Review exercise 1

1 a Any 2 from:
- Used to simplify or represent a real-world problem.
- Cheaper or quicker (than producing the real situation) or more easily modified.
- To improve understanding of the real-world problem.
- Used to predict outcomes from a real-world problem (idea of predictions).

b (3) Model used to make predictions.
(4) Experimental data collected.
(7) Model is refined. (Steps 2 (or 3) to 5 (or 6) are repeated).
You could put 3 and 4 the other way round.

2 $\bar{y} = \dfrac{\bar{x} - 120}{5}$ therefore:

$$\frac{\bar{x} - 120}{5} = 24$$
$$\bar{x} = 240$$

$\sigma_{\bar{y}} = \dfrac{\sigma_{\bar{x}}}{5}$ therefore:

$$\sigma_{\bar{x}} = 5 \times 2.8$$
$$= 14$$

3

$\bar{y} = 1.4\bar{x} - 20$ therefore:

$$1.4\bar{x} - 20 = 60.8$$
$$\bar{x} = \frac{404}{7}$$
$$= 57.7 \text{ (3 s.f.)}$$

$\sigma_y = 1.4\sigma_x$ therefore:

$$\sigma_x = \frac{6.60}{1.4}$$
$$= \frac{33}{7}$$
$$= 4.71 \text{ (3 s.f.)}$$

4 $x = 10s + 1$

$s = \frac{x-1}{10}$

coded mean, $\bar{x} = \frac{\Sigma x}{n} = \frac{947}{30} = 31.6$

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 $\bar{y} = \frac{\bar{x}-720}{1000}$ therefore:

$\frac{\bar{x}-720}{1000} = 18$

$\bar{x} = \$18\,720$

6 a $t = \frac{m+12}{1.25}$

b $n = 28$, $\bar{t} = 52.8$ and $s_{tt} = 7.3$

$t = \frac{m+12}{1.25}$

$\frac{\bar{m}+12}{1.25} = 52.8$

$\bar{m} = 54$

Therefore the mean of the original data is 54

$\sigma_t = \sqrt{\frac{S_{tt}}{n}}$

$= \sqrt{\frac{7.3}{28}}$

$= 0.510...$

$\sigma_t = \frac{\sigma_m}{1.25}$

$\frac{\sigma_m}{1.25} = 0.510...$

$\sigma_m = 0.638$ (3 s.f.)

Therefore the standard deviation of the original data is 0.638 (3 s.f.)

7 a

t	5–10	10–14	14–18	18–25	25–40
Frequency	10	16	24	35	15

b 40

7 c $\bar{t} = \frac{7.5\times10+12\times16+16\times24+21.5\times35+32.5\times15}{100}$

$= 18.91$ minutes

d $\sigma^2 = \frac{7.5^2\times10+12^2\times16+16^2\times24+21.5^2\times35+32.5^2\times15}{100} - 18.91^2$

$= 52.7\ldots$

$\sigma = 7.26$ (3 s.f.)

e Median = 18 minutes
Using interpolation:
The lower quartile lies in the 10–14 group

$Q_1 = 10 + \frac{15}{16}\times4$

Q_1 = 13.75 minutes
The upper quartile lies in the 18–25 group

$Q_3 = 18 + \frac{25}{32}\times7$

Q_1 = 23 minutes

f $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}} = \frac{3(18.91-18)}{7.26...}$

$= 0.376$ (3 s.f.)

Therefore it is positively skewed.

8 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.

9 a Positive skew

b The median lies in the 20–29 group

$Q_2 = 19.5 + \frac{31}{43}\times10$

$Q_2 = 26.7$ km (3 s.f.)

9 c $\sum fx = 3550$ and $\sum fx^2 = 138\,020$

$$\bar{x} = \frac{\sum fx}{n} = \frac{3550}{120} = 29.6 \text{ km (3 s.f.)}$$

$$\sigma^2 = \frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2 = \frac{138\,020}{120} - \left(\frac{3550}{120}\right)^2 = 274.99...$$

$\sigma = 16.6$ (3 s.f.)

d $$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}} = \frac{3(29.58... - 26.70...)}{16.58...} = 0.520 \text{ (3 s.f.)}$$

e Yes as 0.520 > 0

f Use the median since the data is skewed.

g If the data is symmetrical.

