1. The daily takings \( Y \) at a school tuckshop have mean £32-62 and standard deviation £7-32.

   (a) Give two reasons why the distribution of \( Y \) can not follow an exact Normal distribution. (2)

   (b) Explain briefly why the Normal distribution may still provide a useful model for the daily takings. (2)

2. An outline of the stages in developing a statistical model are listed below with stages 4 and 7 missing.

   Stage 1. The recognition of a real-world problem.
   Stage 2. A statistical model is devised.
   Stage 3. The model is used to make predictions
   Stage 4. ____________________________
   Stage 5. Comparisons are made against the model.
   Stage 6. Statistical techniques are used to test how well the model describes the real-world problem.
   Stage 7. ____________________________

   Suggest what stages 4 and 7 are. (2)

3. Give two reasons why simulations are used in the process of designing the wings for a new airplane. (3)
<table>
<thead>
<tr>
<th>Question Number</th>
<th>Scheme</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (a)</td>
<td>It is not continuous and it cannot take negative values.</td>
<td>B1 B1 (2)</td>
</tr>
<tr>
<td>(b)</td>
<td>There are a very large number of possible values so that the approximation of discrete by continuous is not a major problem, and 0 is more than 4 standard deviations below the mean so the truncation at zero is negligible</td>
<td>B1 B1 (2)</td>
</tr>
<tr>
<td>2.</td>
<td>Stage 4 Experimental or observational data is collected Stage 7 Refine the model</td>
<td>B1 B1 2</td>
</tr>
<tr>
<td>3.</td>
<td>It will save time and money and is much safer to conduct trials in simulations than with prototype wings.</td>
<td>B1 B1 B1 3</td>
</tr>
</tbody>
</table>
1. The length of time, in minutes, that customers spend in a department store is recorded. The results are shown below.

4, 15, 7, 26, 34, 22, 3, 42, 26, 19, 6, 18, 21, 8, 11, 39, 28, 6

Draw a stem and leaf diagram to represent these data. (4)

2. The marks for a class of 15 students has \( \sum x = 930 \), and \( \sum x^2 = 59146 \).

(a) Calculate the mean and standard deviation of the marks for this class. (4)

A second class of 20 students on the same examination had a mean of 61.5 and standard deviation of 14.4 marks.

(b) Calculate \( \sum x^2 \) for the second class. (4)

(c) Calculate the mean mark of all the students in the two classes. (3)

(d) Compare the performance of the two classes on this examination. (2)

3. Airlines record the delays in arrival times of flights into each airport. During one week, for one airline into one airport, 32% of all flights had no delay, the longest delay was 50 minutes and half of all flights had delays of no more than 7 minutes. A quarter of all delays were at least 22 minutes, but only one was more than half an hour.

An outlier is an observation that falls either 1.5 \( \times \) (interquartile range) above the upper quartile or 1.5 \( \times \) (interquartile range) below the lower quartile.

(a) On graph paper, draw a box plot to represent these data. (7)

(b) Comment on the distribution of delays. Justify your answer. (2)

The boxplot below summarises the delays for the same airline at a different airport during the same week.

(c) Compare the delays to flights arriving at the two airports during this week. (2)

4. The times, correct to the nearest minute, taken by a group of cadets to complete a route march are summarized in the table.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of cadets</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

(a) Draw a histogram to illustrate these data (4)

(b) Calculate an estimate of the mean time taken by the cadets. (4)
5. The following stem and leaf diagram shows the heights of a sample of plants of a certain species.

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>4</th>
<th>1 means 41 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7 9 9 (3)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 3 5 8 8 (5)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2 4 5 5 7 8 9 (7)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0 0 0 0 2 3 6 7 8 9 (10)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0 1 3 3 3 3 5 6 7 8 9 9 (14)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3 4 4 5 6 7 7 9 (8)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0 0 2 (3)</td>
<td></td>
</tr>
</tbody>
</table>

(a) Write down the modal height. (1)
(b) Find (i) the 3rd decile and (ii) the 43rd percentile of these data. (3)
(c) Find the three quartiles for these data. (3)

The quartile skewness coefficient is defined by \( \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \).
(d) Calculate the quartile skewness coefficient for these data, and comment on its value. (2)

6. The cholesterol levels of a group of 137 athletes are summarised in the table.

<table>
<thead>
<tr>
<th>cholesterol level, c (mg/dL)</th>
<th>Number of athletes, f</th>
</tr>
</thead>
<tbody>
<tr>
<td>160 ( \leq c &lt; 180 )</td>
<td>25</td>
</tr>
<tr>
<td>180 ( \leq c &lt; 200 )</td>
<td>32</td>
</tr>
<tr>
<td>200 ( \leq c &lt; 220 )</td>
<td>26</td>
</tr>
<tr>
<td>220 ( \leq c &lt; 240 )</td>
<td>31</td>
</tr>
<tr>
<td>240 ( \leq c &lt; 260 )</td>
<td>16</td>
</tr>
<tr>
<td>260 ( \leq c &lt; 280 )</td>
<td>7</td>
</tr>
</tbody>
</table>

The midpoints, \( m \), of each interval are calculated and the data is coded by \( C = \frac{m - 170}{20} \), giving summary statistics \( \Sigma Cf = 276 \) and \( \Sigma C^2 f = 846 \).

(a) Calculate the mean and standard deviation of \( C \). (5)
(b) Calculate the mean and standard deviation of \( c \). (4)
<table>
<thead>
<tr>
<th>Question Number</th>
<th>Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Time in minutes</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>2.</td>
<td>(a) Mean = $\frac{930}{15} = 62$</td>
</tr>
<tr>
<td></td>
<td>s.d. = $\sqrt{\frac{59146}{15} - 62^2} = \sqrt{99.0666}$</td>
</tr>
<tr>
<td></td>
<td>= 9.95</td>
</tr>
<tr>
<td></td>
<td>(b) $\text{Var}(X) = \frac{\sum x^2}{20} - \left(\frac{\sum x}{20}\right)^2$</td>
</tr>
<tr>
<td></td>
<td>$\sum x^2 = 20 \times (14.4^2 + 61.5^2)$</td>
</tr>
<tr>
<td></td>
<td>= 79792</td>
</tr>
<tr>
<td></td>
<td>(c) Total marks for second class = 20 $\times$ 61.5 = 1230</td>
</tr>
<tr>
<td></td>
<td>Total marks for both classes = 1230 + 930 = 2160</td>
</tr>
<tr>
<td></td>
<td>Mean mark for both classes = 2160 $\div$ 35 = 61.7</td>
</tr>
<tr>
<td></td>
<td>(d) The two classes have very similar performances on average, but the marks in the second class are much more spread out.</td>
</tr>
<tr>
<td>3.</td>
<td>(a) Least value = 0, $Q_1 = 0$, $Q_2 = 7$, $Q_3 = 22$, greatest value = 50</td>
</tr>
<tr>
<td></td>
<td>Boxplot drawn to scale</td>
</tr>
<tr>
<td></td>
<td>$Q_3 + 1.5 (Q_3 - Q_1) = 55 \rightarrow$ no outlier above upper quartile</td>
</tr>
<tr>
<td></td>
<td>(b) $Q_3 - Q_2 = 15$, $Q_2 - Q_1 = 7$</td>
</tr>
<tr>
<td></td>
<td>$Q_3 - Q_2 &gt; Q_2 - Q_1$ so distribution has positive skew</td>
</tr>
<tr>
<td></td>
<td>(c) The first airport has almost a third of flights on time; the second has none. The first airport has a quarter of flights delayed by at least 22 minutes but the second has no delays over 19 minutes.</td>
</tr>
</tbody>
</table>
4. (a)  

