Representing multiple outcomes: random variables and probability distributions

Skills check

1. Mean = \[ \frac{1\times9 + 2\times7 + 3\times3 + 6\times2 + 11\times1}{9 + 7 + 3 + 2 + 1} = \frac{55}{22} = 2.5 \]

2. Using technology, the solution is \( x = 1.39 \)

Exercise 13A

1. a. Not a discrete distribution as the sum of the probabilities is greater than 1
   b. Not a discrete distribution as one of the probabilities is negative
   c. This is a discrete probability distribution

2. a. \[ P(A = 12) = 1 - (0.5 + 0.05 + 0.04 + 0.1 + 0.2) = 0.11 \]
   b. \[ P(8 < A \leq 10) = P(A = 9) + P(A = 10) = 0.04 + 0.1 = 0.14 \]
   c. \[ P(A \leq 9) = P(A = 5) + P(A = 8) + P(A = 9) = 0.5 + 0.05 + 0.04 = 0.59 \]
   d. \[ P(A \geq 10) = P(A = 10) + P(A = 11) + P(A = 12) = 0.1 + 0.2 + 0.11 = 0.41 \]
   e. \[ P(A > 8 \mid A \leq 11) = \frac{P(9 \leq A \leq 11)}{P(A \leq 11)} = \frac{0.04 + 0.1 + 0.2}{0.5 + 0.05 + 0.04 + 0.1 + 0.2} = 0.38 \]
   f. \[ E(A) = 5\times0.5 + 8\times0.05 + 9\times0.04 + 10\times0.1 + 11\times0.2 + 12\times0.11 = 7.78 \]

3. a. \[ P(B = 1) = 0.0001, P(B = 2) = 0.9999\times0.0001 = 0.00009999, \]
   \[ P(B = 3) = 0.9999\times0.9999\times0.0001 = 0.00009998 \]
   b. To win on your \( b \)-th crisp packet, you need to have had \( (b - 1) \) losses and then a win on the \( b \)-th try, so \[ P(B = b) = P(\text{lose}^{b-1} \times \text{win}) = 0.0001 \times (0.9999)^{b-1} \]
   c. The domain is the set of positive integers, \( 1, 2, ... \) That is, \( b \in \mathbb{Z}^+ \).
   d. \[ P(B \leq 10) = P(B = 1) + \cdots + P(B = 10) = 0.00099955 \]

4. a. \[
\begin{array}{|c|c|c|c|c|}
\hline
& 1 & 2 & 3 & 4 & 5 \\
\hline
P(C = c) & 0.07 & 0.02 & 0.17 & 0.46 & 0.28 \\
\hline
\end{array}
\]

b. \[ E(C) = 1\times0.07 + 2\times0.02 + 3\times0.17 + 4\times0.46 + 5\times0.28 = 3.86 \]
5 a

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>16</td>
</tr>
</tbody>
</table>

d

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>9</th>
<th>12</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(d)</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P(D = d)</td>
<td>1/16 = 0.0625</td>
<td>1/8 = 0.125</td>
<td>1/8 = 0.125</td>
<td>3/16 = 0.1875</td>
<td>1/8 = 0.125</td>
<td>1/8 = 0.125</td>
<td>1/16 = 0.0625</td>
<td>1/16 = 0.0625</td>
<td></td>
</tr>
</tbody>
</table>

b $P(D \text{ is a square number} \mid D < 8) = \frac{P(D \text{ is a square number} \cap D < 8)}{P(D < 8)} = \frac{P(D = 1) + P(D = 4)}{P(D = 1) + \cdots + P(D = 6)} = \frac{0.0625 + 0.1875}{0.0625 + \cdots + 0.125} = \frac{2}{5} = 0.4$

6

<table>
<thead>
<tr>
<th>No of heads</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td>TTT</td>
<td>TTH, THT, HTT</td>
<td>THH, HTH, HHT</td>
<td>HHH</td>
</tr>
<tr>
<td>m</td>
<td>2</td>
<td>γ</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>P(M = m)</td>
<td>1/8</td>
<td>3/8</td>
<td>3/8</td>
<td>1/8</td>
</tr>
</tbody>
</table>

a $E(M) = 2 \times \frac{1}{8} + \gamma \times \frac{3}{8} + 5 \times \frac{3}{8} + 15 \times \frac{13}{8} = 4 + \frac{3}{8} \gamma$

b To be fair, we want $E(M) = \$7$, so $7 = 4 + \frac{3}{8} \gamma \Rightarrow \gamma = 8$

7 a

<table>
<thead>
<tr>
<th>f</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(F = f)</td>
<td>4/7 × 3/6 = 2/7</td>
<td>4/7 × 3/6 + 3/7 × 4/6 = 4/7</td>
<td>3/7 × 2/6 = 1/7</td>
</tr>
</tbody>
</table>

b $E(F) = 0 \times \frac{2}{7} + 1 \times \frac{4}{7} + 2 \times \frac{1}{7} = \frac{6}{7}$

8 a i $P(G = 1) = \frac{1}{5} = \frac{9}{45}$, $P(G = 2) = \frac{8}{10} \times \frac{2}{9} = \frac{8}{45}$, $P(G = 3) = \frac{8}{10} \times 7/9 \times \frac{2}{8} = \frac{7}{45}$, ...
Thus, \( P(G = g) = \frac{10 - g}{45} \) for \( g = 1, \ldots, 9 \) (it is not possible to pick the first black one on the 10th go).

ii \( E(G) = 1 \times \frac{1}{5} + 2 \times \frac{8}{45} + \ldots + 8 \times \frac{2}{45} + 9 \times \frac{1}{45} = \frac{11}{3} \)

b i \( P(G = 1) = \frac{1}{5} \), \( P(G = 2) = \frac{8}{10} \times \frac{2}{10} = \frac{4}{25} \), \( P(G = 3) = \frac{8}{10} \times \frac{8}{10} \times \frac{2}{10} = \frac{16}{125} \), \ldots,

\( P(G = g) = \left( \frac{8}{10} \right)^{g-1} \times \frac{2}{10} \) for any \( g = 1, 2, \ldots \)

ii \( E(G) = 1 \times \frac{1}{5} + 2 \times \frac{4}{25} + \ldots + g \times \left( \frac{8}{10} \right)^{g-1} \times \frac{2}{10} + \ldots = 5 \)

c For a: This is an arithmetic series: \( \sum_{g=1}^{9} P(G = g) = \sum_{g=1}^{9} \left( \frac{10 - g}{45} \right) = \frac{9}{2} \left( \frac{9}{45} + \frac{1}{45} \right) = 1 \)

For b: This is a geometric series: \( \sum_{g=1}^{9} P(G = g) = \sum_{g=1}^{9} \left( \frac{8}{10} \right)^{g-1} \times \frac{2}{10} = \frac{1}{5} \left( \frac{1}{1 - \left( \frac{4}{5} \right)^{10}} \right) = 1 \)

9 a \( P(B \geq x + y | B \geq x) = \frac{P(B \geq x + y)}{P(B \geq x)} = \frac{1 - P(B \leq x + y - 1)}{1 - P(B \leq x - 1)} \)

\( = \frac{1 - (1 - (1 - p)^{y-1})}{1 - (1 - (1 - p)^{x-1})} = (1 - p)^y = 1 - F(y) = P(B \geq y + 1) \)

b This means that the process has no "memory". The probability of finding a golden ticket after at least 15 trials given that no golden ticket was found after at least 10 has the same probability as starting over.

Exercise 13B

1 a \( P(X = 4) = 0.0535 \)

b \( P(X \leq 4) = 0.991 \)

c \( P(1 \leq X < 4) = P(X = 1) + P(X = 2) + P(X = 3) = 0.809 \)

An alternative is to find \( P(X \leq 3) - P(X \leq 0) \)

d \( P(X \geq 2) = 0.558 \)

On some calculators this needs to be calculated as \( 1 - P(X \leq 1) \)

e \( P(X \leq 4 | X \geq 2) = \frac{P(X \leq 4 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(2 \leq X \leq 4)}{P(X \geq 2)} \)

\( = 0.983 \)

Depending on the GDC the numerator in the last expression can be calculated directly or as \( P(X = 2) + P(X = 3) + P(X = 4) \) or as \( P(X \leq 4) - P(X \leq 2) \)

f \( X \leq 4 \) and \( X \geq 2 \) are not independent because \( P(X \leq 4 | X \geq 2) = 0.9833 \neq 0.9907 = P(X \leq 4) \)
2 \[ P(\text{prime}) = \frac{n(2, 3, 5, 7)}{n(1, 2, 3, 4, 5, 6, 7, 8)} = \frac{1}{2}, \]
\[ P(\text{at least three primes}) = P(\text{number of primes} \geq 3) = 0.773 \]

On some GDCs this can be calculated as
\[ P(\text{at least three primes}) = 1 - P(\text{number of primes} \leq 2) = 0.773 \]

3 a \[ E(Y) = 5.2 = n \times 0.4 \Rightarrow n = 13 \Rightarrow \text{Var}(Y) = 13 \times 0.4 \times 0.6 = 3.12 \]

b \[ \text{Var}(Z) = 1.44 = 9 \times p \times (1-p) \Rightarrow \frac{9p^2 - 9p + 36}{25} = 0 \Rightarrow p = \frac{1}{5} \text{ or } p = \frac{4}{5} \]

4 a \[ \text{Distribution is } B(10, 0.1) \]
\[ P(X \geq 5) = 1 - P(X \leq 4) = 0.00163 \]

b Probability of no points in a single game = \( P(X = 0) = 0.3486... \)

Let \( Y \) be the number of times in 6 games that no points are scored.

Distribution of \( Y \) is \( B(6, 0.3486...) \)
\[ P(\text{no points in at least 2 games}) = P(Y \geq 2) = 1 - P(Y \leq 1) \]
\[ = 1 - 0.32156... \]
\[ \approx 0.678 \]

5 a \[ A \sim B(25, 0.2) \]

b \[ P(A \leq 5) = 0.617 \]

c \[ P(A \geq 7) = 1 - P(A \leq 6) = 0.220 \]

d \[ P(A \leq 3) = 0.234 \]

e \[ E(A) = 25 \times 0.2 = 5 \text{ so it is expected that he will guess 5 questions correctly} \]

f \[ P(A > 5) = 1 - P(A \leq 5) = 1 - 0.617 = 0.383 \]

g \[ \text{Expected points} = 4 \times 5 + (-1) \times 20 = 0 \]

h \[ P(\text{at least 2 get 7 or more}) = 0.212 \]

6 a \[ T \sim B(538, 0.91) \text{ assuming that the arrival of a passengers is independent of the arrival of any other passenger.} \]

b \[ P(T = 538) = 0.91^{538} = 9.21 \times 10^{-23} \text{; it is exceedingly unlikely that all passengers will turn up on time to take the flight} \]

c \[ P(T \geq 510) = 0.000672 \text{; it is very likely that there will be at least 28 empty seats on the plane} \]

d Consider the distribution \( T_n \sim B(n, 0.91) \) where \( S \) is the number of passengers who turn up when \( n \) tickets are sold.

We need to find the smallest value of \( n \) such that \( P(T_n \geq 510) \geq 0.1 \)
This can be entered into the function menu of the GCC as $1 - \text{Binomcdf}(x, 0.91, 509)$ (or similar depending on the names given to the cumulative distribution function on the GDC). The Table can then be used to find the smallest value of $x$ (or $n$) for which this is greater than 0.1

$n = 550, P(T \geq 510) = 0.0870$. For $n = 551, P(T \geq 510) = 0.112$.

e $E(T) = 538 = n \times 0.91 \Rightarrow n = 591$

f $P(T = 538) = 0.0573$

$P(T > 538) = P(X \geq 539) = 1 - P(X \leq 538) = 0.468$

It is now much more likely that the plane will be full, and it is also very likely that the plane will be overbooked.

7 $\text{Var}(X) = np(1 - p)$. Differentiating to find the max, $n(1 - 2p) = 0 \Rightarrow p = \frac{1}{2}$

8 a Let $X$ be the number of times 1 is scored when the 3 spinners are spun hence $X \sim B(3, 0.2)$

Expected frequencies:

- 0 1s thrown: $200 \times P(X = 0) = 102.4$
- 1 1s thrown: $200 \times P(X = 1) = 76.8$
- 2 1s thrown: $200 \times P(X = 2) = 19.2$
- 3 1s thrown: $200 \times P(X = 3) = 1.6$

b There are large differences between the expected numbers and the actual recorded numbers, indicating that the dice may not be fair.

