

Introduzione alla Statistica



Il termine statistica deriva da Stato perché è lo Stato che conduce i “censimenti” cioè delle indagini per conoscere il numero degli abitanti, la composizione della popolazione per età, sesso, condizioni economiche (il “censo”) e questo fin dall’antichità.

Si sono poi sviluppate indagini statistiche di vario genere oltre ai “censimenti” dello Stato.

Lo studio statistico dei fenomeni riveste oggi grande importanza per poter risolvere e studiare molti problemi.

Ad esempio uno studio sulla vita media di una popolazione può influenzare le decisioni prese dal governo in campo pensionistico, lo studio degli effetti di un farmaco in via di sperimentazione su un campione di pazienti può far decidere se metterlo in commercio oppure no, in campo medico uno studio statistico può servire a individuare le cause dell’insorgenza di alcune patologie.

Quando si compie un’indagine statistica viene indagata la presenza di un certa caratteristica (carattere) all’interno di una certa “popolazione”.

Il carattere considerato può manifestarsi con modalità diverse e può essere:

un carattere quantitativo se le sue modalità sono espresse da numeri (discreto se può assumere un numero finito di valori o al più un’infinità numerabile o continuo se può assumere tutti i valori di un intervallo reale);

un carattere qualitativo se le sue modalità non sono espresse da numeri.

Tabella statistica e sua rappresentazione

Esempio 1

Qual è lo sport che pratici maggiormente?

Supponiamo di chiedere agli studenti di una classe **quale sport praticano** maggiormente. Il **carattere indagato** (sport preferito) è di **tipo qualitativo**.

La “popolazione statistica” è costituita dagli studenti della classe.

Per ciascuna modalità (calcio, nuoto ecc.) indichiamo il numero degli studenti che hanno indicato quella modalità come sport maggiormente praticato.

Supponiamo di avere ottenuto la seguente tabella:

Sport	n° studenti (frequenza)	frequenza relativa	frequenza relativa percentuale
Calcio	5	$5/20$	25%
Nuoto	2	$2/20=0,1$	10%
Basket	4	$4/20$	20%
pallavolo	6	$6/20$	30%
danza	1	$1/20$	5%
tennis	2	$2/20$	10%

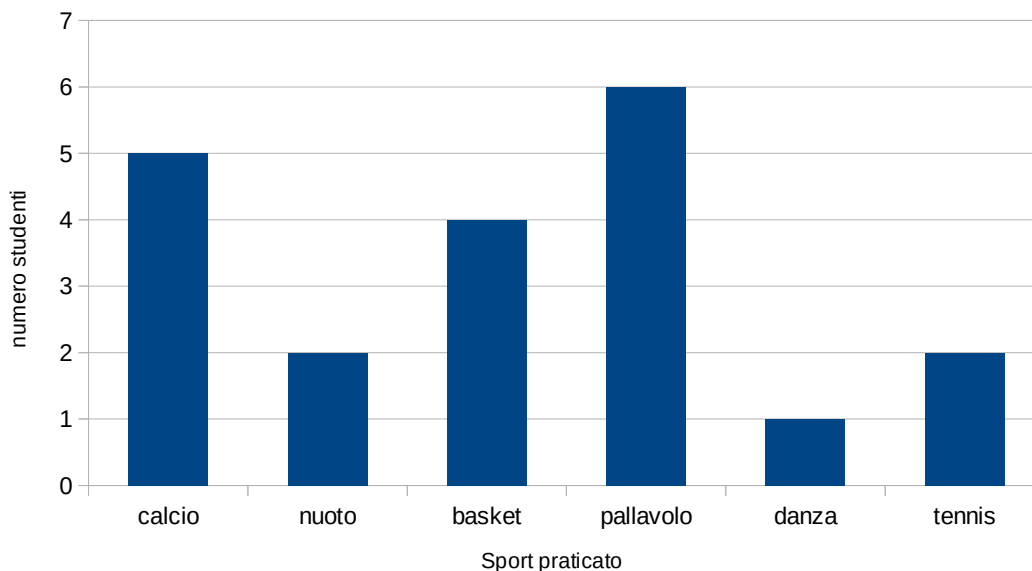
Si chiama **frequenza** (assoluta) di una modalità il numero delle volte che quella data modalità si presenta.

