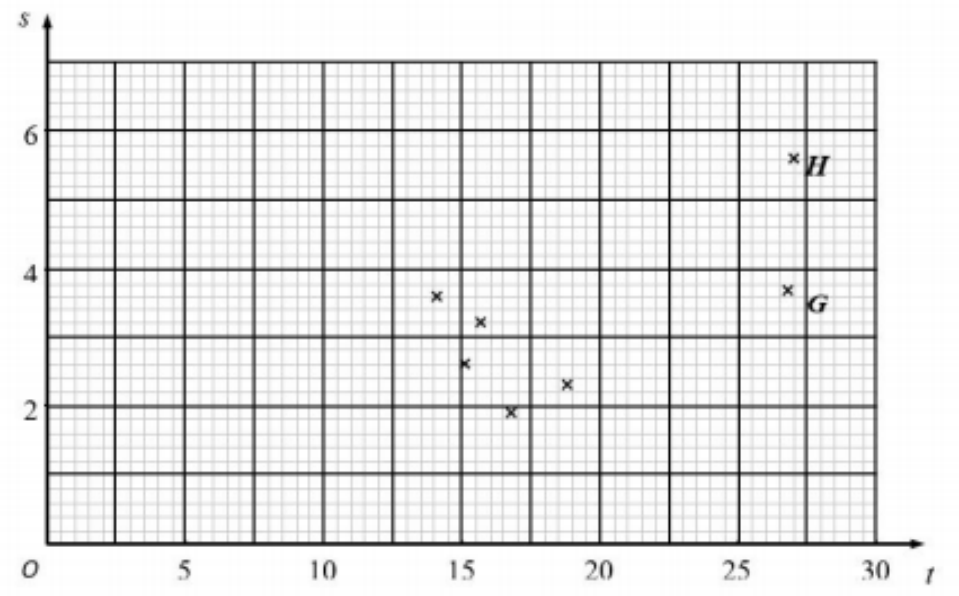
# AS Statistics Practice Paper C
# 45 Marks

45 Minutes

1. Using your knowledge of the large data set, answer the following questions.

a. In what unit is the daily mean pressure measured in? **(1)**

A researcher believes that there is a linear relationship between daily mean temperature and daily total rainfall. The 7 places in the northern hemisphere from the large data set are used. The mean of the daily mean temperatures, $t$ °C, and the mean of the daily total rainfall, $s$ mm, for the month of July in 2015 are shown on the scatter diagram below.




b. With reference to the scatter diagram, explain why a linear regression model may not be suitable for the relationship between $t$ and $s$. **(1)**

c. Using your knowledge of the large data set, suggest the names of the 2 places labelled $G$ and $H$. **(1)**

**(Total Marks: 3)**

2a. Explain what you understand by a population. **(1)**

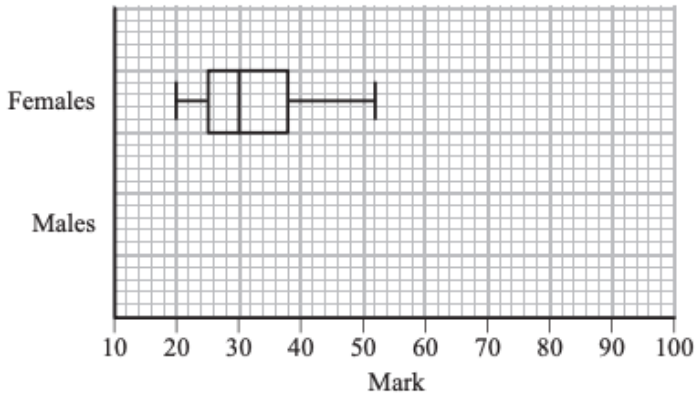
b. Explain what you understand by a statistic. **(1)**

A researcher took a sample of 100 voters from a certain town and asked them who they would vote for in an election. The proportion who said they would vote for Dr Smith was 35%.

c. State the population and the statistic in this case. **(2)**

**(Total Marks: 4)**

3. The marks of a group of female students in a statistics test are summarised in the figure.




a. Write down the mark which is exceeded by 75% of the female students. **(1)**

The marks of a group of male students in a statistics test are summarised by the stem and leaf diagram below.

| Mark | (2\|6 means 26) | Totals |
|---|---|---|
| 1 | 4 | (1) |
| 2 | 6 | (1) |
| 3 | 4 4 7 | (3) |
| 4 | 0 6 6 7 7 8 | (6) |
| 5 | 0 0 1 1 1 3 6 7 7 | (9) |
| 6 | 2 2 3 3 3 8 | (6) |
| 7 | 0 0 8 | (3) |
| 8 | 5 | (1) |
| 9 | 0 | (1) |

b. Find the median and interquartile range of the marks of the male students. **(3)**

