Data Collection 1A

1 a A census observes or measures every member of a population.

b An advantage is it that it will give a completely accurate result.

A disadvantage could be any one from:

It would be time consuming.
It would be expensive.

2 a The testing process will destroy the harness, so a census would destroy *all* the harnesses, meaning that there would be no harnesses left for climbers to use.

b The claim is misleading. 250 kg is the mean and median load at which the harnesses in the sample break. So we would expect half of the harnesses to break at a load of less than 250 kg.

c Test a larger number of harnesses.

3 a Any one from:

It would be time consuming.
It would be expensive.
It would be difficult to process the data.

b A list of residents.

c Each individual resident.

4 a The testing process would destroy the switches, so a census would destroy *all* the switches, meaning that there would be no switches left to sell.

b The mean is 19 615.4, less than the stated average. One of the switches survived significantly fewer operations, which suggests that the median of 22 921 might be a better average to take, as it is not affected by outliers. The data therefore supports the company's claim.

c Test a larger number of switches.

5 a All the mechanics in the garage.

b Everyone's views will be known.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Data Collection 1B

1 a There are 40 + 60 + 80 = 180 pupils altogether.

Year 1: $40 \times 0.2 = 8$ pupils
Year 2: $60 \times 0.2 = 12$ pupils
Year 3: $80 \times 0.2 = 16$ pupils

b Any one from:

A stratified sample accurately reflects the population structure of the school.
A stratified sample guarantees proportional representation of different year groups in the sample.

2 a Taking every 20th person may introduce bias, as the sampling frame is not random.

b A simple random sample using the alphabetical list as the sampling frame.

3 a No, this is not a systematic sample. Any reason from:

The first person is not selected at random.
The required elements of the sample are not being chosen at regular intervals.

b To improve the reliability of the data collected, the gym could use a larger sample.
To reduce bias, take a simple random sample using the list of members as the sampling frame.

4 a Stratified sampling

b There are 70 + 50 + 85 + 75 = 280 students altogether. All answers should be rounded to the nearest whole number as appropriate.

Year 12 Male: $\frac{70}{280} \times 40 = 10$
Year 13 Male: $\frac{50}{280} \times 40 \approx 7$
Year 12 Female: $\frac{85}{280} \times 40 \approx 12$
Year 13 Female: $\frac{75}{280} \times 40 \approx 11$

5 $k = \frac{480}{30} = 16$

Randomly select a number between 1 and 16. Start with the worker having this clock number. Then select the workers that have every 16th clock number after this.

6 a Set up a sampling frame. Use any method to select sampling units in which every member of the population has an equal chance of being selected, e.g. lottery sampling. A disadvantage of this method is that it may not reflect the proportion of members at the club who play each sport.

b The sample will have proportional representation of the members who play the different sports.

c There are 120 + 145 + 105 = 370 members altogether. All answers should be rounded to the nearest whole number as appropriate.

Cricket: $\frac{120}{370} \times 30 \approx 10$
Hockey: $\frac{145}{370} \times 30 \approx 12$
Squash: $\frac{105}{370} \times 30 \approx 9$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Data Collection 1C

1 a i Divide the population into groups according to given characteristics. The size of each group determines the proportion of the sample that should have that characteristic. The interviewer should assess which group people fall into, as part of the interview. Once a quota has been filled, no more people in that group are interviewed.

ii Only sample the people who are available at the time the study is carried out, e.g. the first 40 shoppers who are available to be interviewed.

b Quota sampling, as opportunity sampling is unlikely to reflect the characteristics of the whole population.

2 A similarity between quota sampling and stratified random sampling is that the population is divided according to the characteristics of the whole population (into strata for stratified sampling, and groups for quota sampling). A difference between quota sampling and stratified random sampling is that quota sampling is non-random.

3 a Opportunity sampling

b Opportunity sampling is unlikely to reflect the characteristics of the whole population..

c He could survey people at different times of day.
He could survey people in other parts of the town, not just outside the fish and chip shop.

4 a Mean $= \frac{4+7+6+8+2}{5} = \frac{27}{5} = 5.4$ hours

b She has used opportunity sampling, which is unlikely to reflect the characteristics of the whole population of the town, and has used a very small sample, which is unlikely to be representative.

c Increase the number of people asked.
Ask people at different times and/or in different locations.

5 a Quota sampling

b Any one from:

No sampling frame is required.
It is quick.
It is easy.
It is inexpensive.
It allows for comparison between male and female boars.

c Mean male weight $= \frac{75+80+90+85+82}{5} = \frac{412}{5} = 82$ kg (to the nearest whole number)

Mean female weight $= \frac{67+72+75+68+65}{5} = \frac{347}{5} = 69$ kg (to the nearest whole number)

Male range $= 90 - 75 = 15$
Female range $= 75 - 65 = 10$

Males are heavier on average, but have a greater spread.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

5 d Increase the sample size.
Catch deer at random times during the day.

6 a An example of an opportunity sample would be the first five heights, giving a sample of:

1.8, 1.9, 2.3, 1.7, 2.1

b The second height is 1.9. To take a sample of 5, now choose every 4th height, giving a sample of:

1.9, 2.0, 2.6, 2.3, 2.0

c Mean height for the opportunity sample $= \frac{1.8+1.9+2.3+1.7+2.1}{5} = \frac{9.8}{5} = 1.96$ m (to 2 d.p.)

Mean height for the systematic sample $= \frac{1.9+2.0+2.6+2.3+2.0}{5} = \frac{10.8}{5} = 2.16$ m (to 2 d.p.)

d The systematic sample is likely to be more reliable, because it is random and likely to be more representative. The opportunity sample might just consider all the small values, as it does here.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Data Collection 1D

1 a **Quantitative** as heights are numerical.

b **Qualitative** as colours are not numerical.

c **Quantitative** as time is numerical.

d **Quantitative** as shoe size is numerical.

e **Qualitative** as names are not numerical.

2 a **Discrete –** you cannot have a shoe size of 4.78, for example.

b **Continuous** – you can measure the length of a leaf to any degree of accuracy.

c **Discrete –** you can only have whole numbers of people.

d **Continuous –** you can measure the weight of the sugar to any degree of accuracy.

e **Continuous –** you can measure the time taken to any degree of accuracy.

f **Continuous –** you can measure the lifetime of a battery to any degree of accuracy.

3 a It is descriptive rather than numerical.

b It is discrete because you can only have whole numbers of pupils in a class. It is quantitative as it is numerical.

c Weight can take any value in a given range. Therefore, it is continuous. It is quantitative as it is numerical.

4 a 1.4 and 1.5. There are no gaps, therefore the boundaries are the given boundaries of the class.

b $\frac{1.3+1.4}{2}=\frac{2.7}{2}=1.35$ kg

c $1.3-1.2=0.1$ kg

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Data Collection 1E

1 a Leuchars

b Perth

c Any one from: Leeming, Heathrow, Beijing

d Any one from: Leuchars, Hurn, Camborne, Jacksonville, Perth

e Any one from: Beijing, Jacksonville, Perth

2 Daily maximum relative humidity is a continuous variable as it can take any percentage value between 0 and 100.

3 a i Mean daily total sunshine for Leeming $= \frac{101.4}{10} = 10.1$ hrs (to 1 d.p.)

ii Mean daily total sunshine for Heathrow $= \frac{76}{10} = 7.6$ hrs

b i Leeming range $= 14.6 - 5.1 = 9.5$ hrs

ii Heathrow range $= 14.4 - 1.6 = 12.8$ hrs

c Part (**a**) shows that Leeming has a higher average number of hours of sunshine. Leeming is further north than Heathrow, so the data does not support Supraj's conclusion. (The range calculated in part (**b**) does not affect this.)

4 The rainfall on 2/6/2015, 5/6/2015 and 8/6/2015 was recorded as 'tr', meaning trace, which is between 0 and 0.05 mm of rain. However, we treat the trace amounts as 0 in numerical calculations as anything less than 0.05 mm would be 0.0 to 1 decimal place.

$$\text{Mean daily total rainfall in Heathrow} = \frac{0.6 + 0 \times 3 + 0 \times 5 + 0.8}{10}$$
$$= \frac{1.4}{10}$$
$$= 0.14 \text{ mm}$$

5 a i Using multiple months provides more data.

ii Two days out of each month of 30 or 31 days is a small sample size.

b He is choosing the last day of one month followed by the first day of the next month, which is not random. These consecutive days are likely to be affected by the same weather pattern.

c He could number the days and choose a simple random sample.

6 a Perth is in the Southern hemisphere, where it will be winter in August, and Jacksonville is in the Northern hemisphere, where it will be summer. So it is likely to be hotter in Jacksonville.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 b The lowest temperatures in the UK are at coastal locations – Leuchars and Camborne (14.7 °C and 15.4 °C). The highest temperature is at an inland location – Beijing (26.6 °C). There is some evidence to support this conclusion.

7 1 okta corresponds to 1 eighth of the sky covered by cloud, and so the maximum figure for cloud cover is 8. Brian's answer is more than that and so is incorrect.

8 a Marie should select days at regular intervals from an ordered list. She should put the days into date order. Then she should begin at a random day between 1 and 6 and select every 6th day thereafter until 30 days are selected since $184 \div 30 \approx 6$.

b There may be missing values because the data is not available.

Large data set

1 a Mean pressure $= \frac{31628}{31} = 1020$ hPa (to the nearest whole number)

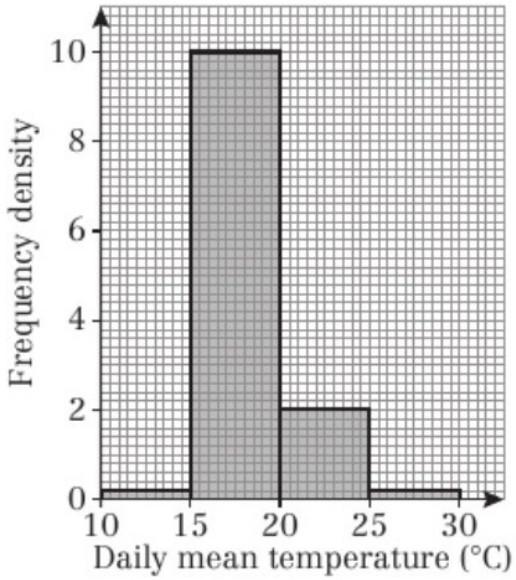
b When the rainfall for July 2015 is in numerical order, the first seventeen days had zero rainfall, so the median daily rainfall is 0.0 mm.

c i

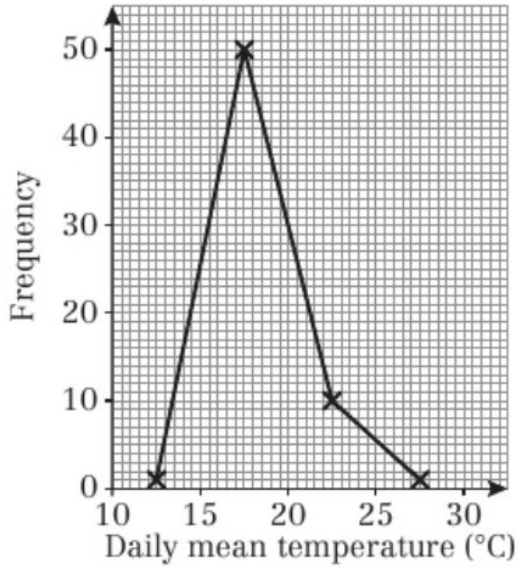
Temperature, t (°C)	**Frequency**
$10 \leq t < 15$	1
$15 \leq t < 20$	50
$20 \leq t < 25$	10
$25 \leq t < 30$	1

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

1 c ii



iii



2 a i Give each day a number from 1 to 184. Use a calculator to generate 10 random numbers, for example:

43, 4, 102, 23, 139, 106, 164, 94, 163, 63

This gives the sample of wind speeds: 4, n/a, 4, 12, 3, 8, 4, 8, 5 and 3.

ii Mean daily wind speed excluding the value n/a is $\frac{51}{9}$ = 6 kn (to the nearest whole number)

b i The first ten available values are 5, 8, 7, 11, 9, 7, 10, 7, 5 and 5.

ii Mean daily mean wind speed = $\frac{74}{10}$ = 7 kn (to the nearest whole number)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

c The simple random sample is likely to be more representative as the opportunity sample records consecutive days, which are likely to affect one another.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

2 d Any two of:

Take a bigger sample, as 10 sample units is likely to be insufficient.
Take a stratified sample, categorising by month.
Take a systematic sample.

Both of the final options guarantee that you will take data spread over a range of months.

e Example: A stratified sample of 30 days.

Month	May	Jun	Jul	Aug	Sep	Oct
Number of days	31	30	31	31	30	31

$\frac{31}{184} \times 30 = 5.05 \approx 5$ days
$\frac{30}{184} \times 30 = 4.89 \approx 5$ days

Therefore, take a simple random sample of 5 days from each month for both Leeming and Leuchars in 1987.

In the sample taken for this example, 4 of the values for Leeming and Leuchars were not available and so the calculations will be evaluated from the remaining 26 sample units.

Mean windspeed for Leeming $= \frac{166}{26} = 6$ kn (to the nearest whole number)
Mean windspeed for Leuchars $= \frac{183}{26} = 7$ kn (to the nearest whole number)

Leuchars is a coastal location and has a higher average wind speed than the inland location of Leeming so the calculation supports the statement 'Coastal locations are likely to have higher average windspeeds than inland locations.'

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Data Collection, Mixed exercise 1

1 a Mean daily maximum temperature $= \frac{14.6 + 8.8 + 7.2 + 7.3 + 10.2}{5} = \frac{48}{5} = 9.6°C$

b The sampling frame is the first 15 days in May 1987. Allocate each date a number from 1 to 15. Use the random number function on a calculator to generate 5 random numbers between 1 and 15 and choose the dates with these numbers.

c For example: The 5 random numbers generated are 12, 8, 13, 6 and 15. These give the temperatures 12.7, 12.1, 8.9, 11.9 and 9.5.

Mean daily max temperature $= \frac{55.1}{5} = 11.0°C$ (to 1 d.p.)

d Mean daily max temperature $= \frac{162.2}{15} = 10.8°C$ (to 1 d.p.)

Using data for the whole period gives a completely accurate result. A sample of one third of the dates will give a less reliable one. This variation is apparent in the answers to a and c, above.

2 a i Advantage: very accurate.
Disadvantages: expensive, time consuming and difficult to process.

ii Advantages: easier data collection, quick and cheap.
Disadvantages: less accurate, less representative and possibly biased.

b Assign unique 3-digit identifiers 000, 001, …, 499 to each member of the population. Use random number tables, a computer or a calculator to generate 3-digit numbers. If these correspond to an identifier then include the corresponding member in the sample, being careful to ignore repeats and numbers greater than 499. Repeat this process until the sample contains 100 members.

3 a i Collection of individual items.

ii List of sampling units, with each unit given an identifying name or number.

b i List of registered owners from the DVLA.

ii List of people visiting a doctor's clinic in Oxford in July 1996.

4 a Advantages:

The results are the most representative of the population since the structure of the sample reflects the structure of the population.
It guarantees proportional representation of groups within a population.

Disadvantages:

You need to know the structure of the population before you can take a stratified sample.
Classification into mutually exclusive strata may be difficult to implement.
The sampling within each strata may suffer from the disadvantages of simple random sampling.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

4 **b** Advantages:

Quick
Cheap.
All units have an equal chance of selection.

Disadvantages:

Can introduce bias (e.g. if the sample, by chance, only includes very tall people in an investigation into heights of students).
A sampling frame is needed first.

5 a People on the shop floor are not represented.

b i Get a list of the 300 workers at the factory. $\frac{300}{30} = 10$ so choose one of the first ten workers on the list at random and every subsequent 10th worker on the list, e.g. if person 7 is chosen, then the sample includes workers 7, 17, 27, …, 287 and 297.

ii The sample should contain $\frac{1}{3} \times 30 = 10$ office workers and $\frac{2}{3} \times 30 = 20$ shop floor workers, as those are the corresponding proportions in the whole population. The 10 office workers in the sample should be selected by a simple random sample of the 100 office workers. The 20 shop floor workers should be selected by a simple random sample of the 200 shop floor workers.

iii Decide the categories e.g. age, gender, office/non-office and set a quota for each in proportion to their numbers in the population. Interview workers until quotas are full.

6 a Allocate a number between 1 and 120 (the total number of pupils) to each pupil. Use random number tables, a computer or a calculator to select 15 different whole numbers between 1 and 120.

Pupils corresponding to these numbers become the sample.

b Allocate numbers 1–64 to girls and 65–120 to boys.

Select $\frac{64}{120} \times 15 = 8$ different random numbers between 1 and 64 for girls.

Select the remaining 7 sampling units using random numbers between 65 and 120 for boys.

Include the corresponding boys and girls in the sample.

7 a Stratified sampling.

b This method uses naturally occurring groupings (strata). The results are more likely to represent the views of the whole population since the sample reflects its structure.

8 a Opportunity sampling

b Any one of:

It is easy to carry out.
It is inexpensive.

c The data is continuous, as weight can take any positive value.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 d Mean weight $= \frac{70+76+82+74+78}{5} = \frac{380}{5} = 76$ kg

e Mean weight $= \frac{79+86+90+68+75}{5} = \frac{398}{5} = 80$ kg (to the nearest whole number)

f The second conservationist is likely to have a more reliable estimate as opportunity sampling is unlikely to provide a representative sample of the whole population, as it does not necessarily reflect its structure.

g The second conservationist could select more springboks at each location.

9 a This sample is not entirely random as the dates are selected at regular intervals. It is actually a systematic sample.

b A systematic sample: select the first date at random and then the same date each month. An advantage of a systematic sample is that each month is covered. A disadvantage of a systematic sample is that there may be patterns, and therefore bias, in the sample data.

A simple random sample: select the six days completely at random. An advantage of a simple random sample is that it avoids the likelihood of patterns and unintentional bias. A disadvantage of a simple random sample is that it may not cover the full range of months.

c The data is continuous as temperature can take any value.

d Mean daily maximum temperature $= \frac{59.1}{6} = 9.9$ °C (to 1 d.p.)

e Any suitable reason e.g. this estimate is unlikely to be reliable as it does not include the winter months / the data is very variable.

Large data set

a $\frac{184}{18} \approx 10$, so take every 10th day, choosing a starting place by getting a calculator to generate a random number, for example, 10. The sample is then:

79, 99, 99, 86, 95, 99, 93, 99, 100, 93, 99, 98, 98, 100, 89, 99, 95, 95

b Any of: easy, quick, more suitable for this relatively large population..

c Mean $= \frac{1715}{18} = 95\%$ (to the nearest whole number)

d The sampling frame is not random (it is in date order) so systematic sampling could introduce bias. This bias could be counteracted by using simple random sampling.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread 2A

1 a 700 g, as this is the most often occurring.

b 500 + 700 + 400 + 300 + 900 + 700 + 700 = 4200

$$\frac{4200}{7} = 600\,\text{g}$$

c 300 400 500 **700** 700 700 900

700 g is the median (the middle value).

d It will increase the mean, as 650 > 600 (the old mean).

The mode will be unchanged.

It will decrease the median. There will now be an *even* number of values, so we take the middle *pair*: 650 and 700. The new median will be half-way between these: 675.

2 a $\frac{256.2}{6} = 42.7$

b It will increase the mean, as the new piece of data (52) is greater than the old mean (42.7).

3 a The mean visibility for May = $\frac{\Sigma h}{n} = \frac{724000}{31} = 23354.8$ (to 1 d.p.)

The mean visibility for June = $\frac{\Sigma h}{n} = \frac{632000}{30} = 21066.7$ (to 1 d.p.)

b The mean visibility for May and June = $\frac{\Sigma h}{n} = \frac{724000 + 632000}{31 + 30} = \frac{1356000}{61} = 22229.5$ (to 1 d.p.)

4 a 8 minutes. Everything else occurs only once, but there are two 8's.

b $\frac{102}{10} = 10.2$ minutes

c 5 6 7 8 **8 9** 10 11 12 26

The median is 8.5 minutes.

d The median would be reasonable. The mean is affected by the extreme value of 26.

In this case the mode is close to the median, so would be acceptable; but this would not always be the case.

5 a 2 breakdowns

b The median is the 18.5th value = 1

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 c $(8 \times 0) + (11 \times 1) + (12 \times 2) + (3 \times 3) + (1 \times 4) + (1 \times 5) = 53$

The mean $= \frac{53}{36} = 1.47$ breakdowns

d The median, since this is the lowest value

6 $(5 \times 8) + (6 \times 57) + (7 \times 29) + (8 \times 3) + (9 \times 1) = 618$ petals

$8 + 57 + 29 + 3 + 1 = 98$ celandines

The mean $= \frac{618}{98} = 6.31$ petals (to 2 d.p.)

7 The mean $= \frac{1 \times 7 + 2 \times p + 3 \times 2}{7 + p + 2}$

$$1.5 = \frac{7 + 2p + 6}{p + 9}$$

$$= \frac{2p + 13}{p + 9}$$

$1.5p + 13.5 = 2p + 13$

$0.5 = 0.5p$

$p = 1$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread 2B

1 a $351 - 400$

b $$\frac{(200\times4)+(263\times8)+(325.5\times18)+(375.5\times28)+(450.5\times7)}{65}$$
$$=\frac{800+2104+5859+10514+3153.5}{65}$$
$$=\frac{22430.5}{65}$$
$$=345.08$$

c There are 65 observations so the median is the 33rd. The 33rd observation will lie in the class 351–400.

2 a $$\frac{(67\times1)+(72\times4)+(77\times6)+(82\times6)+(87\times8)+(92\times4)+(97\times1)}{30}=\frac{2470}{30}=82.3 \text{ decibels}$$

b The answer is an estimate because you don't know the exact data values.

3 a Modal class = $10 \leq t < 12$

b Estimate of the mean $= \frac{(7\times3)+(9\times7)+(11\times9)+(13\times7)+(15\times3)+(17\times2)}{31}$
$$=\frac{353}{31}=11.4 \text{ °C (1 d.p.)}$$

4 Store A $$\frac{(20.5\times5)+(30.5\times16)+(40.5\times14)+(50.5\times22)+(60.5\times26)+(70.5\times14)}{97}$$
$$=\frac{4828.5}{97}=50 \text{ years}$$

Store B $$\frac{(20.5\times4)+(30.5\times12)+(40.5\times10)+(50.5\times28)+(60.5\times25)+(70.5\times13)}{92}$$
$$=\frac{4696}{92}=51 \text{ years}$$

Store B employs older workers but not by a great margin.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread 2C

1 a $Q_2 = \frac{16+1}{2}$ th value = 8.5th value

1009, 1013, 1014, 1017, 1017, 1017, 1018, 1019, 1021, 1022, 1024, 1024, 1025, 1027, 1029, 1031

$Q_2 = \frac{1019+1021}{2} = 1020$ hPa

b Q_1 = 4.5th value
Q_1 = 1017 hPa

Q_3 = 12.5th value
Q_3 = 1024.5 hPa

2 $Q_2 = \frac{95+1}{2}$ th value = 48th value
Q_2 = 37

Q_1 = 24th value
Q_1 = 37

Q_3 = 72nd value
Q_3 = 38

3 In this case, the number of breakdowns is a discrete variable which in this situation has been treated as a continuous variable. While not ideal, it is the best you can do.

