

---

# S1 Chapter 3: Data Representations

## Box Plots and Outliers

# This Chapter Overview

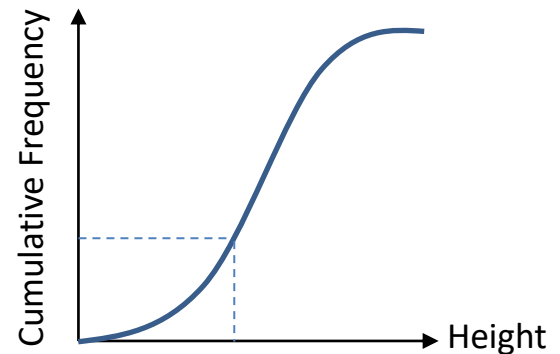
We've seen so far how data is collected and calculations can be made. We now concentrate on how the processed data can be *displayed*.

## BOX PLOTS AND OUTLIERS

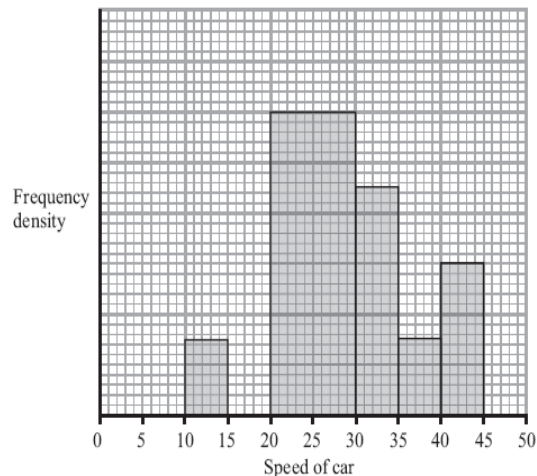


**\*NEW since GCSE!\*** Outliers.

## CUMULATIVE FREQ DIAGRAMS



## HISTOGRAMS



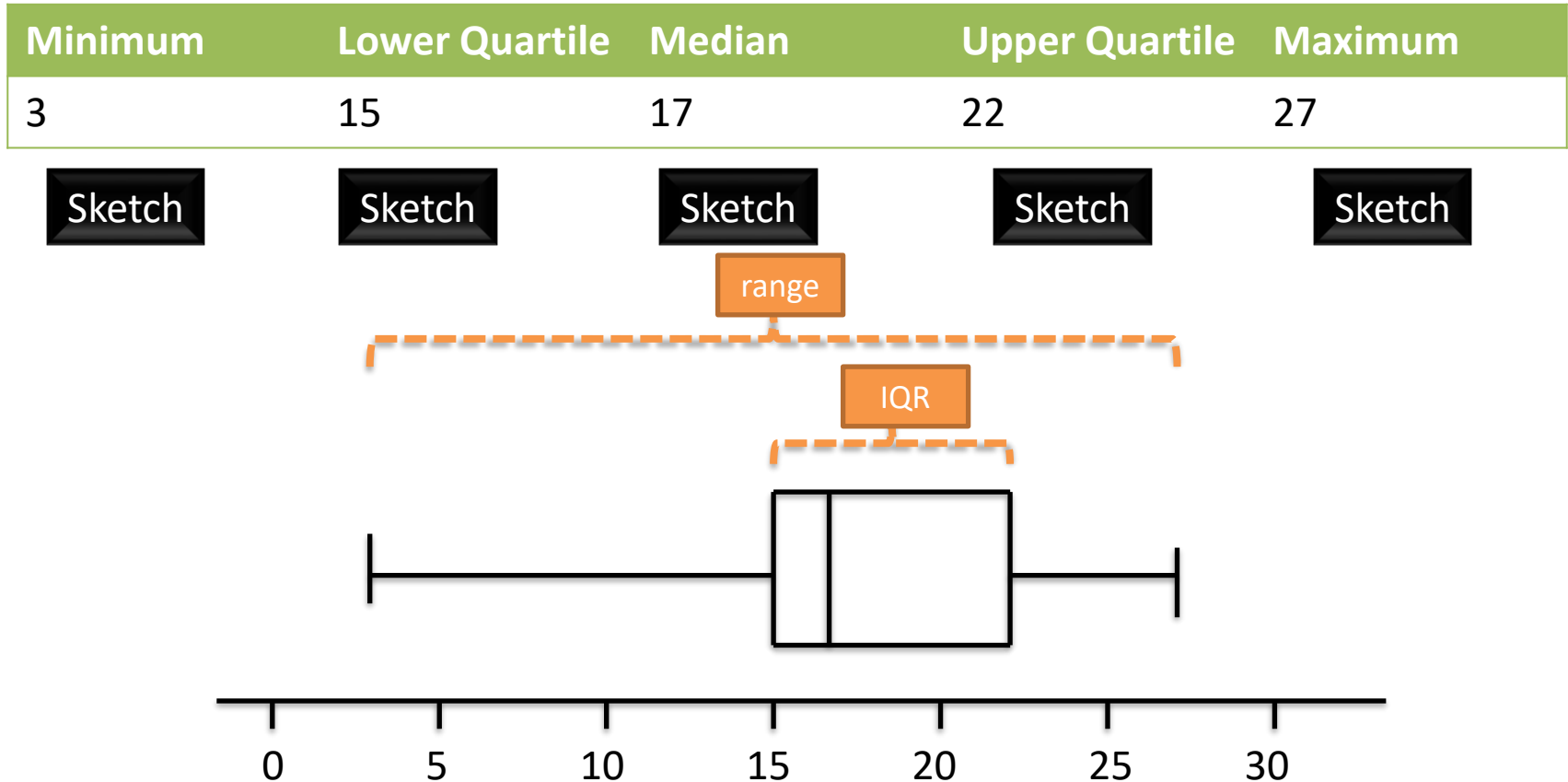
**\*NEW since GCSE!\*** Area is not necessarily equal to frequency.  
Forming a frequency polygon by joining midpoints.