**Exercise 13C**

1 a $P(A = 2) = 0.0390$

b $P(A < 6) = P(A \leq 5) = 0.414$

c $P(A \geq 7 \mid A > 5) = \frac{P(A \geq 7 \cap A > 5)}{P(A > 5)} = \frac{P(A \geq 7)}{1 - P(A \leq 5)} = \frac{1 - P(A \leq 6)}{1 - P(A \leq 5)} = 0.727$

d $P(A \leq 4) = 0.259$

e $P(A > 8 \mid A \geq 3) = \frac{P(A > 8 \cap A \geq 3)}{P(A \geq 3)} = \frac{P(A > 8)}{1 - P(A \leq 2)} = \frac{1 - P(A \leq 8)}{1 - P(A \leq 2)} = 0.184$

2 $\frac{\beta^8e^{-\beta}}{0!} = 0.301 \Rightarrow \beta = 1.201 \Rightarrow \text{Var}(Z) = \beta = 1.20$

or solve $\text{Poissonpdf}(\beta, 0) = 0.301$ using the graph or equation solving facility on your GDC

3 $\frac{\alpha^2e^{-\alpha}}{1!} = 0.15 \Rightarrow \alpha = 0.179$ or 2.99.

or solve $\text{Poissonpdf}(\alpha, 1) = 0.15$ using the graph or equation solving facility on your GDC

For $\alpha = 0.179, P(Y = 1) = 0.15$.

For $\alpha = 2.99, P(Y = 1) = 0.15$. 
The maximum probability will be close to the mean. For $\alpha = 2.99$, $Y = 1$ is before the mean as the probabilities are increasing and for $\alpha = 0.179$, $Y = 1$ is after the mean as they decrease.

4 S $\sim$ Po(27.8). It is assumed that the seeds are distributed randomly throughout the 10 loaves. The most likely number of seeds per loaf is either 27 or 28. Since $P(S = 27) = 0.07566$ and $P(S = 28) = 0.07512$, then the most likely number of seeds in a loaf is 27.

5 To make this fair, need $E(\text{Consequence}) = 0$ so we need to find $x$ such that

$$0 = -10000 \times P(C \geq 4) + 0 \times P(1 \leq C \leq 3) + x \times P(C = 0)$$

$$= -10000 \times (1 - P(C \leq 3)) + x \times P(C = 0)$$

$$= -2026.89 + 0.09926x$$

$$\Rightarrow x = \£20420$$

6 The mean number of calls is $\frac{0 \times 15 + 1 \times 30 + 2 \times 28 + 3 \times 14 + 4 \times 7 + 5 \times 8}{15 + 30 + 28 + 14 + 7 + 8} = \frac{196}{102} = 1.9216$, so the following table shows the expected frequencies if the data follows a Poisson distribution.

<table>
<thead>
<tr>
<th>Number of calls</th>
<th>Expected number of hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$102 \times P(X = 0)$</td>
</tr>
<tr>
<td></td>
<td>$= 14.9$</td>
</tr>
<tr>
<td>1</td>
<td>$102 \times P(X = 1)$</td>
</tr>
<tr>
<td></td>
<td>$= 28.7$</td>
</tr>
<tr>
<td>2</td>
<td>$102 \times P(X = 2)$</td>
</tr>
<tr>
<td></td>
<td>$= 27.6$</td>
</tr>
<tr>
<td>3</td>
<td>$102 \times P(X = 3)$</td>
</tr>
<tr>
<td></td>
<td>$= 17.7$</td>
</tr>
<tr>
<td>4</td>
<td>$102 \times P(X = 4)$</td>
</tr>
<tr>
<td></td>
<td>$= 8.48$</td>
</tr>
<tr>
<td>5 or more</td>
<td>$102 \times P(X = 5)$</td>
</tr>
<tr>
<td></td>
<td>$= 3.26$</td>
</tr>
</tbody>
</table>

Indicating that the data is well modelled by a Poisson distribution except for “5”. This might be because these calls to the helpline are not independent and are due to a problem affecting several computers at the same time.

7 a Let $A \sim \text{Po}(4.2 \times 2)$, then

$$P(A = 0 \cap B = 0) = P(A = 0) \times P(B = 0) = 7.50 \times 10^{-6}$$

b Let $C \sim \text{Po}(4.2 \times 1.7 \times 2)$, then $P(C = 0) = 7.50 \times 10^{-6}$

8 a Let $A \sim \text{Po}(5 \times 0.597)$, then $P(A = 0) = 0.0505$

b Let $B \sim \text{Po}(10 \times 24 \times 0.597)$, then

$$P(B \geq 3) = 1 - P(B \leq 2) = 1$$

c Let $C \sim \text{Po}(24 \times 0.597)$, then $P(C < 11) = P(C \leq 10) = 0.1549$

Let $X$ represent the number of days in a week on which there are fewer than 11 breakdowns. Hence $X \sim B(7,0.1549)$

$$P(X \geq 5) = 1 - P(X \leq 4) = 0.00142$$

9 a For $R \sim \text{Po}(5.1)$, $365 \times P(R = 0) = 2.23$

b $100 \times P(R > 10) = 100 \times (1 - P(R \leq 10)) = 1.56$
Exercise 13D

1 It is known that $P(X > \mu) = P(X < \mu) = 0.5$

From investigation 10 we have

<table>
<thead>
<tr>
<th>If $X \sim N(\mu, \sigma^2)$, then:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$</td>
<td></td>
</tr>
<tr>
<td>$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$</td>
<td></td>
</tr>
<tr>
<td>$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997$</td>
<td></td>
</tr>
</tbody>
</table>

a $P(T < 17.1) = P(T < \mu) = 0.5$

b $P(T < 14) = P(T < \mu - \sigma) = \frac{1 - 0.68}{2} = 0.16$

c $P(T > 20.2) = P(T > \mu + \sigma) = \frac{1 - 0.68}{2} = 0.16$

d $P(14 \leq T < 23.3) = P(\mu - \sigma \leq T < \mu + 2\sigma)$

$= P(T < \mu + 2\sigma) - P(T < \mu - \sigma) = 0.95 + 0.025 - \frac{1 - 0.68}{2} = 0.82$

e $P(T < 7.8) = P(T < \mu - 3\sigma) = \frac{1 - 0.997}{2} = 0.0015$

f $P(T < 23.3 \mid T > 20.2) = \frac{P(\mu + \sigma < T < \mu + 2\sigma)}{P(T > \mu + \sigma)}$

$= \frac{P(T < \mu + 2\sigma) - P(T < \mu + \sigma)}{P(T > \mu + \sigma)} = \frac{0.975 - 0.84}{0.16} = 0.84$

2 a $P(Q < 4) = 0.483$

b $P(Q < 3.4) = 0.184$

c $P(Q > 5) = 0.0829$

On some GDCs this is calculated using $P(Q > 5) = 1 - P(Q < 5)$

d $P(3.5 \leq Q < 4.5) = 0.525$

On some GDCs this is calculated using $P(Q < 4.5) - P(Q < 3.5)$

e $P(Q < 4.9 \mid Q > 2.9) = \frac{P(2.9 < Q < 4.9)}{P(Q > 2.9)} = 0.887$

On some GDCs this is calculated using $P(Q < 4.9) - P(Q < 2.9)$

3 If $F(a) = P(X < a) = p$ then define $F^{-1}(p) = a$ This is normally referred to as the inverse normal function on a GDC. You will need to enter the value for $p$ as well as the mean and standard deviation.

$Q_3 = F^{-1}(0.75) = 22331.3$
\[ Q_i = F^{-1}(0.25) = 21926.7 \]

\[ IQR = Q_3 - Q_1 = 22331.3 - 21926.7 = 405 \]

4  \[ P(|S| > k) = 0.57 \Rightarrow P(S < -k) = 0.285, \text{ so } -k = F^{-1}(0.285) \Rightarrow k = 0.767 \]

5  a  \[ P(X \leq 381) = 0.756 \]

\[ P(X > t) = 0.17 \Rightarrow P(X < t) = 1 - 0.17 = 0.83 \Rightarrow t = F^{-1}(0.83) = 0.384 \]

6  \[ L \sim N(182, 10^2) \]

a  \[ P(L > 190) = 0.212 \]

b  Let \( X \) be the number of batteries in the sample that last longer than 190 days. Hence \( X \sim B(7, 0.2119...) \)

\[ P(X \leq 3) = 0.959 \]

c  \[ P(L \leq 165) = 0.0446 \]

d  \[ 10000 \times 0.04457 = 446 \]

7  \[ D \sim N(16, 5^2) \]

a  \[ P(13 < D < 15.3) = 0.170 \]

b  \[ P(D > x) = 0.13 \Rightarrow P(D < x) = 0.87 \Rightarrow x = F^{-1}(0.87) = 21.6 \]

c  \[ 23109 \times P(D > 14) = 23109 \times 0.6554 = 15146 \]

Of these 15146 employees who live more than 14 km from work, \( 0.91 \times 15146 = 13783 \) will fail to get to work.

8  \[ P(M \leq m) = 0.99 \Rightarrow m = F^{-1}(0.99) = 34.6 \text{ so 35 minutes to be sure of meeting the target.} \]

9  a  Route A is on average shorter than route B, but has more variability, so the nurse will have to allow time for a longer journey. Route B takes on average longer but has less variability so the actual times are closer to the average.

b  \[ A \sim N(42, 8^2), \ B \sim N(50, 3^2) : \]

\[ P(A < 45) = 0.6462 \]

\[ P(B < 45) = 0.04779 \]

He should take route A.
c Let $C$ be the number of days on which he arrives on time. $C \sim B(5, 0.6462)$

i $P(C = 5) = 0.6462^5 = 0.113$

ii $P(C \geq 3) = 1 - P(C \leq 2) = 0.759$

iii Let $T$ represent the event arriving on time. There are three possibilities for the three consecutive days $TTT$, $T'T'T'$ and $T'TT$

Each of these have the same probability so $P(\text{On time on 3 consecutive days}) = 3 \times 0.6462^3 \times 0.3538^2 = 0.101$

Exercise 13E

1 a $6 = E(3F + 1) = 3E(F) + 1 \Rightarrow E(F) = \frac{5}{3}$

b $7 = \text{Var}(3.1 - 2F) = (-2)^2 \times \text{Var}(F) \Rightarrow \text{Var}(F) = \frac{7}{4}$

2 Because these have a Poisson distribution, $E(C) = \text{Var}(C)$ and $E(D) = \text{Var}(D)$.

a $E(9C - 4D) = 9E(C) - 4E(D) = 9 \times 3 - 4 \times 6.1 = 2.6$

b $\text{Var}(D + 0.2C) = \text{Var}(D) + 0.2^2 \times \text{Var}(C) = 6.1 + 0.04 \times 3 = 6.22$

3 a $E(2U + 9) = 2E(U) + 9 = 2 \times 4.01 + 9 = 17.02$

b $E(4X - 0.1) = 4E(X) - 0.1 = 4 \times 7.81 - 0.1 = 31.14$

c $\text{Var}(0.9W + 10) = 0.9^2 \times \text{Var}(W) = 0.81 \times 3 = 2.43$

d $\text{Var}(7 - 3V) = 3^2 \times \text{Var}(V) = 9 \times 0.4 = 3.6$

e $E(U + 4X - 2W) = E(U) + 4E(X) - 2E(W) = 4.01 + 4 \times 7.81 - 2 \times 7.81 = 9.45$

f $\text{Var}(2V + 0.8U - 0.9X + W) = 2^2 \times \text{Var}(V) + 0.8^2 \times \text{Var}(U) + (-0.9)^2 \times \text{Var}(X) + \text{Var}(W) = 4 \times 0.4 + 0.64 \times 1.2 + 0.81 \times 2.11 + 3 = 7.08$

4 $5.1 = E(aU + b) = aE(U) + b = 4.01a + b$ and $2 = \text{Var}(aV + b) = a^2 \text{Var}(V) = 0.4a^2$. Solving these gives $a = 2.24$ and $b = -3.87$

5 $6 = E(aW + bX) = aE(W) + bE(X) = 12.9a + 7.81b$ and $2 = E(aU + bV) = aE(U) + bE(V) = 4.01a + 2.7b$. Solving these gives $a = 0.165$ and $b = 0.496$

6 $15.1 = E(aW + bX) = aE(W) + bE(X) = 12.9a + 7.81b$ and $2 = \text{Var}(aU + bV) = a^2 \text{Var}(U) + b^2 \text{Var}(V) = 1.2a^2 + 0.4b^2$. Solving these gives $a = 1.29$ and $b = -0.191$

7 a

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(T = t)$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
</tr>
</tbody>
</table>
Worked solutions

<table>
<thead>
<tr>
<th>h</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(H = h)</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{2} )</td>
</tr>
</tbody>
</table>

**b i** \( E(T) = 1 \times \frac{1}{3} + 2 \times \frac{2}{3} = \frac{5}{3} \)

**ii** \( E(H) = 3 \times \frac{1}{4} + 4 \times \frac{1}{4} + 5 \times \frac{1}{2} = \frac{17}{4} \)

**c** \( E(S) = E(T + H) = E(T) + E(H) = \frac{5}{3} + \frac{17}{4} = \frac{71}{12} \)

**d**

<table>
<thead>
<tr>
<th>s</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(S = s)</td>
<td>( \frac{1}{12} )</td>
<td>( \frac{1}{4} )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
</tr>
</tbody>
</table>

8 Let \( Z_1 \) and \( Z_2 \) be the results shown on dice 1 and 2, respectively.

Game 1:
\[
E(K) = E(Z_1 + Z_2) = E(Z_1) + E(Z_2) = \frac{3+4+5+5}{4} + \frac{3+4+5+5}{4} = \frac{17}{2}
\]
\[
Var(K) = Var(Z_1 + Z_2) = Var(Z_1) + Var(Z_2) = \frac{11}{16} + \frac{11}{16} = \frac{11}{8}
\]
Outcomes = \{6,7,8,9,10\}

Game 2:
\[
E(L) = E(2Z_i) = 2E(Z_i) = 2 \times \frac{17}{4} = \frac{17}{2}
\]
\[
Var(L) = Var(2Z_i) = 4Var(Z_i) = 4 \times \frac{11}{16} = \frac{11}{4}
\]
Outcomes = \{6,8,10\}

The variances and outcomes are different, so the distributions are different too and Zeinab is right.

