COMPREHENSIVE LESSON NOTEBOOK

AQA GCSE Mathematics

HIGHER TIER • 8300

UNIT 6 | STATISTICS

HIGHER PURPOSE

Collect defensible data, represent grouped distributions correctly, estimate quantiles, compare distributions and make cautious population inferences from evidence.

Specification map and Higher progression

Lesson	AQA	Coverage
1	S1,S4,S5	Data, sampling and inference
2	S2	Statistical tables, diagrams and time series
3	S4	Averages, grouped estimates and modal class
4	S4	Quartiles, IQR, range and outliers
5	S3	Histograms and unequal class intervals
6	S3	Cumulative frequency graphs and estimates
7	S4	Box plots and comparing distributions
8	S4,S5	Comparing distributions and describing populations
9	S6	Scatter graphs, correlation and prediction
10	S1-S6	Critical statistical reasoning and representation choice

HIGHER-ONLY FOCUS

S3 histograms and cumulative-frequency graphs, plus S4 box plots, quartiles and interquartile range, receive full construction and interpretation treatment.

Formula and interpretation bank

Measure	Relationship
Grouped mean estimate	Σ(midpoint×frequency)/Σfrequency
Frequency density	frequency/class width
Interquartile range	Q3-Q1
Outlier fences	Q1-1.5IQR and Q3+1.5IQR
Cumulative readings	Q1 at N/4; median N/2; Q3 at 3N/4
Comparison	centre + spread + context + limitation

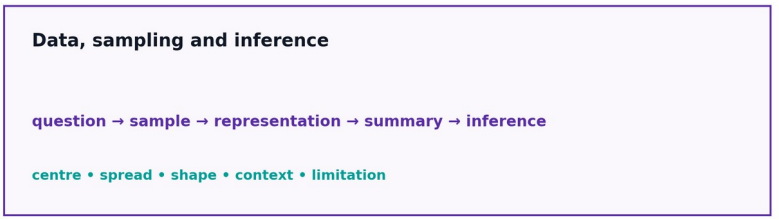
LESSON 01 • S1,S4,S5

Data, sampling and inference

Higher statistics • representation • comparison • inference

Specification focus

AQA references S1,S4,S5. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

Statistics begins with a population question, then collects data from a census or sample. The population is the entire group the conclusion concerns; the sample is the observed subset.

Primary data is collected for the investigation; secondary data already exists. Discrete data is counted, continuous data measured, and categorical data records labels.

A representative sample reflects relevant population features. Simple random sampling gives every member an equal selection chance; systematic sampling uses a fixed interval after a random start.

Stratified sampling preserves group proportions. For stratum size s in population N and sample n, choose approximately s/N×n, resolving rounding so the final sample size is correct.

Build the method

Convenience and voluntary-response samples are vulnerable to bias. A large biased sample estimates the wrong distribution very precisely; size cannot repair exclusion.

Non-response can create bias if respondents differ systematically from non-respondents. Record response rate and consider follow-up or weighting cautiously.

Inference extends sample findings to a population, but conclusions inherit sampling variability, measurement error, question wording and time/place limitations.

A strong statistical conclusion names the target population, quotes evidence and states a limitation rather than claiming proof.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Population for 80 of 1200 pupils surveyed.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain all 1200 pupils.

3 **Interpret:** Return to population, units and limitations.

ANSWER

all 1200 pupils

Worked example 2

QUESTION

Stratified sample 60 from group 180 in population 900, total sample 200.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is 40.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

40

Guided Higher practice

GUIDED QUESTION 1

Why random start in systematic sample?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Classify reaction time.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Population for 80 of 1200 pupils surveyed.	4. Classify reaction time.
2. Stratified sample 60 from group 180 in population 900, total sample 200.	5. Why large convenience sample may fail.
3. Why random start in systematic sample?	6. Difference primary/secondary.

