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.

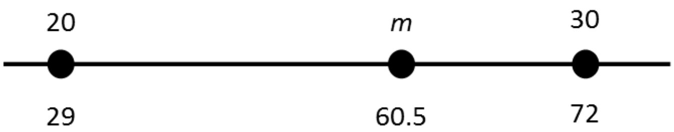
2 a Opportunity sampling is using a sample that is available at the time the study is carried out.
It is unlikely to provide a representative sample of the weather in May.

b 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.

3 b mean = $\frac{\sum fx}{\sum f} = \frac{3610}{120} = 30.1$

$$\sigma = \sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2}$$
$$= \sqrt{\frac{141\,600}{120} - \left(\frac{3610}{120}\right)^2}$$
$$= \sqrt{1180 - 30.1^2}$$
$$= \sqrt{275} = 16.6$$

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, $\bar{x} = \frac{\sum x}{n} = \frac{947}{30} = 31.6$

actual mean, $\bar{s} = \frac{31.6-1}{10} = 3.06$ hours

coded standard deviation, $\sigma_x = \sqrt{\frac{33\,065.37}{30}} = 33.2$

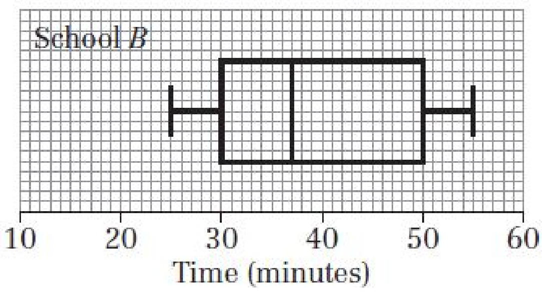
actual standard deviation, $\sigma_s = \frac{33.2}{10} = 3.32$ hours

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).

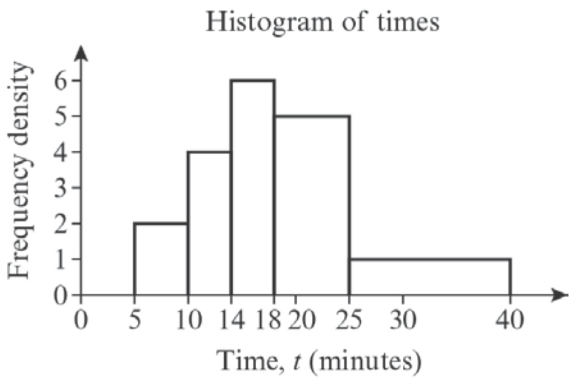
5 d Children from school A generally took less time than those from school B.

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 ÷ 4 = 4
$14 \le x < 18$: height = 24 ÷ 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 ÷ 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 $\sum ft^2 = (10\times 7.5^2) + (16\times 12^2) + (24\times 16^2) + (35\times 21.5^2) + (15\times 32.5^2) = 41033$

$$\sigma = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$= \sqrt{\frac{41\,033}{100} - \left(\frac{1891}{100}\right)^2}$$

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.)

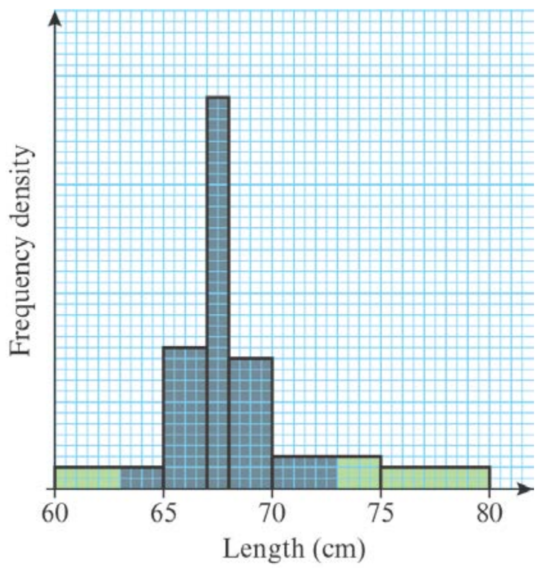
e median, Q_2 is the $\frac{100+1}{2} = 50.5$th term
There are 50 values less than 18 minutes and 50 values greater than or equal to 18 minutes.
The median is therefore 18 minutes.

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.



$$P(63 \le l \le 73) = \frac{(2\times 2) + 26 + 36 + 24 + (3\times 3)}{10 + 26 + 36 + 24 + 15 + 10}$$
$$= \frac{99}{121}$$
$$= 0.82$$