**Exercise 13F**

1 \( I \sim Po(0.7) \) and \( M \sim Po(0.6) \), so \( I + M = W \sim Po(0.7 + 0.6) \)
\[
P(W \geq 1) = 1 - P(W = 0) = 0.728
\]

2 \( M \sim N(85,10^2) \) and \( F \sim N(60,7^2) \), so
\[
M_1 + M_2 + F_1 + F_2 + F_3 = T \sim N(2 \times 85 + 3 \times 60, 2 \times 10^2 + 3 \times 7^2) = N(350,347)
\]
\[
P(T > 375) = 0.0898
\]

3 \( P \sim N(3 \times 300 + 2 \times 180 + 2 \times 100 + 16,3 \times 2^2 + 2 \times 7^2 + 2 \times 2^2 + 0.5^2) = N(1476,118.25), \)
\[
P(P > 1500) = 0.0137
\]

4 \( E \sim Po(24 \times 1.5 + 2) = Po(38) \)

© Oxford University Press 2019
\[ P(E > 40) = 1 - P(E \leq 40) = 0.334 \]

5 a Let \( X \) be the difference in length between the length of one short and one regular together and one long

\[ X \sim N(40 + 80 - 120, 2.1^2 + 3.7^2 + 4^2) = N(0, 34.1) \]

\[ P(X > 0) = 0.5 \]

b \( X \sim N(3 \times 40 - 120, 3 \times 2.1^2 + 4^2) = N(0, 29.23) \)

\[ P(X < 0) = 0.5 \]

6 a By the central limit theorem

\[ E(\bar{A}) = E(A) = 110 \quad \text{and} \quad \text{Var}(\bar{A}) = \frac{\text{Var}(A)}{40} = 15.625 \]

\[ E(\bar{B}) = E(B) = 123 \quad \text{and} \quad \text{Var}(\bar{B}) = \frac{\text{Var}(B)}{40} = 1.640 \]

Thus, \( \bar{A} - \bar{B} \sim N(110 - 123, 15.625 + 1.640) = N(-13, 17.265) \)

\[ P(\bar{A} - \bar{B} > 0) = 0.00878 \]

b \( \bar{A} \sim N(110, 15.625) \), so \( P(108 < \bar{A} < 112) = 0.387 \)

c \( \bar{B} \sim N\left(123, \frac{8.1^2}{50}\right) = N(123, 1.31) \), so \( P(\bar{B} > 125) = 0.0404. \)

So the expected number of samples is \( 150 \times 0.0404 = 6.06 \)

7 \( 2C - H \sim N(2 \times 5.1 - 10, 2^2 \times 0.1^2 + 0.1^2) = N(0.2, 0.05) \)

\[ P(2C - H < 0) = 0.186 \]

8 \( \bar{A} \sim N\left(172, \frac{7^2}{25}\right) \)

\[ P(\bar{A} > 175) = 0.0161 \]

\( \bar{E} \sim N\left(170, \frac{7^2}{25}\right) \)

\[ P(\bar{E} > 175) = 0.000178 \]

9 \( E(\bar{C}) = 170 \) and \( \text{Var}(\bar{A}) = \frac{6^2}{n} \)

This can be solved by educated guessing or by putting a function similar to following into the GDC and using the table to find the value of \( n \).

\[
F = \text{normcdf}\left(172, 170, \frac{6}{\sqrt{n}}\right) - \text{normcdf}\left(168, 170, \frac{6}{\sqrt{n}}\right)
\]
If \( n = 24 \) then \( P(168 < \bar{C} < 172) = 0.898 \)

If \( n = 25 \) then \( P(168 < \bar{C} < 172) = 0.904 \)

Therefore, \( n = 25 \)

**Chapter Review**

1 a  \( E(D) = 1.8 = 0 \times 0.3 + 1 \times (p + q) + 2 \times 0.15 + 3 \times (p - q) + 4 \times (p + 2q) = 0.3 + 8p + 6q \) and 
1 = 0.3 + p + q + 0.15 + p − q + p + 2q = 0.45 + 3p + 2q 
Simplifying 1.5 = 8p + 6q and 0.55 = 3p + 2q 
solving simultaneously gives \( p = 0.15 \) and \( q = 0.05 \)

b \( P(D = 3 \mid D > 1) = \frac{P(D = 3)}{P(D > 1)} = \frac{0.1}{0.5} = 0.2 \)

2 a

<table>
<thead>
<tr>
<th>( t )</th>
<th>( P(T = t) )</th>
<th>( t )</th>
<th>( P(T = t) )</th>
<th>( t )</th>
<th>( P(T = t) )</th>
<th>( t )</th>
<th>( P(T = t) )</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>( \frac{1}{36} )</td>
<td>5</td>
<td>( \frac{1}{9} )</td>
<td>6</td>
<td>( \frac{2}{9} )</td>
<td>7</td>
<td>( \frac{5}{18} )</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>( \frac{2}{9} )</td>
<td>10</td>
<td>( \frac{1}{36} )</td>
<td>Payment</td>
<td>x</td>
<td>-5</td>
<td>0</td>
<td>-7</td>
</tr>
</tbody>
</table>

b \( \frac{1}{36} \times x + \frac{(-5)}{9} + \frac{(-7) \times 5}{18} + \frac{1}{9} \times x = 0 \Rightarrow x = 18 \)

3 a Factors of 120: 1\times120 = 120, 2\times60 = 120, 3\times40 = 120, 4\times30 = 120, 5\times24 = 120, 6\times20 = 120, 8\times15 = 120, 10\times12 = 120 so therefore 16 factors

b let \( F \) be the total number of numbers generated which are factors of 120. \( F \sim B \left( 6, \frac{16}{120} \right) \)

i \( P(F = 3) = 0.0309 \)

ii \( P(F \geq 3) = 1 - P(F \leq 2) = 0.0346 \)

iii \( P(F \leq 3) = 0.996 \)

iv There are 4 ways to get 3 consecutive factors of 120: If \( f \) is a number which is a factor of 120 the ways are \( f ff f f' f' f'', f' f f f f' f'' f', f' f'' f f f f', f' f' f' f f f' \)

\[ P( \text{factor of 120 on 3 consecutive rolls} ) = 4 \times \left( \frac{16}{120} \right)^3 \left( 1 - \frac{16}{120} \right)^3 = 0.00617 \]

4 a \( P(A \geq 3) = 1 - P(A \leq 2) = 0.123 \)

d \( P(A_1 + A_2) = 0.210 \)

c \( P(A = 0) = 0.2982, \) 

Number of days with no accidents in a week is \( W \sim B(7, 0.2982) \)
\[ P(\text{more than 4 days with no accidents}) \]
\[ = P(W > 4) = 1 - P(W \leq 4) = 0.0281 \]

5. a \( P(W > 70) = 0.325 \)

b Let \( Q_1 = a \) and \( Q_3 = b \) \( P(W < a) = 0.25 \Rightarrow a = 57.58 \), using the inverse normal function on the GDC
\[ P(W < b) = 0.75 \Rightarrow b = 72.42 \]
\[ IQR = 72.42 - 57.58 = 14.8 \]

c Need to find \( k \) such that \( P(W > k) = 0.073 \) \( k = 81.0 \) using the inverse normal function on the GDC. On some GDCs it is necessary to first write as \( P(W < k) = 0.927 \)

d Let \( N \) be the number of boys who weigh at least 70kg. \( N \sim B(8, 0.325) \)
\[ P(N \leq 3) = 0.758 \]
\[ 1000 \times P(W < 60) = 325 \]

6. a \( F - 3R \sim N \left( 430 - 3 \times 160, 9^2 + 9 \times 5^2 \right) \)
\[ P(F - 3R < 0) = 0.998 \]

b \( F - R - R - R \sim N \left( 430 - 3 \times 160, 9^2 + 3 \times 5^2 \right) \)
\[ P(F - R - R - R > 0) = 0.0000313 \]

7. Let \( Y \) be the value where the ball touches the \( y \)-axis. \( Y = 2F + 3 \sim N(7.6, 4 \times 0.3) \)
\[ E(P) = P(5 < Y \leq 6) \times 1 + P(6 < Y \leq 8) \times 5 + P(8 < Y \leq 9) \times 10 \]
\[ = 5.484... \]
\[ 127 \times E(P) = 127 \times 5.484... = 696 \]

8. a \( E(C) = E(A - B) = E(A) - E(B) = 3.1 - 4.7 = -1.6 \) and
\[ \text{Var}(C) = \text{Var}(A - B) = \text{Var}(A) + \text{Var}(B) = 3.1 + 4.7 = 7.8 \] so \( C \) is not Poisson because \( E(C) \neq \text{Var}(C) \).

\( D \) is not Poisson as its domain includes only values greater than or equal to 9.
\( E \) is Poisson because \( E(E) = E(A + B) = E(A) + E(B) = \text{Var}(A) + \text{Var}(B) = \text{Var}(A + B) = \text{Var}(E) \)
\[ E(F) = E(2.9B) = 2.9E(B) = 13.63 \text{ and } \text{Var}(F) = \text{Var}(2.9B) = 2.9^2 \text{Var}(B) = 39.527 \] so \( F \) is not Poisson because \( E(F) \neq \text{Var}(F) \).

b Here, \( B \sim Po(7) \). Need to find \( k \) such that \( P(B < k) = 0.99 \). If \( k = 13 \) then
\[ P(B < k) = 0.987 \] but if \( k = 14 \) then \( P(B < k) = 0.994 \) so 14 batteries should be in stock.

c Need to find \( \lambda \) such that
\[ P(U > 2) = 1 - P(U \leq 2) = 0.401, \]
This can be entered as an equation into the GDC and solved to give \( \lambda = 2.2888 \)
9 a \[ E(T) = E(X + Y + 2Z) = E(X) + E(Y) + 2E(Z) = 3.07 + 5 \times 0.27 + 2 \times 3.81 = 12.04 \]
\[ \text{Var}(T) = \text{Var}(X + Y + 2Z) = \text{Var}(X) + \text{Var}(Y) + 4\text{Var}(Z) = 0.8^2 + 5 \times 0.27 \times 0.73 + 4 \times 3.81 = 16.9 \]

b By the central limit theorem, \( T \sim N\left(12.04, \frac{16.9}{35}\right) \)
\[ P(11 \leq T < 13) = 0.850 \]

c This can be solved by educated guessing or by putting a function similar to following into the GDC and using the table to find the value of \( n \).
\[ F = \text{normcdf}\left(13, 12.04, \frac{16.9}{\sqrt{x}}\right) - \text{normcdf}\left(11, 12.04, \frac{16.9}{\sqrt{x}}\right) \]
If \( n = 65 \), \( P(11 \leq T < 13) = 0.9497 \) and if \( n = 66 \), \( P(11 \leq T < 13) = 0.9514 \)

Exam style questions

10 i \[ E(X) = 0 = (-3) \times \frac{1}{20} + (-2) \times \frac{2}{20} + (-1) \times a + 0 \times b + \frac{1}{20} \times 3 + \frac{2}{20} \times 2 + \frac{3}{20} \times 1 \]
\[ a = 0.15 \]

ii \[ 1 = \frac{1}{20} + \frac{2}{20} + a + b + \frac{3}{20} + \frac{2}{20} + \frac{1}{20} \]
\[ b = 0.4 \]

11 a \[ P(X = 5) = \frac{5^5 e^{-5}}{5!} = 0.1755 \]
\[ P(X < 5) = \frac{5^0 e^{-5}}{0!} + \cdots + \frac{5^4 e^{-5}}{4!} = 0.4405 \]

b \[ P(X < 5) = \frac{5^0 e^{-5}}{0!} + \cdots + \frac{5^4 e^{-5}}{4!} = 0.4405 \]

\[ P(W > 33) = 1 - P(W \leq 33) = 1 - \left(\frac{35 e^{-35}}{0!} + \cdots + \frac{35^{35} e^{-35}}{33!}\right) = 0.5898 \]

c \[ P(X + Y \leq 8) = \frac{9^6 e^{-9}}{0!} + \cdots + \frac{9^8 e^{-9}}{8!} = 0.4557 \]

d \[ E(Z) = E(3X + 2Y) = 3E(X) + 2E(Y) = 3 \times 5 + 2 \times 4 = 23 \]
\[ \text{Var}(Z) = \text{Var}(3X + 2Y) = 3^2\text{Var}(X) + 2^2\text{Var}(Y) = 9 \times 5 + 4 \times 4 = 61 \]

f Does not satisfy Poisson distribution as \( E(Z) \neq \text{Var}(Z) \)

