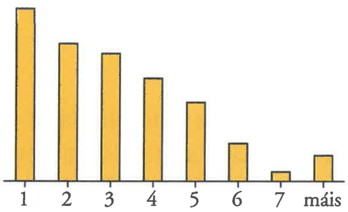
*estatura*, *idade* → cuantitativas continuas

Se se lles fai unha enquisa a 387 alumnos desta universidade, o colectivo formado por eses alumnos é unha **mostra.**

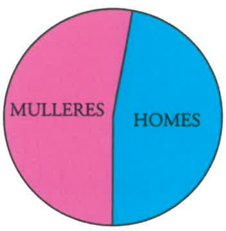

FACULTADES



ESTATURAS




NÚMERO DE ANOS MATRICULADO




SEXO

## Actividades

2. Actividades para **repasar** os conceptos básicos de estatística.

**1** Desexamos facer un estudo comparativo dalgúns aspectos dos distintos países do mundo (número de habitantes, renda per cápita, relixión predominante e número de cidades con máis de 500 000 habitantes). Neste estudo estatístico, cal é a **poboación**? Cales son os **individuos**? Di cales son as **variables** e de que tipo son.

# 2 Dúas ramas da estatística

A estatística ten por obxecto o desenvolvemento de técnicas para o coñecemento numérico dun conxunto de datos empíricos (recollidos mediante experimentos ou enquisas). Segundo o colectivo a partir do cal se obteña a información e o obxectivo que se persiga á hora de analizar eses datos, a estatística chámase descritiva ou inferencial.

## Estatística descritiva

*Primeiro recóllense os datos.*

*Despois clasifícanse e organízanse os datos.*

A **estatística descritiva** trata de describir e analizar algúns caracteres dos individuos dun grupo dado (poboación) sen extraer conclusións para un grupo maior.

Para este estudo danse os seguintes pasos:

- Selección dos caracteres que interesa estudar.
- Análise de cada carácter: deseño da enquisa ou do experimento e recollida de datos.
- Clasificación e organización dos resultados en táboas de frecuencias.
- Elaboración de gráficos, se convén, para divulgalos a un público amplo (non expertos).
- Obtención de **parámetros:** valores numéricos que resumen a información obtida.

**▶ Exemplo**

Supoñamos que por orde do reitor, un funcionario dunha universidade organiza, tabula, representa graficamente e obtén parámetros dalgúns caracteres de todos os alumnos (entre outros, idades e resultados académicos) para comparalos con estudos similares feitos en anos anteriores. Este estudo é **estatística descritiva,** pois realízase sobre a totalidade da poboación.

## Estatística inferencial

A **estatística inferencial** traballa con mostras e pretende, a partir delas, "inferir" características de toda a poboación. É dicir, preténdense tomar como xerais propiedades que só se verificaron para casos particulares. Nese proceso hai que operar con moita cautela: Como se elixe a mostra? Que grao de confianza se pode ter no resultado obtido?

**▶ Exemplo**

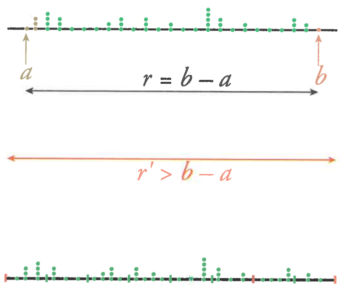
Unha editorial realiza unha enquisa a 387 universitarios sobre as súas preferencias na lectura, co fin de extraer consecuencias válidas para todos os universitarios. Isto é **estatística inferencial,** pois, a partir dunha mostra, deséxase obter información sobre algún aspecto da poboación.

# 3 Táboas de frecuencias

3. Podes **reforzar,** con algunhas actividades, a elaboración de táboas de frecuencias.

Tras a recollida de datos, a elaboración dunha táboa de frecuencias é o seguinte paso. Cando a variable toma poucos valores, a elaboración da táboa é sumamente sinxela. Non hai máis que facer o reconto dos resultados.

## Táboa cos datos agrupados




Cando nunha distribución estatística o número de valores que toma a variable é moi grande, convén elaborar unha táboa de frecuencias agrupándoos en intervalos. Para iso:

- Localízanse os valores extremos, $a$ e $b$, e áchase a súa diferenza, $r = b - a$.
- Decídese o número de intervalos que se quere formar, tendo en conta a cantidade de datos que se posúen. O número de intervalos non debe ser inferior a 6 nin superior a 15.
- Tómase un intervalo, $r'$, de lonxitude algo maior que o percorrido $r$ e que sexa múltiplo do número de intervalos, co obxecto de que estes teñan unha lonxitude enteira.
- Fórmanse os intervalos, de modo que o extremo inferior do primeiro sexa algo menor que $a$ e o extremo superior do último sexa algo superior a $b$. É desexable que os extremos dos intervalos non coincidan con ningún dos datos. Para iso, convén que os extremos dos intervalos teñan unha cifra decimal máis que os datos.

O punto medio de cada intervalo chámase **marca de clase.** É o valor que representa a todo o intervalo para o cálculo dalgúns parámetros.

### Non o esquezas

Cando se elabora unha táboa con datos agrupados, pérdese algo de información (pois nela ignórase cada valor concreto, que se esvaece dentro dun intervalo). A cambio, gáñase en claridade e eficacia.

### Exercicio resolto

**Elaborar unha táboa de frecuencias coas estaturas de 40 adolescentes:**

| | | | | |
|---|---|---|---|---|
| 168 | 160 | 167 | 175 | 175 |
| 167 | 168 | 158 | 149 | 160 |
| 178 | 166 | 158 | 163 | 171 |
| 162 | 165 | 163 | 156 | 174 |
| 160 | 165 | 154 | 163 | 165 |
| 161 | 162 | 166 | 163 | 159 |
| 170 | 165 | 150 | 167 | 164 |
| 165 | 173 | 164 | 169 | 170 |

- Valores extremos: $a = 149$, $b = 178$. Percorrido: $r = 178 - 149 = 29$
- Tomaremos só 6 intervalos. Un múltiplo de 6 maior que 29 e próximo a el é 30. Lonxitude de cada intervalo: 5
- Formamos os intervalos comezando por un número algo menor que $a = 149$ e terminando nun número algo maior que $b = 178$.
- Repartimos os datos nos intervalos:

| INTERVALOS | 148,5-153,5 | 153,5-158,5 | 158,5-163,5 | 163,5-168,5 | 168,5-173,5 | 173,5-178,5 |
|---|---|---|---|---|---|---|
| M. DE CLASE | 151 | 156 | 161 | 166 | 171 | 176 |
| FRECUENCIAS | 2 | 4 | 11 | 14 | 5 | 4 |

### Actividades

**1** Reparte os corenta datos do exercicio resolto anterior en 10 intervalos co mesmo percorrido total.

**2** Reparte os corenta datos do exercicio resolto anterior en 8 intervalos. Para iso, toma $r' = 32$.

# 4 Parámetros estatísticos: $\bar{x}$ e $\sigma$

| $x_i$ | $f_i$ |
|---|---|
| $x_1$ | $f_1$ |
| $x_2$ | $f_2$ |
| . | . |
| . | . |
| . | . |
| $x_n$ | $f_n$ |

A táboa de frecuencias da esquerda **pode corresponder** a:

- Unha distribución de datos illados que toma os valores $x_1, x_2, \dots x_n$.
- Unha distribución de datos agrupados en intervalos, dos cales $x_1, x_2, \dots x_n$ son as marcas de clase.