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

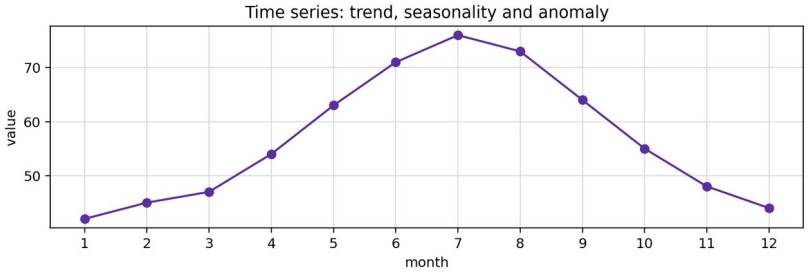
LESSON 02 • S2

Statistical tables, diagrams and time series

Higher statistics • representation • comparison • inference

Specification focus

AQA references S2. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

A display should match data structure: bar or pie chart for categorical data, vertical line chart for ungrouped discrete values, and line graph for time-ordered observations.

Frequency tables must have mutually exclusive and exhaustive categories. Grouped intervals should be unambiguous, for example $0<t\leq10$ followed by $10<t\leq20$.

Bar-chart gaps show separate categories. Equal-width bars and a consistent frequency scale are essential; truncated axes can exaggerate small differences.

A pie sector angle is frequency/total×360°. Pie charts show composition but make close comparison difficult and conceal sample size unless totals are stated.

Build the method

Time-series points are plotted in chronological order. Describe trend, seasonal pattern and anomalies separately; joining points does not prove smooth change between observations.

Dual charts require common scales for fair comparison. A pictogram requires a key and accurately proportioned partial symbols.

Choose a diagram for the intended question. A visually attractive chart may still be unsuitable if it hides distribution shape or time order.

Critique titles, labels, units, intervals, missing data and source before interpreting a displayed pattern.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Chart for monthly temperature.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain time-series line graph.

3 **Interpret:** Return to population, units and limitations.

ANSWER

time-series line graph

Worked example 2

QUESTION

Sector for 18 of 45.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is 144°.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

144°

Guided Higher practice

GUIDED QUESTION 1

Why gaps in bar chart?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Problem with equal spacing labels 0,10,20,50.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Chart for monthly temperature.	4. Problem with equal spacing labels 0,10,20,50.
2. Sector for 18 of 45.	5. Meaning of anomaly.
3. Why gaps in bar chart?	6. Why state pie-chart total?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

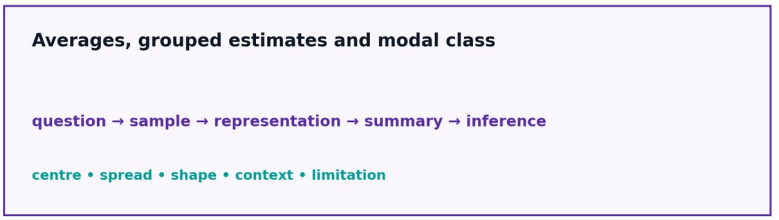
LESSON 03 • S4

Averages, grouped estimates and modal class

Higher statistics • representation • comparison • inference

Specification focus

AQA references S4. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

The mean uses every value: sum/count. The median is the ordered middle, and the mode is most frequent. Choice depends on context and distribution shape.

For a frequency table, mean=Σfx/Σf. The product fx represents the total contribution of value x across its frequency.

For grouped continuous data, exact values are unknown. Use class midpoints to estimate the total, then divide by total frequency; label the result an estimate.

The modal class has greatest frequency for equal-width classes, but for unequal widths the tallest histogram bar corresponds to greatest frequency density, not necessarily greatest frequency.

Build the method

The mean is sensitive to extreme values; the median is resistant. The mode can describe categories, while mean and median require ordered numerical data.

A weighted mean multiplies each value by its weight before dividing by total weight. Frequencies are a particular kind of weight.

Combining groups requires reconstructing totals: group total=mean×size. Averaging two means directly is valid only for equal group sizes.

Check estimated means lie within the data range and use units. Retain midpoint precision until the final step.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Mean of 2,4,4,5,10.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain 5.

3 **Interpret:** Return to population, units and limitations.

