Review exercise 1

1 a Produce a table for the values of log s and log t:

<table>
<thead>
<tr>
<th>log s</th>
<th>0.3010</th>
<th>0.6532</th>
<th>0.7924</th>
<th>0.8633</th>
<th>0.9590</th>
</tr>
</thead>
<tbody>
<tr>
<td>log t</td>
<td>−0.4815</td>
<td>0.0607</td>
<td>0.2455</td>
<td>0.3324</td>
<td>0.4698</td>
</tr>
</tbody>
</table>

which produces \( r = 0.9992 \)

b Since \( r \) is very close to 1, this indicates that log s by log t is very close to being linear, which means that \( s \) and \( t \) are related by an equation of the form \( t = ax^r \) (beginning of Section 1.1).

c Rearranging the equation:

\[
\log t = -0.9051 + 0.4437 \log s
\]

\[
\Rightarrow t = 10^{-0.9051 + 0.4437 \log s} = 10^{-0.9051} \times 10^{0.4437 \log s}
\]

\[
\Rightarrow t = 10^{-0.9051} \times s^{1.4437}
\]

and so \( a = 10^{-0.9051} = 0.1244 \) (4 s.f.) and \( n = 1.4437 \)

2 a Rearranging the equation:

\[
y = -0.2139 + 0.0172x
\]

\[
\Rightarrow \log t = -0.2139 + 0.0172P
\]

\[
\Rightarrow t = 10^{-0.2139 + 0.0172P} = 10^{-0.2139} \times 10^{0.0172P}
\]

\[
\Rightarrow t = 10^{-0.2139} \times (10^{0.0172})^P
\]

Therefore \( a = 10^{-0.2139} = 0.611 \) (3 s.f.) and \( b = 10^{0.0172} = 1.04 \) (3 s.f.).

b Not in the range of data (extrapolation).

3 a \[
r = \frac{59.524}{\sqrt{152.444 \times 26.589}} = 0.93494\]

(\( \text{the formulae for this is under S1 in the formula book.} \))

b Make sure your hypotheses are clearly written using the parameter \( \rho \):

\[
H_0 : \rho = 0, \quad H_1 : \rho > 0
\]

Test statistic: \( r = 0.935 \)

Critical value at 1% = 0.7155

(\( \text{Look up the value under 0.01 in the table for product moment coefficient; quote the figure in} \)

\( \text{full.} \))

\( 0.935 > 0.7155 \)

Draw a conclusion in the context of the question:

So reject \( H_0 \) : levels of serum and disease are positively correlated.

4 \( r = -0.4063 \), critical value for \( n = 6 \) is \( -0.6084 \), so no evidence.
5 a  \( H_0: \rho = 0 \)
\( H_1: \rho < 0 \)
From the data, \( r = -0.9313 \). Since the critical value for \( n = 5 \) is \(-0.8783\), there is sufficient evidence to reject \( H_0 \), i.e. at the 2.5\% level of significance, there is sufficient evidence to say that there is negative correlation between the number of miles done by a one-year-old car and its value.

b If a 1\% level of significance was used, then the critical value for \( n = 5 \) is \(-0.9343\) and so there would not be sufficient evidence to reject \( H_0 \).

6 a  \[
P(\text{tourism}) = \frac{50}{148} = \frac{25}{74} = 0.338 \text{ (3 s.f.)}
\]

b The words ‘given that’ in the question tell you to use conditional probability:
\[
P(\text{no glasses} \mid \text{tourism}) = \frac{P(G' \cap T)}{P(T)} = \frac{23}{148} = \frac{23}{50} = 0.46
\]

c It often helps to write down which combinations you want:
\[
P(\text{right-handed}) = P(E \cap RH) + P(T \cap RH) + P(C \cap RH)
\]
\[
= \frac{30}{148} \times 0.8 + \frac{50}{148} \times 0.7 + \frac{68}{148} \times 0.75
\]
\[
= \frac{55}{74} = 0.743 \text{ (3 s.f.)}
\]

d The words ‘given that’ in the question tell you to use conditional probability:
\[
P(\text{engineering} \mid \text{right-handed}) = \frac{P(E \cap RH)}{P(RH)} = \frac{30}{148} \times 0.8 = \frac{12}{55} = 0.218 \text{ (3 s.f.)}
\]
7 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.

\[\begin{array}{c}
G \\
\quad7
\end{array} \quad \begin{array}{c}
L \\
\quad6
\end{array} \quad \begin{array}{c}
D \\
\quad12
\end{array}
\]

\[\begin{array}{c}
9 \\
\quad10
\end{array} \quad \begin{array}{c}
10 \\
\quad5
\end{array} \quad \begin{array}{c}
41
\end{array}
\]