Median value is the 13th value. This is in the first class.
Let m be the median.

$\frac{m-0}{1.5-0} = \frac{13-0}{18-0}$ so $m \approx 1.08$ (3 sf)

4 a Median: $\frac{31}{2}$ = 15.5th value
Using interpolation:

$\frac{Q_2 - 399.5}{449.5-399.5} = \frac{15.5-9}{19-9}$

Q_2 = 432 kg

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

4 b $Q_1 : \frac{31}{4} = 7.75$th value, so Q_1 is in class $350 - 399$

$$\frac{Q_1 - 349.5}{399.5 - 349.5} = \frac{7.75 - 3}{9 - 3}$$

$$\frac{Q_1 - 349.5}{50} = \frac{4.75}{6}$$

$$Q_1 = 39.58 + 349.5 = 389$$

c $Q_3 : 3 \times \frac{31}{4} = 23.25$th value, so Q_3 is in class $450 - 499$

$$\frac{Q_3 - 449.5}{499.5 - 449.5} = \frac{23.25 - 19}{26 - 19}$$

$$\frac{Q_3 - 449.5}{50} = \frac{4.25}{7}$$

$$Q_3 = 30.36 + 449.5 = 480$$

d Three-quarters of the cows weigh less than 480 kg.

5 a Estimate for the mean $= \frac{(25 \times 6) + (35 \times 10) + (45 \times 18) + (55 \times 13) + (65 \times 2)}{49}$

$$= \frac{2155}{49}$$

$$= 44.0 \text{ minutes } (\text{to 3 s.f.})$$

b $65\text{th} : \frac{65}{100} \times 49 = 31.85$

$$\frac{P_{65} - 40}{50 - 40} = \frac{31.85 - 16}{34 - 16}$$

$P_{65} = 48.8$

c $$\frac{P_{90} - 50}{60 - 50} = \frac{44.1 - 34}{47 - 34}$$

$P_{90} = 57.8$

90th percentile = 57.8 minutes, so more than 10% of customers have to wait longer then 57.8 minutes – not 56 minutes as stated by the firm.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

6 a $\dfrac{P_{80}-2.5}{3.0-2.5}=\dfrac{80-61}{89-61}$

$P_{80} = 2.84$ (to 2 d.p.)

80th percentile = 2.84 m, so 80% of condors have a wingspan of less than 2.84 m.

b The 90th percentile is in the $3.0 \leq w$ class. There is no upper boundary for this class, so it is not possible to estimate the 90th percentile.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread 2D

1 a CF = 4 8 10 17 37 61 71

71 slow worms were measured.

b $Q_1: \frac{71}{4} = 17.75$th value, so Q_1 is in class $185 - 199$

$$\frac{Q_1 - 184.5}{199.5 - 184.5} = \frac{17.75 - 17}{37 - 17}$$

$$Q_1 - 184.5 = 0.5625$$

$$Q_1 = 185.0625$$

$$Q_3 : 3 \times \frac{71}{4} = 53.25\text{th value}$$

so Q_3 is in class $200 - 214$

$$\frac{Q_3 - 199.5}{214.5 - 199.5} = \frac{53.25 - 37}{61 - 37}$$

$$Q_3 - 199.5 = \frac{243.75}{24}$$

$$Q_3 = 209.656$$

$$\text{IQR} = 209.656 - 185.0625$$
$$= 24.6 \text{ (3 s.f.)}$$

c $$\bar{x} = \frac{(132 \times 4) + (147 \times 4) + (162 \times 2) + (177 \times 7) + (192 \times 20) + (207 \times 24) + (222 \times 10)}{71}$$

$$= \frac{13707}{71}$$
$$= 193.1 \text{mm (to 1 d.p.)}$$

d $$\bar{x} + \text{IQR} = 193.1 + 24.6$$
$$= 217.7$$

217.7 is in the class interval 215–229

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

1 d Using interpolation:

$$\frac{217.7-214.5}{229.5-214.5}=\frac{y-61}{71-61}$$

$$y = 63.13...$$

$$71 - y = 7.87$$

8 slow worms have that length.

2 a 34th: $\frac{34}{100}\times 70 = 23.8$

$$\frac{P_{34}-1000}{1100-1000}=\frac{23.8-3}{27-3}$$

$$P_{34} = 1086.7$$

66th: $\frac{66}{100}\times 70 = 46.2$

$$\frac{P_{66}-1100}{1200-1100}=\frac{46.2-27}{55-27}$$

$$P_{66} = 1168.6$$

34% to 66% interpercentile range $= P_{66} - P_{34} = 1168.6 - 1086.7 = £81.90$

b $46.2 - 23.8 = 22.4$
So 22 data values

3 a 5th: $\frac{5}{100}\times 60 = 3$

$$\frac{P_5-14.5}{16.5-14.5}=\frac{3-0}{5-0}$$

$$P_5 = 15.7$$

95th: $\frac{95}{100}\times 60 = 57$

$$\frac{P_{95}-20.5}{22.5-20.5}=\frac{57-50}{60-50}$$

$$P_{95} = 21.9$$

5% to 95% interpercentile range $= 21.9 - 15.7 = 6.2$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 b $57 - 3 = 54$
So 54 data values

4 a 9.4, 10.3, 10.3, 10.6, 10.9, 12.1, 12.4, 12.7, 13.2, 14.3

Q_2 = 5.5th value = $\frac{10.9+12.1}{2} = 11.5$
Q_1 = 3rd value = 10.3
Q_3 = 8th value = 12.7

IQR = $12.7 - 10.3 = 2.4$

b On average, the temperature was higher in June than in May as the median is higher. However, the temperature was more variable in May than in June, as the IQR is higher.

c 10th: $\frac{10}{100} \times 31 = 3.1$
90th: $\frac{90}{100} \times 31 = 27.9$

$27.9 - 3.1 = 24.8$

So 24 days

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread 2E

1 a Mean $= \frac{24}{8} = 3$

b Variance $= \frac{78}{8} - 3^2 = 0.75$

c Standard deviation $= \sqrt{0.75} = 0.866$

2 Standard deviation $= \sqrt{\frac{5905}{10} - \left(\frac{241}{10}\right)^2} = 3.11$ kg

3 a $\sum h = 165 + 170 + 190 + 180 + 175 + 185 + 176 + 184 = 1425$

Mean $= \frac{1425}{8} = 178.125 \approx 178$

b Variance $= \frac{254307}{8} - 178.125^2 = 59.9$

c Standard deviation $= \sqrt{59.9} = 7.74$

4 $\sum x = 50 + 86 = 136$

$\sum x^2 = 310 + 568 = 878$

Mean $= \frac{136}{25} = 5.44$

Standard deviation $= \sqrt{\frac{878}{25} - \left(\frac{136}{25}\right)^2} = 2.35$

5 a Mean $= \frac{869}{85} = 10.22$

Standard deviation $= \sqrt{\frac{9039}{85} - \left(\frac{869}{85}\right)^2} = 1.35$

b 10.22 + 1.35 = £11.57

$\frac{11.57 - 11.50}{12.50 - 11.50} = \frac{s - 65}{85 - 65}$

$s = 66.4$

85 − 66.4 = 18.6
So 19 students

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 Standard deviation $= \sqrt{\frac{203}{54} - \left(\frac{81}{54}\right)^2} = 1.23$

7 Mean $= \frac{805}{50} = 16.1$ hours

Standard deviation $= \sqrt{\frac{14062.5}{50} - \left(\frac{805}{50}\right)^2} = 4.69$ hours

One standard deviation below mean $= 16.1 - 4.69 = 11.41$ hours.

$$\frac{11.41-10}{15-10} = \frac{p-5}{19-5}$$

$p = 8.948$

$50 - p = 41.052$

41 parts tested (82%) lasted longer than one standard deviation below the mean.

According to the manufacturers, this should be 45 parts (90%), so the claim is false.

8 a Mean $= \frac{243}{30} = 8.1$ kn

Standard deviation $= \sqrt{\frac{2317}{30} - \left(\frac{243}{30}\right)^2} = 3.41$ kn

b $8.1 + 3.41 = 11.51$ kn

$$\frac{11.51-4}{17-4} = \frac{d-0}{30-0}$$

$d = 17.33$

$30 - d = 12.67$

So 12 days

c The windspeeds are equally distributed throughout the range.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread 2F

1 a $11 + 9 + 5 + 8 + 3 + 7 + 6 = 49$

b Mean $= \frac{49}{7} = 7$

c $7 = \frac{\bar{x}}{10}$ so $\bar{x} = 70$

2 a $7 + 10 + 4 + 10 + 5 + 11 + 2 + 3 = 52$

b Mean $= \frac{52}{8} = 6.5$

c $6.5 = \frac{\bar{x} - 3}{7}$ so $\bar{x} = 48.5$

3 $(1.5 \times 200) + 65 = 365$

4 Standard deviation = 2.34

5 a Mean $= \frac{(1\times 3) + (1.1\times 12) + (1.2\times 40) + (1.3\times 10) + (1.4\times 5)}{70}$ $\frac{84.2}{70}$

$= \frac{84.2}{70}$

$= 1.2$ hours

b $\frac{84.2}{70} = \frac{\bar{x} - 1}{20}$ so $\bar{x} = 25.1$ hours

c Standard deviation of coded data $= \sqrt{\frac{101.82}{70} - \left(\frac{84.2}{70}\right)^2}$

$= 0.0877845...$

Standard deviation = 20 × 0.0877845... = 1.76 hours

6 Standard deviation of coded data $= \sqrt{\frac{176.84}{100} - \left(\frac{131}{100}\right)^2} = 0.229$

Standard deviation $= 0.229 \times 100 = 22.9$

7 Standard deviation of coded data = $= \sqrt{\frac{147.03}{6} - \left(\frac{16.1}{6}\right)^2} = 4.16$

Standard deviation $= \frac{4.16}{0.01} = 416$

8 a $t = 0.8(m + 12)$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 b Mean of the standardised marks = $\bar{t} = 52.8$

$\bar{t} = 0.8(\bar{m} + 12)$

$\bar{m} = \frac{52.8}{0.8} - 12$

Mean of the original marks = 54

Standard deviation of the standardised marks = $\sqrt{\frac{S_{tt}}{n}} = \sqrt{\frac{7.3}{28}} = 0.5106...$

Standard deviation of the original marks = $\frac{0.5106...}{0.8} = 0.64$

9 Coded mean = 10.15

Mean of the daily mean pressure = 2(10.15 + 500) = 1020.3 hPa

Coded standard deviation = $\sqrt{\frac{S_{cc}}{n}} = \sqrt{\frac{296.4}{30}} = 3.1432...$

Standard deviation of the daily mean pressure = 2 × 3.1432... = 6.28 hPa

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Measures of location and spread, Mixed Exercise 2

1 $(8\times 65)+(12\times 72)=1384$

$\frac{1384}{20}=69.2$ marks

2 a 10, 12, 9, 2, 2.5, 9.5

b coded mean $=\frac{45}{6}=7.5$

c mean $= 7.5 \times 80 + 7 = 607$

3 $18\times 1000+720=£18720$

4 a Group *A*:

$(1 \times 24.5) + (3 \times 34.5) + (6 \times 44.5) + (6 \times 54.5) + (11 \times 64.5) + (10 \times 74.5) + (8 \times 84.5)$

Mean $=\frac{2852.5}{45}=63.39$ marks

Group *B*:

$(1 \times 24.5) + (2 \times 34.5) + (4 \times 44.5) + (13 \times 54.5) + (15 \times 64.5) + (6 \times 74.5) + (3 \times 84.5)$

Mean $=\frac{2648}{44}=60.18$ marks

b The method used to teach group *A* is best as the mean mark is higher.

5 a Modal Class is $21-25$ hours

b We need the 40th value. This is in the $21-25$ class.

$\frac{m-20.5}{25.5-20.5}=\frac{40-30}{75-30}$

$45m=50+922.5=972.5$

median = 21.6 hours

c mean $= 20.6$ hours

d $12\times 22.3=267.6$ hours

Total hours for all 92 batteries is $267.6+1645=1912.6$
Mean life for 92 batteries is 20.8 hours.

6 CF = 5 15 30 42 50

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 Data is continuous, so:

$Q_1 : \frac{50}{4} = 12.5$th value, so Q_1 is in class $21 - 40$

$$\frac{Q_1 - 20.5}{40.5 - 20.5} = \frac{12.5 - 5}{15 - 5}$$

$$\frac{Q_1 - 20.5}{20} = \frac{7.5}{10}$$

$$Q_1 = 35.5$$

$Q_3 : \frac{3 \times 50}{4} = 37.5$th value, so Q_3 is in class $61 - 80$

$$\frac{Q_3 - 60.5}{80.5 - 60.5} = \frac{37.5 - 30}{42 - 30}$$

$$Q_3 = 73$$

$$\text{IQR} = 73 - 35.5 = 37.5$$

7 a $30\text{th} : \frac{30}{100} \times 100 = 30$

$$P_{30} = 20.5$$

b $70\text{th} : \frac{70}{100} \times 100 = 70$

$$\frac{P_{70} - 30.5}{40.5 - 30.5} = \frac{70 - 60}{84 - 60}$$

$$P_{66} = 34.7$$

c 30% to 70% interpercentile $= 34.7 - 20.5 = 14.2$

8 a $Q_1 : \frac{80}{4} = 20$th value, so Q_1 is in class $40 - 49$

$$\frac{Q_1 - 39.5}{49.5 - 39.5} = \frac{20 - 15}{51 - 15}$$

$$\frac{Q_1 - 39.5}{10} = \frac{5}{36}$$

$$Q_1 = 40.9$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 a Q_3: $\frac{3\times 80}{4} = 60$th value, so Q_3 is in class $50-59$

$$\frac{Q_3 - 49.5}{59.5 - 49.5} = \frac{60-51}{71-51}$$

$$Q_3 = 54$$

$$\text{IQR} = 54 - 40.9 = 13.1$$

b $\text{Variance} = \frac{183040}{80} - \left(\frac{3740}{80}\right)^2 = 102.4375 = 102$

$\text{Standard deviation} = \sqrt{102.4375} = 10.1$

9 CF = 5 15 41 49 50

a Data is continuous, so:

Q_1: $\frac{50}{4} = 12.5$th value, so Q_1 is in class $95-100$

$$\frac{Q_1 - 95}{100-95} = \frac{12.5-5}{15-5}$$

$$Q_1 = 98.75$$

b Q_3: $3\times\frac{50}{4} = 37.5$th value, so Q_3 is in class $100-105$

$$\frac{Q_3 - 100}{105-100} = \frac{37.5-15}{41-15}$$

$$Q_3 = 104.33$$

c $\text{IQR} = 104.33 - 98.75 = 5.58$

d $\text{Standard deviation} = \sqrt{\frac{516112.5}{50} - \left(\frac{5075}{50}\right)^2} = 4.47$

10 a $\text{Mean} = \frac{(12\times 12)+(14\times 14)+(16\times 4)}{30} = \frac{202}{15} = 13.5$ (1 d.p)

$\text{Standard deviation} = \sqrt{\frac{916}{5} - \left(\frac{202}{15}\right)^2} = 1.36$ (3 s.f.)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

10 b $10\text{th}: \frac{10}{100} \times 30 = 3$

$$\frac{P_{10} - 11}{13 - 11} = \frac{3 - 0}{12 - 0}$$

$$P_{10} = 11.5$$

$90\text{th}: \frac{90}{100} \times 30 = 27$

$$\frac{P_{90} - 15}{17 - 15} = \frac{27 - 26}{30 - 26}$$

$$P_{90} = 15.5$$

10% to 90% interpercentile range $= 15.5 - 11.5 = 4°C$

c $13.5 + 1.36 = 14.86$

Using interpolation:

$$\frac{14.86 - 13}{15 - 13} = \frac{d - 12}{26 - 12}$$

$d = 25.02$

$30 - d = 4.98$

So 5 days

11 a Coded mean $= \frac{106}{31} = 3.419...$

Coded standard deviation $= \sqrt{\frac{80.55}{31}} = 1.6119...$

b Mean $= 3.419... \times 2 + 3 = 9.84$ kn

Standard deviation $= 1.6119 \times 2 = 3.22$ kn

12 a Mean $= \frac{316}{20} = 15.8$

Standard deviation $= \sqrt{\frac{5078}{20} - \left(\frac{316}{20}\right)^2} = 2.06$

b It will decrease the mean wing span since $13 < 15.8$

c Coded mean $= \frac{104}{20} = 5.2$

Mean $= 5.2 \times 10 + 5 = 57$ cm

Coded standard deviation $= \sqrt{\frac{1.8}{20}} = 0.3$

Standard deviation $= 0.3 \times 10 = 3$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Challenge

Total = $3.1 \times 20 = 62$
New total = $62 - 2.3 + 3.2 = 62.9$
New mean = $\frac{62.9}{20} = 3.145$

$\sigma = 1.4$, $\sigma^2 = 1.96$

$1.96 = \frac{\sum x^2}{20} - \left(\frac{62}{20}\right)^2$
$\sum x^2 = 231.4$
New $\sum x^2 = 231.4 - 2.3^2 + 3.2^2 = 236.35$
New standard deviation = $\sqrt{\frac{236.35}{20} - \left(\frac{62.9}{20}\right)^2} = 1.39$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Representations of data 3A

1 IQR = 68−46 = 22

$46 - 1.5 \times 22 = 13$
$68 + 1.5 \times 22 = 101$

a 7 is an outlier as 7 < 13.

b 88 is not an outlier as 13 < 88 < 101.

c 105 is an outlier as 105 > 101.

2 a Outliers are < 400 – 180 = 220 or > 580 + 180 = 760. So there are no outliers.

b Outliers are < 260 – 80 = 180 or > 340 + 80 = 420. So 170 g and 440 g are both outliers.

c 760 g

3 a Mean = 6.1 kg
Standard deviation = $\sqrt{4.2}$
Mean − 2 × standard deviation = $6.1 - 2 \times \sqrt{4.2} = 2.00$ (to 3 s.f.)
Mean + 2 × standard deviation = $6.1 + 2 \times \sqrt{4.2} = 10.2$ (to 3 s.f.)
So 11.5 kg is an outlier.

b The smallest is 2.00 kg.
The largest is 10.2 kg.

4 a $\text{Mean} = \frac{\sum x}{n} = \frac{92}{9} = 10.2$ (to 3 s.f.)

$$\text{Standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$
$$= \sqrt{\frac{1428}{9} - \left(\frac{92}{9}\right)^2}$$
$$= 7.36 \text{ (to 3 s.f.)}$$

b Mean − 2 × standard deviation = −4.50 (to 3 s.f.)
Mean + 2 × standard deviation = 24.9 (to 3 s.f.)
30 is an outlier, as it is more than 2 standard deviations above the mean (30 > 24.9).

c It could be the age of a parent at the party.

d $\sum x - 30 = 92 - 30 = 62$

$\text{Mean} = \frac{62}{8} = 7.75$

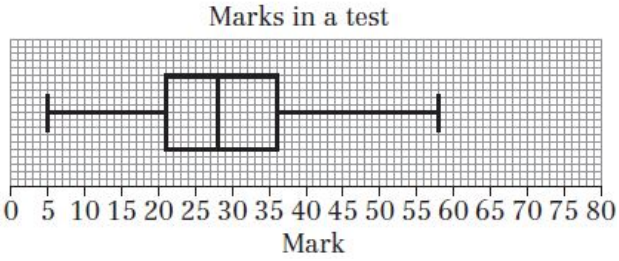
$\sum x^2 - 30^2 = 1428 - 900 = 528$

$\text{Standard deviation} = \sqrt{\frac{528}{8} - \left(\frac{62}{8}\right)^2} = 2.44$ (to 3 s.f.)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Representations of data 3B

1



2 a Upper quartile is 47 marks
Lower quartile is 32 marks

b 38 marks

c $\text{IQR} = 47 - 32 = 15$ marks

d $\text{Range} = 76 - 12 = 64$ marks

3 a Male turtles have a higher median mass.

Males have a bigger range (or IQR).

b It is more likely to be a female. Very few male turtles weigh as little as this, but roughly three-quarters of the female turtles weigh even less.

c 500 g despite this data point being an outlier.

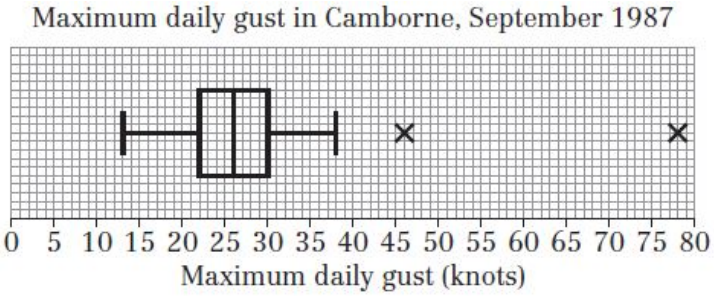
4 a Q_1: $\frac{25}{4} = 6.25$, so we pick the 7th term: 22

Q_2: $\frac{25+1}{2} = 13$, so we pick the 13th term: 26

Q_3: $\frac{3}{4} \times 25 = 18.75$, so we pick the 19th term: 30

b IQR = 30 − 22 = 8
Outliers are values less than 22 − 1.5 × 8 = 10 and values greater than 30 + 1.5 × 8 = 42.
46 and 78 are both greater than 42, so 46 and 78 are both outliers.