<table>
<thead>
<tr>
<th>Class width</th>
<th>50</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. cadets</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Freq. density</td>
<td>0.24</td>
<td>0.8</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

(b) Histogram drawn correct and to scale

Mid points 114.5 144.5 154.5 164.5 174.5 204.5

Mean = \((114.5 \times 12 + 144.5 \times 8 + \ldots + 204.5 \times 5) \div 41 = 150.6\) mins

5. (a) 73 cm

(b) (i) \(0.3 \times 50 = 15\), so 3rd decile is between 15th and 16th value

\[ D_3 = 59.5 \]

(ii) \(0.43 \times 50 = 21.5\), so 43rd percentile is 22nd value

\[ P_{43} = 66 \]

(c) \(\frac{1}{4} \times 50 = 12.5\), so Q1 is 13th value

\[ Q_1 = 57 \]

\(\frac{1}{2} \times 50 = 25\), so Q2 is between 25th and 26th values

\((69 + 70) \div 2 = 69.5\), so Q2 = 69.5

\(\frac{3}{4} \times 50 = 37.5\), so Q3 is 38th value

\[ Q_3 = 79 \]

\[ \frac{Q_3 - 2Q_2 + Q_1}{Q_1 - Q_3} = \frac{79 - 2 \times 69.5 + 57}{79 - 57} = -0.14 \]

Slight negative skew
6. (a) Mean of \( C = \frac{276}{137} = 2.02 \)  
Use of \( \frac{\sum Cf}{\sum f} \)

Variance of \( C = \frac{846}{137} - 2.02^2 = 2.1166 \)  
Use of \( \frac{\sum C^2f}{\sum f} - \bar{C}^2 \)

s.d. = \( \sqrt{2.1166} = 1.455 \)

(b) \( \bar{C} = \frac{\bar{m} - 170}{20} \)  
\( \bar{m} = 20\bar{C} + 170 = 20 \times 2.02 + 170 = 210.4 \)

s.d. of \( m \) = s.d. of \( C \times 20 = 1.455 \times 20 = 29.1 \)
1. A bag contains 7 green discs and 3 red discs. A disc is selected at random from the bag and its colour is recorded. The disc is not replaced. A second disc is selected at random and its colour is recorded.

(a) Draw a tree diagram to represent the information.  

Find the probability that

(b) the second disc selected is green,  

(c) the discs selected are different colours, given that the first disc selected is green.  

2. Of the pupils who took Mathematics in a certain school one year, 50% of them took Physics, 35% of them took Chemistry and 40% took both neither Physics nor Chemistry.

One of the pupils taking Mathematics is chosen at random.

(a) Find the probability that the pupil took both Physics and Chemistry.

(b) Given that the pupil took exactly one of Physics and Chemistry, find the probability that it was Physics.

3. The events $A$ and $B$ are independent such that $P(A) = 0.6$ and $P(B) = 0.4$.

Find

(a) $P(A \cap B)$,  

(b) $P(A' \cap B)$,  

(c) $P(A|B)$.

4. A blue bag has discs numbered 1, 2, 4, 8 and 16 in it. A red bag has discs numbered 1, 2 and 3 in it.

A coin is tossed, and a disc is drawn at random from the blue bag if it shows a Head, otherwise a disc is drawn from the red bag.

Find the probability that

(a) the disc drawn has the number  

(i) 8  

(ii) 2  

(b) the red bag was used given that the disc had the number 2.
5. A pharmaceutical company is trialling a new drug. Patients are chosen at random to be given the existing drug, the new drug or a placebo [a similar looking tablet containing no drug]. The doctor notes how the patient is at the end of the course of treatment, and the results are summarised in the table.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Worse</th>
<th>No change</th>
<th>Improved</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing drug</td>
<td>7</td>
<td>87</td>
<td>363</td>
<td>457</td>
</tr>
<tr>
<td>New drug</td>
<td>16</td>
<td>5</td>
<td>391</td>
<td>412</td>
</tr>
<tr>
<td>Placebo</td>
<td>93</td>
<td>68</td>
<td>214</td>
<td>375</td>
</tr>
<tr>
<td>Total</td>
<td>116</td>
<td>160</td>
<td>968</td>
<td>1244</td>
</tr>
</tbody>
</table>

A patient is selected at random.

A is the event that the patient was given the new drug.
B is the event that the patient improved during the course of treatment.
C is the event that the patient was worse after the course of treatment.
A' is the event ‘Not A’

(a) Write down the value of
(i) $P(A)$
(ii) $P(A \cap B)$
(iii) $P(B|A)$

(b) Find the probability that the patient improved if they had been given the existing drug.

(c) Dr Jeckyll says that a much higher proportion of patients improve with the new drug, but Mr Hyde, the managing director, refuses to sanction further trials of the drug in its existing form. Give a reason why Mr Hyde might have made this decision.
### Question 1

**(a)**

- Green: \( \frac{7}{10} \)
- Red: \( \frac{3}{10} \)

- Green: \( \frac{3}{9} \)
- Red: \( \frac{6}{9} \)

- Green: \( \frac{7}{9} \)
- Red: \( \frac{2}{9} \)

- Green: \( \frac{21}{90} \)
- Red: \( \frac{6}{90} \)

**Diagram**

- First branches
- Second branches

- Use of \( P(A) \times P(B) \)

**Scheme Marks**

- Diagram: M1
- First branches: A1
- Second branches: A1

**Marks**

- 3

**(b)**

\[
\frac{\frac{7}{10} \times \frac{3}{9} + \frac{3}{10} \times \frac{7}{9}}{\frac{90}{90}} = \frac{7}{10}
\]

**Marks**

- 2

**(c)**

\[
\frac{\frac{7}{10} \times \frac{3}{9}}{\frac{7}{10} + \frac{3}{9}} = \frac{1}{3}
\]

**Marks**

- 2

### Question 2

**(a)**

- \( A \) is event student takes Physics
- \( B \) takes Chemistry

\[
P(A \cup B) = 1 - P(A' \cap B') = 1 - 0.4 = 0.6
\]

\[
P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.5 + 0.35 - 0.6 = 0.25
\]

