Question 1

(Q06 6683/01, June 2013)

Worked Solution

Data: times t (minutes, nearest minute). $n = 62 + 88 + 16 + 13 + 11 + 10 = 200$.

(a) Mean and standard deviation

Midpoints: 15.5, 23, 28, 33, 40.5, 53.

$$\sum ft = 62(15.5) + 88(23) + 16(28) + 13(33) + 11(40.5) + 10(53)$$
$$= 961 + 2024 + 448 + 429 + 445.5 + 530 = 4837.5$$
$$\bar{t} = \frac{4837.5}{200} = 24.1875 \approx \mathbf{24.2}$$

$$\sigma = \sqrt{\frac{\sum ft^2}{n} - \bar{t}^2} = \sqrt{\frac{134281.25}{200} - 24.1875^2} = \sqrt{671.40625 - 585.03...} = \sqrt{86.37...} \approx 9.29$$

Mean $\approx$ 24.2 minutes; standard deviation $\approx$ 9.29 minutes.

(b) Median by linear interpolation

$n/2 = 100$. Cumulative frequencies: ≤ 20: 62; ≤ 25: 150. Median lies in the 21–25 class.

$$Q_2 = 20.5 + \frac{100 - 62}{88} \times 5 = 20.5 + \frac{38}{88} \times 5 = 20.5 + 2.159... \approx \mathbf{22.7}$$

Median $\approx$ 22.7 minutes.

(c) Lower quartile

$n/4 = 50$. Cumulative frequency ≤ 20: 62, so Q_1 lies in the first class (11–20).

$$Q_1 = 10.5 + \frac{50 - 0}{62} \times 10 = 10.5 + 8.065... = 18.565... \approx 18.6 \quad \checkmark$$

(d) Interquartile range

$3n/4 = 150$. Q_3 lies in the 21–25 class (cumulative up to 20 is 62; up to 25 is 150, exactly 150).

$$Q_3 = 20.5 + \frac{150 - 62}{88} \times 5 = 20.5 + 5 = 25.5$$
$$\text{IQR} = Q_3 - Q_1 = 25.5 - 18.6 = 6.9$$

IQR = 6.9 minutes.

(e) Why mean and standard deviation are not most appropriate

The data are skewed (positively skewed), so the mean and standard deviation are not the most appropriate summary statistics; the median and IQR would be more suitable.

(f) Effect of subtracting 5 minutes from all times

The mean would decrease by 5 (to ≈ 19.2) and the standard deviation would remain the same (≈ 9.29), since standard deviation measures spread which is unchanged by a constant shift. The median and lower quartile would each decrease by 5. The IQR would remain unchanged, since IQR depends only on spread.

Question 2

(Q02 8MA0/21, Oct 2021)

Worked Solution

Ages of airline passengers. Classes (years): $[0, 5)$: 5; $[5, 20)$: 45; $[20, 40)$: 90; $[40, 65)$: ?; $[65, 80)$: ?; $[80, 90)$: 1.

(a) Complete the histogram

From the $[5, 20)$ bar: frequency density $= 45/15 = 3$ (fd units per year). Use the existing bars to establish the scale:

The $[0, 5)$ bar has fd $= 5/5 = 1$ and the $[20, 40)$ bar has fd $= 90/20 = 4.5$.

From the histogram, the $[40, 65)$ bar has fd $= 130/25 = 5.2$ (reading 130 passengers from the scale), and the $[65, 80)$ bar has fd $= 60/15 = 4$ (60 passengers). These values are read from the partial histogram and confirm total $= 5+45+90+130+60+1 = 331$.

The missing frequencies are: $[40, 65)$: 130 passengers (fd $= 5.2$); $[65, 80)$: 60 passengers (fd $= 4.0$). The $[0, 5)$ bar has fd $= 1$ and the $[20, 40)$ bar has fd $= 4.5$.

(b) Median age by linear interpolation

Total $n = 331$, so $n/2 = 165.5$.

Cumulative frequencies: < 5: 5; < 20: 50; < 40: 140; < 65: 270.

Median lies in the $[40, 65)$ class.

$$\text{Median} = 40 + \frac{165.5 - 140}{130} \times 25 = 40 + \frac{25.5}{130} \times 25 = 40 + 4.904... \approx 44.9$$

Median age ≈ 44.9 years.

(c) Is the oldest passenger an outlier?

$Q_1 = 27.3$, $Q_3 = 58.9$, IQR $= 58.9 - 27.3 = 31.6$.

