Chapter review 3

1 a Q_1: $\frac{31}{4} = 7.75$ so we pick the 8th value: 178

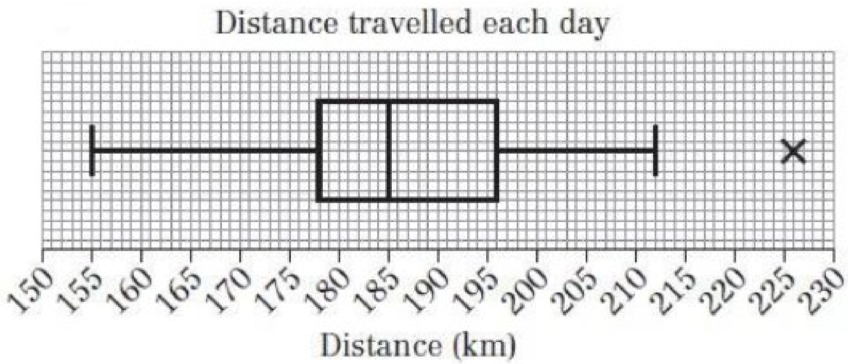
Q_2: $\frac{31+1}{2} = 16$ so we pick the 16th value: 185

Q_3: $\frac{3}{4} \times 31 = 23.25$ so we pick the 24th value: 196

b $Q_1 - 1.5(Q_3 - Q_1) = 178 - 1.5(196 - 178) = 151$
$Q_3 + 1.5(Q_3 - Q_1) = 196 + 1.5(196 - 178) = 223$

So 226 km is an outlier.

c



d The distribution is positively skewed.

2 a 45 minutes

b 60 minutes

c This is an outlier that does not fit the pattern.

d The Runners Club had the highest median, so overall they had the slowest runners.

The IQR ranges were about the same, with the Runners Club slightly more spread out.

e With the exception of the outlier, the Marathon Club runners were faster in every respect. Their minimum, Q_1, Q_2, Q_3 and maximum times were all lower than the corresponding times for the runners from the Runners Club.

2 f Advantages, any one from:

It helps us to see the spread of the data easily.
The plot is clear and easy to understand.
It uses the range and the median values.
It is easy to compare the stratified data.

Disadvantages, any one from:

Original data is not clearly shown in the box plot.
Mean and mode cannot be identified using the box plot.
It can be easily misinterpreted.
If large outliers are present, the box plot is more likely to give an incorrect representation.

3 a The median is the $\frac{n+1}{2}=\frac{35+1}{2}=18$th piece of data.
Therefore the median is 39 years.
To find the lower quartile
$\frac{n}{4}=\frac{35}{4}=8.75$
Since this is not a whole number round up, so the lower quartile is the 9th piece of data, therefore $Q_1 = 31$ years
To find the upper quartile
$\frac{3n}{4}=\frac{105}{4}=26.25$
Since this is not a whole number round up, so the upper quartile is the 27th piece of data, therefore $Q_3 = 55$ years

b $1.5(Q_3 - Q_1) = 1.5(55 - 31)$
$= 36$
$Q_3 + 36 = 55 + 36 = 91$ years, so the tortoise aged 99 years is an outlier
$Q_1 - 36 = 31 - 36 = -5$ years, so there are no outliers below Q_1

c For Zoo 1
$Q_2 - Q_1 = 64 - 44 = 20$
$Q_3 - Q_2 = 76 - 64 = 12$
$Q_2 - Q_1 > Q_3 - Q_2$ therefore the distribution of Zoo 1 is negatively skewed
For Zoo 2
$Q_2 - Q_1 = 39 - 31 = 8$
$Q_3 - Q_2 = 55 - 39 = 16$
$Q_2 - Q_1 < Q_3 - Q_2$ therefore the distribution of Zoo 2 is positively skewed

4 a The areas of the bars are proportional to the frequencies represented.

