

Statistics

AQA · GCSE · Mathematics

AQA GCSE Mathematics 8300 topic 3.6 (specification Version 1.0, 12 September 2014, first exams 2017). Rows S1–S6 run in three tiers; the additional foundation rows are inferring population properties from a sample while knowing the limitations of sampling (S1) and tables and line graphs for time series data (S2); the Higher-only row is histograms with equal and UNEQUAL class intervals and cumulative frequency graphs (S3), with box plots, quartiles and the inter-quartile range as the Higher addition to S4. The June 2024 Higher set supplies verified anchors below.

Must-know for topic 3.6:

Family	Formulae
centre	mean = $\frac{\sum fx}{\sum f}$ (midpoints for grouped data); median by cumulative count; mode / modal class
spread	range = max – min; interquartile range = $Q_3 - Q_1$
displays	pie angle = $\frac{f}{n} \times 360^\circ$; frequency density = $\frac{\text{frequency}}{\text{class width}}$
bivariate	correlation direction + strength; estimated line of best fit; interpolate, extrapolate with care

Sampling and populations

Infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling (S1, additional foundation): a **sample** is a subset; a census covers the whole **population**; random selection reduces selection bias but never removes sampling variability or non-response. Of a random sample of 80 students, 24 walk: proportion $\frac{24}{80} = 0.3$, applied to a school of 600 suggests about 180 walkers – an estimate, not a count. A convenience sample from a sports club overrepresents active students; a voluntary online poll misses non-respondents however many responses arrive. Apply statistics to describe a population (S5): state what the sample supports and what might make extrapolating unreliable. Know the terms **primary data**, secondary data, discrete data and continuous data (number of siblings is discrete; travel time is continuous even when recorded to whole minutes).

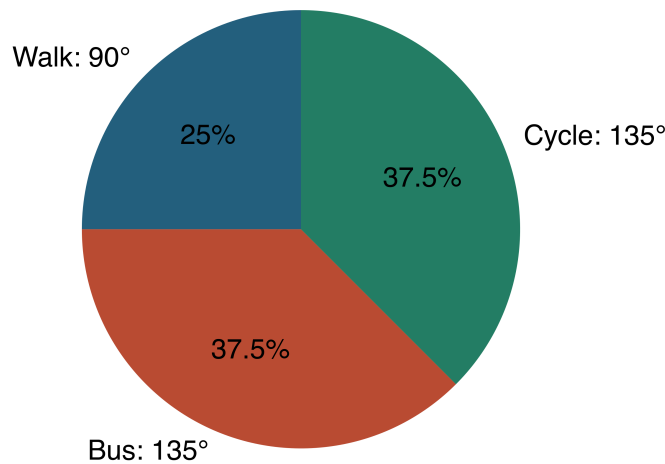
Tables, charts and time series

Interpret and construct tables, charts and diagrams (S2): frequency tables, bar charts, pie charts and pictograms for categorical data, vertical line charts for ungrouped discrete numerical data – and, at additional foundation, tables and line graphs for **time series** data – knowing their appropriate use, including choosing suitable statistical diagrams. A pie-chart sector angle is $\frac{f}{n} \times 360^\circ$: of 40 students with 10 walking, 15 bussing and 15 cycling, the angles are 90° , 135° , 135° . Time series plot in