**Marks**

- 2

**(b)**

\[
P(A' \cap B') = P(A) - P(A \cap B) = 0.5 - 0.25 = 0.25
\]

\[
P(A' \cap B) = P(B) - P(A \cap B) = 0.35 - 0.25 = 0.1
\]

\[
\frac{P(A \cap B')}{P(A \cap B) + P(A' \cap B)} = \frac{0.25}{0.25 + 0.1} = 0.714
\]

**Marks**

- 4

### Question 3

**(a)**

\[
0.6 \times 0.4 = 0.24
\]

**Use of** \( P(A \cap B) = P(A) \times P(B) \)

**Marks**

- 2

**(b)**

\[
0.4 - 0.24 = 0.16
\]

**Use of** \( P(B) = P(A \cap B) + P(A' \cap B) \)

**Marks**

- 2

**(c)**

\[
\frac{0.24}{0.4} = 0.6
\]

**Use of** \( P(A/B) = \frac{P(A \cap B)}{P(B)} \)

**Marks**

- 2
### 4. (a)

(i) \( P(\text{blue } \cap \text{8}) = \frac{1}{2} \times \frac{1}{5} = 0.1 \)  
Use of \( P(A \cap B) = P(A/B) \times P(B) \)  
M1 A1 (2)

(ii) \( P(\text{blue } \cap \text{2}) = \frac{1}{2} \times \frac{1}{5} = 0.1 \)  
\( P(\text{red } \cap \text{2}) = \frac{1}{2} \times \frac{1}{5} = \frac{1}{10} \)  
\( P(2) = \frac{1}{10} + \frac{1}{6} = \frac{4}{15} \)  
M1 A1 (3)

(b) \( P(\text{red } / \text{2}) = \frac{\frac{1}{6}}{\frac{1}{10}} = \frac{5}{8} \)  
Use of \( P(A/B) = \frac{P(A \cap B)}{P(B)} \)  
M1 A1  
A1 (3)  
8

### 5. (a)

(i) \( \frac{412}{1244} \)  
(ii) \( \frac{391}{1244} \)  
(iii) \( \frac{391}{412} \)  
M1 A1  
M1 A1  
M1 A1 (6)

(b) \( \frac{363}{457} = 0.794 \)  
M1 A1 (2)

(c) \( P(\text{worse after existing drug}) = \frac{7}{457} = 0.0153 \)  
M1  
\( P(\text{worse after new drug}) = \frac{16}{412} = 0.388 \)  
A1  
More than double the proportion get worse after new drug  
B1 (3)  
11
1. The CensusAtSchool project has data on various physical measurements of pupils who have completed their questionnaires.

   The table shows the height, \( x \) (cm), and the foot length, \( y \) (cm), for a sample of pupils.

   \[
   \begin{array}{cccccccccc}
   x & 150 & 139 & 133 & 147 & 146 & 135 & 140 & 136 & 133 & 132 & 138 \\
   y & 23 & 21 & 18 & 21 & 22 & 22 & 23 & 20 & 19 & 21 & 21 \\
   \end{array}
   \]

   \[
   \begin{align*}
   \sum x &= 1529, & \sum x^2 &= 212913, & \sum y &= 231, & \sum y^2 &= 4875, & \sum xy &= 32169 \\
   \end{align*}
   \]

   (a) Draw a scattergraph to represent this information. (2)

   (b) Calculate the correlation coefficient between \( x \) and \( y \). (4)

   (c) Interpret your answer to part (b). (2)

2. The score on training, \( (T) \), and on the first performance assessment after starting work, \( (W) \), for a group of 12 new employees at a firm gave the following summary statistics.

   \[
   \begin{align*}
   \sum (t - \bar{T})^2 &= 1268.25, & \sum (w - \bar{w})^2 &= 404.92, & \sum (t - \bar{T})(w - \bar{w}) &= 137.75 \\
   \end{align*}
   \]

   (a) Calculate the correlation coefficient between the training score and the employees’ first performance assessments. (4)

   (b) Comment on whether the score on training is a good indicator of how well an employee will perform when they start work. (2)

   (c) The manager of the training programme says that the maximum score in the two assessments are not the same. He claims that the correlation would be higher if the marks were scaled so both were out of 100. Explain why he is wrong. (2)

3. A specialist physical trainer is brought in by the manager of a boxing team who feels there are weaknesses in the boxers’ preparation.

   Before planning a training programme, the trainer decided to see how many repetitions of a hand strengthening exercise the boxers were able to do without stopping for a rest.

   The results were as follows:

   \[
   \begin{array}{cccccccccc}
   \text{Boxer} & \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} & \text{H} \\
   \text{Left hand, } x & 27 & 35 & 28 & 41 & 36 & 22 & 28 & 37 \\
   \text{Right hand, } y & 32 & 39 & 31 & 38 & 37 & 25 & 27 & 40 \\
   \end{array}
   \]

   The data in the table can be summarised as follows.

   \[
   \begin{align*}
   \sum x &= 254, & \sum y &= 269, & \sum x^2 &= 8352, & \sum y^2 &= 9273, & \sum xy &= 8773 \\
   \end{align*}
   \]

   (a) Calculate the correlation coefficient. (4)

   (b) Comment briefly on its value. (2)
4. Two trainee wine buyers are given ten wines to taste on their first day at work and asked to score the wines according to various criteria. The scores for each trainee, \( x \) and \( y \), for each wine is given in the table.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>58</td>
<td>79</td>
<td>52</td>
<td>68</td>
<td>56</td>
<td>67</td>
<td>53</td>
<td>88</td>
<td>71</td>
<td>64</td>
</tr>
<tr>
<td>y</td>
<td>67</td>
<td>75</td>
<td>65</td>
<td>65</td>
<td>71</td>
<td>69</td>
<td>59</td>
<td>65</td>
<td>75</td>
<td>69</td>
</tr>
</tbody>
</table>

The data in the table can be summarised as follows.

\[
\sum x = 656, \quad \sum y = 680, \quad \sum x^2 = 44,248, \quad \sum y^2 = 46,458, \quad \sum xy = 44,798
\]

(a) Calculate the correlation coefficient. (4)

(b) Comment briefly on its value. (2)

(c) Explain why you would expect the correlation coefficient to be higher if a similar test was given to the two trainees at the end of their training programme. (2)
### Question 1

1. (a) Axes; Points

2. (b) \[ S_{xy} = 32169 - \frac{1529 \times 231}{11} = 60 \]
   \[ S_{xx} = 212913 - \frac{1529^2}{11} = 382 \]
   \[ S_{yy} = 4875 - \frac{231^2}{11} = 24 \]
   \[ r = \frac{60}{\sqrt{382 \times 24}} = 0.627 \]

(c) There is moderately strong (positive) correlation so pupils with larger feet tend to be taller.

### Question 2

2. (a) \[ S_{nw} = 137.75 \quad S_{tt} = 1268.25 \quad S_{ww} = 404.92 \]
   \[ r = \frac{137.75}{\sqrt{1268.25 \times 404.92}} = 0.192 \]

(b) If the two scores are measuring the same thing you would expect a strong positive correlation, so this weak correlation suggests that the training score gives very little indication of how well they will do when they start work.