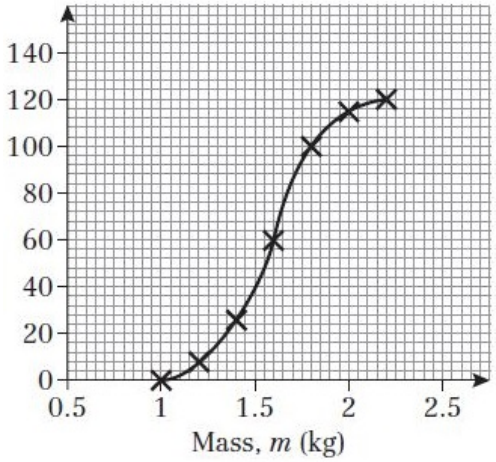
c



© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Representations of data 3C

1 a Cumulative frequencies are: 7, 25, 59, 100, 115, 120

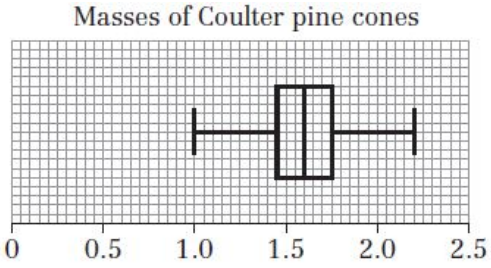


b 1.6 kg

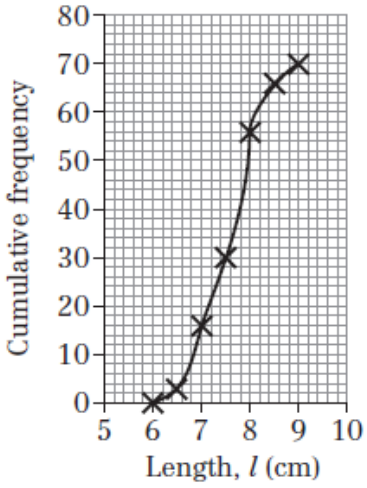
c $Q_1 = 1.43$ kg
$Q_3 = 1.75$ kg
IQR $= 1.75 - 1.43 = 0.32$ kg

$P_{90} - P_{10} = 1.9 - 1.25 = 0.65$ kg

d



2 a Cumulative frequencies are:
3, 16, 30, 56, 66, 70

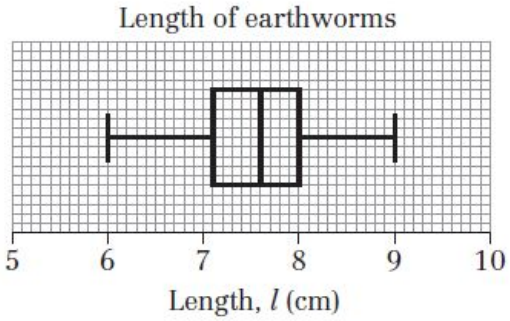


© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

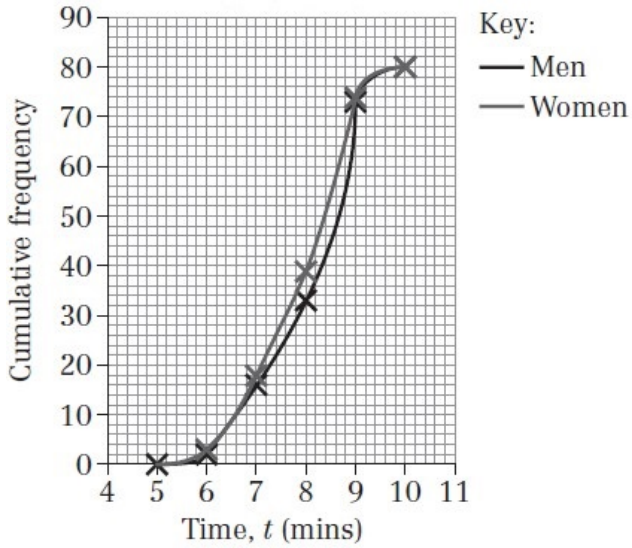
2 b Median = 7.6 cm
Q_1 = 7.1 cm
Q_3 = 8 cm

c i $70 - 62 = 8$
ii 24

d



3 a Cumulative frequencies for men are: 2, 16, 33, 73, 80
Cumulative frequencies for women are: 3, 18, 39, 74, 80



b Median for men is 8.2 mins
Median for women is 8 mins

Women have the lower median time.

c IQR for men is $8.9 - 7.2 = 1.7$ mins
IQR for women is $8.6 - 7.1 = 1.5$ mins

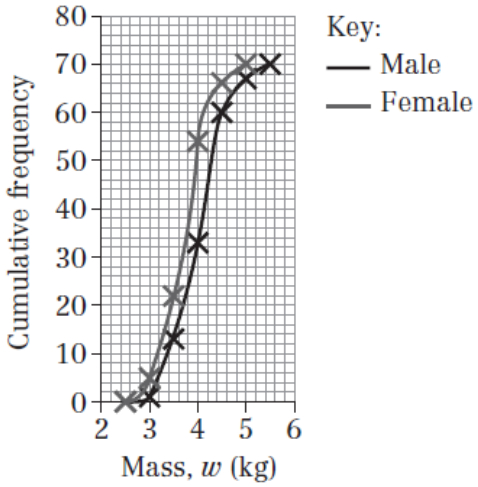
Men have the greater spread of times.

d 24 men and 28 women

4 a Cumulative frequencies for males are: 1, 13, 33, 60, 67, 70
Cumulative frequencies for females are: 5, 22, 54, 66, 70, 70

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

4 a



b IQR for males is $4.4 - 3.6 = 0.8$ kg
IQR for females is $4 - 3.4 = 0.6$ kg

Males have the greater spread of masses.

c There are approximately 12 underweight females, and 25 underweight males.

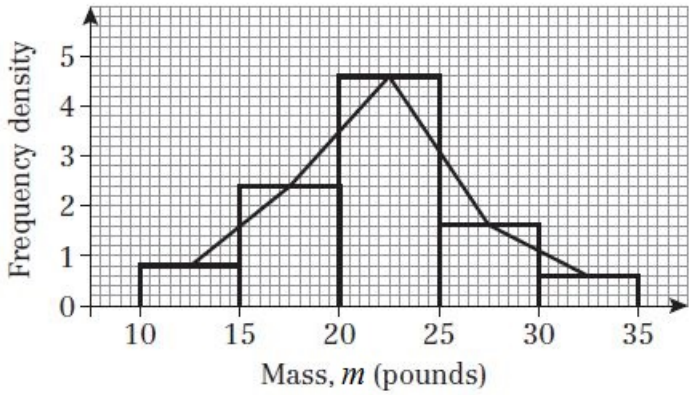
There are fewer female underweight cats.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Representations of data 3D

1 a Class widths are all 5.

Frequency densities are: 0.8, 2.4, 4.6, 1.6, 0.6



2 a Time is a continuous variable.

b Frequency = frequency density × class width
= 5 × 20 = 100

So there were 100 students who took between 40 and 60 seconds.

c 100 students took between 40 and 60 seconds (from part **b**)
6 × 10 = 60 students took between 60 and 70 seconds.
8.6 × 10 = 86 students took between 70 and 80 seconds.

So 100 + 60 + 86 = 246 students took 80 seconds or less.

d 246 students took 80 seconds of less (from part **c**)
14 × 5 = 70 students took between 80 and 85 seconds.
12 × 5 = 60 students took between 85 and 90 seconds.
3 × 30 = 90 students took between 90 and 120 seconds.

Total: 246 + 70 + 60 + 90 = 466
There are 466 students in total.

3 a Distance is a continuous variable.

b Frequency = frequency density × class width
0 m to 20 m: 2 × 20 = 40
20 m to 35 m: 5 × 15 = 75
35 m to 45 m: 10 × 10 = 100
45 m to 60 m: 6 × 15 = 90
60 m to 65 m: 1 × 5 = 5

Total: 40 + 75 + 100 + 90 + 5 = 310

So 310 people entered the competition.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 c Frequency = frequency density × class width
30 m to 35 m: 5 × 5 = 25
35 m to 40 m: 10 × 5 = 50

Total: 25 + 50 = 75

So 75 people threw between 30 and 40 metres.

d From part **b**:
45 m to 60 m: 90 people
60 m to 65 m: 5 people

Total: 90 + 5 = 95

So 95 people threw between 45 and 60 metres.

e From part **b**, 40 people threw between 0 and 20 metres.
20 m to 25 m: 5 × 5 = 25

Total: 40 + 25 = 65

So 65 people threw between 0 and 25 metres.

4 a The bar for $28 \leqslant m < 32$ has an area of 10 × 10 = 100 squares.
If 100 squares represents 32 lambs then
$\frac{100}{4}$ squares represents $\frac{32}{4}$ lambs.
i.e. 25 squares represents 8 lambs.

b The class $24 \leqslant m < 26$ contains 5 × 20 = 100 squares.
As above, this represents 32 lambs.

c The class $20 \leqslant m < 24$ contains 10 × 10 = 100 squares which represents 32 lambs.

The class $24 \leqslant m < 26$ contains 5 × 20 = 100 squares which represents 32 lambs.

The class $26 \leqslant m < 28$ contains 5 × 40 = 200 squares which represents 64 lambs.

The class $28 \leqslant m < 32$ contains 10 × 10 = 100 squares which represents 32 lambs.

The class $32 \leqslant m < 34$ contains 5 × 5 = 25 squares which represents 8 lambs.

So in total we have 32 + 32 + 64 + 32 + 8 = 168 lambs.

d Class $25 \leqslant m < 26$ is approximately $\frac{1}{2}$ of class $24 \leqslant m < 26$ which equates to 16 lambs.
Class $26 \leqslant m < 28$ represents 64 lambs.
Class $28 \leqslant m < 29$ is approximately $\frac{1}{4}$ of class $28 \leqslant m < 32$ which equates to 8 lambs.

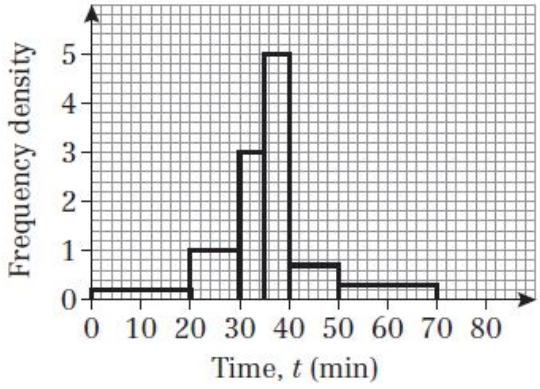
So in total we have 16 + 64 + 8 = 88 lambs.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 a i Use extra columns to help, using the frequency densities given in the histogram:

Time, t (min)	Frequency	Class width	Frequency density
$0 \leqslant t < 20$	4	20	0.2
$20 \leqslant t < 30$	$10 \times 1 = 10$	10	1
$30 \leqslant t < 35$	15	5	3
$35 \leqslant t < 40$	25	5	5
$40 \leqslant t < 50$	$10 \times 0.7 = 7$	10	0.7
$50 \leqslant t < 70$	$20 \times 0.3 = 6$	20	0.3

5 a ii



b $\left(\frac{5}{10}\times 10\right)+15+\left(\frac{3}{5}\times 25\right)=35$ passengers.

6 a 12.5 and 14.5 are the class boundaries, as we are dealing with continuous data.

b i The class boundaries for the 15–17 class are 14.5 and 17.5.
This width is 1.5 times the width of the 13–14 class, since 17.5 – 14.5 = 3 = 1.5 x 2.
So the width of the class is $1.5 \times 4 = 6$ cm.

ii The frequency density for the 13–14 class is $\frac{24}{2}=12$.

The frequency density of this class is 6, which is 0.5 times the frequency density above: 12.
So the height of the class is 0.5 x 6 = 3 cm.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

7 a Width is half of the $8 \leqslant t < 10$ class, which is 0.5 cm.

Height is double the frequency density, so must be $\frac{7}{1} \times 2 = 14$ cm.

b Mean = $\frac{\sum fx}{\sum f} = \frac{645}{62} = 10.4$°C (to 3 s.f.) where x is taken as the midpoint of each class.

Standard deviation = $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{113.89 - \left(\frac{645}{62}\right)^2} = 2.4$ (to 2 s.f.)

c Q_1 is the $\frac{31}{4} = 7.75^{\text{th}}$ piece of data.

$\frac{Q_1 - 8}{10 - 8} = \frac{7.75 - 4}{12 - 4}$ using linear interpolation on the $8 \leq t < 10$ class.

$Q_1 = 8.94$°C

d Mean + standard deviation = 12.8°C

$\frac{12.8 - 12}{15 - 12} = \frac{d - 25}{30 - 25}$

$d = 26.33$ days
So 4.7 days

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Representations of data 3E

1 Median for motorway A = 76 mph
Median for motorway B = 72 mph

IQR for motorway A = 80 − 75 = 5 mph
IQR for motorway B = 75 − 65 = 10 mph

The median speed is greater on motorway A than on motorway B. The spread of speeds for motorway B is greater than the spread of speeds for motorway A.

2 Mean for class 2B = $\frac{\sum x}{n} = \frac{650}{20} = 32.5$ minutes

Mean for class 2F = $\frac{\sum x}{n} = \frac{598}{22} = 27.2$ minutes (to 3 s.f.)

Standard deviation for class 2B = $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{22000}{20} - \left(\frac{650}{20}\right)^2}$

$= 6.61$ (to 3 s.f.)

Standard deviation for class 2F = $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{19100}{22} - \left(\frac{598}{22}\right)^2}$

$= 11.4$ (to 3 s.f.)

The mean time for Class 2B is higher than the mean time for Class 2F, showing that Class 2F are generally faster at completing the puzzle. The standard deviation for Class 2F is bigger than for Class 2B, showing that the times are more spread out.

3 Boys median = 163 cm
Girls median = 158 cm

Boys IQR = 171 − 154.5 = 16.5 cm
Girls IQR = 164.5 − 149.5 = 15 cm

The median height for boys is higher than the median height for girls, showing that boys are gnerally taller. Comparing interquartile ranges,,the spread of heights for boys is greater than the spread of heights for girls.

4 a Leuchars: 89, 91, 94, 98, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100

Q_1 = 4th value = 98

Median = 8th value = 100

Q_3 = 12th value = 100

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

4 a Camborne: 81, 90, 91, 92, 95, 96, 98, 98, 99, 99, 99, 100, 100, 100, 100

Q_1 = 4th value = 92

Median = 8th value = 98

Q_3 = 12th value = 100

b The median for Leuchars is higher than for Camborne, showing that the humidity in Leuchars is higher. Comparing interquartile ranges, the spread of humidity in Camborne is greater than the spread of humidity in Leuchars.

Large data set

1 a Hurn 1987 mean $= \frac{\sum x}{n}$

$= \frac{1103}{167}$

$= 6.60$ kn (to 3 s.f.)

Hurn 2015 mean $= \frac{\sum x}{n}$

$= \frac{1418}{184}$

$= 7.71$ kn (to 3 s.f.)

b Hurn 1987 mode = 4 kn
Hurn 2015 mode = 7 kn

c Hurn 1987 standard deviation $= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{8793}{167} - \left(\frac{1103}{167}\right)^2}$

$= 3.00$ kn (to 3 s.f.)

Hurn 2015 standard deviation $= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{12410}{184} - \left(\frac{1418}{184}\right)^2}$

$= 2.84$ kn (to 3 s.f.)

2 The mean and modal windspeeds were higher in 2015 than in 1987. The spread of the speeds was greater in 1987 than in 2015 due to a higher standard deviation.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Representations of data, Mixed Exercise 3

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

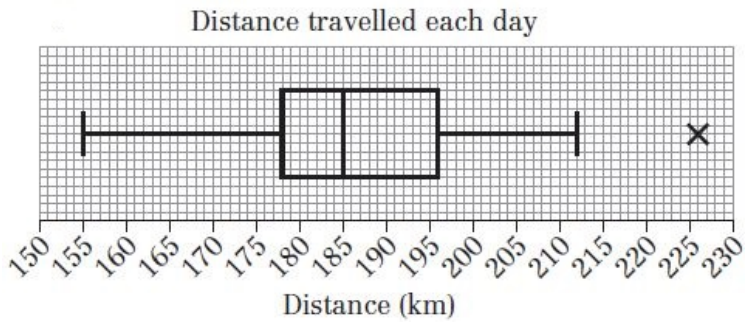
Q_2: $\frac{31+1}{2} = 16$ so we pick the 16th value: 185

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



2 a 45 minutes

b 60 minutes

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

d The Irt club had the highest median, so overall they had the slowest runners.

The IQR ranges were about the same, with the Irt club slightly more spread out.

e With the exception of the outlier, the Esk 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 Irt Club runners.

f Advantages, any one from:

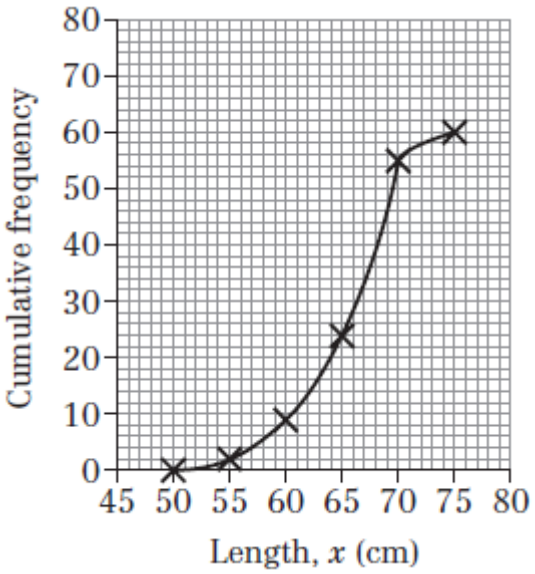
It helps us to see the spread of the data easily.
The plot is clear and easy to understand.
It uses the range and the median values.
It is easy to compare the stratified data.

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.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 a Cumulative frequencies are: 2, 9, 24, 55, 60



b Median = 66 cm

c IQR = 69 − 62.5 = 6.5 cm

d Median = 65.5 cm

IQR = 70 − 63.25 = 6.75

The median length of the honey badgers is slightly higher than the median length for the European badgers, showing that honey badgers are only a little longer on average than European badgers. Comparing interquartiles ranges, the spread of lengths of European badgers is slightly greater than the spread of lengths of honey badgers.

e The cumulative frequency diagram shows us how the lengths of the badgers is spread and enables us to estimate the median and quartiles, but tells us little about individual data points.

4 a The areas of the bars is proportional to the frequency 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 $=2\left((4.5\times2)+(1\times4)\right)=26$ girls

b Number of girls who took between 52 and 55 seconds $=2\left((1.5\times2)+(\frac{1}{2}\times2\times5.5)\right)=17$ girls

5 $1.5\times5.7\times k=2565$, so $k = 300$

a $\text{width}=1\times1.5=1.5\,\text{cm}$

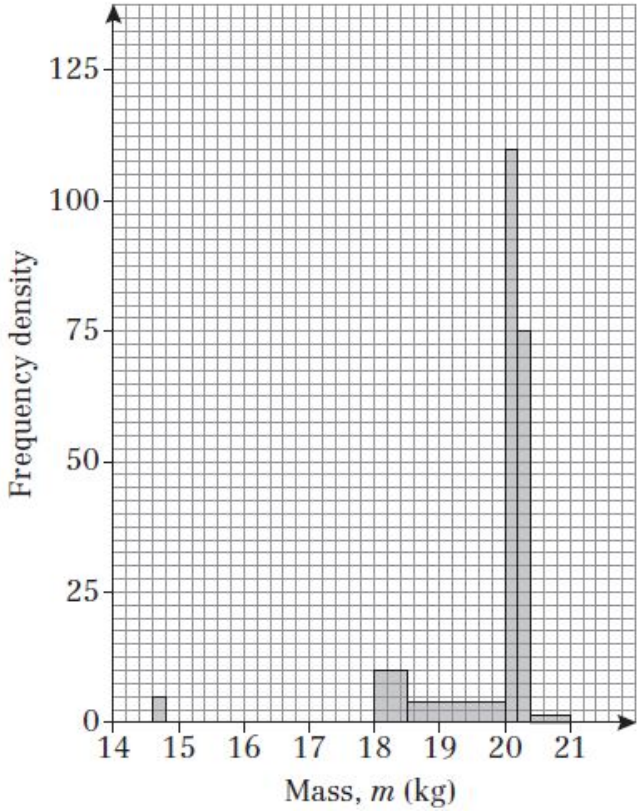
$$\text{height}=\frac{\text{frequency}}{k\times\text{width}}=\frac{1170}{300\times1.5}=2.6\,\text{cm}$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 b width $= 5 \times 1.5 = 7.5\,\text{cm}$

height $= \dfrac{630}{300 \times 7.5} = 0.28\,\text{cm}$

6 a



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

Standard deviation $= \sqrt{\dfrac{\Sigma fy^2}{n} - \mu^2} = \sqrt{\dfrac{19602.84}{50} - 19.777^2} = \sqrt{0.927} = 0.963$ (to 3 s.f.)

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

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

b Median is the $\dfrac{14+1}{2} = 7.5$th piece of data: 20

Q_1 is the $\dfrac{14}{4} = 3.5$th piece of data, so we choose the 4th: 13

Q_3 is the $\dfrac{3 \times 14}{4} = 10.5$th piece of data, so we choose the 11th: 31

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

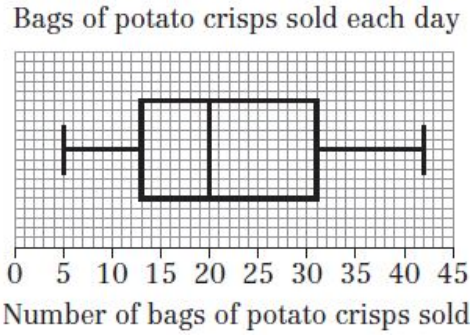
7 c $\text{IQR} = 31 - 13 = 18$ so $1.5 \times \text{IQR} = 27$

$13 - 27 = -14$

$31 + 27 = 58$

So there are no outliers.

d



8 a The maximum gust is a continuous variable and the data is given in a grouped frequency table.

b $10 \leqslant g < 15$ bar:

Width = 2.5, class width = 5, so width = class width × 0.5
Area = 2.5 × 1.8 = 4.5 cm^2, frequency = 3, so area = frequency × 1.5

$18 \leqslant g < 20$ bar:

Width = class width × 0.5 (from above) = 2 × 0.5 = 1 cm
Area = frequency × 1.5 (from above) = 9 × 1.5 = 13.5 cm^2

$$\text{Height} = \frac{\text{Area}}{\text{Width}} = 13.5 \text{ cm}$$

c $$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{1334.5}{57} = 23.4 \text{ kn (to 3 s.f.)}$$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2} = \sqrt{\frac{34299.25}{57} - \left(\frac{1334.5}{57}\right)^2} = 7.32 \text{ (to 3 s.f.)}$$

d Using exact figures throughout:

lower boundary = mean – standard deviation = 16.1 (to 3 s.f.)
higher boundary = mean + standard deviation = 30.7 (to 3 s.f.)

$$\frac{\text{lower boundary} - 15}{18 - 15} = \frac{g - 3}{12 - 3} \Rightarrow g = 6.27 \text{ (to 3 s.f.)}$$

$$\frac{30.72 - 30}{50 - 30} = \frac{h - 50}{57 - 50}$$

$h - g = 44.0$ (to 3 s.f.)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

9 a Mean for 1987 = $\frac{\sum x}{n} = \frac{356.1}{30} = 11.9$ °C (to 1 d.p.)

Mean for 2015 = $\frac{\sum x}{n} = \frac{364.1}{30} = 12.1$ °C (to 1 d.p.)

b Standard deviation for 1987 = $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{4408.9}{30} - \left(\frac{356.1}{30}\right)^2} = 2.46$ °C (to 3 s.f.)

The mean temperature was slightly higher in 2015 than in 1987. The standard deviation of temperatures was higher in 1987 than in 2015, so the temperatures were more spread out.

c In all calculations, exact figures are used.
critical temperature = mean − 1.02 = 11.08 °C

Let a = the number of the day with this temperature.