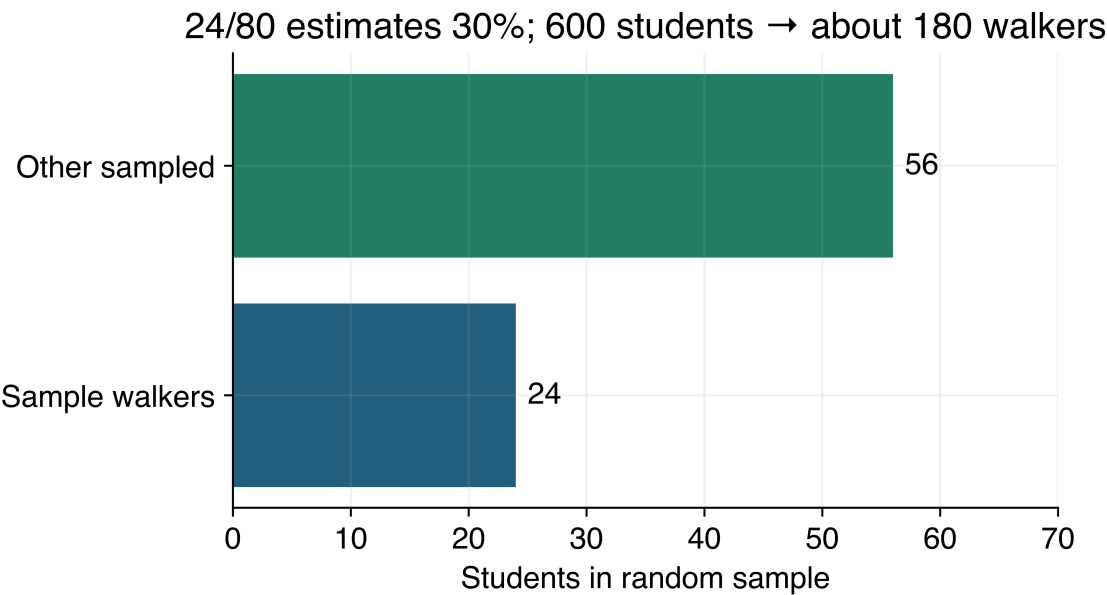
order and join point-to-point; a scatter graph never joins in observation order.

Worked example (3H June 2024 Q8, adapted). A table of six values has mean $(8+9+9+6+9+10) \div 6 = 8.5$; a pie chart's 162° sector is $\frac{162}{360} \times 100 = 45\%$. A question combining "four times the mean" with the percentage: $4 \times 8.5 + 45 = 34 + 45 = 79$.

40 students: frequencies determine sector angles



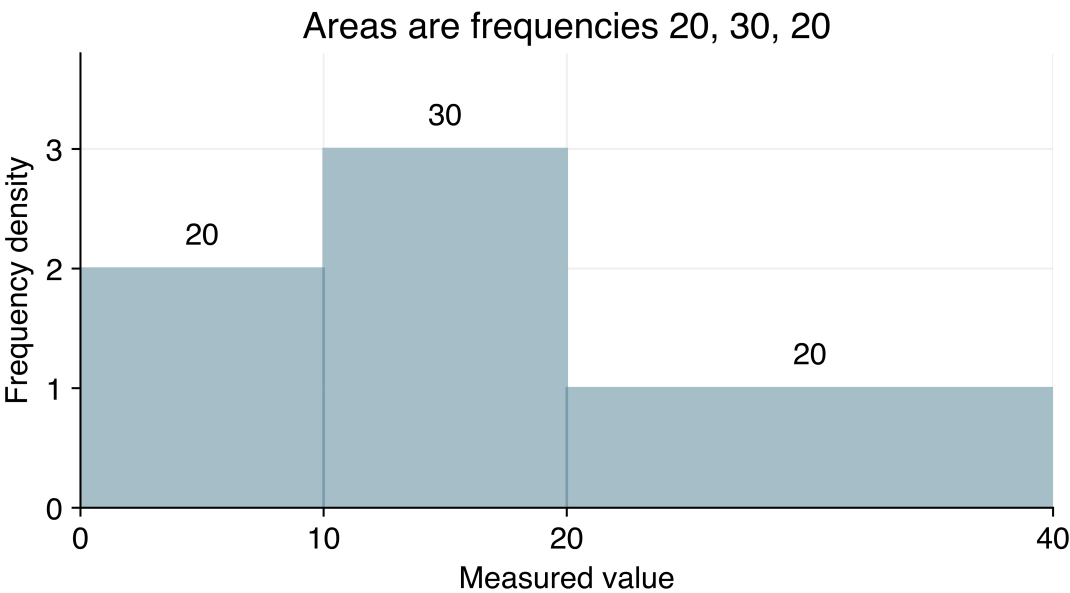
Original native-lesson illustration; labels belong to its worked example.



