Past-paper walk-through

Sampling and estimation ▪ 6.4

eXercise

Questions in this set

- 9709/61 N25 Q4 ▪ 4 marks
- 9709/62 N25 Q3 ▪ 8 marks
- 9709/61 J25 Q2 ▪ 7 marks
- 9709/61 J25 Q3 ▪ 8 marks
- 9709/61 J25 Q6 ▪ 7 marks

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 4

9709/61 N25 ▪ Q4 ▪ 4 marks

The masses of a certain species of animal are normally distributed with standard deviation σ kg. A researcher uses a random sample of n animals to find two confidence intervals ($\alpha\%$ and 90

(a) Find α. **[3]**

(b) Find the probability that the 90

Solution 4

(a)

Mark scheme

- Width $\propto$ z-value: $\frac{z_\alpha}{z_{90}} = 1.414 \Rightarrow z_\alpha = 1.414 \times 1.645 = 2.326$. $\Phi(2.326) = 0.99 \Rightarrow \frac{1+\alpha/100}{2} = 0.99 \Rightarrow \alpha = 98$.

Solution 4

(b)

Mark scheme

- The 90

Question 3

9709/62 N25 ▪ Q3 ▪ 8 marks

The times, in minutes, taken by students to complete a test have mean μ and standard deviation σ. The times taken by a random sample of 100 students are noted and are used to calculate a 95

(a) Given that the end points of the 95% confidence interval are 31.02 and 33.98, correct to 4 significant figures, calculate the value of σ. **[3]**

(b) The calculation of the confidence interval required the use of the Central Limit theorem. Explain why it is valid to use the Central Limit theorem in this case. **[1]**

(c) A researcher calculates a number, r, of 95

Solution 3

(a)

Mark scheme

- Half-width $= 1.96 \dfrac{\sigma}{\sqrt{100}} = \dfrac{33.98 - 31.02}{2} = 1.48 \Rightarrow \sigma = \dfrac{1.48 \times 10}{1.96} = 7.55.$

Solution 3

(b)

Mark scheme

- Because the sample is large ($n = 100$), the distribution of the sample mean is approximately normal whatever the underlying distribution of the times.

Solution 3

(c)

Mark scheme

- The intervals are independent, each with probability 0.95 of containing μ: $0.95^r > 0.5 \Rightarrow r < \dfrac{\ln 0.5}{\ln 0.95} = 13.5$. Largest $r = 13$.

Question 2

9709/61 J25 ▪ Q2 ▪ 7 marks

The random variable X has the distribution $B(8, \frac{3}{4})$. A random sample of 100 values of X is chosen, and the sample mean, $\overline{X}$, is found.

(a) Find $\mathrm{P}(\overline{X} > 6.2)$. You are not expected to use a continuity correction. **[6]**

(b) State why the Central Limit Theorem was needed in the calculation in part (a). **[1]**

Solution 2

(a)

Mark scheme

- $X \sim B(8, \frac{3}{4})$: mean 6, variance 1.5. $\overline{X} \approx N(6, \frac{1.5}{100})$;
  $$\mathrm{P}(\overline{X} > 6.2) = 1 - \Phi\left(\frac{6.2-6}{\sqrt{1.5/100}}\right) = 1 - \Phi(1.633) = 0.0512.$$

Solution 2

(b)

Mark scheme

- The population X is not normally distributed (it has a binomial distribution), so the CLT is needed for $\overline{X}$ to be approximately normal.

Question 3

9709/61 J25 ▪ Q3 ▪ 8 marks

The time, T minutes, for a certain daily bus journey is normally distributed. The bus company claims that the mean of T is 45. A passenger believes that the mean of T is actually greater than 45. She notes the times taken for this journey on a random sample of 60 days. The results are summarised below: $n = 60$, $\sum t = 2750$, $\sum t^2 = 127000$.

(a) Calculate unbiased estimates of the population mean and variance. **[3]**

(b) Test the passenger's belief at the 5% significance level. **[5]**

Solution 3

(a) Worked steps

Two formulae, and the divisor is what makes each estimate **unbiased**.

