Data, sampling and graphs

Skills sheet · SL AA 4.1 · 4.2

KEY FACTS

Discrete: counted; continuous: measured D1

Stratified: $\frac{\text{stratum}}{\text{population}} \times \text{sample size}$ P1

Median at $\frac{n}{2}$, Q_1 at $\frac{n}{4}$, Q_3 at $\frac{3n}{4}$ D2

Outlier: beyond $Q_1 - 1.5\,\text{IQR}$ or $Q_3 + 1.5\,\text{IQR}$ P2

Convenience: easy, often biased W1.

Fences use the IQR, not the range W2.

Quartiles: read on the vertical axis W2.

WHAT UNLOCKS

- "stratified" → each group in proportion to its size.
- "name the technique" → how the members were chosen.
- a cumulative frequency graph → across from $\frac{n}{2}$, $\frac{n}{4}$, $\frac{3n}{4}$, then down.
- "outlier" → the fence, then the comparison in words.

WORKED EXAMPLE ①

A club has 300 juniors, 500 adults and 400 seniors. Find the numbers in a stratified sample of 60. no GDC

$\frac{300}{1200} \times 60 = 15$ (M1)A1

25 adults, 20 seniors A1

Method

A stratified sample

① The total of the population.
② Each share times the sample size.
③ Check the sizes add up.

WORKED EXAMPLE ②

$Q_1 = 10$, $Q_3 = 18$. Is 32 an outlier? no GDC

$\text{IQR} = 8$, $18 + 1.5 \times 8 = 30$ M1

$32 > 30$: an outlier A1

Method

Testing an outlier

① $\text{IQR} = Q_3 - Q_1$.
② The fence $Q_3 + 1.5\,\text{IQR}$ or $Q_1 - 1.5\,\text{IQR}$.
③ Compare, conclude in words.

The fence, not the maximum.

FLUENCY

●○○ **1.** The mass of a parcel: discrete or continuous? no GDC

●○○ **2.** The number of cars in a car park: discrete or continuous? no GDC

●○○ **3.** 360 of 900 students are in Year 1. How many in a stratified sample of 50? no GDC

●○○ **4.** Every 10th name on a list, from a random start: the technique? no GDC

●○○ **5.** $Q_1 = 14$, $Q_3 = 26$: find the IQR no GDC

●○○ **6.** $Q_1 = 14$, $Q_3 = 26$: find the upper fence no GDC

REASONING

●●● **Question 1.** The times, t minutes, of 100 calls: cumulative frequencies 12, 40, 78, 100 for $t \leq 5$, 10, 15, 20. [no GDC]

(a) Find the number of calls with $10 < t \leq 15$. [1]

(b) Write down the class containing the median. [1]

●●● **Question 2.** A set of data has minimum 2, $Q_1 = 9$, median 12, $Q_3 = 15$, maximum 27. [no GDC]

(a) Show that 27 is an outlier. [2]

(b) Determine whether 2 is an outlier. [2]

●●● **Question 3.** A researcher interviews the first 20 men and the first 20 women she meets in a street. [no GDC]

(a) Name the sampling technique and give one disadvantage. [2]

PROBLEMS

●●● **Question 4.** 120 values: cumulative frequencies 15, 50, 96, 120 for $x \leq 10$, 20, 30, 40, the points joined by straight segments. [no GDC]

(a) Estimate the median. [2]

●●● **Question 5.** The data 4, 6, 7, 9, 10, 12, k are in order, with $k > 12$. [no GDC]

(a) Find the least whole number k for which k is an outlier. [3]

ANSWERS

1 continuous

2 discrete

3 $\frac{360}{900} \times 50 = 20$

4 systematic

5 12

6 $26 + 18 = 44$

Q1 (a) $78 - 40 = 38$ (b) $10 < t \leq 15$ (the 50th and 51st: $40 < 50 \leq 78$)

Q2 (a) $15 + 1.5 \times 6 = 24$, $27 > 24$ (b) $9 - 9 = 0$, $2 > 0$: not an outlier

Q3 quota sampling; not random, so it may be biased (only people in that street)

Q4 60th value: $20 + 10 \times \frac{60-50}{96-50} = 22.2$

Q5 $Q_1 = 6$, $Q_3 = 12$, fence $12 + 9 = 21$; $k > 21$, so $k = 22$

Written and verified by a practising IB mathematics teacher, produced with AI assistance; every result recomputed by software, every figure computed from its function. Not affiliated with or endorsed by the International Baccalaureate Organization. © 2026 Hence