E' molto importante, per poter confrontare le varie frequenze, determinare la **frequenza relativa** cioè **il rapporto tra la frequenza assoluta e il numero delle unità statistiche**, in questo caso il numero degli studenti della classe (20).

La **frequenza relativa percentuale** si ottiene semplicemente moltiplicando per 100 la frequenza relativa.

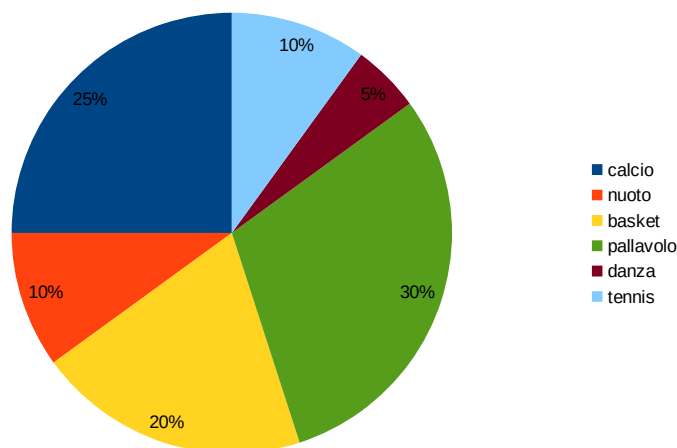
Possiamo rappresentare graficamente i nostri dati con:

- 1) un **diagramma a barre** in cui le basi dei rettangoli distanziati corrispondono alle varie modalità e le altezze sono proporzionali alle **frequenze**



Nota: se i rettangoli sono affiancati si chiama **istogramma**.

- 2) un **aerogramma** in cui un cerchio viene suddiviso in settori circolari corrispondenti alle varie modalità e ampiezza proporzionale alla **frequenza percentuale**:



Nota

Per determinare l'ampiezza α del settore corrispondente ad una data frequenza percentuale f_p basta impostare una proporzione $\alpha : 360^\circ = f_p : 100$.

Se per esempio $f_p = 25$ abbiamo $\alpha : 360^\circ = 25 : 100$ e otteniamo $\alpha = 90^\circ$.

Esempio 2

Qual è la tua altezza?

Supponiamo di chiedere agli studenti di una classe di 20 alunni qual è la propria altezza (espressa in cm).

In questo caso il carattere indagato è **quantitativo**.

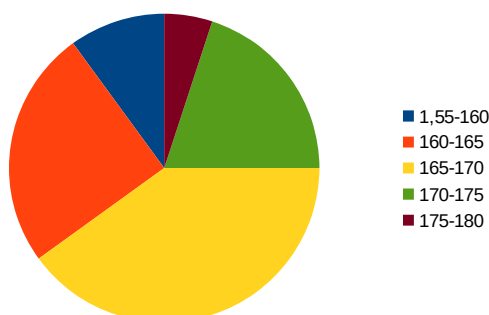
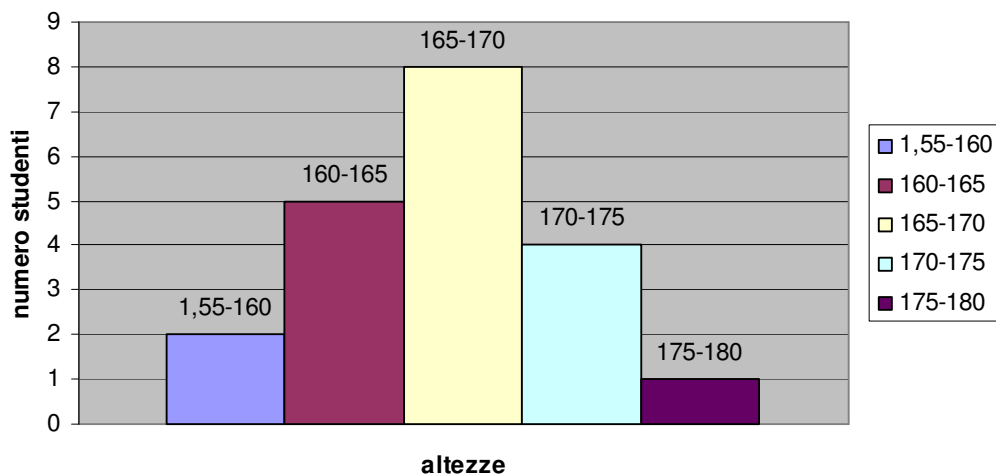
Supponiamo di aver ottenuto i seguenti dati:

156 ; 164 ; 162 ; 168 ;169 ; 176 ;175 ;172 ;161 ; 169 ;165 ;173 ;168 ;167 ;166 ;160 ;164 ;167 ;174 ;170

In questo caso **risulta più opportuno dividere i dati in “classi”** considerando per esempio e di solito l’estremo superiore della classe è incluso nella classe e quello inferiore è escluso (quindi per esempio nella classe 150 – 160 ci saranno 2 studenti).

Classe	Frequenza	Frequenza relativa percentuale
155 - 160	2	10%
160 - 165	5	25%
165 – 170	8	40%
170 - 175	4	20%
175 - 180	1	5%

Per le “classi di valori” conviene utilizzare al posto del diagramma a barre quello che viene chiamato **istogramma** in cui **i rettangoli sono affiancati** e le aree sono proporzionali alle frequenze ma se le “classi hanno la stessa “ampiezza” (quindi le basi dei rettangoli sono uguali) si ritrova che le altezze sono proporzionali alle rispettive frequenze.



Esempio 3

Supponiamo di aver rilevato le seguenti temperature massime nei vari giorni dei mesi di marzo e luglio di un dato anno:

Giorno	Temp. Max. Marzo	Temp. Max Luglio
1	16	28
2	19	29
3	20	29
4	22	27
5	21	26
6	22	24
7	22	26
8	23	26
9	20	28
10	20	28
11	21	30
12	18	30
13	16	31
14	16	32
15	14	32
16	19	30
17	20	31
18	18	29
19	19	28
20	22	32
21	24	33
22	24	32
23	20	30
24	24	30
25	25	29
26	25	32
27	24	33
28	22	33
29	21	30
30	17	30
31	16	30

Per una serie di dati quantitativi può essere interessante determinare la **media aritmetica** \bar{x} cioè la somma di tutti i dati x_1, \dots, x_n divisa per il numero dei dati cioè

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

Nota

Nel nostro caso per calcolarla possiamo sommare tutte le temperature oppure determinare prima la frequenza di ciascuna temperatura: se per esempio nel mese di Luglio la temperatura 24 ha frequenza 1, la temperatura 26 ha frequenza 3, la temperatura 27 frequenza 1, la temperatura 28 frequenza 4...possiamo scrivere

$$media_aritmetica = \frac{24 \cdot 1 + 26 \cdot 3 + 27 \cdot 1 + 28 \cdot 4 + \dots}{31}$$

Otteniamo:

Temp max media Marzo 20,3	Temp max media Luglio 29,6
---------------------------------	----------------------------------

Nota

Un altro “indice” che a volte viene richiesto è la modalità che si presenta con la massima frequenza (si parla di **moda della distribuzione**)

Nel nostro caso abbiamo

Temp moda Marzo 20	Temp moda Luglio 30
--------------------------	---------------------------

Scheda di lavoro 1

Apriamo **Open Office** e poi il Foglio elettronico: comparirà un foglio con righe e colonne e una serie di icone-strumenti.