$2k(1 + 1.5 + 5.5 + 4.5) + 4k(1) = 58$

$29k = 58$ so $k = 2$

Number of girls who took longer than 56 seconds $= 2((4.5 \times 2) + (1 \times 4)) = 26$ girls

4 b Number of girls who took between 52 and 55 seconds = $2((1.5 \times 2) + (\frac{1}{2} \times 2 \times 5.5)) = 17$ girls

5 $1.5 \times 5.7 \times k = 2565$, so $k = 300$

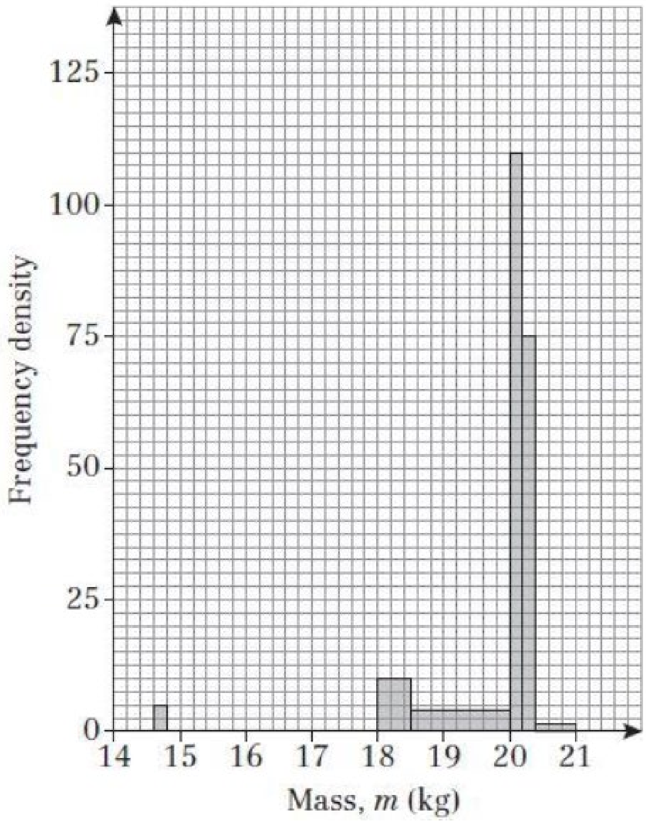
a Width = $1 \times 1.5 = 1.5$ cm

Height = $\frac{\text{frequency}}{k \times \text{width}} = \frac{1170}{300 \times 1.5} = 2.6$ cm

b Width = $5 \times 1.5 = 7.5$ cm

Height = $\frac{630}{300 \times 7.5} = 0.28$ cm

6 a



b Mean = $\frac{\Sigma fy}{n} = \frac{988.85}{50} = 19.777$ kg

Standard deviation = $\sqrt{\frac{\Sigma fy^2}{n} - u^2} = \sqrt{\frac{19\ 602.84}{50} - 19.777^2} = \sqrt{0.927} = 0.963$ (to 3 s.f.)

c Median = $20.0 + \frac{13.5}{22} \times 0.2 = 20.123$ (to 3 d.p.)

d $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}} = \frac{3(19.777 - 20.123)}{0.96285}$

$= -1.078$ (4 s.f.)

e The distribution of the weights of bags of compost is negatively skewed.

7 a $\frac{312}{14} = 22.286$ (to 3 d.p.)

b Key: 1|5 means 15 bags

0	5				
1	5	0	1	3	7
2	0	0	5		
3	0	3	1		
4	0	2			

Key: 1|5 means 15 bags

0	5				
1	0	1	3	5	7
2	0	0	5		
3	0	1	3		
4	0	2			

The median is the $\frac{n+1}{2} = \frac{14+1}{2} = 7.5$th piece of data.
Therefore the median lies halfway between the 7th and 8th pieces of data.
The median is $\frac{20+20}{2} = 20$
To find the lower quartile
$\frac{n}{4} = \frac{14}{4} = 3.5$
Since this is not a whole number round up, so the lower quartile is the 4th piece of data, therefore $Q_1 = 13$ bags
To find the upper quartile
$\frac{3n}{4} = \frac{42}{4} = 10.5$
Since this is not a whole number round up, so the upper quartile is the 11th piece of data, therefore $Q_3 = 31$ bags