ANSWER

5

Worked example 2

QUESTION

$\Sigma fx=386$, $\Sigma f=40$. Mean.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is 9.65.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

9.65

Guided Higher practice

GUIDED QUESTION 1

Midpoint of $20<x\leq30$.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Why grouped mean estimated?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Mean of 2,4,4,5,10.	4. Why grouped mean estimated?
2. $\Sigma fx=386$, $\Sigma f=40$. Mean.	5. Combine mean 12 size 20 and mean 15 size 30.
3. Midpoint of $20<x\leq 30$.	6. Can categorical data have mean?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

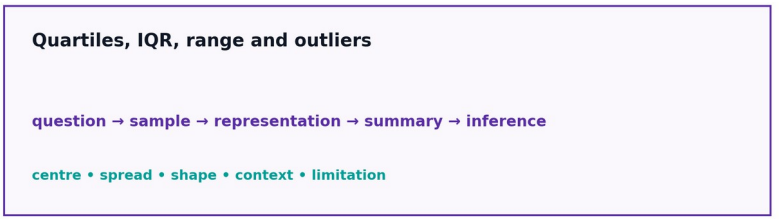
LESSON 04 • S4

Quartiles, IQR, range and outliers

Higher statistics • representation • comparison • inference

Specification focus

AQA references S4. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

Quartiles divide ordered data into four equal parts. Q1 is the lower quartile, Q2 the median and Q3 the upper quartile. State or follow the position convention implied by the question.

Range=max-min measures total spread but depends only on two values. Interquartile range IQR=Q3-Q1 measures spread of the middle 50% and resists extremes.

A smaller IQR indicates greater consistency in the central half. A larger median indicates a higher typical value, but context decides whether higher is better.

A common outlier rule flags values below Q1-1.5IQR or above Q3+1.5IQR. If a question supplies a different definition, use it.

Build the method

An outlier may be error or genuine exceptional data. Investigate before removal and report whether conclusions change with inclusion.

When data count is odd or even, order first and locate quartiles carefully. Do not calculate quartiles from unsorted values.

Comparisons need centre and spread: quote medians/means and IQR/range, then interpret in context.

Range and IQR describe different parts of a distribution; neither alone captures skew, clusters or gaps.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Q1=12,Q3=28. IQR.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain 16.

3 **Interpret:** Return to population, units and limitations.

ANSWER

16

Worked example 2

QUESTION

Outlier fences.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is -12 and 52.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

-12 and 52

Guided Higher practice

GUIDED QUESTION 1

Median A=54,B=61. Typical higher?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

IQR A=8,B=15. More consistent?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Q1=12,Q3=28. IQR.	4. IQR A=8,B=15. More consistent?
2. Outlier fences.	5. Why IQR robust?
3. Median A=54,B=61. Typical higher?	6. Should every outlier be deleted?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

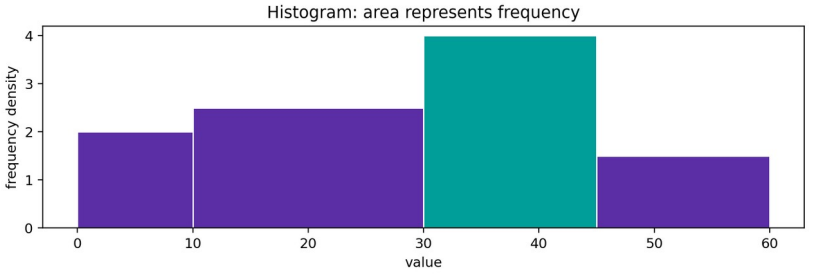
LESSON 05 • S3

Histograms and unequal class intervals

Higher statistics • representation • comparison • inference

Specification focus

AQA references S3. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

A histogram represents continuous or grouped numerical data. Adjacent bars touch because class intervals form a continuous scale.

When class widths differ, bar area—not height—must represent frequency. Frequency density=frequency/class width, so frequency=class width×frequency density.

The vertical axis is frequency density. Its units are frequency per horizontal unit; a missing label makes interpretation ambiguous.

To construct, calculate each class width and density, draw bars across exact boundaries, and choose a linear density scale.

Build the method

To read frequency, find the bar area numerically. Never compare heights alone when widths differ.