Supponiamo di voler riportare in questo foglio **la distribuzione delle provenienze degli studenti della classe nel corrente anno scolastico.**

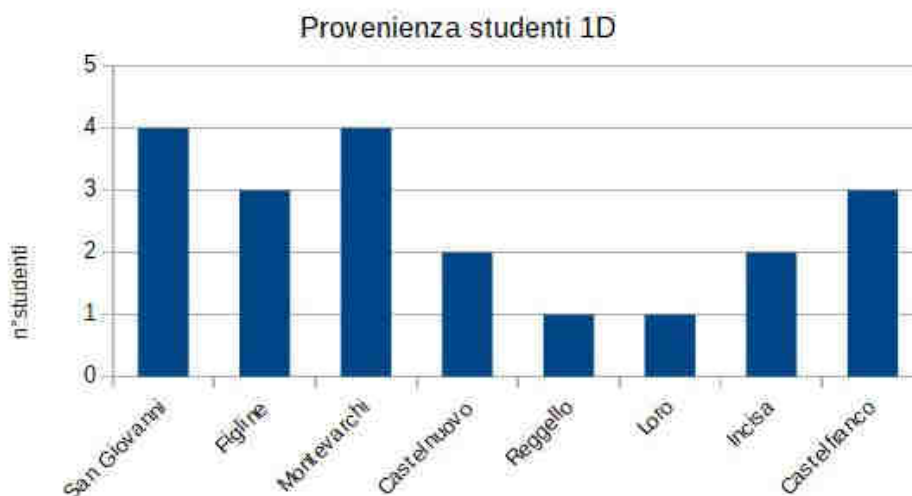
Inseriamo nella prima colonna (A) i nomi dei paesi di provenienza e nella colonna accanto (B) il rispettivo numero di studenti che provengono da quel paese.

Possiamo visualizzare questi dati disegnando il corrispondente diagramma a barre seguendo questa procedura:

- selezionare la zona dei nostri dati (trascinando il mouse fino a che la zona non risulta evidenziata in azzurro);
- scegliere dalla barra delle applicazioni l'icona con le colonne colorate (o inserisci grafico);
- scegliere il tipo di grafico, nel nostro caso "Colonna".

Cliccando su "Avanti" arriviamo ad "Elementi del grafico" e possiamo inserire il titolo del grafico e i sotto-titoli per l'asse x e l'asse y (per esempio scrivere come titolo "Provenienza alunni classe... a.s. ..." e come sotto-titoli degli assi "Paesi" e "n° studenti").

Paesi	n°studenti	Frequenza %
San Giovanni	4	20%
Figline	3	15%
Montevarchi	4	20%
Castelnuovo	2	10%
Reggello	1	5%
Loro	1	5%
Incisa	2	10%
Castelfranco	3	15%



Nota: possiamo stampare il nostro foglio di lavoro con file-stampa ma è meglio controllare prima con anteprima di stampa per poter modificare eventualmente la posizione del grafico (basta fare clic vicino ad un angolo e quando compare una crocina trascinare il grafico nella posizione che vogliamo).

Se vogliamo cancellare un grafico basta cliccare nell'area del grafico e premere il tasto Canc.

Calcoliamo le frequenze **percentuali** di studenti provenienti dai vari paesi.

Ricordiamo che se gli studenti provenienti da San Giovanni sono 4 su un totale di 20 studenti della 1D, la frequenza relativa percentuale risulta:

$$\frac{4}{20} \cdot 100 = 0,2 \cdot 100 = 20 \rightarrow 20\%$$

Il foglio elettronico può essere usato per **ripetere lo stesso tipo di calcolo per tutti i paesi** se procediamo in questo modo: dobbiamo usare non il numero 4 ma il **nome della cella** (casella) in cui si trova il dato, nel nostro caso B2, e **far precedere il calcolo dal segno di =** per indicare che stiamo inserendo una “formula” che poi “estenderemo” anche alle altre righe.