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

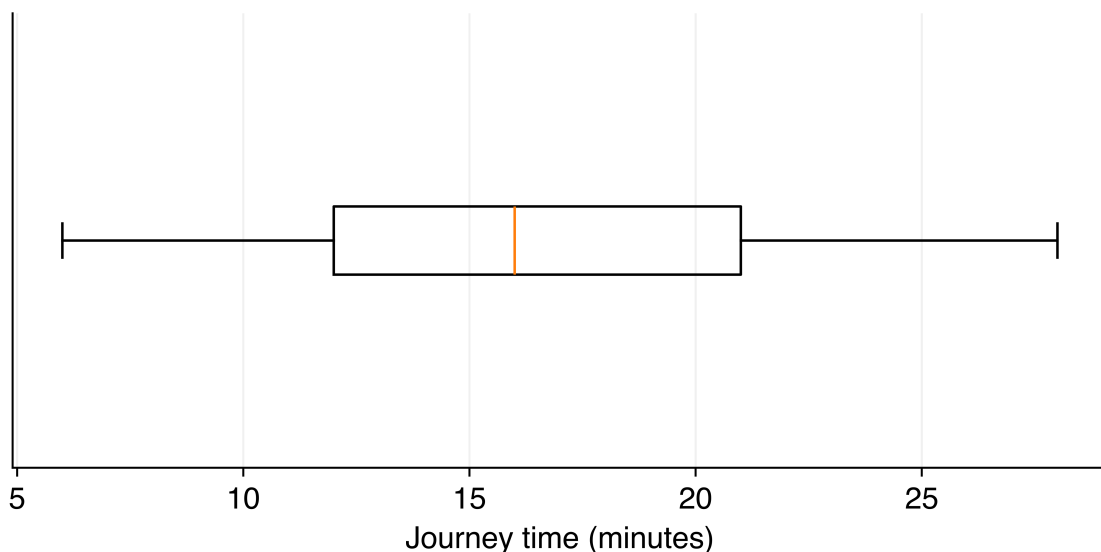
Histograms and cumulative frequency

Higher (S3): construct and interpret diagrams for grouped discrete and continuous data – histograms with equal and unequal class intervals and **cumulative frequency** graphs. Histogram AREA is frequency, so height (density) is frequency $\div$ class width: classes $0 \leq x < 10$, $10 \leq x < 20$, $20 \leq x < 40$ with frequencies 20, 30, 20 have densities 2, 3, 1 – the wide last bar is the shortest yet still holds 20. A cumulative-frequency graph plots each upper class boundary against the running total, starting at zero: for those classes the points are (0,0), (10,20), (20,50), (40,70); the median reads at 35, giving $10 + \frac{35-20}{30} \times 10 = 15$. Values within classes are interpolation estimates.