10 a Mode = 56

b There are 27 pieces of data therefore the median is the 14th piece of data.
Median = 52
Q_1 is the 7th piece of data so $Q_1 = 35$
Q_3 is the 21st piece of data so $Q_3 = 60$

c $\sum fx = 1335$ and $\sum fx^2 = 71\,801$

$$\bar{x} = \frac{\sum fx}{n} = \frac{1335}{27} = 49.4 \text{ (3 s.f.)}$$

$$\sigma^2 = \frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2 = \frac{71\,801}{27} - \left(\frac{1335}{27}\right)^2 = 214.54...$$

$\sigma = 14.6$ (3 s.f.)

10 d $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}} = \frac{3(49.44... - 52)}{14.64...}$
$= -0.533$ (3 s.f.)

e For a negative skew;
Mean < median < mode (49.4 < 52 < 56)
$Q_2 - Q_1 > Q_3 - Q_2$ (17 > 8)

11 a Distance is a continuous variable.

b Frequency density = $\frac{\text{class width}}{\text{frequency}}$

Class	Frequency	Frequency density
41–45	4	0.8
46–50	19	3.8
51–60	53	5.3
61–70	37	3.7
71–90	15	0.75
91–150	6	0.1

c The median is in the 51-60 group
$Q_2 = 50.5 + \frac{44}{53} \times 10$
$Q_2 = 58.8$ (3 s.f.)
The lower quartile is in the 51–60 group
$Q_1 = 50.5 + \frac{10.5}{53} \times 10$
$Q_1 = 52.5$ (3 s.f.)
The upper quartile is in the 61–70 group
$Q_3 = 60.5 + \frac{24.5}{37} \times 10$
$Q_3 = 67.1$ (3 s.f.)

11 d $\sum fx = 8379.5$ and $\sum fx^2 = 557\,489.75$

$$\bar{x} = \frac{\sum fx}{n} = \frac{8379.5}{134} = 62.5 \text{ (3 s.f.)}$$

$$\sigma^2 = \frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2 = \frac{557\,489.75}{134} - \left(\frac{8379.5}{134}\right)^2 = 249.92...$$

$\sigma = 15.8$ (3 s.f.)

e $$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} = \frac{67.12... - 2(58.80...) + 52.48...}{67.12... - 52.48...} = 0.137 \text{ (3 s.f.)}$$

So positively skewed.

f $Q_3 - Q_2 > Q_2 - Q_1$ (8.3 > 6.3)

or

$$\frac{3(62.53... - 58.80...)}{15.80...} = 0.708 \text{ (3 s.f.)}$$

0.708 > 0 so positively skewed.

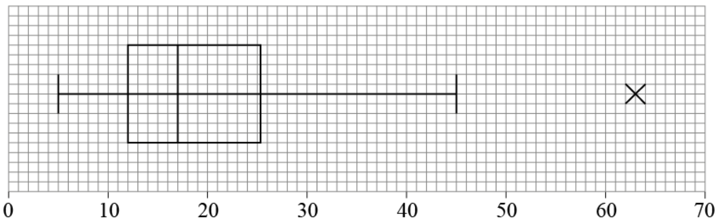
12 a Time is a continuous variable.

b Area is proportional to frequency.

c Area on diagram is 7.2 cm^2 which represents 9 students; therefore 1 student is represented by 0.8 cm^2.

d The total area is 24 cm^2. Therefore the number of students is 24 ÷ 0.8 = 30 students.

13 a



b Distribution is positively skewed since $Q_2 - Q_1 < Q_3 - Q_2$ (5 < 11)

c Many delays are short and passengers should find them acceptable.

14 a 17 males and 15 females

b £48

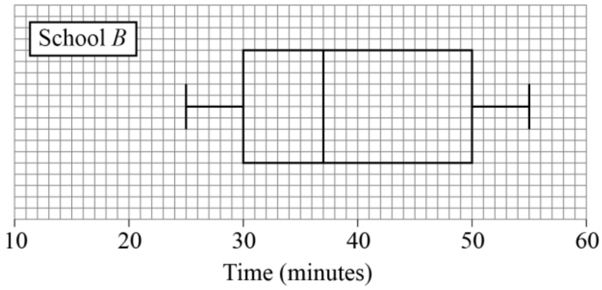
c Males tend to earn more.

