

GCSE MATHEMATICS

FOUNDATION TIER

COMPREHENSIVE LESSON NOTEBOOK

UNIT 6 | STATISTICS

TEACH-FROM-SCRATCH • PUBLICATION EDITION

6 detailed chapters • original diagrams • worked examples • guided practice • answers

INDEPENDENT PUBLICATION NOTICE

Independent GCSE Mathematics learning resource aligned to EDUQAS C300. It is not produced, approved or endorsed by Eduqas or WJEC CBAC Ltd. Eduqas identifies the qualification only. No official examination question, mark scheme, textbook page, logo, trade dress or third-party artwork is reproduced.

© 2026 GCSE Online Tutoring and RS Remote Tutoring. All rights reserved.

COMPREHENSIVE LESSON NOTEBOOK

How to learn from this notebook

Meaning first. Method second. Practice third.

FIVE-PASS ROUTINE

1 Read the key idea without a calculator. 2 Cover each worked solution and reproduce it. 3 Complete practice independently. 4 Mark in a different colour and repair errors. 5 Retest after 2-7 days.

What a complete mathematics answer contains

- A clear model or calculation, not only a calculator display.
- Enough working for another person to follow the reasoning.
- Exact values until approximation is requested.
- Units and sensible rounding in context.
- A final statement when interpretation is required.

Error loop

REPAIR, DO NOT MERELY CORRECT

Label each error: concept, operation, arithmetic, calculator entry, accuracy, unit or interpretation. Write the corrected step and one sentence explaining the change.

Foundation does not mean superficial

Eduqas Foundation questions can combine familiar skills in unfamiliar contexts. Grade 4-5 success depends on selecting a method, connecting several steps and deciding whether the answer is reasonable.

COMPREHENSIVE LESSON NOTEBOOK

Unit 6 coverage map

Eduqas C300 Foundation Statistics content organised into a teachable sequence.

CH 01 **Eduqas C300 FS1**

Sampling and populations

CH 02 **Eduqas C300 FS2**

Questionnaires and fairness

CH 03 **Eduqas C300 FS3**

Tables, charts and time series

CH 04 **Eduqas C300 FS4**

Averages, spread and distributions

CH 05 **Eduqas C300 FS5**

Using statistics to describe populations

CH 06 **Eduqas C300 FS6**

Scatter graphs, correlation and prediction

IMPORTANT TIER NOTE

This notebook covers Eduqas C300 references FS1-FS6 at Foundation Tier: sampling, questionnaires, data displays, distribution summaries, population estimates and scatter graphs. Specification references are identifiers, not copied specification wording.

COMPREHENSIVE LESSON NOTEBOOK

Fast-reference method bank

Understand each relationship before memorising it.

Mean

sum of values / number of values

Frequency-table mean

sum of value x frequency / total frequency

Median

middle ordered value

Range

maximum - minimum

Grouped estimated mean

sum of midpoint x frequency / total frequency

Pie-chart angle

frequency / total x 360 degrees

Population estimate

sample proportion x population size

Correlation

association does not prove causation

Interpolation

prediction inside observed range

Extrapolation

prediction outside observed range

CURRENT ASSESSMENT NOTE

Eduqas states that an additional formula sheet is integrated into examination papers from Summer 2026. Check the official Eduqas qualification page for the arrangement applying to your examination series.

COMPREHENSIVE LESSON NOTEBOOK

Sampling and populations

Infer properties of populations from samples while recognising the limitations of sampling.

SPECIFICATION FOCUS

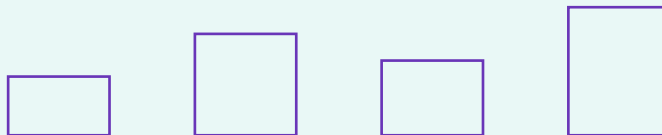
Specification reference Eduqas C300 FS1. This lesson develops representation, accurate calculation, reasoning and interpretation within Foundation Tier.

Understand the idea

A sample supports conclusions about a population only when its selection gives the population a fair chance of being represented.

KEY VOCABULARY

population • sample • random • representative • bias • inference



Why the method works

- The method preserves the link between the observations and the conclusion.
- A suitable display reveals the relevant pattern without changing the underlying data.
- Matched summaries and explicit limitations make an inference auditable.

CHECK BEFORE MOVING ON

Answer mentally: An unbiased sample of 50 people contains 22 who prefer option A. Estimate how many of a population of 600 prefer A and state why the answer is only an estimate. Then define one key word from this page in your own words.

COMPREHENSIVE LESSON NOTEBOOK

Sampling and populations - method and examples

Follow the reasoning; do not copy steps mechanically.

Reliable method

- Name the target population before judging a sample.
- A representative sampling method matters more than size alone.
- Scale a sample proportion to the population only when the sample can support the inference.

WORKED EXAMPLE 1

An unbiased sample of 50 people contains 22 who prefer option A. Estimate how many of a population of 600 prefer A and state why the answer is only an estimate.

SOLUTION

1. Sample proportion = $22/50$. Estimate = $600 \times 22/50 = 264.00$, which is about 264 people. Sampling variation means the true population value may differ.

ANSWER: About 264; a sample may not represent the population exactly

WORKED EXAMPLE 2

A school wants pupils' views on homework. Survey 2 asks only pupils attending an after-school homework club. Identify the sampling problem and suggest an improvement.

SOLUTION

1. Homework-club pupils may have different views from other pupils. A random or stratified sample from the full school population is more representative.

ANSWER: The sample is biased; select pupils randomly from the whole school

SELF-CHECK QUESTION

Which step carries the mathematical reason, and which step carries only arithmetic? Being able to distinguish them makes a solution easier to repair.

COMPREHENSIVE LESSON NOTEBOOK

Sampling and populations - apply and reason

Connect the method to decisions and common errors.

Where this mathematics appears

opinion polling and market research rely on transparent data collection and proportionate conclusions.

Common misconceptions

- Generalising beyond the population represented by the sample.
- Using overlapping categories, distorted scales or an unsuitable average.
- Reporting a statistic without units, context, spread or uncertainty.

