4.3 HL: data analysis and interpretation

Name: _____ Class: _____ Date: _____ **Total: 30 marks**

KEY WORDS mean • median • standard deviation • p-value p • statistically significant • correlation coefficient • content analysis

Objective

Build the skills to describe and interpret data for HL Paper 3: averages and spread, **standard deviation**, graphs, **statistical significance** and p-values, correlation coefficients, and thematic and content analysis.

You must be able to:

- calculate and choose the mean, median, mode and range
- judge a difference between means against the standard deviations
- explain p-values, significance and the difference between significance and size
- describe a correlation from r, and choose a suitable graph
- distinguish thematic analysis and content analysis

1 Key knowledge and worked examples

■ Describing numbers

measure	meaning	use
mean	total ÷ number of scores	most data; pulled by extreme scores
median	middle score	data with extreme scores
mode	most common score	categories
range	highest − lowest	quick idea of spread
standard deviation	typical distance of scores from the mean	how much scores spread

A difference between two means matters only if it is **large compared with the standard deviation**.

exercises.lol exercises.com.cn

- **Inferential test → p-value**: the chance of a difference this big if the null hypothesis were true. **Significant** usually means $p < 0.05$. A significant result can still be a **small** effect; a result from a small sample may not repeat.
- **Correlation coefficient** r: from -1 to $+1$. Near 0: no linear relationship; about ± 0.5: moderate; ± 1: perfect. Negative: one goes down as the other goes up.
- **Graphs**: bar chart for means of conditions; histogram for a distribution; scatter graph for a correlation.

Describing words

Thematic analysis: find themes in interview transcripts. **Content analysis**: count how often coded categories appear, turning words into numbers. Two coders check each other.

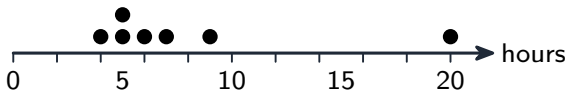
Paper 3 answer shapes: Q1 (3 marks) one issue and its effect on the conclusion; Q2 (6) describe the pattern with numbers → limits → conclusion linked to the claim; Q3 (6) specific improvements to credibility; Q4 (15) weigh all sources against the claim.

2 Practice

2.1 The number of hours seven students slept before an exam were: 4, 5, 5, 6, 7, 9, 20 (one student recorded a whole weekend by mistake).

(a) Calculate the mean, median and mode. [3]

(b) State which average best represents these data, and why. [1]

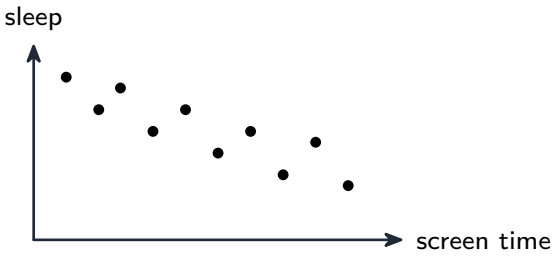


2.2 Two groups had mean test scores of 12.4 and 13.0, with standard deviations of 4.8 and 5.1. Explain why the difference may not be meaningful. [2]



exercises.lol exercises.com.cn

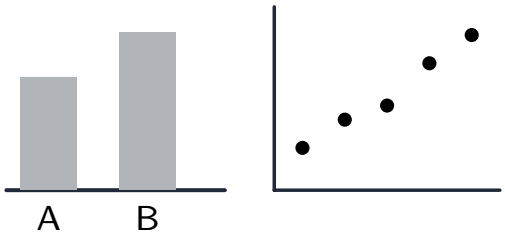
2.3 A study found $r = -0.56$ between daily screen time and hours of sleep. Describe this correlation. [2]



2.4 A study reports a difference between two groups with $p = 0.03$. Explain what this means, and **one** reason to be careful about the result. [3]

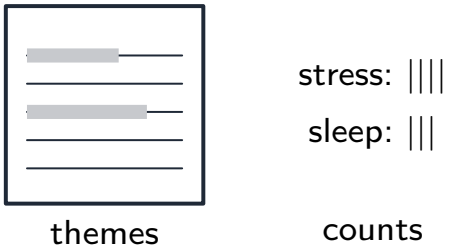
$$p = 0.03 < 0.05$$

2.5 Identify a suitable graph for: (a) the mean recall in two conditions; (b) the relationship between hours of sleep and test score. [2]



exercises.lol exercises.com.cn

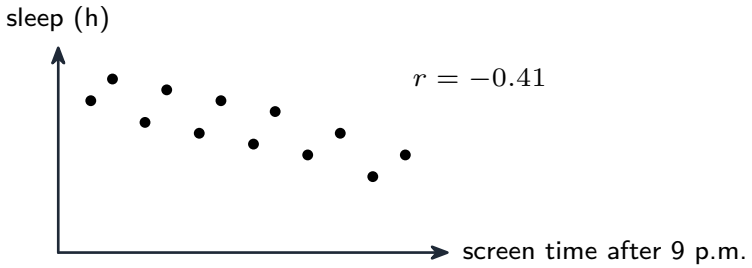
2.6 Distinguish between **thematic analysis** and **content analysis**. [2]



3 Exam-style questions

3.1 *(Paper 3 style.)* Claim: **"Using screens late at night harms students' sleep."**

Source A (fictional data for practice). A researcher asked 60 Year 12 students in one school to record their screen time after 9 p.m. and their hours of sleep for one week, in a diary. The correlation between mean screen time after 9 p.m. and mean sleep was $r = -0.41$. Sleep ranged from about 6 to 8.5 hours.



(a) Explain **one** issue that limits the interpretation of the data in Source A. [3]

exercises.lol exercises.com.cn

(b) Analyse the findings in Source A and state a conclusion linked to the claim. [6]

(c) Discuss how the researcher could improve the credibility of the findings. [6]

exercises.lol exercises.com.cn