No primeiro caso, a táboa reflicte exactamente a distribución real. No segundo a táboa é unha boa aproximación á realidade.

Lembremos como se obteñen os **parámetros** a partir dunha táboa:

PUNTUACIÓNS NUN TEST

| $x_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|
| 0 | 12 | 0 |
| 1 | 31 | 31 |
| 2 | 86 | 172 |
| 3 | 92 | 276 |
| 4 | 48 | 192 |
| 5 | 19 | 95 |
| | 288 | 766 |

■ **MEDIA:** $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$  $\Sigma f_i = N \rightarrow$ n.º total de individuos; $\Sigma f_i x_i \rightarrow$ suma de todos os datos

Por exemplo, na distribución que temos na marxe:

$\Sigma f_i = 288$. Hai 288 individuos (que realizaron o test).

$\Sigma f_i x_i = 766$. É a suma das puntuacións de todos os individuos.

A media é $\bar{x} = 766/288 = 2,66$.

■ **VARIANZA:** $Var = \dfrac{\Sigma f_i (x_i - \bar{x})^2}{N}$, ou ben $Var = \dfrac{\Sigma f_i x_i^2}{N} - \bar{x}^2$

As dúas expresións coinciden.

— Na primeira delas, vese claro o significado da varianza: suma dos cadrados das desviacións á media.

— A segunda é máis cómoda para os cálculos, como se pode apreciar no exemplo (táboa do marxe):

$$Var = \frac{2\,446}{288} - 2,66^2 = 1,42$$

| $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|
| 0 | 12 | 0 | 0 |
| 1 | 31 | 31 | 31 |
| 2 | 86 | 172 | 344 |
| 3 | 92 | 276 | 828 |
| 4 | 48 | 192 | 768 |
| 5 | 19 | 95 | 475 |
| | 288 | 766 | 2 446 |

■ **DESVIACIÓN TÍPICA:** $\sigma = \sqrt{varianza}$

A desviación típica é un parámetro máis razoable que a varianza, pois exprésase na mesma magnitude que os datos e que a media (por exemplo, se os datos veñen en centímetros, a desviación típica vén en centímetros; porén, a varianza daríase en centímetros cadrados).

No exemplo: $\sigma = \sqrt{1,42} = 1,19$

■ **COEFICIENTE DE VARIACIÓN:** $C.V. = \dfrac{\sigma}{\bar{x}}$

O coeficiente de variación serve para comparar as dispersións de poboación heteroxéneas, pois indica a *variación relativa*.

No exemplo: $C.V. = \dfrac{1,19}{2,66} = 0,447$. Ou ben 44,7%.

**FOLLA DE CÁLCULO.** Aplicación para confeccionar táboas de frecuencias, representar o gráfico correspondente e calcular $\bar{x}$, $\sigma$ e C.V.

## Exercicio resolto

**Calcular $\bar{x}$, $\sigma$ e C.V. na seguinte distribución:**

| DISTRIBUCIÓN DE PESOS (EN kg) | |
|---|---|
| INTERVALOS | FRECUENCIAS |
| 42,5-53,5 | 4 |
| 53,5-64,5 | 19 |
| 64,5-75,5 | 86 |
| 75,5-86,5 | 72 |
| 86,5-97,5 | 41 |
| 97,5-108,5 | 7 |

Empezamos substituíndo os intervalos polas súas marcas de clase:

| $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|
| 48 | 4 | 192 | 9 216 |
| 59 | 19 | 1 121 | 66 139 |
| 70 | 86 | 6 020 | 421 400 |
| 81 | 72 | 5 832 | 472 392 |
| 92 | 41 | 3 772 | 347 024 |
| 103 | 7 | 721 | 74 263 |
| | 229 | 17 658 | 1 390 434 |

$N = \Sigma f_i = 229$

$\Sigma f_i x_i = 17\,658$

$\Sigma f_i x_i^2 = 1\,390\,434$

Os números da 3.ª columna, $f_i x_i$, obtéñense multiplicando os números das columnas anteriores $(x_i \cdot f_i = f_i x_i)$. Por exemplo, $59 \cdot 19 = 1\,121$.

Analogamente, os da 4.ª columna obtéñense multiplicando os da 1.ª polos da 3.ª $(x_i \cdot f_i x_i = f_i x_i^2)$. Por exemplo, $59 \cdot 1\,121 = 66\,139$.

Coas sumas das columnas da táboa, obtemos os parámetros:

MEDIA: $$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{17\,658}{229} = 77{,}1 \text{ kg}$$

DESVIACIÓN TÍPICA: $$\sigma = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \bar{x}^2} = \sqrt{\frac{1\,390\,434}{229} - 77{,}1^2} = 11{,}2 \text{ kg}$$

COEF. DE VARIACIÓN: $$\text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{11{,}2}{77{,}1} = 0{,}145 = 14{,}5\%$$

**CON CALCULADORA**

1. Preparamos a calculadora para que traballe no **MODO SD.**
2. Borramos os datos que puidera haber acumulados doutras ocasións: INV AC
3. Introducimos os datos: 48 ⊠ 4 DATA

   59 ⊠ 19 DATA

   ...

   103 ⊠ 7 DATA
4. Resultados obtidos:

   N.º DE INDIVIDUOS $\Sigma f_i$   $n$ → 229

   SUMA DE VALORES $\Sigma f_i x_i$   $\Sigma x$ → 17 658

   SUMA DE CADRADOS $\Sigma f_i x_i^2$   $\Sigma x^2$ → 1 390 434

   MEDIA $\bar{x}$   $\bar{x}$ → 77.10917031

   DESV. TÍPICA $\sigma$   $\sigma_n$ → 11.22230132

CALCULADORA. No teu CD encontrarás explicacións máis minuciosas sobre o uso da calculadora en estatística.

4. No teu CD hai máis actividades para **reforzar** o cálculo de $\bar{x}$, $\sigma$ e C.V.

## Actividades

**1** Acha, manualmente e con calculadora, $\bar{x}$, $\sigma$ e C.V. no exercicio resolto da páxina 191:

| $x_i$ | 151 | 156 | 161 | 166 | 171 | 176 |
|---|---|---|---|---|---|---|
| $f_i$ | 2 | 4 | 11 | 14 | 5 | 4 |

**2** Acha, manualmente e con calculadora, $\bar{x}$, $\sigma$ e C.V. na distribución dos exercicios 1 e 2 da páxina 191.

Compara os resultados entre si e cos do exercicio 1 desta páxina.