GUIDED TRY

Two random samples from the same population contain 35 and 150 people. Which sample is generally more reliable, and give one limitation that may remain? First predict the form or approximate size of the answer. Then show the complete method.

REASONING PROMPT

A school wants pupils' views on homework. Survey 8 asks only pupils attending an after-school homework club. Identify the sampling problem and suggest an improvement. Explain why each condition or operation you use is valid.

EXAM COMMUNICATION

Use one mathematical statement per line. Label units and restrictions. If the result is approximate, show the unrounded value before the final rounded answer.

COMPREHENSIVE LESSON NOTEBOOK

Independent practice

Sampling and populations

Complete without notes. Show all working, then use the answer section to diagnose and repair errors.

1. An unbiased sample of 50 people contains 22 who prefer option A. Estimate how many of a population of 600 prefer A and state why the answer is only an estimate. [4]

2. A school wants pupils' views on homework. Survey 2 asks only pupils attending an after-school homework club. Identify the sampling problem and suggest an improvement. [3]

3. Two random samples from the same population contain 35 and 150 people. Which sample is generally more reliable, and give one limitation that may remain? [3]

4. An unbiased sample of 80 people contains 34 who prefer option A. Estimate how many of a population of 900 prefer A and state why the answer is only an estimate. [4]

5. A school wants pupils' views on homework. Survey 5 asks only pupils attending an after-school homework club. Identify the sampling problem and suggest an improvement. [3]

6. Two random samples from the same population contain 50 and 180 people. Which sample is generally more reliable, and give one limitation that may remain? [3]

7. An unbiased sample of 110 people contains 46 who prefer option A. Estimate how many of a population of 1200 prefer A and state why the answer is only an estimate. [4]

8. A school wants pupils' views on homework. Survey 8 asks only pupils attending an after-school homework club. Identify the sampling problem and suggest an improvement. [3]

REFLECTION

Circle one: secure / almost / reteach. Record the question number of your most useful error and the rule that repairs it.

COMPREHENSIVE LESSON NOTEBOOK

Questionnaires and fairness

Design and criticise questionnaire questions, response options and sampling choices.

SPECIFICATION FOCUS

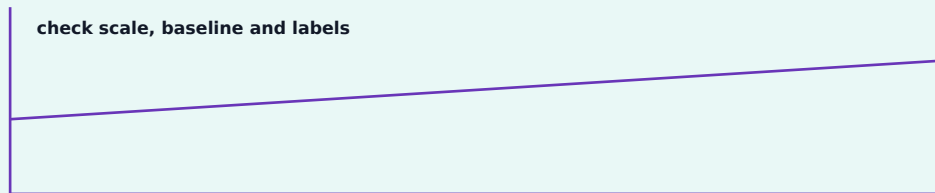
Specification reference Eduqas C300 FS2. This lesson develops representation, accurate calculation, reasoning and interpretation within Foundation Tier.

Understand the idea

Fair questionnaire design reduces measurement bias by using neutral language and unambiguous response categories.

KEY VOCABULARY

questionnaire • leading • response option • boundary • pilot



Why the method works

- The method preserves the link between the observations and the conclusion.
- A suitable display reveals the relevant pattern without changing the underlying data.
- Matched summaries and explicit limitations make an inference auditable.

CHECK BEFORE MOVING ON

Answer mentally: Questionnaire item 1 asks: “Don’t you agree that our excellent new library should stay open later?” Explain two problems and rewrite it fairly. Then define one key word from this page in your own words.

COMPREHENSIVE LESSON NOTEBOOK

Questionnaires and fairness - method and examples

Follow the reasoning; do not copy steps mechanically.

Reliable method

- Use neutral wording that does not suggest a preferred response.
- Give exhaustive, mutually exclusive answer options with a stated time period.
- Pilot the question to find ambiguity before collecting the main data.

WORKED EXAMPLE 1

Questionnaire item 1 asks: "Don't you agree that our excellent new library should stay open later?" Explain two problems and rewrite it fairly.

SOLUTION

1. Remove language that encourages one answer and provide balanced, exhaustive response options.

ANSWER: It is leading and uses emotive wording; e.g. "Should the library stay open later? Yes / No / Not sure"

WORKED EXAMPLE 2

A questionnaire asks weekly study time using: 0-3 hours, 3-5 hours, 5-7 hours. Identify the boundary problem and give non-overlapping intervals.

SOLUTION

1. A value at a shared endpoint could fit two boxes. Use one included and one excluded boundary, then add $t \geq 7$ so the categories are exhaustive.

ANSWER: The endpoints overlap; e.g. $0 \leq t < 3$, $3 \leq t < 5$, $5 \leq t < 7$, $t \geq 7$

SELF-CHECK QUESTION

Which step carries the mathematical reason, and which step carries only arithmetic? Being able to distinguish them makes a solution easier to repair.

COMPREHENSIVE LESSON NOTEBOOK

Questionnaires and fairness - apply and reason

Connect the method to decisions and common errors.

Where this mathematics appears

surveys and service evaluation rely on transparent data collection and proportionate conclusions.

Common misconceptions

- Generalising beyond the population represented by the sample.
- Using overlapping categories, distorted scales or an unsuitable average.
- Reporting a statistic without units, context, spread or uncertainty.

GUIDED TRY

A survey numbered 3 asks “How often do you exercise?” with options Never, Sometimes and Often. Give two improvements that would make responses easier to interpret. First predict the form or approximate size of the answer. Then show the complete method.

REASONING PROMPT

A questionnaire asks weekly study time using: 0-9 hours, 9-11 hours, 11-13 hours. Identify the boundary problem and give non-overlapping intervals. Explain why each condition or operation you use is valid.

EXAM COMMUNICATION

Use one mathematical statement per line. Label units and restrictions. If the result is approximate, show the unrounded value before the final rounded answer.

COMPREHENSIVE LESSON NOTEBOOK

Independent practice

Questionnaires and fairness

Complete without notes. Show all working, then use the answer section to diagnose and repair errors.