A partial interval estimate assumes observations are evenly distributed within the class. State that assumption when estimating how many lie below a value inside a bar.

The modal class has highest density, corresponding to the tallest bar. This differs from greatest raw frequency for unequal widths.

Check total bar area in scale units corresponds to total frequency and intervals have no gaps or overlap.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Class 0-10 freq 20: density.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain 2.

3 **Interpret:** Return to population, units and limitations.

ANSWER

2

Worked example 2

QUESTION

Class 10-30 freq 50: density.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is 2.5.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

2.5

Guided Higher practice

GUIDED QUESTION 1

Which taller?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Bar density 4 width 15: frequency.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Class 0-10 freq 20: density.	4. Bar density 4 width 15: frequency.
2. Class 10-30 freq 50: density.	5. Estimate within half class under uniformity.
3. Which taller?	6. Why bars touch?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

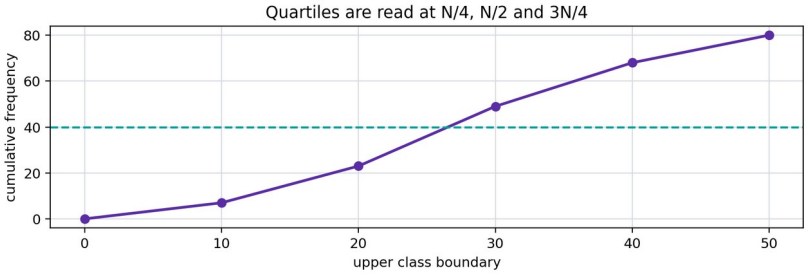
LESSON 06 • S3

Cumulative frequency graphs and estimates

Higher statistics • representation • comparison • inference

Specification focus

AQA references S3. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

Cumulative frequency is a running total. For each class, add its frequency to previous classes and plot against the upper class boundary.

The graph begins at the lowest boundary with cumulative frequency zero and ends at total frequency. Points are joined with a smooth increasing curve or appropriate segments.

Median is read at N/2, lower quartile at N/4 and upper quartile at 3N/4. Move horizontally to curve, then vertically to the data axis.

IQR is Q3-Q1. Percentiles generalise this: the pth percentile is read at p/100×N.

Build the method

To estimate number above a value x, subtract cumulative frequency at x from total N. For an interval, subtract the two cumulative frequencies.

Graph readings are estimates affected by plotting accuracy and the assumption about distribution within classes. State sensible precision.

A cumulative curve cannot decrease; a downward segment signals an error. A steeper section indicates greater concentration per horizontal interval.

Comparing curves horizontally at the same cumulative proportion compares quantiles; comparing vertically at the same x compares counts below x.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

N=80: median reading level.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain 40.

3 **Interpret:** Return to population, units and limitations.

ANSWER

40

Worked example 2

QUESTION

Q1 level.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is 20.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

20

Guided Higher practice

GUIDED QUESTION 1

Q3 level.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

If CF at 35 is 52, number above 35.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. N=80: median reading level.	4. If CF at 35 is 52, number above 35.
2. Q1 level.	5. CF 20 at x=10 and 58 at x=30: number 10-30.
3. Q3 level.	6. Why graph never decreases?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

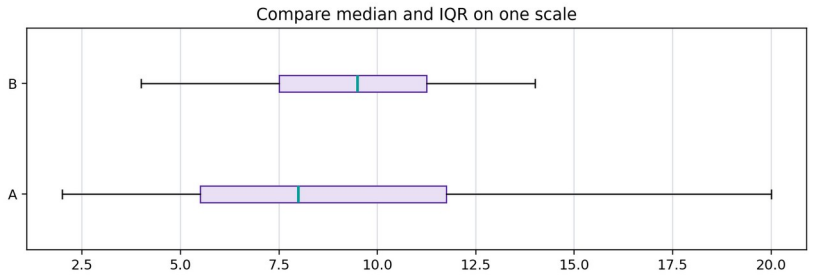
LESSON 07 • S4

Box plots and comparing distributions

Higher statistics • representation • comparison • inference

Specification focus

AQA references S4. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

A box plot displays minimum, Q1, median, Q3 and maximum on a numerical scale. The box length is IQR; whiskers show outer spread.