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

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

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

e The word ‘given’ in the question tells you to use conditional probability:
\(P(G|L \cap D) = \frac{P(G \cap L \cap D)}{P(L \cap D)}\)
\(= \frac{10}{100} \div \frac{15}{100} = \frac{10}{15} = \frac{2}{3} = 0.667 (3 \text{ s.f.})\)

8 a \(P(B \cup T) = P(B) + P(T) - P(B \cap T)\)
\(0.6 = 0.25 + 0.45 - P(B \cap T)\)
\(P(B \cap T) = 0.1\)

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 \(B = 0.25\) and the total in circle \(T = 0.45\).
8 c  The words ‘given that’ in the question tell you to use conditional probability:

\[ P(B \cap T' \mid B \cup T) = \frac{0.15}{0.6} = \frac{1}{4} = 0.25 \]

9 a

There are two different situations where the second counter drawn is blue. These are BB and RB. Therefore the probability is:

\[ P(\text{both blue} \mid \text{2nd blue}) = \frac{\frac{3}{8} \times \frac{5}{7} + \frac{5}{8} \times \frac{3}{7}}{\frac{21}{56} + \frac{3}{8}} = \frac{3}{8} \]

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

ii  \( P(\text{both blue} \mid \text{2nd blue}) = \frac{\frac{3}{8} \times \frac{2}{7}}{\frac{21}{56}} = \frac{2}{7} \)

10 a  The first two probabilities allow two spaces in the Venn diagram to be filled in.

\[ P(\text{both blue} \mid \text{2nd blue}) = \frac{\text{P(both blue and 2nd blue)}}{\text{P(2nd blue)}} = \frac{\frac{3}{8} \times \frac{2}{7}}{\frac{21}{56}} = \frac{2}{7} \]

Finally, \( P(\text{both blue} \mid \text{2nd blue}) = \frac{3}{8} \).

The completed Venn diagram is therefore:

b  \( P(A) = 0.34 + 0.15 = 0.49 \) and \( P(B) = 0.13 + 0.15 = 0.28 \)

c  \( P(A \mid B') = \frac{P(A \cap B')}{P(B')} = \frac{0.34}{0.72} = \frac{0.34}{0.72} = 0.472 \) (3 d.p.).

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

d  \( P(A \cap B) = P(A) \times P(B) \Rightarrow P(A) = P(A \cap B) \div P(B) = 0.15 \div 0.3 = 0.5 \)
11 b \[ P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.3 - 0.15 = 0.65 \Rightarrow P(A' \cap B') = 1 - 0.65 = 0.35 \]

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

![Venn Diagram](image)

d i \[ P(A | C) = \frac{P(A \cap C)}{P(C)} = \frac{0.1}{0.4} = 0.25 \]

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, \( P(A \cap (B \cup C')) = 0.15 + 0.25 = 0.4 \)

iii From part ii, \( P(B \cup C') = 0.15 + 0.15 + 0.25 + 0.05 = 0.6 \). Therefore \[ P(A | (B \cup C')) = \frac{P(A \cap (B \cup C'))}{P(B \cup C')} = \frac{0.4}{0.6} = \frac{2}{3} \]

12 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 \( P(O) = 0.15 \) and so \( P(O' \text{ does not oversleep}) = P(O') = 0.85 \). The other two statements can be interpreted as \( \frac{P(L \cap O)}{P(O)} = 0.75 \) and \( \frac{P(L \cap O')}{P(O')} = 0.1 \)

Filling in the first one: \[ \frac{P(L \cap O)}{P(O)} = 0.75 \Rightarrow \frac{P(L \cap O)}{0.15} = 0.75 \Rightarrow P(L \cap O) = 0.1125 \]

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

Therefore, \( P(L) = P(L \cap O) + P(L \cap O') = 0.1125 + 0.085 = 0.1975 \)

b \[ P(L | O) = \frac{P(L \cap O)}{P(O)} = \frac{0.1125}{0.1975} = \frac{45}{79} = 0.5696 \text{ (4 s.f.)} \]
13 a Drawing a diagram will help you to work out the correct area:

Using \( z = \frac{x - \mu}{\sigma} \). As 91 is to the left of 100, your \( z \) value should be negative.

\[
P(X < 91) = P\left( Z < \frac{91 - 100}{15} \right)
= P(Z < -0.6)
= 1 - 0.7257
= 0.2743
\]

(The tables give \( P(Z < 0.6) = P(Z > -0.6) \), so you want 1 – this probability.)

b

As 0.2090 is not in the table of percentage points you must work out the larger area:

\[
1 - 0.2090 = 0.7910
\]

Use the first table or calculator to find the \( z \) value. It is positive as \( 100 + k \) is to the right of 100.

\[
P(X > 100 + k) = 0.2090 \text{ or } P(X < 100 + k) = 0.791
\]

\[
\frac{100 + k - 100}{15} = 0.81
\]

\[
k = 12
\]
14 a Let $H$ be the random variable ~ height of athletes, so $H \sim N(180, 5.2^2)$