- Questionnaire item 1 asks: "Don't you agree that our excellent new library should stay open later?" Explain two problems and rewrite it fairly. [4]

- A questionnaire asks weekly study time using: 0-3 hours, 3-5 hours, 5-7 hours. Identify the boundary problem and give non-overlapping intervals. [3]

- A survey numbered 3 asks "How often do you exercise?" with options Never, Sometimes and Often. Give two improvements that would make responses easier to interpret. [4]

- Questionnaire item 4 asks: "Don't you agree that our excellent new library should stay open later?" Explain two problems and rewrite it fairly. [4]

- A questionnaire asks weekly study time using: 0-6 hours, 6-8 hours, 8-10 hours. Identify the boundary problem and give non-overlapping intervals. [3]

- A survey numbered 6 asks "How often do you exercise?" with options Never, Sometimes and Often. Give two improvements that would make responses easier to interpret. [4]

- Questionnaire item 7 asks: "Don't you agree that our excellent new library should stay open later?" Explain two problems and rewrite it fairly. [4]

- A questionnaire asks weekly study time using: 0-9 hours, 9-11 hours, 11-13 hours. Identify the boundary problem and give non-overlapping intervals. [3]

REFLECTION

Circle one: secure / almost / reteach. Record the question number of your most useful error and the rule that repairs it.

COMPREHENSIVE LESSON NOTEBOOK

Tables, charts and time series

Interpret and construct frequency tables, bar charts, pie charts, pictograms, vertical line charts and time-series graphs.

SPECIFICATION FOCUS

Specification reference Eduqas C300 FS3. This lesson develops representation, accurate calculation, reasoning and interpretation within Foundation Tier.

Understand the idea

Statistical displays encode count, proportion or change through carefully labelled position, length, angle or symbol.

KEY VOCABULARY

frequency table • bar chart • pie chart • pictogram • time series



Why the method works

- The method preserves the link between the observations and the conclusion.
- A suitable display reveals the relevant pattern without changing the underlying data.
- Matched summaries and explicit limitations make an inference auditable.

CHECK BEFORE MOVING ON

Answer mentally: A survey records four categories with frequencies 4, 6, 8 and 6. Find the pie-chart angle for the third category. Then define one key word from this page in your own words.

COMPREHENSIVE LESSON NOTEBOOK

Tables, charts and time series - method and examples

Follow the reasoning; do not copy steps mechanically.

Reliable method

- Choose a display that matches the data type and purpose.
- Label axes, units, keys and scales so another reader can reconstruct the data.
- Pie-chart angles use frequency divided by total, multiplied by 360 degrees.

WORKED EXAMPLE 1

A survey records four categories with frequencies 4, 6, 8 and 6. Find the pie-chart angle for the third category.

SOLUTION

1. Angle=frequency/total $\times$ 360= $8/24 \times 360=120$ degrees.

ANSWER: 120 degrees

WORKED EXAMPLE 2

On a pictogram, one complete symbol represents 4 people. A category has 5 complete symbols and one half-symbol. Find its frequency.

SOLUTION

1. Frequency= $(5+1/2) \times 4=22$.

ANSWER: 22

SELF-CHECK QUESTION

Which step carries the mathematical reason, and which step carries only arithmetic? Being able to distinguish them makes a solution easier to repair.

COMPREHENSIVE LESSON NOTEBOOK

Tables, charts and time series - apply and reason

Connect the method to decisions and common errors.

Where this mathematics appears

public reports and monitoring rely on transparent data collection and proportionate conclusions.

Common misconceptions

- Generalising beyond the population represented by the sample.
- Using overlapping categories, distorted scales or an unsuitable average.
- Reporting a statistic without units, context, spread or uncertainty.

GUIDED TRY

A time series has consecutive values [15, 18, 14, 21, 23]. State the overall change from first to last and identify the interval with the largest increase. First predict the form or approximate size of the answer. Then show the complete method.

REASONING PROMPT

Scores 1, 2, 3, 4 and 5 have frequencies [9, 11, 12, 10, 8]. State an appropriate chart and the total number of observations. Explain why each condition or operation you use is valid.

EXAM COMMUNICATION

Use one mathematical statement per line. Label units and restrictions. If the result is approximate, show the unrounded value before the final rounded answer.

COMPREHENSIVE LESSON NOTEBOOK

Independent practice

Tables, charts and time series

Complete without notes. Show all working, then use the answer section to diagnose and repair errors.

1. A survey records four categories with frequencies 4, 6, 8 and 6. Find the pie-chart angle for the third category. [3]

2. On a pictogram, one complete symbol represents 4 people. A category has 5 complete symbols and one half-symbol. Find its frequency. [2]

3. A time series has consecutive values [15, 18, 14, 21, 23]. State the overall change from first to last and identify the interval with the largest increase. [3]

4. Scores 1, 2, 3, 4 and 5 have frequencies [5, 7, 8, 6, 4]. State an appropriate chart and the total number of observations. [2]

5. A survey records four categories with frequencies 8, 10, 12 and 10. Find the pie-chart angle for the third category. [3]

6. On a pictogram, one complete symbol represents 4 people. A category has 9 complete symbols and one half-symbol. Find its frequency. [2]

7. A time series has consecutive values [19, 22, 18, 25, 27]. State the overall change from first to last and identify the interval with the largest increase. [3]

8. Scores 1, 2, 3, 4 and 5 have frequencies [9, 11, 12, 10, 8]. State an appropriate chart and the total number of observations. [2]

REFLECTION

Circle one: secure / almost / reteach. Record the question number of your most useful error and the rule that repairs it.

COMPREHENSIVE LESSON NOTEBOOK

Averages, spread and distributions

Interpret, analyse and compare discrete, continuous and grouped data using suitable diagrams, averages and range.

SPECIFICATION FOCUS

Specification reference Eduqas C300 FS4. This lesson develops representation, accurate calculation, reasoning and interpretation within Foundation Tier.

Understand the idea

Measures of centre describe typical value in different ways, while spread and graphs reveal variation and unusual values.

KEY VOCABULARY

mean • median • mode • modal class • range • outlier



Why the method works

- The method preserves the link between the observations and the conclusion.
- A suitable display reveals the relevant pattern without changing the underlying data.
- Matched summaries and explicit limitations make an inference auditable.