# 5 Medidas de posición

## ■ MEDIANA

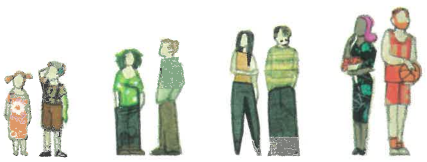
↑
*Me*

Se os individuos dunha poboación están colocados en orde crecente segundo a variable que estudamos, o que ocupa o valor central chámase individuo mediano, e o seu valor, a **mediana.** A mediana, $Me$, está situada de modo que antes dela está o 50% da poboación, e detrás, o outro 50%.

Por exemplo: 6, 7, 7, 7, **8**, 9, 10, 12, 15 → mediana: $Me = 8$

Se o número de individuos é par, a mediana é o valor medio dos dous centrais. Por exemplo: 6, 7, 7, 7, **8**, **9**, 10, 12, 15, 16 → $Me = 8{,}5$

## ■ CUARTÍS



↑ ↑ ↑
$Q_1$ $Me$ $Q_3$

Se, en lugar de partir a totalidade dos individuos en dúas metades, o facemos en catro partes iguais (todas elas co mesmo número de individuos), os dous novos puntos de separación chámanse **cuartís.**

**Cuartil inferior,** $Q_1$, é un valor da variable que deixa por debaixo del o 25% da poboación, e por riba, o 75%.

O **cuartil superior,** $Q_3$, deixa debaixo o 75% e encima o 25%.

Desígnanse por $Q_1$ e $Q_3$, porque a mediana sería o $Q_2$.

Por exemplo, na distribución

$$\underbrace{1, 2, 2}_{25\%} \;,\; \underbrace{3, 4, 5}_{25\%} \;,\; \underbrace{5, 5, 6}_{25\%} \;,\; \underbrace{8, 9, 10}_{25\%}$$

↑ $Q_1$ ↑ $Me$ ↑ $Q_3$

estes parámetros toman os valores seguintes: $Q_1 = 2{,}5$; $Me = 5$; $Q_3 = 7$

### Ten en conta

En xeral, as cousas non son tan fáciles como neste exemplo. Obsérvao no exercicio resolto.

## ■ CENTÍS OU PERCENTÍS

Se partimos a poboación en 100 partes e sinalamos o lugar que deixa debaixo $k$ delas, o valor da variable correspondente a ese lugar desígnase por $p_k$ e denomínase **centil** $k$ ou **percentil** $k$.

A mediana é $Me = p_{50}$, e os cuartís, $Q_1 = p_{25}$, $Q_3 = p_{75}$.

Mediana, cuartís e centís chámanse **medidas de posición.**

5. No teu CD hai máis actividades para **reforzar** o cálculo das medidas de posición.

### Exercicio resolto

**Calcular $Me$, $Q_1$, $Q_3$, $p_{10}$ e $p_{80}$ na distribución:**

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

Hai 17 individuos. 17/2 = 8,5 → A $Me$ é o valor do individuo 9.°, $Me = 5$

17/4 = 4,25 → (5.° lugar) $Q_1 = 4$

17 · (3/4) = 12,75 → (13.° lugar) $Q_3 = 7$

17 · (10/100) = 1,7 → (2.° lugar) $p_{10} = 1$

17 · (80/100) = 13,6 → (14.° lugar) $p_{80} = 7$

# Frecuencias acumuladas

**FOLLA DE CÁLCULO** que che permite obter e visualizar medidas de posición da distribución que desexes.

Para calcular a mediana, os cuartís e os demais percentís en distribucións dadas por táboas de frecuencias, necesítase o concepto de **frecuencia acumulada.**

Nunha distribución de frecuencias, chámaselle **frecuencia acumulada,** $F_i$, correspondente ao valor *i*-ésimo, $x_i$, á suma da frecuencia dese valor con todas as anteriores:

$$F_i = f_1 + f_2 + \dots + f_i$$

NÚMERO DE FILLOS DE 110 PARELLAS

| $x_i$ | $f_i$ |
|---|---|
| 0 | 10 |
| 1 | 20 |
| 2 | 41 |
| 3 | 29 |
| 4 | 14 |
| 5 | 5 |
| 6 | 1 |

Por exemplo, na distribución de frecuencias dada na marxe, obtemos:

| $x_i$ | $f_i$ | $F_i$ | | EN % |
|---|---|---|---|---|
| 0 | 10 | 10 | = 10 | 8,3 |
| 1 | 20 | 10 + 20 | = 30 | 25 |
| 2 | 41 | 10 + 20 + 41 | = 30 + 41 = 71 | 59,2 |
| 3 | 29 | 10 + 20 + 41 + 29 | = 71 + 29 = 100 | 83,3 |
| 4 | 14 | 10 + 20 + 41 + 29 + 14 | = 100 + 14 = 114 | 95 |
| 5 | 5 | 10 + 20 + 41 + 29 + 14 + 5 | = 114 + 5 = 119 | 99,2 |
| 6 | 1 | 10 + 20 + 41 + 29 + 14 + 5 + 1 | = 119 + 1 = 120 | 100 |

A expresión en % das frecuencias acumuladas permítenos obter facilmente os percentís.

## Lembra

Para achar que porcentaxe (%) son respecto a 120 diversas cantidades, procédese así:

100 ÷ 120 = × ×

# Obtención de percentís en táboas de frecuencias

Para achar o percentil $p_k$ nunha táboa de frecuencias, obtéñense as frecuencias acumuladas e exprésanse en %. O percentil $p_k$ é o valor para o cal a frecuencia acumulada correspondente supera o $k$%.

No caso de que unha delas coincida con $k$%, tómase como $p_k$ o valor intermedio entre ese valor de $x$ e o seguinte.

| $x_i$ | $f_i$ | $F_i$ | EN % |
|---|---|---|---|
| 0 | 10 | 10 | 8,3 |
| 1 | 20 | 30 | 25 |
| 2 | 41 | 71 | 59,2 |
| 3 | 29 | 100 | 83,3 |
| 4 | 14 | 114 | 95 |
| 5 | 5 | 119 | 99,2 |
| 6 | 1 | 120 | 100 |

Por exemplo, na táboa da marxe, obtemos $Me$, $Q_1$, $Q_3$, $p_{99}$ e $p_{95}$:

$Me = p_{50} = 2$ porque para $x_i = 2$ a $F_i$ supera o 50%

$Q_1 = p_{25} = 1{,}5$ porque para $x_i = 1$ a $F_i$ é exactamente o 25%

$Q_3 = p_{75} = 3$ porque para $x_i = 3$ a $F_i$ supera o 75%

$p_{99} = 5$ porque para $x_i = 5$ a $F_i$ supera o 99%

$p_{95} = 4{,}5$ porque para $x_i = 4$ a $F_i$ é exactamente o 25%

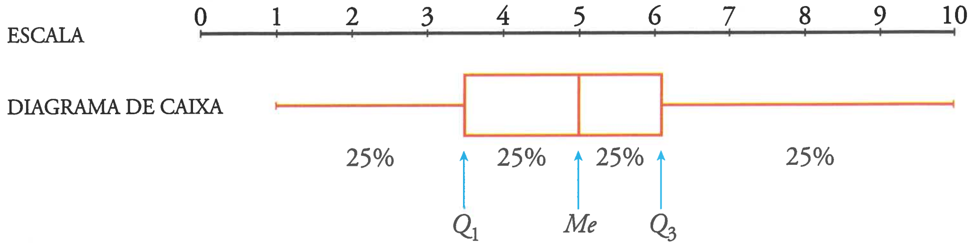
## Actividades

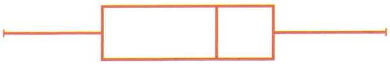
**1** Na seguinte distribución de notas, acha $Me$, $Q_1$, $Q_3$, $p_{80}$, $p_{90}$ e $p_{99}$.

| NOTAS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| N.° DE ALUMNOS | 7 | 15 | 41 | 52 | 104 | 69 | 26 | 13 | 19 | 14 |

# 6 Diagramas de caixa

Observa a seguinte forma de representar distribucións estatísticas.



Este diagrama chámase tamén de **caixa e bigotes.**

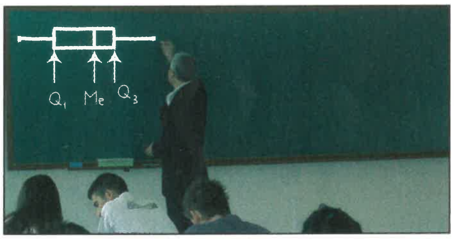
A gráfica corresponde á distribución de notas nun certo exame. Na parte alta púxose a escala sobre a que se move a variable. Debaixo ponse o diagrama propiamente dito, que consiste no seguinte:

— A poboación total pártese en catro anacos, cada un deles co 25% dos individuos, previamente ordenados de menor a maior.

— O 50% dos valores centrais destácanse mediante un rectángulo (**caixa**).

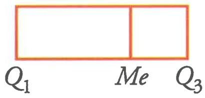
— Os valores extremos (o 25% dos menores e o 25% dos maiores) represéntanse mediante co cadanseu segmento (**bigotes**).

Os puntos que separan os catro anacos son, obviamente, os cuartís e a mediana ($Q_1$, $Me$, $Q_3$).

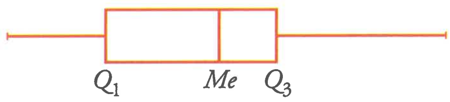



Os **diagramas de caixa** (ou caixa e bigote) constrúense do seguinte modo:

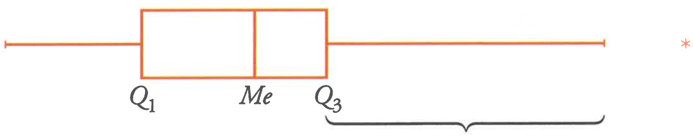
- A **caixa** abarca o intervalo $Q_1$, $Q_3$ (chamado percorrido intercuartílico) e nela sinálase expresamente o valor da mediana, $Me$.




- Os bigotes trázanse ata abarcar a totalidade dos individuos, coa condición de que cada lado non se alongue máis dunha vez e media a lonxitude da caixa.




- Se un (ou máis) dos individuos quedara por debaixo ou por arriba desa lonxitude, o correspondente lado do bigote debuxaría con esa limitación e engadiríase, mediante asterisco, o individuo no lugar que lle corresponda. Por exemplo:




A lonxitude deste lado do bigote é 1,5 veces a da caixa. Neste lado non está incluído o individuo extremo que se representa mediante un asterisco.

## Problemas resoltos

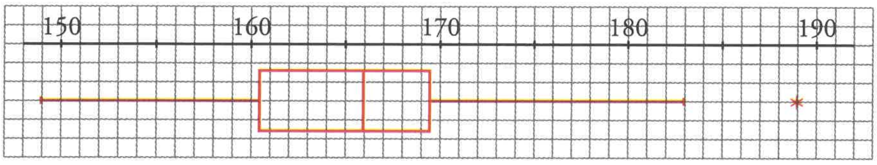
**1. As estaturas dos 40 alumnos e alumnas dunha clase son, dadas ordenadamente:**

| | | | | |
|---|---|---|---|---|
| 149 | 150 | 154 | 156 | 157 |
| 158 | 159 | 160 | 160 | 160 |
| 161 | 162 | 162 | 163 | 163 |
| 163 | 163 | 164 | 165 | 166 |
| 166 | 166 | 167 | 167 | 167 |
| 168 | 168 | 168 | 169 | 169 |
| 170 | 170 | 170 | 171 | 172 |
| 173 | 174 | 175 | 175 | 189 |

**Representar a distribución mediante un diagrama de caixa.**

1. Posto que o número de individuos é múltiplo de catro, $Q_1$, $Me$ e $Q_3$ serán os valores que hai entre os individuos 10.° e 11.°, entre 20.° e 21.°, entre 30.° e 31.°, respectivamente. É dicir,

$Q_1 = 160{,}5$ $\qquad$ $Me = 166$ $\qquad$ $Q_3 = 169{,}5$




A lonxitude da caixa é 169,5 – 160,5 = 9.

Unha vez e media esta lonxitude é 1,5 · 9 = 13,5.

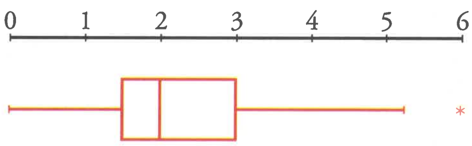
O altísimo estudante que mide 189 cm sepáranse do extremo superior da caixa 189–160,5 = 18,5. Esa distancia é maior que unha vez e media a lonxitude da caixa. Por iso, puxemos á dereita un bigote de lonxitude 13,5 e engadimos un asterisco que sinala a situación do individuo excepcional.

**2. Representar, mediante un diagrama de caixa, a seguinte distribución.**

| $x_i$ | $f_i$ |
|---|---|
| 0 | 10 |
| 1 | 20 |
| 2 | 41 |
| 3 | 29 |
| 4 | 14 |
| 5 | 5 |
| 6 | 1 |

2. Na páxina 207 calculamos algunhas medidas de posición correspondentes a esta distribución. En concreto:

$Q_1 = 1{,}5$ $\qquad$ $Me = 2$ $\qquad$ $Q_3 = 3$




A **caixa** abarca o intervalo [1,5; 3]. A súa lonxitude é 3 – 1,5 = 1,5.

Os segmentos do bigote deben ter, como moito, 1,5 · 1,5 = 2,25.