12 a \[ S \sim B(10, 0.25) \]
\[ P(S = 3) = \binom{10}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^7 = 0.2503 \]

b \[ S \sim B(100, 0.25) \]
\[ P(S \leq 27) = \binom{100}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^{100} + \cdots + \binom{100}{27} \left(\frac{1}{4}\right)^{27} \left(\frac{3}{4}\right)^{73} = 0.7224 \]
c  \( S \sim B(n, 0.25) \)
\[
P(S \geq 50) = 1 - P(S \leq 49) = 0.75
\]

This can be entered as an equation into the GDC and the table function used to find the value of \( n \)

If \( n = 215 \), \( P(S \geq 50) = 0.7460 \) and if \( n = 216 \), \( P(S \geq 50) = 0.7582 \) so \( n = 216 \)

13 a  \( T \sim N\left(3 \times 60, 3^2 \times 3^2\right) \)
\[
P(T > 175) = 0.711
\]

b
\[
E(H - 2.5M) = 160 - 2.5 \times 60 = 10
\]
\[
\text{Var}(H - 2.5M) = \text{Var}(H) + 2.5^2 \text{Var}(M) = 5^2 + 2.5^2 \times 3^2 = 81.25
\]
\[
P(H - 2.5M < 0) = 0.1336
\]

14 a  \( P(E > 55) = 0.1057 \Rightarrow 10.57\%
\]

b Let \( L \) be the number of large eggs in a box. \( L \sim B(6, 0.1057) \)
\[
P(L \geq 1) = 1 - P(L = 0) = 0.488
\]

15 Enter both expressions into the GDC and use the equation solving function
\[
P(X = 9) = \frac{1}{2} P(X = 7)
\]
\[
\mu = 6
\]

16 a i  \( E(X) = 1 \times \frac{1}{4} = \frac{1}{4} \)

ii  \( \text{Var}(X) = 1 \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{16} \)

b i  \( E(\bar{X}) = E(X) = \frac{1}{4} \)

ii  \( \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{N} = \frac{3}{1600} = \frac{3}{1600} \)

c  \( P(\bar{X} > 0.305) = 0.1020 \)

d  \( P\left(\sum_{i=1}^{100} X_i > 30.5\right) = P(\bar{X} > 0.305) = 0.1020 \)

e i  \( T \sim B\left(100, \frac{1}{4}\right) \)

ii  \( P(T > 30.5) = P(T \geq 31) = 1 - P(T \leq 30) = 0.1038 \)

17 a  \( P(X \leq 2) = P(X = 1) + P(X = 2) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6} \)

b  \( E(X) = 1 \times \frac{1}{2} + 2 \times \frac{1}{3} + 3 \times \frac{1}{6} = \frac{5}{3} \)

c i  \( E(Y) = E(4X) = 4E(X) = \frac{20}{3} \)

ii  \( \text{Var}(Y) = \text{Var}(4X) = 4^2 \text{Var}(X) = \frac{80}{9} \)

d i  \( E(T) = E(X_1 + X_2 + X_3) = E(X) + E(X) + E(X) = 3E(X) = 5 \)
ii \[ \text{Var}(T) = \text{Var}(X_1 + X_2 + X_3) \]
\[ = \text{Var}(X) + \text{Var}(X) + \text{Var}(X) \]
\[ = 3\text{Var}(X) = \frac{5}{3} \]

\[ \text{e i } E(R) = \frac{1}{1} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{6} = \frac{13}{18} \]

ii Not true because \[ \frac{13}{18} \neq \frac{1}{5} = \frac{3}{5} \]
### Skills check

1. 
   \[ P(S) = \frac{n(\{1, 4\})}{n(\{1, 2, 3, 4, 5, 6\})} = \frac{1}{3}, \quad P(E) = \frac{n(\{2, 4, 6\})}{n(\{1, 2, 3, 4, 5, 6\})} = \frac{1}{2}, \]
   
   \[ P(E \cap S) = \frac{n(\{4\})}{n(\{1, 2, 3, 4, 5, 6\})} = \frac{1}{6} = \frac{1}{3} \times \frac{1}{2} = P(S) \times P(E) \] so independent

2. 
   \[ P(W < 36) = 0.6306 \]

3. Using the function on the GDC
   \[ r = 0.7719 \]

### Exercise 14A

1. 
   \[ a \quad 1 \quad \text{because the data is strictly increasing} \]
   \[ b \quad 0.99 \quad \text{because the data is strictly increasing other than the first two points which are non-decreasing} \]
   \[ c \quad -1 \quad \text{because the data is strictly decreasing} \]
   \[ d \quad \text{Approximately 0 because though there is a relation it is neither increasing or decreasing} \]

2. 
   \[
   \begin{array}{|c|c|c|c|}
   \hline
   x & y & R_x & R_y \\
   \hline
   0 & 23 & 1 & 7 \\
   5 & 18 & 2 & 6 \\
   10 & 10 & 3 & 5 \\
   15 & 9 & 4 & 4 \\
   20 & 7 & 5 & 2 \\
   25 & 7 & 6 & 2 \\
   30 & 7 & 7 & 2 \\
   \hline
   \end{array}
   \]

   \[ r_s = -0.9636 \]
b

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>R_x</th>
<th>R_y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>14</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>9</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

$r_s = 0.8929$

3 a Not appropriate as the plot indicates that the relationship is not linear

b

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>R_x</th>
<th>R_y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>30.4</td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>0.51</td>
<td>25.1</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>0.69</td>
<td>20</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>1.09</td>
<td>15.6</td>
<td>4</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>1.52</td>
<td>13.4</td>
<td>5</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>2.02</td>
<td>13.4</td>
<td>6</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>2.43</td>
<td>11.2</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2.67</td>
<td>10.2</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2.93</td>
<td>10.8</td>
<td>9</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3.17</td>
<td>10.3</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

$r_s = -0.9605$

3 c The result indicates that there is a strong inverse relationship between velocity and force.

4 a

$r = 0.6699$

Moderate positive correlation. This indicates generally students who do better in Maths tend to do better in English.
Graph indicates that the data does not follow a linear model throughout the whole range of scores. Those with a high score in Maths have very variable scores in English. This is possibly due to the Mathematics test not separating students at the higher ability level, or students who are able at Mathematics not having English as a mother tongue. A linear model is probably not appropriate over the whole range.

### c

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$R_x$</th>
<th>$R_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>44</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>25</td>
<td>47</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>37</td>
<td>42</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>49</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>60</td>
<td>52</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>72</td>
<td>44</td>
<td>6</td>
<td>2.5</td>
</tr>
<tr>
<td>74</td>
<td>54</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>78</td>
<td>59</td>
<td>8.5</td>
<td>8</td>
</tr>
<tr>
<td>78</td>
<td>69</td>
<td>8.5</td>
<td>9</td>
</tr>
<tr>
<td>79</td>
<td>78</td>
<td>10.5</td>
<td>10</td>
</tr>
<tr>
<td>79</td>
<td>89</td>
<td>10.5</td>
<td>11</td>
</tr>
</tbody>
</table>

Repeated ranks, so need to use repeated rank formula:

$$r_s = 0.8833$$

The Spearman’s correlation coefficient is higher and indicates more strongly that those who do better in Maths do better in English. This is supported by the graph.

### d

Spearman’s is the more useful measure of correlation because the relationship between the variables is not linear.

### 5 a

Because the ranks are given rather than quantifiable data.
Generally, the preferred coffees are the more expensive ones.

\[ r_s = -0.8857 \]

Exercise 14B

1 a Assume that \( H_0 \) is true, then

\[
P(X \geq 7) = \binom{12}{7} \times 0.3^7 \times 0.7^5 + \binom{12}{12} \times 0.3^{12} \times 0.7^0 = 0.03860 < 0.05 \quad \text{this is significant, so therefore reject } H_0
\]
b Assume that $H_0$ is true, then

$$P(X \leq 6) = \binom{20}{0} \times 0.4^0 \times 0.6^{20} + \binom{20}{6} \times 0.4^6 \times 0.6^{14} = 0.2500 < 0.1$$

this is not significant, so therefore no reason to reject $H_0$.

c Assume that $H_0$ is true, then

$$P(X \geq 9) = \binom{10}{9} \times 0.7^9 \times 0.3^1 + \binom{10}{10} \times 0.7^{10} \times 0.3^0 = 0.1493 < 0.1$$

this is not significant, so therefore no reason to reject $H_0$.

2 a $H_0 : p = 0.6, H_1 : p < 0.6$

b Need to find $r$ such that $P(X \leq r) = 0.05 \Rightarrow r = 13 \Rightarrow P(X \leq r) = 0.04811$,

$$r = 14 \Rightarrow P(X \leq r) = 0.09706$$

so the critical region is $X \leq 13$.

c As $14 \leq 13$, the result is not significant, so therefore no reason to reject $H_0$. Thus, the treatment leads to a significant reduction in symptoms.

3 a Need to find $r$ such that $P(X \geq r) = 0.05 \Rightarrow r = 3 \Rightarrow P(X \geq r) = 0.05297$,

$$r = 4 \Rightarrow P(X \geq r) = 0.00833$$

so the critical region is $X \geq 4$.

b Binomial may not be appropriate as tripping on one fence may mean tripping on others (i.e. not independent).

4 $H_0 : p = 0.3, H_1 : p > 0.3$. Assume that $H_0$ is true, then

$$P(X \geq 8) = \binom{20}{8} \times 0.3^8 \times 0.7^{12} + \binom{20}{20} \times 0.3^{20} \times 0.7^0 = 0.2277 < 0.05$$

this is not significant, so therefore no reason to reject $H_0$. There is insufficient evidence to suggest that the course helped.

5 $H_0 : p = 0.1, H_1 : p > 0.1$. Assume that $H_0$ is true, then

$$P(X \geq 2) = \binom{5}{2} \times 0.1^2 \times 0.9^3 + \binom{5}{5} \times 0.1^5 \times 0.9^0 = 0.08146 < 0.05$$

this is not significant, so therefore no reason to reject $H_0$. There is insufficient evidence to suggest that the bus is late more than 10% of the time.

Exercise 14C

1 a Assume that $H_0$ is true, then $P(X \leq 5) = 0.2759 > 0.05$ this is not significant, so therefore no reason to reject $H_0$.

b Assume that $H_0$ is true, then $P(X \geq 10) = 0.07721 > 0.05$ this is not significant, so therefore no reason to reject $H_0$.

c Assume that $H_0$ is true, then $P(X \geq 23) = 1 - P(X \leq 22) = 0.005876 < 0.01$, this is significant, so therefore reject $H_0$.

2 $H_0 : \lambda = 2 \times 1.8 = 3.6, H_1 : \lambda < 3.6$. Assume that $H_0$ is true, then $P(X \leq 2) = 0.3028 > 0.05$. This is not significant, so therefore no reason to reject $H_0$. There is insufficient evidence to suggest that the bus is late more than 1.8 times per week.

3 a $H_0 : \lambda = 5 \times 2.2 = 11, H_1 : \lambda < 11$
b Need to find \( r \) such that \( P(X \leq r) = 0.05 \). \( r = 5 \Rightarrow P(X \leq r) = 0.03752 \).
\( r = 6 \Rightarrow P(X \leq r) = 0.07861 \) so the critical region is \( X \leq 5 \). The actual significance level for the test is 3.75%.

c A constant rate is assumed for the Poisson distribution, which may not be appropriate as the number of goals conceded depends on the strength of the opposition, the players available, the weather, etc.

4 a \( H_0 : \lambda = 0.25 \times 16 = 4.0 \), \( H_1 : \lambda > 4.0 \). Need to find \( r \) such that \( P(X \geq r) = 0.05 \)
\( r = 8 \Rightarrow P(X \geq r) = 0.05113 \), \( r = 9 \Rightarrow P(X \geq r) = 0.02136 \) so the critical region is \( X \geq 9 \).

b Need to assume that the infections occur independently of each other.

Exercise 14D

1 a i \( r = 0.7206 \)

ii \( p \)-value = 0.4878 > 0.1 so no reason to reject the null hypothesis so there is insufficient evidence that there is a linear correlation.

b i \( r = 0.7246 \)

ii \( p \)-value = 0.1033 > 0.1 so no reason to reject the null hypothesis so there is insufficient evidence that there is a linear correlation

c i \( r = 0.7176 \)

ii \( p \)-value = 0.008597 < 0.1 so reason to reject the null hypothesis that there is no correlation between the two variables

iii \( y = 42.7981 + 0.5205x \)

2 a \( r = -0.9013 \)

b i \( p \)-value = 0.00112 < 0.05, which is highly significant and so the null hypothesis is rejected and the alternative hypothesis that there is a negative correlation between price and number of sales is accepted.

ii \( s = 20.2 - 2.76p \)

iii \( s = 20.2 - 2.76p \Rightarrow s = 20.2 - 2.76 \times 5.50 = $5.02 \)

c
The relationship is not linear, so even though the \( p \)-value was very small, the linear model may not be appropriate.