### Changes since the old 'S1' syllabus:

- Stem and leaf diagrams have been cut. (THANK GOD FOR THAT)
- 'Skew' has been cut.
- Cumulative frequency diagrams have been added.
- Turning histogram into frequency polygon.

# Box Plot recap

Box Plots allow us to visually represent the distribution of the data.



How is the **IQR** represented in this diagram?

Sketch

How is the **range** represented in this diagram?

Sketch

# Interpreting a Box Plot



True or false: (click your answer)

**“The right box represents more people than the left box.”**

False

True

Each box represents 25% of people, i.e. the same number of people!

**“The ages are more spread out above the median.”**

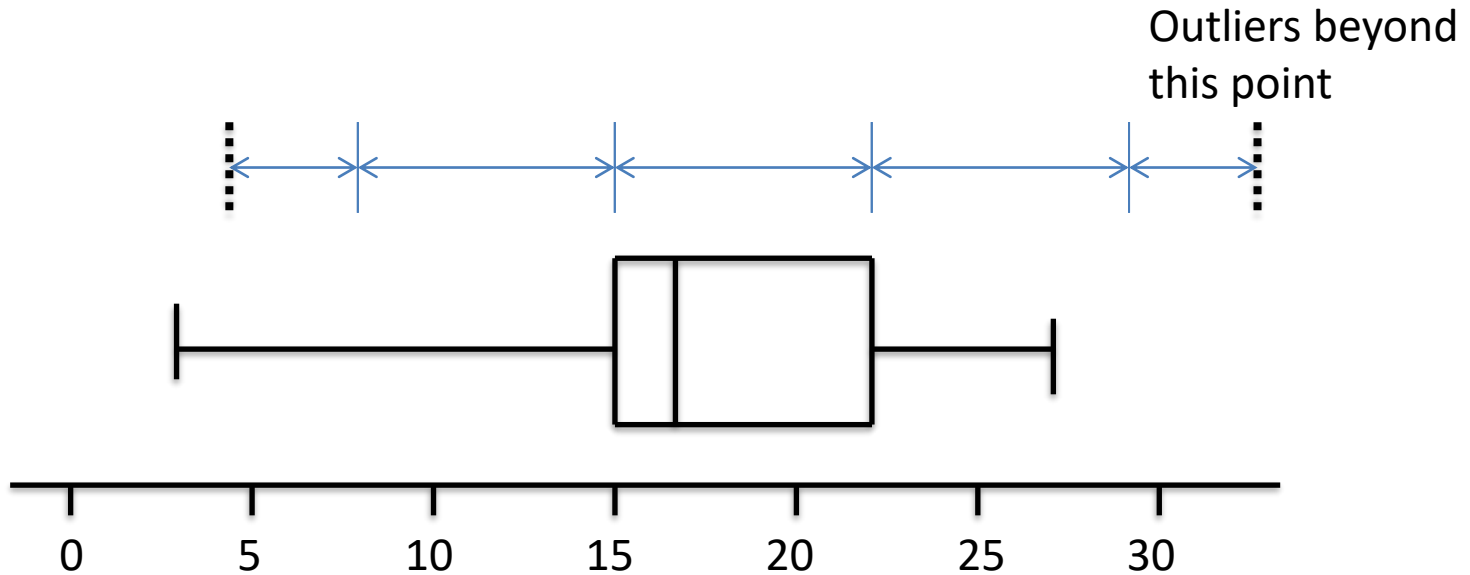
False

True

The wider the box or whisker, the more spread out the values are within that 25% of the data. We'd say that the data has **“positive skew”**, but you are not required to know this term.

# Outliers

An outlier is: **an extreme value.**



One common definition of an outlier is when we're **1.5 IQRs** beyond the lower and upper quartiles.  
(But you will be told in the exam if the rule differs from this)

# Examples

The diameters of 11 different Roman coins are measured in centimetres:

2.2 2.5 2.7 2.7 2.8 3.0 3.1 3.1 3.2 4.0 4.7

Determine the quartiles and hence any outliers.

?

[Textbook] The lengths, in cm, of 12 giant African land snails are given below:

17 18 18 19 20 20 20 20 21 23 24 32

- Calculate the mean and standard deviation, given that  $\Sigma x = 252$  and  $\Sigma x^2 = 5468$ .
- An outlier is an observation which lies  $\pm 2$  standard deviations from the mean. Identify any outliers for this data.

? a

? b

**Context:** Recall that the standard deviation is, roughly speaking, the average distance of each value from the mean. So the outlier definition is saying we're at least twice this average distance, which seems like a sensible definition.