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

The box contains the middle half



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

Averages, spread and box plots

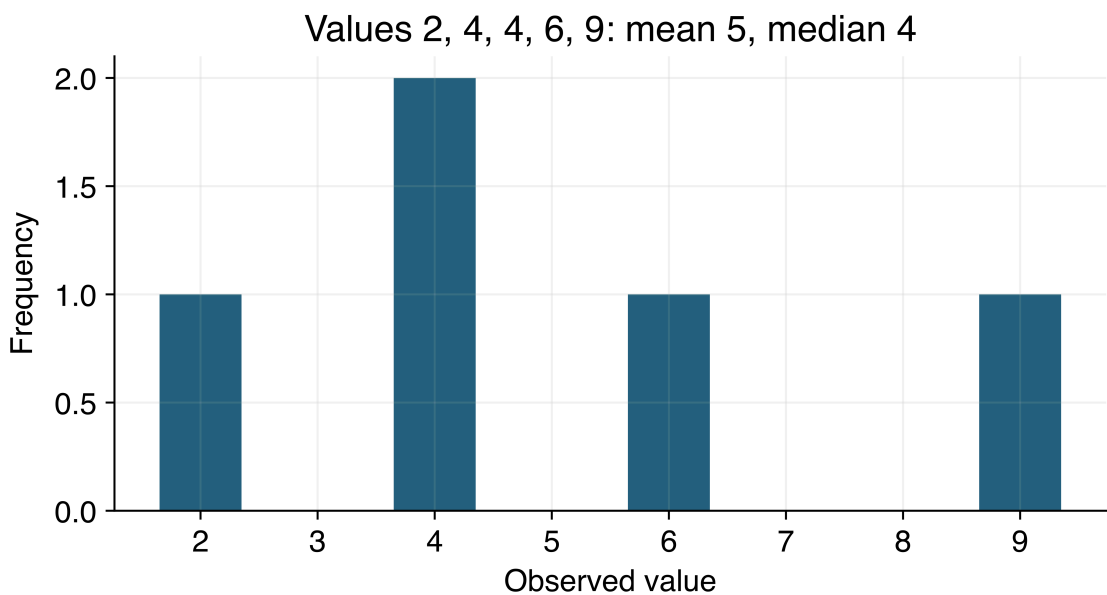
Interpret, analyse and compare distributions of data sets from univariate empirical distributions through appropriate graphical representation and appropriate measures of central tendency – median, mean, mode and modal class – and spread: range, including consideration of **outliers** (S4). From a frequency table, mean is $\frac{\sum fx}{\sum f}$: values 1, 2, 3 with frequencies 2, 5, 3 give $\frac{2+10+9}{10} = 2.1$; grouped data use midpoints, so the result is an estimate. The unusual value in 2, 4, 4, 6, 24 lifts the mean to 8 while the median stays 4.

Worked example (2H June 2024 Q13, adapted). A table of lengths: 15 values of 7.2, 18 of 7.6 and 7 of 8 give $\sum fx = 108 + 136.8 + 56 = 300.8$; mean = $300.8 \div 40 = 7.52$ – keep the products exact (approximated fx values lose the method mark).

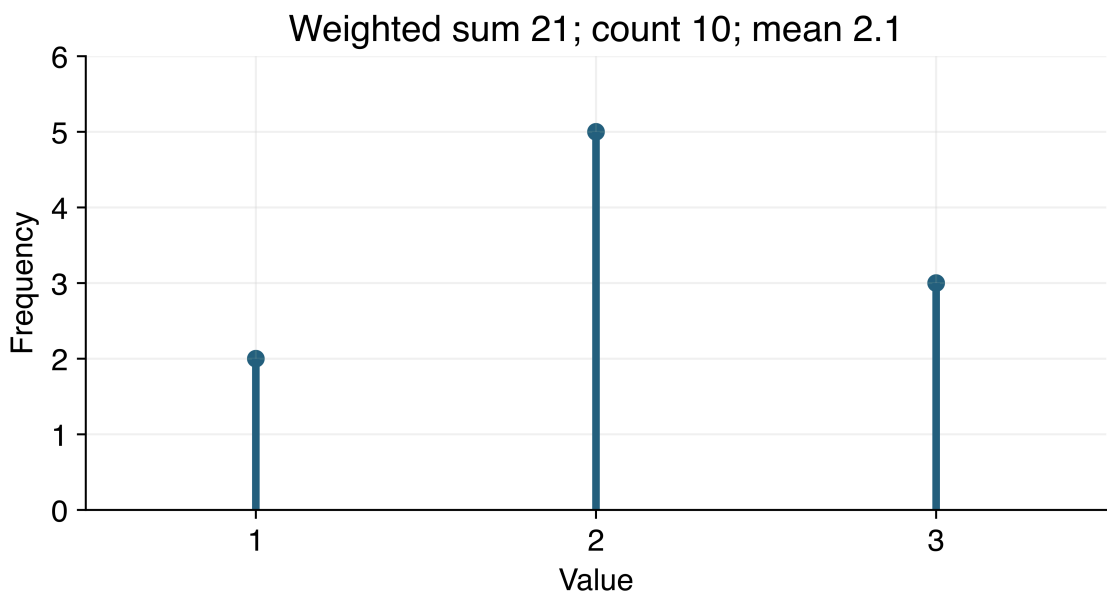
Higher adds box plots, quartiles and the inter-quartile range: a box plot marks minimum, Q_1 , median, Q_3 , maximum; $IQR = Q_3 - Q_1$ describes the middle half. Compare the MEDIAN for a typical value and the IQR for consistency, in context, with units.

