Higher Unit 3— Interpreting and representing data

Key point 2

To draw a **frequency polygon** you can join the midpoints of the tops of the bars in a frequency diagram with straight lines.

Key point 3

To draw a frequency polygon, plot the frequency against the midpoints for each group.

Key point 1

A **back-to-back stem and leaf diagram** compares two sets of results.

Example 1

The annual salaries of employees working in an ICT company are displayed in the back-to-back stem and leaf diagram.

Key Male Female

8 | 1 represents a salary of £18000 1 | 9 represents a salary of £19000

Male					Female			
			8	1	9	9		
9	5	2	0	2	1	2	6	7
8	7	3	0	3	0	4	4	
				4	5	6		
				5	4	8		

Compare the distribution of salaries of the male and female employees.

Male range: 38000 – 18000 = £20000

Female range: 58000 – 19000 = £39000

There are 9 males, so median male salary is: $\frac{9+1}{2}$ = 5th value = £29000

There are 13 females so median female salary is: $\frac{13+1}{2}$ = 7th value = £30000

Female employees' salaries have a larger range but the median salaries of men and women are similar.

Write a sentence comparing ranges and medians.

Key point 4

A **time series** graph is a line graph with time plotted on the horizontal axis.

Example 2

The table shows the quarterly price of a tonne of wheat (in dollars) during the last three years.

2012				2013				2014			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
250	279	101	157	348	371	230	264	451	477	322	347

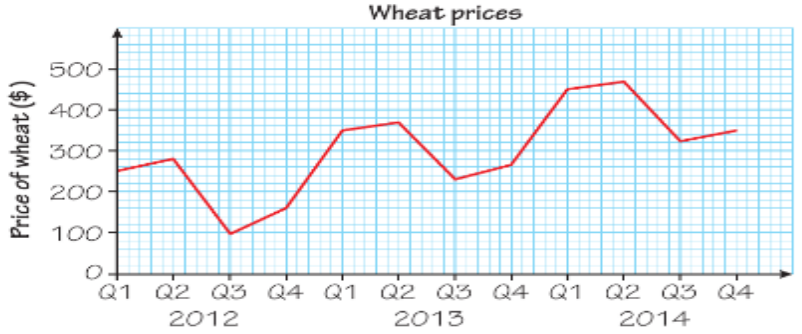
Communication hint
Prices are recorded every 3 months so the first quarter covers January, February and March.

a What is the price in the third quarter (Q3) of 2013?
b In which quarter is the price the lowest?
c Draw a time series graph of the data.
d Describe the variation in prices during this period and comment on the overall trend.

a $230

b The lowest price is $101 which occurs in the third quarter of 2012.

c



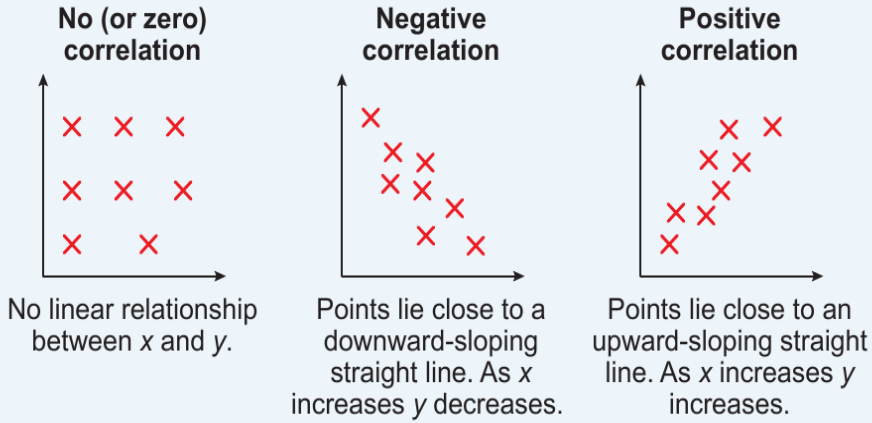
d The price of wheat fluctuates up and down during the course of each year.