CHECK BEFORE MOVING ON

Answer mentally: For the data [1, 3, 4, 6, 6], calculate the mean, median, mode and range. Then define one key word from this page in your own words.

COMPREHENSIVE LESSON NOTEBOOK

Averages, spread and distributions - method and examples

Follow the reasoning; do not copy steps mechanically.

Reliable method

- Order raw data before locating the median.
- Use frequency-weighted totals for means from tables.
- Compare distributions with a matched measure of centre and spread, then interpret in context.

WORKED EXAMPLE 1

For the data [1, 3, 4, 6, 6], calculate the mean, median, mode and range.

SOLUTION

1. Sum=20, so mean= $20/5=4$. The middle value is 4, the repeated value is 6, and range= $(6)-1=5$.

ANSWER: Mean 4; median 4; mode 6; range 5

WORKED EXAMPLE 2

Group A has mean 22 and range 14. Group B has mean 24 and range 8. Compare their typical values and consistency.

SOLUTION

1. Compare centres using the means and spread using the ranges. A smaller range indicates less variation.

ANSWER: Group B has the higher mean and smaller range, so it is typically higher and more consistent

SELF-CHECK QUESTION

Which step carries the mathematical reason, and which step carries only arithmetic? Being able to distinguish them makes a solution easier to repair.

COMPREHENSIVE LESSON NOTEBOOK

Averages, spread and distributions - apply and reason

Connect the method to decisions and common errors.

Where this mathematics appears

performance and quality comparison rely on transparent data collection and proportionate conclusions.

Common misconceptions

- Generalising beyond the population represented by the sample.
- Using overlapping categories, distorted scales or an unsuitable average.
- Reporting a statistic without units, context, spread or uncertainty.

GUIDED TRY

Grouped intervals $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 30$ have frequencies [5, 7, 6]. Estimate the mean. First predict the form or approximate size of the answer. Then show the complete method.

REASONING PROMPT

The ordered data are [12, 13, 14, 15, 38]. Identify the outlier and explain whether the mean or median better represents a typical value. Explain why each condition or operation you use is valid.

EXAM COMMUNICATION

Use one mathematical statement per line. Label units and restrictions. If the result is approximate, show the unrounded value before the final rounded answer.

COMPREHENSIVE LESSON NOTEBOOK

Independent practice

Averages, spread and distributions

Complete without notes. Show all working, then use the answer section to diagnose and repair errors.

1. For the data [1, 3, 4, 6, 6], calculate the mean, median, mode and range.

[4]

2. Group A has mean 22 and range 14. Group B has mean 24 and range 8. Compare their typical values and consistency.

[4]

3. Grouped intervals $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 30$ have frequencies [5, 7, 6]. Estimate the mean.

[4]

4. The ordered data are [8, 9, 10, 11, 34]. Identify the outlier and explain whether the mean or median better represents a typical value.

[4]

5. For the data [5, 7, 8, 10, 10], calculate the mean, median, mode and range.

[4]

6. Group A has mean 26 and range 18. Group B has mean 28 and range 12. Compare their typical values and consistency.

[4]

7. Grouped intervals $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 30$ have frequencies [9, 11, 10]. Estimate the mean.

[4]

8. The ordered data are [12, 13, 14, 15, 38]. Identify the outlier and explain whether the mean or median better represents a typical value.

[4]

REFLECTION

Circle one: secure / almost / reteach. Record the question number of your most useful error and the rule that repairs it.

COMPREHENSIVE LESSON NOTEBOOK

Using statistics to describe populations

Use sample statistics appropriately to estimate and describe features of a population.

SPECIFICATION FOCUS

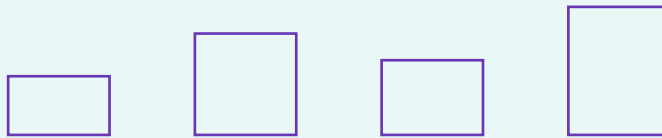
Specification reference Eduqas C300 FS5. This lesson develops representation, accurate calculation, reasoning and interpretation within Foundation Tier.

Understand the idea

A representative sample provides an evidence-based estimate of a population, but sampling variation remains.

KEY VOCABULARY

sample statistic • population parameter • estimate • uncertainty



Why the method works

- The method preserves the link between the observations and the conclusion.
- A suitable display reveals the relevant pattern without changing the underlying data.
- Matched summaries and explicit limitations make an inference auditable.

CHECK BEFORE MOVING ON

Answer mentally: A random sample of 60 packets has mean mass 21 g. State a suitable estimate for the population mean and explain one uncertainty. Then define one key word from this page in your own words.

COMPREHENSIVE LESSON NOTEBOOK

Using statistics to describe populations - method and examples

Follow the reasoning; do not copy steps mechanically.

Reliable method

- A sample statistic estimates a population feature; it is not the exact population value.
- Preserve units and round counts sensibly.
- Report sampling limitations alongside the numerical estimate.

WORKED EXAMPLE 1

A random sample of 60 packets has mean mass 21 g. State a suitable estimate for the population mean and explain one uncertainty.

SOLUTION

1. For a representative random sample, the sample mean is the best available estimate of the population mean, but sampling variation remains.

ANSWER: 21 g; another random sample could give a different mean

WORKED EXAMPLE 2

In a random sample of 100 components, 6 are defective. Estimate the number defective in a production run of 2200. Give one reason the actual number may differ.

SOLUTION

1. Estimated defective proportion = $6/100$. Multiply by 2200: $2200 \times 6/100 = 132.00$, so estimate about 132 components.

ANSWER: About 132; random sampling variation or changing production conditions

SELF-CHECK QUESTION

Which step carries the mathematical reason, and which step carries only arithmetic? Being able to distinguish them makes a solution easier to repair.

COMPREHENSIVE LESSON NOTEBOOK

Using statistics to describe populations - apply and reason

Connect the method to decisions and common errors.

Where this mathematics appears

manufacturing and population planning rely on transparent data collection and proportionate conclusions.