(c) Correlation coefficients are independent of scale, so this will not make a difference.
3. (a) \[ S_{xy} = 8773 - \frac{254 \times 269}{8} = 232.25 \] 
\[ S_{xx} = 8352 - \frac{254^2}{8} = 287.5 \] 
\[ S_{yy} = 9273 - \frac{269^2}{8} = 227.875 \] 
\[ r = \frac{232.25}{\sqrt{287.5 \times 227.875}} = 0.907 \] 
(b) very strong correlation, so boxers who are strong in one hand are likely to be strong in the other as well.

4. (a) \[ S_{xy} = 44798 - \frac{656 \times 680}{10} = 190 \] 
\[ S_{xx} = 44248 - \frac{656^2}{10} = 1214.4 \] 
\[ S_{yy} = 46458 - \frac{680^2}{10} = 218 \] 
\[ r = \frac{190}{\sqrt{1214.4 \times 218}} = 0.369 \] 
(b) This is only moderate correlation, but they are supposed to be measuring the same thing, so you would expect it to be higher.
(c) With training, the trainees should understand the criteria better and be more consistent in applying them, so the correlation should be stronger, and the pmcc should therefore be higher.
1. A company manufactures soap powder. They give a number of identical blouses to some customers and ask them to keep a record of how often the blouses are washed. After 3 months, the company goes back to the customers and notes how many times the blouse had been washed, and the strength of colour in the blouse.

<table>
<thead>
<tr>
<th>Number of washes, $n$</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>4</th>
<th>10</th>
<th>13</th>
<th>8</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colour strength, $s$</td>
<td>17</td>
<td>14</td>
<td>9</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

The data in the table can be summarised as follows.

\[ \sum n = 63, \quad \sum s = 94, \quad \sum n^2 = 553, \quad \sum s^2 = 1188, \quad \sum ns = 685 \]

(a) Explain why $n$ is the explanatory variable in this situation. (1)

(b) Draw a scatter diagram for the data. (3)

(c) Calculate exact values of $S_{ns}$ and $S_{nn}$ (4)

(d) Calculate the equation for the regression line of colour strength on the number of washes. Draw the regression line on your scatter diagram. (4)

(e) Give an interpretation of the slope and intercept of the regression line. (2)

(f) One of the customers who was given a blouse had washed it 8 times, but her daughter was wearing it when the company rep called so he could not get the colour strength. Estimate what the colour strength would be, and comment on the reliability of your estimate. (3)

(g) Explain why the equation in part (d) should not be used to predict the colour strength when one of the blouses has been washed 25 times. (2)

2. The number of hours of maintenance, $t$, required during a week is recorded for 10 machines in a factory, along with the age, $d$, of the machine in days.

The data are coded such that $x = \frac{(d - 450)}{30}$ and $y = t - 5$.

\[ \sum x = 275.9 \quad \sum y = 18.8 \quad \sum xy = 656.7 \quad S_{xx} = 1684.7 \]

(a) Find the equation of the regression line of $y$ on $x$ in the form $y = a + bx$. (6)

(b) Hence find the equation of the regression line of $t$ on $d$. (3)

(c) Predict the time taken on maintenance for another machine which is 1200 days old. (2)

(d) Give an interpretation of the slope and intercept of the regression line of $t$ on $d$. (2)

3. As part of an investigation into organised crime, the fraud squad purchase a number of copies of designer items from street traders. The table shows the price of the designer original, $x$, and the price paid for the copy, $y$ (both in £).

<table>
<thead>
<tr>
<th>Designer, $x$</th>
<th>1000</th>
<th>350</th>
<th>700</th>
<th>2500</th>
<th>380</th>
<th>165</th>
<th>590</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy, $y$</td>
<td>240</td>
<td>50</td>
<td>250</td>
<td>300</td>
<td>80</td>
<td>60</td>
<td>70</td>
</tr>
</tbody>
</table>

The data in the table can be summarised as follows (see next page).
\[ \Sigma x = 5685, \quad \Sigma y = 1050, \quad \Sigma x^2 = 8382225, \quad \Sigma y^2 = 227500, \quad \Sigma xy = 264100. \]

(a) Write down which is the explanatory variable in this investigation. 

(b) Draw a scatter diagram to illustrate these data.

(c) Find the equation of the appropriate regression line.

(d) Draw the regression line on your scatter diagram.

4. The goals for, \( x \), and against, \( y \), for each football team in the FA Carling Premiership in the 2006 – 07 season is shown in the table.

<table>
<thead>
<tr>
<th>Goals for, ( x )</th>
<th>83</th>
<th>64</th>
<th>57</th>
<th>63</th>
<th>57</th>
<th>52</th>
<th>47</th>
<th>52</th>
<th>45</th>
<th>52</th>
<th>43</th>
<th>44</th>
<th>38</th>
<th>37</th>
<th>32</th>
<th>34</th>
<th>29</th>
<th>35</th>
<th>38</th>
<th>37</th>
<th>32</th>
<th>34</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals against, ( y )</td>
<td>27</td>
<td>24</td>
<td>27</td>
<td>35</td>
<td>54</td>
<td>36</td>
<td>52</td>
<td>47</td>
<td>42</td>
<td>54</td>
<td>41</td>
<td>49</td>
<td>47</td>
<td>44</td>
<td>49</td>
<td>47</td>
<td>44</td>
<td>49</td>
<td>47</td>
<td>44</td>
<td>49</td>
<td>47</td>
<td>44</td>
</tr>
</tbody>
</table>

The data in the table can be summarised as follows.

\[ \Sigma x = 931, \quad \Sigma y = 931, \quad \Sigma x^2 = 46927, \quad \Sigma y^2 = 45959, \quad \Sigma xy = 41077. \]

(a) Draw a scatter diagram to illustrate these data.

(b) Find the equation of the regression line of \( y \) on \( x \).

(c) Draw the regression line on your scatter diagram.

(d) Explain why the regression line should not be used to predict the number of ‘goals against’ for a team who score 50 goals in the FA Carling Premiership in the 2008 – 09 season.

5. The variables \( R \) and \( S \) are known to be approximately linearly related. Forty pairs of values \((r, s)\) of the variables gave the following results:

\[ \Sigma r = 73.1, \quad \Sigma r^2 = 351.5, \quad \Sigma s = 356.2, \quad \Sigma s^2 = 2847.3, \quad \Sigma rs = 497.2 \]

(a) Find the equation of the regression line of \( S \) on \( R \).