$\frac{\text{critical temperature} - 10.1}{14.1 - 10.1} = \frac{a - 0}{30 - 0}$ so $a = 7.35$

So there are 7 days where the temperature $\leqslant$ 11.08 °C.

critical temperature = mean + 1.02 = 13.12 °C

Let b = the number of the day with this temperature.

$\frac{\text{critical temperature} - 10.1}{14.1 - 10.1} = \frac{b - 0}{30 - 0}$ so $b = 22.65$

So there are 22 days where the temperature $\leq$ 13.12 °C.

30 − (b − a) = 14.7 (to 3 s.f.)

So there are 15 'abnormal' days.

Assumption: the temperatures are equally distributed throughout the range.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Challenge

Length (mins)	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 $x = \frac{k}{5}$

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

Using substitution

$10 \times \frac{k}{5} + 30 = 17k$, so $2k + 30 = 17k$, i.e. $k = 2$

$x = \frac{k}{5} = 0.4$

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

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

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Correlation 4A

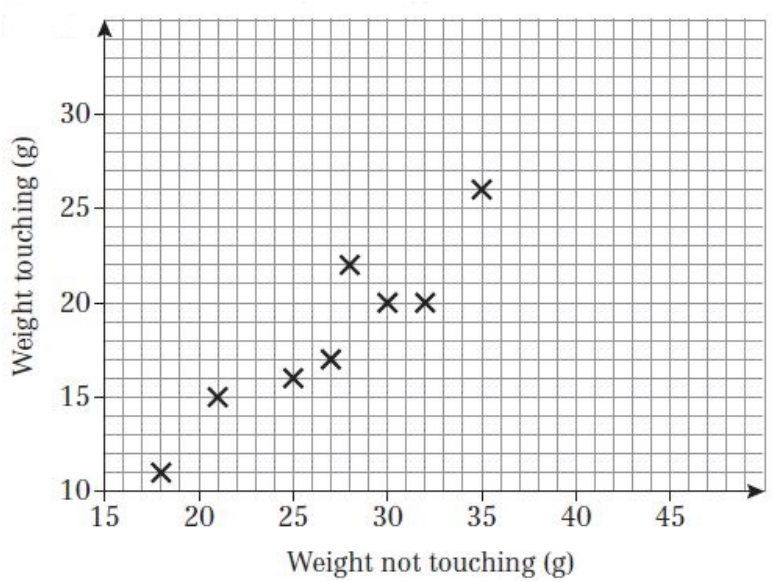
1 a Positive correlation.

b The longer the treatment the greater the loss of weight.

2 a There is no correlation.

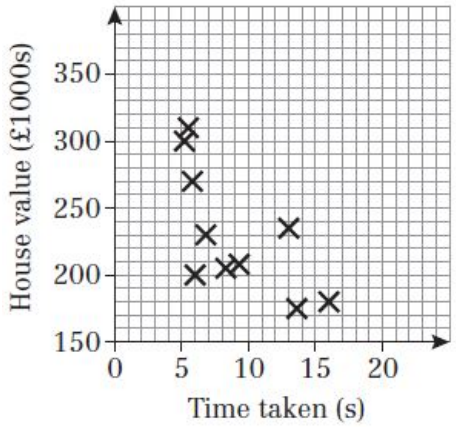
b The scatter graph does not support the statement that hotter cities have less rainfall.

3 a



b It shows positive correlation. Each student's tendency to guess lower or higher than the mean was the same in both tests.

4 a



b There is weak negative correlation

c For example, there may be a third variable that influences both house value and internet connection, such as the distance from built up areas.

5 a $Q_3 + 1.5(Q_3 - Q_1) = 4.85 + 1.5(4.85 - 0.1) = 11.975$

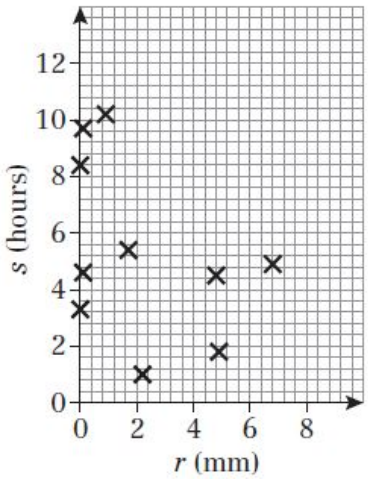
As $21.7 > 11.975$, r is an outlier.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

5 b i The outlier is part of the distribution of the data.
It is unlikely to be an anomaly.

ii 21.7 is an outlier so may not be representative of the typical rainfall

c



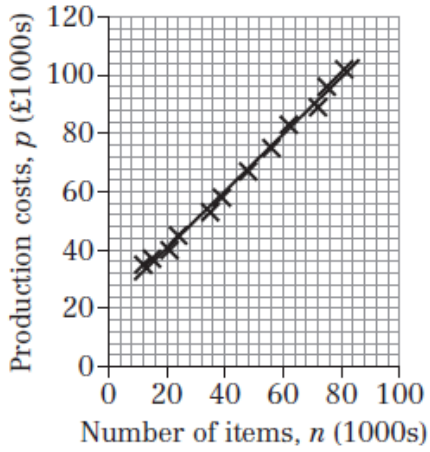
d There is weak negative correlation

e There could be a causal relationship, as days with more rainfall will have more clouds and therefore less sunshine.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Correlation 4B

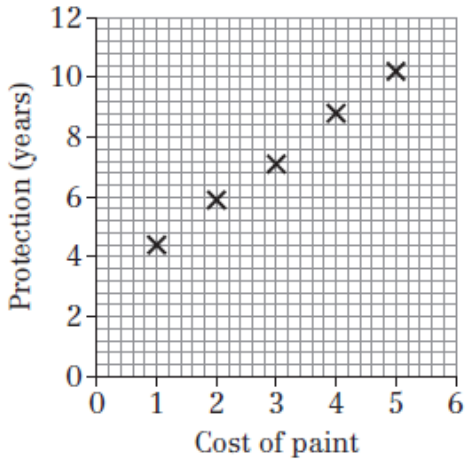
1 **a** and **b**



c If the number of items produced is zero, the production costs will be approximately £21 000. If the number of items produced increases by 1000, the production costs increase by approximately £980.

d The prediction for 74 000 is within the range of the data (interpolation) so is more likely to be reliable The prediction for 95 000 is outside the range of the data (extrapolation) so is less likely to be reliable.

2 **a**



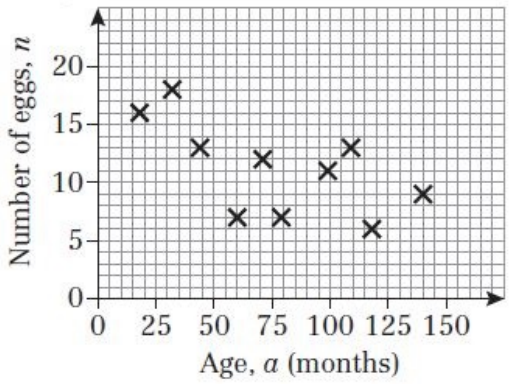
b There are two key problems with Helen's statement:

First, 10 coats of paint is very far outside our range of given data, and we cannot assume that this linear relationship continues as we extrapolate, so using the regression line is not necessarily valid.

Second, even if we accept the extrapolation as valid, a gradient of 1.45 means that, for every extra coat of paint, the protection will increase by 1.45 years. Therefore, if 10 coats of paint are applied, the protection will be 14.5 years longer than if no paint were applied. Helen has, however, forgotten to include the constant 2.93 years, which is the weather resistance if no paint were applied. After 10 coats of paint the protection will last approximately 2.93 + 14.5 = 17.43 years.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 a



b The scatter diagram shows weak negative correlation, therefore the gradient in the regression equation, given as 0.063, should be negative.

4 This is not a reasonable statement as there are unlikely to be any houses with no bedrooms, so she is extrapolating outside of the range of data, where the linear relationship is unlikely to continue.

5 a If the humidity increases by 1%, the decrease in visibility is approximately 106 m.

b High humidity (> 95%) can give rise to misty and foggy conditions, which in turn will cause lower visibility. Hence there is likely to be a causal relationship.

c i The prediction for 100% is outside the range of the data (extrapolation) so is less likely to be accurate.

ii This particular regression equation should only be used to predict a value for v given h.

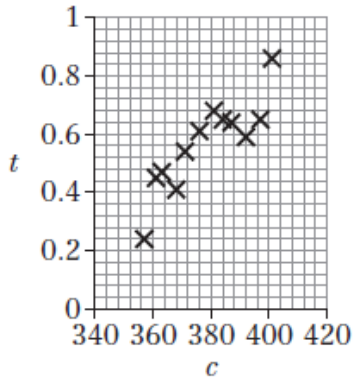
d The data provided is only useful for analysing the first two weeks of September. Random values throughout September should be used and analysis made of the whole month. The sample size could also be increased across multiple months as data between May and October is available.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Correlation, Mixed Exercise 4

1 The data shows that the number of serious road accidents in a week strongly correlates with the number of fast food restaurants. However, it does not show whether the relationship is causal. Both variables could correlate with a third variable, e.g. the number of roads coming into a town.

2 a



b There is strong positive correlation.

c As mean CO_2 concentration in the atmosphere increased, mean temperatures also increased.

3 a There is strong positive correlation.

b If the number of items increases by 1, the time taken increases by approximately 2.64 minutes.

4 The answer is likely to be unreliable as it involves extrapolation. 3500 is well outside the limits of the data set used.

The economist used the regression line of y on x. Estimating energy consumption (x) from Gross National Product (y) would need the regression line of x on y.

5 a $15.2 + 2 \times 11.4 = 38$

As $50 > 38$, $t = 50$ °C is an outlier.

b The outlier should be omitted, as it is very unlikely that the average temperature was 50 °C in a climate where people need to buy gloves, and so this data point is likely an anomaly.

c If the temperature increases by approximately 1 °C, the number of pairs of gloves sold each month decreases by 5.2.

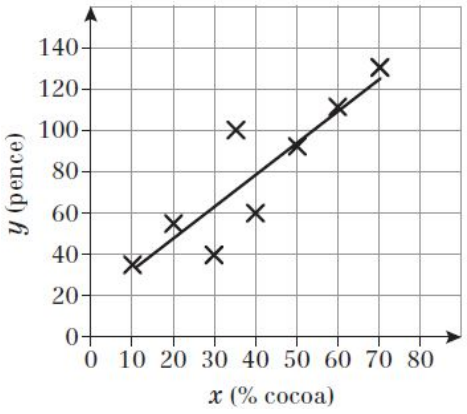
6 a 44 cm is the length of the spring with no mass attached. If a mass of 1 g is attached, the spring will increase in length by approximately 0.2 cm.

b i $m = 150$ is outside the range of the data (extrapolation) so is less likely to be accurate.

ii This particular regression equation should only be used to predict a value of s given m. To predict a value of m given s, you should use the regression equation of m on s.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

7 a and **b**

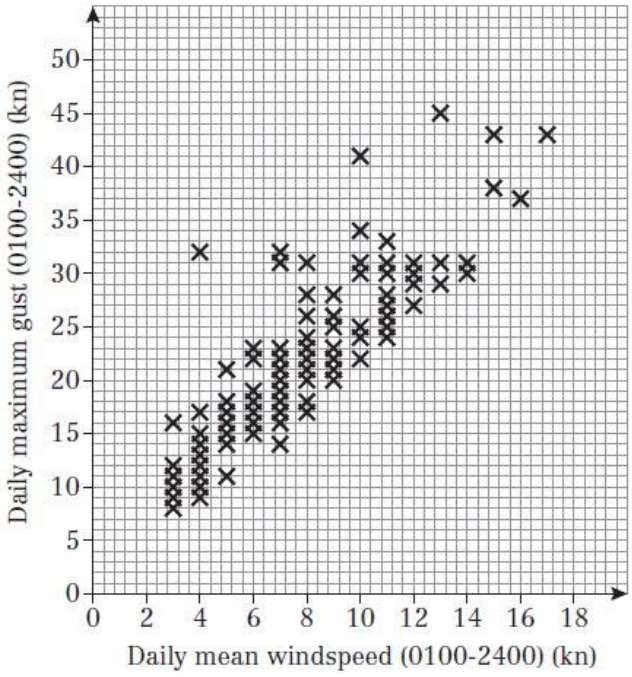


c Brand D is overpriced, since it's price is much more than you would expect (the data point is far above the regression line)..

d The regression equation should be used to predict a value for y given x, i.e. the price given the percentage of cocoa solids. So the student's method is a valid one.

Large data set

1 a



b There is moderate positive correlation

c The relationship is causal, as the maximum gust is related to the mean windspeed. A higher mean windspeed is likely to result in a higher maximum gust.

d i $g = 4.97 + 2.15 \times 0.5$
$= 6.045$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

d **ii** $g = 4.97 + 2.15 \times 5$
$= 15.72$ kn

iii $g = 4.97 + 2.15 \times 12$
$= 30.77$ kn

iv $g = 4.97 + 2.15 \times 40$
$= 90.97$ kn

e Parts **ii** and **iii** are within the range of the data (interpolation), so are more likely to be accurate.
Parts **i** and **iv** are outside the range of the data (extrapolation), so are less likely to be accurate.

f Using the 'SLOPE' and 'INTERCEPT' function in a spreadsheet:

$w = 0.05 + 0.35\,g$

$w = 0.05 + 0.35 \times 30$
$= 10.55$ kn

2 **a** Using the 'SLOPE' and 'INTERCEPT' function in a spreadsheet:

$s = 14.6 - 1.7c$ (values taken to 2 d.p.)

The missing total sunshine data values are approximately:

2.7, 7.8, 6.1, 11.2, 2.7. 7.8, 11.2, 14.6, 9.5, 6.1. 2.7, 4.4, 7.8, 2.7, 4.4, 4.4, 1.0

b The relationship is likely to be causal: greater cloud cover is likely to result in less sunshine.

Probability 5A

1

		Coin 1	
		H	**T**
Coin 2	**H**	HH	TH
	T	HT	TT

$$\text{P(same)} = \frac{2}{4} = \frac{1}{2}$$

2 a

		Second roll					
		1	2	3	4	5	6
First roll	1	1	2	3	4	5	6
	2	2	4	6	8	10	12
	3	3	6	9	12	15	18
	4	4	8	12	16	20	24
	5	5	10	15	20	25	30
	6	6	12	18	24	30	36

b i $\text{P}(X = 24) = \frac{2}{36} = \frac{1}{18}$

ii $\text{P}(X < 5) = \frac{8}{36} = \frac{2}{9}$

iii $\text{P}(X \text{ is even}) = \frac{27}{36} = \frac{3}{4}$

3 a $\text{P}(m \geqslant 54) = \frac{33+21+2}{140} = \frac{56}{140} = \frac{2}{5}$

b $\text{P}(48 \leqslant m < 57) = \frac{25+42+33}{140} = \frac{100}{140} = \frac{5}{7}$

c Let B = the number of Bullmastiffs with mass less than 53 kg.

Using interpolation:

$$\frac{53-51}{54-51} = \frac{B-(17+25)}{42}$$

So $B = 70$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 c $P(m < 53) = \frac{70}{140} = 0.5$, so half of the Bullmastiffs are estimated to have a mass less than 53 kg.

This probability is lower than the probability of 0.54 for Rottweilers, and so it is less likely.

The assumption made is that the frequency is uniformly distributed throughout the class.

4 a $P(\text{female}) = \frac{14+15+32+27+26}{240} = \frac{114}{240} = \frac{19}{40}$

b $P(l < 80) = \frac{4+14+20+15+24+32}{240} = \frac{109}{240}$

c $P(\text{male and } 75 \leqslant l < 85) = \frac{24+47}{240} = \frac{71}{240}$

d Using interpolation for males:

The number of male juvenile koalas is approximately $4 + \frac{72-70}{75-70} \times 20 = 4 + 8 = 12$.

The number of female juvenile koalas is approximately $14 + \frac{72-70}{75-70} \times 15 = 14 + 6 = 20$.

So $P(\text{juvenile}) = \frac{12+20}{240} = \frac{32}{240} = \frac{2}{15}$

The assumption is that the distribution of lengths of koalas between 70 and 75 cm is uniform.

5 a $P(m > 5) = \frac{(1 \times 24) + (2 \times 4)}{70} = \frac{32}{70} = \frac{16}{35}$

b Start with the probability that the cat has a mass *greater* than 6.5.

$P(m > 6.5) = \frac{\frac{3}{4} \times (2 \times 4)}{70} = \frac{6}{70} = \frac{3}{35}$

So $P(m < 6.5) = 1 - \frac{3}{35} = \frac{32}{35}$

The fact that we have ignored the case 6.5 is not a problem in this estimate. We are assuming that the class is continuous when we interpolate, and that the probability of being exactly equal to any individual value is negligible.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Challenge

		A		
B	×	**2**	**7**	**5**
	4	8	28	20
	x	$2x$	$7x$	$5x$

If x is even, all the products are even, so P(Y is even) = 1

But P($Y \geqslant 20$) = 1 is impossible, as the product of 2 and 4 is only 8, so x cannot be even.

If x is odd, there are four even values of Y: 8, 28, 20 and $2x$.

But P(Y is even) = P($Y \geqslant 20$), so there must also be four values where $Y \geqslant 20$.

Two of them are in the top row: 28 and 20, leaving two in the bottom row.

Given that exactly two of these three values are greater than or equal to 20:

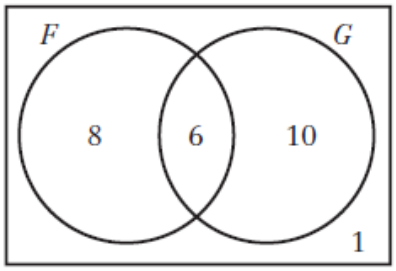
$2x < 20$ and $5x \geqslant 20$, i.e. $x < 10$ and $x \geqslant 4$.

Hence $4 \leqslant x < 10$ and x is odd so the possible values of x are 5, 7 and 9.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Probability 5B

1 a



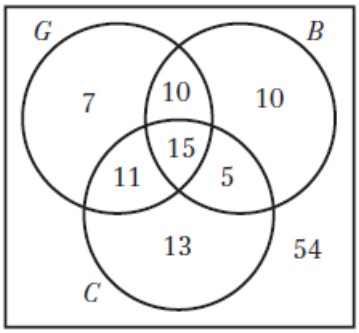
b i $P(F) = \frac{14}{25}$

ii $P(F \cap G) = \frac{6}{25}$

iii $P(\text{French but not German}) = \frac{8}{25}$

iv $P(\text{Neither French nor German}) = \frac{1}{25}$

2 a



b i $P(\text{All three}) = \frac{15}{125} = \frac{3}{25}$

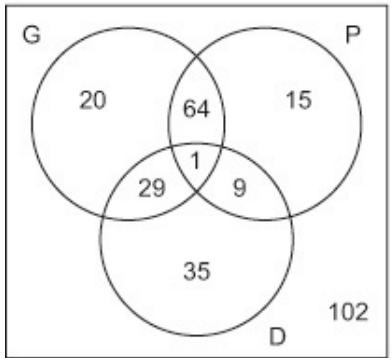
ii $P(\text{Beer but not cheesecake and not garlic bread}) = \frac{10}{125} = \frac{2}{25}$

iii $P(\text{Garlic bread and beer but not cheesecake}) = \frac{10}{125} = \frac{2}{25}$

iv $P(\text{None}) = \frac{54}{125}$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 a



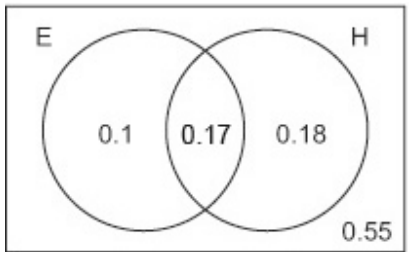
b i $\mathrm{P}(\text{Plays piano}) = \dfrac{89}{275}$

ii $\mathrm{P}(\text{At least 2}) = \dfrac{64+9+29+1}{275} = \dfrac{103}{275}$

iii $\mathrm{P}(\text{Plays exactly one}) = \dfrac{20+15+35}{275} = \dfrac{14}{55}$

iv $\mathrm{P}(\text{Plays none}) = \dfrac{102}{275}$

4

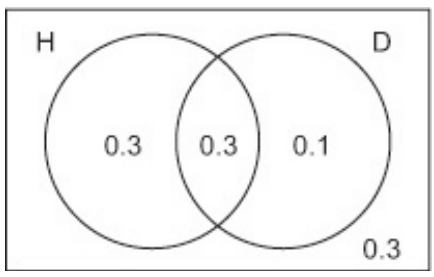


a $\mathrm{P}(E \cap H) = \mathrm{P}(E) + \mathrm{P}(H) - \mathrm{P}(E \cup H)$
$= 0.27 + 0.35 - 0.45$
$= 0.17$

b $\mathrm{P}(\text{Blonde hair but not Blue eyes}) = 0.35 - 0.17 = 0.18$

c $\mathrm{P}(\text{Neither}) = 1 - \mathrm{P}(E \cup H) = 1 - 0.45 = 0.55$

5



© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 a $P(H \cap D) = P(H) + P(D) - P(H \cup D)$
$= 0.6 + 0.4 - 0.7$
$= 0.3$

b P(Hiya only) $= 0.6 - 0.3 = 0.3$

6 a $P(B) = x + 0.1 + 0.2 = 0.45$

So $x = 0.45 - 0.3 = 0.15$.

b $y = 1 - (0.35 + 0.15 + 0.1 + 0.2 + 0.05) = 1 - 0.85 = 0.15$

7 $P(M) = 0.32 + p$

$P(P) = p + q + 0.07$

As P(M) = P(P), $0.32 + p = p + q + 0.07$.

So, rearranging, $q = 0.32 - 0.07 = 0.25$.

$p = 1 - (0.32 + 0.25 + 0.07 + 0.13 + 0.1) = 0.13$

$p = 0.13$, $q = 0.25$

Challenge

$P(B) = p + q + 0.05$

$P(A) = 0.15 + p$

As $P(B) = 2P(A)$, $p + q + 0.05 = 2(0.15 + p)$, or $p + q + 0.05 = 0.3 + 2p$

So our first equation relating p and q is: $q = 0.25 + p$

As P(not C) = 0.83

$0.15 + p + q + 0.2 = 0.83$, so our second equation results: $p + q = 0.48$

Using substitution to solve simultaneously:

$p + (0.25 + p) = 0.48$, so $2p = 0.23$ and therefore $p = 0.115$

$q = 0.25 + 0.115 = 0.365$

$P(C) = 1 - P(\text{not } C) = 1 - 0.83 = 0.17$

Hence $r + 0.05 = 0.17$, so $r = 0.12$

$p = 0.115$, $q = 0.365$, $r = 0.12$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Probability 5C

1 a



b $P(A \cup B) = 0.7$

c $P(A' \cap B') = 0.3$

2 $P(\text{Sum of 4}) = \frac{3}{36} = \frac{1}{12}$

$P(\text{Same number}) = \frac{6}{36} = \frac{1}{6}$

$P(\text{Sum of 4}) + P(\text{Same number}) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}$

$P(\text{Sum of 4 or same number}) = \frac{8}{36} = \frac{2}{9}$

$P(\text{Sum of 4}) + P(\text{Same number}) \neq P(\text{Sum of 4 or same number})$,
so the events are not mutually exclusive.