Digitiamo nella cella C1 “frequenza %” ed inseriamo nella cella C2

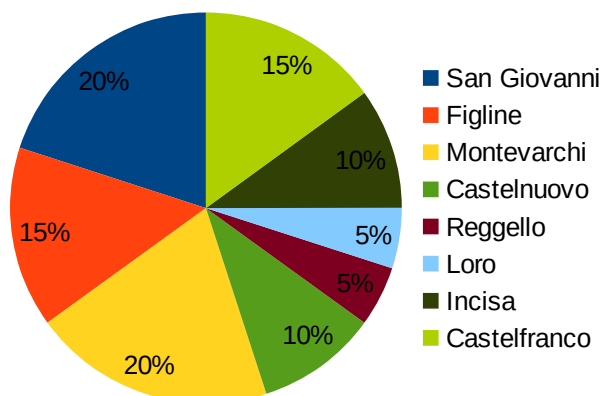
$$= B2 / 20$$

e poi con il tasto destro formatta celle- percentuale- posizioni decimali 1.

A questo punto per “**estendere**” la formula e calcolare le altre percentuali basta posizionare il cursore sull'angolo in basso a destra della cella finché non compare una crocetta e tenendo premuto il tasto del mouse trascinare il cursore fino alla cella desiderata : vedremo comparire tutte le altre frequenze percentuali!

Possiamo visualizzare in modo significativo queste percentuali utilizzando il **grafico a torta**:

- selezioniamo la colonna A e la colonna C (per selezionare colonne di dati non adiacenti occorre tenere premuto il tasto CTRL);
- scegliamo dallo strumento grafico il tipo “torta”;
- andando avanti possiamo inserire il titolo;
- cliccando su un qualsiasi settore della torta con il tasto destro abbiamo la possibilità di inserire l'indicazione delle percentuali scegliendo “inserisci etichette dati”.



Scheda di lavoro 2

Riprendiamo le temperature massime rilevate nel mese di Marzo di un dato anno come sono riportate nell'esempio 2 ed inseriamole in un foglio elettronico di Open Office.

Nota: per inserire i numeri da 1 a 31 possiamo inserire nella cella A2 il numero 1 e nella cella A3 la formula

$$=A2+1$$

ed “estenderla” poi fino ad avere 31.

Possiamo calcolare la temperatura massima “media” utilizzando la funzione **MEDIA**: basta scrivere, se i dati relativi alle temperature si trovano nelle celle da B2 a B32