Common misconceptions

- Generalising beyond the population represented by the sample.
- Using overlapping categories, distorted scales or an unsuitable average.
- Reporting a statistic without units, context, spread or uncertainty.

GUIDED TRY

A report numbered 3 states: "The mean income in a sample is high, so most people have a high income." Explain why this conclusion may be misleading and name another statistic to inspect. First predict the form or approximate size of the answer. Then show the complete method.

REASONING PROMPT

In a random sample of 160 components, 12 are defective. Estimate the number defective in a production run of 2800. Give one reason the actual number may differ. Explain why each condition or operation you use is valid.

EXAM COMMUNICATION

Use one mathematical statement per line. Label units and restrictions. If the result is approximate, show the unrounded value before the final rounded answer.

COMPREHENSIVE LESSON NOTEBOOK

Independent practice

Using statistics to describe populations

Complete without notes. Show all working, then use the answer section to diagnose and repair errors.

1. A random sample of 60 packets has mean mass 21 g. State a suitable estimate for the population mean and explain one uncertainty. [3]

2. In a random sample of 100 components, 6 are defective. Estimate the number defective in a production run of 2200. Give one reason the actual number may differ. [4]

3. A report numbered 3 states: "The mean income in a sample is high, so most people have a high income." Explain why this conclusion may be misleading and name another statistic to inspect. [3]

4. A random sample of 90 packets has mean mass 24 g. State a suitable estimate for the population mean and explain one uncertainty. [3]

5. In a random sample of 130 components, 9 are defective. Estimate the number defective in a production run of 2500. Give one reason the actual number may differ. [4]

6. A report numbered 6 states: "The mean income in a sample is high, so most people have a high income." Explain why this conclusion may be misleading and name another statistic to inspect. [3]

7. A random sample of 120 packets has mean mass 27 g. State a suitable estimate for the population mean and explain one uncertainty. [3]

8. In a random sample of 160 components, 12 are defective. Estimate the number defective in a production run of 2800. Give one reason the actual number may differ. [4]

REFLECTION

Circle one: secure / almost / reteach. Record the question number of your most useful error and the rule that repairs it.

COMPREHENSIVE LESSON NOTEBOOK

Scatter graphs, correlation and prediction

Interpret scatter graphs, correlation and lines of best fit, including interpolation, extrapolation and causation.

SPECIFICATION FOCUS

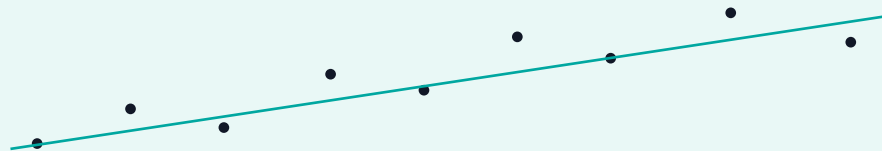
Specification reference Eduqas C300 FS6. This lesson develops representation, accurate calculation, reasoning and interpretation within Foundation Tier.

Understand the idea

Scatter graphs describe association between paired variables and support cautious prediction without establishing cause.

KEY VOCABULARY

bivariate • correlation • causation • line of best fit • interpolation • extrapolation



Why the method works

- The method preserves the link between the observations and the conclusion.
- A suitable display reveals the relevant pattern without changing the underlying data.
- Matched summaries and explicit limitations make an inference auditable.

CHECK BEFORE MOVING ON

Answer mentally: Scatter graph 1 shows points rising closely from left to right. Describe the correlation and explain what it does not prove. Then define one key word from this page in your own words.

COMPREHENSIVE LESSON NOTEBOOK

Scatter graphs, correlation and prediction - method and examples

Follow the reasoning; do not copy steps mechanically.

Reliable method

- Describe correlation by direction and, where justified, strength.
- A best-fit line should balance the point cloud and is safest for interpolation.
- Correlation does not prove causation; extrapolation assumes the pattern continues.

WORKED EXAMPLE 1

Scatter graph 1 shows points rising closely from left to right. Describe the correlation and explain what it does not prove.

SOLUTION

1. An upward pattern indicates positive correlation. Correlation alone cannot establish causation because other factors may affect both variables.

ANSWER: Strong positive correlation; it does not prove that one variable causes the other

WORKED EXAMPLE 2

A line of best fit is $y=4x+7$. Estimate y when $x=8$ and state whether this is interpolation if observed x -values range from 2 to 12.

SOLUTION

1. Substitute $x=8$: $y=4 \times 8 + 7 = 39$. Since 8 lies inside the observed range, this is interpolation.

ANSWER: 39; interpolation

SELF-CHECK QUESTION

Which step carries the mathematical reason, and which step carries only arithmetic? Being able to distinguish them makes a solution easier to repair.

COMPREHENSIVE LESSON NOTEBOOK

Scatter graphs, correlation and prediction - apply and reason

Connect the method to decisions and common errors.

Where this mathematics appears

science, economics and forecasting rely on transparent data collection and proportionate conclusions.

Common misconceptions

- Generalising beyond the population represented by the sample.
- Using overlapping categories, distorted scales or an unsuitable average.
- Reporting a statistic without units, context, spread or uncertainty.

GUIDED TRY

Data on a scatter graph cover x-values from 1 to 13. A line of best fit is used to predict at $x=21$. Name this process and explain why the prediction is less reliable. First predict the form or approximate size of the answer. Then show the complete method.

REASONING PROMPT

A scatter graph numbered 8 shows almost no upward or downward pattern. State the correlation and explain why drawing a prediction line would be of little value. Explain why each condition or operation you use is valid.

EXAM COMMUNICATION

Use one mathematical statement per line. Label units and restrictions. If the result is approximate, show the unrounded value before the final rounded answer.

COMPREHENSIVE LESSON NOTEBOOK

Independent practice

Scatter graphs, correlation and prediction

Complete without notes. Show all working, then use the answer section to diagnose and repair errors.