Alternatively: A roll of 2 followed by another roll of 2 fits both conditions, so the intersection is not empty, and the events are not mutually exclusive.

3 $P(A \text{ and } B) = P(A) \times P(B) = 0.5 \times 0.3 = 0.15$

4 $P(A \text{ and } B) = P(A) \times P(B)$

$P(B) = P(A \text{ and } B) \div P(A) = 0.045 \div 0.15 = 0.3$

5 a The closed curves representing bricks and trains do not overlap and so they are mutually exclusive.

b $P(B \text{ and } F) = \dfrac{1}{3+1+4+6+2+5} = \dfrac{1}{21}$

$P(B) \times P(F) = \dfrac{3+1}{21} \times \dfrac{1+4+6}{21} = \dfrac{4}{21} \times \dfrac{11}{21} = \dfrac{44}{441}$

As $P(B \text{ and } F) \neq P(B) \times P(F)$, 'plays with bricks' and 'plays with action figures' are not independent events.

6 a $0.4 + x + 0.3 + 0.05 = 1$

$x = 0.25$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 b P(A and B) = x = 0.25

P(A) × P(B) = 0.65 × 0.55 = 0.3575

As P(A and B) ≠ P(A) × P(B), the two events 'like pasta' and 'like pizza' are not independent.

7 a P(S and T) = P(S) – P(S but not T) = 0.3 − 0.18 = 0.12

P(S) × P(T) = 0.3 × 0.4 = 0.12

As P(S and T) = P(S) × P(T), S and T are independent events.

b i P(S and T) = 0.12, as above.

ii P(neither S nor T) = 1 – P(S or T) = 1 – (P(S but not T) + P(T)) = 1 – (0.18 + 0.4) = 0.42

8 P(W and X) = P(W) – P(W and not X) = 0.5 − 0.25 = 0.25

P(X) = 1 – (P(W and not X) + P(neither W nor X)) = 1 – (0.25 + 0.3) = 0.45

P(W) × P(X) = 0.5 × 0.45 = 0.225

As P(W and X) ≠ P(W) × P(X), the two events W and X are not independent.

9 a P(A or R) = P(A) + P(R) = 0.6 because A and R are mutually exclusive.

0.2 + 0.25 + x = 0.6, so x = 0.15

y = 1 − (0.2 + 0.25 + 0.15 + 0.1) = 0.3

(x, y) = (0.15, 0.3)

b P(R and F) = x = 0.15

P(R) × P(F) = 0.4 × 0.45 = 0.18

As P(R and F) ≠ P(R) × P(F), the two events R and F are not independent.

10 P(A and B) = p

P(A) × P(B) = (0.42 + p) × (p + 0.11)

= (p + 0.42)(p + 0.11)

As the events A and B are independent, P(A and B) = P(A) × P(B), so

(p + 0.42)(p + 0.11) = p

p^2 + 0.53p + 0.0462 = p

p^2 − 0.47p + 0.0462 = 0, a quadratic in p, which we can solve with the quadratic formula

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

10 $p = \dfrac{0.47 \pm \sqrt{(-0.47)^2 - 4(1)(0.0462)}}{2(1)}$

$p = \dfrac{0.47 \pm 0.19}{2}$

$p = 0.33$ or 0.14

When $p = 0.14$, $q = 1 - (0.42 + 0.14 + 0.11) = 0.33$

When $p = 0.33$, $q = 1 - (0.42 + 0.33 + 0.11) = 0.14$

$(p, q) = (0.14, 0.33)$ or $(0.33, 0.14)$

Challenge

a Set $P(A) = p$ and $P(B) = q$

As A and B are independent events, $P(A \text{ and } B) = P(A) \times P(B) = pq$

$P(A \text{ and not } B) = P(A) - P(A \text{ and } B) = p - pq$, and notice $P(\text{not } B) = 1 - P(B) = 1 - q$

Then $P(A) \times P(\text{not } B) = p(1 - q) = p - pq = P(A \text{ and not } B)$

As $P(A \text{ and not } B) = P(A) \times P(\text{not } B)$, the events A and 'not B' are independent.

b Still using p and q as above,

$P(\text{not } A \text{ and not } B) = 1 - P(A \text{ or } B)$

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Meaning $P(\text{not } A \text{ and not } B) = 1 - P(A) - P(B) + P(A \text{ and } B) = 1 - p - q + pq = (1 - p)(1 - q)$

Remember $P(\text{not } A) = 1 - p$ and $P(\text{not } B) = 1 - q$

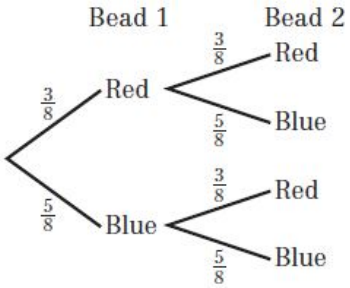
So $P(\text{not } A) \times P(\text{not } B) = (1 - p)(1 - q) = P(\text{not } A \text{ and not } B)$

As $P(\text{not } A \text{ and not } B) = P(\text{not } A) \times P(\text{not } B)$, the events 'not A' and 'not B' are independent.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Probability 5D

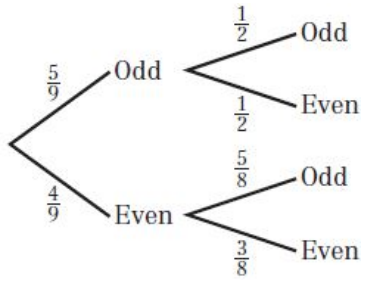
1 a



b $P(\text{both blue}) = \frac{5}{8} \times \frac{5}{8} = \frac{25}{64}$

c $P(\text{second blue}) = \frac{3}{8} \times \frac{5}{8} + \frac{5}{8} \times \frac{5}{8} = \frac{5}{8}$

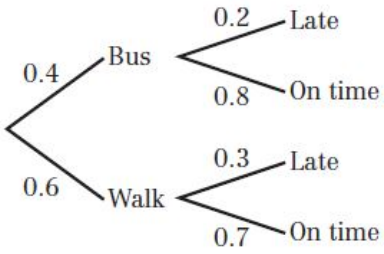
2 a



b $P(\text{both even}) = \frac{4}{9} \times \frac{3}{8} = \frac{12}{72} = \frac{1}{6}$

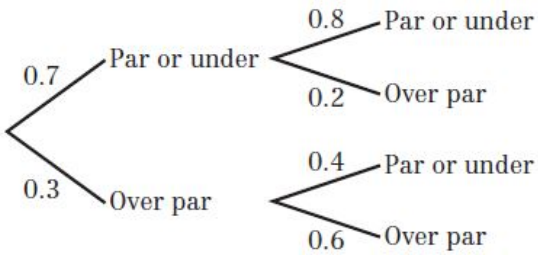
c $P(\text{different parity}) = P(\text{odd then even}) + P(\text{even then odd}) = \frac{5}{9} \times \frac{1}{2} + \frac{4}{9} \times \frac{5}{8} = \frac{40}{72} = \frac{5}{9}$

3 a



b $P(\text{late}) = P(\text{bus and late}) + P(\text{walk and late}) = 0.4 \times 0.2 + 0.6 \times 0.3 = 0.26$

4 a



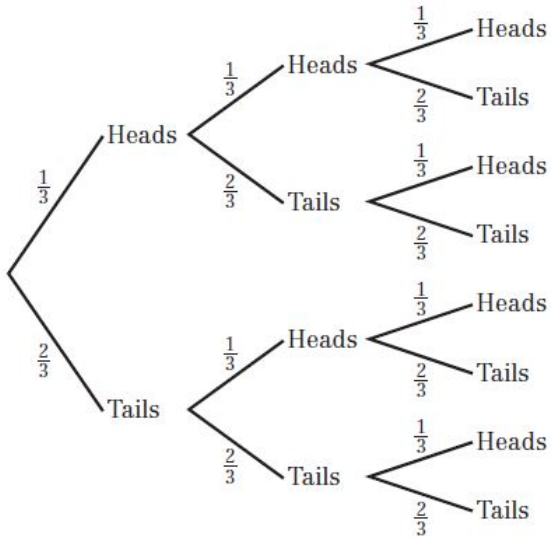
© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

4 b The events are not independent.

c P(par or under on exactly one hole) = P(par or under then over par) + P(over par then par or under)

$= 0.7 \times 0.2 + 0.3 \times 0.4 = 0.26$

5 a



b $P(HHH) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$

c $P(\text{one H only}) = P(HTT) + P(THT) + P(TTH) = \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} + \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} + \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{4+4+4}{27} = \frac{4}{9}$

d $P(\text{HHH or TTT}) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} + \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{1+8}{27} = \frac{1}{3}$

So $P(\text{HHH or TTT in both trials}) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

6 a $P(\text{both yellow}) = P(\text{first } Y) \times P(\text{second } Y) = \frac{6}{13} \times \frac{5}{12} = \frac{30}{156} = \frac{5}{26}$

b $P(\text{third } Y) = \frac{4}{11}$

c $P(\text{all different}) = P(BRY) + P(BYR) + P(RYB) + P(RBY) + P(YBR) + P(YRB)$

$= \frac{4}{13} \times \frac{3}{12} \times \frac{6}{11} + \frac{4}{13} \times \frac{6}{12} \times \frac{3}{11} + \frac{3}{13} \times \frac{6}{12} \times \frac{4}{11} + \frac{3}{13} \times \frac{4}{12} \times \frac{6}{11} + \frac{6}{13} \times \frac{4}{12} \times \frac{3}{11} + \frac{6}{13} \times \frac{3}{12} \times \frac{4}{11}$

$= 6 \times \left(\frac{4 \times 3 \times 6}{13 \times 12 \times 11} \right) = \frac{432}{1716} = \frac{36}{143}$

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Probability, Mixed Exercise 5

1 a $P(RRB \text{ or } RRG) = \left(\frac{7}{15}\times\frac{7}{15}\times\frac{3}{15}\right)+\left(\frac{7}{15}\times\frac{7}{15}\times\frac{5}{15}\right)$

$= \frac{392}{3375}$

b $P(RBG) + P(RGB) + P(BGR) + P(BRG) + P(GBR) + P(GRB)$

$=\left(\frac{7}{15}\times\frac{3}{15}\times\frac{5}{15}\right)+\left(\frac{7}{15}\times\frac{5}{15}\times\frac{3}{15}\right)+\left(\frac{3}{15}\times\frac{5}{15}\times\frac{7}{15}\right)+\left(\frac{3}{15}\times\frac{7}{15}\times\frac{5}{15}\right)+\left(\frac{5}{15}\times\frac{3}{15}\times\frac{7}{15}\right)+\left(\frac{5}{15}\times\frac{7}{15}\times\frac{3}{15}\right)$

$= 6\times\left(\frac{7\times3\times5}{15^3}\right) = \frac{630}{3375} = \frac{14}{75}$

2 a $P(HHH) = 0.341 \times 0.341 \times 0.341 = 0.0397$ (to 3 s.f.)

b $P(NNN) = 0.659 \times 0.659 \times 0.659 = 0.286$ (to 3 s.f.)

c $P(\text{at least one } H) = 1 - P(NNN) = 1 - 0.28619118 = 0.714$ (to 3 s.f.)

3 a $P(\text{female}) = \frac{8+13+19+30+26+32}{250} = \frac{128}{250} = \frac{64}{125}$

b $P(s < 35) = \frac{7+8+15+13+18+19}{250} = \frac{80}{250} = \frac{8}{25}$

c $P(\text{male with score between 25 and 34}) = \frac{15+18}{250} = \frac{33}{250}$

d Using interpolation:

$\text{Number of students passing} = \frac{40-37}{40-35}\times(25+30)+30+26+27+32$

$=\left(\frac{3}{5}\times55\right)+30+26+27+32=148$

$P(\text{pass}) = \frac{148}{250} = \frac{74}{125}$

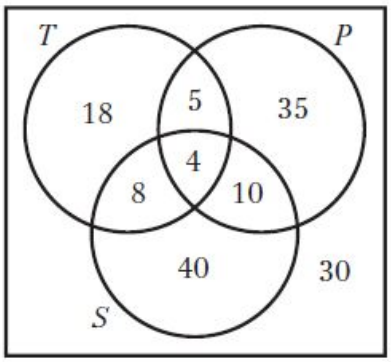
The assumption is that the marks between 35 and 40 are uniformly distributed.

4 a $P(>3) = \frac{0.5\times50+0.5\times30+2\times2}{1\times6+0.5\times50+0.5\times30+2\times2} = \frac{44}{50} = \frac{22}{25}$

b $P(<3.75) = \frac{(1\times6)+(0.5\times50)+0.5\times(0.5\times30)}{50} = \frac{38.5}{50} = 0.77$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 a



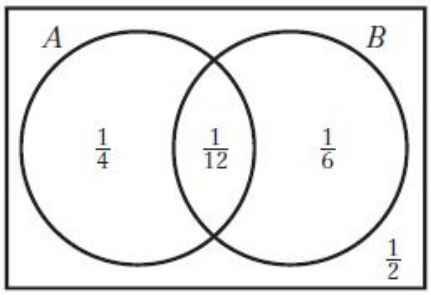
b i $\mathrm{P}(\text{None}) = \frac{30}{150} = \frac{1}{5}$

ii $\mathrm{P}(\text{No more than one}) = \frac{30+40+18+35}{150} = \frac{123}{150} = \frac{41}{50}$

6 a $\mathrm{P}(A \text{ and } B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \text{ or } B \text{ or both}) = \frac{1}{3} + \frac{1}{4} - \frac{1}{2} = \frac{1}{12}$

$\mathrm{P}(A \text{ and not } B) = \frac{1}{3} - \frac{1}{12} = \frac{1}{4}$

$\mathrm{P}(B \text{ and not } A) = \frac{1}{4} - \frac{1}{12} = \frac{1}{6}$



b $\mathrm{P}(A \text{ and } B) = \frac{1}{12}$

$\mathrm{P}(A) \times \mathrm{P}(B) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}$

As $\mathrm{P}(A \text{ and } B) = \mathrm{P}(A) \times \mathrm{P}(B)$, A and B are independent events.

7 a Cricket and swimming do not overlap so are mutually exclusive.

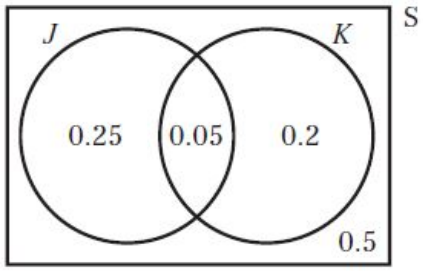
b $\mathrm{P}(C \text{ and } F) = \frac{13}{38}$

$\mathrm{P}(C) \times \mathrm{P}(F) = \frac{21}{38} \times \frac{22}{38} = \frac{462}{1444} = \frac{231}{722}$

As $\mathrm{P}(C \text{ and } F) \neq \mathrm{P}(C) \times \mathrm{P}(F)$, the events 'likes cricket' and 'likes football' are not independent.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 a



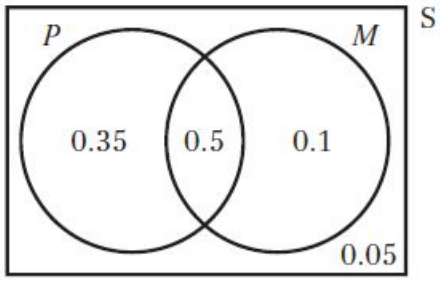
b P(J and K) = 0.05

P(J) × P(K) = 0.3 × 0.25 = 0.075

As P(J and K) ≠ P(J) × P(K), the events J and K are not independent.

9 a P(Phone and MP3) = 0.85 + 0.6 − (1 − 0.05) = 0.5 = 50%

b



c P(only P) = 0.35

d P(P and M) = 0.5

P(P) × P(M) = 0.85 × 0.6 = 0.51

As P(P and M) ≠ P(P) × P(M), the events P and M are not independent.

10 $x = 1 - (0.3 + 0.4 + 0.15) = 0.15$

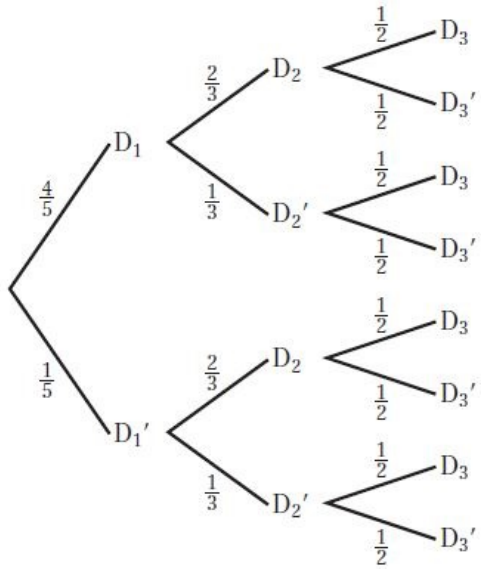
P(A and B) = x = 0.15

P(A) × P(B) = 0.45 × 0.55 = 0.2475

As P(A and B) ≠ P(A) × P(B), the events A and B are not independent.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

11 a



11 b i $P(D_1D_2D_3) = \frac{4}{5}\times\frac{2}{3}\times\frac{1}{2} = \frac{4}{15}$

ii Where D means a diamond and D' means no diamond,

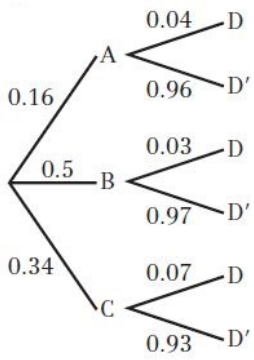
P (exactly one diamond) = P(D, D', D') + P(D', D, D') + P(D', D', D)

$$= \left(\frac{4}{5}\times\frac{1}{3}\times\frac{1}{2}\right) + \left(\frac{1}{5}\times\frac{2}{3}\times\frac{1}{2}\right) + \left(\frac{1}{5}\times\frac{1}{3}\times\frac{1}{2}\right) = \frac{7}{30}$$

11 c P(at least two diamonds) = 1 – P(at most one diamond) = 1 – (P(none) + P(exactly one diamond))

$$= 1 - \left(\frac{1}{5}\times\frac{1}{3}\times\frac{1}{2} + \frac{7}{30}\right) = 1 - \frac{4}{15} = \frac{11}{15}$$

12 a



b i $P(B \text{ and defective}) = 0.5\times0.03 = 0.015$

ii $P(\text{defective}) = 0.16\times0.04 + 0.5\times0.03 + 0.34\times0.07 = 0.0452$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Challenge

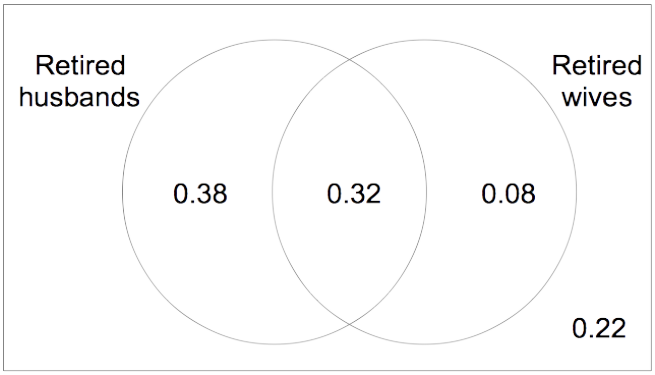
The probability that a wife is retired is 0.4.

Given that she is retired, the probability that her husband is also retired is 0.8.

Hence the probability that both are retired is $0.4 \times 0.8 = 0.32$.

The probability that a husband is retired is 0.7.

From this data you can deduce the following Venn diagram of the probabilities:



Let H = husband retired, H′ = husband not retired, W = wife retired, W′ = wife not retired.

The permutations where only one husband and only one wife is retired are:

Couple 1	Probability	Couple 2	Probability	Combined probability
H W′	0.38	H′ W	0.08	0.38×0.08
H′ W	0.08	H W′	0.38	0.08×0.38
H W	0.32	H′ W′	0.22	0.32×0.22
H′ W′	0.22	H W	0.32	0.22×0.32

$$\text{P(only one husband and only one wife is retired)} = (0.38 \times 0.08 + 0.32 \times 0.22) \times 2 = 0.2016$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Statistical distributions 6A

1 a This is not a discrete random variable, since height is a continuous quantity.

b This is a discrete random variable, since it is takes whole number values at random.

c This is not a discrete random variable, since the number of days in a given week is always 7; the result is predetermined and so not random.