$$=media(b2:b32)$$

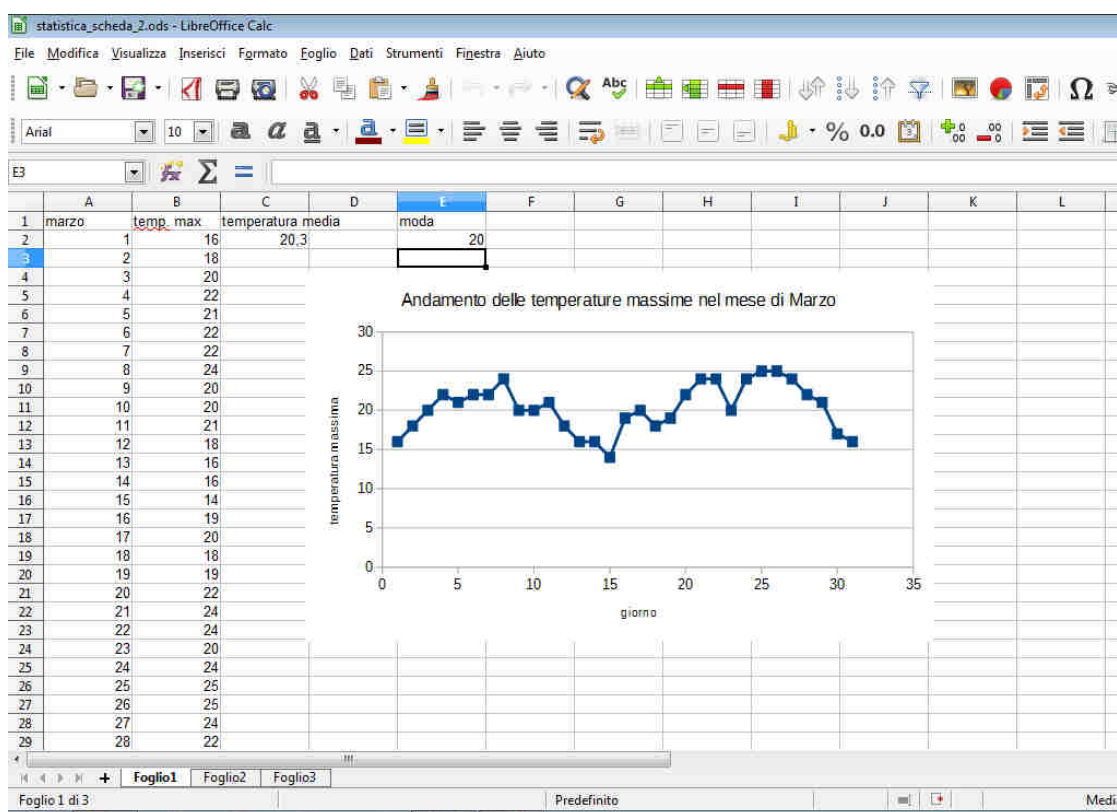
Possiamo calcolare la temperatura che ha avuto la massima frequenza usando la funzione **MODA**

$$=moda(b2:b32)$$

In questo caso per rappresentare graficamente l'andamento delle temperature conviene utilizzare il tipo di grafico XY: dopo aver selezionato le celle delle temperature b2...b32 e attivato “inserisci grafico” scegliamo **XY dispersione - punti e linee**.

Facendo clic con il tasto destro del mouse nella zona del grafico si possono aggiungere titolo del grafico e degli assi.

Il grafico può essere spostato andando negli angoli finché non compare una crocetta e trascinando.



ESERCIZI

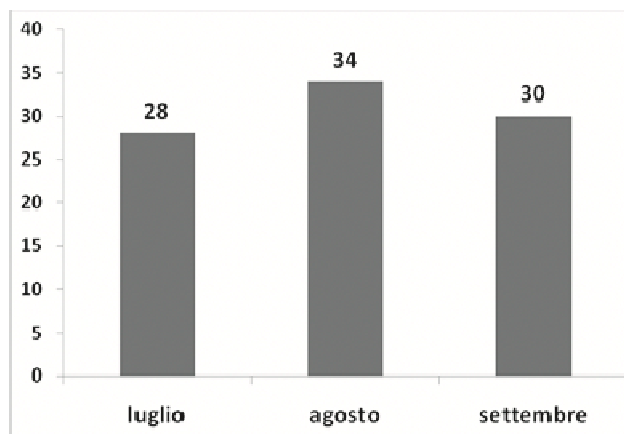
- 1) Con una bilancia si è misurata 10 volte la massa di una lastra di alluminio ottenendo le seguenti misure in chilogrammi:

10,55	10,76	10,60	10,87	10,64	10,67	10,84	10,46	10,55	10,70
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Determina la media aritmetica (approssimata alla seconda cifra decimale).

[10,66 Kg]

- 2) Il grafico riporta il numero di e-book reader (lettori di libri elettronici) venduti nei mesi di luglio, agosto e settembre da un negozio di informatica. Negli altri nove mesi lo stesso negozio ha venduto in media 18 e-book reader al mese. Qual è il numero medio **mensile** di e-book reader venduti in quell'anno in negozio?



[circa 21]

- 3) Un sondaggio condotto su un gruppo di 51 studenti sul numero di televisori presenti in casa ha dato i seguenti risultati

N° televisori	N° studenti
1	10
2	15
3	18
4	7

Qual è la percentuale di studenti che ha in casa meno di tre televisori?

