Question 1

Worked Solution

Part (a): Mean and standard deviation

The 12 raw data values are: 160, 390, 169, 175, 125, 420, 171, 250, 210, 258, 186, 243.

Enter all 12 values into your calculator in statistics mode. Read off $\bar{x}$ and σ_x directly.

$\bar{x} = 229.75 \approx 230$ (awrt 230)
$\sigma_x = 87.3$ (awrt 87.3) [Accept $s_x = 91.2$ if using $n-1$]

Part (b): Median, lower quartile, upper quartile

Write the 12 values in ascending order:

$$125,\ 160,\ 169,\ 171,\ 175,\ 186,\ 210,\ 243,\ 250,\ 258,\ 390,\ 420$$

Median Q_2: With $n = 12$, take the average of the 6th and 7th values:

$$Q_2 = \frac{186 + 210}{2} = 198$$

Lower quartile Q_1: Median of the lower six values — average of 3rd and 4th:

$$Q_1 = \frac{169 + 171}{2} = 170$$

Upper quartile Q_3: Median of the upper six values — average of 9th and 10th:

$$Q_3 = \frac{250 + 258}{2} = 254$$

$Q_1 = 170$, $Q_2 = 198$, $Q_3 = 254$

Part (c): Identify outliers using $Q_3 + 1.5(Q_3 - Q_1)$

$$Q_3 + 1.5(Q_3 - Q_1) = 254 + 1.5(254 - 170) = 254 + 1.5 \times 84 = 254 + 126 = 380$$

Any value greater than 380 is a potential outlier. From the data, 390 (patient B) and 420 (patient F) both exceed 380.

Patients F (cotinine 420) and B (cotinine 390) may have been smoking more than a packet of cigarettes a day.

Question 2

Worked Solution

Part (a): Mean and standard deviation for the 20 timed people

The 20 raw data values are:

$$17,\ 19,\ 22,\ 26,\ 28,\ 31,\ 34,\ 36,\ 38,\ 39,\ 41,\ 42,\ 43,\ 47,\ 50,\ 51,\ 53,\ 55,\ 57,\ 58$$

Enter all 20 values into your calculator in statistics mode. Read off $\bar{x}$ and σ_x directly.

Mean $\bar{x} = 39.35$ (awrt 39.3–39.4)
Standard deviation $\sigma_x = 12.36$ (awrt 12.3–12.7) [Accept $s_x = 12.68$]

Part (b): Median and IQR for all 23 people

The 3 extra people who did not finish within 60 s are recorded only as > 60. So the full ordered dataset for all 23 people has the original 20 values followed by three values each exceeding 60.

Median ($n = 23$): Position $\frac{23+1}{2} = 12$. Counting through the ordered list, the 12th value is **42**.

Lower quartile Q_1: Position $\frac{23+1}{4} = 6$. The 6th value is **31**.

Upper quartile Q_3: Position $\frac{3\times 24}{4} = 18$. The 18th value is **55**.

$$\text{IQR} = Q_3 - Q_1 = 55 - 31 = 24$$

Median $= 42$, IQR $= 24$ (allow IQR in range 21–27)

Part (c): Why mode and range are not appropriate for all 23 people

(i) Mode: All 23 values are distinct — there are no repeated values, so the mode does not exist.

(ii) Range: The three extra values are recorded only as > 60; the true maximum is unknown, so the range cannot be determined.

(i) The mode does not exist — no value is repeated (all values distinct / sparse data).
(ii) The maximum value is unknown (only > 60 is recorded), so the range cannot be calculated.

Question 3

Worked Solution

Part (a): Modal mark

From the stem and leaf diagram, the value 60 appears most frequently (four times in the row 6 | 0 0 0 0 1 3 4 4 4).