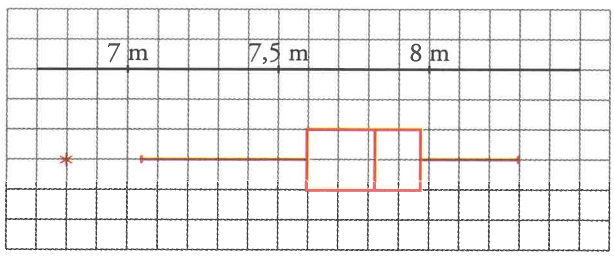
A rama esquerda mide menos de 2,25. A da dereita, de 2,25, non abarca o elemento maior (unha familia con 6 fillos), que se representa mediante un asterisco.

## Actividades

**1** Fai o diagrama de caixa correspondente a esta distribución de notas.

| $x_i$ | $f_i$ |
|---|---|
| 1 | 6 |
| 2 | 15 |
| 3 | 22 |
| 4 | 24 |
| 5 | 33 |
| 6 | 53 |
| 7 | 22 |
| 8 | 16 |
| 9 | 8 |
| 10 | 1 |

**2** Interpreta o seguinte diagrama de caixa relativo a marcas de saltadores de lonxitude.



# 7 Estatística inferencial

## Por que se recorre ás mostras

I. Se desexamos coñecer algúns datos anatómicos (estatura, peso, perímetro torácico...) dos 843 estudantes dun centro docente, pódese conseguir con facilidade medíndoos. Con todo, se quixésemos as mesmas medidas de todos os mozos europeos de idades comprendidas entre 18 e 30 anos, a tarefa sería desmesurada. Teriamos que recorrer a unha mostra.

II. Para estudar a duración dunha lámpada hai que deixala acesa e medir o tempo transcorrido ata que se funda. Como é natural, non se pode facer iso coa totalidade das lámpadas dunha produción. Debe recorrerse a unha mostra.

III. Desexamos coñecer a opinión que teñen sobre as rebaixas as persoas que, nun certo momento, se encontran nuns grandes almacéns. É imposible preguntarlles a todas elas, hai que recorrer a unha mostra, pois non é posible controlar, nin aproximadamente, cales son os individuos da poboación, nin a cantos deles non se enquisou.

Na práctica, é moi frecuente ter que recorrer a unha mostra para inferir datos da poboación por algún ou varios dos seguintes motivos:

- A poboación é excesivamente numerosa (caso I).
- A poboación é moi difícil, ou imposible, de controlar (caso III).
- O proceso de medición é destrutivo (caso II) ou demasiado caro.
- Deséxase coñecer rapidamente certos datos da poboación e tardaríase demasiado en consultar a todos (por exemplo, as sondaxes electorais).

## Actividades

**1** Un fabricante de parafusos desexa facer un control de calidade. Recolle un de cada 100 parafusos fabricados e analízao.

O conxunto de parafusos analizados é poboación ou mostra? Por que?

**2** Un fabricante de vasos de vidro quere estudar a resistencia que presentan á rotura. O procedemento consiste en sometelos a presións paulatinamente crecentes ata que parten.

Pode facer o estudo sobre a poboación ou debe recorrer a unha mostra? Por que?

**3** Un campesiño posúe 127 galiñas. Para probar a eficacia dun novo tipo de alimentación, pésaas todas antes e despois dos vinte días que dura o tratamento.

O conxunto desas 127 galiñas é poboación ou mostra? Por que?

## Tamaño da mostra

Respecto do tamaño, está claro que, se a mostra é demasiado pequena, non poderemos extraer dela ningunha conclusión que pague a pena. Con todo, con mostras aparentemente moi pequenas conséguense estimacións sorprendentemente boas na realidade.

Máis adiante analizaremos a relación que hai entre o tamaño da mostra e o tipo de conclusións que obteñamos dela.

## A mostra debe elixirse ao chou

Ao substituír o estudo da poboación polo da mostra, cométense erros. Pero con eles contamos de antemán e poden controlarse.

Non obstante, se a mostra está mal elixida (**non é representativa**), prodúcense erros adicionais imprevistos e incontrolados (**nesgos**).

O proceso mediante o cal se confecciona a mostra chámase **mostraxe.** Como debe ser a mostraxe para que nos proporcione unha mostra representativa, non nesgada? Talvez che resulte chocante, pero é imprescindible que a mostra se elixa ao chou. É dicir, a *mostraxe debe ser aleatoria.*

**▶ Exemplo**

Nun centro escolar deséxase coñecer o número de horas que estudan, por termo medio, os 1 000 alumnos e alumnas. Pero en vez de indagalo enquisando a todos eles, vaise extraer unha mostra de 100. Reflexionemos sobre a validez de cada un dos seguintes métodos de mostraxe:

a) O director elixe a mostra procurando que haxa alumnos de todo tipo.

b) Elíxense os 100 primeiros que cheguen ao centro un certo día.

c) Numéranse do 1 ao 1 000 e elíxense, ao chou, 100 deles.

O procedemento a) ten o grande inconveniente de que depende da subxectividade do director. Nunca se debe elixir unha mostra con criterios subxectivos, pois as ideas de quen a elixa influirán enormemente nas conclusións que se extraian.

O método b) tampouco é bo. É posible que os primeiros alumnos que chegan ao centro sexan os máis responsables e isto inflúa sobre os seus métodos de estudo.

Destes tres procedementos, o único método válido é o c).

Dise que unha **mostraxe é aleatoria** cando os individuos da mostra se elixen ao chou, de modo que todos os individuos da poboación teñen a mesma probabilidade de ser elixidos.

A mostraxe aleatoria é o único que garante a fiabilidade das conclusións que se obteñan.

## Conclusións que se obteñen dunha mostra

Imos centrarnos nun exemplo concreto para estudar as conclusións que podemos sa
car dunha mostra e o grao de fiabilidade que teñen as devanditas conclusións:

Para estimar a estatura de todos os mozos homes de 21 anos, extraemos unha mos
tra aleatoria de 200 deles. A media das estaturas dos individuos da mostra
$\bar{x}$ = 173 cm. A partir destes datos realízase a seguinte estimación:

"A estatura media dos varóns de 21 anos é, aproximadamente, 173 cm".

Coa expresión *aproximadamente* queremos indicar "algo máis ou algo menos"
Por exemplo, medio centímetro máis ou menos. Se así fose, diriamos:

"A estatura media dos homes de 21 anos encóntrase no intervalo (172,5; 173,5)"

Seguro? Por suposto que non. É, só, *probable*. A afirmación podería ser así:

"A estatura media dos varóns de 21 anos encóntrase no intervalo(172,5; 173,5)
E esta afirmación facémola cun nivel de confianza do 90%". (Isto significa que
probabilidade de que sexa certo é de 0,9).

Observa que **canto maior sexa o intervalo, maior será o nivel de confianza qu teremos** (canto maior sexa o tamaño da diana, máis probable é que atinemos). E
lugar da anterior, a afirmación podería ser estoutra:

"A estatura media dos homes de 21 anos encóntrase no intervalo (172, 174).
isto dicímolo cun nivel de confianza do 99,5%".

Canto máis precisos queiramos ser na estimación, menor será o nivel de confian
za que teñamos nela.