15 a i 37 minutes

ii upper quartile

b They are outliers. Outliers are values that are much greater or much less than the other values.

c



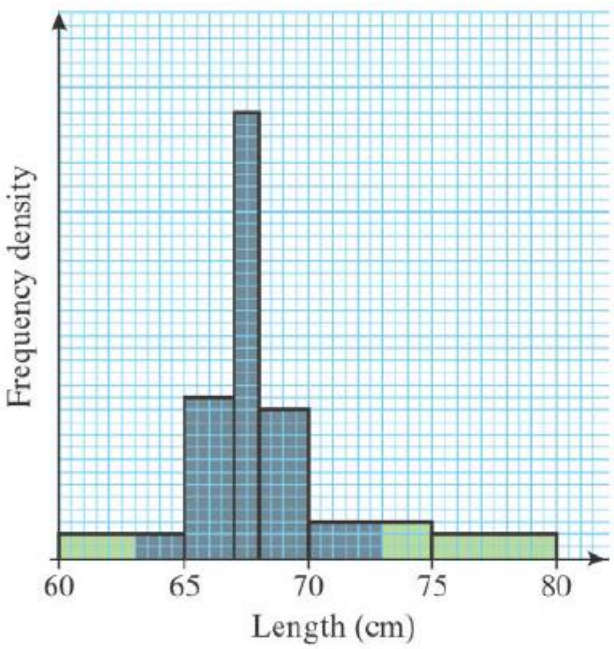
d The children from school *A* generally took less time than those from school *B*. The median for *A* is less than the median for *B*. *A* has outliers, but *B* does not. The interquartile range for *A* is less than the interquartile range for *B*, suggesting that the times for school A are less spread out. However, the total range for *A* is greater than the total range for *B* (although this includes the outliers).

16 Area of 65 to 67 cm class = 26
Frequency density = 26 ÷ 2 = 13

Using this information:

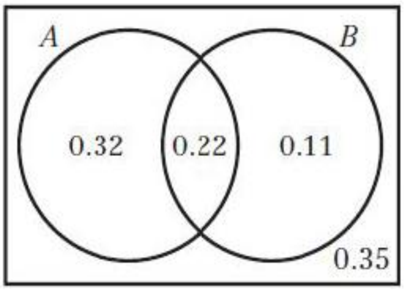
Length, *l* (cm)	**Frequency**	**Class width**	**Frequency density**
60 to 65	10	5	2
65 to 67	26	2	13
67	36	1	36
68 to 70	24	2	12
70 to 75	15	5	3
75 to 80	10	5	2

The number of owls with wing length between 63 and 73 cm is given by the shaded area on the graph.



$$P(63 \le l \le 73) = \frac{(2\times2)+26+36+24+(3\times3)}{10+26+36+24+15+10}$$
$$= \frac{99}{121}$$
$$= 0.82$$

17 a P(*A* or *B*) = P(*A* but not *B*) + P(*B* but not *A*) + 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.

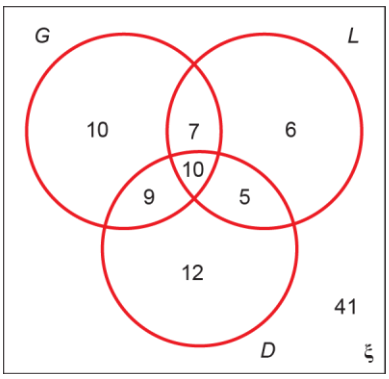
18 a Magazines and Television are mutually exclusive preferences as the sets do not overlap.

b P(*M* and *B*) = $\frac{13}{38} = 0.34$

P(*M*) × 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.