Drawing a diagram will help you to work out the correct area:

Using $z = \frac{x - \mu}{\sigma}$. As 188 is to the right of 180 your $z$ value should be positive. The tables give $P(Z < 1.54)$ so you want $1 -$ this probability:

$$P(H > 188) = P \left( Z > \frac{188 - 100}{5.2} \right)$$
$$= P \left( Z > 1.54 \right)$$
$$= 1 - 0.9382$$
$$= 0.0618$$

b Let $W$ be the random variable ~ weight of athletes, so $W \sim N(85, 7.1^2)$

Using $z = \frac{x - \mu}{\sigma}$. As 97 is to the right of 85, your $z$ value should be positive.

$$P(W < 97) = P \left( Z < \frac{97 - 85}{7.1} \right)$$
$$= P \left( Z < 1.69 \right)$$
$$= 0.9545$$

c $P(W > 97) = 1 - P(W < 97)$, so

$$P(H > 188 \& W > 97) = 0.618(1 - 0.9545)$$
$$= 0.00281$$

d Use the context of the question when you are commenting:

The evidence suggests that height and weight are positively correlated.linked, so assumption of independence is not sensible.
15 a Use the table of percentage points or calculator to find $z$. You must use at least the four decimal places given in the table.

$P(Z > a) = 0.2$

$a = 0.8416$

$P(Z < b) = 0.3$

$b = -0.5244$

0.5244 is negative since 1.65 is to the left of the centre. 0.8416 is positive as 1.78 is to the right of the centre.

Using $z = \frac{x - \mu}{\sigma}$:

$\frac{1.78 - \mu}{\sigma} = 0.8416 \Rightarrow 1.78 - \mu = 0.8416\sigma$ (1)

$\frac{1.65 - \mu}{\sigma} = -0.5244 \Rightarrow 1.65 - \mu = 0.5244\sigma$ (2)

Solving simultaneously, (1) − (2):

$0.13 = 1.366\sigma$

$\sigma = 0.095$ m

Substitute in (1): $1.78 - \mu = 0.8416 \times 0.095$

$\mu = 1.70$ m

b

Using $z = \frac{x - \mu}{\sigma}$:

$P(\text{height} > 1.74) = P \left( z > \frac{1.74 - 1.70}{0.095} \right)$

$= P(z > 0.42)$ (the tables give $P(Z < 0.42)$ so you need $1 - \text{this probability}$)

$= 1 - 0.6628$

$= 0.3372$ (calculator gives 0.3369)

16 a $P(D < 21.5) = 0.32$ and $P(Z < a) = 0.32 \Rightarrow a = -0.467$. Therefore

$\frac{21.5 - \mu}{\sigma} = -0.467 \Rightarrow 21.5 - 22 = -0.467\sigma \Rightarrow \sigma = \frac{0.5}{0.467} = 1.071$ (4 s.f.).

b $P(21 < D < 22.5) = P(D < 22.5) - P(D < 21) = 0.5045$ (4 s.f.).

c $P(B \geq 10) = 1 - P(B \leq 9) = 1 - 0.01899 = 0.98101$ (using 4 s.f. for the value given by the binomial distribution) or 0.981 (4 s.f.).
17 a  Let \( W \) be the random variable ‘the number of white plants’. Then \( W \sim B(12, 0.45) \) (‘batches of 12’: \( n = 12 \); ‘45% have white flowers’: \( p = 0.45 \)).

\[
P(W = 5) = \binom{12}{5} 0.45^5 0.55^7 \quad (\text{you can also use tables: } P(W \leq 5) - P(W \leq 4))
\]

\[
= 0.2225
\]

\( b \)  Batches of 12, so: 7 white, 5 coloured; 8 white, 4 coloured; etc.

\[
P(W \geq 7) = 1 - P(W \leq 6)
\]

\[
= 1 - 0.7393
\]

\[
= 0.2607
\]

c  Use your answer to part \( b \): \( p = 0.2607 \), \( n = 10 \):

\[
P(\text{exactly 3}) = \binom{10}{3} (0.2607)^3 (1 - 0.2607)^7
\]

\[
= 0.2567
\]

d  A normal approximation is valid, since \( n \) is large (> 50) and \( p \) is close to 0.5. Therefore

\[
\mu = np = 150 \times 0.45 = 67.5 \quad \text{and} \quad \sigma = \sqrt{np(1-p)} = \sqrt{67.5 \times 0.55} = \sqrt{37.125} = 6.093 \quad (4 \text{ s.f.})
\]

\[
P(X > 75) \approx P(N > 75.5) = 0.0946 \quad (3 \text{ s.f.})
\]

18 a  Using the binomial distribution, \( P(B = 35) = \binom{80}{35} 0.48^{35} \times 0.52^{45} = 0.06703 \).