Construct by drawing a scale, vertical markers at five-number summary, a box from Q1 to Q3, median line and whiskers to extremes.

Box plots compare centre and spread compactly but do not reveal sample size, clusters, gaps or detailed shape.

A larger median indicates higher typical value. Smaller IQR indicates greater consistency in the middle 50%; smaller range indicates smaller overall spread.

Build the method

Skew can be suggested by unequal box halves or whiskers, but use cautious language because the plot is highly summarised.

Plots must share the same scale for visual comparison. Measure from axis values, not apparent lengths across differently scaled diagrams.

If outliers are plotted separately under a stated convention, whiskers may extend to non-outlier extremes; follow the definition supplied.

Write comparisons in context and quote values: centre statement plus spread statement, followed by a practical interpretation.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Five-number 2,6,9,14,20: IQR.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain 8.

3 **Interpret:** Return to population, units and limitations.

ANSWER

8

Worked example 2

QUESTION

Range.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is 18.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

18

Guided Higher practice

GUIDED QUESTION 1

Median A 45,B52: higher typical.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

IQR A12,B7: more consistent.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Five-number 2,6,9,14,20: IQR.	4. IQR A12,B7: more consistent.
2. Range.	5. One limitation box plot.
3. Median A 45,B52: higher typical.	6. Why common scale?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

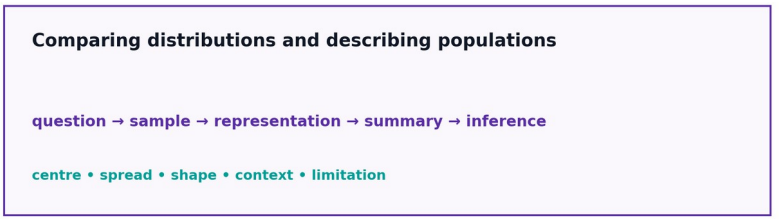
LESSON 08 • S4,S5

Comparing distributions and describing populations

Higher statistics • representation • comparison • inference

Specification focus

AQA references S4,S5. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

A distribution includes centre, spread and shape. No single statistic fully describes it, so select measures suited to the data and question.

Symmetric data without strong outliers often suit mean and range or standard graphical description; skewed data or outliers often suit median and IQR.

Comparisons must quote values for both groups using matched measures. “Group A is better” without centre and spread evidence is incomplete.

Greater consistency means smaller spread, not necessarily smaller values. Higher average and greater consistency are separate claims.

Build the method

Sample conclusions refer to the sampled population only when sampling is representative. Distinguish observed sample facts from inferred population claims.

A difference may be practically small even if numerically present. Consider units, measurement accuracy and context.

Aggregation can hide subgroup patterns. Where relevant, compare like with like and investigate confounding variables.

A balanced conclusion acknowledges overlap: one group may have a higher median while distributions still overlap substantially.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

A mean 68 range 12; B mean72 range20: compare.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain B typically higher; A more consistent..

3 **Interpret:** Return to population, units and limitations.

ANSWER

B typically higher; A more consistent.

Worked example 2

QUESTION

Which measure with severe outlier?

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is median and IQR.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

median and IQR

Guided Higher practice

GUIDED QUESTION 1

What does smaller spread mean?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Can sample prove whole population?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. A mean 68 range 12; B mean72 range20: compare.	4. Can sample prove whole population?
2. Which measure with severe outlier?	5. Why quote units?
3. What does smaller spread mean?	6. What can aggregation hide?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

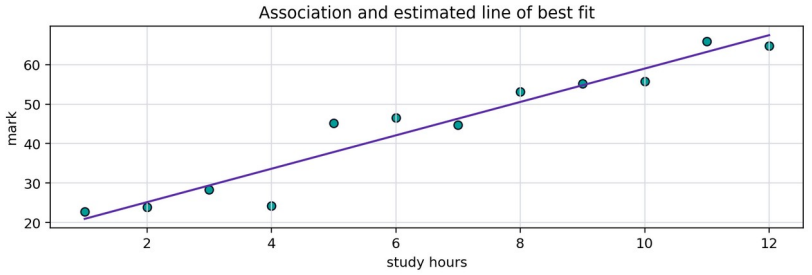
LESSON 09 • S6

Scatter graphs, correlation and prediction

Higher statistics • representation • comparison • inference

Specification focus

AQA references S6. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

A scatter graph pairs two measurements from each individual. Each point must preserve the pairing; separate unpaired lists cannot create meaningful bivariate data.