[50%]

- 4) In un grande magazzino è stata compiuta un'indagine sulle taglie di pantaloni da uomo venduti durante una settimana e si sono ottenuti i dati riassunti nella seguente tabella:

Taglia	44	46	48	50	52	54	56
Frequenza	2	7	6	12	10	2	1

Qual è la frequenza percentuale della taglia 48?

Qual è la frequenza percentuale di pantaloni di taglia superiore alla 50?

[15 % ; 32,5 %]

- 5) In una scuola di 240 studenti il 20% ha 14 anni, il 40% ha 15 anni, il 25% ha 16 anni e il 15% ha 17 anni. Quanti studenti hanno 17 anni? Quanti studenti hanno meno di 16 anni? Quanti studenti hanno più di 15 anni?

[36 ; 144 ; 96]

- 6) I voti dell'ultimo compito di matematica in una classe sono stati i seguenti:

5 ; 6 ; 6 ; 7 ; 7,5 ; 6,5 ; 8 ; 4 ; 7 ; 7 ; 10 ; 3 ; 4,5 ; 5,5 ; 8 ; 8 ; 7 ; 9 ; 4 ; 6

Compila la tabella delle frequenze dei vari voti e determina la media e la moda della distribuzione.

[6,45 ; 7]

- 7) Considera la seguente tabella che riporta i pesi degli studenti di una classe (indicati con il numero corrispondente al loro nome nell'elenco di classe).

Studente	Peso (Kg)
1	56
2	58
3	60
4	64
5	56
6	62
7	70
8	74
9	71
10	70
11	65
12	60
13	55
14	72
15	65

Determina il peso medio (approssimando alla prima cifra decimale).

Rappresenta i dati con un istogramma dividendoli nelle "classi" di valori 55-60 ; 60-65 ecc.

[63,9 Kg]

TEST
STATISTICA

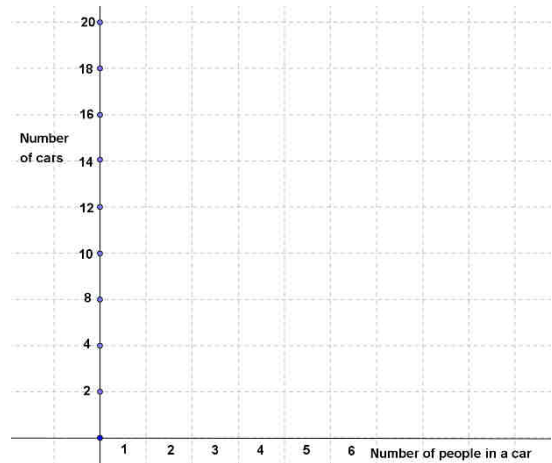
1) A travel brochure contains 24 pictures from different countries. The table shows how many pictures there are from each country.

- a) Complete the table
b) Complete the pie chart accurately and label the sectors for South Africa, Australia and New Zealand.

Country	Numbers of pictures	Angle in pie chart
Argentina	6	90°
South Africa	10	150°
Australia	3	
New Zealand		

2) Marie counts the number of people in each of 60 cars one morning.

Number of people in a car	Number of cars
1	6
2	17
3	8
4	9
5	11
6	9



On the grid,

draw a bar chart to show the information for the 60 cars.

- a) Write down the mode
b) Manuel uses Marie's result to draw a pie chart. Work out the sector angle for the number of cars with 5 people.

3) The table shows how many books were borrowed by the 126 members of a library group in a month.

Number of books	11	12	13	14	15	16
Number of members (frequency)	35	28	22	18	14	9

Find the mode, the median and the mean for the number of books borrowed.

4) A school has 350 students.

- a) On the school sports day 96% of the students were present. Calculate how many students were absent.
b) The table shows the number of students attending school in one week.

Monday	Tuesday	Wednesday	Thursday	Friday
334	329	348	341	323

For these values, calculate the mean, the median and the range.