Modal mark = 60

Part (b): Lower quartile, median, upper quartile ($n = 45$)

Read values off the stem and leaf in order and build cumulative counts:

Row	Freq	Cumulative
3 | 6 9 9	3	3
4 | 0 1 2 2 3 4	6	9
4 | 5 6 6 6 8	5	14
5 | 0 2 3 3 4 4	6	20
5 | 5 5 6 7 7 9	6	26
6 | 0 0 0 0 1 3 4 4 4	9	35
6 | 5 5 6 7 8 9	6	41
7 | 1 2 3 3	4	45

Median Q_2: 23rd value. Rows 1–4 total 20 values; the 21st–26th values (row 5 | 5 5 6 7 7 9) are 55, 55, 56, 57, 57, 59. The 23rd value is **56**.

Q_1**:** Average of 11th and 12th values. Rows 1–3 total 14 values; the 10th–14th values (row 4 | 5 6 6 6 8) are 45, 46, 46, 46, 48. 11th = 46, 12th = 46.

$$Q_1 = 46$$

Q_3**:** Average of 34th and 35th values. Rows 1–5 total 26 values; the 27th–35th values (row 6 | 0 0 0 0 1 3 4 4 4) are 60, 60, 60, 60, 61, 63, 64, 64, 64. 34th = 64, 35th = 64.

$$Q_3 = 64$$

$Q_1 = 46, \quad Q_2 = 56, \quad Q_3 = 64$

Part (c): Mean and standard deviation from summary statistics

Summary statistics $\sum x = 2497$ and $\sum x^2 = 143{,}369$ are given, so use the formula directly.

Mean:

$$\bar{x} = \frac{\sum x}{n} = \frac{2497}{45} = 55.49\ldots$$

Standard deviation:

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{143{,}369}{45} - \left(\frac{2497}{45}\right)^2} = \sqrt{3185.98 - 3079.08} = \sqrt{106.90} = 10.34\ldots$$

$\bar{x} = 55.5$ (awrt), $\sigma = 10.3$ (awrt; accept $s = 10.5$ using $n - 1$)

Question 4

Worked Solution

Frequency distribution for time to first cigarette (100 heavy smokers). Classes and midpoints: $[1, 4]$ $m = 2.5$; $[5, 14]$ $m = 9.5$; $[15, 29]$ $m = 22$; $[30, 59]$ $m = 44.5$; $[60, 240)$ $m = 150$. Frequencies: 31, 27, 19, 14, 9.

The data is grouped, so individual values are not available. Enter (m, f) pairs into your calculator in statistics mode (using midpoints as x-values and frequencies as f). The calculator gives the mean and standard deviation directly.

Part (i): Mean and standard deviation

Mean $\approx$ 27.3 minutes (awrt 27.2) Standard deviation $\approx$ 41.0 minutes (awrt 40.5–41.1)

Check: $\bar{x} = \sum fm / \sum f = 2725/100 = 27.25$; $\sigma = \sqrt{2420.5 - 27.25^2} = \sqrt{1677.94} = 40.96$

Part (ii): Estimate of IQR by linear interpolation

Cumulative frequencies: 31 (≤ 4.5), 58 (≤ 14.5), 77 (≤ 29.5), 91 (≤ 59.5), 100.

Q_1**:** 25th value lies in the first class $[0.5, 4.5)$ with 31 values (starts from cumulative 0):

$$Q_1 = 0.5 + \frac{25 - 0}{31} \times 4 = 0.5 + 3.23 = 3.73 \text{ min}$$

Q_3**:** 75th value lies in the third class $[14.5, 29.5)$ with 19 values (starts from cumulative 58):

$$Q_3 = 14.5 + \frac{75 - 58}{19} \times 15 = 14.5 + 13.42 = 27.9 \text{ min}$$

$$\text{IQR} = 27.9 - 3.73 \approx 24 \text{ min}$$

IQR $\approx$ 24 minutes (awrt 23–25)

Part (iii): Effect of changing 'at least 60' to 'at least 60 but less than 480'

(a) Mean: The midpoint of the final class increases from 150 to 270, so the mean **increases**.

(b) Standard deviation: The final class midpoint moves further from the mean, increasing the spread. The standard deviation **increases**.

(c) IQR: Both quartiles lie in the lower classes, unaffected by the upper boundary. The IQR is **unchanged**.