Positive correlation rises, negative falls and no correlation has no clear trend. Strength describes how tightly points follow a pattern.

Correlation does not imply causation. A lurking variable, reverse direction or coincidence may explain association.

A line of best fit represents the central trend with roughly balanced points above and below. It need not pass through the origin or every point.

Build the method

Interpolation predicts within the observed x-range and is usually safer. Extrapolation extends beyond evidence and may fail if the relationship changes.

An outlier may strongly affect correlation and the line. Investigate whether it is error, special case or important evidence.

Predictions are estimates. Read with appropriate precision and do not predict an individual with certainty from a population trend.

Curved association may be strong despite weak linear correlation; inspect the plot before fitting a straight line.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Hours study vs mark rises tightly: describe.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain strong positive correlation.

3 **Interpret:** Return to population, units and limitations.

ANSWER

strong positive correlation

Worked example 2

QUESTION

Why correlation not causation?

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is Other variables or reverse causation may explain it..

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

Other variables or reverse causation may explain it.

Guided Higher practice

GUIDED QUESTION 1

Safer: interpolation/extrapolation.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Should line join points?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Hours study vs mark rises tightly: describe.	4. Should line join points?
2. Why correlation not causation?	5. Effect of outlier.
3. Safer: interpolation/extrapolation.	6. Can strong correlation predict exactly?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

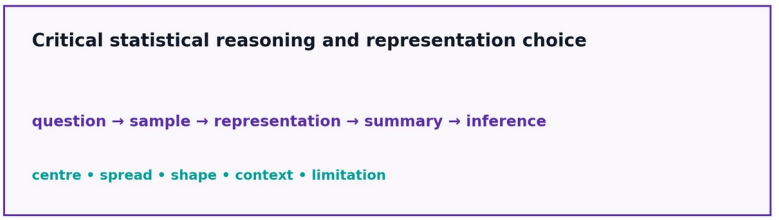
LESSON 10 • S1-S6

Critical statistical reasoning and representation choice

Higher statistics • representation • comparison • inference

Specification focus

AQA references S1-S6. This lesson develops the data model, construction or calculation, interpretation and limitations expected at Higher tier.



Understand the idea

Statistical reasoning follows a chain: question, population, collection method, representation, summary, inference and limitation. Weakness at any stage affects the conclusion.

Choose diagrams by variable type and purpose: category comparison, time trend, grouped distribution, cumulative quantiles, five-number comparison or bivariate association.

A misleading graph may truncate an axis, distort area, use unequal intervals without density, omit sample size or select a favourable time window.

Question wording can lead respondents, overlap response categories or omit realistic options. Pilot questions and ensure categories are exhaustive and mutually exclusive.

Build the method

Missing data are not automatically random. If absence relates to the measured variable, available cases may be systematically biased.

Averages can be manipulated by choosing mean, median or mode. Ask which measure is reported, why it was chosen and what spread accompanies it.

Causation requires more than correlation: timing, mechanism, control of confounding and ideally experimental evidence.

Write a final conclusion proportionate to evidence, separating calculation, interpretation and limitation.

STATISTICAL HABIT

Identify variable type, sample, scale and denominator; quote centre and spread with units, then state what can and cannot be inferred.

COMMON TRAP

Do not confuse histogram height with frequency, treat grouped estimates as exact, infer causation from correlation or generalise from a biased sample.

Worked examples

Worked example 1

QUESTION

Best graph for unequal grouped continuous data.

1 **Represent:** Choose the correct table, graph or summary measure.

2 **Calculate:** Apply the definition or scale carefully to obtain histogram.

3 **Interpret:** Return to population, units and limitations.

