Stats1 Chapter 3: Data Representations

Chapter Practice

Key Points

1. A common definition of an outlier is any value that is:
   - either greater than $Q_3 + k(Q_3 - Q_1)$
   - or less than $Q_1 - k(Q_3 - Q_1)$
2. The process of removing anomalies from a data set is known as cleaning the data.
3. On a histogram, to calculate the height of each bar (the **frequency density**) use the formula

   area of bar = $k \times$ frequency.
4. Joining the middle of the top of each bar in a histogram forms a frequency polygon.
5. When comparing data sets you can comment on:
   - a measure of location
   - a measure of spread

Chapter Exercises

1 Jason and Perdita decided to go on a touring holiday on the continent for the whole of July. They recorded the distance they travelled, in kilometres, each day:

155, 164, 168, 169, 173, 175, 177, 178, 178, 178, 179, 179, 179, 184, 184, 185, 185, 188, 192, 193, 194, 195, 195, 196, 204, 207, 208, 209, 211, 212, 226

a Find Q_1, Q_2 and Q_3

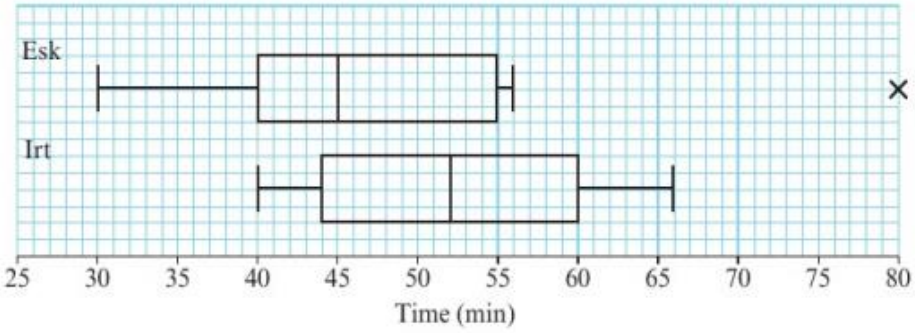
Outliers are values that lie outside $Q_1 - 1.5(Q_3 - Q_1)$ and $Q_3 + 1.5(Q_3 - Q_1)$.

b Find any outliers.

c Draw a box plot of this data.

2 Fell runners from the Esk Club and the Irt Club were keen to see which club had the faster runners overall. They decided that all the members from both clubs would take part in a fell run. The time each runner took to complete the run was recorded.

The results are summarised in the box plot.



a Write down the time by which 50% of the Esk Club runners had completed the run.

b Write down the time by which 75% of the Irt Club runners had completed the run.

c Explain what is meant by the cross (×) on the Esk Club box plot.

d Compare and contrast these two box plots.

e What conclusions can you draw from this information about which club has the faster runners?

f Give one advantage and one disadvantage of comparing distributions using box plots.

Chapter Exercises

3 The table shows the lengths, in cm, of 60 honey badgers.

a Draw a cumulative frequency diagram for this data.

b Find the median length of a honey badger.

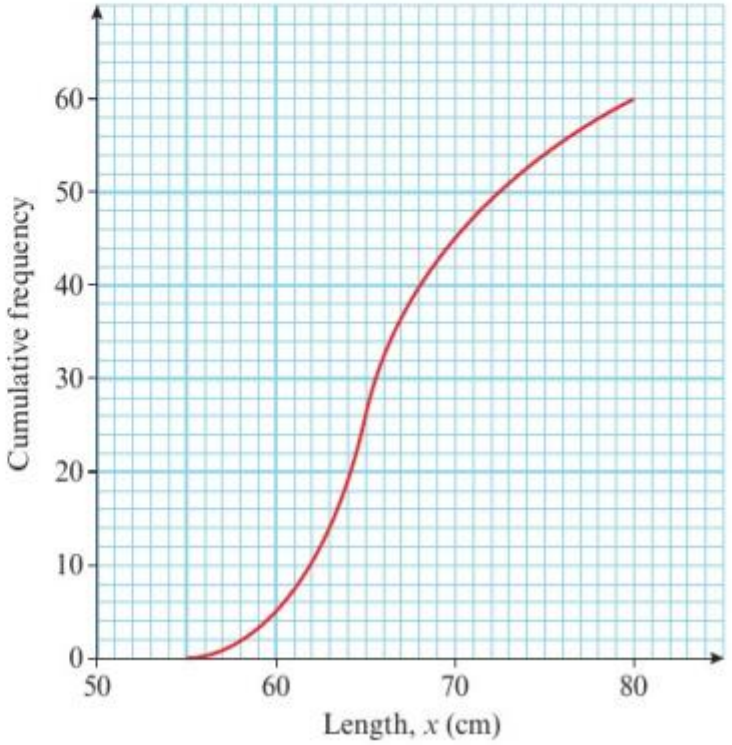
c Find the interquartile range.