In Year 2, you will encounter the **normal distribution**, which can be used to model data which is **clustered about some mean and tails off symmetrical in either direction**. If this data was approximately normally distributed, then there is a 5% chance a random observation would fall outside 2 standard deviations within the mean. You will learn then how to make such probability calculations.

# Examples

The diameters of 11 different Roman coins are measured in centimetres:

2.2 2.5 2.7 2.7 2.8 3.0 3.1 3.1 3.2 4.0 4.7

Determine the quartiles and hence any outliers.

$$Q_1 = 3^{\text{rd}} \text{ item} = 2.7 \quad Q_3 = 9^{\text{th}} \text{ item} = 3.2$$

$$IQR = 3.2 - 2.7 = 0.5$$

$$\text{Lower outlier boundary} = 2.7 - 1.5 \times 0.5 = 1.95$$

$$\text{Upper outlier boundary} = 3.2 + 1.5 \times 0.5 = 3.95$$

Therefore outliers are 4.0cm and 4.7cm.

[Textbook] The lengths, in cm, of 12 giant African land snails are given below:

17 18 18 19 20 20 20 20 21 23 24 32

a) Calculate the mean and standard deviation, given that  $\Sigma x = 252$  and  $\Sigma x^2 = 5468$ .

b) An outlier is an observation which lies  $\pm 2$  standard deviations from the mean.  
Identify any outliers for this data.

$$\bar{x} = \frac{252}{12} = 21 \text{ cm}$$

$$\sigma = \sqrt{\frac{5468}{12} - 21^2} = 3.83 \text{ (3sf)}$$

$$\text{Outlier boundaries: } 21 - 2 \times 3.83 = 13.34$$

$$21 + 2 \times 3.83 = 28.66$$

Outlier is just 32.

**Context:** Recall that the standard deviation is, roughly speaking, the average distance of each value from the mean. So the outlier definition is saying we're at least twice this average distance, which seems like a sensible definition.

In Year 2, you will encounter the **normal distribution**, which can be used to model data which is **clustered about some mean and tails off symmetrical in either direction**. If this data was approximately normally distributed, then there is a 5% chance a random observation would fall outside 2 standard deviations within the mean. You will learn then how to make such probability calculations.