3 a \( r = 0.8683 \)

\( p \)-value = 0.001119 < 0.05, which is highly significant and so the null hypothesis is rejected and the alternative hypothesis that there is a positive correlation between sales and temperature is accepted.

b Let \( x \) be the temperature (the independent variable) and \( y \) be the ice cream (the dependent variable). \( y = 3.94x + 78.1 \)

c \( y = 3.94 \times 23 + 78.1 \Rightarrow y = €169 \)

d 35°C is outside the domain of the data provided, and it is not wise to extrapolate.

Exercise 14E

1 These questions should be answered using the Z-test function on your GDC

a \( p \)-value = 0.166 > 0.05 not significant, so do not reject \( H_0 \)

b \( p \)-value = 0.0302 < 0.05 so significant, reject \( H_0 \)

c \( p \)-value = 0.412 > 0.10 not significant, so do not reject \( H_0 \)

2 a The sample size is large enough for the central limit theorem to apply

b \( H_0 : \mu = 24 \), \( H_1 : \mu < 24 \)

c \( p \)-value = 0.01391 < 0.05, significant, so enough evidence to reject \( H_0 \) that the amount of area covered is 24 m\(^2\)

d Need to find \( r \) such that \( P(\bar{X} \leq r) = 0.05 \).

Under \( H_0 \)  \( \bar{X} \sim N(24, \frac{1.81^2}{32}) \)

To find the critical value use the inverse normal function on your GDC with the appropriate mean and standard deviation

\( P(\bar{X} \leq r) = 0.05 \Rightarrow r = invnorm(0.05) \Rightarrow r = 23.48 \), so the critical region is  \( \bar{X} \leq 23.48 \), thus confirming the result of the test as we had a value of 23.3 < 23.48

3 a \( H_0 : \mu = 8.3 \), \( H_1 : \mu > 8.3 \)

b \( \bar{x} = \frac{8 + 8.7 + 9.2 + 8.4 + 8.5}{5} = 8.56 \)

c \( \bar{x} \sim N(8.3, \frac{2.1^2}{5}) = N(8.3, 0.882) \)

d \( P(\bar{X} > \bar{x}) = P(\bar{X} > 8.56) = 0.391 \)

e 0.391 > 0.1 so not enough evidence to reject the null hypothesis at the 10% significance level that the mean time between buses is 8.3 minutes.

f \( p \)-value = 0.391
g Need to find $r$ such that $P(\bar{X} \geq r) = 0.05$, using the inverse normal function on the GDC

$P(\bar{X} \geq r) = 0.05 \Rightarrow P(\bar{X} < r) = 0.95 \Rightarrow r = \text{invnorm}(0.95) \Rightarrow r = 9.8448$, so the critical region is $\bar{X} \geq 9.8448$

Exercise 14F
These questions should be done using the t-test function on the GDC

1 a p-value = 0.1001 > 0.05, so not significant so do not reject $H_0$
   b p-value = 0.421 > 0.05 not significant, so do not reject $H_0$
   c p-value = 0.0239 < 0.05 significant, so reject $H_0$

2 a p-value = 0.058 > 0.05 not significant, so do not reject $H_0$
   b p-value = 0.066 > 0.05 not significant, so do not reject $H_0$
   c p-value = 0.0952 < 0.10 significant, reject $H_0$

3 a It is possible because the sample size is large enough for the central limit theorem to apply
   b $H_0 : \mu = 28.2$, $H_1 : \mu > 28.2$
   c p-value = 0.233 > 0.05 not significant, so insufficient evidence

4 a $H_0 : \mu = 83$, $H_1 : \mu > 83$
   b i 86.0 ii 3.41 iii 3.81
   c 0.0764 > 0.05, not significant so insufficient evidence to reject the null hypothesis that the mean journey time is 83 minutes.

Exercise 14G
These questions should be done using the Z or T confidence interval function on the GDC

1 a (10.4,14.4)
   b (61.9,62.7)
   c Note that if your GDC requires you to enter the unbiased estimator you need to enter $\sqrt{\frac{10 \times 5.2}{9}} = (4.58,8.02)$
   d Note that if your GDC requires you to enter the unbiased estimator you need to enter $\sqrt{\frac{10 \times 0.2^2}{9}} = (2.10,2.50)$
   e As the population standard deviation is known find the Z (normal) interval rather than the T interval (4.47,4.73)

2 These are small samples so use Student’s $t$.
   a (10.24,14.28)
   b (−7.85,6.89)
3  a  i  (19.76,22.64)
   ii  (20.30,22.10)
   iii  (20.67,21.73)
   iv  (20.83,21.57)
  b  The larger the sample the smaller the width of the confidence interval
4  a  (11.57,15.23)
  b  That the population can be modelled by a normal distribution
  c  15.3 is outside the range for the confidence interval so is unlikely to be the population mean.

Exercise 14H
In this exercise use the two sample T-test function on your GDC

1  \[ H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 < \mu_2 \]
p-value = 0.251
0.251 > 0.10, not significant, do not reject the null hypothesis. There is no difference in the weights of the apples.

2  \[ H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 > \mu_2 \]
p-value = 0.00539
0.00539 < 0.01, significant, so reject the null hypothesis. Those using the new remedy do lose more weight.

Exercise 14I

1  

<table>
<thead>
<tr>
<th>Differences</th>
<th>7</th>
<th>1</th>
<th>7</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>4</th>
<th>−5</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
</table>

\[ H_0 : \mu = 0, \quad H_1 : \mu > 0 \]
p-value  0.00233 < 0.01, significant so reject \( H_0 \) that there has been no improvement.

2  a  \[ H_0 : \mu = 5, \quad H_1 : \mu > 5 \]

b  

<table>
<thead>
<tr>
<th>Differences</th>
<th>5</th>
<th>10</th>
<th>6</th>
<th>8</th>
<th>3</th>
<th>−1</th>
</tr>
</thead>
</table>

p-value  0.460 > 0.05, not significant so insufficient evidence to reject \( H_0 \) that the company on average increases a client’s investment by 5%.

3  a  

<table>
<thead>
<tr>
<th>Differences</th>
<th>−0.6</th>
<th>−1.7</th>
<th>−0.2</th>
<th>−1.5</th>
<th>−1.7</th>
<th>−2</th>
<th>−1.4</th>
<th>−0.7</th>
<th>−0.6</th>
<th>−0.8</th>
</tr>
</thead>
</table>

\[ H_0 : \mu = 0, \quad H_1 : \mu < 0 \]
p-value = 0.000127<0.05 so the result is significant and the drug has a positive effect.
b \( H_0 : \mu_0 = 0.7 \), \( H_1 : \mu_0 < 0.7 \)

\[ p\text{-value} = 0.0285 < 0.05 \], so the result is significant and the drug has a positive effect on top of the healthy diet.

c  
1. Yes because there is some good evidence the drug is working.
2. Larger sample, group chosen to eliminate other factors e.g. age, gender.

Exercise 14J

1 \( H_0 : \) favourite flavour of chocolate is independent of gender
\( H_1 : \) favourite flavour of chocolate is not independent of gender

From the inbuilt function on the GDC

Either: the \( \chi^2 \) test statistic = 9.52. 9.52 > 9.210, significant so reject the null hypothesis that favourite chocolate and gender are independent.

Or: the p-value = 0.00856. 0.00856 < 0.01, significant so reject the null hypothesis that favourite chocolate and gender are independent.

2 \( a \) \( H_0 : \) GPA is independent of number of hours on social media
\( H_1 : \) GPA is not independent of number of hours on social media

\[ 85 \times 99 \frac{270}{270} \approx 31.1667 = 31.2 \]

b Degrees of freedom = \((3 - 1)\times(3 - 1) = 2 \times 2 = 4\)

d The \( \chi^2 \) test statistic = 78.5 and the p-value = \( 3.59 \times 10^{-16} \).

The critical value is 7.779.

e Either: 78.5 > 7.779,

Or: \( 3.59 \times 10^{-16} < 0.10 \)

Significant so reject the null hypothesis that GPA is independent of number of hours on social media.

3 \( H_0 : \) number of people walking their dog is independent of the time of day
\( H_1 : \) number of people walking their dog is not independent of the time of day

Either: The \( \chi^2 \) test statistic = 5.30

5.30 < 9.488 so the result is not significant so no reason to reject the null hypothesis that number of people walking their dog is independent of the time of day.

Or: the p-value = 0.257

0.257 > 0.05 so the result is not significant so no reason to reject the null hypothesis that number of people walking their dog is independent of the time of day.

4 \( a \) \( H_0 : \) type of degree that a person has is independent of their annual salary
\( H_1 : \) type of degree that a person has is not independent of their annual salary
b For the cell containing the number of people with a BA who earn less than $60,000:
\[
\frac{18}{100} \times \frac{27}{100} \times 100 = 4.86 < 5
\]

c

<table>
<thead>
<tr>
<th>Observed</th>
<th>BA</th>
<th>MA</th>
<th>PhD</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; $120,000</td>
<td>20</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>&gt; $120,000</td>
<td>7</td>
<td>13</td>
<td>24</td>
</tr>
</tbody>
</table>

d p-value = 0.00403

e 0.00403 < 0.01 The result is significant so reject the null hypothesis that type of degree and salary are independent.

Exercise 14K

1 a

<table>
<thead>
<tr>
<th>Colour</th>
<th>Yellow</th>
<th>Orange</th>
<th>Red</th>
<th>Purple</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>104</td>
<td>132</td>
<td>98</td>
<td>129</td>
<td>137</td>
</tr>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Expected</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
</tbody>
</table>

b 4 degrees of freedom

c \( H_0 \): the colours follow a uniform distribution
\( H_1 \): the colours do not follow a uniform distribution

\[
\chi^2 = 10.45 \text{ and the p-value } = 0.0335
\]
Either 10.45 > 9.488 or 0.0335 < 0.05

The result is significant so reject the null hypothesis that the distribution is uniform.

2 a

<table>
<thead>
<tr>
<th>Last number</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>44</td>
<td>53</td>
<td>49</td>
<td>61</td>
<td>47</td>
<td>52</td>
<td>39</td>
<td>58</td>
<td>42</td>
<td>45</td>
</tr>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Expected</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

b \( H_0 \): the last number follows a uniform distribution
\( H_1 \): the last number does not follow a uniform distribution
The p-value = 0.430

0.430 > 0.10 not significant so no reason to reject the null hypothesis that the last number on the lottery tickets follows a uniform distribution.
3 a

<table>
<thead>
<tr>
<th>x &lt; 50</th>
<th>50 ≤ x &lt; 60</th>
<th>60 ≤ x &lt; 70</th>
<th>70 ≤ x &lt; 80</th>
<th>80 ≤ x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.02275</td>
<td>0.2297</td>
<td>0.4950</td>
<td>0.2297</td>
</tr>
<tr>
<td>Expected</td>
<td>6.83</td>
<td>68.92</td>
<td>148.50</td>
<td>68.92</td>
</tr>
</tbody>
</table>

b \( H_0 \): the grades fit a normal distribution with mean 65% and s.d. of 7.5%.
\( H_1 \): the grades don’t fit a normal distribution with mean 65% and s.d. of 7.5%.
The p-value = 0.947

0.947 > 0.10 so the result is not significant and hence there is no reason to reject the null hypothesis that the exam results follow the given distribution.

4 a

<table>
<thead>
<tr>
<th>h &lt; 235</th>
<th>235 ≤ h &lt; 245</th>
<th>245 ≤ h &lt; 255</th>
<th>255 ≤ h &lt; 265</th>
<th>265 ≤ h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.08634</td>
<td>0.2384</td>
<td>0.3506</td>
<td>0.2384</td>
</tr>
<tr>
<td>Expected</td>
<td>21.6</td>
<td>59.6</td>
<td>87.6</td>
<td>59.6</td>
</tr>
</tbody>
</table>

b \( H_0 \): the heights fit a normal distribution with mean 250 and s.d. of 11.
\( H_1 \): the heights don’t fit a normal distribution with mean 250 and s.d. of 11.
p-value = 0.0906

0.0906 > 0.05 not significant so no reason to reject the null hypothesis that the sample of elephants was taking from a population whose heights are fit a \( N(250,11^2) \) distribution.

5 a \( H_0 \): the scores are normally distributed with mean of 100 and s.d. of 10.
\( H_1 \): the scores are not normally distributed with mean of 100 and s.d. of 10.

b Because if the expected values for the scores outside the range of the observed data are not zero then they would contribute to the test statistic.

c

<table>
<thead>
<tr>
<th>x &lt; 90</th>
<th>90 ≤ x &lt; 100</th>
<th>100 ≤ x &lt; 110</th>
<th>110 ≤ x &lt; 120</th>
<th>120 ≤ x &lt; 130</th>
<th>130 ≤ x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1587</td>
<td>0.3413</td>
<td>0.3413</td>
<td>0.1359</td>
<td>0.02140</td>
</tr>
<tr>
<td>Expected</td>
<td>31.73</td>
<td>68.27</td>
<td>68.27</td>
<td>27.18</td>
<td>4.28</td>
</tr>
</tbody>
</table>

d

<table>
<thead>
<tr>
<th>x &lt; 90</th>
<th>90 ≤ x &lt; 100</th>
<th>100 ≤ x &lt; 110</th>
<th>110 ≤ x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>18</td>
<td>39</td>
<td>78</td>
</tr>
<tr>
<td>Probability</td>
<td>0.1587</td>
<td>0.3413</td>
<td>0.3413</td>
</tr>
<tr>
<td>Expected</td>
<td>31.73</td>
<td>68.27</td>
<td>68.27</td>
</tr>
</tbody>
</table>

e 3 degrees of freedom
Worked solutions

\( f \quad \chi^2_{\text{calc}} = 54.8 \) and \( p\text{-value} = 7.40 \times 10^{-12} \)

58.4 > 11.3 or \( 7.40 \times 10^{-12} < 0.01 \) hence the result is significant and the null hypothesis that the IQs of the students have been taken from this distribution should be rejected.

g i Mean = 105
Standard deviation = 13.56

ii The data is fairly symmetrical and most of the data is within two standard deviations of the mean which indicates a normal distribution is possible. The high value of the chi-squared statistic is probably due to the mean and the standard deviation being somewhat higher for those being tested.

iii Redo the test with a different mean and standard deviation.