Length, x (cm)	Frequency
$50 \leqslant x < 55$	2
$55 \leqslant x < 60$	7
$60 \leqslant x < 65$	15
$65 \leqslant x < 70$	31
$70 \leqslant x < 75$	5

This diagram shows the distribution of lengths of European badgers.

d Compare the distributions of lengths of honey badgers and European badgers.

e Comment on the suitability of using cumulative frequency diagrams to compare these distributions.

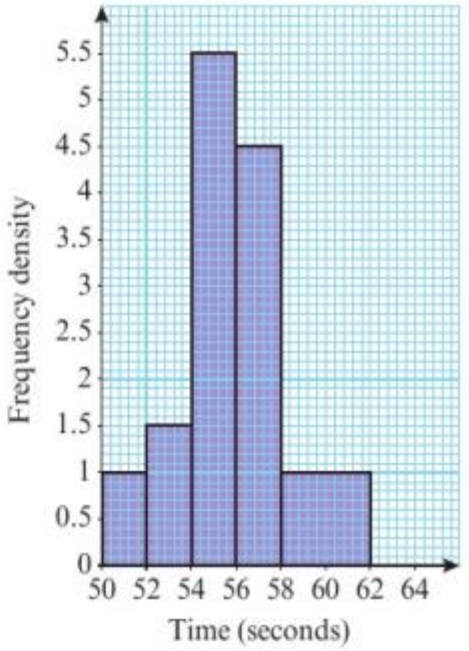


Chapter Exercises

4 The histogram shows the time taken by a group of 58 girls to run a measured distance.

a Work out the number of girls who took longer than 56 seconds.

b Estimate the number of girls who took between 52 and 55 seconds.



5 The table gives the distances travelled to school, in km, of the population of children in a particular region of the United Kingdom.

Distance, *d* (km)	$0 \leqslant d < 1$	$1 \leqslant d < 2$	$2 \leqslant d < 3$	$3 \leqslant d < 5$	$5 \leqslant d < 10$	$10 \leqslant d$
Number	2565	1784	1170	756	630	135

A histogram of this data was drawn with distance along the horizontal axis. A bar of horizontal width 1.5 cm and height 5.7 cm represented the 0–1 km group.

Find the widths and heights, in cm, to one decimal place, of the bars representing the following groups:

a $2 \leqslant d < 3$ **b** $5 \leqslant d < 10$ **(5 marks)**

Chapter Exercises

6 The labelling on bags of garden compost indicates that the bags have a mass of 20 kg.
The masses of a random sample of 50 bags are summarised in the table opposite.

a On graph paper, draw a histogram of this data.

b Estimate the mean and standard deviation of the mass of a bag of compost.

[You may use $\Sigma fy = 988.85$, $\Sigma fy^2 = 19\,602.84$]

c Using linear interpolation, estimate the median.

Mass, m (kg)	Frequency
$14.6 \leqslant m < 14.8$	1
$14.8 \leqslant m < 18.0$	0
$18.0 \leqslant m < 18.5$	5
$18.5 \leqslant m < 20.0$	6
$20.0 \leqslant m < 20.2$	22
$20.2 \leqslant m < 20.4$	15
$20.4 \leqslant m < 21.0$	1

7 The number of bags of potato crisps sold per day in a bar was recorded over a two-week period. The results are shown below.

20 15 10 30 33 40 5 11 13 20 25 42 31 17

a Calculate the mean of this data.

b Find the median and the quartiles of this data.