# Test Your Understanding

The ages of 15 Lib Dem MPs are given:

11 18 20 27 30 31 32 32 35 36 37 58 63 78 105

- a) If an outlier is considered to be 1.5 interquartile ranges below the lower quartile or above the upper quartile, determine any outliers.
- b) If instead an outlier is considered to be outside 2 standard deviations within the mean, determine any outliers. Note that  $\Sigma x = 613$  and  $\Sigma x^2 = 33815$



? a

? b



# Test Your Understanding

The ages of 15 Lib Dem MPs are given:

11 18 20 27 30 31 32 32 35 36 37 58 63 78 105

- a) If an outlier is considered to be 1.5 interquartile ranges below the lower quartile or above the upper quartile, determine any outliers.
- b) If instead an outlier is considered to be outside 2 standard deviations within the mean, determine any outliers. Note that  $\Sigma x = 613$  and  $\Sigma x^2 = 33815$



$$Q_1 = 27, \quad Q_3 = 58, \quad IQR = 31$$

$$\text{Boundaries: } 27 - 1.5 \times 31 = -19.5$$

$$58 + 1.5 \times 31 = 104.5$$

Therefore the 105 year old is an outlier.

$$\bar{x} = \frac{613}{15} = 40.9$$

$$\sigma = \sqrt{\frac{33815}{15} - \left(\frac{613}{15}\right)^2} = 24.2 \text{ (3sf)}$$

$$\text{Boundaries: } 40.9 + 24.2 \times 2 = 89.3$$

$$40.9 - 24.2 \times 2 = -7.5$$

Therefore the 105 year old is an outlier.

# Box Plot Example

Smallest values	Largest values	Lower Quartile	Median	Upper Quartile
0, 3	21, 27	8	10	14

Draw a box plot to represent the above data.

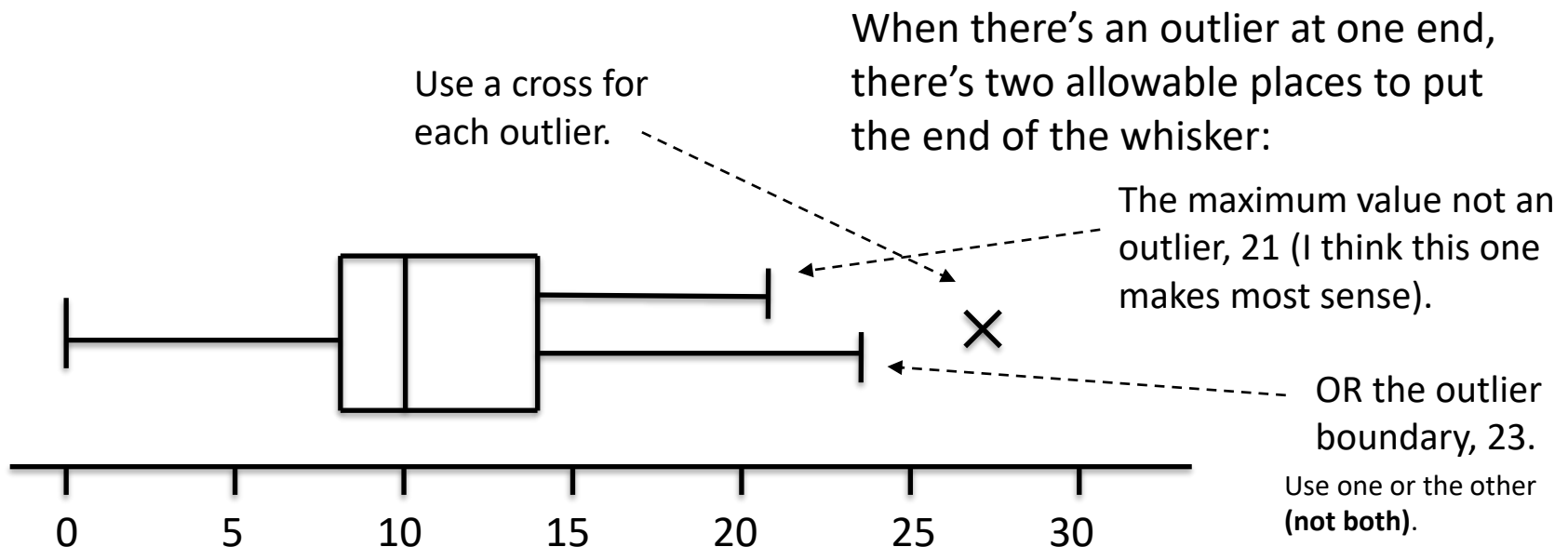
$$IQR = 14 - 8 = 6$$

Outlier boundaries:

$$14 + (1.5 \times 6) = 23$$

$$8 - (1.5 \times 6) = -1$$

**Fro Exam Tip:** You MUST show your outlier boundary calculations.



# Test Your Understanding

[Jan 2011 Q3] Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls either  $1.5 \times$  interquartile range above the upper quartile or  $1.5 \times$  interquartile range below the lower quartile.

- (a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers. (5)

(a)

a ?

M1  
A1

- (c) The company claims that for 75% of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer. (2)

M1

A1ft (c)

B1

c ?

# Test Your Understanding

[Jan 2011 Q3] Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

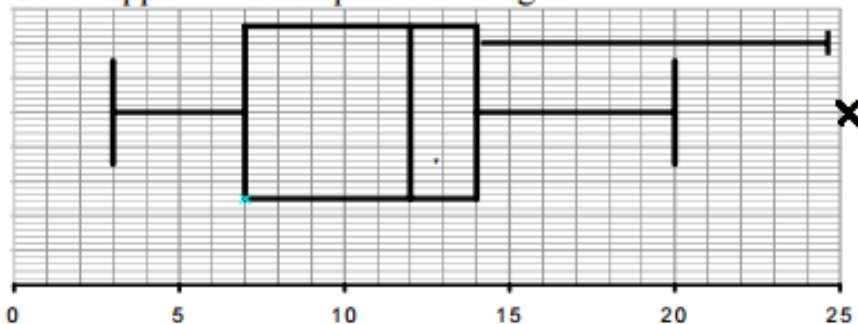
	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls either  $1.5 \times$  interquartile range above the upper quartile or  $1.5 \times$  interquartile range below the lower quartile.

- (a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers. (5)

- (a) Outliers  
 $14 + 1.5 \times (14 - 7) = 24.5$   
 $7 - 1.5 \times (14 - 7) = -3.5$

Outlier 25  
 either upper limit acceptable on diagram



- (c) The company claims that for 75% of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer. (2)

M1  
 A1

M1

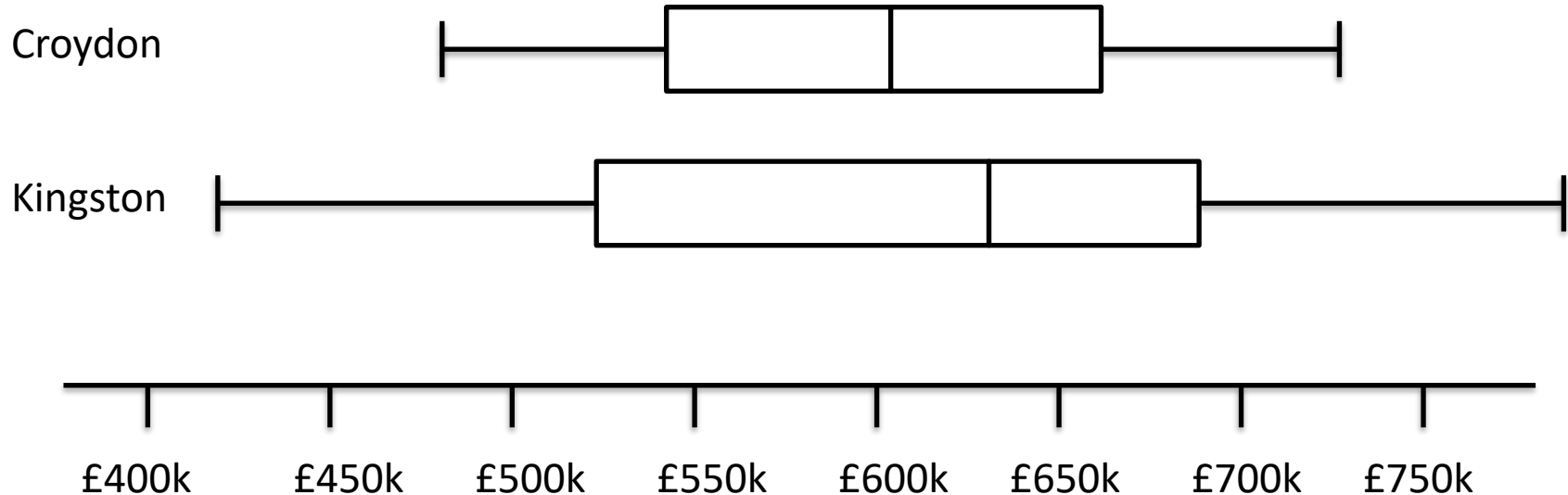
A1ft (c) not true  
 since the lower quartile is 7000 and therefore 75% above 7000 not 10000 or 10 is inside the box or any other sensible comment