1. Scatter graph 1 shows points rising closely from left to right. Describe the correlation and explain what it does not prove. [3]

2. A line of best fit is $y=4x+7$. Estimate y when $x=8$ and state whether this is interpolation if observed x -values range from 2 to 12. [3]

3. Data on a scatter graph cover x -values from 1 to 13. A line of best fit is used to predict at $x=21$. Name this process and explain why the prediction is less reliable. [3]

4. A scatter graph numbered 4 shows almost no upward or downward pattern. State the correlation and explain why drawing a prediction line would be of little value. [3]

5. Scatter graph 5 shows points rising closely from left to right. Describe the correlation and explain what it does not prove. [3]

6. A line of best fit is $y=2x+11$. Estimate y when $x=12$ and state whether this is interpolation if observed x -values range from 2 to 16. [3]

7. Data on a scatter graph cover x -values from 1 to 17. A line of best fit is used to predict at $x=25$. Name this process and explain why the prediction is less reliable. [3]

8. A scatter graph numbered 8 shows almost no upward or downward pattern. State the correlation and explain why drawing a prediction line would be of little value. [3]

REFLECTION

Circle one: secure / almost / reteach. Record the question number of your most useful error and the rule that repairs it.

COMPREHENSIVE LESSON NOTEBOOK

Practice answers and repair notes

Sampling and populations

Use only after a genuine attempt. Rework incorrect questions from a blank line rather than copying the final value.

1. An unbiased sample of 50 people contains 22 who prefer option A. Estimate how many of a ...

ANSWER: About 264; a sample may not represent the population exactly

Method cue: Sample proportion= $\frac{22}{50}$. Estimate= $600 \times \frac{22}{50}=264.00$, which is about 264 people. Sampling variation means the true population value may differ. Check sign, size, accuracy and units.

2. A school wants pupils' views on homework. Survey 2 asks only pupils attending an after-s...

ANSWER: The sample is biased; select pupils randomly from the whole school

Method cue: Homework-club pupils may have different views from other pupils. A random or stratified sample from the full school population is more representative. Check sign, size, accuracy and units.

3. Two random samples from the same population contain 35 and 150 people. Which sample is g...

ANSWER: The sample of 150; it can still be unrepresentative by chance or through non-response

Method cue: A larger random sample usually reduces sampling variation, but size alone does not guarantee perfect representation. Check sign, size, accuracy and units.

4. An unbiased sample of 80 people contains 34 who prefer option A. Estimate how many of a ...

ANSWER: About 383; a sample may not represent the population exactly

Method cue: Sample proportion= $\frac{34}{80}$. Estimate= $900 \times \frac{34}{80}=382.50$, which is about 383 people. Sampling variation means the true population value may differ. Check sign, size, accuracy and units.

5. A school wants pupils' views on homework. Survey 5 asks only pupils attending an after-s...

ANSWER: The sample is biased; select pupils randomly from the whole school

Method cue: Homework-club pupils may have different views from other pupils. A random or stratified sample from the full school population is more representative. Check sign, size, accuracy and units.

6. Two random samples from the same population contain 50 and 180 people. Which sample is g...

ANSWER: The sample of 180; it can still be unrepresentative by chance or through non-response

Method cue: A larger random sample usually reduces sampling variation, but size alone does not guarantee perfect representation. Check sign, size, accuracy and units.

7. An unbiased sample of 110 people contains 46 who prefer option A. Estimate how many of a...

ANSWER: About 502; a sample may not represent the population exactly

Method cue: Sample proportion= $\frac{46}{110}$. Estimate= $1200 \times \frac{46}{110}=501.82$, which is about 502 people. Sampling variation means the true population value may differ. Check sign, size, accuracy and units.

8. A school wants pupils' views on homework. Survey 8 asks only pupils attending an after-s...

ANSWER: The sample is biased; select pupils randomly from the whole school

Method cue: Homework-club pupils may have different views from other pupils. A random or stratified sample from the full school population is more representative. Check sign, size, accuracy and units.

COMPREHENSIVE LESSON NOTEBOOK

Practice answers and repair notes

Questionnaires and fairness

Use only after a genuine attempt. Rework incorrect questions from a blank line rather than copying the final value.

1. Questionnaire item 1 asks: "Don't you agree that our excellent new library should stay o..."

ANSWER: It is leading and uses emotive wording; e.g. "Should the library stay open later? Yes / No / Not sure"

Method cue: Remove language that encourages one answer and provide balanced, exhaustive response options. Check sign, size, accuracy and units.

2. A questionnaire asks weekly study time using: 0-3 hours, 3-5 hours, 5-7 hours. Identify ...

ANSWER: The endpoints overlap; e.g. $0 \leq t < 3$, $3 \leq t < 5$, $5 \leq t < 7$, $t \geq 7$

Method cue: A value at a shared endpoint could fit two boxes. Use one included and one excluded boundary, then add $t \geq 7$ so the categories are exhaustive. Check sign, size, accuracy and units.

3. A survey numbered 3 asks "How often do you exercise?" with options Never, Sometimes and ...

ANSWER: Specify a time period and use numerical, exhaustive, non-overlapping response intervals

Method cue: Words such as "sometimes" mean different things to different people. A period such as "in the last 7 days" and numerical intervals produce clearer data. Check sign, size, accuracy and units.

4. Questionnaire item 4 asks: "Don't you agree that our excellent new library should stay o..."

ANSWER: It is leading and uses emotive wording; e.g. "Should the library stay open later? Yes / No / Not sure"

Method cue: Remove language that encourages one answer and provide balanced, exhaustive response options. Check sign, size, accuracy and units.

5. A questionnaire asks weekly study time using: 0-6 hours, 6-8 hours, 8-10 hours. Identify...

ANSWER: The endpoints overlap; e.g. $0 \leq t < 6$, $6 \leq t < 8$, $8 \leq t < 10$, $t \geq 10$

Method cue: A value at a shared endpoint could fit two boxes. Use one included and one excluded boundary, then add $t \geq 10$ so the categories are exhaustive. Check sign, size, accuracy and units.

6. A survey numbered 6 asks "How often do you exercise?" with options Never, Sometimes and ...

ANSWER: Specify a time period and use numerical, exhaustive, non-overlapping response intervals

Method cue: Words such as "sometimes" mean different things to different people. A period such as "in the last 7 days" and numerical intervals produce clearer data. Check sign, size, accuracy and units.