(b) Predict the value of \( S \) when \( R \) takes the value 7.3.
Regression

Exam-style mark scheme

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Scheme</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (a)</td>
<td>The colour strength is affected by the number of washes rather than the other way round.</td>
<td>B1 (1)</td>
</tr>
<tr>
<td>(b)</td>
<td><img src="image" alt="Graph" /></td>
<td>Axes, points, labels B1 B1 B1 (3)</td>
</tr>
<tr>
<td>(c)</td>
<td>$S_{ns} = 685 - \frac{63 \times 94}{8} = -55.25$</td>
<td>M1 A1</td>
</tr>
<tr>
<td></td>
<td>$S_{nn} = 553 - \frac{63^2}{8} = 56.875$</td>
<td>M1 A1 (4)</td>
</tr>
<tr>
<td>(d)</td>
<td>$\beta = \frac{-55.25}{56.875} = -0.971$ Use of $\beta = \frac{S_{ns}}{S_{nn}}$</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>$\alpha = \frac{94}{8} - \left(-0.971 \times \frac{63}{8}\right) = 19.40$ Use of $\alpha = \bar{y} - \beta \bar{x}$</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>$s = -0.971n + 19.4$ Line drawn</td>
<td>A1 B1 (4)</td>
</tr>
<tr>
<td>(e)</td>
<td>The slope is the average loss of colour strength per extra wash, the intercept is the estimate of the colour strength when the blouses were new.</td>
<td>B1 B1 (2)</td>
</tr>
<tr>
<td>(f)</td>
<td>12 (11.6). The observed measurements are not very strongly correlated (see the scatter diagram) so this can only be regarded as a ballpark estimate.</td>
<td>M1 B1 B1 (3)</td>
</tr>
<tr>
<td>(g)</td>
<td>If you substitute 25 for $n$ you get a negative colour strength which does not make sense - the problem lies in trying to extrapolate a long way outside the observed data (the most was 13 washes). Also when the data are only weakly correlated as here, the estimate would be so unreliable as to be worthless when trying to extrapolate even if the same relationship continued to hold.</td>
<td>B1 B1 (2)</td>
</tr>
</tbody>
</table>
2. (a) \[ S_{xy} = 656.7 - \frac{275.9 \times 18.8}{10} = 138.0 \]
\[ \beta = \frac{138.0}{1684.7} = 0.0819 \]
\[ \alpha = \frac{18.8}{10} - 0.0819 \times \frac{275.9}{10} = -0.380 \]
\[ y = 0.0819x - 0.380 \]

(b) \[ t - 5 = 0.0819 \times \left( \frac{d - 450}{30} \right) - 0.380 \]
\[ t = 3.39 + 0.00273d \]

(c) \[ t = 3.39 + 0.00273 \times 1200 = 6.67 \text{ hours} \]

(d) the slope is the average extra weekly maintenance for each day’s use of a machine, and the intercept is an estimate of the weekly maintenance time for a new machine.

3. (a) The price of the original – the price of the copy follows after this.

(b) The scatter diagram is shown below:

\[ y = 0.1093x + 61.273 \]

(c) \[ S_{xy} = 264100 - \frac{5685 \times 1050}{7} = -588650 \]
\[ S_{xx} = 8382225 - \frac{5685^2}{7} = 3765193 \]
\[ \beta = \frac{-588650}{3765193} = -0.1563 \]
\[ \alpha = \frac{1050}{7} - (-0.1563) \times \frac{5685}{7} = 277 \]
\[ y = 277 - 0.156x \]

(d) see scatter diagram
4. (a) 

\[ y = -0.63x + 75.877 \]

(b) \[ S_{xy} = 41077 - \frac{931 \times 931}{20} = -2261.05 \]
\[ S_{xx} = 46927 - \frac{931^2}{20} = 3588.95 \]
\[ \beta = \frac{-2261.05}{3588.95} = -0.630 \]
\[ \alpha = \frac{931}{20} - (-0.630) \times \frac{931}{20} = 75.88 \]
\[ y = 75.88 - 0.630x \]

(c) see scatter diagram

(d) The data used to construct the model is for the season 2006–07 and this is 2 seasons later, so it is risky to assume that the same model will apply. The teams and players will almost certainly be different.

5. (a) \[ S_{rs} = 497.2 - \frac{73.1 \times 356.2}{40} = -153.7555 \]
\[ S_{rr} = 351.5 - \frac{73.1^2}{40} = 217.90975 \]
\[ \beta = \frac{-153.7555}{217.90975} = -0.7056 \]
\[ \alpha = \frac{356.2}{40} - (-0.7056) \times \frac{73.1}{40} = 10.19 \]
\[ s = 10.2 - 0.706r \]

(b) \[ s = 10.2 - 0.706 \times 7.3 = 5.04 \]
1. The random variable $X$ has probability function
   \[ P(X = x) = \frac{k}{x}, \quad x = 1, 2, 3, 4, 5 \]
   (a) Show that $k = \frac{60}{137}$. (2)
   (b) Find $\mu = E(X)$. (2)
   (c) $P(X < \mu)$. (3)
   (d) $E(5 - 2X)$. (3)

2. A discrete random variable $X$ has the probability function shown in the table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>2</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x)$</td>
<td>0.2</td>
<td>$a$</td>
<td>$0.8 - a$</td>
</tr>
</tbody>
</table>

(a) Given that $E(X) = 6.9$, find $a$. (3)

(b) Find $\text{Var}(X)$. (3)

(c) Find $P(X < \mu - \sigma)$, where $\mu = E(X)$, $\sigma^2 = \text{Var}(X)$ (4)

3. I have an unbiased die. For each of the following, state with a reason whether
   the random variable is a discrete uniform distribution.
   (a) $X = \text{the number of times I get a six if I roll it 30 times}$. (3)
   (b) $Y = \text{the larger of the scores showing on the top face when I roll it twice}$. (3)
   (c) $Z = 6 - \text{the score showing on the top face when I roll it}$. (3)

4. A discrete random variable $X$ has the cumulative distribution function shown in the table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>3</th>
<th>7</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(x)$</td>
<td>0.2</td>
<td>0.5</td>
<td>0.9</td>
<td>1</td>
</tr>
</tbody>
</table>

Find
   (a) $P(X = 9)$, (2)
   (b) $E(X)$, (4)
   (c) $\text{Var}(X)$, (3)
   (d) $E(5 - 2X)$, (2)
   (e) $\text{Var}(5 - 2X)$ (2)

5. The random variable $X$ has probability function
   \[ P(X = x) = \frac{(9 - 2x)}{16}, \quad x = 1, 2, 3, 4. \]
   (a) Construct a table giving the probability distribution of $X$. (3)
Find
(b) \( P(2 \leq X < 4) \) \( (2) \)
(c) \( E(X) \) \( (2) \)
(d) \( \text{Var}(X) \) \( (4) \)
(e) \( \text{Var}(2 - 3X) \) \( (2) \)

6. The random variable \( X \) has the discrete uniform distribution
\[ P(X = x) = \frac{1}{5}, \quad x = 1, 2, 3, 4, 5. \]
(a) Write down the value of \( E(X) \) and show that \( \text{Var}(X) = 2 \). \( (3) \)
Find
(b) \( E(3X - 4) \) \( (2) \)
(c) \( \text{Var}(6 - 5X) \) \( (2) \)

