

4.1 Data summaries, histograms and interpretation

Name: _____ Class: _____ Date: _____ **Total: 13 marks**

KEY WORDS frequency density

Objective

Data summaries, histograms and interpretation. Supported HL concept preparation; original classroom practice, not an official IB paper or assessed learner work.

1 Key knowledge and worked examples

■ Can one average tell the whole story?

- Two groups have the same median but different spread. One summary cannot describe both location and consistency.
- This lesson studies **frequency density** : Frequency divided by class width, used as histogram height.

■ Choose the mathematical structure

- Compare an appropriate average and spread in context. A histogram uses area for frequency, so height=frequency/class width. Grouped estimates assume representative values within intervals.
- State the allowed inputs and units before calculating. An equation should express the relationship, not just record a calculator entry.

■ Work through a checked case

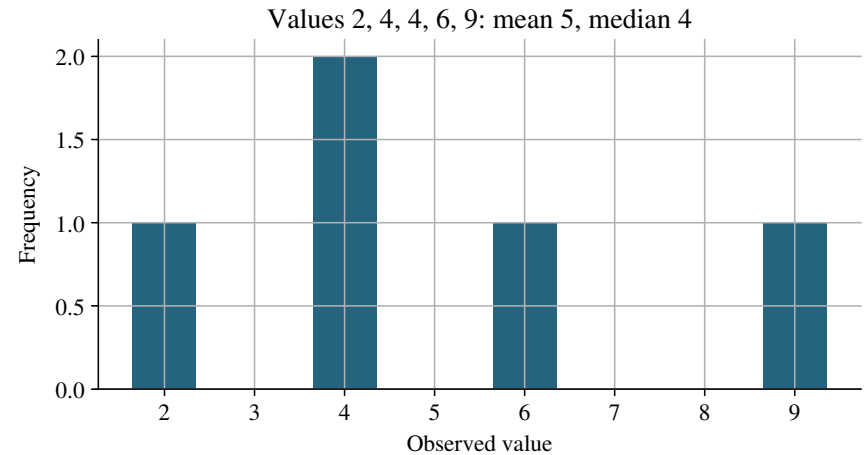
$$\bar{x} = \frac{\sum x_i}{n}, \quad \text{density} = \frac{\text{frequency}}{\text{class width}}$$

- Check the result against the starting quantities. Substitute into the original relation, or compare the graph and numerical answer where appropriate.

Example:

A class from 10 to 20 with frequency 30 has density 30/10=3. A class from 20 to 40 with frequency 20 has density 20/20=1. Its wider bar must not be mistaken for a larger

density. For values 2,4,4,6,9, the median is 4 and mean is 5.



■ Test a tempting shortcut

- The tallest histogram bar need not contain the most observations. A grouped mean is an estimate. Correlation does not prove causation, and extrapolation extends beyond the observed range.
- When a shortcut fails, identify the assumption it breaks. Keep an exact value until the requested final rounding.

Warn:

A histogram bar's height always equals its frequency, even with unequal class widths. This claim is false. Explain which definition or assumption it violates.

■ Interpret a new situation

- Choose a display that fits the data type. Give both a numerical comparison and what it means for the population; do not infer more precision than the sample supports.
- A complete solution gives the mathematical result and explains what it means. Check that it is possible in the stated context.

■ Use this in your course

- Current first-assessment-2021 Analysis and Approaches HL. This is authored concept support; the full guide is needed to certify every objective.
- Give the method before the final answer, and use the paper's calculator and formula rules. Review a wrong answer by locating the first invalid step.

Key:

Frequency divided by class width, used as histogram height. Choose the relationship, show the method, check its assumptions and interpret the result.

2 Practice

2.1 Find the density for frequency 30 and class width 10. Show the relation, substitution and units. [2]

2.2 Find the mean of 2,4,4,6,9. Show the relation, substitution and units. [2]

2.3 Find the median of 2,4,4,6,9. Show the relation, substitution and units. [2]

2.4 Decide whether this claim is true and justify: A histogram bar’s height always equals its frequency, even with unequal class widths. [2]

2.5 Which description correctly defines frequency density? Choose one answer.

- A Frequency divided by class width, used as histogram height.
- B A numerical answer that needs no assumptions or interpretation
- C A rule that remains valid even when its required conditions fail

[1]

3 Exam-style questions

3.1 Classes 0–10 and 10–30 have frequencies 20 and 30. Find histogram densities and compare bar heights with frequencies; can the exact mean be recovered? [4]

Solutions

2.1 Density=frequency/class width= $30/10=3$.

2.2 The total is 25 across five observations: mean= $25/5=5$.

2.3 The sorted list is 2,4,4,6,9. Its third observation is 4.

2.4 False. The tallest histogram bar need not contain the most observations. A grouped mean is an estimate. Correlation does not prove causation, and extrapolation extends beyond the observed range.

2.5 A Frequency divided by class width, used as histogram height.

3.1 Densities 2 and 1.5, so the higher bar has fewer observations. Areas represent frequency. Only a grouped midpoint estimate, not the exact mean, follows without raw values; state the endpoint convention.