2 {0, 1, 2, 3, 4}

3 a (2, 2), (2, 3), (3, 2), (3, 3)

b i

x	4	5	6
$\mathbf{P}(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

ii

$$\mathrm{P}(X = x) = \begin{cases} \frac{1}{4}, \text{ if } x = 4, 6 \\ \frac{1}{2}, \text{ if } x = 5 \\ 0, \text{ otherwise} \end{cases}$$

4

$$\frac{1}{3} + \frac{1}{3} + k + \frac{1}{4} = 1$$

$$k = 1 - \left(\frac{1}{3} + \frac{1}{3} + \frac{1}{4}\right)$$

$$= 1 - \frac{11}{12}$$

$$= \frac{1}{12}$$

5

x	1	2	3	4
$\mathbf{P}(X = x)$	k	$2k$	$3k$	$4k$

$$k + 2k + 3k + 4k = 1$$

$$10k = 1$$

$$k = \tfrac{1}{10}$$

6 a

x	1	2	3	4
$\mathbf{P}(X = x)$	k	k	$3k$	$3k$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 a Using the fact that the probabilities add up to 1:

$$k + k + 3k + 3k = 1$$
$$8k = 1$$
$$k = \tfrac{1}{8}$$

b The probability distribution is:

x	1	2	3	4
$\mathbf{P}(X = x)$	0.125	0.125	0.375	0.375

$\mathrm{P}(X > 1) = 0.125 + 0.375 + 0.375 = 0.875$

7 a

x	−2	−1	0	1	2
$\mathbf{P}(X = x)$	0.1	0.1	β	β	0.2

The probabilities add up to 1.

$$0.1 + 0.1 + \beta + \beta + 0.2 = 1$$
$$2\beta = 1 - 0.4 = 0.6$$
$$\beta = 0.3$$

b

x	−2	−1	0	1	2
$\mathbf{P}(X = x)$	0.1	0.1	0.3	0.3	0.2

c $\mathrm{P}(-1 \leqslant X < 2) = 0.1 + 0.3 + 0.3 = 0.7$

8 $$\frac{1}{4} - a + a + \frac{1}{2} + a = 1$$

$$\frac{3}{4} + a = 1$$

$$a = \frac{1}{4}$$

9 a $\mathrm{P}(X = 1) = \dfrac{1}{50}$

since each of the 50 individual outcomes is equally likely.

b $\mathrm{P}(X \geqslant 28) = 1 - \dfrac{27}{50} = \dfrac{23}{50}$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

9 c $P(13 < X < 42) = P(14 \leqslant X \leqslant 41) = \frac{28}{50} = \frac{14}{25}$

10 a $P(1 < X \leq 3) = P(X = 2) + P(X = 3) = \frac{1}{2} + \frac{1}{8} = \frac{5}{8}$

b $P(X < 2) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$

c $P(X > 3) = 0$

11 a

s	1	2	3	4
$P(S = s)$	$\frac{2}{3}$	$\frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$	$\frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{27}$	$\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} + \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$

b $P(S > 2) = \frac{2}{27} + \frac{1}{27} = \frac{1}{9}$

12 a

x	0	1	2	3	4	5
$P(X = x)$	$0.6^5 = 0.07776$	$0.4 \times 0.6^4 \times 5 = 0.2592$	$0.4^2 \times 0.6^3 \times 10 = 0.3456$	$0.4^3 \times 0.6^2 \times 10 = 0.2304$	$0.4^4 \times 0.6 \times 5 = 0.0768$	$0.4^5 = 0.01024$

b

y	0	1	2	3	4	5
$P(Y = y)$	$0.8^5 = 0.32768$	$0.2 \times 0.8^4 \times 5 = 0.4096$	$0.2^2 \times 0.8^3 \times 10 = 0.2048$	$0.2^3 \times 0.8^2 \times 10 = 0.0512$	$0.2^4 \times 0.8 \times 5 = 0.0064$	$0.2^5 = 0.00032$

c

z	1	2	3	4	5
$P(Z = z)$	0.4	$0.4 \times 0.6 = 0.24$	$0.4 \times 0.6^2 = 0.144$	$0.4 \times 0.6^3 = 0.0864$	$0.4 \times 0.6^4 + 0.6^5 = 0.1296$

13 a

x	2	3	4
$P(X = x)$	$\frac{1}{2}$	$\frac{2}{9}$	$\frac{1}{8}$

$\frac{1}{2} + \frac{2}{9} + \frac{1}{8} = \frac{61}{72}$

The sum of the probabilities is not 1.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

13 b

x	2	3	4
$\mathbf{P}(X = x)$	$\frac{k}{4}$	$\frac{k}{9}$	$\frac{k}{16}$

$$\frac{k}{4} + \frac{k}{9} + \frac{k}{16} = 1$$

$$\frac{61k}{144} = 1$$

$$k = \frac{144}{61} = 2\frac{22}{61}$$

Challenge

x	1	2	3	4	5	6	7	8
$\mathbf{P}(X = x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

y	2	3	6
$\mathbf{P}(Y = y)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

$\mathrm{P}(X > Y) = \mathrm{P}(X > 2 \text{ and } Y = 2) + \mathrm{P}(X > 3 \text{ and } Y = 3) + \mathrm{P}(X > 6 \text{ and } Y = 6)$

$$= \frac{6}{8} \times \frac{1}{2} + \frac{5}{8} \times \frac{1}{3} + \frac{2}{8} \times \frac{1}{6} = \frac{5}{8}$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Statistical distributions 6B

1 a $P(X=2)=\binom{8}{2}\times\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^6$

$=0.273$ (to 3 s.f.)

b $P(X=5)=\binom{8}{5}\times\left(\frac{1}{3}\right)^5\times\left(\frac{2}{3}\right)^3$

$=0.0683$ (to 3 s.f.)

c $P(X\le 1)=P(X=1)+P(X=0)$

$$=8\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^7+\left(\frac{2}{3}\right)^8$$
$$=\left(\frac{2}{3}\right)^7\left(\frac{8}{3}+\frac{2}{3}\right)$$
$$=\left(\frac{2}{3}\right)^7\times\frac{10}{3}$$

$=0.195$ (to 3 s.f.)

2 a $P(T=5)=\binom{15}{5}\times\left(\frac{2}{3}\right)^5\times\left(\frac{1}{3}\right)^{10}=0.00670$ (to 3 s.f.)

b $P(T=10)=\binom{15}{10}\times\left(\frac{2}{3}\right)^{10}\times\left(\frac{1}{3}\right)^5=0.214$ (to 3 s.f.)

c $P(3\le T\le 4)=P(T=3)+P(T=4)=\binom{15}{3}\times\left(\frac{2}{3}\right)^3\times\left(\frac{1}{3}\right)^{12}+\binom{15}{4}\times\left(\frac{2}{3}\right)^4\times\left(\frac{1}{3}\right)^{11}$

$$=0.00025367...+0.00152206...$$
$$=0.00178 \text{ (to 3 s.f.)}$$

3 a $X=$ 'number of defective bolts in a sample of 20'

$X\sim B(20,0.01)$

$n=20$
$p=0.01$

Assume bolts are defective independently of one another.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

3 b $X =$ 'number of times wait or stop in 6 lights'

$X \sim \mathrm{B}(6, 0.52)$

$n = 6$
$p = 0.52$

Assume the lights operate independently and the time lights are on/off is constant.

c $X =$ 'number of aces in Stephanie's next 30 serves'

$X \sim \mathrm{B}(30, \frac{1}{8})$

$n = 30$
$p = \frac{1}{8}$

Assume serving an ace occurs independently and the probability of an ace is constant.

4 a $X =$ 'number of people in class of 14 who are Rh–'

$X \sim \mathrm{B}(14, 0.15)$ is a reasonable model if we assume that being Rh– is independent from pupil to pupil - so no siblings.

b This is not binomial since the number of trials or tosses is not known and fixed. The probability of a head at each toss is constant $(p = 0.5)$ but there is no value for n.

c Assuming, reasonably, that the colours of the cars are independent,

$X =$ 'number of red cars out of 15'

$X \sim \mathrm{B}(15, 0.12)$

5 a Let $X =$ 'number of balloons that do not burst'

$\mathrm{P}(X = 0) = (0.95)^{20}$

$= 0.358$ (to 3 s.f.)

b Let Y = 'number of balloons that do burst'

$$\mathrm{P}(Y = 2) = \binom{20}{2}(0.95)^{18}(0.05)^2$$

$= 0.189$ (to 3 s.f.)

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

6 a There are two possible outcomes of each trial: faulty or not faulty. There are a fixed number of trials, 10, and fixed probability of success: 0.08. Assuming each member in the sample is independent, a suitable model is $X \sim \text{B}(10, 0.08)$.

b $\text{P}(X = 4) = \binom{10}{4}(0.08)^4(0.92)^6 = \frac{10!}{4!6!}(0.08)^4(0.92)^6 = 0.00522$ (to 3 s.f.)

7 a Assumptions are that there is a fixed sample size, that there are only two outcomes for the genetic marker (present or not present), and that there is a fixed probability of people having the marker.

b $X \sim \text{B}(50, 0.04)$

$$\text{P}(X = 6) = \binom{50}{6}(0.04)^6(0.96)^{44} = \frac{50!}{6!44!}(0.04)^6(0.96)^{44} = 0.0108 \text{ (to 3 s.f.)}$$

8 a The random variable can take two values, 6 or not 6. There are a fixed number of trials (15) and a fixed probability of success (0.3). We are assuming that each roll of the dice is independent. A suitable model is $X \sim \text{B}(15, 0.3)$.

b $X \sim \text{B}(15, 0.3)$

$$\text{P}(X = 4) = \binom{15}{4}(0.3)^4(0.7)^{11} = \frac{15!}{4!11!}(0.3)^4(0.7)^{11} = 0.219 \text{ (to 3 s.f.)}$$

c $\text{P}(X \leq 2) = \text{P}(X = 0) + \text{P}(X = 1) + \text{P}(X = 2)$

$$= (0.7)^{15} + \binom{15}{1}(0.3)^1(0.7)^{14} + \binom{15}{2}(0.3)^2(0.7)^{13}$$

$$= (0.7)^{15} + \frac{15!}{1!14!}(0.3)^1(0.7)^{14} + \frac{15!}{2!13!}(0.3)^2(0.7)^{13}$$

$$= 0.127 \text{ (to 3 s.f.)}$$

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Statistical distributions 6C

1 $X \sim \mathrm{B}(9, 0.2)$

a $\mathrm{P}(X \leq 4) = 0.9804$ (tables)

b $\mathrm{P}(X < 3) = \mathrm{P}(X \leq 2) = 0.7382$ (tables)

c $\mathrm{P}(X \geq 2) = 1 - \mathrm{P}(X \leq 1) = 1 - 0.4362 = 0.5638$ (tables)

d $\mathrm{P}(X = 1) = \mathrm{P}(X \leq 1) - \mathrm{P}(X = 0) = 0.4362 - 0.1342 = 0.3020$ (tables)

2 $X \sim \mathrm{B}(20, 0.35)$

a $\mathrm{P}(X \leq 10) = 0.9468$ (tables)

b $\mathrm{P}(X > 6) = 1 - \mathrm{P}(X \leq 6) = 1 - 0.4166 = 0.5834$ (tables)

c $\mathrm{P}(X = 5) = \mathrm{P}(X \leq 5) - \mathrm{P}(X \leq 4) = 0.2454 - 0.1182 = 0.1272$ (tables)

d $\mathrm{P}(2 \leq X \leq 7) = \mathrm{P}(X \leq 7) - (X \leq 1) = 0.6010 - 0.0021 = 0.5989$ (tables)

3 a Using the binomial cumulative function on a calculator where $x = 19$, $n = 40$ and $p = 0.47$,

$\mathrm{P}(X < 20) = \mathrm{P}(X \leq 19) = 0.5888$

b Using the binomial cumulative function on a calculator where $x = 16$, $n = 40$ and $p = 0.47$,

$\mathrm{P}(X > 16) = 1 - \mathrm{P}(X \leq 16) = 0.7662$

c Using the binomial cumulative function on a calculator where $x = 10$ and 15, $n = 40$ and $p = 0.47$,

$\mathrm{P}(11 \leq X \leq 15) = \mathrm{P}(X \leq 15) - \mathrm{P}(X \leq 10) = 0.1478 - 0.0036 = 0.1442$

d Using the binomial cumulative function on a calculator where $x = 10$ and 16, $n = 40$ and $p = 0.47$,

$\mathrm{P}(10 < X < 17) = \mathrm{P}(X \leq 16) - \mathrm{P}(X \leq 10) = 0.2338 - 0.0036 = 0.2302$

4 a Using the binomial cumulative function on a calculator where $x = 20$, $n = 37$ and $p = 0.65$,

$\mathrm{P}(X > 20) = 1 - \mathrm{P}(X \leq 20) = 0.8882$

b Using the binomial cumulative function on a calculator where $x = 26$, $n = 37$ and $p = 0.65$,

$\mathrm{P}(X \leq 26) = 0.7992$

c Using the binomial cumulative function on a calculator where $x = 19$ and 14, $n = 37$ and $p = 0.65$,

$\mathrm{P}(15 \leq X < 20) = \mathrm{P}(X \leq 19) - \mathrm{P}(X \leqslant 14) = 0.06061 - 0.00068 = 0.05993$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

4 d Using the binomial cumulative function on a calculator where $x = 23$ and 22, $n = 37$ and $p = 0.65$,

$P(X = 23) = P(X \leq 23) - P(X \leq 22) = 0.4184 - 0.2926 = 0.1258$

5 $X =$ 'number of heads'
$X \sim B(8, 0.5)$ (coins are fair so $p = 0.5$)

a $P(X = 0) = (0.5)^8 = 0.0039$ (tables)

b $P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.0352 = 0.9648$ (tables)

c $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.6367 = 0.3633$ (tables)

6 $X =$ 'number of plants with blue flowers on tray of 15'

a $P(X = 4) = P(X \leq 4) - P(X \leq 3) = 0.6865 - 0.4613 = 0.2252$ (tables)

b $P(X \leq 3) = 0.4613$ (tables)

c $P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 2) = 0.9434 - 0.2361 = 0.7073$ (tables)

7 $X \sim B(50, 0.40)$

a $P(X \leq 13) = 0.0280$

$P(X \leq 14) = 0.0540$ (tables)

$\therefore k = 13$

b $P(X \leq 27) = 0.9840$

$\Rightarrow P(X > 27) = 0.0160 > 0.01$

$P(X \leq 28) = 0.9924$

$\Rightarrow P(X > 28) = 0.0076 < 0.01$

$\therefore r = 28$

8 $X \sim B(40, 0.10)$

a $P(X = 0) = 0.0148 < 0.02$

$P(X \leq 1) = 0.0805 > 0.02$ (tables)

$P(X < 1) = 0.0148 < 0.02$

$\therefore k = 1$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 b $P(X \leq 8) = 0.9845$ (tables)

$\Rightarrow P(X > 8) = 0.0155 > 0.01$

$P(X \leq 9) = 0.9949$

$\Rightarrow P(X > 9) = 0.0051 < 0.01$

$r = 9$

c $P(k \leq X \leq r) = P(X \leq r) - P(X \leq k-1)$

$= P(X \leq 9) - P(X = 0)$

$= 0.9949 - 0.0148$

$= 0.9801$

9 a A suitable distribution is $X \sim B(10, 0.30)$. Assumptions: There are two possible outcomes of each trial listen or don't listen. There is a fixed number of trials, 10, and fixed probability of success: 0.3. Each member in the sample is assumed to listen independently.

b $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.8497 = 0.1503$ (tables)

c $P(X \leq 6) = 0.9894$ so $P(X \geq 7) = 1 - 0.9894 = 0.0106 > 0.01$

$P(X \leq 7) = 0.9984$ so $P(X \geq 8) = 1 - 0.9984 = 0.0016 < 0.01$ (tables)

Therefore $s = 8$ is the smallest such value.

10 $X =$ number of defects in 50 components

$X \sim B(50, 0.05)$

a $P(X < 2) = P(X \leq 1) = 0.2794$ (tables)

b $P(X > 5) = 1 - P(X \leq 5) = 1 - 0.9622 = 0.0378$ (tables)

c Seek smallest d such that $P(X > d) < 0.05$

$P(X \leq 4) = 0.8964$ so $P(X > 4) = 0.1036 > 0.05$

$P(X \leq 5) = 0.9622$ so $P(X > 5) = 0.0378 < 0.05$

$\therefore d = 5$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Statistical distributions, Mixed Exercise 6

1 a

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

b $P(2 < X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{3}{21} + \frac{4}{21} + \frac{5}{21} = \frac{12}{21} = \frac{4}{7}$

2 a $0.1 + 0.2 + 0.3 + r + 0.1 + 0.1 = 1$

$r = 1 - 0.8$

$= 0.2$

b $P(-1 \leq X < 2) = P(-1) + P(0) + P(1) = 0.2 + 0.3 + 0.2 = 0.7$

3 a

x	1	2	3	4
$P(X = x)$	$\frac{2}{26}$	$\frac{5}{26}$	$\frac{8}{26}$	$\frac{11}{26}$

b $P(2 < X \leq 4) = P(X = 3) + P(X = 4) = \frac{19}{26}$

4 a For a discrete uniform distribution, the probability of choosing each counter must be equal.

b i $P(X = 5) = \frac{1}{16}$

ii The prime numbers are 2, 3, 5, 7, 11 and 13

$P(X \text{ is prime}) = \frac{6}{16} = \frac{3}{8}$

iii $P(3 \leq X < 11) = \frac{8}{16} = \frac{1}{2}$

5 a

y	1	2	3	4	5
$P(Y = y)$	$\frac{1}{k}$	$\frac{2}{k}$	$\frac{3}{k}$	$\frac{4}{k}$	$\frac{5}{k}$

$\frac{1}{k} + \frac{2}{k} + \frac{3}{k} + \frac{4}{k} + \frac{5}{k} = 1$

$\frac{15}{k} = 1, k = 15$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 b

y	1	2	3	4	5
$\mathbf{P}(Y = y)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15} = \frac{1}{5}$	$\frac{4}{15}$	$\frac{5}{15} = \frac{1}{3}$

c $\mathrm{P}(Y > 3) = \mathrm{P}(Y = 4) + \mathrm{P}(Y = 5) = \frac{4}{15} + \frac{5}{15} = \frac{9}{15} = \frac{3}{5}$

6 a

t	0	1	2	3	4
$\mathbf{P}(T = t)$	$\frac{81}{256}$	$\frac{108}{256}$	$\frac{54}{256}$	$\frac{12}{256}$	$\frac{1}{256}$

b $\mathrm{P}(T < 3) = \mathrm{P}(T = 0) + \mathrm{P}(T = 1) + \mathrm{P}(T = 2) = \frac{243}{256}$

c

S	1	2	3	4	5
$\mathbf{P}(S = s)$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{9}{64}$	$\frac{27}{256}$	$\frac{81}{256}$

d $\mathrm{P}(S > 2) = \mathrm{P}(S = 3) + \mathrm{P}(S = 4) + \mathrm{P}(S = 5) = \frac{9}{16}$

7 a $\mathrm{P}(X = 20) = \binom{30}{20}(0.73)^{20}(0.27)^{10} = \frac{30!}{20!10!}(0.73)^{20}(0.27)^{10} = 0.114$ (to 3 s.f.)

b Using the binomial cumulative function on a calculator where $x = 13$, $n = 30$ and $p = 0.73$,

$\mathrm{P}(X \leq 13) = 0.000580$ (to 3 s.f.)

c Using the binomial cumulative function on a calculator where $x = 11$ and 25, $n = 30$ and p = 0.73,

$\mathrm{P}(11 < X \leq 25) = \mathrm{P}(X \leq 25) - \mathrm{P}(X \leq 11) = 0.937302995 - 0.000033512 = 0.937$ (to 3 s.f.)

8 a Sequence is: H H H H H T

Probability: $\left(\frac{2}{3}\right)^5 \times \frac{1}{3} = \frac{32}{729}$ or 0.0439 (to 3 s.f.)

b Let X = 'number of tails in the first 8 tosses', then

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

$$P(X=2)=\binom{8}{2}\left(\frac{1}{3}\right)^2\times\left(\frac{2}{3}\right)^6=0.273 \text{ (to 3 s.f.)}$$

9 $X=$ number of patients waiting more than $\frac{1}{2}$ hour

$X\sim B(12,0.3)$

a $P(X=0)=(0.7)^{12}=0.01384...=0.0138$ (to 3 s.f.)

b $P(X>2)=1-P(X\leq 2)=1-0.2528=0.7472=0.747$ (to 3 d.p.)

10 a **i** There are n independent trials.
ii n is a fixed number.
iii The outcome of each trial is a success or a failure.
iv The probability of success at each trial is constant.
v The outcome of any trial is independent of any other trial.

b $X=$ number of successes

$X\sim B(10,0.05)$

$P(X\geq 2)=1-P(X\leq 1)=1-0.9139=0.0861$ (to 3 s.f.)

c $Y_n=$ 'number of successes in n houses'

$Y_n\sim B(n,0.05)$

Looking for smallest n such that $P(Y_n\geq 1)>0.99$ or, equivalently, $P(Y_n=0)<0.01$.

$P(Y_n=0)=0.95^n<0.01$

So $n=90$ using logarithms.

11 $X=$ 'number of correctly answered questions' and $X\sim B(10,0.5)$

a $P(X=10)=(0.5)^{10}=0.00097656...=0.000977$ (to 3 s.f.)

b $P(X\geq 8)=1-P(X\leq 7)=1-0.9453=0.0547$ (to 3 s.f. using tables)

12

x	1	2	3	4	5	6
$P(X=x)$	p	p	p	p	$2p$	p

$7p=1\Rightarrow p=\frac{1}{7}$

a Sequence: $\overline{5}\ \overline{5}\ \overline{5}\ \overline{5}\ \overline{5}\ 5$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Probability: $\left(\frac{5}{7}\right)^5\left(\frac{2}{7}\right) = 0.0531$ (to 3 s.f.)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

12 b $Y =$ 'number of 5s in 8 throws'

$Y \sim \text{B}(8, \frac{2}{7})$

$$\text{P}(Y = 3) = \binom{8}{3} \times \left(\frac{2}{7}\right)^3 \times \left(\frac{5}{7}\right)^5 = 0.24285 = 0.243 \text{ (to 3 s.f.)}$$

13 $X =$ 'number of green chairs in sample of 10'

a $X \sim \text{B}(10, 0.15)$

b $\text{P}(X \geq 5) = 1 - \text{P}(X \leq 4) = 1 - 0.9901 = 0.0099$ (tables)

c $\text{P}(X = 2) = \text{P}(X \leq 2) - \text{P}(X \leq 1) = 0.8202 - 0.5443 = 0.2759$ (tables)

14 $X =$ 'number of yellow beads in sample of 20' and assume $X \sim \text{B}(20, 0.45)$

a $\text{P}(X < 12) = \text{P}(X \leq 11) = 0.8692$ (tables)

b $\text{P}(X = 12) = \text{P}(X \leq 12) - \text{P}(X \leq 11) = 0.9420 - 0.8692 = 0.0728$ (tables)

15 a $\text{P}(X \geq 10) = 1 - \text{P}(X \leq 9) = 1 - 0.1275 = 0.8725$ (tables)

b $\text{P}(X \geq 10 \text{ in 7 out of 12 sets}) = \binom{12}{7} (0.8725)^7 (0.1275)^5$

$= 0.0103$ (to 3 s.f.)

c Let $Y =$ 'number of sets out of 12 that she hits the bullseye with at least 50% of her arrows', then

$Y \sim \text{B}(12, 0.8725)$

Using the binomial cumulative function on a calculator where $x = 5$, $n = 12$ and $p = 0.8725$,

$\text{P}(Y < 6) = \text{P}(Y \leq 5) = 0.0002407$

Challenge

$Y \sim \text{B}(18, 0.25)$

$\text{P}(Y \geq 11) = 1 - \text{P}(Y \leq 10) = 1 - 0.9988 = 0.0012$ (tables)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Hypothesis testing 7A

1 a A hypothesis is a statement made about the value of a population parameter. A hypothesis test uses a sample or an experiment to determine whether or not to reject the hypothesis.

b The null hypothesis (H_0) is what we assume to be correct and the alternative hypothesis (H_1) tells us about the parameter if our assumption is shown to be wrong.

c The test statistic is used to test the hypothesis. It could be the result of the experiment or statistics from a sample.

2 a one-tailed test

b two-tailed test

c one-tailed test

3 a The test statistic is the number of sixes rolled in the 60 trials.

b $H_0 : p = \frac{1}{6}$

c $H_1 : p > \frac{1}{6}$

4 a Shell is describing what her experiment wants to test rather than the test statistic. The test statistic is the number of times she gets a head in 100 tosses.

b $H_0 : p = \frac{1}{2}$

c $H_1 : p \neq \frac{1}{2}$

5 a A suitable test statistic is the number of faulty articles found in a sample of 100.

b $H_0 : p = 0.1 \qquad H_1 : p < 0.1$

c If the probability of that number being 8 or less is less than 5%, the null hypothesis is rejected.