7. The discrete random variable \( X \) has probability distribution

<table>
<thead>
<tr>
<th>( x )</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X = x) )</td>
<td>0.1</td>
<td>0.25</td>
<td>0.2</td>
<td>0.15</td>
<td>0.3</td>
</tr>
</tbody>
</table>

(a) Calculate \( P(-1 < X \leq 2) \) \( (2) \)
(b) Find \( E(X) \) \( (3) \)
(c) Find \( E(4X^2 + 1) \) \( (3) \)

8. When two fair dice are thrown, the probability distribution of the larger score showing, \( X \), is

<table>
<thead>
<tr>
<th>( x )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X = x) )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{3}{36} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{7}{36} )</td>
<td>( \frac{9}{36} )</td>
<td>( \frac{11}{36} )</td>
</tr>
</tbody>
</table>

In a board game, a coin is tossed to decide how many dice are thrown in a player’s turn. If it shows Heads then one dice is used. If it shows Tails, then two dice are thrown and the larger score is taken.

(a) Find the probability distribution of \( Y \), the score on a player’s turn. \( (3) \)
(b) Calculate \( E(Y) \) \( (3) \)

9. Two fair dice are thrown. Let \( X = \) the product of the scores on the top faces.
(a) Give the probability distribution of \( X \). \( (4) \)
(b) Calculate \( E(X) \). \( (3) \)
## Discrete random variables

### Exam-style mark scheme

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Scheme</th>
<th>Marks</th>
</tr>
</thead>
</table>
| 1. (a) | \[k\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right) = 1\]  
\[k \times \frac{137}{60} = 1 \quad k = \frac{60}{137}\] | M1 |
| (b) | \[E(X) = \frac{60}{137}\left(1 \times \frac{1}{1} + 2 \times \frac{1}{2} + \ldots + 5 \times \frac{1}{5}\right) = \frac{60}{137} \times 5 = 2.19\] | M1 A1 (2) |
| (c) | \[P(X < \mu) = P(X < 2.2) = \frac{60}{137}\left(1 + \frac{1}{2}\right) = \frac{90}{137}\]  
\[M1 \quad M1 \quad A1 (3)\] |
| (d) | \[E(5 - 2X) = 5 - 2E(X) = 5 - 2 \times \frac{300}{137} = \frac{87}{137}\] | M1 M1 A1 (3) |
| 2. (a) | \[E(X) = 2 \times 0.2 + 5a + 10(0.8 - a) = 6.9\]  
\[8.4 - 5a = 6.9 \quad a = 0.3\] | M1 A1 |
| | \[E(X^2) = 2^2 \times 0.2 + 5^2 \times 0.3 + 10^2 \times 0.5 = 58.3 = \mu\]  
\[Var(X) = 58.3 - 6.9^2 = 10.69 = \sigma^2\] | M1 A1 (3) |
| | \[P(X < \mu - \sigma) = P(X < 6.9 - 3.3) = P(X < 3.6)\]  
\[= P(X = 2) = 0.2\] | M1 A1 |
| 3. (a) | No – you can get 0 to 30 but they are not equally likely outcomes. | B1 |
| (b) | No – the values 1 to 6 are not equally likely. | B1 |
| (c) | Yes – the values 0–5 are equally likely. | B1 |
| 4. (a) | \[P(X = 9) = F(9) - F(7) = 0.4\] | M1 A1 (2) |
| (b) | \[P(X = 3) = 0.2 \quad P(X = 7) = 0.3 \quad P(X = 10) = 0.1\]  
\[E(X) = 3 \times 0.2 + 7 \times 0.3 + 9 \times 0.4 + 10 \times 0.1 = 7.3\] | M1 A1 (4) |
| (c) | \[E(X^2) = 3^2 \times 0.2 + 7^2 \times 0.3 + 9^2 \times 0.4 + 10^2 \times 0.1 = 58.9\]  
\[Var(X) = 58.9 - 7.3^2 = 5.61\] | M1 A1 (3) |
| (d) | \[E(5 - 2X) = 5 - 2E(X) = 5 - 2 \times 7.3 = -9.6\]  
\[Var(5 - 2X) = 2^2 Var(X) = 4 \times 5.61 = 22.44\] | M1 A1 (2) |
5. (a) \[ P(X = 1) = \frac{9 - 2 \times 1}{16} = \frac{7}{16} \]
\[
\begin{array}{c|cccc}
 x & 1 & 2 & 3 & 4 \\
P(X = x) & \frac{7}{16} & \frac{5}{16} & \frac{3}{16} & \frac{1}{16} \\
\end{array}
\]

(b) \[ P(2 \leq X < 4) = P(X = 2) + P(X = 3) = \frac{5}{16} + \frac{3}{16} = \frac{1}{2} \]

(c) \[ E(X) = 1 \times \frac{7}{16} + 2 \times \frac{5}{16} + 3 \times \frac{3}{16} + 4 \times \frac{1}{16} = 1.875 \]

(d) \[ E(X^2) = 1^2 \times \frac{7}{16} + 2^2 \times \frac{5}{16} + 3^2 \times \frac{3}{16} + 4^2 \times \frac{1}{16} = 4.375 \]

\[ \text{Var}(X) = 4.375 - 1.875^2 = 0.859375 \]

(e) \[ \text{Var}(2 - 3X) = 9 \times 0.859375 = 7.734375 \]

6. (a) \[ E(X) = 3 \]

\[ E(X^2) = \frac{1}{5} (1^2 + 2^2 + 3^2 + 4^2 + 5^2) = 11 \]

\[ \text{Var}(X) = 11 - 3^2 = 2 \]

(b) \[ E(3X - 4) = 3E(X) - 4 = 3 \times 3 - 4 = 5 \]

(c) \[ \text{Var}(6 - 5X) = 5^2 \text{Var}(X) = 25 \times 2 = 50 \]

7. (a) \[ P(-1 < X \leq 2) = P(0) + P(1) + P(2) = 0.6 \]

(b) \[ E(X) = -1 \times 0.1 + 0 \times 0.25 + \ldots + 3 \times 0.3 = 1.3 \]

(c) \[ E(X^2) = (-1)^2 \times 0.1 + 0^2 \times 0.25 + \ldots + 3^2 \times 0.3 = 3.6 \]

\[ E(4X^2 + 1) = 4E(X^2) + 1 = 4 \times 3.6 + 1 = 16 \]

8. (a) \[ P(1 \text{ dice used}) = P(2 \text{ dice used}) = \frac{1}{2} \]

\[ P(Y = 1) = P(1 \text{ dice used and score is 1}) + P(2 \text{ dice used and score is 1}) \]
\[ = \frac{1}{2} \times \frac{1}{6} + \frac{1}{2} \times \frac{1}{36} = \frac{7}{32} \]

\[ P(Y = 2) = \frac{9}{72}, P(Y = 3) = \frac{11}{72}, P(Y = 4) = \frac{13}{72}, P(Y = 5) = \frac{15}{72}, P(Y = 6) = \frac{17}{72} \]