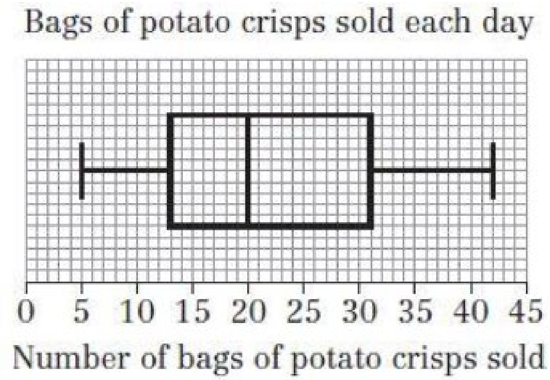
c $\text{IQR} = 31 - 13 = 18$ so $1.5 \times \text{IQR} = 27$

$13 - 27 = -14$

$31 + 27 = 58$

So there are no outliers.

7 d



e $Q_2 - Q_1 = 20 - 13 = 7$
$Q_3 - Q_2 = 31 - 20 = 11$
$Q_2 - Q_1 < Q_3 - Q_2$ therefore the distribution is positively skewed

8 a 22

b For Suha
To find the lower quartile
$$\frac{n}{4} = \frac{21}{4} = 5.25$$
Since this is not a whole number round up, so the lower quartile is the 6th piece of data, therefore $Q_1 = 11$ bicycles
To find the upper quartile
$$\frac{3n}{4} = \frac{63}{4} = 15.75$$
Since this is not a whole number round up, so the upper quartile is the 16th piece of data, therefore $Q_3 = 22$ bicycles

For Jameela
The median is the $\frac{n+1}{2} = \frac{21+1}{2} = 11$th piece of data.
Therefore the median is 27 bicycles.

So $X = 11$, $Y = 27$ and $Z = 22$

9 **a** For 1987

$$\bar{x} = \frac{\sum x}{\sum n}$$
$$= \frac{356.1}{30}$$
$$= 11.87$$
$$= 11.9°C \text{ (3 s.f.)}$$

For 2015

$$\bar{x} = \frac{\sum x}{\sum n}$$
$$= \frac{364.1}{30}$$
$$= 12.136...$$
$$= 12.1°C \text{ (3 s.f.)}$$

b For 1987

$$\sigma^2 = \frac{\sum x^2}{\sum n} - \left(\frac{\sum x}{\sum n}\right)^2$$
$$= \frac{4408.9}{30} - (11.87)^2$$
$$= 6.066...$$
$$\sigma = 2.463...$$
$$= 2.46°C \text{ (3 s.f.)}$$

The mean temperature was slightly higher in 2015 than in 1987. The standard deviation of temperatures was higher in 1987 (2.46°C) than in 2015 showing that the temperatures were more spread out in 1987.

c $\bar{x} - \sigma = 12.136... - 1.02$
$$= 11.116...$$
$$= 11.12 \text{ (to 4 s.f.)}$$

Using linear interpolation

$$\frac{11.12 - 10.1}{14.1 - 10.1} = \frac{y_1 - 0}{30 - 0}$$
$$y_1 = 7.65$$
$$\bar{x} + \sigma = 12.136... + 1.02$$
$$= 13.156...$$
$$= 13.16 \text{ (to 4 s.f.)}$$
$$\frac{13.16 - 10.1}{14.1 - 10.1} = \frac{y_2 - 0}{30 - 0}$$
$$y_2 = 22.95$$
$$y_2 - y_1 = 22.95 - 7.65$$
$$= 15.3$$
$$= 15 \text{ days (to the nearest day)}$$

Assume that the temperatures are equally distributed throughout the range.

Challenge

Length (min)	Frequency	Area of bar	Class width	Bar height
70–89	4	$4k$	20	x
90–99	17	$17k$	10	$x + 3$
100–109	20	$20k$	10	
110–139	9	$9k$	30	
140–179	2	$2k$	40	

Area of 70–89 bar $= 20x = 4k$, so $x = \dfrac{k}{5}$

Area of 90–99 bar $= 10(x + 3) = 10x + 30 = 17k$

Using substitution

$10 \times \dfrac{k}{5} = 0.4$

Area of 110–139 class $= 9k = 9 \times 2 = 18 \text{ cm}^2$

Height $= \dfrac{\text{Area}}{\text{Class width}} = \dfrac{18}{30} = 0.6 \text{ cm}$