6 a A suitable test statistic is the number of supporters found in a sample of 20.

b $H_0: p = 0.55 \qquad H_1: p < 0.55$

c If the probability of that number being 7 is 2% or more, the null hypothesis is accepted.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Hypothesis testing 7B

1 a The critical value is the first value to fall inside of the critical region.

b A critical region is a region of the probability distribution which, if the test statistic falls within it, would cause you to reject the null hypothesis.

c The acceptance region is the area in which we accept the null hypothesis.

2 B (10, 0.2)

$$\begin{aligned} P(X \geq 4) &= 1 - P(X \leq 3) \\ &= 1 - 0.8791 \\ &= 0.1209 > 0.05 \end{aligned}$$

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 4) \\ &= 1 - 0.9672 \\ &= 0.0328 < 0.05 \end{aligned}$$

The critical value is $x = 5$ and the critical region is $X \geq 5$ since $P(X \geq 5) = 0.0328 < 0.05$

3 B (20, 0.15)

$$P(X \leq 1) = 0.1756 > 0.05$$
$$P(X = 0) = 0.0388 < 0.05$$

The critical value is $x = 0$ and the critical region is $X = 0$

4 a B (20, 0.4)

$$P(X \leq 4) = 0.0510 > 0.025$$
$$P(X \leq 3) = 0.0160 < 0.025$$

The critical value is $x = 3$

$$P(X \geq 13) = 1 - P(X \leq 12) = 1 - 0.9790 = 0.0210 < 0.025$$
$$P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9435 = 0.0565 > 0.025$$

The critical value is $x = 13$

The critical region is $X \geq 13$ and $X \leq 3$

b The actual significance level is 0.021 + 0.016 = 0.037 = 3.7%

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

5 B (20, 0.18)

$B(X = 0) = 0.0189 < 0.05$
$B(X \leq 1) = 0.1018 > 0.05$

The critical value is $x = 0$
The critical region is $X = 0$

6 a B (10, 0.22)

$P(X < 5) = 0.952$
$P(X \geq 5) = 0.0478 > 0.005$

The critical value is $x = 5$
The critical region is $X \geq 5$

b The actual significance level is 0.0478 = 4.78%

7 a The test statistic is the number of components in the sample that fail.

b H_0: $p = 0.3$
H_1: $p < 0.3$

c Assume H_0 is true then $X \sim B(20, 0.3)$

$P(X \leq 2) = 0.0355$ (closer to 0.05)
$P(X \leq 3) = 0.1071$

The critical region is $X \leq 2$

d 0.0355 = 3.55%

8 a The test statistic is the number of seedlings that survive.

H_0: $p = \frac{1}{3}$,
H_1: $p > \frac{1}{3}$

b Assume H_0 is true then $X \sim B(36, \frac{1}{3})$

Using a calculator

$P(X \geq 16) = 1 - P(X \leq 15) = 1 - 0.8906 = 0.1094 > 0.1$

$P(X \geq 17) = 1 - P(X \leq 16) = 1 - 0.9416 = 0.0584 < 0.1$

The critical region is $X \geq 17$

c 0.0584 = 5.84%

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

9 a In a given time, the number of customers choosing lasagne out of the total number.

$H_0: p = 0.2$
$H_1: p \neq 0.2$

b Assume H_0 is true then $X \sim B(25, 0.2)$

Consider the lower tail:

$P(X \leq 0) = 0.0038$
$P(X \leq 1) = 0.0274$ (closer to 0.025)

Consider the upper tail:

$P(X \geq 9) = 1 - P(X \leq 8) = 1 - 0.9532 = 0.0468$
$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9827 = 0.0173$ (closer to 0.025)

The critical region is $X \leq 1$ and $X \geq 10$.

c The probability of incorrectly rejecting H_o is 0.0274 + 0.0173 = 0.0447 = 4.47%

Challenge

a Assume H_0 is true then $X \sim B(50, 0.7)$

Consider the lower tail:

$P(X \leq 29) = 0.0478$ (closer to 0.05)
$P(X \leq 30) = 0.0848$

Consider the upper tail:

$P(X \geq 41) = 1 - P(X \leq 40) = 1 - 0.9598 = 0.0402$ (closer to 0.05)
$P(X \geq 40) = 1 - P(X \leq 39) = 1 - 0.9211 = 0.0789$

The critical region is $X \leq 29$ and $X \geq 41$

b The probability of one observation falling within the critical region is 0.0478 + 0.0402 = 8.8%

The probability of two observations falling within the critical region is $0.088^2 = 0.007744 = 0.77\%$

The probability that Chloe has incorrectly rejected H_0 is 0.77%

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Hypothesis testing 7C

1 Distribution, B(10, 0.25)

$H_0 : p = 0.25 \qquad H_1 : p > 0.25$

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 4) \\ &= 1 - 0.9219 \\ &= 0.0781 > 0.05 \end{aligned}$$

There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.25$

2 Distribution, B(10, 0.40)

$H_0 : p = 0.40 \qquad H_1 : p < 0.40$

$P(X \leq 1) = 0.0464 < 0.05$

There is sufficient evidence to reject H_0 so $p < 0.04$

3 Distribution, B(20, 0.30)

$H_0 : p = 0.30 \qquad H_1 : p > 0.30$

$$\begin{aligned} P(X \geq 10) &= 1 - P(X \leq 9) \\ &= 1 - 0.9520 \\ &= 0.0480 < 0.05 \end{aligned}$$

There is sufficient evidence to reject H_0 so $p > 0.3$

4 Distribution, B(20, 0.45)

$H_0 : p = 0.45 \qquad H_1 : p < 0.45$

$P(X \leq 3) = 0.0049 < 0.01$

There is sufficient evidence to reject H_0 so $p < 0.45$

5 Distribution, B(20, 0.28)

$H_0 : p = 0.28 \qquad H_1 : p < 0.28$

p-value $= P(X \leq 2) = 0.0526 > 0.05$

There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.28$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 Distribution, B(8, 0.32)

$H_0 : p = 0.32 \qquad H_1 : p < 0.32$

$$\begin{aligned} P(X \geq 7) &= 1 - P(X \leq 6) \\ &= 1 - 0.9980 \\ &= 0.002 < 0.05 \end{aligned}$$

There is sufficient evidence to reject H_0 so $p < 0.32$

7 Distribution, B(12, $\frac{1}{6}$)

$H_0 : p = \frac{1}{6} \qquad H_1 : p < \frac{1}{6}$

$P(X \leq 1) = 0.3813 > 0.05$

There is insufficient evidence to reject H_0 so there is no evidence that the probability of a 6 on this dice is less than $\frac{1}{6}$

8 a Distribution, B(n, 0.68)

Reasons:
- Fixed number of trials.
- Outcomes of the trials are independent.
- There are two outcomes, success and failure.
- The probability of success is constant.

b Distribution, B(10, 0.68)

$H_0 : p = 0.68 \qquad H_1 : p < 0.68$

$P(X \leq 3) = 0.0155 < 0.05$

There is sufficient evidence to reject H_0 so $p < 0.68$. The treatment is not as effective as claimed.

9 a X is the number of seeds in the trial for which the germination method was successful.
p is the probability of success for each seed.

$X \sim B(20, p)$

$H_0: p = 0.4 \qquad H_1: p > 0.4$

$P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9435 = 0.0565 \geq 0.05$
$P(X \geq 13) = 1 - P(X \leq 12) = 1 - 0.9790 = 0.021 \leq 0.05$

The critical region is $X \geq 13$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

9 b 14 lies within the critical region, so we can reject the null hypothesis. There is evidence that the new technique has improved the number of plants that germinate.

10 a The test statistic is the number of people who support the candidate.

$H_0: p = 0.35$ $H_1: p > 0.35$

b $X \sim B(50, 0.35)$

$P(X \geq 23) = 1 - P(X \leq 22) = 1 - 0.9290 = 0.071 > 0.05$
$P(X \geq 24) = 1 - P(X \leq 23) = 1 - 0.9604 = 0.0396 < 0.05$

The critical region is $X \geq 24$

c 28 lies in the critical region, so we can reject the null hypothesis. There is evidence that the candidate's level of popularity has increased.

Hypothesis testing 7D

1 $H_0 : p = 0.5$ $H_1 : p \neq 0.5$

If H_0 is true $X \sim B(30, 0.5)$

Expected value would be $30 \times 0.5 = 15$.
The observed value, 10, is less than this so consider $P(X \leq 10)$

$P(X \leq 10) = 0.0494 > 0.025$ (two-tailed)

There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.5$

2 $H_0: p = 0.3$ $H_1: p \neq 0.3$

If H_0 is true $X \sim B(25, 0.3)$

Expected value would be $25 \times 0.3 = 7.5$.
The observed value, 10, is more than this so consider $P(X \geq 10)$

$P(X \geq 10) = 0.1894... > 0.05$ (two-tailed)

There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.3$

3 $H_0: p = 0.75$ $H_1: p \neq 0.75$

If H_0 is true $X \sim B(10, 0.75)$

Expected value would be $10 \times 0.75 = 7.5$.
The observed value, 9, is more than this so consider $P(X \geq 9)$

$P(X \geq 9) = 0.2440... > 0.025$ (two-tailed)

There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.75$

4 $H_0: p = 0.6$ $H_1: p \neq 0.6$

If H_0 is true $X \sim B(20, 0.6)$

Expected value would be $20 \times 0.6 = 12$
The observed value, 1, is less than this so consider $P(X \leq 1)$

$P(X \leq 1) = 0.00000034.. < 0.005$ (two-tailed)

Reject H_0. There is evidence that $p \neq 0.6$.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

5 $H_0: p = 0.02$ $H_1: p \neq 0.02$

If H_0 is true $X \sim B(50, 0.02)$

Expected value would be $50 \times 0.02 = 1$
The observed value, 4, is more than this so consider $P(X \geq 4)$

$P(X \geq 4) = 0.01775... > 0.01$ (two-tailed)

There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.02$

6 The probability that an unbiased coin lands on heads is 0.5
X is the number of times the coin being tested lands on heads
p is the probability that the coin being tested lands on heads.

$H_0 : p = 0.5$ $H_1 : p \neq 0.5$

If H_0 is true $X \sim B(20, 0.5)$

Expected value would be $20 \times 0.5 = 10$
The observed value, 6, is less than this so consider $P(X \leq 6)$

$P(X \leq 6) = 0.0577 > 0.025$ (two-tailed)

There is insufficient evidence to reject H_0 so there is no reason to think that the coin is biased.

7 a $H_0 : p = 0.20$ $H_1 : p \neq 0.20$

If H_0 is true $X \sim B(20, 0.20)$

$P(X \leq 1) = 0.0692$
$P(X \leq 0) = 0.0115$ (closer to 0.025)

critical value $= 0$

$P(X \geq 9) = 1 - P(X \leq 8) = 1 - 0.9900 = 0.0100$
$P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.9679 = 0.0321$ (closer to 0.025)

Critical region $X = 0$ and $X \geq 8$

b Actual significance level is $0.0115 + 0.0321 = 0.0436 = 4.36\%$

c $X = 8$ is in the critical region. There is enough evidence to reject H_0. The hospital's proportion of complications differs from the national figure.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 a The probability that a glass bowl made using the original process is cracked is 0.1
X is the number of bowls in the sample using the new process that are cracked.
p is the probability that a bowl made using the new process is cracked.

$H_0 : p = 0.1$ $H_1 : p \neq 0.1$

If H_0 is true $X \sim B(20, 0.1)$

Expected value would be $20 \times 0.1 = 2$
The observed value, 1, is less than this so consider $P(X \leq 1)$

$P(X \leq 2) = 0.3917... > 0.05$ (two-tailed)

There is insufficient evidence to reject H_0 so there is no reason to think that the proportion of cracked bowls has changed.

b Double the calculated probability to find the p-value
p- value = 0.3917…+ 0.3917… = 0.7835

9 The probability that a carrot grown in the original fertiliser is longer than 7 cm is 0.25
X is the number of carrots in the sample grown in the new fertiliser that are longer than 7 cm.
p is the probability that a carrot grown in the new fertiliser is longer than 7 cm.

$H_0 : p = 0.25$ $H_1 : p \neq 0.25$

If H_0 is true $X \sim B(30, 0.25)$

Expected value would be $30 \times 0.25 = 7.5$
The observed value, 13, is more than this so consider $P(X \geq 13)$

$P(X \geq 13) = 0.02159... < 0.025$ (two-tailed)

There is evidence to reject H_0. Therefore, there is reason to doubt $p = 0.25$.
So the probability of a carrot being longer than 7 cm has changed.

10 The probability that a standard blood test diagnoses the disease is 0.96
X is the number of patients correctly diagnosed in the sample using the new process.
p is the probability that a patient is correctly diagnosed using the new process.

$H_0 : p = 0.96$ $H_1 : p \neq 0.96$

If H_0 is true $X \sim B(75, 0.96)$

Expected value would be $75 \times 0.96 = 72$
The observed value, 63, is less than this so consider $P(X \leq 63)$

$P(X \leq 63) = 0.0000417... < 0.05$ (two-tailed)

There is evidence to reject H_0. Therefore, there is reason to doubt $p = 0.96$.
So the new test does not have the same probability of success as the old test.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

Hypothesis testing, Mixed Exercise 7

1 $X \sim \mathrm{B}(10, 0.20)$

$\mathrm{H}_0 : p = 0.20 \qquad \mathrm{H}_1 : p > 0.20$

$$\begin{aligned}\mathrm{P}(X \geq 3) &= 1 - \mathrm{P}(X \leq 2) \\ &= 1 - 0.6778 \\ &= 0.3222 > 0.05\end{aligned}$$

There is insufficient evidence to reject H_0

There is no evidence that the trains are late more often.

2 $X \sim \mathrm{B}(5, 0.5)$

$\mathrm{H}_0 : p = 0.50 \qquad \mathrm{H}_1 : p > 0.50$

$$\begin{aligned}\mathrm{P}(X \geq 4) &= 1 - \mathrm{P}(X \leq 3) \\ &= 1 - 0.8125 \\ &= 0.1875 > 0.05\end{aligned}$$

There is insufficient evidence to reject H_0

There is insufficient evidence that the company's claims are true.

3 a Fixed number; independent trials; two outcomes (pass or fail); p constant for each car.

b $X \sim \mathrm{B}(5, 0.30)$

$\mathrm{P}(\text{all pass}) = 0.70^5 = 0.16807$

c $X \sim \mathrm{B}(10, 0.30)$

$\mathrm{H}_0 : p = 0.30 \qquad \mathrm{H}_1 : p < 0.30$

$\mathrm{P}(X \leq 2) = 0.3828 > 0.05$

There is insufficient evidence to reject H_0.

There is no evidence that the garage fails fewer than the national average.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

4 a $X \sim \mathrm{B}(50, 0.1)$

$\mathrm{H}_0 : p = 0.10 \qquad \mathrm{H}_1 : p \neq 0.10$

$\mathrm{P}(X \leq 1) = 0.0338$ (closer to 0.025)
$\mathrm{P}(X = 0) = 0.0052$
critical value = 1

$\mathrm{P}(X \geq 9) = 1 - \mathrm{P}(X \leq 8) = 1 - 0.9421 = 0.0579$
$\mathrm{P}(X \geq 10) = 1 - \mathrm{P}(X \leq 9) = 1 - 0.9755 = 0.0245$ (closer to 0.025)

critical value = 10

Critical region $X \leq 1$ and $X \geq 10$

b Actual significance level = 0.0338 + 0.0245 = 0.0583 = 5.83%

c $X \sim \mathrm{B}(20, 0.1)$

$\mathrm{H}_0 : p = 0.1 \qquad \mathrm{H}_1 : p > 0.1$

$$\begin{aligned}\mathrm{P}(X \geq 4) &= 1 - \mathrm{P}(X \leq 3) \\ &= 1 - 0.8670 \\ &= 0.133 > 0.1\end{aligned}$$

p-value = 0.133

Accept H_0. There is no evidence that the proportion of defective articles has increased.

5 $X \sim \mathrm{B}(20, 0.5)$
$\mathrm{H}_0 : p = 0.50 \qquad \mathrm{H}_1 : p \neq 0.50$

8 used Oriels powder.

$\mathrm{P}(X \leq 8) = 0.2517 > 0.025$

There is insufficient evidence to reject H_0.
There is no evidence that the claim is wrong.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

6 $X \sim \mathrm{B}(50, 0.2)$

a $P(X \leq 4) = 0.0185$ (closer to 0.025)
$\mathrm{P}(X \leq 5) = 0.0480$

$c_1 = 4$

$\mathrm{P}(X \geq 16) = 1 - \mathrm{P}(X \leq 15) = 1 - 0.9692 = 0.0308$ (closer to 0.025)
$\mathrm{P}(X \geq 17) = 1 - \mathrm{P}(X \leq 16) = 1 - 0.9856 = 0.0144$

$c_2 = 16$

Critical region $X \leq 4$ and $X \geq 16$

b Actual significance level = 0.0185 + 0.0308 = 0.0493 = 4.93%

c This is not in the critical region. Therefore, there is insufficient evidence to reject H_0. There is no evidence to suggest that the proportion of people buying that certain make of computer differs from 0.2.

7 a i A hypothesis is a statement made about the value of a population parameter. A hypothesis test uses a sample or an experiment to determine whether or not to reject the hypothesis.

ii The critical value is the first value to fall inside of the critical region.

iii The acceptance region is the region where we accept the null hypothesis.

b $H_0: p = 0.2$ $H_1: p \neq 0.2$

If H_0 is true $X \sim \mathrm{B}(20, 0.2)$

Let c_1 and c_2 be the two critical values so $\mathrm{P}(X \leq c_1) \leq 0.05$ and $\mathrm{P}(X \geq c_2) \leq 0.05$

For the lower tail:

$\mathrm{P}(X = 0) = 0.0115 < 0.05$
$\mathrm{P}(X \leq 1) = 0.0692 > 0.05$

So $c_1 = 0$

For the upper tail:

$\mathrm{P}(X \geq 7) = 1 - \mathrm{P}(X \leq 6) = 1 - 0.9133 = 0.0978 > 0.05$
$\mathrm{P}(X \geq 8) = 1 - \mathrm{P}(X \leq 7) = 1 - 0.9679 = 0.0321 < 0.05$

So $c_2 = 8$

So the critical region is $X = 0$ and $X \geq 8$

c Actual significance level = 0.0115 + 0.0321 = 0.0436 = 4.36%

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

7 d As 7 does not lie in the critical region, H_0 is not rejected. Therefore, the proportion of times that Johan is late for school has not changed.

8 X is the number of days with zero or a trace of rain.

$X \sim B(30, 0.5)$

$H_0: p = 0.5$ $H_1: p > 0.5$

$P(X \geq 19) = 1 - P(X \leq 18) = 1 - 0.8998 = 0.1002 > 0.05$
$P(X \geq 20) = 1 - P(X \leq 19) = 1 - 0.9506 = 0.0494 < 0.05$

The critical region is $X \geq 20$

21 lies in the critical region, so we can reject the null hypothesis. There is evidence that the likelihood of a rain-free day in 2015 has increased.

9 a $H_0: p = 0.35$ $H_1: p \neq 0.35$

If H_0 is true $X \sim B(30, 0.35)$

Let c_1 and c_2 be the two critical values so $P(X \leq c_1) \leq 0.025$ and $P(X \geq c_2) \leq 0.025$

For the lower tail:

$P(X \leq 5) = 0.0233 < 0.025$
$P(X \leq 6) = 0.0586 > 0.025$

So $c_1 = 5$

For the upper tail:

$P(X \geq 16) = 1 - P(X \leq 15) = 1 - 0.9699 = 0.0301$, $0.0301 - 0.025 = 0.0051$
$P(X \geq 17) = 1 - P(X \leq 16) = 1 - 0.9876 = 0.0124$, $0.025 - 0.0124 = 0.0126$

So $c_2 = 16$

So the critical region is $X \leq 5$ and $X \geq 16$

b Actual significance test is $0.0233 + 0.0301 = 0.0534 = 5.34\%$

c $X = 4$ lies in the critical region so there is enough evidence to reject H_0.

10 a $X \sim B(20, 0.85)$

b $P(X = 16) = \binom{20}{16} 0.85^{16} 0.15^4 = 0.18$

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

10 c X is the number of patients that recover
p is the probability that a patient recovers

$H_0 : p = 0.85$ $H_1 : p < 0.85$

If H_0 is true $X \sim B(30, 0.85)$

Expected value would be $30 \times 0.85 = 25.5$
The observed value, 20, is less than this so consider $P(X \leq 20)$

p-value = $P(X \leq 20) = 0.009657... < 0.05$ (one-tailed)

There is evidence to reject H_0. The percentage of patients who recover after treatment with the new ointment is lower than 85%.

Large Data Set

1 a X is the number of days with a recorded daily mean temperature greater than 15 °C.

$X \sim (10, 0.163)$

H_0: $p = 0.163$ H_1: $p > 0.163$

$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.935 = 0.065 > 0.05$
$P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.9959 = 0.0141 < 0.05$

The critical region is $X \geq 5$

b A random sample of temperatures, 15.6, 17.3, 12.5, 14.1, 11.9, 14.5, 13.0, 9.1, 14.1, 10.0, 9.3 has 2 days with a daily mean temperature greater than 15 °C.

c 2 does not lie in the critical region so H_0 is not rejected. Therefore, there is no reason to suggest that $p \neq 0.163$.

2 X is the number of days with daily mean temperature greater than 25 °C.

$X \sim (10, 0.23)$

H_0: $p = 0.23$ H_1: $p > 0.23$

$P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.9431 = 0.0569 > 0.05$
$P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9870 = 0.0130 < 0.05$

The critical region is $X \geq 6$

A random sample of temperatures 17.5, 18.9, 25.9, 27.7, 30.4, 26.6, 27.4, 27.0, 19.4 and 13.9 has 6 days with a mean temperature greater than 25 °C.

6 does lie in the critical region so H_0 is rejected. Therefore, there is reason to suggest that $p > 0.23$.

© Pearson Education Ltd 2019. Copying permitted for purchasing institution only. This material is not copyright free.

Review Exercise 1

1 a A census observes every member of a population.
A disadvantage of a census is it would be time-consuming to get opinions from all the employees.
OR It would be difficult/time-consuming to process the large amount of data from a census.

b Opportunity sampling.

c It is not a random sample. The sample only includes cleaners, there are no types of other employees such as managers.
The first 50 cleaners to leave may be in the same group/shift so may share the same views.

d i Allocate a number from 1–550 to all employees.
For a sample of 50, you need every eleventh person since 550 ÷ 50 = 11.
Select the first employee using a random number from 1 - 11, then select every eleventh person from the list; e.g. if person 8 is then the sample is 8, 19, 30, 41...

ii For this sample, you need $\frac{55}{550}\times 50 = 5$ managers and $\frac{495}{550}\times 50 = 45$ cleaners.
Label the managers 1–55 and the cleaners 1–495.
Use random numbers to select 5 managers and 45 cleaners.