Key point 7

A **line of best fit** is the line that passes as close as possible to the points on a scatter graph.

Key point 6

A scatter graph shows a relationship or correlation between variables.



No linear relationship between x and y.

Points lie close to a downward-sloping straight line. As x increases y decreases.

Points lie close to an upward-sloping straight line. As x increases y increases.

Key point 11

Line graphs are useful for tracking changes over time. Pie charts are good when comparing parts of a whole. Bar charts are used to compare the frequencies of two data sets.

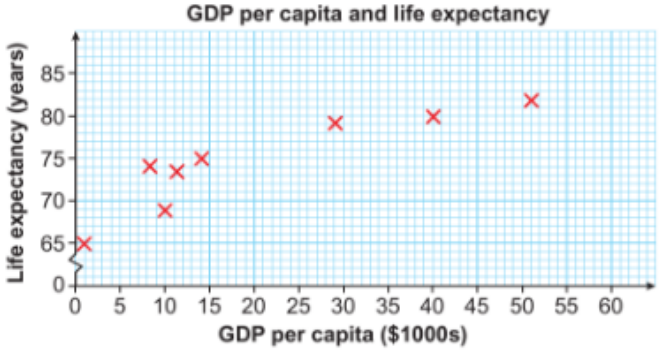
Key point 10

If the total frequency in a grouped frequency table is n, then the median lies in the class containing the $\frac{n+1}{2}$th item of data.

The **modal class** has the highest frequency.

Example 3

The scatter graph shows the GDP per capita (in $1000s) and life expectancy (in years) for eight countries.

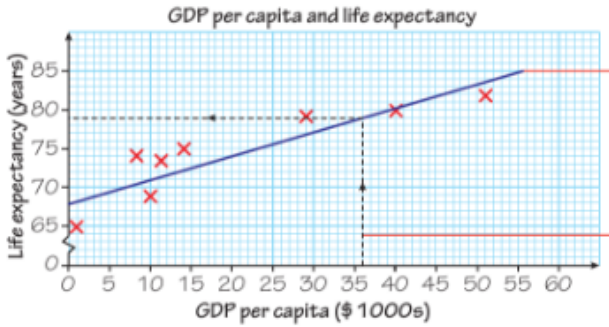


Communication hint
The gross domestic product (GDP) measures the value of goods and services produced by a country. The GDP per capita is the GDP divided by the number of people in that country.

a Draw a line of best fit.

b The GDP per capita in the UK is $36000. Estimate the life expectancy of a baby born in the UK.

a



Position a transparent ruler over your scatter graph so it follows the overall trend. Move it slightly so you have roughly the same number of points above and below the line.

To estimate life expectancy, start at $36000 on the horizontal axis, go up to the line of best fit and read off the answer on the vertical axis.

b Estimated life expectancy in the UK is 79 years.

Key point 8

Using a line of best fit to predict data values within the range of the data given is called **interpolation** and is usually reasonably accurate.

Using a line of best fit to predict data values outside the range of the data given is called **extrapolation** and many not be accurate.

Key point 9

Individual points which are outside the overall pattern of a scatter graph are called **outliers**. If they are likely to be from incorrect readings you can ignore them.

Example 4

The table shows the times, T, taken for 100 people to queue for a rollercoaster at a theme park.

a Estimate the mean waiting time.

b Explain why the mean is only an estimate.

a

Time, T (mins)	Frequency, f	Class midpoint, x	xf
$0 \leqslant T < 20$	14	10	$10 \times 14 = 140$
$20 \leqslant T < 40$	55	30	$30 \times 55 = 1650$
$40 \leqslant T < 60$	31	50	$50 \times 31 = 1550$
Total	100		3340

The third column gives an estimate of the waiting time in each class.

The fourth column gives an estimate of the total waiting time in each class.

$$\text{Mean} = \frac{\text{sum of waiting times}}{\text{total number of people}} = \frac{3340}{100}$$

$$= 33.4 \text{ minutes}$$

b The mean is an estimate because we don't know the exact times taken.

Discussion What assumptions have been made about the data?