Upper outlier limit $= Q_3 + 1.5 \times \text{IQR} = 58.9 + 1.5 \times 31.6 = 58.9 + 47.4 = 106.3$.

The oldest passenger is in the $[80, 90)$ class, so their age $< 90 < 106.3$.

The upper outlier limit is 106.3. Since the oldest passenger is under 90, which is less than 106.3, the oldest passenger is **not** an outlier.

Question 3

(Q04 8MA0/21, June 2019)

Worked Solution

Daily total rainfall in Hurn, May–Oct 2015. $n = 121 + 10 + 24 + 12 + 17 = 184$.

(a) Why data needs cleaning

The large data set records trace rainfall as "tr" (a non-numeric entry). These must be converted to a number (e.g. 0) before the mean can be calculated.

(b) Upper quartile by linear interpolation

$3n/4 = 138$. Cumulative frequencies: < 0.5: 121; < 1.0: 131; < 5.0: 155. Q_3 lies in $[1.0, 5.0)$.

$$Q_3 = 1 + \frac{138 - 131}{24} \times 4 = 1 + \frac{7}{24} \times 4 = 1 + \frac{7}{6} = 1 + 1.1\overline{6} \approx 2.17$$

Upper quartile ≈ 2.17 mm.

(c) Standard deviation

$$\sigma = \sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2} = \sqrt{\frac{7704.1875}{184} - \left(\frac{539.75}{184}\right)^2}$$

$$= \sqrt{41.87... - (2.933...)^2} = \sqrt{41.87... - 8.60...} = \sqrt{33.27...} \approx 5.77$$

Standard deviation ≈ 5.77 mm.

(d)(i) Assumption when using class midpoints

It is assumed that all values within each class are uniformly distributed (i.e. concentrated at the class midpoint).

(d)(ii) Why this assumption does not hold

In the large data set, the majority of values in the first class $[0, 0.5)$ are 0 (trace rainfall recorded as 0 after cleaning), so values are not uniformly spread across that class.

(d)(iii) Effect on the actual mean

The actual mean is likely to be **smaller** than the estimate, because the majority of data in the first class ($[0, 0.5)$) are at or close to 0, which is below the midpoint of 0.25 used in the estimate.

Question 4

(Q06 6683/01, June 2015)

Worked Solution

60 students drawing a 20° angle (box plot given) and a 70° angle (frequency table given).

(a) Range for 20° data

From the box plot: minimum = 9, maximum = 48.

Range $= 48 - 9 = 39$.

(b) Interquartile range for 20° data

From the box plot: $Q_1 = 12$, $Q_3 = 25$.

IQR $= 25 - 12 = 13$.

(c) Median of 70° data by linear interpolation

$n = 60$, $n/2 = 30$. Table: $[55, 60)$: 6; $[60, 65)$: 15; cumulative to 65: 21. $[65, 70)$: 13; cumulative to 70: 34.

Median lies in $[65, 70)$:

$$Q_2 = 65 + \frac{30 - 21}{13} \times 5 = 65 + \frac{9}{13} \times 5 = 65 + 3.461... \approx \mathbf{68.5}^\circ$$

Median $\approx 68.5°$.

(d) Show lower quartile is 63°

$n/4 = 15$. Cumulative to 60: 6; cumulative to 65: 21. Q_1 lies in $[60, 65)$:

$$Q_1 = 60 + \frac{15 - 6}{15} \times 5 = 60 + \frac{9}{15} \times 5 = 60 + 3 = \mathbf{63}^\circ \quad \checkmark$$

(e)(i) Show there are no outliers

IQR $= 75 - 63 = 12$.

Lower fence: $63 - 1.5 \times 12 = 63 - 18 = 45$. Minimum $= 55 > 45$. No lower outliers.

Upper fence: $75 + 1.5 \times 12 = 75 + 18 = 93$. Maximum $= 84 < 93$. No upper outliers.

Lower fence $= 45$, upper fence $= 93$. Since $55 > 45$ and $84 < 93$, there are no outliers. ✓

(e)(ii) Box plot

Box from 63 to 75, median line at 68.5, left whisker to 55, right whisker to 84 (no outliers), on scale 40–90.

(f) Which angle were students more accurate at drawing?

Students were more accurate at drawing the 70° angle. The median (68.5°) is closer to the target of 70° than the 20° angle median is to 20°. The IQR (12) is similar to the 20° IQR (13), indicating comparable spread, but the 70° data is more centred on the target value.

End of Worked Solutions