ANSWER

histogram

Worked example 2

QUESTION

Best graph for quartile comparison.

1 **Read:** Use class boundaries, cumulative levels or the shared axis.

2 **Execute:** The result is box plot.

3 **Compare:** Quote matched measures and explain practical meaning.

ANSWER

box plot

Guided Higher practice

GUIDED QUESTION 1

Problem with truncated axis.

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

GUIDED QUESTION 2

Why pilot questionnaire?

STRATEGIC HINT

Write the denominator or graphical scale first; label estimates and support every comparison with two quoted values.

Independent and Grade 7–9 reasoning

Questions 1–3	Questions 4–6
1. Best graph for unequal grouped continuous data.	4. Why pilot questionnaire?
2. Best graph for quartile comparison.	5. Why report spread with average?
3. Problem with truncated axis.	6. What extra evidence supports causation?

REASONING EXTENSION

Critique the sampling or display, compare another valid measure and explain how the limitation changes the strength—not merely the arithmetic—of the conclusion.

Complete answer and reasoning guide

Lesson 1: Data, sampling and inference

1. all 1200 pupils

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. 40

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. To avoid a repeating-list pattern selecting a biased subset.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. continuous numerical

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. Selection bias remains.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. Primary is collected first-hand; secondary was collected previously.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 2: Statistical tables, diagrams and time series

1. time-series line graph

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. 144°

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. Categories are separate.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. Unequal numerical intervals are shown as equal distances.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. An unusually different observation.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. Equal sectors can represent very different sample sizes.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 3: Averages, grouped estimates and modal class

1. 5

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. 9.65

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. 25

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. Original values within classes are unknown.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. 13.8

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. No

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 4: Quartiles, IQR, range and outliers

1. 16

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. -12 and 52

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. B

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. A

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. It ignores the lowest and highest quarters.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. No; determine whether it is error or valid.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 5: Histograms and unequal class intervals

1. 2

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. 2.5

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. 10-30 class

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. 60

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. half the class frequency

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. The variable is continuous and intervals are adjacent.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 6: Cumulative frequency graphs and estimates

1. 40

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. 20

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. 60

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. 28

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. 38

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. Cumulative totals can only stay same or increase.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 7: Box plots and comparing distributions

1. 8

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. 18

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. B

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. B

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. It hides detailed distribution and sample size.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. Otherwise visual lengths are not comparable.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 8: Comparing distributions and describing populations

1. B typically higher; A more consistent.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. median and IQR

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. greater consistency

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. No

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. They give magnitude and context.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. Different subgroup patterns or confounding.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 9: Scatter graphs, correlation and prediction

1. strong positive correlation

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. Other variables or reverse causation may explain it.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. interpolation

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. No

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. It can distort trend and predictions.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. No

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Lesson 10: Critical statistical reasoning and representation choice

1. histogram

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

2. box plot

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

3. It exaggerates visual differences.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

4. To detect ambiguity, leading wording and missing options.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

5. Same centre can accompany very different variability.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

6. Controlled evidence, mechanism and correct time order.

Check variable type, class boundaries, scale, units and whether the value is exact or estimated; infer only as far as the sampling and representation justify.

Unit 6 mastery record

AQA	I can demonstrate…	Secure?
S1,S4,S5	Data, sampling and inference	☐
S2	Statistical tables, diagrams and time series	☐
S4	Averages, grouped estimates and modal class	☐
S4	Quartiles, IQR, range and outliers	☐
S3	Histograms and unequal class intervals	☐
S3	Cumulative frequency graphs and estimates	☐
S4	Box plots and comparing distributions	☐
S4,S5	Comparing distributions and describing populations	☐
S6	Scatter graphs, correlation and prediction	☐
S1-S6	Critical statistical reasoning and representation choice	☐

Official specification source

AQA GCSE Mathematics (8300), Subject Content 3.6 Statistics: https://www.aqa.org.uk/subjects/mathematics/gcse/mathematics-8300/specification/subject-content/3.6-statistics

Independently written and not endorsed by AQA or any other examination board. All explanations, examples and diagrams are original and do not reproduce the supplied commercial guide or past-paper questions.