7. Questionnaire item 7 asks: "Don't you agree that our excellent new library should stay o..."

ANSWER: It is leading and uses emotive wording; e.g. "Should the library stay open later? Yes / No / Not sure"

Method cue: Remove language that encourages one answer and provide balanced, exhaustive response options. Check sign, size, accuracy and units.

8. A questionnaire asks weekly study time using: 0-9 hours, 9-11 hours, 11-13 hours. Identify...

ANSWER: The endpoints overlap; e.g. $0 \leq t < 9$, $9 \leq t < 11$, $11 \leq t < 13$, $t \geq 13$

Method cue: A value at a shared endpoint could fit two boxes. Use one included and one excluded boundary, then add $t \geq 13$ so the categories are exhaustive. Check sign, size, accuracy and units.

COMPREHENSIVE LESSON NOTEBOOK

Practice answers and repair notes

Tables, charts and time series

Use only after a genuine attempt. Rework incorrect questions from a blank line rather than copying the final value.

1. A survey records four categories with frequencies 4, 6, 8 and 6. Find the pie-chart angl...

ANSWER: 120 degrees

Method cue: Angle=frequency/total x 360=8/24 x 360=120 degrees. Check sign, size, accuracy and units.

2. On a pictogram, one complete symbol represents 4 people. A category has 5 complete symbo...

ANSWER: 22

Method cue: Frequency=(5+1/2) x 4=22. Check sign, size, accuracy and units.

3. A time series has consecutive values [15, 18, 14, 21, 23]. State the overall change from...

ANSWER: Increase of 8; interval 3 to 4

Method cue: Overall change=23-15=8. Consecutive changes are 3, -4, 7 and 2; the largest rise is from 3 to 4. Check sign, size, accuracy and units.

4. Scores 1, 2, 3, 4 and 5 have frequencies [5, 7, 8, 6, 4]. State an appropriate chart and...

ANSWER: Vertical line chart; 30

Method cue: The scores are ungrouped discrete numerical data, so a vertical line chart is suitable. Add the frequencies for the total. Check sign, size, accuracy and units.

5. A survey records four categories with frequencies 8, 10, 12 and 10. Find the pie-chart a...

ANSWER: 108 degrees

Method cue: Angle=frequency/total x 360=12/40 x 360=108 degrees. Check sign, size, accuracy and units.

6. On a pictogram, one complete symbol represents 4 people. A category has 9 complete symbo...

ANSWER: 38

Method cue: Frequency=(9+1/2) x 4=38. Check sign, size, accuracy and units.

7. A time series has consecutive values [19, 22, 18, 25, 27]. State the overall change from...

ANSWER: Increase of 8; interval 3 to 4

Method cue: Overall change=27-19=8. Consecutive changes are 3, -4, 7 and 2; the largest rise is from 3 to 4. Check sign, size, accuracy and units.

8. Scores 1, 2, 3, 4 and 5 have frequencies [9, 11, 12, 10, 8]. State an appropriate chart ...

ANSWER: Vertical line chart; 50

Method cue: The scores are ungrouped discrete numerical data, so a vertical line chart is suitable. Add the frequencies for the total. Check sign, size, accuracy and units.

COMPREHENSIVE LESSON NOTEBOOK

Practice answers and repair notes

Averages, spread and distributions

Use only after a genuine attempt. Rework incorrect questions from a blank line rather than copying the final value.

1. For the data [1, 3, 4, 6, 6], calculate the mean, median, mode and range.

ANSWER: Mean 4; median 4; mode 6; range 5

Method cue: Sum=20, so mean=20/5=4. The middle value is 4, the repeated value is 6, and range=(6)-1=5. Check sign, size, accuracy and units.

2. Group A has mean 22 and range 14. Group B has mean 24 and range 8. Compare their typical...

ANSWER: Group B has the higher mean and smaller range, so it is typically higher and more consistent

Method cue: Compare centres using the means and spread using the ranges. A smaller range indicates less variation. Check sign, size, accuracy and units.

3. Grouped intervals $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 30$ have frequencies [5, 7, 6]. Estimate th...

ANSWER: 15.56

Method cue: Use midpoints 5, 15 and 25. Estimated mean=(5x5+7x15+6x25)/18=15.56. Check sign, size, accuracy and units.

4. The ordered data are [8, 9, 10, 11, 34]. Identify the outlier and explain whether the me...

ANSWER: Outlier 34; median 10 is more representative

Method cue: The value 34 is far from the cluster. It pulls the mean to 14.4, while the median remains 10, within the main cluster. Check sign, size, accuracy and units.

5. For the data [5, 7, 8, 10, 10], calculate the mean, median, mode and range.

ANSWER: Mean 8; median 8; mode 10; range 5

Method cue: Sum=40, so mean=40/5=8. The middle value is 8, the repeated value is 10, and range=(10)-5=5. Check sign, size, accuracy and units.

6. Group A has mean 26 and range 18. Group B has mean 28 and range 12. Compare their typica...

ANSWER: Group B has the higher mean and smaller range, so it is typically higher and more consistent

Method cue: Compare centres using the means and spread using the ranges. A smaller range indicates less variation. Check sign, size, accuracy and units.

7. Grouped intervals $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 30$ have frequencies [9, 11, 10]. Estimate ...

ANSWER: 15.33

Method cue: Use midpoints 5, 15 and 25. Estimated mean=(9x5+11x15+10x25)/30=15.33. Check sign, size, accuracy and units.

8. The ordered data are [12, 13, 14, 15, 38]. Identify the outlier and explain whether the ...

ANSWER: Outlier 38; median 14 is more representative

Method cue: The value 38 is far from the cluster. It pulls the mean to 18.4, while the median remains 14, within the main cluster. Check sign, size, accuracy and units.

COMPREHENSIVE LESSON NOTEBOOK

Practice answers and repair notes

Using statistics to describe populations

Use only after a genuine attempt. Rework incorrect questions from a blank line rather than copying the final value.