19 a Start in the middle of the Venn diagram and work outwards. Remember the rectangle and those not in any of the circles. Your numbers should total 100.



b $\mathrm{P}(G \cap L' \cap D') = \frac{10}{100} = \frac{1}{10} = 0.1$

c $\mathrm{P}(G' \cap L' \cap D') = \frac{41}{100} = 0.41$

d $\mathrm{P}(\text{only two attributes}) = \frac{9+7+5}{100} = \frac{21}{100} = 0.21$

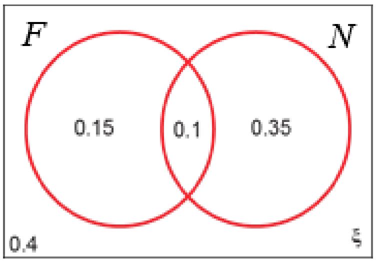
19 e The word 'given' in the question tells you to use conditional probability:

$$\mathrm{P}(G \mid L \cap D) = \frac{\mathrm{P}(G \mid L \cap D)}{\mathrm{P}(\mathrm{L} \mid D)} = \frac{\frac{10}{100}}{\frac{15}{100}} = \frac{10}{15} = \frac{2}{3} = 0.667 \text{ (3 s.f.)}$$

20 a Let F be the event that a student reads fiction books on a regular basis, and N the event that they read non-fiction books.

$$\mathrm{P}(F \cup N) = \mathrm{P}(F) + \mathrm{P}(N) - \mathrm{P}(F \cap N)$$
$$0.6 = 0.25 + 0.45 - \mathrm{P}(F \cap N)$$
$$\mathrm{P}(F \cap N) = 0.1$$

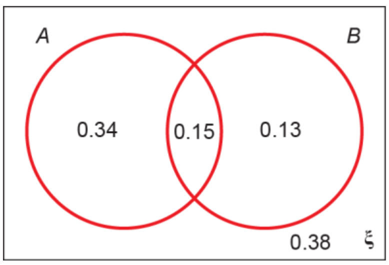
b When drawing the Venn diagram remember to draw a rectangle around the circles and add the probability 0.4.
Remember the total in circle $F = 0.25$ and the total in circle $N = 0.45$.



c The words 'given that' in the question tell you to use conditional probability:

$$\mathrm{P}(F \cap N' \mid F \cup N) = \frac{0.15}{0.6} = \frac{1}{4} = 0.25$$

21 a The first two probabilities allow two spaces in the Venn diagram to be filled in.
$\mathrm{P}(A \cup B) = \mathrm{P}(A \cap B') + \mathrm{P}(A' \cap B) + \mathrm{P}(A \cap B)$, and this can be rearranged to see that
$P(A \cap B) = 0.15$
Finally, $\mathrm{P}(A \cup B) = 0.62 \Rightarrow \mathrm{P}((A \cup B)') = 0.38$. The completed Venn diagram is therefore:



b $\mathrm{P}(A) = 0.34 + 0.15 = 0.49$ and $\mathrm{P}(B) = 0.13 + 0.15 = 0.28$

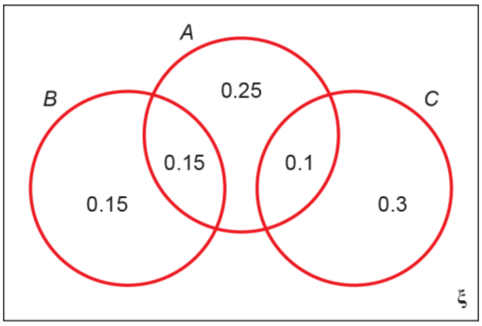
c $\mathrm{P}(A \mid B') = \dfrac{\mathrm{P}(A \cap B')}{\mathrm{P}(B')} = \dfrac{0.34}{1 - \mathrm{P}(B)} = \dfrac{0.34}{0.72} = 0.472$ (3 d.p.).

d If A and B are independent, then $\mathrm{P}(A) = \mathrm{P}(A \mid B) = \mathrm{P}(A \mid B')$. From parts **b** and **c**, this is not the case. Therefore they are not independent.