(b) \[ E(Y) = 1 \times \frac{7}{32} + 2 \times \frac{9}{72} + \ldots + 6 \times \frac{17}{72} = \frac{287}{72} = 3.99 \]
9. (a) First dice

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 1 & 2 & 3 & 4 & 5 & 6 \\
2 & 2 & 4 & 6 & 8 & 10 & 12 \\
3 & 3 & 6 & 9 & 12 & 15 & 18 \\
4 & 4 & 8 & 12 & 16 & 20 & 24 \\
5 & 5 & 10 & 15 & 20 & 25 & 30 \\
6 & 6 & 12 & 18 & 24 & 30 & 36 \\
\end{array}
\]

Second dice

\[
\begin{array}{cccccc}
x & 1 & 2 & 3 & 4 & 5 & 6 \\
P(x) & \frac{1}{36} & \frac{3}{36} & \frac{2}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \\
\end{array}
\]

(b) \(E(X) = \frac{1}{36}(1 \times 1 + 2 \times 2 + 3 \times 2 + \ldots + 36 \times 1) = \frac{441}{36} = 12.25\)
1. A psychologist produces a new measure she claims measures a person’s intelligence. Measurements on the scale, $Y$, have a mean of 200 and standard deviation 25. Assuming that the measurements can be modelled by a normal random variable, find

(a) $P(Y > 240)$, (3)
(b) $P(170 \leq Y \leq 195)$, (4)
(c) the value of $k$, to 1 decimal place, such that $P(Y \leq k) = 0.03$ (4)

2. Jam is sold in jars containing a stated weight of 250 g of jam. The jars are filled by a machine. The actual weight of jam in each jar is normally distributed with mean 257 g and standard deviation 5 g.

(a) (i) Find the probability of a jar containing less than the stated weight.  
(ii) In a box of 24 jars, find the expected number of jars containing less than the stated weight. (5)

The mean weight of jam is changed so that 1% of the jars contain less than the stated weight. The standard deviation stays the same.

(b) Find the new mean weight of jam. (4)

A sales representative for the machines which fill the jars claims that the latest model will have only 1% of the jars contain less than the stated weight when the mean weight of jam is 253 g.

(c) What is the standard deviation of the amount of jam in the jars filled by the latest model? (4)

3. The random variable $X \sim N(\mu, \sigma^2)$. It is known that $P(X \leq 25) = 0.0668$ and $P(X \leq 44.5) = 0.9599$

(a) (i) Show that the value of $\sigma$ is 6. 
(ii) Find the value of $\mu$. (8)

(b) Find $P(37 \leq X \leq 39)$. (3)

4. The length of time it takes a plumber to complete a job on a callout may be modelled by a Normal distribution with mean 11 minutes and standard deviation 150 seconds.

Find the time, $t$ minutes, such that 1 job in 7 will take the plumber longer than $t$ minutes to complete. (6)
5. The breaking strains of cables produced by a factory are normally distributed with mean $\mu$ Newtons and standard deviation $\sigma$ Newtons. Only 2.5% of all cables break at a strain of less than 6416 Newtons and 20% break at more than 7537 Newtons.

(a) Show that $\mu - 6416 = 1.96\sigma$. (2)
(b) Obtain a second equation in $\mu$ and $\sigma$. (3)
(c) Find the values of $\mu$ and $\sigma$ correct to 2 significant figures. (3)
(d) Find the values between which the middle 90% of breaking strains lie. (6)

6. A student scores 81% in a statistics practical examination, and 87% in a statistics theory examination.

The table shows the mean and standard deviations of the marks on these two examinations.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistics practical</td>
<td>73</td>
<td>6</td>
</tr>
<tr>
<td>statistics theory</td>
<td>81</td>
<td>9</td>
</tr>
</tbody>
</table>

Explain why the student’s performance in the practical is better than their performance in the theory. (4)

7. The random variable $X$ is normally distributed with mean 7830 and variance 1225.

Find

(a) $P(X < 7760)$. (3)
(b) $P(7800 < X < 7900)$. (3)

It is known that $P(a \leq X) = 0.18$. Find the value of $a$. (3)

8. The length of bolts produced in a factory can be modelled by a Normal distribution with mean 37.2 mm and standard deviation 0.8 mm. Any bolts which are less than 36 mm in length are scrapped.

(a) What proportion of bolts are scrapped? (3)
(b) What is the median length of the bolts which are not scrapped? (Give your answer in mm correct to 2 decimal places.) (4)
### The Normal distribution