1. A random sample of 60 packets has mean mass 21 g. State a suitable estimate for the popu...

ANSWER: 21 g; another random sample could give a different mean

Method cue: For a representative random sample, the sample mean is the best available estimate of the population mean, but sampling variation remains. Check sign, size, accuracy and units.

2. In a random sample of 100 components, 6 are defective. Estimate the number defective in ...

ANSWER: About 132; random sampling variation or changing production conditions

Method cue: Estimated defective proportion = $6/100$. Multiply by 2200: $2200 \times 6/100 = 132.00$, so estimate about 132 components. Check sign, size, accuracy and units.

3. A report numbered 3 states: "The mean income in a sample is high, so most people have a ...

ANSWER: A few very high values may raise the mean; inspect the median and the spread

Method cue: The mean is sensitive to extreme values and does not reveal how values are distributed. Median, range and the graph add context. Check sign, size, accuracy and units.

4. A random sample of 90 packets has mean mass 24 g. State a suitable estimate for the popu...

ANSWER: 24 g; another random sample could give a different mean

Method cue: For a representative random sample, the sample mean is the best available estimate of the population mean, but sampling variation remains. Check sign, size, accuracy and units.

5. In a random sample of 130 components, 9 are defective. Estimate the number defective in ...

ANSWER: About 173; random sampling variation or changing production conditions

Method cue: Estimated defective proportion = $9/130$. Multiply by 2500: $2500 \times 9/130 = 173.08$, so estimate about 173 components. Check sign, size, accuracy and units.

6. A report numbered 6 states: "The mean income in a sample is high, so most people have a ...

ANSWER: A few very high values may raise the mean; inspect the median and the spread

Method cue: The mean is sensitive to extreme values and does not reveal how values are distributed. Median, range and the graph add context. Check sign, size, accuracy and units.

7. A random sample of 120 packets has mean mass 27 g. State a suitable estimate for the popu...

ANSWER: 27 g; another random sample could give a different mean

Method cue: For a representative random sample, the sample mean is the best available estimate of the population mean, but sampling variation remains. Check sign, size, accuracy and units.

8. In a random sample of 160 components, 12 are defective. Estimate the number defective in...

ANSWER: About 210; random sampling variation or changing production conditions

Method cue: Estimated defective proportion = $12/160$. Multiply by 2800: $2800 \times 12/160 = 210.00$, so estimate about 210 components. Check sign, size, accuracy and units.

COMPREHENSIVE LESSON NOTEBOOK

Practice answers and repair notes

Scatter graphs, correlation and prediction

Use only after a genuine attempt. Rework incorrect questions from a blank line rather than copying the final value.

1. Scatter graph 1 shows points rising closely from left to right. Describe the correlation...

ANSWER: Strong positive correlation; it does not prove that one variable causes the other

Method cue: An upward pattern indicates positive correlation. Correlation alone cannot establish causation because other factors may affect both variables. Check sign, size, accuracy and units.

2. A line of best fit is $y=4x+7$. Estimate y when $x=8$ and state whether this is interpolatio...

ANSWER: 39; interpolation

Method cue: Substitute $x=8$: $y=4 \times 8 + 7 = 39$. Since 8 lies inside the observed range, this is interpolation. Check sign, size, accuracy and units.

3. Data on a scatter graph cover x -values from 1 to 13. A line of best fit is used to predi...

ANSWER: Extrapolation; the pattern may not continue beyond the observed data

Method cue: The prediction is outside the measured x -range, so it extends an assumed trend without supporting observations. Check sign, size, accuracy and units.

4. A scatter graph numbered 4 shows almost no upward or downward pattern. State the correla...

ANSWER: No correlation; one variable gives little useful information for predicting the other

Method cue: With no clear relationship, a line would not represent a dependable trend and predictions would have large uncertainty. Check sign, size, accuracy and units.

5. Scatter graph 5 shows points rising closely from left to right. Describe the correlation...

ANSWER: Strong positive correlation; it does not prove that one variable causes the other

Method cue: An upward pattern indicates positive correlation. Correlation alone cannot establish causation because other factors may affect both variables. Check sign, size, accuracy and units.

6. A line of best fit is $y=2x+11$. Estimate y when $x=12$ and state whether this is interpolat...

ANSWER: 35; interpolation

Method cue: Substitute $x=12$: $y=2 \times 12 + 11 = 35$. Since 12 lies inside the observed range, this is interpolation. Check sign, size, accuracy and units.

7. Data on a scatter graph cover x -values from 1 to 17. A line of best fit is used to predi...

ANSWER: Extrapolation; the pattern may not continue beyond the observed data

Method cue: The prediction is outside the measured x -range, so it extends an assumed trend without supporting observations. Check sign, size, accuracy and units.

8. A scatter graph numbered 8 shows almost no upward or downward pattern. State the correla...

ANSWER: No correlation; one variable gives little useful information for predicting the other

Method cue: With no clear relationship, a line would not represent a dependable trend and predictions would have large uncertainty. Check sign, size, accuracy and units.

COMPREHENSIVE LESSON NOTEBOOK

Independent authorship and release check

Eduqas Foundation Mathematics • Unit 6 Statistics

ORIGINAL-CONTENT STATEMENT

This publication contains independently written teaching explanations, examples, practice questions, numerical data, diagrams and solutions. Eduqas is referenced solely to identify the C300 qualification. No Eduqas or WJEC logo, official assessment question, mark scheme, examiner-report text or protected assessment layout is reproduced.

Release checklist

- 6 publisher-organised chapters mapped to Eduqas C300 references FS1-FS6.
- Every chapter begins on a new page and contains conceptual explanation.
- Original vector diagrams and two worked examples appear in every chapter.
- Applications, misconceptions, guided reasoning and independent practice are included.
- Answers are collected at the end with method cues.
- Copyright, trademark and non-endorsement wording included.

USE OF OFFICIAL MATERIALS

The current Eduqas specification, qualification guidance and public assessment materials informed curriculum scope and general assessment characteristics only. This independent publication is not an official Eduqas paper, mark scheme or endorsed resource. Consult Eduqas for the current specification, assessment materials and formula-sheet arrangements.