

4.5 Confidence intervals, chi-square and regression

Name: _____ Class: _____ Date: _____

Total: 13 marks

KEY WORDS confidence interval

Objective

Confidence intervals, chi-square and regression. Supported HL concept preparation; original classroom practice, not an official IB paper or assessed learner work.

1 Key knowledge and worked examples

■ How uncertain is the sample mean?

- Two samples give different means. We need a statement about sampling uncertainty, not a promise that one sample equals the population.
 - This lesson studies **confidence interval** : An interval calculated by a procedure designed to cover the population parameter at a stated long-run rate.
-

■ Choose the mathematical structure

- For a normal mean with known σ , a 95% interval is sample mean $\pm 1.96\sigma/\sqrt{n}$. For chi-square, sum (observed-expected)²/expected with the correct degrees of freedom. Choose a test that fits the data and assumptions.
 - 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

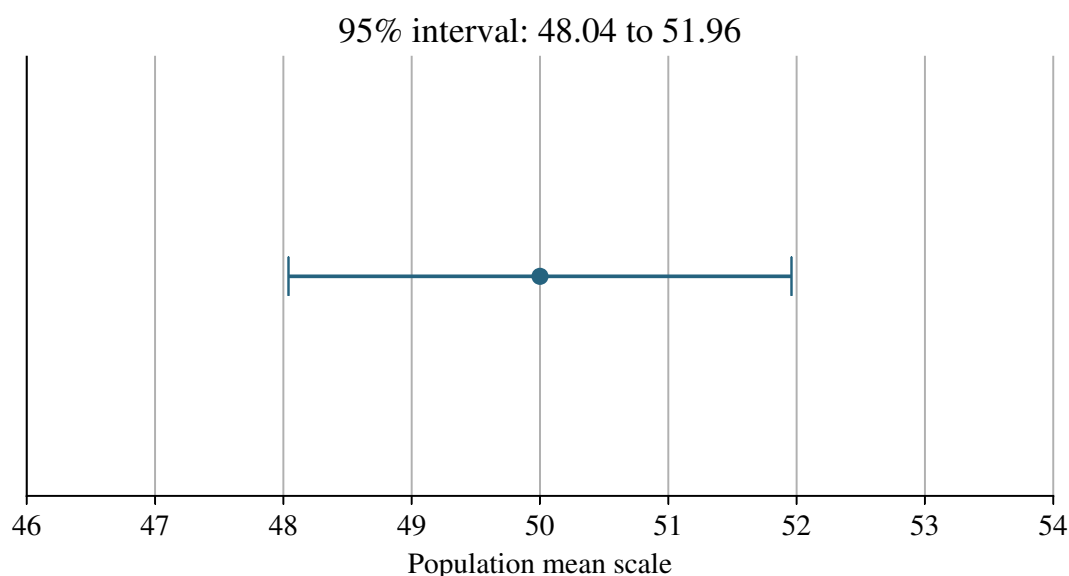
$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}, \quad \chi^2 = \sum \frac{(O - E)^2}{E}$$

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

Example:

With mean=50, $\sigma=10$, $n=100$, standard error=1 and the 95% interval is [48.04, 51.96]. For observed counts 30 and 20 against expected 25 and 25, chi-square=25/25+25/25=2,

before deciding degrees of freedom and the rejection threshold.



■ Test a tempting shortcut

- A 95% confidence procedure does not give a 95% probability that a fixed population mean moves within this one interval. Expected counts and estimated parameters affect the validity and degrees of freedom of chi-square tests.
- When a shortcut fails, identify the assumption it breaks. Keep an exact value until the requested final rounding.

Warn:

A strong correlation alone establishes a causal relationship. This claim is false. Explain which definition or assumption it violates.

■ Interpret a new situation

- Report uncertainty in the context. In regression, inspect residuals, avoid unjustified extrapolation, and distinguish a fitted relationship from a causal explanation. S3 and IB AI HL require different verified scope boundaries.
 - 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 Applications and Interpretation 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:

An interval calculated by a procedure designed to cover the population parameter at a stated long-run rate. Choose the relationship, show the method, check its assumptions and interpret the result.

2 Practice

2.1 Find the standard error when $\sigma=10$ and $n=100$. Show the relation, substitution and units. [2]

2.2 Find the lower 95% confidence bound when $\text{mean}=50, \sigma=10, n=100$. Show the relation, substitution and units. [2]

2.3 Find chi-square for observed 30,20 and expected 25,25. Show the relation, substitution and units. [2]

2.4 Decide whether this claim is true and justify: A strong correlation alone establishes a causal relationship. [2]

2.5 Which description correctly defines confidence interval? Choose one answer.

- **A** An interval calculated by a procedure designed to cover the population parameter at a stated long-run rate.
 - **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 A known- σ normal model has sample mean 50, $\sigma 10$, $n 100$. Give the 95% interval, then the interval if n quadruples with mean and σ unchanged. Explain the repeated-sampling interpretation. [4]

Solutions

2.1 Standard error= $\sigma/\sqrt{n}=10/10=1$.

2.2 Lower endpoint= $50-1.96\times 1=48.04$ under the stated known- σ normal model.

2.3 Chi-square= $(30-25)^2/25+(20-25)^2/25=1+1=2$.

2.4 False. A 95% confidence procedure does not give a 95% probability that a fixed population mean moves within this one interval. Expected counts and estimated parameters affect the validity and degrees of freedom of chi-square tests.

2.5 A An interval calculated by a procedure designed to cover the population parameter at a stated long-run rate.

3.1 SE 1:48.04 to 51.96. At $n=400$, SE 0.5:49.02 to 50.98. The procedure covers the fixed population mean in 95% of repeated samples under assumptions; it does not assign 95% probability to the fixed mean after observing this interval.