6.1 Sampling, averages and spread

Name: _____ Class: _____ Date: _____ **Total: 25 marks**

KEY WORDS modal class • interquartile range

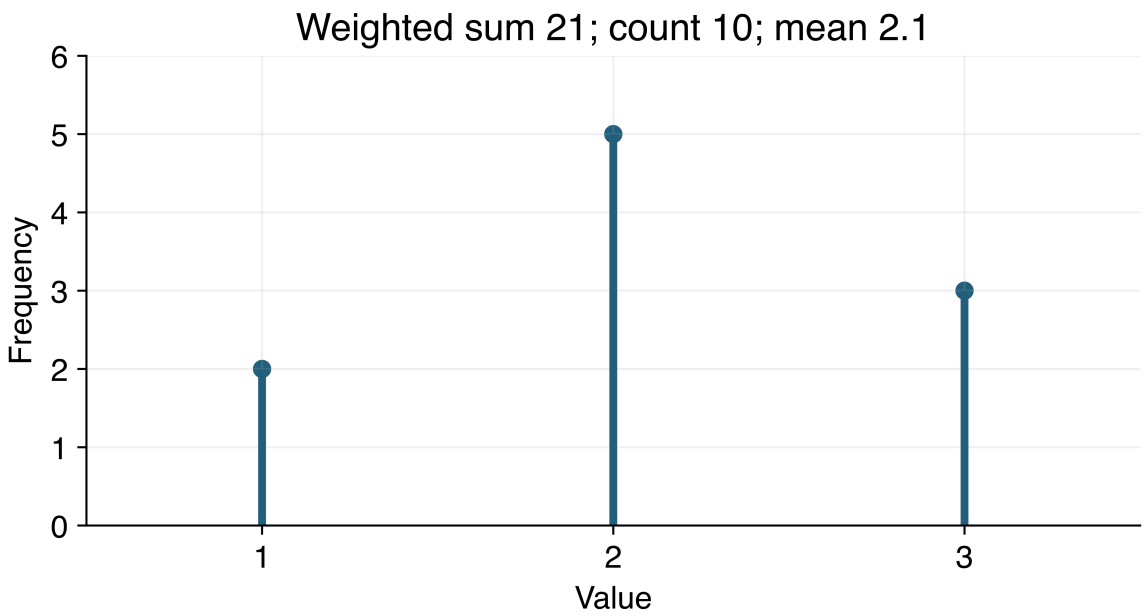
Sampling, averages and spread

Objective

Infer population properties from samples with their limitations, and interpret, analyse and compare distributions through means, medians, modes, modal classes, range, quartiles, inter-quartile range, outliers and box plots (rows S1, S4–S5).

1 Key knowledge and worked examples

Centre and spread. Mean $= \frac{\sum fx}{\sum f}$ (midpoints for grouped data – an estimate); median by cumulative count; mode or modal class for the most common value. Range $= \max - \min$; inter-quartile range $= Q_3 - Q_1$ describes the middle half. A random sample estimates a population proportion; voluntary or convenience samples carry selection bias that size alone cannot fix.



Original native-lesson illustration; labels belong to its worked example.

2 Practice

2.1 Find the mean of $2, 4, 4, 6, 9$. [1]

2.2 Values 1, 2 and 3 have frequencies 2, 5 and 3. Find the mean and the mode. [2]

2.3 In a random sample of 80 students, 24 walk to school. Estimate the number of walkers in a school of 600. [2]

2.4 For the data $2, 4, 4, 6, 24$, find the mean and the median, and explain why they differ. [2]

2.5 Classify each as discrete or continuous: number of siblings; travel time to school. [2]

2.6 Classes $0 \leq x < 10$, $10 \leq x < 20$, $20 \leq x < 30$ have frequencies 2, 5, 3. Estimate the mean and write down the modal class. [3]

3 Exam-style questions

3.1 *(June 2024 2H Q13, adapted.)* A table of lengths shows 15 values of 7.2, 18 values of 7.6 and 7 values of 8. Work out the mean. [3]

3.2 *(June 2024 2H Q16, adapted.)* Draw a box plot for: shortest value 2.4 cm, lower quartile 4 cm, median 6 cm, upper quartile 7 cm, longest value 9 cm. [3]

3.3 *(June 2024 1H Q16, adapted.)* The swimming club's median age is lower than the cycling club's, and its inter-quartile range is higher. Write two sentences comparing the clubs' ages. [2]

3.4 *(Original, Higher.)* A cumulative-frequency graph of 80 observations reads $Q_1 = 12$ and $Q_3 = 21$. State where these readings were taken on the cumulative axis, and work out the inter-quartile range. [3]

3.5 *(Original.)* An optional online poll finds that 70% of 400 respondents cycle to school. Explain why this does not establish that 70% of the school cycles. [2]

Solutions

2.1 $(2+4+4+6+9) \div 5 = 25 \div 5 = 5$.

2.2 $\sum fx = 2+10+9 = 21$ over 10 values: mean $= 2.1$; mode $= 2$.

2.3 $\frac{24}{80} = 0.3$; $600 \times 0.3 = 180$ walkers.

2.4 Mean $= 40 \div 5 = 8$; median $= 4$. The single large value 24 pulls the mean up while the median resists it.

2.5 Siblings: discrete (counted); travel time: continuous (measured, even if recorded to whole minutes).

2.6 Midpoints 5, 15, 25: $\sum fm = 10+75+75 = 160$, mean $\approx 160 \div 10 = 16$; modal class $10 \leq x < 20$.

3.1 $\sum fx = 15\times7.2+18\times7.6+7\times8 = 108+136.8+56 = 300.8$; mean $= 300.8 \div 40 = 7.52$.

3.2 A rectangle from 4 to 7 with a median line at 6, whiskers to 2.4 and 9 – five plotted points.

3.3 The swimming club's average age is lower (its median is lower); its ages are less consistent (its inter-quartile range is higher).

3.4 Quartiles read at cumulative frequency $80 \div 4 = 20$ and $3 \times 80 \div 4 = 60$; $IQR = 21 - 12 = 9$.

3.5 The poll is voluntary, so keen cyclists respond more – selection bias that more responses cannot remove; the sample is not random.

Local review guidance: exact $\sum fx$ products before dividing (approximated products forfeit the mark); comparisons state an average AND a spread in context; sampling limits named as selection bias or non-response, not "not enough people". This is authored classroom guidance.