Sampling and confidence intervals - Paper 3

10 questions | 60 marks | 75 minutes

A community transport project combines survey design, estimates and careful reporting. Use supplied two-sided critical values: 90% 1.645, 95% 1.96, 98% 2.326 and 99% 2.576. Retain intermediate precision; give final probabilities to four decimal places and state assumptions.

Equivalent correct methods earn the corresponding marks.

1. A town has 2,400 residents in its east zone and 1,600 in its west zone. Plan a proportional random sample of 100. [6 marks]

(a) Give the stratum sizes and selection method. [3 marks]

Multiply each population fraction by 100.

Answer: 60 east and 40 west; randomly choose within each complete zone list.

(b) Why can surveying 100 people at the central bus stop distort a town-wide estimate? [3 marks]

The sampling frame is linked to the characteristic being studied.

Answer: Bus-stop users can differ from residents who do not use buses.

2. The selected sample contains 60 east and 40 west residents. Only 30 east and all 40 west residents reply. [6 marks]

(a) Find the east-zone proportion among respondents and compare it with the town proportion. [3 marks]

There are 70 respondents.

$\frac{30}{70}$ differs from $\frac{2,400}{4,000}$.

Answer: 0.4286 among respondents versus 0.6 in the town.

(b) Explain what this reveals without assuming a direction of opinion bias. [3 marks]

Do not invent which group is more supportive.

Answer: Response has changed zone representation; an opinion bias is possible but its direction is unknown without zone/opinion information.

3. A separate representative random sample of 625 residents has 375 regular public-transport users. [6 marks]

(a) Find the estimate and estimated standard error. [3 marks]

$\frac{375}{625} = 0.6$.

Take the square root of $\frac{0.24}{625}$.

Answer: 0.6 and 0.0196.

(b) Construct a 95% interval for the town's regular-use proportion. [3 marks]

Use the estimated standard error and critical value 1.96.

Large observed success/failure counts support the approximation.

Answer: $(0.5616, 0.6384)$

4. Normal waiting times in a controlled model have mean 12 minutes and SD 6 minutes. A mean is formed from 36 independent waits. [6 marks]

(a) Find the model for the mean. [3 marks]

Standard error $= \frac{6}{6} = 1$.

Answer: $\bar{X} \sim N(12, 1^2)$.

(b) Find the probability that the mean is below 10.5 minutes. [3 marks]

Standardise to $z = -1.5$.

Evaluate the lower normal tail.

Answer: 0.0668

5. A large random independent waiting-time sample of 100 has mean 11.8 minutes and sample standard deviation 5 minutes. Population spread is unknown. Use a large-sample normal approximation. [6 marks]

(a) Construct an approximate 90% confidence interval for the population mean. [3 marks]

Estimate standard error as $\frac{s}{\sqrt{n}} = \frac{5}{10} = 0.5$ minutes.

Margin $= 1.645 \times 0.5 = 0.8225$ minutes.

Using sample rather than known population spread makes this a large-sample approximate normal interval.

Answer: $(10.9775, 12.6225)$

(b) The underlying wait distribution is right-skewed. Explain the normal-model justification and its limitation. [3 marks]

The individual distribution need not be normal, but the approximation is not automatic.

Answer: A large independent sample supports an approximate normal mean model through the central limit theorem, and sample spread estimates unknown population spread; severe tails or dependence can still be problematic.

6. Assume a population proportion $p = 0.6$ and $n = 625$ independent observations. [6 marks]

(a) Find the approximate probability that the sample proportion lies between 0.56 and 0.64. [3 marks]

Use mean 0.6 and SD $\sqrt{\frac{0.24}{625}}$.

Apply a continuous normal approximation without continuity correction.

Answer: 0.9588

(b) Explain why this probability calculation uses 0.6 inside the standard deviation. [3 marks]

Sampling probabilities and confidence intervals use different information.

Answer: 0.6 is the stated model probability here, not an unknown parameter being estimated by an interval.

7. An online press report describes a 95% mean interval $(10.9, 12.7)$ minutes. [6 marks]

(a) Correct “95% of residents wait between 10.9 and 12.7 minutes.” [3 marks]

Individual variability requires a different distribution.

Answer: The interval estimates the population mean wait, not individual residents’ waits.

(b) Can the interval establish an average wait below 12 minutes? [3 marks]

An average below 12 is not established by the entire interval.

Answer: No; it contains values on both sides of 12.

8. Planners require a 95% proportion margin at most 0.025 with no defensible prior estimate. [6 marks]

(a) Find the conservative sample size. [3 marks]

$\frac{1.96^2 \times 0.25}{0.025^2} = 1{,}536.64$; round up.

Answer: 1537

(b) If the expected response rate is 80%, estimate the contact budget and state one limitation. [3 marks]

Divide required valid completions by 0.8, then round up.

Answer: 1922 contacts; neither completions nor absence of non-response bias is guaranteed.

9. Two separate 95% support intervals are east $(0.55, 0.63)$ and west $(0.57, 0.65)$. [6 marks]

(a) Recover each estimate and margin. [3 marks]

Use midpoints and half-widths.

Answer: East estimate 0.59, west 0.61; both margins 0.04.

(b) Evaluate “Their overlap proves there is no difference.” [3 marks]

These intervals estimate two separate parameters.

Answer: Overlap does not prove equality. A direct difference analysis needs its own model.

10. A report claims “The proportion will still be 60% in five years because today’s interval is narrow.” [6 marks]

(a) Identify the unsupported step. [3 marks]

Time change is not quantified by the current sampling margin.

Answer: An interval for the current population does not establish a future population preference.

(b) Write a more responsible conclusion. [3 marks]

Do not turn present precision into a guarantee about change over time.

Answer: Report the current estimate and confidence interval, the survey date and sampling limitations; describe any future prediction as requiring a separate model.