22 a $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B) \Rightarrow \mathrm{P}(A) = \mathrm{P}(A \cap B) \div \mathrm{P}(B) = 0.15 \div 0.3 = 0.5$

b $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = 0.5 + 0.3 - 0.15 = 0.65 \Rightarrow \mathrm{P}(A' \cap B') = 1 - 0.65 = 0.35$

c Since B and C are mutually exclusive, they do not intersect.
The intersection of A and C should be 0.1 but $\mathrm{P}(A) = 0.5$, allowing $\mathrm{P}(A \cap B' \cap C')$ to be calculated. The filled-in probabilities sum to 0.95, and so $\mathrm{P}(A' \cap B' \cap C') = 0.05$.
Therefore, the filled-in Venn diagram should look like:



d i $\mathrm{P}(A \mid C) = \dfrac{\mathrm{P}(A \cap C)}{\mathrm{P}(C)} = \dfrac{0.1}{0.4} = 0.25$

ii The set $A \cap (B \cup C')$ must be contained within A. First find the set $B \cup C'$: this is made up from four distinct regions on the above Venn diagram, with labels 0.15, 0.15, 0.25 and 0.05.
Restricting to those regions that are also contained within A leaves those labelled 0.15 and 0.25.
Therefore, $\mathrm{P}(A \cap (B \cup C')) = 0.15 + 0.25 = 0.4$

iii From part **ii**, $\mathrm{P}(B \cup C') = 0.15 + 0.15 + 0.25 + 0.05 = 0.6$. Therefore

$$\mathrm{P}(A \mid (B \cup C')) = \frac{\mathrm{P}(A \cap (B \cup C'))}{\mathrm{P}(B \cup C')} = \frac{0.4}{0.6} = \frac{2}{3}$$

23 a $\mathrm{P}(\text{tourism}) = \dfrac{50}{148} = \dfrac{25}{74} = 0.338$ (3 s.f.)

b The words 'given that' in the question tell you to use conditional probability:

$$\mathrm{P}(\text{no glasses} \mid \text{tourism}) = \frac{\mathrm{P}(G' \cap T)}{\mathrm{P}(T)} = \frac{\frac{23}{148}}{\frac{50}{148}} = \frac{23}{50} = 0.46$$

c It often helps to write down which combinations you want:

$$\begin{aligned}\mathrm{P}(\text{right-handed}) &= \mathrm{P}(E \cap RH) + \mathrm{P}(T \cap RH) + \mathrm{P}(C \cap RH)\\ &= \frac{30}{148} \times 0.8 + \frac{50}{148} \times 0.7 + \frac{68}{148} \times 0.75\\ &= \frac{55}{74} = 0.743 \text{ (3 s.f.)}\end{aligned}$$

d The words 'given that' in the question tell you to use conditional probability:

$$\mathrm{P}(\text{engineering} \mid \text{right-handed}) = \frac{\mathrm{P}(E \cap RH)}{\mathrm{P}(RH)} = \frac{\frac{30}{148} \times 0.8}{\frac{55}{74}} = \frac{12}{55} = 0.218 \text{ (3 s.f.)}$$

24 a There are two different events going on: 'Joanna oversleeps' (O) and 'Joanna is late for college' (L). From the context, we cannot assume that these are independent events.
Drawing a Venn diagram, none of the regions can immediately be filled in. We are told that $\text{P}(O) = 0.15$ and so $\text{P}(J$ does not oversleep$) = \text{P}(O') = 0.85$. The other two statements can be interpreted as $\frac{\text{P}(L \cap O)}{\text{P}(O)} = 0.75$ and $\frac{\text{P}(L \cap O')}{\text{P}(O')} = 0.1$

Filling in the first one:

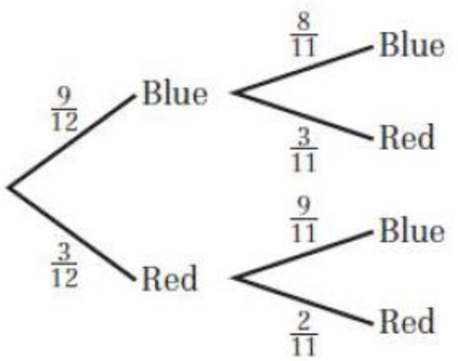
$$\frac{\text{P}(L \cap O)}{\text{P}(O)} = 0.75 \Rightarrow \frac{\text{P}(L \cap O)}{0.15} = 0.75 \Rightarrow \text{P}(L \cap O) = 0.1125$$

Also, $\frac{\text{P}(L \cap O')}{0.85} = 0.1 \Rightarrow \text{P}(L \cap O') = 0.085$

Therefore, $\text{P}(L) = \text{P}(L \cap O) + \text{P}(L \cap O') = 0.1125 + 0.085 = 0.1975$

b $\text{P}(L \mid O) = \frac{\text{P}(L \cap O)}{\text{P}(O)} = \frac{0.1125}{0.1975} = \frac{45}{79} = 0.5696$ (4 s.f.).