B1

B1  
 dB1

# Comparing Box Plots

Box Plot comparing house prices of Croydon and Kingston-upon-Thames:



“Compare the prices of houses in Croydon with those in Kingston”. (2 marks)

For 1 mark, one of:

- In **interquartile range** of house prices in Kingston is greater than Croydon.
- The **range** of house prices in Kingston is greater than Croydon.

Include some measure of **spread**.

For 1 mark:

“The **median** house price in Kingston was greater than that in Croydon.”

Include some measure of **location** (median is best).

# Exercise 3.1 / 3.2

Pearson Pure Mathematics Year 1/AS

- 1) Pages 17 - 18
  - 2) Pages 18 - 19
-

# Homework Exercise

- 1 Some data is collected.  $Q_1 = 46$  and  $Q_3 = 68$ .

A value greater than  $Q_3 + 1.5 \times (Q_3 - Q_1)$  or smaller than  $Q_1 - 1.5 \times (Q_3 - Q_1)$  is defined as an outlier.

Work out whether the following are outliers using this rule:

- a 7                      b 88                      c 105

- 2 The masses of male and female turtles are given in grams. For males, the lower quartile was 400 g and the upper quartile was 580 g. For females, the lower quartile was 260 g and the upper quartile was 340 g.

An outlier is an observation that falls either  $1 \times$  (interquartile range) above the upper quartile or  $1 \times$  (interquartile range) below the lower quartile.

- a Which of these male turtle masses would be outliers?

400 g      260 g      550 g      640 g

- b Which of these female turtle masses would be outliers?

170 g      300 g      340 g      440 g

- c What is the largest mass a male turtle can be without being an outlier?

**Hint** The definition of an outlier here is different from that in question 1. You will be told which rule to use in the exam.

- 3 The masses of arctic foxes are found and the mean mass was 6.1 kg. The variance was 4.2.

An outlier is an observation which lies  $\pm 2$  standard deviations from the mean.

- a Which of these arctic fox masses are outliers?

2.4 kg      10.1 kg      3.7 kg      11.5 kg

- b What are the smallest and largest masses that an arctic fox can be without being an outlier?

# Homework Exercise

4 The ages of nine people at a children's birthday party are recorded.  $\Sigma x = 92$  and  $\Sigma x^2 = 1428$ .

a Calculate the mean and standard deviation of the ages. (3 marks)

An outlier is an observation which lies  $\pm 2$  standard deviations from the mean.

One of the ages is recorded as 30.

b State, with a reason, whether this is an outlier. (2 marks)

c Suggest a reason why this age could be a legitimate data value. (1 mark)

d Given that all nine people were children, clean the data and recalculate the mean and standard deviation. (3 marks)

## Problem-solving

After you clean the data you will need to find the new values for  $n$ ,  $\Sigma x$  and  $\Sigma x^2$ .

5 A group of students did a test. The summary data is shown in the table.

Lowest value	Lower quartile	Median	Upper quartile	Highest value
5	21	28	36	58

Given that there were no outliers, draw a box plot to illustrate this data.