Worked example (2H June 2024 Q16, adapted). A box plot with shortest value 2.4 cm, lower quartile 4 cm, median 6 cm, upper quartile 7 cm and longest value 9 cm – five points, one rectangle with whiskers.

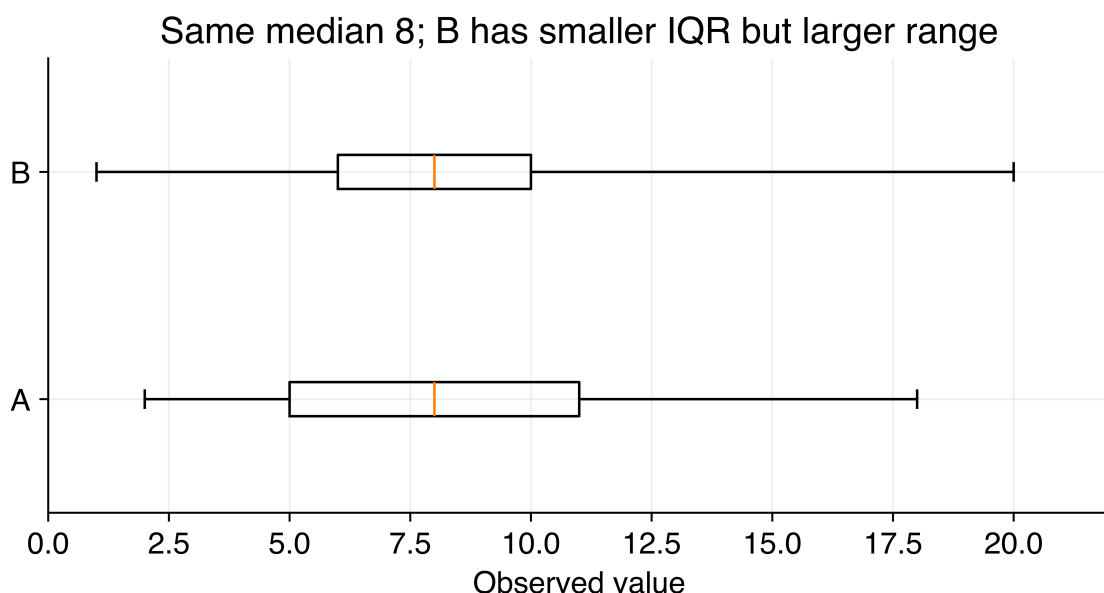
Worked example (1H June 2024 Q16, adapted). Two clubs' ages: "the swimming club's median age is lower, so its average age is lower" AND "the swimming club's inter-quartile range is higher, so its ages are less consistent" – a comparison needs BOTH an average and a spread statement.