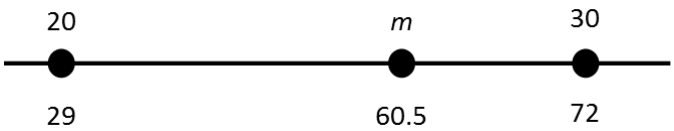
2 a Opportunity sampling is using a sample that is available at the time the study is carried out.
It is unlikely to provide a representative sample of the weather in May.

b $\text{mean} = \frac{\sum h}{n} = \frac{100+91+77+83+86}{5} = \frac{437}{5} = 87.4$

c The daily maximum relative humidity is how saturated the air saturation is with water vapour. Mist and fog occur only when the relative humidity is above 95%. Since 87.4% is less than this, Joanna is correct. 'However 5 days is likely not to be a representative sample for the whole of May.

3 a There are 120 observations, so the median is the 60.5th value. This lies in the class $20 < x \leq 30$.

Using interpolation:



$$\frac{m-20}{30-20} = \frac{60.5-29}{72-29}$$
$$\frac{m-20}{10} = \frac{31.5}{43}$$
$$m = 27.3$$

The median distance is 27.3 miles.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

3 b $\text{mean} = \dfrac{\sum fx}{\sum f} = \dfrac{3610}{120} = 30.1$

$$\sigma = \sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2}$$
$$= \sqrt{\frac{141\,600}{120} - \left(\frac{3610}{120}\right)^2}$$
$$= \sqrt{1180 - 30.1^2}$$
$$= \sqrt{275} = 16.6$$

The mean distance travelled is 30.1 miles, with a standard deviation of 16.6 miles.

4 $x = 10s + 1$

$$s = \frac{x-1}{10}$$

coded mean, $\bar{x} = \dfrac{\sum x}{n} = \dfrac{947}{30} = 31.6$

actual mean, $\bar{s} = \dfrac{31.6-1}{10} = 3.06$ hours

coded standard deviation, $\sigma_x = \sqrt{\dfrac{33\,065.37}{30}} = 33.2$

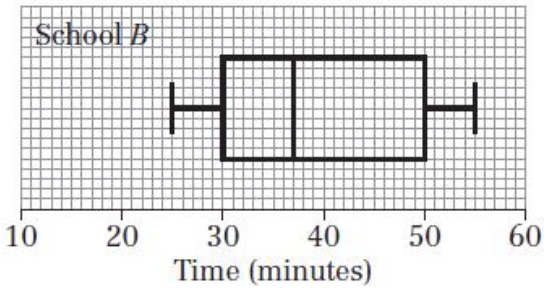
actual standard deviation, $\sigma_s = \dfrac{33.2}{10} = 3.32$ hours

5 a i 37 minutes

ii Upper quartile / third quartile / 75th percentile

b The crosses represent outliers; i.e. observations that are very different from the others and need to be treated with caution. These two children took a lot longer so probably walked.

c Defining an outlier as an observation that lies either 1.5 × interquartile range (IQR) above the upper quartile (Q_3) or 1.5 × IQR below the lower quartile (Q_1):
IQR = 20, $Q_1 = 30$, $Q_3 = 50$
$30 - 1.5 \times 20 = 0$ therefore no outliers
$50 + 1.5 \times 20 = 80$ therefore no outliers



d The median time for children from school A is less than that of children from school B.
The interquartile range is less for school A, although the overall range is greater and there are also outliers (there are none for school B).

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

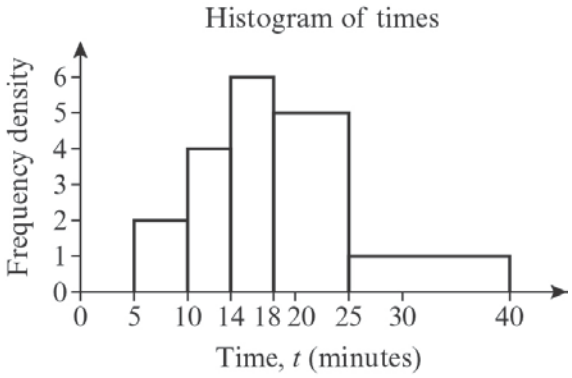
5 d Children from school A generally took less time than those from school B.

Both plots have a positive skew, showing that there was more variation in the time taken by slower children than in the time taken by the faster children.

75% of those from A took less than 37 minutes whereas only 50% of those from B completed in this time.

50% of those from A took less than 30 minutes whereas only 25% of those from B completed in this time.

6 a Missing frequencies:
$18 \le x < 25$: area $= 7 \times 5 = 35$
$25 \le x < 40$: area $= 15 \times 1 = 15$
Missing frequency densities:
$10 \le x < 14$: height $= 16 \div 4 = 4$
$14 \le x < 18$: height $= 24 \div 4 = 6$



b Sample size, $n = 100$
The number who took longer than 20 minutes is given by the shaded area:
$(25-20)\times 5 + (40-25)\times 1 = 40$
$40 \div 100 = 0.4$
The probability that a person chosen at random took more than 20 minutes to swim 500 m is 0.4

c Mid points are 7.5, 12, 16, 21.5, 32.5

$$\sum f = 100$$

$$\sum ft = (10\times 7.5) + (16\times 12) + (24\times 16) + (35\times 21.5) + (15\times 32.5)$$
$$= 1891$$

$$\frac{\sum ft}{\sum f} = \frac{1891}{100} = 18.91$$

The mean time taken is 18.9 minutes.

d $\sum ft^2 = (10\times 7.5^2) + (16\times 12^2) + (24\times 16^2) + (35\times 21.5^2) + (15\times 32.5^2) = 41033$

$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$= \sqrt{\frac{41\,033}{100} - \left(\frac{1891}{100}\right)^2}$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

6 d $\sigma = \sqrt{410.33 - 357.59}$
$= \sqrt{52.74}$
$= 7.26$

The standard deviation of t is approximately 7.26 minutes (3 s.f.)

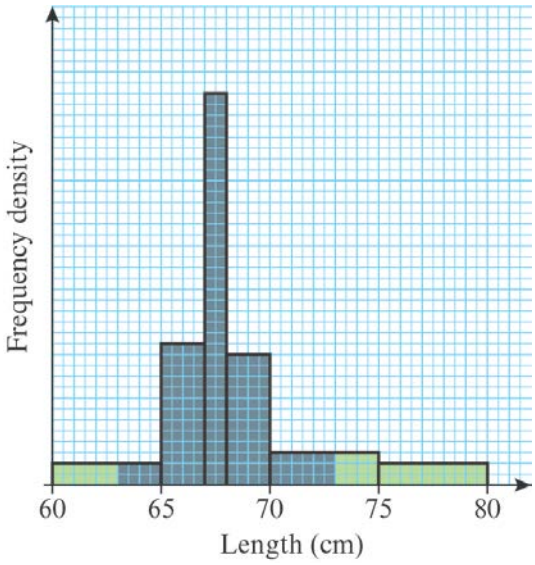
e median, Q_2 is the $\frac{100+1}{2} = 50.5$th term
There are 50 values less than 18 minutes and 50 values greater than or equal to 18 minutes.
The median is therefore 18 minutes.

7 Area of 65 to 67 cm class = 26
Frequency density = 26 ÷ 2 = 13

Using this information:

Length, l (cm)	Frequency	Class width	Frequency density
60 to 65	10	5	2
65 to 67	26	2	13
67	36	1	36
68 to 70	24	2	12
70 to 75	15	5	3
75 to 80	10	5	2

The number of owls with wing length between 63 and 73 cm is given by the shaded area on the graph.



$$P(63 \le l \le 73) = \frac{(2\times 2) + 26 + 36 + 24 + (3\times 3)}{10 + 26 + 36 + 24 + 15 + 10}$$
$$= \frac{99}{121}$$
$$= 0.82$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

8 a 20th percentile : $\frac{20}{100} \times 31 = 6.2$

$$\frac{P_{20} - 13}{16 - 13} = \frac{6.2 - 1}{8 - 1}$$

$$P_{20} - 13 = \frac{5.2 \times 3}{7}$$

$$P_{20} = 2.23 + 13 = 15.23$$

80th percentile : $\frac{80}{100} \times 31 = 24.8$

$$\frac{P_{80} - 19}{22 - 19} = \frac{24.8 - 21}{31 - 21}$$

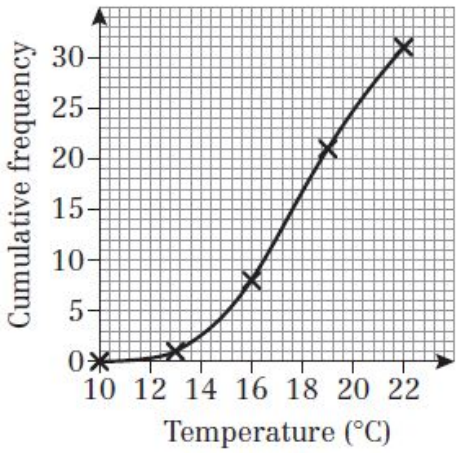
$$P_{80} - 19 = \frac{3.8 \times 3}{10}$$

$$P_{80} = 1.14 + 19$$

$$= 20.14$$

$P_{80} - P_{20} = 20.14 - 15.23 = 4.9$ (2 s.f.)

b Cumulative frequencies are 1, 8, 21, 31



c $P_{80} - P_{20} = 19.8 - 15.4 = 4.8$
The 20% to 80% interpercentile range is slightly bigger using interpolation than using the diagram. Interpolation is likely to be more accurate.

d Using the graph:
$31 - 5 = 26$ days

OR Using interpolation:

$$\frac{15 - 13}{16 - 13} = \frac{d - 1}{8 - 1}$$

$$\frac{2}{3} = \frac{d - 1}{7}$$

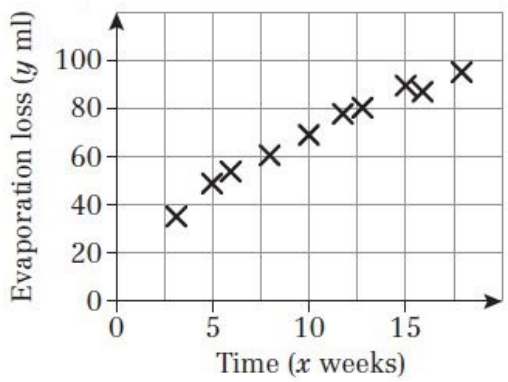
$$d = \frac{2 \times 7}{3} + 1$$

$$= 5.67$$

$31 - 5.67 = 25.33$, so 25 days

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

9 a



b Points lie close to a straight line.

c For every extra week in storage, another 3.90 ml of chemical evaporates.

d The prediction for 19 weeks is likely to be reasonably reliable as it is close to the range investigated.
The prediction for 35 weeks is likely to be unreliable, since the time is well outside range of x and there is no evidence that model will continue to hold.

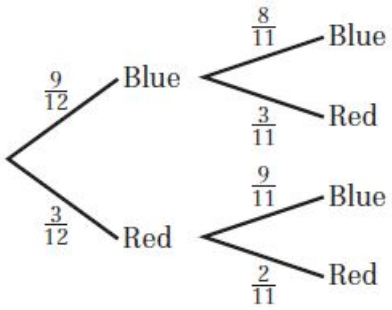
10 a mean + 2 standard deviations = $15.3 + 2 \times 10.2 = 35.7$
$45 > 35.7$ so $t = 45$ is an outlier

b A temperature of 45 °C is very high so it is likely this value was recorded incorrectly. Therefore, this outlier should be omitted from the data.

c In the regression equation, 2.81 represents the number of additional ice creams (in hundreds) sold each month for each degree Celsius increase in average temperature.

d A temperature of 2 °C is outside the range of the data so a value calculated using the equation for the regression line involves extrapolation and is likely to be inaccurate.

11 a



b P (second ball red) = P(blue then red) + P(red then red)

$$= \frac{9}{12} \times \frac{3}{11} + \frac{3}{12} \times \frac{2}{11}$$

$$= \frac{27+6}{132} = \frac{1}{4}$$

The probablility the second ball is red is 0.25.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

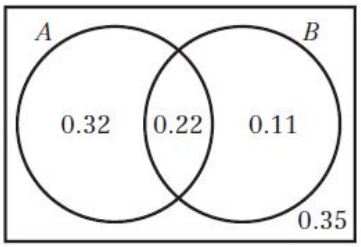
11 c P (balls are different colours) = P(blue then red) + P(red then blue)

$$= \frac{9}{12} \times \frac{3}{11} + \frac{3}{12} \times \frac{9}{11}$$

$$= \frac{27+27}{132} = \frac{54}{132}$$

The probability the balls are different colours is 0.409.

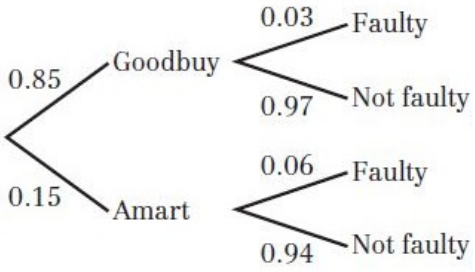
12 a P(A or B) = P(A) + P(B) − P(A and B)
0.65 = 0.32 + 0.11 − P(A and B)
P(A and B) = 0.65 − 0.32 − 0.11= 0.22
P(neither A nor B) = 1 – 0.65 = 0.35



b $P(A) = 0.32 + 0.22 = 0.54$
$P(B) = 0.33$

c For independence P(A and B) = P(A) × P(B)
Here: P(A) × P(B) = 0.54 × 0.33 = 0.1782
$0.1782 \neq 0.22$
So these events are not independent.

13 a



b G = Goodbuy, A = Amart, NF = Not faulty
P(NF) = P(G and NF) + P(A and NF)

$$= (0.85 \times 0.97) + (0.15 \times 0.94)$$

$$= 0.9655$$

14 a Comics and Television are mutually exclusive preferences as the sets do not overlap.

b P(C and B) = $\frac{13}{38} = 0.34$

P(C) × P(B) = $\frac{21}{38} \times \frac{11}{19} = \frac{231}{722} = 0.32$

$0.34 \neq 0.32$ so these preferences are not independent.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

15 a For four coin tosses there are $2 \times 2 \times 2 \times 2 = 2^4 = 16$ possible outcomes.
There is only way of achieving 0 or 4 heads (= 0 tails).
1 head or 3 heads (= 1 tail) may appear on the 1st, 2nd, 3rd or 4th toss; i.e. in one of 4 ways.

Therefore, if X is the number of heads in the outcome:

No. of heads, x	0	1	2	3	4
P($X = x$)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{n}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$\frac{1}{16}+\frac{4}{16}+\frac{n}{16}+\frac{4}{16}+\frac{1}{16}=1$$
$$10+n=1$$
$$n=6$$

The probability of an equal number of heads and tails is $\frac{6}{16}=\frac{3}{8}=0.375$

Alternative solution:
Let X be the random variable 'the number of heads'.
$X \sim \text{B}(4, 0.5)$

$$\text{P}(X=2)=\binom{4}{2}0.5^2\times0.5^2$$
$$=\frac{4!}{2!2!}0.5^2\times0.5^2$$
$$=0.375$$

b $\text{P}(X = 0 \text{ or } 4) = \text{P}(X=0)+\text{P}(X=4)$
$$=\frac{1}{16}+\frac{1}{16}=\frac{1}{8}$$
$$=0.125$$

c $$\text{P}(HHT)=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}$$
$$=0.125$$

16 a Let X be the random variable 'the number of faulty bolts'.
$X \sim \text{B}(20, 0.3)$

$$\text{P}(X=2)=\binom{20}{2}\times(0.3)^2(0.7)^{18}$$
$$=\frac{20!}{2!18!}(0.3)^2(0.7)^{18}$$
$$=0.0278$$

b $\text{P}(X>3)=1-\text{P}(X\le3)$
From calculator or tables, $\text{P}(X \le 3) = 0.1071$
$$\text{P}(X>3)=1-0.1071$$
$$=0.8929$$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

16 c Let Y be the random variable 'number of bags containing more than 3 faulty bolts'.
$Y \sim B(10,0.8929)$

$$P(Y=6) = \binom{10}{6} \times (0.8929)^6 (0.1071)^4$$
$$= \frac{10!}{6!4!}(0.8929)^6(0.1071)^4$$
$$= 0.0140$$

17a

x	1	2	3	4	5	6
P ($X = x$)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$
P ($X = x$)	0.0278	0.0833	0.1389	0.1944	0.25	0.3056

b $P(2 < X \leq 5) = P(Y=3) + P(Y=4) + P(Y=5)$
$P(Y=3) + P(Y=4) + P(Y=5) = \frac{21}{36}$
$= \frac{7}{12}$
$= 0.5833...$

18 a $P(X = x) = 0.2$

b The spinner has 3 odd numbers and two even numbers so:
$P(X = \text{even}) = 0.4$
$P(X = \text{odd}) = 0.6$

y	1	2	3	4
P($Y = y$)	0.6	$0.4 \times 0.6 = 0.24$	$0.4^2 \times 0.6 = 0.096$	$0.4^3 \times 0.6 + 0.4^4 = 0.064$

c $P(Y > 2) = P(Y = 3) + P(Y = 4) = 0.096 + 0.064 = 0.16$

19 For $X \sim B(15, 0.32)$, using calculator or tables:

a $P(X = 7) = 0.101$ (3 s.f.)

b $P(X \leq 4) = 0.448$ (3 s.f.)

c $P(X < 8) = P(X \leq 7) = 0.929$ (3 s.f.)

d $P(X > 6) = 1 - P(X \leq 5) = 1 - 0.6607 = 0.339$ (3 s.f.)

20 a $H_0: p = 0.3$ $H_1: p \neq 0.3$
Significance level 2.5%
If H_0 is true $X \sim B(40, 0.3)$
Let c_1 and c_2 be the two critical values.
$P(X \leq c_1) = 0.0125$ and $P(X \geq c_2) = 0.0125$
For the lower tail:
Finding values of $P(X)$ from tables or calculator, those either side of 0.0125 are:
$P(X \leq 5) = 0.0086$
$P(X \leq 6) = 0.0238$
$0.0125 - 0.0086 = 0.0039$ ($X \leq 5$) and $0.0238 - 0.0125 = 0.0113$ ($X \leq 6$)
So $c_1 = 5$ as this gives the value closest to 0.0125.

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

20 a In the same way, for the upper tail:
$P(X \geq 19) = 1 - P(X \leq 18) = 1 - 0.9852 = 0.0148$
$P(X \geq 20) = 1 - P(X \leq 19) = 1 - 0.9937 = 0.0063$
$0.0148 - 0.0125 = 0.0023$ $(X \geq 19)$ and
$0.0125 - 0.0063 = 0.0062$ $(X \geq 20)$
So $c_2 = 19$
$P(X \leq 5)$ and $P(X \geq 19)$, so the critical region is $0 \leq X \leq 5$ and $19 \leq X \leq 40$

b The probability of incorrectly rejecting the null hypothesis is the same as the probability that X falls within the critical region.
$P(X \leq 5) + P(X \geq 19) = 0.0086 + 0.0148 = 0.0234$

21 a $X \leq B(10, 0.75)$ where X is the random variable 'number of patients who recover when treated'.

b Using tables or calculator:
$P(X = 6) = 0.146$
OR

$$\begin{aligned} P(X = 6) &= P(X \leq 6) - P(X \leq 5) \\ &= 0.9219 - 0.7759 \\ &= 0.146 \end{aligned}$$

OR, Calculating:

$$\begin{aligned} P(X = 6) &= \binom{10}{6} \times (0.75)^6 (0.25)^4 \\ &= \frac{10!}{6!4!} \times (0.75)^6 (0.25)^4 \\ &= 0.146 \end{aligned}$$

c $H_0 : p = 0.75$
$H_1 : p < 0.75$
$X \sim B(20, 0.75)$

$$\begin{aligned} P(X \leq 13) &= 1 - 0.7858 \\ &= 0.2142 \end{aligned}$$

As the probability is greater than 5%, there is insufficient evidence to reject the null hypothesis that 75% of patients will recover. Therefore the doctor's belief that fewer patients than this will recover.

d Using tables/calculator to find values either side of $1 - 0.01 = 0.99$:
$P(X \leq 9) = 1 - 0.9961 = 0.0039 < 0.01$
i.e. if 9 patients recover, the null hypothesis is accepted.
$P(X \leq 10) = 1 - 0.9861 = 0.0139 > 0.01$
i.e. if 10 patients recover, the null hypothesis is rejected.
Therefore, no more than 9 patients should recover for the test to be significant at this level.

22 a H_0: p = 0.3 i.e the fertiliser will have no effect.
H_1: p > 0.3 i.e. the fertiliser will increase the probability of the tomatoes having a dimeter > 4 cm.

b Significance level 5%
If H_0 is true $X \sim$ B(40, 0.3)

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.

22 b Let c be the critical value.
$P(X \geq 17) = 1 - P(X \leq 16) = 1 - 0.9367 = 0.0633 > 0.05$
$P(X \geq 18) = 1 - P(X \leq 17) = 1 - 0.9680 = 0.032 < 0.05$
So $c = 18$
$P(X \geq 18)$, so the critical region is $18 \leq X \leq 40$

c Actual significance level = 0.032 = 3.2%

d The observed value of 18 lies in the critical region so it is reasonable to reject the null hypothesis that the fertiliser has no effect. Dhiriti's claim that the fertiliser works is supported.

Challenge

1 $P(C) = \frac{z+7}{50}$

$P(A) = \frac{y+1}{50}$

$\frac{z+7}{50} = 3\left(\frac{y+1}{50}\right)$

$z + 7 = 3y + 3$

$z + 4 = 3y$ **(1)**

$P(\text{not } B) = 0.76 = \frac{38}{50}$

$P(\text{not } B) = \frac{y+z+18}{50}$

So $\frac{y+z+18}{50} = \frac{38}{50}$

$y + z + 18 = 38$

$y = 20 - z$ **(2)**

Use **(2)** to substitute for y in **(1)**:
$z + 4 = 3(20 - z)$
$z + 4 = 60 - 3z$
$4z = 56$
$z = 14$
Substituting this value for z in **(2)**:
$y = 20 - 14 = 6$
Referring to the diagram:
$x = 50 - (6 + 1 + 7 + 14 + 18) = 4$
$x = 4, y = 6, z = 14$

2 a Significance level 10%
If H_0 is true $X \sim B(30, 0.65)$
Let c be the critical value.
$P(X \leq 15) = 0.0652$
$P(X \leq 16) = 0.1263$
$0.1 - 0.0652 = 0.0348$ $(X \leq 15)$
$0.1263 - 0.1 = 0.0263$ $(X \leq 16)$
So $c = 16$
$P(X \leq 16)$, so the critical region is $0 \leq X \leq 16$

b $P(X \leq 16) = 0.1263$
$P(X \leq 16 \text{ and } X \leq 16) = 0.1263^2 = 0.0160$

© Pearson Education Ltd 2017. Copying permitted for purchasing institution only. This material is not copyright free.