### Nivel de confianza

Unha confianza do 90% significa que, por termo medio, de cada 100 veces que fixeramos unha estimación coas precaucións tomadas nesta, en 90 delas acertariamos e en 10 equivocariámonos.

As valoracións numéricas danse mediante intervalos, acompañados dunha probabilidade (nivel de confianza).

Canto máis amplo sexa o intervalo, maior é o nivel de confianza que teremos. E, ao contrario, se se quere afinar moito nas previsións reducindo o intervalo, perderemos confianza no que dicimos.

O tamaño da mostra tamén inflúe. Aumentándoo poderemos:

- Mellorar o nivel de confianza mantendo a amplitude do intervalo.
- Reducir a amplitude do intervalo mantendo o nivel de confianza.

### Non o esquezas

As conclusións que se extraen dunha mostra para a poboación son só aproximadas.

6. No teu CD hai algunha actividade máis para **relacionar** "tamaño da mostra-nivel de confianza-amplitude do intervalo de confianza".

### Actividades

**4** Pasóuselles un test aos 64 individuos dunha mostra seleccionada aleatoriamente. Cos resultados obtidos chegouse á seguinte conclusión:

"A puntuación media que alcanzarían os individuos de toda a poboación se se lles pasara este test estaría entre 42,7 puntos e 44,1 puntos. E isto podémolo afirmar cun nivel de confianza do 95%".

a) Se o intervalo que se dera fose 42-44,8, o nivel de confianza sería do...

- 90%
- 95%
- 98%

b) Se quixésemos un nivel de confianza do 99% e un intervalo da mesma amplitude, como tería que ser a mostra?

- De menos de 64
- De 64
- De máis de 64

# Exercicios e problemas

## PRACTICA

### Táboas de frecuencias

**1** O número de faltas de ortografía que cometeu un grupo de estudantes nun ditado foi:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 1 | 2 | 0 | 2 | 1 | 3 | 0 | 4 |
| 0 | 1 | 1 | 4 | 3 | 5 | 3 | 2 | 4 | 1 |
| 5 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 2 | 1 |
| 2 | 1 | 0 | 0 | 3 | 0 | 5 | 3 | 2 | 1 |

a) Di cal é a variable e de que tipo é.

b) Fai unha táboa de frecuencias e representa os datos nun diagrama adecuado.

**2** As urxencias atendidas durante un mes nun centro de saúde foron:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 3 | 2 | 1 | 6 | 4 | 2 | 2 | 3 |
| 4 | 3 | 5 | 1 | 0 | 1 | 5 | 3 | 3 | 6 |
| 2 | 4 | 6 | 3 | 2 | 4 | 3 | 2 | 1 | 5 |

a) Cal é a variable e de que tipo é?

b) Fai unha táboa de frecuencias e representa os datos.

**3** Nunha maternidade tomáronse os pesos (en quilogramos) de 50 nenos e nenas acabados de nacer:

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

a) Cal é a variable e de que tipo é?

b) Constrúe unha táboa cos datos agrupados en 6 intervalos de 1,65 a 4,05.

c) Representa graficamente esta distribución.

**4** A un grupo de 30 persoas tomóuselles o número de pulsacións por minuto (ritmo cardíaco) e obtivéronse os seguintes resultados:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 87 | 85 | 61 | 51 | 64 | 75 | 80 | 70 | 69 | 82 |
| 80 | 79 | 82 | 74 | 92 | 76 | 72 | 73 | 63 | 65 |
| 67 | 71 | 88 | 76 | 68 | 73 | 70 | 76 | 71 | 86 |

Representa graficamente esta distribución agrupando os datos en 6 intervalos (desde 50,5 a 92,5).

### Media, desviación típica e C.V.

Acha a media, a desviación típica e o coeficiente de variación nas seguintes distribucións:

**5**

| $x_i$ | $f_i$ |
|---|---|
| 0 | 12 |
| 1 | 9 |
| 2 | 7 |
| 3 | 6 |
| 4 | 3 |
| 5 | 3 |

**6**

| $x_i$ | $f_i$ |
|---|---|
| 0 | 1 |
| 1 | 5 |
| 2 | 6 |
| 3 | 7 |
| 4 | 4 |
| 5 | 4 |
| 6 | 3 |

**7**

| INTERVALO | $f_i$ |
|---|---|
| 1,65-2,05 | 4 |
| 2,05-2,45 | 5 |
| 2,45-2,85 | 13 |
| 2,85-3,25 | 17 |
| 3,25-3,65 | 8 |
| 3,65-4,05 | 3 |

**8**

| INTERVALO | $f_i$ |
|---|---|
| 50,5-57,5 | 1 |
| 57,5-64,5 | 3 |
| 64,5-71,5 | 8 |
| 71,5-78,5 | 8 |
| 78,5-85,5 | 6 |
| 85,5-92,5 | 4 |

**9** Os gastos mensuais dunha empresa *A* teñen unha media de 100 000 euros e unha desviación típica de 12 500 euros. Noutra empresa *B* a media é 15 000 euros, e a desviación típica, 2 500 euros. Calcula o coeficiente de variación e di cal das dúas ten máis variación relativa.

**10** O peso medio dos alumnos dunha clase é de 58,2 kg, e a súa desviación típica, 3,1 kg. O das alumnas desa clase é 52,4 kg e a súa desviación típica é 5,2 kg. Calcula o coeficiente de variación e compara a dispersión de ambos os grupos.

**11** Pedíronse os pesos e as alturas de 6 persoas e obtivéronse os seguintes datos:

| PESO (kg) | ALTURA (m) |
|---|---|
| 65 | 1,70 |
| 60 | 1,50 |
| 63 | 1,70 |
| 63 | 1,70 |
| 68 | 1,75 |
| 68 | 1,80 |

Calcula o coeficiente de variación e di se están máis dispersos os pesos ou as alturas.

# Exercicios e problemas

## Medidas de posición

**12** A mediana e os cuartís da distribución de "Aptitude para a música" (escala 1-100) nun colectivo de persoas son $Q_1 = 31$, $Me = 46$ e $Q_3 = 67$.

Completa as seguintes afirmacións:

a) O 75% ten unha aptitude superior ou igual a ____.

b) O 25% ten unha aptitude superior ou igual a ____.

c) O ____% ten unha aptitude igual ou menor a 46 puntos.

d) O ____% ten unha aptitude superior ou igual a 46 e inferior ou igual a 67.

e) O ____% ten unha aptitude superior ou igual a 31 e inferior ou igual a 67.

**13** A altura, en centímetros, dun grupo de alumnos e alumnas dunha mesma clase é:

| | | | | | | |
|---|---|---|---|---|---|---|
| 150 | 169 | 171 | 172 | 172 | 175 | 181 |
| 182 | 183 | 177 | 179 | 176 | 184 | 158 |

Calcula a mediana e os cuartís e explica o significado destes parámetros.