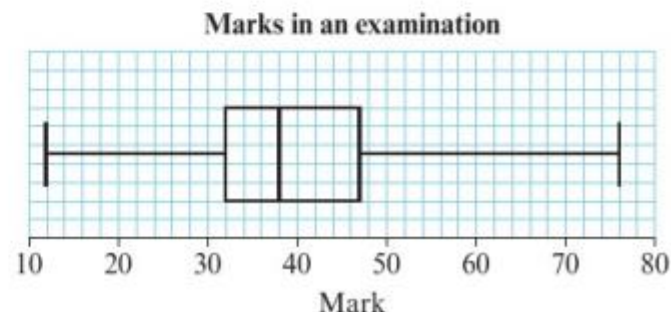
6 Here is a box plot of marks in an examination.

a Write down the upper and lower quartiles.

b Write down the median.

c Work out the interquartile range.

d Work out the range.

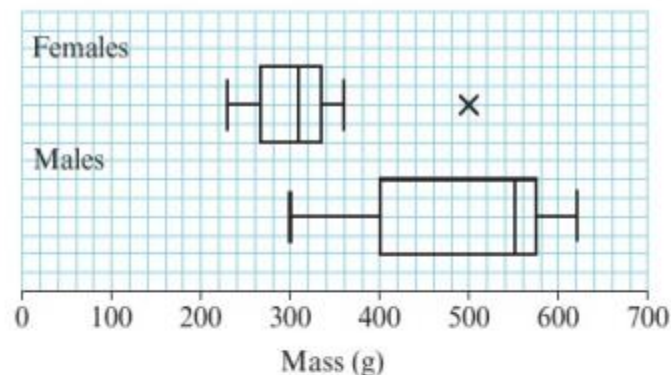




# Homework Exercise

- 7 The masses of male and female turtles are given in grams. Their masses are summarised in the box plots.

- Compare and contrast the masses of the male and female turtles.
- A turtle was found to have a mass of 330 grams. State whether it is likely to be a male or a female. Give a reason for your answer.
- Write down the size of the largest female turtle.



- 8 Data for the maximum daily gust (in knots) in Camborne in September 1987 is taken from the large data set:

13	17	19	20	21
21	22	23	24	25
25	25	26	26	26
27	29	30	30	30
33	35	38	46	78

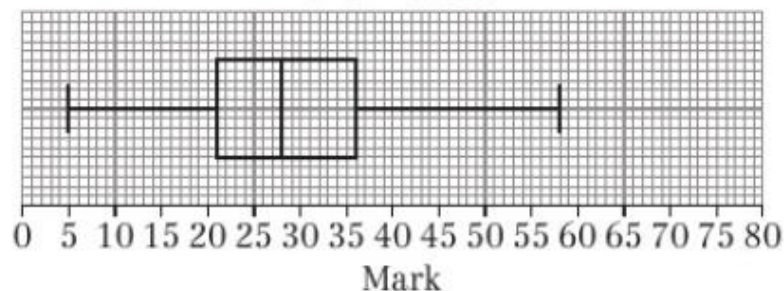
© Crown Copyright Met Office

- Calculate  $Q_1$ ,  $Q_2$  and  $Q_3$ . (3 marks)
- An outlier is defined as a value which lies either  $1.5 \times$  the interquartile range above the upper quartile or  $1.5 \times$  the interquartile range below the lower quartile.
- Show that 46 and 78 are outliers. (1 mark)
  - Draw a box plot for this data. (3 marks)

# Homework Answers

- 1 a 7 is an outlier                      b 88 is not an outlier  
c 105 is an outlier
- 2 a No outliers                          b 170 g and 440 g  
c 760 g
- 3 a 11.5 kg  
b Smallest 2.0 kg, largest 10.2 kg
- 4 a Mean 10.2, standard deviation 7.36  
b It is an outlier as it is more than 2 standard deviations above the mean.  
c e.g. It could be the age of a parent at the party.  
d Mean 7.75, standard deviation 2.44

5 Marks in a test



- 6 a 47, 32                      b 38                      c 15                      d 64

- 7 a The male turtles have a higher median mass, a greater interquartile range and a greater total range.  
b It is more likely to have been female. Very few of the male turtles had a mass this low, but more than a quarter of the female turtles had a mass of more than this.  
c 500 g
- 8 a  $Q_1 = 22$  knots,  $Q_2 = 26$  knots,  $Q_3 = 30$  knots  
b  $IQR = 8$   
 $1.5 \times IQR$  above  $Q_3 = 42$   
 $46 > 42$  and  $78 > 42$ , so 46 and 78 are outliers.
- c Maximum daily gust in Camborne, September 1987