#### Exam-style mark scheme

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Scheme</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (a)</td>
<td>$P(Y &gt; 240) = P\left(Z &gt; \frac{240 - 200}{25}\right)$&lt;br&gt;$= 1 - \Phi(1.6) = 1 - 0.9452 = 0.0548$</td>
<td>M1&lt;br&gt;B1 A1 (3)</td>
</tr>
<tr>
<td></td>
<td>$P(170 \leq Y \leq 195) = \Phi\left(\frac{195 - 200}{25}\right) - \Phi\left(\frac{170 - 200}{25}\right)$&lt;br&gt;$= \Phi(-0.2) - \Phi(-1.2)$&lt;br&gt;$= (1 - 0.5793) - (1 - 0.8849) = 0.3056$</td>
<td>M1 M1&lt;br&gt;B1 A1 (4)</td>
</tr>
<tr>
<td></td>
<td>$P(Y \leq k) = 0.03$&lt;br&gt;$P\left(Z &lt; \frac{k - 200}{25}\right) = 0.03$&lt;br&gt;$\Phi\left(\frac{200 - k}{25}\right) = 0.97$&lt;br&gt;$\frac{200 - k}{25} = 1.881 \quad k = 153.0$</td>
<td>M1&lt;br&gt;M1&lt;br&gt;B1 A1 (4)</td>
</tr>
<tr>
<td>2. (a)</td>
<td>(i) $X \sim N(257, 5^2)$&lt;br&gt;$P(X &lt; 250) = P\left(Z &lt; \frac{250 - 257}{5}\right) = \Phi(-1.4) = 0.0808$</td>
<td>M1 B1 A1 (3)</td>
</tr>
<tr>
<td></td>
<td>(ii) Expected number of jars $= 0.0808 \times 24 = 1.94 = 2$ to 1 s.f.</td>
<td>M1 A1 (2)</td>
</tr>
<tr>
<td></td>
<td>$Y \sim N(\mu, 5^2)$&lt;br&gt;$P(Y &lt; 250) = 0.01$&lt;br&gt;$P\left(Z &lt; \frac{250 - \mu}{5}\right) = 0.01$&lt;br&gt;$P\left(Z &gt; \frac{250 - \mu}{5}\right) = 0.01$&lt;br&gt;$2.3263 = \frac{\mu - 250}{5} \quad \mu = 261.6$</td>
<td>B1&lt;br&gt;M1&lt;br&gt;B1 A1 (4)</td>
</tr>
<tr>
<td></td>
<td>$M$ is the weight with new machine. $M \sim N(253, \sigma^2)$&lt;br&gt;$P(M &lt; 250) = 0.01$&lt;br&gt;$P\left(Z &lt; \frac{250 - 253}{\sigma}\right) = 0.01$&lt;br&gt;$P\left(Z &gt; \frac{3}{\sigma}\right) = 0.01$&lt;br&gt;$\frac{3}{\sigma} = 2.3263 \quad \sigma = 1.29$</td>
<td>B1&lt;br&gt;M1&lt;br&gt;B1 A1 (4)</td>
</tr>
<tr>
<td>3. (a)</td>
<td>(i) $P(X \leq 25) = 0.0668$&lt;br&gt;$P\left(Z &lt; \frac{25 - \mu}{\sigma}\right) = 0.0668 \Rightarrow P\left(Z &lt; \frac{\mu - 25}{\sigma}\right) = 0.9332$&lt;br&gt;$\frac{\mu - 25}{\sigma} = 1.5$</td>
<td>M1&lt;br&gt;B1 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P(X ≤ 44.5) = 0.9599</td>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>$P \left( Z &lt; \frac{44.5 - \mu}{\sigma} \right) = 0.9599$</td>
<td></td>
<td>B1 A1</td>
</tr>
<tr>
<td>$\frac{44.5 - \mu}{\sigma} = 1.75$</td>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>$\mu - 25 = 1.5\sigma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$44.5 - \mu = 1.75\sigma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma = 6$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) $\mu = 34$</td>
<td></td>
<td>A1</td>
</tr>
<tr>
<td>$P(37 \leq X \leq 39) = P \left( Z &lt; \frac{39 - 34}{6} \right) - P \left( Z &lt; \frac{37 - 34}{6} \right)$</td>
<td>M1</td>
<td></td>
</tr>
<tr>
<td>$0.79767 - 0.69146 = 0.1062$</td>
<td>B1 A1</td>
<td></td>
</tr>
<tr>
<td>(as interpolation is not required, answers which round to 0.11)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$X \sim N(11, 2.5^2)$</td>
<td>B1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(X &gt; t) = \frac{1}{7}$</td>
<td>M1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(X &lt; t) = \frac{6}{7}$</td>
<td>M1 B1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P \left( Z &lt; \frac{t - 11}{2.5} \right) = 0.8571$</td>
<td>M1 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1.07 = \frac{t - 11}{2.5}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t = 13.7$ minutes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. (a)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X &lt; 6416) = 0.025$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P \left( Z &lt; \frac{6416 - \mu}{\sigma} \right) = 0.025$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P \left( Z &gt; \frac{\mu - 6416}{\sigma} \right) = 0.025$</td>
<td>M1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{\mu - 6416}{\sigma} = 1.96 \Rightarrow \mu - 6416 = 1.96\sigma$</td>
<td>B1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(X &gt; 7537) = 0.2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P \left( Z &gt; \frac{7537 - \mu}{\sigma} \right) = 0.2$</td>
<td>M1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{7537 - \mu}{\sigma} = 0.8416 \Rightarrow 7537 - \mu = 0.8416\sigma$</td>
<td>B1 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) Solving both equations simultaneously gives:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma = 400$</td>
<td>M1 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu = 7200$</td>
<td>A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(X &lt; x_1) = 0.95$</td>
<td>B1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P(X &gt; x_1) = 0.05$</td>
<td>M1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P \left( Z &gt; \frac{x_1 - 7200}{400} \right) = 0.05$</td>
<td>M1 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{x_1 - 7200}{400} = 1.6449 \Rightarrow x_1 = 7858$</td>
<td>B1 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_1$ and $x_2$ are equidistant from 7200</td>
<td>M1 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\therefore x_2 = 7200 - (7858 - 7200) = 6542$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Standardised practical mark is \( \frac{81 - 73}{6} = 1.3 \)
Standardised theory mark is \( \frac{87 - 81}{9} = 0.6 \)
Therefore practical mark is better.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 6. | Standardised practical mark is \( \frac{81 - 73}{6} = 1.3 \)  
Standardised theory mark is \( \frac{87 - 81}{9} = 0.6 \)  
Therefore practical mark is better. | M1 A1 |

7. (a) \( P(X < 7760) = P\left(Z < \frac{7760 - 7830}{\sqrt{1225}}\right) = \Phi(-2) = 1 - 0.9772 = 0.0228 \)
(b) \( P(7800 < X < 7900) = P\left(Z < \frac{7900 - 7830}{\sqrt{1225}}\right) - P\left(Z < \frac{7800 - 7830}{\sqrt{1225}}\right) = \Phi(2) - \Phi(-0.8571) = 0.9772 - (1 - 0.8043) = 0.7815 \) (awrt 1.782)  
(c) \( P(a \leq X) = 0.18 \Rightarrow P(X \geq a) = 0.18 \Rightarrow P(X \leq a) = 0.82 \)
\[ a - \frac{7830}{35} = 0.915 \]
\[ a = 7860 \]

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 7. (a) | \( P(X < 7760) = P\left(Z < \frac{7760 - 7830}{\sqrt{1225}}\right) = \Phi(-2) = 1 - 0.9772 = 0.0228 \)  
(b) \( P(7800 < X < 7900) = P\left(Z < \frac{7900 - 7830}{\sqrt{1225}}\right) - P\left(Z < \frac{7800 - 7830}{\sqrt{1225}}\right) = \Phi(2) - \Phi(-0.8571) = 0.9772 - (1 - 0.8043) = 0.7815 \) (awrt 1.782)  
(c) \( P(a \leq X) = 0.18 \Rightarrow P(X \geq a) = 0.18 \Rightarrow P(X \leq a) = 0.82 \) | M1 B1 |

8. (a) \( P(X < 36) = P\left(Z < \frac{36 - 37.2}{0.8}\right) = \Phi(-1.5) = 1 - 0.9332 = 0.0668 = 6.7\% \)
(b) Require \( a \) such that \( P(X < a) = 0.0668 + 0.5 (1 - 0.0668) \)
\[ P(X < a) = 0.5334 \Rightarrow P\left(Z < \frac{a - 37.2}{0.8}\right) = 0.5334 \]
\[ \frac{a - 37.2}{0.8} = 0.0838 \Rightarrow a = 37.27 \]
awrt 37.3 accepted because interpolation not required

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 8. (a) | \( P(X < 36) = P\left(Z < \frac{36 - 37.2}{0.8}\right) = \Phi(-1.5) = 1 - 0.9332 = 0.0668 = 6.7\% \)  
(b) Require \( a \) such that \( P(X < a) = 0.0668 + 0.5 (1 - 0.0668) \) | M1 |

© Oxford University Press 2008