**14** Calcula a mediana e os cuartís da seguinte distribución:

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $f_i$ | 12 | 9 | 7 | 6 | 3 | 3 |

**15** Acha a mediana, os cuartís e o percentil 60 en cada unha das seguintes distribucións, correspondentes ás notas obtidas nun test que fixeron dous grupos de estudantes:

A: 25 – 22 – 27 – 30 – 23 – 22 – 31 – 18
24 – 25 – 32 – 35 – 20 – 28 – 30

B: 27 – 32 – 19 – 22 – 25 – 30 – 21
29 – 23 – 31 – 21 – 20 – 18 – 27

**16** Na fabricación de certo tipo de lámpadas detectáronse algunhas defectuosas. Estudáronse 200 caixas de 100 lámpadas cada unha e obtívose a seguinte táboa:

| DEFECTUOSAS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| N.° DE CAIXAS | 5 | 15 | 38 | 42 | 49 | 31 | 18 | 2 |

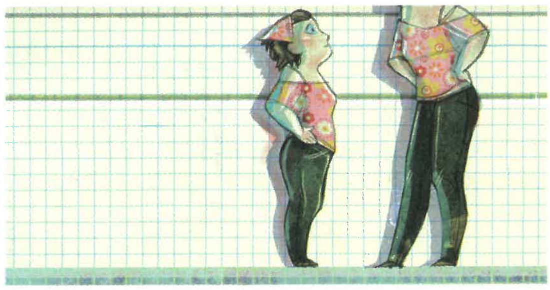
Calcula a mediana, os cuartís e os percentís $p_{10}$, $p_{90}$ e $p_{95}$.

## Diagramas de caixa

**17** As puntuacións obtidas por 87 persoas teñen os seguintes parámetros de posición: $Q_1 = 4{,}1$; $Me = 5{,}1$ e $Q_3 = 6{,}8$. Todas as puntuacións están no intervalo 1 a 9. Fai o diagrama de caixa.

**18** As estaturas de 35 alumnos dunha clase están comprendidas entre 153 e 188. Os tres restantes miden 151, 152 e 190. Coñecemos os seguintes parámetros: $Q_1 = 161$; $Me = 166$ e $Q_3 = 176$.

Fai un diagrama de caixa para esta distribución.



**Fai o diagrama de caixa correspondente ás seguintes distribucións.**

**19** A mesma que a do exercicio 13 anterior.

**20** A mesma que a do exercicio 14 anterior.

**21** A A e a B que se propuxeron no exercicio 15 anterior.

**22** A mesma que a do exercicio 16 anterior.

## Mostraxe

**23** Quérense realizar os seguintes estudos:

I. Tipo de transporte que utilizan os veciños dun barrio para acudir ao seu traballo.

II. Estudos que pensan seguir os alumnos e alumnas dun centro escolar ao rematar a ESO.

III. Idade das persoas que viron unha obra de teatro nunha cidade.

IV. Número de horas diarias que ven a televisión os nenos e nenas da túa comunidade autónoma con idades comprendidas entre 5 e 10 anos.

a) Di en cada un de estes casos cal é a poboación.

b) En cales deles é necesario recorrer a unha mostra? Por que?

**24** Como se pode contar o número aproximado de palabras que ten un certo libro?

— Selecciónanse, abrindo ao chou, unhas cantas páxinas e cóntanse as palabras en cada unha.

— Calcúlase o número medio de palabras por páxina.

— Dáse un intervalo no que poida estar comprendido o número total de palabras.

Faino con algún libro. Ou se non, imaxina que o fixeches e inventa os resultados.

**25** Para facer unha sondaxe electoral nunha poboación de 400 electores, aproximadamente, vaise elixir unha mostra de 200 individuos. Di se che parece válido cada un dos seguintes modos de seleccionalos e explica por que.

a) Pregúntaselle ao alcalde, que coñece a todos os veciños, que individuos lle parecen máis representativos.

b) Elíxense 200 persoas ao chou entre as que acoden á verbena o día do patrón.

c) Selecciónanse ao chou na guía telefónica e son enquisadas por teléfono.

d) Acódese ás listas electorais e selecciónanse ao chou 200 deles.

## PENSA E RESOLVE

**26** Desexamos facer unha táboa con datos agrupados a partir de 384 datos, cuxos valores extremos son 19 e 187.

a) Se queremos que sexan 10 intervalos de amplitude 17, cales serán eses intervalos?

b) Fai outra distribución en 12 intervalos da amplitude que creas conveniente.

**27** Nunha urbanización de 25 familias observouse a variable "número de coches que ten a familia" e obtivéronse os seguintes datos:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 1 | 0 | 1 | 2 | 3 | 1 |
| 0 | 1 | 1 | 1 | 4 | 0 | 1 | 1 | 1 | 4 |
| 3 | 2 | 2 | 1 | 1 | | | | | |

a) Constrúe a táboa de frecuencias da distribución.

b) Fai o diagrama de barras.

c) Calcula a media e a desviación típica.

d) Acha a mediana e os cuartís.

e) Fai o diagrama de caixa.

**28** O número de persoas que acudiron cada día ás clases de natación dunha piscina municipal foron:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 38 | 31 | 54 | 47 | 50 | 56 | 52 | 48 | 55 | 60 |
| 58 | 46 | 47 | 55 | 60 | 53 | 43 | 52 | 46 | 55 |
| 43 | 60 | 45 | 48 | 40 | 56 | 54 | 48 | 39 | 50 |
| 53 | 59 | 48 | 39 | 48 | | | | | |

a) Fai unha táboa de frecuencias agrupando os datos en intervalos.

b) Representa graficamente a distribución.

c) Acha $\bar{x}$ e $\sigma$.

**29** Un dentista observa o número de caries en cada un dos 100 nenos dun colexio e obtén os resultados resumidos nesta táboa:

| N.º DE CARIES | F. ABSOLUTA | F. RELATIVA |
|---|---|---|
| 0 | 25 | 0,25 |
| 1 | 20 | 0,2 |
| 2 | $y$ | $z$ |
| 3 | 15 | 0,15 |
| 4 | $x$ | 0,05 |

a) Completa a táboa obtendo $x$, $y$, $z$.

b) Calcula o número medio de caries.

**30** O número de erros cometidos nun test por un grupo de persoas vén reflectido na seguinte táboa:

| NÚMERO DE ERROS | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| NÚMERO DE PERSOAS | 10 | 12 | 8 | 7 | 5 | 4 | 3 |

a) Acha a mediana e os cuartís inferior e superior, e explica o seu significado.

b) Cal é o número medio de erros por persoa?

**31** Ao preguntarlle a un grupo de persoas canto tempo dedicaron a ver televisión durante unha fin de semana, obtivéronse estes resultados:

| TEMPO EN HORAS | N.º DE PERSOAS |
|---|---|
| [0; 0,5) | 10 |
| [0,5; 1,5) | 10 |
| [1,5; 2,5) | 18 |
| [2,5; 4) | 12 |
| [4; 8) | 12 |

Debuxa o histograma correspondente e acha a media e a desviación típica.