\( b \)  A normal approximation is valid, since \( n \) is large (> 50) and \( p \) is close to 0.5. Therefore

\[
\mu = np = 80 \times 0.48 = 38.4 \quad \text{and} \quad \sigma = \sqrt{np(1-p)} = \sqrt{38.4 \times 0.52} = \sqrt{19.968} = 4.469 \quad (4 \text{ s.f.})
\]

\[
P(B = 35) \approx P(34.5 < N < 35.5) = 0.0668 \quad (3 \text{ s.f.})
\]

Percentage error is

\[
\frac{0.06703 - 0.0668}{0.06703} = 0.34%.
\]

19  Remember to identify which is \( H_0 \) and which is \( H_1 \). This is a one-tail test since we are only interested in whether the time taken to solve the puzzle has reduced. You must use the correct parameter (\( \mu \)):

\( H_0 \): \( \mu = 18 \) \quad \( H_1 \): \( \mu < 18 \)

Using

\[
z = \frac{x - \mu}{\sigma}, \quad z = \frac{(16.5 - 18)}{\frac{1}{\sqrt{15}}} = -1.9364...
\]

Using the percentage point table and quoting the figure in full:

5% one tail c.v. is \( z = -1.6449 \)

\(-1.9364 < -1.6449\), so significant or reject \( H_0 \) or in critical region.

State your conclusion in the context of the question:

There is evidence that the (mean) time to complete the puzzles has reduced.

Or Robert is getting faster (at doing the puzzles).
20 a \( P(Z < a) = 0.05 \Rightarrow -1.645. \) Using that \( P(L < 1.7) = 0.05 \) means that
\[
\frac{1.7 - \mu}{0.4} = -1.645 \Rightarrow 1.7 - \mu = -0.658 \Rightarrow \mu = 2.358
\]

b \( P(L > 2.3) = 0.5576 \) (4 s.f.) and so, using the binomial distribution,
\[
P(B \geq 6) = 1 - P(B \leq 5) = 1 - 0.4758 = 0.5242 \) (4 s.f.).

c It is thought that the mean length of the female rattlesnakes is 1.9 m, and a hypothesis test is needed to conclude whether the mean length is not equal to 1.9 m. Therefore,
\[H_0 : \mu = 1.9\]
\[H_1 : \mu \neq 1.9\]
Sample size: 20. Therefore, the sample population is initially thought to have distribution
\[
\bar{M} \sim N\left(1.9, \frac{0.3^2}{20}\right).
\]
By using the inverse normal distribution,
\[
P(\bar{M} < 1.768) = 0.025\)
\[
P(\bar{M} > 2.032) = 0.025\), meaning that the critical region is below 1.768 and above 2.032

d There is sufficient evidence to reject \( H_0 \), since 2.09 > 2.032; i.e. there is sufficient evidence to say, at the 5% level, that the mean length of the female rattlesnakes is not equal to 1.9 metres.

21 It is thought that the daily mean temperature in Hurn is less than 12 °C, and so a hypothesis test is needed to conclude whether, at the 5% level of significance, the mean temperature is less than 12 °C. Therefore,
\[H_0 : \mu = 12\]
\[H_1 : \mu < 12\]
Sample size: 20. Therefore, the sample population is initially thought to have distribution
\[
\bar{T} \sim N\left(12, \frac{2.3^2}{20}\right).
\]
By using the inverse normal distribution,
\[
P(\bar{T} < 11.154) = 0.05,\) meaning that the critical region consists of all values below 11.154. Since 11.1 < 11.154, there is sufficient evidence to reject \( H_0 \); i.e. there is sufficient evidence to say, at the 5% level, that the mean daily temperature in Hurn is less than 12 °C.

Challenge

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

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

2 a  We wish to use a hypothesis test to determine (at the 10% significance level) whether the support for the politician is 53%. A normal distribution is suitable, and we use the model given by

\[ \mu = np = 300 \times 0.53 = 159 \quad \text{and} \quad \sigma = \sqrt{np(1-p)} = \sqrt{159 \times 0.47} = \sqrt{74.73} = 8.645 \quad (4 \text{ s.f.}). \]

Therefore,

\[ H_0 : \mu = 159 \]

\[ H_1 : \mu \neq 159 \]

By using the inverse normal distribution, \( P(\bar{X} < 144.78) = 0.05 \) and \( P(\bar{X} > 173.22) = 0.05 \) (2 d.p.) and so the critical region consists of the values below 144.78 and above 173.22.

b Since 173 is not within the critical region, there is not sufficient evidence to reject \( H_0 \) at the 10% significance level; i.e. there is not sufficient evidence to say, at the 10% level, that the politician’s claim that they have support from 53% of the constituents is false.