(a) Mean **increases**. (b) Standard deviation **increases**. (c) IQR **unchanged**.

Question 5

Worked Solution

Frequency distribution for goals per match in 2004/05, $n = 380$. Values $x = 0, 1, \dots, 9$ with frequencies 30, 79, 99, 68, 60, 24, 11, 6, 2, 1.

Part (a)(i): Median and IQR

Build cumulative frequencies:

Goals x	Freq f	Cumulative F
0	30	30
1	79	109
2	99	208
3	68	276
4	60	336

Median: 190th and 191st values — both in $x = 2$ row (cumulative 109 to 208). **Median** $= 2$.

Q_1**:** 95th and 96th values — both in $x = 1$ row (cumulative 30 to 109). $Q_1 = 1$.

Q_3**:** 285th and 286th values — both in $x = 4$ row (cumulative 277 to 336). $Q_3 = 4$.

$$\text{IQR} = Q_3 - Q_1 = 4 - 1 = 3$$

Median $= 2$, IQR $= 3$

Part (a)(ii): Mean and standard deviation

Enter the (x, f) pairs into your calculator in statistics mode. Read off $\bar{x}$ and σ_x directly.

Mean $\bar{x} = 2.56$ (awrt 2.56), Standard deviation $\sigma_x = 1.66$ (awrt 1.66–1.67)

Part (b)(i): Compare 2004/05 and 2005/06

For 2005/06: median $= 2$, IQR $= 2$, mean $= 2.48$, standard deviation $= 1.59$.

Average: The median is the same (2) in both seasons; the mean is slightly higher in 2004/05 (2.56 vs 2.48), so the average goals per match was similar or marginally higher in 2004/05.

Spread: Both IQR (3 vs 2) and standard deviation (1.66 vs 1.59) are larger in 2004/05, indicating greater variation in goals per match that season.

Average: similar in both seasons (same median of 2; mean slightly greater in 2004/05).
Spread: greater in 2004/05 (larger IQR and larger standard deviation).

Question 6

Worked Solution

Frequency distribution for number of matches per box, $n = 85$. Most entries are exact values; two open-ended classes have extreme values 227 and 271.

Cumulative frequencies: ≤ 227: 1; ≤ 243: 2; ≤ 246: 4; ≤ 247: 7; ≤ 248: 11; ≤ 249: 17; ≤ 250: 27; ≤ 251: 40; ≤ 252: 56; ≤ 253: 76; ≤ 254: 81; ≤ 259: 84; ≤ 271: 85.

Part (a)(i): Modal value

Mode = 253 (frequency 20, the highest)

Part (a)(ii): Median and IQR

Median: $\frac{85+1}{2} = 43$rd value. Cumulative through 251 is 40, through 252 is 56. The 43rd value = **252**.

Q_1: $\frac{86}{4} = 21.5$th — average of 21st and 22nd values. Cumulative through 249 is 17, through 250 is 27. Both are 250. $Q_1 = 250$.

Q_3: $\frac{3 \times 86}{4} = 64.5$th — average of 64th and 65th values. Cumulative through 252 is 56, through 253 is 76. Both are 253. $Q_3 = 253$.

$$\text{IQR} = 253 - 250 = 3$$

Median = 252, $Q_1 = 250$, $Q_3 = 253$, IQR = 3

Part (b)(i): Range

$$\text{Range} = 271 - 227 = 44$$

Range = 44

Part (b)(ii): Mean and standard deviation

Enter all (x, f) pairs into your calculator in statistics mode, using the actual extreme values 227 and 271 (and midpoints 241, 257 for the classes 239–243 and 255–259). Read off $\bar{x}$ and σ_x directly.

Mean ≈ 251.2 (awrt 251–251.4), Standard deviation ≈ 4.22 (awrt 4.21–4.28)

Part (c): Most appropriate measure of spread

The **interquartile range (IQR)** is most appropriate. The data contains extreme/outlying values (227 and 271) which would inflate both the range and the standard deviation; the IQR is unaffected by these outliers.

End of Worked Solutions