Exercise 14L

1 a i This is a binomial distribution with \( n = 2 \) and \( p = \frac{1}{6} \)

<table>
<thead>
<tr>
<th>Number of sixes</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.694</td>
<td>0.278</td>
<td>0.0278</td>
</tr>
<tr>
<td>Expected</td>
<td>173.61</td>
<td>69.44</td>
<td>6.94</td>
</tr>
</tbody>
</table>

ii In table above

b \( H_0 : \) the dice are fair
\( H_1 : \) the dice aren’t fair
\( \chi^2 = 28.1461 > 5.991. \) The result is significant so there is strong evidence to reject the null hypothesis that the dice are fair.

2 a

There will be several methods for generating these values on your GDC. For example entering the Poisson probability density function (Poissonpdf or similar) as a function and using the table function to read off the values. The expected frequency is the probability multiplied by 50

<table>
<thead>
<tr>
<th>Goals</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0907</td>
<td>4.53</td>
</tr>
<tr>
<td>1</td>
<td>0.218</td>
<td>10.89</td>
</tr>
<tr>
<td>2</td>
<td>0.261</td>
<td>13.07</td>
</tr>
<tr>
<td>3</td>
<td>0.209</td>
<td>10.45</td>
</tr>
<tr>
<td>4</td>
<td>0.125</td>
<td>6.27</td>
</tr>
<tr>
<td>5</td>
<td>0.0602</td>
<td>3.01</td>
</tr>
<tr>
<td>( \geq 6 )</td>
<td>0.0357</td>
<td>1.78</td>
</tr>
</tbody>
</table>
b  
H₀ : the number of goals follows a Poisson distribution with a mean of 2.4  
H₁ : the number of goals doesn’t follow a Poisson distribution with a mean of 2.4  

Need to first combine the final three rows of the table.

<table>
<thead>
<tr>
<th>Goals</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>≥4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>7</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Expected</td>
<td>4.53</td>
<td>10.89</td>
<td>13.07</td>
<td>10.45</td>
<td>11.06</td>
</tr>
</tbody>
</table>

\[ X^2 = 1.8299, \ p-value = 0.767 > 0.1. \]  
The result is not significant and so there is insufficient evidence to reject the null hypothesis that the sample is taken from a population that follows a Po(2.4) distribution.

3  a  
Using the probability density function on the GDC for B(3, 0.75)

<table>
<thead>
<tr>
<th>Seeds</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0156</td>
<td>0.78</td>
</tr>
<tr>
<td>1</td>
<td>0.1406</td>
<td>7.03</td>
</tr>
<tr>
<td>2</td>
<td>0.4219</td>
<td>21.09</td>
</tr>
<tr>
<td>3</td>
<td>0.4219</td>
<td>21.09</td>
</tr>
</tbody>
</table>

b  
H₀ : the number of seeds germinating fit a B(3, 0.75) distribution  
H₁ : the number of seeds germinating doesn’t fit a B(3, 0.75) distribution  

c  Need to combine first two columns of the table.

<table>
<thead>
<tr>
<th>Seeds germinating</th>
<th>0 or 1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>15</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Probability</td>
<td>0.1562</td>
<td>0.4219</td>
<td>0.4219</td>
</tr>
<tr>
<td>Expected</td>
<td>7.81</td>
<td>21.09</td>
<td>21.09</td>
</tr>
</tbody>
</table>

d  2 degrees of freedom  
e  \ p-value = 0.01474 < 0.05.  
The result is significant so reject the null hypothesis that the seeds fit a B(3, 0.75) distribution.

4  a  
No of 5-minute periods = \[ \frac{5 \times 60}{5} = 60 \]  
Generate the Poisson probabilities using the inbuilt functions on the GDC then multiply by 60 to obtain the expected values.
<table>
<thead>
<tr>
<th>Number of People</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0150</td>
<td>0.90</td>
</tr>
<tr>
<td>1</td>
<td>0.0630</td>
<td>3.78</td>
</tr>
<tr>
<td>2</td>
<td>0.1323</td>
<td>7.94</td>
</tr>
<tr>
<td>3</td>
<td>0.1852</td>
<td>11.11</td>
</tr>
<tr>
<td>4</td>
<td>0.1944</td>
<td>11.66</td>
</tr>
<tr>
<td>5</td>
<td>0.1633</td>
<td>9.80</td>
</tr>
<tr>
<td>6</td>
<td>0.1143</td>
<td>6.86</td>
</tr>
<tr>
<td>7</td>
<td>0.0686</td>
<td>4.12</td>
</tr>
<tr>
<td>≥8</td>
<td>0.0639</td>
<td>3.84</td>
</tr>
</tbody>
</table>

Combine the first three and last two rows:

| ≤2 | 0.0150 + 0.0630 + 0.1323 = 0.2103 | 12.62 |
| ≥7 | 0.0686 + 0.0639 = 0.1325          | 7.96 |

p-value = 0.00308 < 0.05

The result is significant so reject the null hypothesis that the number of people joining the queue follows a Po(4.2) distribution.

b It is assumed that the average number of people arriving for the cell ≥8 people is 8.

i Mean = 4.23

Obtain the value for the standard deviation from the GDC then square it to obtain the variance = 6.81

ii The parameter for the Poisson is the mean and the mean of the sample is close to the mean for the distribution so it is likely a correct mean was chosen.

For the Poisson distribution the mean and variance of the sample should be similar, that is not the case here which indicates the distribution is not Poisson.

5 a $X \sim B(5, 0.25)$

\begin{tabular}{|c|c|c|}
\hline
X & Probability & Expected Frequency \\
\hline
0 & 0.2373 & 118.65 \\
1 & 0.3955 & 197.75 \\
2 & 0.2637 & 131.84 \\
3 & 0.0879 & 43.95 \\
\hline
\end{tabular}
c Need to combine the final two rows of the table

| 4 or 5 | 0.0146 + 0.0010 = 0.0156 | 7.81 |

\[ H_0 : \text{the students are guessing randomly} \]
\[ H_1 : \text{the students are not guessing randomly} \]

There are 4 degrees of freedom. The p-value = 0.0495

0.0495 < 0.05 just significant so some evidence to reject the null hypothesis that the students are guessing randomly.

Exercise 14M

1 a
Standard deviation
\[ = 121.22 \approx 121 \]

b Using the \[ N(1200, 121^2) \] distribution

<table>
<thead>
<tr>
<th>Lifespan, ( h ) hours</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h &lt; 1000 )</td>
<td>( P(H &lt; 1000) = 0.0492 )</td>
<td>19.7</td>
</tr>
<tr>
<td>( 1000 \leq h &lt; 1100 )</td>
<td>( P(1000 \leq H &lt; 1100) = 0.1551 )</td>
<td>62.0</td>
</tr>
<tr>
<td>( 1100 \leq h &lt; 1200 )</td>
<td>( P(1100 \leq H &lt; 1200) = 0.2957 )</td>
<td>118.3</td>
</tr>
<tr>
<td>( 1200 \leq h &lt; 1300 )</td>
<td>( P(1200 \leq H &lt; 1300) = 0.2957 )</td>
<td>118.3</td>
</tr>
<tr>
<td>( 1300 \leq h &lt; 1400 )</td>
<td>( P(1300 \leq H &lt; 1400) = 0.1551 )</td>
<td>62.0</td>
</tr>
<tr>
<td>( 1400 \leq h )</td>
<td>( P(H \geq 1400) = 0.0492 )</td>
<td>19.7</td>
</tr>
</tbody>
</table>

c Degrees of freedom: \( 6 - 1 - 1 = 4 \)

d \( H_0 : \) the lifespan of lightbulbs follows a \( N(1200,121^2) \) distribution
\[ H_1 : \text{the lifespan of lightbulbs doesn't follow a } N(1200,121^2) \text{ distribution} \]

\[ p\text{-value} = 5.83 \times 10^{-7} < 0.05 \text{. This result is significant and so we reject the null hypothesis that the data follows a } N(1200,121^2) \text{ distribution.} \]

2 a Mean = \( np = 1.5319 \), so as \( n = 3 \Rightarrow p = 0.5106 \)

b \( H_0 : \) the number of boys in a family has a \( B(3,0.5106) \) distribution
\[ H_1 : \text{the number of boys in a family doesn’t have a } B(3,0.5106) \text{ distribution} \]

<table>
<thead>
<tr>
<th>Number of boys</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1172</td>
<td>11.72</td>
</tr>
<tr>
<td>1</td>
<td>0.3669</td>
<td>36.69</td>
</tr>
</tbody>
</table>

© Oxford University Press 2019
No of degrees of freedom $= 4 - 1 - 1 = 2$.

$p$-value $= 0.0732 > 0.01$. The result is not significant and so there isn’t enough evidence to reject the null hypothesis that the number of boys in these families follows a $B(3,0.5106)$ distribution.

3 a Mean $= \frac{0 \times 7 + 1 \times 10 + 2 \times 15 + 3 \times 21 + 4 \times 14 + 5 \times 9 + 6 \times 4}{7 + 10 + 15 + 21 + 14 + 9 + 4} = 2.85$

b $H_0$: the number of fish caught has a $Po(2.85)$ distribution

$H_1$: the number of fish caught doesn’t have a $Po(2.85)$ distribution

<table>
<thead>
<tr>
<th>Number of fish</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0578</td>
<td>4.6</td>
</tr>
<tr>
<td>1</td>
<td>0.1649</td>
<td>13.2</td>
</tr>
<tr>
<td>2</td>
<td>0.2349</td>
<td>18.8</td>
</tr>
<tr>
<td>3</td>
<td>0.2232</td>
<td>17.9</td>
</tr>
<tr>
<td>4</td>
<td>0.159</td>
<td>12.7</td>
</tr>
<tr>
<td>5</td>
<td>0.0906</td>
<td>7.3</td>
</tr>
<tr>
<td>$\geq 6$</td>
<td>0.0696</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Need to combine the first two rows:

| 0 or 1         | 0.0578 + 0.1649 = 0.2227 | 17.8 |

Degrees of freedom $= 6 - 1 - 1 = 4$.

$p$-value $= 0.676 > 0.05$. The result is not significant so no reason to reject the null hypothesis that the number of fish caught follows a $Po(2.85)$ distribution.

4 a Mean $= 52.6$

$s_{n,1} = 5.15$

b $H_0$: the weights of the children has a $N(52.6, 5.15^2)$ distribution

$H_1$: the weights of the children doesn’t have a $N(52.6, 5.15^2)$ distribution

<table>
<thead>
<tr>
<th>Weight, $w$ kg</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w &lt; 45$</td>
<td>$P(W &lt; 45)$ = 0.07001</td>
<td>14.00</td>
</tr>
<tr>
<td>$45 \leq w &lt; 50$</td>
<td>$P(45 \leq W &lt; 50)$ = 0.2368</td>
<td>47.36</td>
</tr>
<tr>
<td>$50 \leq w &lt; 55$</td>
<td>$P(50 \leq W &lt; 55)$ = 0.3726</td>
<td>74.51</td>
</tr>
</tbody>
</table>
Degrees of freedom = 5 – 1 – 2 = 2

\[ p\text{-value} = 5.24 \times 10^{-5} < 0.05. \text{ The result is significant so the sample is very unlikely to have come from a population with a } N(52.6, 5.15^2) \text{ distribution.} \]

c The data has two peaks, which suggests two populations. For example, the sample may have included a mixture of boys and girls.

5 a \( X \sim B(50, p) \)

b Mean \( np = 1.9 \Rightarrow p = \frac{1.9}{50} = 0.038 \)

c \( H_0: \) the sample is from a \( B(50,0.038) \) population  
\( H_1: \) the sample isn’t from a \( B(50,0.038) \) population

<table>
<thead>
<tr>
<th>Number of Prizes</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.65</td>
</tr>
<tr>
<td>1</td>
<td>17.08</td>
</tr>
<tr>
<td>2</td>
<td>16.53</td>
</tr>
<tr>
<td>3</td>
<td>10.45</td>
</tr>
<tr>
<td>4</td>
<td>4.85</td>
</tr>
<tr>
<td>( \geq 5 )</td>
<td>2.45</td>
</tr>
</tbody>
</table>

Need to combine final two rows:

| \( \geq 4 \) | 7.3 |

Degrees of freedom = 5 – 1 – 1 = 3.

\[ p\text{-value} = 0.67 > 0.05 \]

The result is not significant so no reason to reject the null hypothesis that the prizes are distributed randomly and independently.