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



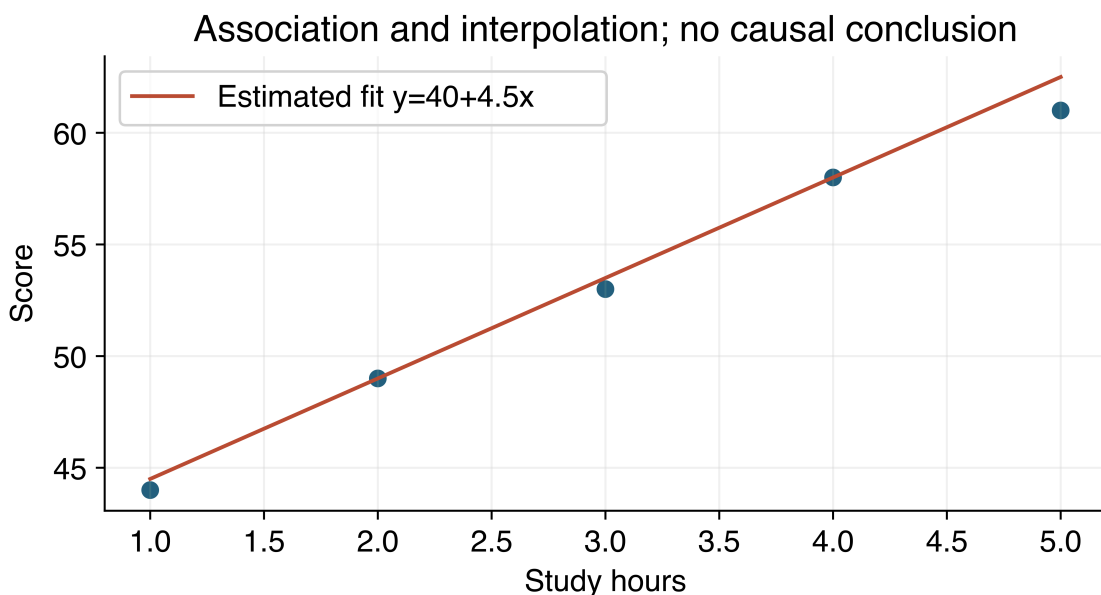
Original native-lesson illustration; labels belong to its worked example.



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

Scatter graphs and correlation

Use and interpret scatter graphs of bivariate data (S6): recognise positive, negative, no, weak and strong correlation; additional foundation knows that correlation does NOT indicate causation; draw estimated lines of best fit through the middle of the pattern (balancing points, not necessarily through the origin) and make predictions; interpolate and extrapolate apparent trends whilst knowing the dangers. Study hours 1–5 against scores 44, 49, 53, 58, 61 with line $\hat{y} = 40 + 4.5x$: predicting at $x = 3.6$ interpolates ($\hat{y} = 56.2$); at $x = 10$ the line gives 85, but gains may flatten – state the observed input range with every prediction.



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

Checklist

- Samples vs census; inference to populations with limitations stated; primary/secondary, discrete/continuous data (S1, S5).
- Frequency tables, bar/pie/pictogram, vertical line charts, time-series tables and line graphs; choosing suitable diagrams (S2; Q8 anchor).
- Histograms with equal and unequal widths (area = frequency, density = $f \div \text{width}$) and cumulative frequency graphs at boundaries (S3).
- Mean from tables with exact products; median/mode/modal class; range and outliers; box plots, quartiles, IQR; median-and-IQR comparisons (S4; Q13, Q16, 1H Q16 anchors).
- Scatter graphs: correlation types and strength; correlation is not causation; estimated best-fit lines; interpolation safe, extrapolation dangerous (S6).

Boundary notes: nothing for this topic appears on the enclosed Formulae Sheet – the mean, range, pie angle and density relationships must be known. Comparisons state an average AND a spread with context and units. Worked examples are adapted from the June 2024 Higher papers (see the meta’s assessment evidence) or original.