25 a



b P(second ball red) = P(blue then red) + P(red then red)

$$= \tfrac{9}{12} \times \tfrac{3}{11} + \tfrac{3}{12} \times \tfrac{2}{11} = \tfrac{27+6}{132} = \tfrac{1}{4}$$

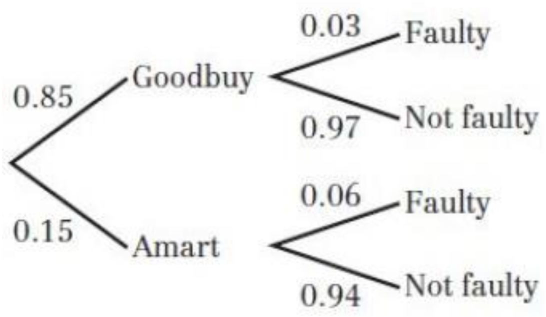
The probability the second ball is red is 0.25.

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

$$= \tfrac{9}{12} \times \tfrac{3}{11} + \tfrac{3}{12} \times \tfrac{9}{11} = \tfrac{27+27}{132} = \tfrac{54}{132}$$

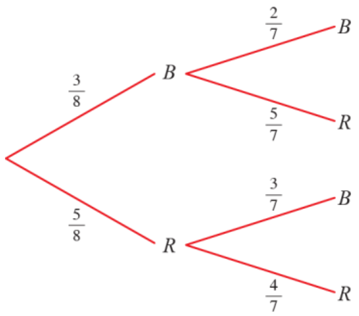
The probability the balls are different colours is 0.409.

26 a



b G = Goodbuy, A = Amart, NF = Not faulty
P(NF) = P(G and NF) + P(A and NF)
= (0.850.97) + (0.150.94)
= 0.9655

27 a



b i There are two different situations where the second counter drawn is blue. These are BB and RB. Therefore the probability is: $\left(\frac{3}{8}\times\frac{2}{7}\right)+\left(\frac{5}{8}\times\frac{3}{7}\right)=\frac{6+15}{56}=\frac{21}{56}=\frac{3}{8}=0.375$.

ii $\text{P(both blue} \mid \text{2nd blue)} = \dfrac{\text{P(both blue and 2nd blue)}}{\text{P(2nd blue)}} = \dfrac{\text{P(both blue)}}{\text{P(2nd blue)}} = \dfrac{\frac{3}{8}\times\frac{2}{7}}{\frac{3}{8}} = \dfrac{2}{7}$

Challenge

1 P(C) = $\frac{z+7}{50}$

P(A) = $\frac{y+1}{50}$

$\frac{y+7}{50} = 3\left(\frac{y+1}{50}\right)$

$z + 7 = 3y + 3$

$z + 4 = 3y$ **(1)**

P(not B) = 0.76 = $\frac{38}{50}$

P(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 Since A and B could be mutually exclusive, $\mathrm{P}(A \cap B) \geq 0$. Since $\mathrm{P}(A \cap B) \leq \mathrm{P}(B) = 0.3$, we have that $0 \leq \mathrm{P}(A \cap B) \leq 0.3$ and so $p = \mathrm{P}(A \cap B') = \mathrm{P}(A) - \mathrm{P}(A \cap B)$. Therefore $0.4 \leq p \leq 0.7$

b First, $\mathrm{P}(B \cap C) \leq \mathrm{P}(B) = 0.3$ and so $q \leq \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B \cap C) \leq 0.25$. Moreover, it is possible to draw a Venn diagram where $q = 0$, and so $0 \leq q \leq 0.25$