6 a \( H_0: \) the sample is from a \( \text{Po}(2.0) \) distribution  
\( H_1: \) the sample isn’t from a \( \text{Po}(2.0) \) distribution

Number of 5-minute periods \( \frac{5 \times 60}{5} = 60 \)

<table>
<thead>
<tr>
<th>Number of calls</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1353</td>
<td>8.12</td>
</tr>
</tbody>
</table>
Need to combine the final two rows:

\[ \begin{align*}
\geq 4 & : 0.0902 + 0.0527 = 0.1429 \quad 8.57 \\
\end{align*} \]

Degrees of freedom \( = 5 - 1 = 4 \).

\( p \)-value \( = 0.0133 < 0.05 \). This is significant, so reject the null hypothesis that the data follows a \( \text{Po}(2.0) \) distribution.

\( b \) Mean \( = 2.57 \)

\( \text{H}_0 : \text{the sample is from a Po}(2.57) \) distribution

\( \text{H}_1 : \text{the sample isn't from a Po}(2.57) \) distribution

<table>
<thead>
<tr>
<th>Number of calls</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0765</td>
<td>4.59</td>
</tr>
<tr>
<td>1</td>
<td>0.1967</td>
<td>11.80</td>
</tr>
<tr>
<td>2</td>
<td>0.2528</td>
<td>15.17</td>
</tr>
<tr>
<td>3</td>
<td>0.2165</td>
<td>12.99</td>
</tr>
<tr>
<td>4</td>
<td>0.1391</td>
<td>8.35</td>
</tr>
<tr>
<td>( \geq 5 )</td>
<td>0.1184</td>
<td>7.10</td>
</tr>
</tbody>
</table>

Degrees of freedom \( = 6 - 1 - 1 = 4 \).

\( p \)-value \( = 0.335 > 0.05 \). hence the result is not significant and there is no reason to reject the null hypothesis that the calls arriving at the business follow \( \text{Po}(2.57) \).

Exercise 14N

1 a The first data shows random error and the second shows systematic error. The second dataset follows \( y = 1 + 2.2x \)

\( b \) The second data would have a correlation of 1, the first data would be close to but not equal to 1.

\( c \) A high correlation normally means a line of regression is useful. In this case the systematic error has resulted in a perfect correlation but the line of regression would be increasingly inaccurate as \( x \) increases.

2 a Perform the \( \chi^2 \) test for independence using the inbuilt function on your GDC
<table>
<thead>
<tr>
<th>Positive</th>
<th>30.8</th>
<th>16.1</th>
<th>23.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>13.2</td>
<td>6.9</td>
<td>9.9</td>
</tr>
</tbody>
</table>

Degrees of freedom = \((3 - 1) \times (2 - 1) = 2\)

\(p\)-value = 0.0300 < 0.05 so significant and the two factors are not independent.

\(b\) Each of the 10 attributes can be paired with 9 others, to give 90 pairings but each pair will be counted twice so you need to divide by 2. Number of pairings is \(\frac{10 \times 9}{2} = 45\) or you could reason the first attribute can be paired with 9 others, the second with 8 more and so on to give \(9+8+7+\ldots+2+1=45\)

\(c\) If all attributes are independent, then the probability of obtaining a significant result by chance is 0.05. Let \(X\) be the number of pairs that yield significant results then \(X \sim B(45,0.05)\) and \(P(X \geq 1) = 1 - P(X = 0) = 0.9006\).

\(d\) Need to know if there are other reasons to suspect the two attributes given in part \(a\) were not independent. If there are no other reasons, need to do further tests as not enough evidence otherwise.

3 \(a\) i Taking a random sample or a sample stratified by gender. If measuring which school is better it is important the sample is representative of the school.

ii The ratio of boys to girls in each sample should be as equal as possible, so the improvement is due to the teaching method and not to the gender of the student.

iii A sample of girls should be compared with a sample of boys for each of the schools. The samples could also be pooled to give boys and girls from both schools. However, to avoid the results being affected by the teaching method, in the pooled sample, the ratio of boys to girls from each school should be equal.

\(b\) A t-test on the difference between the average improvement in each school to see if there is a significant difference. Assume the populations are normally distributed or the sample size is large enough for the central limit theorem to apply.

4 \(a\) Good: Any reason such as: The census contains details of most of the population. It will be relatively cheap and easy to collect. Because it contains other information, focused sampling could take place if required.

Bad: Any reason such as: The data is 6 years out of date. It records who was living in the house on census day so if taken during a school holiday it may include students who do not normally live at home.

\(b\) i Using \(X \sim B(1000,0.15)\),

\[
P(140 \leq X \leq 160) = P(X \leq 160) - P(X \leq 139) = 0.8242 - 0.1765 = 0.648
\]

\[
P(100 \leq X \leq 200) = P(X \leq 200) - P(X \leq 100) = 1.0000 - 2.039 \times 10^{-4} = 1
\]

The survey is almost certain to find a proportion correct to within 50 households or 5%, but there is a about a 1 – 0.648 = 0.352 chance it will not be within 10 households or 1%. If this level of accuracy is needed, then more households will need to be surveyed.

\(c\) Do you have any children who have left school? If so, how many live at home and how many live away from home? (A single question such as 'Do you have any children who are still
living at home?’ is ambiguous as it would be answered ‘no’ both by those with no children and by those with children who are not living at home.)

5 a The method of obtaining the sample, to ensure it is appropriate (the two groups should be as uniform as possible) and unbiased. Any confounding factors such as age, employment status or household income. Any assumptions made about the distributions, for example could the distribution be assumed to be normal.

b Let \( X \) be the number of significant results in the 8 tests. Then \( X \sim B(8, 0.05) \) and

\[
P(X \geq 1) = 1 - P(X = 0) = 0.3366
\]

This means that finding one significant result is not in itself significant, so the result is not meaningful.

Exercise 14O

1 a i Two outcomes, each trial independent and identical.

ii Events are independent and occur at a uniform average rate during the period of interest.

iii Data comes from a normal population or the sample size is large so that the central limit theorem applies

b Chi-squared goodness of fit test

2 a The Poisson distribution assumes that the number of infections occur independently and uniformly.

However, new infections are unlikely to be independent as the people already infected may come from the same wards or be receiving treatment from the same people.

The new infections are unlikely to be uniform, because over time more people will have been infected and the number of people susceptible will reduce so that the rate of new infections will decrease. The hospital may also take precautions such as isolating infected wards which would reduce the rate of infections.

b A chi-squared goodness of fit. A significant result might be because the distribution is not Poisson or because the rate of infection is not \( r \).

3 a Chi-squared test or test for \( \rho = 0 \)

b Chi-squared as we do not know if distance or weekly allowance are normally distributed. If it is suspected they might be normally distributed a test should be done to check this.

c The distances and allowances need to be categorised into groups so that the expected values are all greater than 5.

4 a Difficult to find parallel forms for questions about food. Test-retest will also indicate any change over time as well as considering the reliability of the data.

b \( r = 0.907 \)

which indicates the survey is reliable as a source of information

c Using the statistical summary data from your GDC

Test 1 quartiles 4.5 and 5.5, median 4.9

Test 2 quartiles 5.2 and 5.7, median 5.5
Test 1:

---

**d** The box plots indicate there has been an improvement as the median and lower quartile have increased considerably, and the upper quartile has increased slightly though the maximum has not changed.

**e** Box plots are reasonably symmetrical so no reason to assume not normal, so reasonable to use the t-test.

<table>
<thead>
<tr>
<th>Differences</th>
<th>0.8</th>
<th>0</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>0.6</th>
<th>-0.2</th>
</tr>
</thead>
</table>

$H_0 : \mu = 0$, $H_1 : \mu > 0$

$p$-value $= 0.00518 < 0.05$, so the result is significant and there is significant evidence to say that the canteen has improved.

**5 a** $H_0 : \rho = 0$, $H_1 : \rho > 0$

$p$-value $= 0.00246$,

If testing $H_1 : \rho \neq 0$

$p$-value is $0.004916 < 0.05$.

In either case this is significant so very strong evidence to reject $H_0$ that there is no correlation between height and salary.

**b** The mix of males and females implies that heights are not likely to have been normal (i.e. they may be bimodal). Females on average are shorter than males, and this factor has not been considered. The positive relationship between height and salary might reflect that women are paid less than men on average. It would be better to stratify by gender and alter the hypothesis accordingly.

**c** Group the sample into salary bands and use a chi-squared test for independence between gender and salary or use a 2-sample t-test on average salary with the null hypothesis that there is no difference between the salaries of men and women.

---

**Exercise 14P**

**1 a** Under $H_0$ that $\mu = 7$, $P(X \geq a) = 0.05 \Rightarrow a = 7.12$

$P(X \leq 7.12 \mid \mu = 7.1) = 0.608$

**b** Under $H_0$ that $\mu = 12.1$, $P(X \leq a) = 0.05 \Rightarrow a = 12.026$,

$P(X \geq 12.026 \mid \mu = 12.0) = 0.28$

**2 a i** Need to find smallest $a$ such that $P(X \geq a) \leq 0.05 \Rightarrow P(X \geq 22) = 0.0940 > 0.05$,

$P(X \geq 23) = 0.0435 \leq 0.05$ so critical region is $X \geq 23$ and $P(X \geq 23 \mid \rho = 0.6) = 0.0435$

**ii** $P(X < 23 \mid \rho = 0.68) = 0.792$

**b i** Need to find largest $a$ such that $P(X \leq a) \leq 0.05 \Rightarrow P(X \leq 12) = 0.0386 \leq 0.05$,

$P(X \leq 13) = 0.0751 > 0.05$ so critical region is $X \leq 12$ and $P(X \leq 12 \mid \rho = 0.45) = 0.0386$
ii $P(X > 12 | \mu = 0.43) = 0.935$

3 a i Here, $X \sim \text{Po}(30 \times 4.6) = \text{Po}(138)$.

Need to find largest $a$ such that $P(X \leq a) \leq 0.05$. $P(X \leq 118) = 0.0458 \leq 0.05$

$P(X \leq 119) = 0.0551 > 0.05$ so critical region is $X \leq 118$ and $P(X \leq 118 | \mu = 4.6) = 0.0458$

ii Here, $X \sim \text{Po}(30 \times 4.5) = \text{Po}(135)$.

$P(X > 118 | \mu = 4.5) = 0.925$

b i Here, $X \sim \text{Po}(20 \times 2.1) = \text{Po}(42)$.

Need to find smallest $a$ such that $P(X \geq a) \leq 0.05$. $P(X \geq 53) = 0.0422 \leq 0.05$ so critical region is $X \geq 54$ and $P(X \geq 54 | \mu = 2.1) = 0.0422$

ii Here, $X \sim \text{Po}(20 \times 2.8) = \text{Po}(56)$.

$P(X < 54 | \mu = 2.8) = 0.377$

4 a Need to find $a$ such that $P(4.5 \leq X \leq 54) = 0.95$.

$P(54 \leq X) = 0.025 \Rightarrow c = 48.3$

$P(54 \leq X) = 0.025 \Rightarrow c = 53.7$. So critical regions are $X < 48.3$ and $X > 53.7$.

b $P(48.3 \leq X \leq 53.7 | \mu = 51.5) = 0.9364$

5 a Under $H_0$ that $\mu = 50$, Need to find $a$ such that $P(54 \geq a) = 0.05$.

$P(X \geq a) = 0.05 \Rightarrow a = 53.121$. So critical region is $X > 53.121$

b $P(X \geq 53.121 | \mu = 54) = 0.322$

c Increase the sample size or increase the significance level of the test.

6 a $H_0$ : the number of people with a car follows a $\text{B}(50,0.3)$ distribution

b Need to find smallest $a$ such that $P(X \geq a) \leq 0.05$. $P(X \geq 20) = 0.0848 \leq 0.05$

$P(X \geq 21) = 0.0478 \leq 0.05$ so critical region is $X \geq 21$

c i $P(X < 21 | \mu = 0.4) = 0.561$  

ii $P(X < 21 | \mu = 0.5) = 0.101$

7 a The hours chosen for the sample need to be independent. If they were adjacent to each other, then an event such as a large car crash could mean that several of the hours had more arrivals than usual.

b $H_0 : \mu = 82, H_1 : \mu < 82$

c A type I error would occur if the average number of patients has not fallen but the sample mean $\leq 75$ such that the null hypothesis is rejected. This would mean the phone line is continued unnecessarily. For $X \sim \text{Po}(82)$, $P(X \leq 75 | \mu = 82) = 0.239$.

d $P(X \geq 76 | \mu = 78) = 1 - P(X \leq 75 | \mu = 78) = 0.605$.

This is the probability of accepting the null hypothesis that the mean has not fallen and discontinuing the phone line even though it has had a positive effect.