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 : $\frac{80}{100} \times 31 = 24.8$

$$\frac{P_{80} - 19}{22 - 19} = \frac{24.8 - 21}{31 - 21}$$

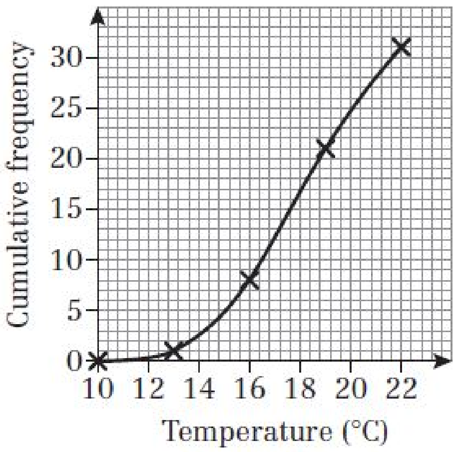
$$P_{80} - 19 = \frac{3.8 \times 3}{10}$$

$$P_{80} = 1.14 + 19$$

$$= 20.14$$

$P_{80} - P_{20} = 20.14 - 15.23 = 4.9$ (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:

$$\frac{15 - 13}{16 - 13} = \frac{d - 1}{8 - 1}$$

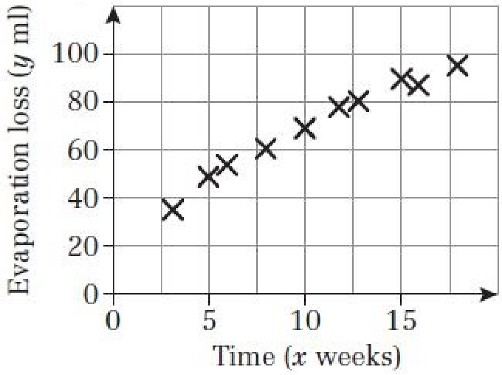
$$\frac{2}{3} = \frac{d - 1}{7}$$

$$d = \frac{2 \times 7}{3} + 1$$

$$= 5.67$$

31 − 5.67 = 25.33, so 25 days

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.

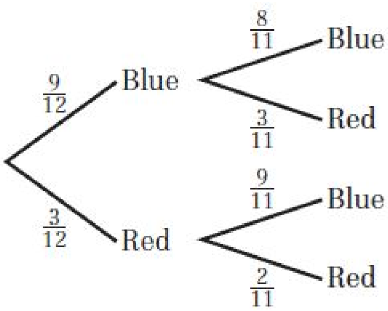
10 a mean + 2 standard deviations = 15.3 + 2 × 10.2 = 35.7
45 > 35.7 so t = 45 is an outlier

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.

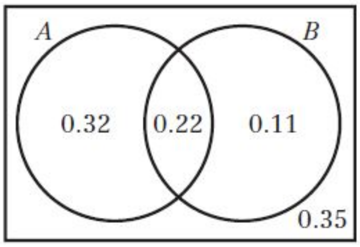
11 c P (balls are different colours) = P(blue then red) + P(red then blue)

$$= \frac{9}{12}\times\frac{3}{11}+\frac{3}{12}\times\frac{9}{11}$$

$$= \frac{27+27}{132} = \frac{54}{132}$$

The probability the balls are different colours is 0.409.

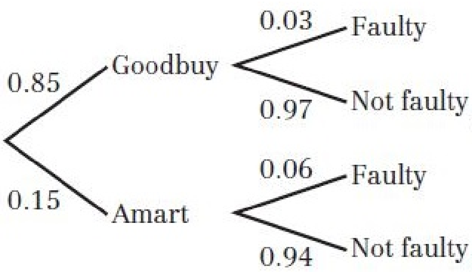
12 a P(*A* or *B*) = P(*A*) + P(B) − P(*A* and *B*)
0.65 = 0.32 + 0.11 − P(*A* and *B*)
P(*A* and *B*) = 0.65 − 0.32 − 0.11= 0.22
P(neither *A* nor *B*) = 1 – 0.65 = 0.35



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*)

$$= (0.85\times0.97)+(0.15\times0.94)$$

$$= 0.9655$$

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.

15 a For four coin tosses there are 2 × 2 × 2 × 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'.
$X \sim \mathrm{B}(4, 0.5)$

$$\mathrm{P}(X=2)=\binom{4}{2}0.5^2\times 0.5^2$$
$$=\frac{4!}{2!2!}0.5^2\times 0.5^2$$
$$=0.375$$

b $\mathrm{P}(X=0 \text{ or } 4) = \mathrm{P}(X=0)+\mathrm{P}(X=4)$
$$=\frac{1}{16}+\frac{1}{16}=\frac{1}{8}$$
$$=0.125$$

c $\mathrm{P}(HHT)=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}$
$$=0.125$$

16 a Let X be the random variable 'the number of faulty bolts'.
$X \sim \mathrm{B}(20, 0.3)$

$$\mathrm{P}(X=2)=\binom{20}{2}\times(0.3)^2(0.7)^{18}$$
$$=\frac{20!}{2!18!}(0.3)^2(0.7)^{18}$$
$$=0.0278$$

b $\mathrm{P}(X>3)=1-\mathrm{P}(X\le 3)$
From calculator or tables, $\mathrm{P}(X \le 3) = 0.1071$
$\mathrm{P}(X>3)=1-0.1071$
$$=0.8929$$

16 c Let Y be the random variable 'number of bags containing more than 3 faulty bolts'.
Y ~ 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 $P(2<X\leq 5)=P(Y=3)+P(Y=4)+P(Y=5)$

$$P(Y=3)+P(Y=4)+P(Y=5)=\tfrac{21}{36}$$
$$=\tfrac{7}{12}$$
$$=0.5833...$$

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	$0.4\times0.6=0.24$	$0.4^2\times0.6=0.096$	$0.4^3\times0.6+0.4^4=0.064$

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.

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 $X \sim B(40, 0.3)$

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 $P(X \leq 16) = 0.1263$
$P(X \leq 16 \text{ and } X \leq 16) = 0.1263^2 = 0.0160$