An outlier is a mark that is either more than 1.5 × interquartile range above the upper quartile or more than 1.5 × interquartile range below the lower quartile.

c. List any outliers and draw a box plot of the data **(5)**

**(Total Marks: 9)**

4. The following table summarises the times, $t$ minutes to the nearest minute, recorded for a group of students to complete an exam.

| Time (minutes) $t$ | 11 – 20 | 21 – 25 | 26 – 30 | 31 – 35 | 36 – 45 | 46 – 60 |
|---|---|---|---|---|---|---|
| Number of students f | 62 | 88 | 16 | 13 | 11 | 10 |

[You may use $\Sigma ft^2 = 134281.25$]

a. Estimate the mean and standard deviation of these data. **(5)**

b. Use linear interpolation to estimate the value of the median. **(2)**

c. Use linear interpolation to estimate the value of the lower quartile. **(1)**

d. Estimate the interquartile range of this distribution. **(2)**

**(Total Marks: 10)**

5. The probability of Richard winning a prize in a game at the fair is 0.15

Richard plays a number of games.

a. Find the probability of Richard winning his second prize on his 8th game. **(2)**

The probability of Richard winning his second prize on his 8th game is calculated and is 0.0594

c. State two assumptions that have to be made, for this model to be valid. **(2)**

Mary plays the same game, but has a different probability of winning a prize. She plays until she has won $r$ prizes. The random variable $G$ represents the total number of games Mary plays.

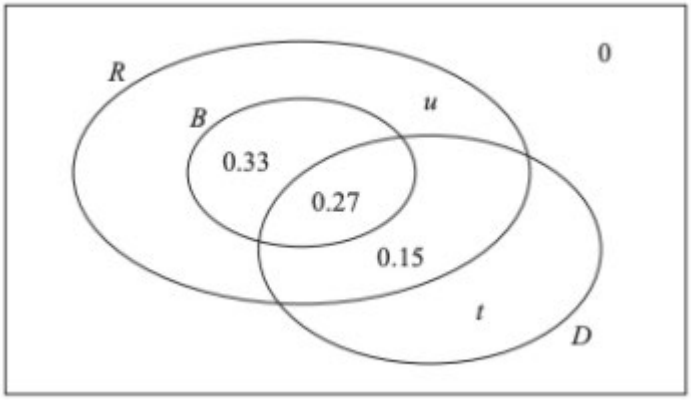
b. Given that the mean and standard deviation of $G$ are 18 and 6 respectively, determine whether Richard or Mary has the greater probability of winning a prize in a game.

**(4)**

**(Total Marks: 8)**

6. The Venn diagram shows the probabilities of customer bookings at Harry's hotel.

$R$ is the event that a customer books a room $B$ is the event that a customer books breakfast $D$ is the event that a customer books dinner $u$ and $t$ are probabilities.




a. Write down the probability that a customer books breakfast but does not book a room. **(1)**

b. Given that the events $B$ and $D$ are independent, find the value of $t$ **(4)**

c. Hence find the value of $u$ **(2)**

A coach load of 77 customers arrive at Harry's hotel.

Of these 77 customers

40 have booked a room and breakfast
37 have booked a room without breakfast

d. Estimate how many of these 77 customers will book dinner **(2)**

**(Total Marks: 9)**

7. A group of 100 adults recorded the amount of time, $t$ minutes, they spent exercising each day. Their results are summarised in the table below.

| Time ($t$ minutes) | Frequency (f) | Time midpoint ($x$) |
|---|---|---|
| $0 \leqslant t < 15$ | 25 | 7.5 |
| $15 \leqslant t < 30$ | 17 | 22.5 |
| $30 \leqslant t < 60$ | 28 | 45 |
| $60 \leqslant t < 120$ | 24 | 90 |
| $120 \leqslant t \leqslant 240$ | 6 | 180 |

[You may use $\sum fx^2 = 455\,512.5$]

A histogram is drawn to represent these data.

The bar representing the time $0 \leq t < 15$ has width 0.5 cm and height 6 cm.