**32** Estas táboas recollen a frecuencia de cada signo nas quinielas durante as 20 primeiras xornadas:

| XORNADA | 1 | X | 2 |
|---|---|---|---|
| 1.ª | 4 | 4 | 6 |
| 2.ª | 9 | 3 | 2 |
| 3.ª | 11 | 2 | 1 |
| 4.ª | 10 | 2 | 2 |
| 5.ª | 8 | 4 | 2 |
| 6.ª | 9 | 4 | 1 |
| 7.ª | 10 | 4 | 0 |
| 8.ª | 8 | 4 | 2 |
| 9.ª | 9 | 5 | 0 |
| 10.ª | 5 | 6 | 3 |
| 11.ª | 9 | 3 | 2 |
| 12.ª | 5 | 6 | 3 |
| 13.ª | 7 | 5 | 2 |
| 14.ª | 4 | 9 | 1 |
| 15.ª | 7 | 3 | 4 |
| 16.ª | 6 | 4 | 4 |
| 17.ª | 8 | 2 | 4 |
| 18.ª | 6 | 7 | 1 |
| 19.ª | 7 | 4 | 3 |
| 20.ª | 7 | 5 | 2 |

a) Fai unha táboa de frecuencias para o número de veces que sae o "1" en cada unha das 20 xornadas:

| $x_i$ | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|
| $f_i$ | | | | | | | | |

Acha a súa media e a súa desviación típica.

b) Fai o mesmo para o "X" e para o "2".

c) Acha o C.V. nos tres casos e compáraos.

**33** Cada alumno dun grupo conta o número de persoas e o número de cans que viven no seu portal.

Suman os seus resultados e obteñen unha mostra coa que se pode estimar o número de cans que hai na súa cidade.

Por exemplo, supoñamos que na súa observación obteñen un total de 747 persoas e 93 cans. E saben que na súa cidade viven 75 000 persoas.

a) Cantos cans estiman que haberá na cidade?

b) Como é de fiable esta estimación?

c) É aleatoria a mostra que utilizaron?

**34** Para facer un estudo sobre os hábitos ecolóxicos das familias dunha cidade, seleccionáronse por sorteo os enderezos, rúa e número, que serán visitados.

Se nun portal vive máis dunha familia, sortearase entre elas a que será seleccionada.

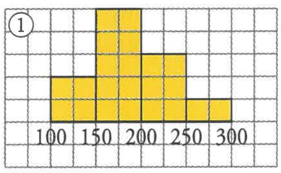
Obteremos con este procedemento unha mostra aleatoria?

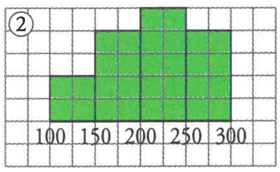
☞ *Pensa se ten a mesma probabilidade de ser incluída na mostra unha familia que vive nunha vivenda unifamiliar que outra que vive, por exemplo, nun bloque de 32 vivendas.*

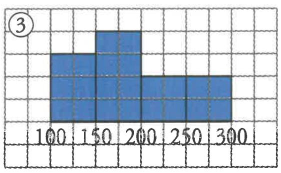
**35** Mediuse o nivel de colesterol en catro grupos de persoas sometidas a diferentes dietas. As medias e as desviacións típicas son as que figuran nesta táboa:

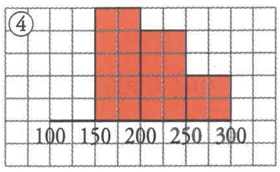
| DIETA | A | B | C | D |
|---|---|---|---|---|
| $\bar{x}$ | 211,4 | 188,6 | 211,7 | 188,6 |
| σ | 37,5 | 52,6 | 49,9 | 43,1 |

As gráficas son, non respectivamente:




Asocia a cada dieta a gráfica que lle corresponde.

## Infórmate e reflexiona

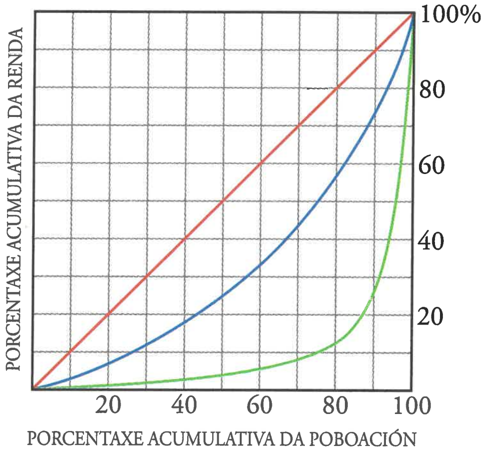
### Cantos abarcan canto?

Estas gráficas representan a distribución da riqueza entre a poboación en tres países diferentes:

A → País desenvolvido: distribución relativamente razoable da renda.

B → País do Terceiro Mundo: uns poucos teñen a maior parte da riqueza.

C → País ideal: a riqueza distribúese por igual entre todos.




- Podes identificar o país que corresponde a cada gráfica?

### As curvas de Lorenz

Estas gráficas chámanse **curvas de Lorenz.** E un dos seus usos máis frecuentes é precisamente a realización de estudos comparativos da distribución da riqueza en distintos colectivos.

## Utiliza o teu enxeño

### Adiviña a clave!

Este cadro serve para cifrar e descifrar mensaxes.


Por exemplo: "*ieoueuooeuuiie ie eoeaieoouu*" significa "*adiviña a clave*".

Escribe en clave: "Podo codificar esta mensaxe".

## Autoavaliación

### Reflexiona sobre a túa aprendizaxe

- Coñeces os parámetros estatísticos $\bar{x}$, $\sigma$ e C.V.? Sábelos calcular e interpretar?
- Coñeces as medidas de posición, mediana, cuartís e centís? Sábelos calcular e interpretar? Sabes utilizalos para construír ou interpretar un diagrama de caixa?

### Verifícao resolvendo exercicios

**1** A idade dos visitantes dunha exposición está recollida na seguinte táboa:

| IDADE | [15-25) | [25-35) | [35-45) | [45-55) | [55-65) | [65-75) |
|---|---|---|---|---|---|---|
| N.° DE VIS. | 63 | 95 | 189 | 243 | 175 | 105 |

a) Representa os datos nun gráfico adecuado.

b) Acha $\bar{x}$, $\sigma$ e C.V.

**2** Os beneficios, en millóns de euros, de dúas empresas en seis anos consecutivos foron os seguintes:

| | | | | | | |
|---|---|---|---|---|---|---|
| A | 5,9 | 2,5 | 7,4 | 8,1 | 4,8 | 3,7 |
| B | 4,5 | 3,8 | 5,7 | 3,5 | 5,5 | 4,6 |

Cal das dúas empresas ten maior variación?

**3** Acha a mediana e os cuartís da seguinte distribución:

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $f_i$ | 27 | 73 | 193 | 62 | 38 | 4 | 2 | 0 | 1 |

Fai o correspondente diagrama de caixa.

7. No teu CD-ROM tes unha **autoavaliación moito máis ampla e completa.** Nel atoparás, ademais, orientacións e, se o desexas, as solucións dos exercicios.