- $\widehat{\mu} = \bar{x} = \dfrac{2750}{60} = \dfrac{275}{6} = \mathbf{45.8}$
- $\widehat{\sigma}^2 = \dfrac{n}{n-1}\left(\dfrac{\sum x^2}{n} - \bar{x}^2\right) = \dfrac{60}{59}\left(\dfrac{127\,000}{60} - \left(\dfrac{275}{6}\right)^2\right) = \mathbf{16.2}$

The $\dfrac{n}{n-1}$ factor is the whole point of the word **unbiased**, and it is the mark: the sample variance systematically understates the population's, and multiplying by $\dfrac{n}{n-1}$ corrects it.

Solution 3

Keep $\bar{x}$ as the exact fraction $\frac{275}{6}$ while squaring it. Rounding to 45.8 first and then squaring loses accuracy that shows up in the third significant figure of the variance — and the test statistic in part (b) is sensitive to it. **Mark scheme**

- $\widehat{\mu} = \dfrac{2750}{60} = \dfrac{275}{6} = 45.8$; $\widehat{\sigma}^2 = \dfrac{60}{59}\left(\dfrac{127000}{60} - \left(\dfrac{275}{6}\right)^2\right) = 16.2$ (i.e. $\frac{2875}{177}$).

Solution 3

(b) Worked steps

Four stages, all marked.

- H_0: mean $= 45$; H_1: mean > 45 — **one-tailed**, since the passenger believes the journey takes *longer*.
- $z = \dfrac{45.83 - 45}{\sqrt{16.24/60}} = \dfrac{0.833}{0.520} = 1.60$
- Critical value at the 5% level, one-tailed: 1.645.
- $1.60 < 1.645$, so **do not reject** H_0: there is insufficient evidence, at the 5% level, that the mean journey time exceeds 45 minutes.

Solution 3

Three points:

- **Use the estimates from part (a)**, unrounded — with 16.2 rounded down the statistic drifts, and this test is close enough to the boundary for that to matter.
- **Divide by $\sqrt{\widehat{\sigma}^2/n}$**, the standard error, not by the standard deviation.
- **Conclude in context, and do not overclaim.** Failing to reject is not proof that the mean is 45; it is a failure to find evidence against it. With $z = 1.60$ against a boundary of 1.645 the result is close, and saying so is a fair observation.

Solution 3

Mark scheme

- H_0: mean $= 45$, H_1: mean > 45. $z = \dfrac{45.83 - 45}{\sqrt{16.24/60}} = 1.60$. Since $1.60 < 1.645$, do not reject H_0: insufficient evidence that the mean journey time exceeds 45 minutes.

Question 6

9709/61 J25 ▪ Q6 ▪ 7 marks

A manufacturer of cell phones claims that 25% of students own a Pumpkin phone. Jeyeraj thinks that the proportion of students at his large college who own a Pumpkin phone is less than 25%. He plans to test the manufacturer's claim. He chooses a random sample of 30 students at his college. If the number of students who own a Pumpkin phone is less than 5, Jeyeraj will reject the manufacturer's claim.

(a) State suitable hypotheses for the test. **[1]**

(b) Given that the true proportion of students at the college who own a Pumpkin phone is 10%, use a binomial distribution to find the probability of a Type II error. **[3]**

(c) At Florence's college, in a random sample of 40 students, it was found that 5 own a Pumpkin phone. Calculate an approximate 95% confidence interval for the proportion of students at Florence's college who own a Pumpkin phone. **[3]**

Solution 6

(a)

Mark scheme

- H_0: $p = 0.25$; H_1: $p < 0.25$ (proportion of students owning a Pumpkin phone).

Solution 6

(b)

Mark scheme

- True $p = 0.1$, $X \sim B(30, 0.1)$. Rejection region is $X < 5$, so Type II error $= \mathrm{P}(X \geqslant 5) = 1 - \mathrm{P}(X \leqslant 4) = 0.175$.

Solution 6

(c)

Mark scheme

- $\hat{p} = \frac{5}{40} = 0.125.$ 95