Calculate the width and height of the bar representing a time of $60 \leq t < 120$. **(3)**

**(Total Marks: 3)**

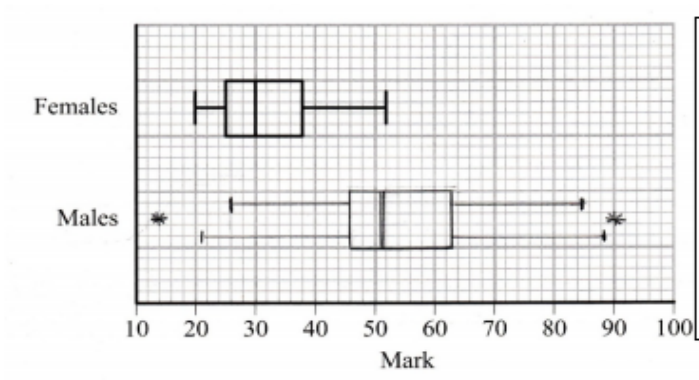
---

**Total Marks for Paper: 45**

# Mark Scheme

| | | |
|---|---|---|
| 1a | hPa | **B1** |
| 1b | The points do not lie close to a straight line | **B1** |
| 1c | Beijing or Jacksonville | **B1** |

| | | |
|---|---|---|
| 2a | A collection of all items | **B1** |
| 2b | A calculation based only on known data from a sample | **B1** |
| 2c | The voter in the town | **B1** |
| | Percentage/poportion voting for Dr. Smith | **B1** |

| | | |
|---|---|---|
| 3a | 25<br>Allow any value in the range of 24 to 26. | **B1** |
| 3b | $Q_2 = 51$ | **B1** |
| | IQR = 63 – 46 = 17 | **M1**<br>**A1** |
| 3c | Outliers given by 46 – 1.5 x 17 = 20.5<br>And, 63 + 1.5 x 17 = 88.5 | **M1** |
| | Outlier limits are 20.5 and 88.5 | **A1** |
| |  <br>Allow lower whisker to 20.5 and upper whisker to 88.5<br>Do **not** allow a mix of whiskers e.g 20.5 and 85<br>Do **not** allow both sets of whiskers<br>M1<br>A1ft<br>B1 | |

| | | |
|---|---|---|
| 4a | $\sum ft = 4837.5$ | **B1** |
| | Mean = $\frac{4837.5}{200} = 24.1875$ | **M1**<br>**A1** |
| | $\sigma = \sqrt{\frac{134281.25}{200} - \left(\frac{4837.5}{200}\right)^2}$ | **M1** |
| | $\sigma = 9.293 \ldots = 9.29$ | **A1** |
| 4b | $Q_2 = 20.5 + \frac{\left(\frac{100}{100.5-62}\right)}{88} \times 5$ | **M1** |
| | $Q_2 = 22.7$ | **A1** |
| 4c | $Q_1 = 10.5 + \frac{\frac{50}{50.25}}{62} \times 10$<br>$Q_1 = 18.6$ | **B1** |
| 4d | $Q_3 = 25.5$ | **B1** |
| | IQR = 6/9 | **B1** |

| | | |
|---|---|---|
| 5a | $\binom{7}{1} \times 0.15^2 \times (0.85)^6$ | **M1** |
| | = 0.0594 | **A1** |
| 5b | The model is only valid if,<br>the grame trials are independent | **B1** |
| | The probability of winning a prize, 0.15, is constant for each game | **B1** |
| 5c | $18 = \frac{r}{p}$ | **M1**<br>**A1** |

| | | |
|---|---|---|
| | $6^2 = \frac{r(1-p)}{p^2}$ | |
| | $2p = 1 - p$ | **M1** |
| | $p = \frac{1}{3} (> 0.15)$<br>So Mary has the greater chance of winning a prize | |

| | | |
|---|---|---|
| 6a | 0 | **B1** |
| 6b | $P(B) = 0.27 + 0.33 = 0.6$<br>$P(D) = 0.27 + 0.15 + t$<br>$P(B \; n \; D) = 0.27$ | **M1** |
| | 0.6 x (0.42 + t) = 0.27 | **M1** |
| | $0.42 + t = \frac{0.27}{0.6}$ | **A1** |
| | $t = 0.03$ | **A1** |
| 6c | $1 - (0.6 + 0.15 + t) = u$ | **M1** |
| | $u = 0.22$ | **A1** |
| 6d | $40 \times 0.45$<br>$37 \times \frac{15}{37}$ | **M1** |
| | = 33 | **A1** |

| | | |
|---|---|---|
| 7 | Width: 15 minutes is 0.5cm so 60 mins will be $4 \times 0.5 = 2$ | **B1** |
| | Height: Frequence of 25 represented by $6 \times 0.5 = 3$<br>So frequency of 24 = $\frac{24}{25} \times 3$ | **M1** |
| | So height = $\frac{1}{2} \times \frac{24}{25} \times 3$ = 1.44cm | **A1** |