An outlier is an observation that falls either 1.5 × (interquartile range) above the upper quartile or 1.5 × (interquartile range) below the lower quartile.

c Determine whether or not any items of data are outliers.

d On graph paper draw a box plot to represent this data. Show your scale clearly.

Chapter Exercises

8 From the large data set, the daily maximum gust (knots) is measured at Hurn throughout May and June 2015. The data is summarised in the table.

Daily maximum gust, g (knots)	Frequency
$10 \leqslant g < 15$	3
$15 \leqslant g < 18$	9
$18 \leqslant g < 20$	9
$20 \leqslant g < 25$	20
$25 \leqslant g < 30$	9
$30 \leqslant g < 50$	7

A histogram is drawn to represent this data. The bar representing the $10 \leqslant g < 15$ class is 2.5 cm wide and 1.8 cm high.

a Give a reason to support the use of a histogram to represent this data. **(1 mark)**

b Calculate the width and height of the bar representing the $18 \leqslant g < 20$ class. **(3 marks)**

c Use your calculator to estimate the mean and standard deviation of the maximum gusts. **(3 marks)**

d Use linear interpolation to find an estimate for the number of days the maximum gust was within one standard deviation of the mean. **(4 marks)**

9 From the large data set, data was gathered in September 1987 and in September 2015 for the mean daily temperature in Leuchars. Summary statistics are given in the table.

	Min	Max	Median	Σx	Σx^2
1987	7.0	17.0	11.85	356.1	4408.9
2015	10.1	14.1	12.0	364.1	4450.2

a Calculate the mean of the mean daily temperatures in each of the two years. **(2 marks)**

b In 2015, the standard deviation was 1.02. Compare the mean daily temperatures in the two years. **(2 marks)**

c A recorded temperature is considered 'normal' for the time of year if it is within one standard deviation of the mean. Estimate for how many days in September 2015 a 'normal' mean daily temperature was recorded. State one assumption you have made in making the estimate. **(3 marks)**

Chapter Exercises

The table shows the lengths of the films in a film festival, to the nearest minute.

Length (min)	Frequency
70–89	4
90–99	17
100–109	20
110–139	9
140–179	2

A histogram is drawn to represent the data, and the bar representing the 90–99 class is 3 cm higher than the bar representing the 70–89 class.

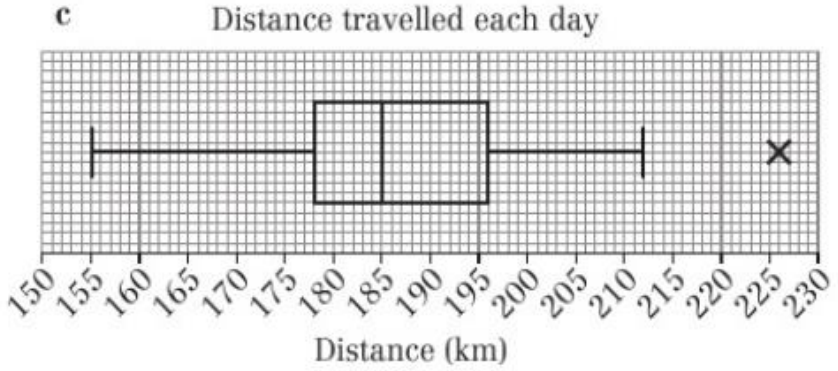
Find the height of the bar representing the 110–139 class.

Chapter Answers

1 **a** $Q_1 = 178$, $Q_2 = 185$, $Q_3 = 196$

b 226

c



2 **a** 45 minutes **b** 60 minutes

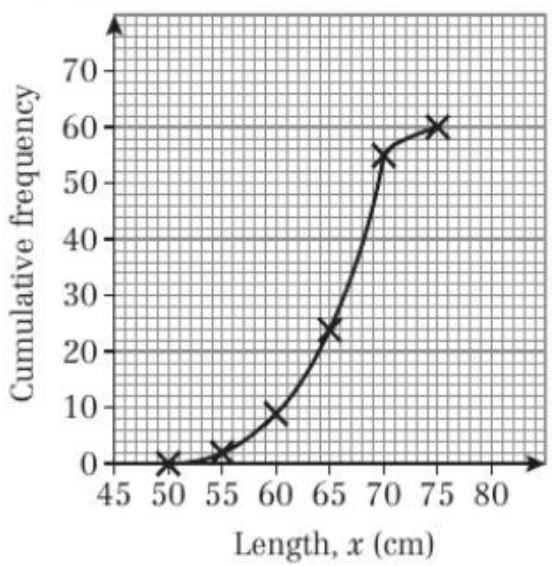
c This represents an outlier.

d Irt has a higher median than Esk.
The interquartile ranges were about the same.

e Esk had the fastest runners.

f Advantages: easy to compare quartiles, median and spread. Disadvantages: cannot compare mean or mode.