8 a $H_0 : \mu = 1.2, H_1 : \mu \neq 1.2$  

b 0.05
c  Need to find $b$ and $c$ such that $P(b \leq \bar{X} \leq c) = 0.95$. When $\mu = 1.2$, $\sigma = \frac{0.1}{\sqrt{20}}$

\[ P(\bar{X} \leq b) = 0.025 \Rightarrow b = 1.1562. \quad P(\bar{X} \leq c) = 0.975 \Rightarrow c = 1.2438 \]

So critical regions are $\bar{X} < 1.1562$ and $\bar{X} > 1.2438$.

$P(1.1562 \leq \bar{X} \leq 1.2438 | \mu = 1.17) = 0.731$

Exercise 14Q

1 a  Under the null hypothesis that the shapes are guessed randomly the probability of 3 correct is $0.2^3 = 0.008$

b  The probability he is quoting is the probability of not getting all three cards right if guessing randomly, not the probability he is not just guessing randomly.

c  How many previous tests had Bruno taken and what were the results? Is there any evidence of trickery or conspiracy between Bruno and the tester? Are the cards marked? Is the distribution of cards in the pack even? Were the cards randomly drawn?

d

e i  $P(D) = P(H_0) \times P(D | H_0) + P(H_1) \times P(D | H_1) = 0.99 \times 0.008 + 0.01 \times 1 = 0.1792$

ii  $P(H_1 | D) = \frac{P(H_1) \times P(D | H_1)}{P(D)} = \frac{0.01 \times 1}{0.1792} = 0.5580$

f i  The researcher’s belief that the probability of ESP existing is 0.01.

ii  $P(D) = P(H_0) \times P(D | H_0) + P(H_1) \times P(D | H_1) = 0.999 \times 0.008 + 0.001 \times 1 = 0.008992$

$P(H_1 | D) = \frac{P(H_1) \times P(D | H_1)}{P(D)} = \frac{0.001 \times 1}{0.008992} = 0.111$

2 a i  $X \sim B(100,0.01)$,

<table>
<thead>
<tr>
<th>$P$</th>
<th>$X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0794</td>
<td>3</td>
</tr>
<tr>
<td>0.0794</td>
<td>2</td>
</tr>
</tbody>
</table>

ii  No, more information is needed.

b i  $P(X \geq 3) = P(H_0) \times P(X \geq 3 | H_0) + P(H_1) \times P(X \geq 3 | H_1)$,

\[ = 0.9 \times 0.0794 + 0.1 \times 1 = 0.171 \]

$P(H_1 | X \geq 3) = \frac{P(H_1) \times P(X \geq 3 | H_1)}{P(X \geq 3)} = \frac{0.1 \times 1}{0.171} = 0.583$
The probability of bad practice is just under 60%. Though not significant, further checks might be advisable.

3 a Need to find \( a \) such that \( P(\bar{X} \leq a) = 0.05 \), given \( \mu = 5.2, \sigma = \frac{1.2}{\sqrt{16}} \)

\[
P(\bar{X} \leq a) \Rightarrow a = 4.7065 . \text{ So critical region is } \bar{X} < 4.71 .
\]

b \( P(\bar{X} \geq 4.7065 | \mu = 4.6) = 0.361 \)

c \( P(X \leq 4.7065) = P(\text{inside}) \times P(X \leq 4.7065 | \text{inside}) + P(\text{outside}) \times P(X \leq 4.7065 | \text{outside}) \)

\[
= 0.9 \times 0.05 + 0.1 \times (1 - 0.361) = 0.109
\]

d \( P(\text{inside} | X \leq 4.7065) = \frac{P(\text{inside}) \times P(X \leq 4.7065 | \text{inside})}{P(X \leq 4.7065)} = \frac{0.9 \times 0.05}{0.109} = 0.413 \)

Chapter review

1 a

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
<th>( R_x )</th>
<th>( R_y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>151</td>
<td>17.5</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>153</td>
<td>18</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>153</td>
<td>16.5</td>
<td>2.5</td>
<td>10</td>
</tr>
<tr>
<td>154</td>
<td>16</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>155</td>
<td>15.4</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>159</td>
<td>13.2</td>
<td>6</td>
<td>4.5</td>
</tr>
<tr>
<td>162</td>
<td>14</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>164</td>
<td>13.7</td>
<td>8.5</td>
<td>6</td>
</tr>
<tr>
<td>164</td>
<td>13.2</td>
<td>8.5</td>
<td>4.5</td>
</tr>
<tr>
<td>168</td>
<td>12.5</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>175</td>
<td>12</td>
<td>11</td>
<td>1.5</td>
</tr>
<tr>
<td>181</td>
<td>12</td>
<td>12</td>
<td>1.5</td>
</tr>
</tbody>
</table>

\( r_s = -0.9525 \)

b The correlation is strong and negative, so, the taller the person, the less time they take to run the 100 metres. The sample is fairly small. The time at which someone runs the 100m will also depend on factors such as their age, gender and physical fitness.

2 a \( H_0 : \) egg colour is independent of type of hen

\( H_1 : \) egg colour isn't independent of type of hen

b \( \frac{30 \times 42}{90} \times \frac{90}{90} = 14 \)

c Degrees of freedom = \((3 - 1) \times (2 - 1) = 2 \times 1 = 2 \)

d

<table>
<thead>
<tr>
<th>Expected</th>
<th>Leghorn</th>
<th>Brahma</th>
<th>Sussex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Worked solutions

| White eggs | 14 | 14 | 14 |
| Brown eggs | 16 | 16 | 16 |

\[ \chi^2 = 21.7 \]

\( p \)-value = 0.0000194

e 21.7 > 5.991 and 0.0000194 < 0.05. Significant so reject the null hypothesis. The colour of the eggs is not independent of the type of hen.

3  a  \( P(0 \text{ tails}) = \frac{1}{2} \times \frac{1}{2} = 0.25 \) so expected frequency for tossing 0 tails is

\[ 60 \times 0.25 = 15 \]

<table>
<thead>
<tr>
<th>Number of tails</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.25</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Expected Frequency</td>
<td>15</td>
<td>30</td>
<td>15</td>
</tr>
</tbody>
</table>

c 2 degrees of freedom

d \( H_0: \) the data fits a binomial distribution

\( H_1: \) the data doesn't fit a binomial distribution

\[ \chi^2 = 1.2 \text{ or } p \text{-value} = 0.549 \]

e 1.2 < 5.991 or 0.549 > 0.05, so not significant and no reason to reject the null hypothesis that the data fits a binomial distribution.

4  a \( H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2 \)

b Two tailed

c 0.678 > 0.05, not significant so no reason to reject the null hypothesis that there is no difference between the two groups.

5 Using the confidence interval function on the GDC

a \( (1.3733, 1.741) \)

b \( (21.611, 22.589) \)

6 \( H_0: \lambda = 15.0, H_1: \lambda > 15.0 \), assume \( H_0 \) is true, then \( P(X \geq 19) = 0.1805 > 0.05 \), not significant so insufficient evidence to reject \( H_0 \) that the number of hurricanes is still 7.5 on average.

7  a \( X \sim B(42, 0.82) \)

There are just two possible outcomes and the recovery of patients is likely to be independent of all other patients.

b \( H_0: p = 0.82, H_1: p < 0.82 \)

c Need to find \( r \) such that \( P(X \leq r) = 0.05 \). \( r = 29 \Rightarrow P(X \leq r) = 0.0293 \), \( r = 30 \Rightarrow P(X \leq r) = 0.0625 \) so the critical region is \( X \leq 29 \)

d 0.0293
8 a i \(0.05\) ii \(0.01\)

b i Need to find \(r\) such that \(P(\bar{X} \geq r) = 0.05\) when \(\mu = 30, \sigma = \frac{3}{\sqrt{4}}\)

\[
P(\bar{X} \leq r) = 0.95 \Rightarrow r = 32.4674, \text{ so the critical region is } \bar{X} \geq 32.47 \text{ and}
\]

\[
P(\bar{X} < 32.4674 | \mu = 32) = 0.622
\]

ii Need to find \(r\) such that \(P(\bar{X} \geq r) = 0.01\).

\[
P(\bar{X} \leq r) = 0.99 \Rightarrow r = 33.4895, \text{ so the critical region is } \bar{X} \geq 33.49 \text{ and}
\]

\[
P(\bar{X} < 33.4895 | \mu = 32) = 0.840
\]

c i Decrease ii Increase iii Decrease

9 a Test-retest means the same test is given to the same people after a period of time. If the test is reliable there should be a good correlation between the two sets of results.

b \[
r = 0.968
\]

This indicates that the test is very reliable.

c Let \(\mu_0\) be average difference between the scores on the first and second tests.

<table>
<thead>
<tr>
<th>Differences</th>
<th>0.2</th>
<th>0.4</th>
<th>0.4</th>
<th>0</th>
<th>-0.5</th>
<th>0.6</th>
<th>-0.3</th>
<th>0.6</th>
</tr>
</thead>
</table>

\[H_0 : \mu_0 = 0, \ H_1 : \mu_0 \neq 0\]

\[p\text{-value} = 0.133 > 0.05\], not significant at 10% so no reason to reject \(H_0\) that there has been no increase in the overall level of satisfaction.

Assumptions: The differences can be modeled by a normal distribution and the responses of those surveyed were independent of each other.

Exam-style questions

10 Let \(X\) be the number of heads.

a Probability of getting 4 heads with a fair coin is \(P(X = 4 | p = 0.5) = \left(\frac{1}{2}\right)^4 = 0.0625\).

b 0

11 a Use a chi-squared test for independence

\[H_0 : \text{favourite TV channel is independent of age}\]

\[H_1 : \text{favourite TV channel isn’t independent of age}\]

<table>
<thead>
<tr>
<th>Expected</th>
<th>Alpha</th>
<th>Beta</th>
<th>Peppa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 5 years old</td>
<td>12</td>
<td>21</td>
<td>27</td>
</tr>
<tr>
<td>Between 6 and 10 years</td>
<td>14</td>
<td>24.5</td>
<td>31.5</td>
</tr>
</tbody>
</table>
Degrees of freedom = \((3 - 1) \times (3 - 1) = 2 \times 2 = 4\)

\(p\)-value = 0.000279

0.000279 < 0.01. Significant so reject the null hypothesis. The favourite TV channel is not independent of age.

b Table above: values are valid as they are all >5

12a i Welsh mean \(\bar{x}_1 = 173.75\)

ii Scottish mean \(\bar{x}_2 = 177.8\)

b 2-sample t-test since variance is unknown.

\[ H_0: \mu_{\text{Welsh}} = \mu_{\text{Scottish}} \]
\[ H_1: \mu_{\text{Welsh}} < \mu_{\text{Scottish}} \]

\(p\)-value = 0.0214 < 0.05, so we reject the null hypothesis and conclude that there is sufficient evidence at the 5% level to conclude that Welsh policemen are shorter than Scottish policemen.

c 0.0214 > 0.01 so at the 1%, level we would accept \(H_0\).

13 One tailed test. \(H_0: \lambda = 20.0, \ H_1: \lambda < 20.0\).

Here \(X \sim \text{Po}(20 \times 6) = \text{Po}(120)\).

Assume \(H_0\) is true, then \(P(X \leq 100) = 0.0347 < 0.05\), significant so sufficient evidence to reject \(H_1\), suggesting that Narcissus is exaggerating.

14a If \(X\) has a mean of \(\mu\) and a standard deviation of \(\sigma\) then the mean of a sample that is sufficiently large (>30), has distribution \(\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\).

b Using the inbuilt function on the GDC for a confidence interval when the variance is known (Z) (4.804, 5.196)

15 A paired t test is used because we wish to directly compare measurements associated with the same cars. Let \(\mu_0\) be average difference between the scores on the first and second tests.

<table>
<thead>
<tr>
<th>Differences</th>
<th>0.2</th>
<th>0</th>
<th>0.4</th>
<th>-0.1</th>
<th>0.6</th>
<th>0</th>
<th>0.2</th>
<th>0.1</th>
<th>-0.1</th>
<th>0.3</th>
</tr>
</thead>
</table>

\(H_0: \mu_0 = 0, \ H_1: \mu_0 > 0\)

\(p\)-value = 0.0264 < 0.05, significant at 5% so evidence to reject \(H_0\) that front wheels wear at the same rate as the rear wheels.

16 \(H_0: \text{there is not a linear relationship}, \ \left(\rho = 0\right)\ \ H_1: \text{there is a linear relationship}, \ \left(\rho \neq 0\right)\)

\(p\)-value is 0.7996 > 0.05 so not significant, so not enough evidence to reject the null hypothesis that there is no linear relationship between a female’s height and the number of hours spent on social media.
17 Use a Chi squared goodness of fit test.

$H_0$ : toys appear in the colour ratio 3:4:2:1

$H_1$ : toys don’t appear in the distribution stated

Degrees of freedom is 4-1=3

<table>
<thead>
<tr>
<th>Colour</th>
<th>Blue</th>
<th>Pink</th>
<th>Purple</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed</strong></td>
<td>32</td>
<td>37</td>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td><strong>Expected</strong></td>
<td>30</td>
<td>40</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

$p$-value = 0.7510

0.7510 > 0.05. Not significant so do not reject the null hypothesis. The colour of the toys follows the stated distribution.