# 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. Original classroom practice for IB HL; not an official IB paper.

## 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.

---

### The method

- For a normal mean with known $\sigma$, 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.

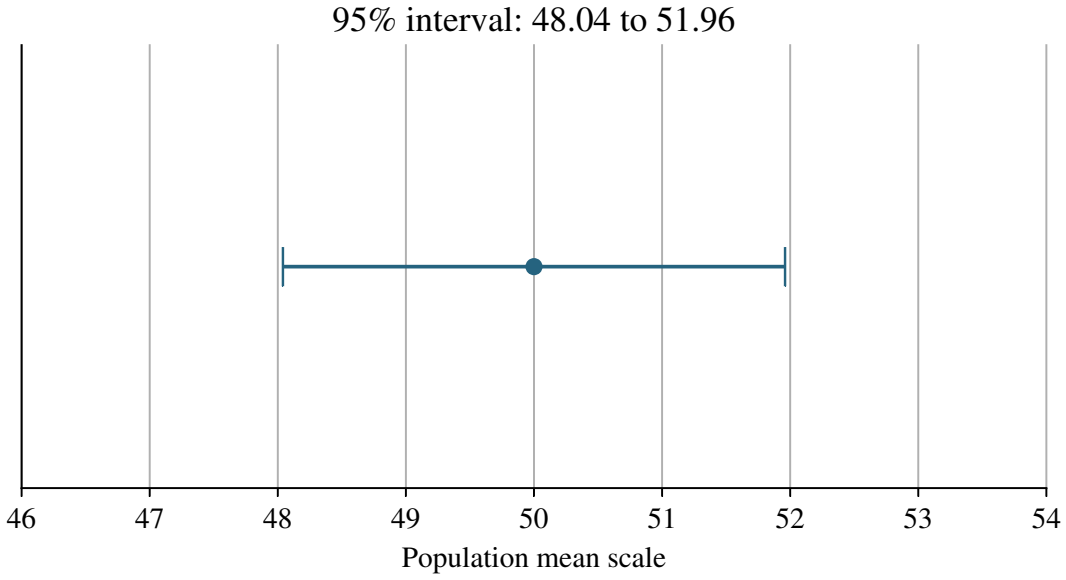
---

### A worked case

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

**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.




---

## Where it goes wrong

- 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.

**Not true:** A strong correlation alone establishes a causal relationship. Correlation shows association; a third variable or chance can explain it, so causation needs an experiment or other evidence.

---

## Applying it

- 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.

---

## In your course

- Applications and Interpretation HL, the course first assessed in 2021.
- Every paper allows a calculator, but method marks still need the expression or equation written before its value.

**Key:**

- **Confidence interval**: An interval calculated by a procedure designed to cover the population parameter at a stated long-run rate.
- For a normal mean with known $\sigma$, a 95% interval is sample mean$\pm 1.96\sigma/\sqrt{n}$.

- Watch for: A 95% confidence procedure does not give a 95% probability that a fixed population mean moves within this one interval.

■ **Write from a chi-square value of 6**

- Observed counts are 18 and 6. Expected counts are 12 and 12.
- Find the chi-square statistic. A strong correlation does not prove a cause.

## 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 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** What does a 95% confidence interval procedure guarantee? Choose one answer.

- **A** That 95% of the data lie inside the interval
- **B** 95% of intervals built this way cover the true parameter in the long run
- **C** A 95% probability that THIS interval holds the parameter [1]

## 3 Exam-style questions

**3.1** A known-$\sigma$ normal model has sample mean 50, $\sigma$10, n 100. Give the 95% interval, then the interval if n quadruples with mean and$\sigma$unchanged. Explain the repeated-sampling interpretation. [4]

## Solutions

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

**2.2** Lower endpoint=50-1.96×1=48.04 under the stated known-$\sigma$ normal model.

**2.3** Chi-square=$(30\text{-}25)^2/25+(20\text{-}25)^2/25=1+1=2$.

**2.4** False. Correlation shows association; a third variable or chance can explain it, so causation needs an experiment or other evidence.

**2.5 B** The 95% describes the procedure's long-run success rate over repeated samples. The population parameter is fixed, and one computed interval either contains it or not.

**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.