3 **a**



b ≈ 66 cm

c ≈ 6.5 cm

d The distributions have very similar medians and quartiles. Maximum length of the European badgers is greater than the maximum length of the honey badgers.

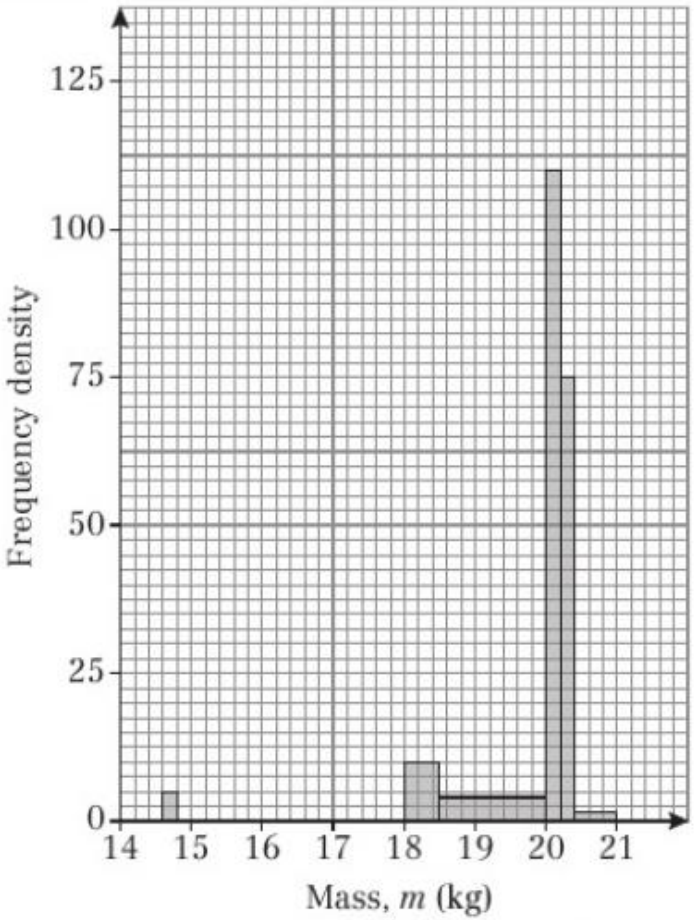
e Do not have exact data values so cannot compare the median, quartiles or range accurately.

Chapter Answers

4 **a** 26
b 17

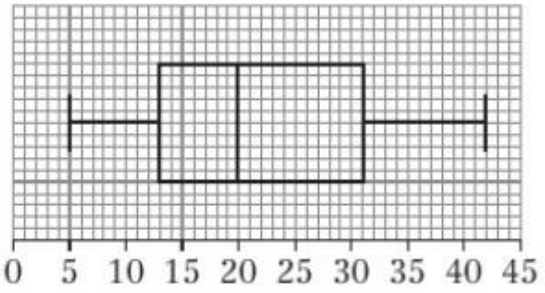
5 **a** width = 1.5 cm, height = 2.6 cm
b width = 7.5 cm, height = 0.28 cm

6 **a**



b Mean 19.8 kg, standard deviation 0.963 kg
c 20.1 kg

7 **a** 22.3
b Median 20; quartiles 13, 31
c No outliers.
d Bags of potato crisps sold each day



8 **a** The maximum gust is continuous data and the data is given in a grouped frequency table.
b 1 cm wide and 13.5 cm tall
c Mean 23.4, standard deviation 7.32
d 44 days

9 **a** 1987: 11.9 °C, 2015: 12.1 °C
b 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.
c 15 days assuming that the temperatures are equally distributed throughout the range.

Challenge:
0.6 cm