Unbiased Estimates 1

Further A

Part A 14 observations of a random variable V

The results of 14 observations of a random variable V are summarised by

$$n = 14,\quad \Sigma v = 3752,\quad \Sigma v^2 = 1\,007\,448.$$

Calculate an unbiased estimate of $\mathrm{E(V)}$.

Calculate an unbiased estimate of $\mathrm{Var(V)}$, giving your answer to 3 s.f.

Part B 50 observations of a random variable X

A random sample of 50 observations of the random variable X is summarised by

$$n = 50,\quad \Sigma x = 182.5,\quad \Sigma x^2 = 739.625.$$

Calculate an unbiased estimate of the expectation of X.

Calculate an unbiased estimate of the variance of X.

Adapted with permission from UCLES, A Level, June 2016 and January 2012, Papers 4733/01 and 4733, Questions 1 and 1

Hypothesis Testing: Animals

Further A

P P P

It is known that the lifetime of a certain species of animal in the wild has mean $13.3\,\mathrm{years}$. A zoologist reads a study of 50 randomly chosen animals of this species that have been kept in zoos. According to the study, for these 50 animals the sample mean lifetime is $12.48\,\mathrm{years}$ and the population variance is $12.25\,\mathrm{years}^2$.

Part A Distribution of sample means

What distribution can be used to model the distribution of sample means? Explain your answer. Drag and drop into the boxes below to complete the explanation.

The distribution of the lifetime of the animal in the wild is not stated However, for [ ] samples, the central limit theorem states that the distribution of sample means will be approximately [ ]. For a population variance $\sigma =$ [ ] years^2, the variance of the distribution of sample means when $n = 50$ will be [ ] years^2.

Let X be the mean lifetime of one animal. Then $\overline{X} \sim \mathrm{N}(\mu,$ [ ]$)$.

Items:

geometric 25 **normal** 612.5 2.45 0.245 **10** 1.225 **large** **5** **binomial** **1** 12.25 **small**

Part B **Hypothesis test**

Test at the 5% significance level whether these results provide evidence that animals of this species that have been kept in zoos have a shorter expected lifetime than those in the wild.

The mean lifetime of animals in the wild is $13.3\,\text{years}$. We are testing whether there is evidence that animals in zoos have a shorter life expectancy. The null and alternative hypotheses are:

$$\mathrm{H}_0 : \mu = 13.3\,\text{years} \qquad \mathrm{H}_1 : \boxed{}\ \text{years}$$

Assuming that the null hypothesis is true, we have that $\overline{X} \sim \mathrm{N}(13.3, \boxed{})$.

The probability of selecting a sample with a mean of $12.48\,\text{years}$ or less is found to be

$$\mathrm{P}(\overline{X} \leq 12.48) = \boxed{}$$

The calculated probability for the sample $\boxed{}$ 0.05.

Therefore, $\boxed{}$ the null hypothesis. There $\boxed{}$ evidence to suggest that animals kept in zoos have a shorter life expectancy.

Items:

$\mu < 13.3$ | 0.951 | $\mu = 12.48$ | 0.245 | **is not** | **is** | 0.0488 | **is equal to** | $\mu \neq 13.3$ | **reject** | $\mu > 13.3$

is less than | **is greater than** | 0.407 | 12.25

Part C Population and sample variance

Subsequently the zoologist discovered that there had been a mistake in the study. The quoted variance of 12.25 years^2 was in fact the sample variance. Determine whether this makes a difference to the conclusion of the test.

The sample variance is a biased estimator of the population variance. An unbiased estimate of the population variance can be obtained by multiplying the sample variance by [ ], which gives a value of [ ].

Repeating the calculation from the previous part with this new value for the population variance gives $\mathrm{P}(\overline{X} \leq 12.48) =$ [ ]. This is [ ] 0.05. Therefore, [ ] the null hypothesis. There [ ] significant evidence to suggest that animals kept in zoos have a shorter life expectancy.

Items:

$\frac{n}{n-1}$ | **do not reject** | **greater than** | **less than** | **is not** | $\frac{n}{n+1}$ | $\frac{n-1}{n}$ | 0.125 | 0.0505 | **is** | 0.49 | **reject** | 0.949 | 0.25 | **equal to** | 0.0488

Used with permission from UCLES, A Level, June 2016, Paper 4733/01, Question 8

Gameboard:

STEM SMART Double Maths 42 - The Central Limit Theorem & Estimation

Hypothesis Testing: Motorist

Further A

A motorist records the time taken, T minutes, to drive a particular stretch of road on each of 64 occasions. Her results are summarised by

$$\Sigma t = 876.8, \quad \Sigma t^2 = 12\,657.28.$$

Part A **Hypothesis test**

Test, at the 5% significance level, whether the mean time for the motorist to drive the stretch of road is greater than 13.1 minutes.

Available items

From the summary data, we find that $\overline{T} \sim \mathrm{N}(13.7, 0.1575)$

Therefore we reject the null hypothesis. There is evidence to suggest that the mean time for the motorist to drive the stretch of road is greater than $13.1\,\mathrm{minutes}$.

Therefore we do not reject the null hypothesis. There is insufficient evidence to suggest that the mean time for the motorist to drive the stretch of road is greater than $13.1\,\mathrm{minutes}$.

The probability that the sample would have a mean of 13.7 or more is

$$\mathrm{P}(\overline{T} \geq 13.7) = 0.0668$$

The null and alternative hypotheses are:

$$H_0 : \mu = 13.7 \quad H_1 : \mu > 13.7$$

From the summary data, we find that $\overline{T} \sim \mathrm{N}(13.1, 10.08)$

Comparing this probability to the significance level, we see that it is greater than 0.05.

The null and alternative hypotheses are:

$$H_0 : \mu = 13.1 \quad H_1 : \mu > 13.1$$

Comparing this probability to the significance level, we see that is is less than 0.05.

From the summary data, we find that $\overline{T} \sim \mathrm{N}(13.1, 0.16)$

The probability that the sample would have a mean of 13.7 or more is

$$\mathrm{P}(\overline{T} \geq 13.7) = 0.0653$$

Part B Central Limit Theorem

Explain whether it is necessary to use the Central Limit Theorem in your test.

- No, because the population already has a normal distribution.
- Yes, because the sample is large, we must use the Central Limit Theorem to show that the distribution of sample means will have a normal distribution.
- No, because times are continuous variables and hence always follow a normal distribution.
- Yes, because we do not know the distribution of the population. The Central Limit Theorem tells us that since the sample is large, the distribution of sample means will have normal distribution.

Used with permission from UCLES, A Level, January 2009, Paper 4733, Question 7

Gameboard:

STEM SMART Double Maths 42 - The Central Limit Theorem & Estimation

Confidence Intervals: Smartphones

Further A P P P

A survey was conducted to find the proportion of people owning smartphones. It was found that 203 out of a random sample of 278 people owned a smartphone.

Part A 97% confidence interval

Calculate a 97% confidence interval for the true proportion of people who own a smartphone.

Give the lower endpoint of the interval to 3 s.f.

Give the upper endpoint of the interval to 3 s.f.

Part B Unsatisfactory sample

A second survey to find the proportion of people owning smartphones was conducted at 10 o'clock on a Thursday morning in a shopping centre.

Give one reason why this is not a satisfactory sample.

- The people may not answer truthfully
- The people there at that time are not representative of the population
- The same person might be asked multiple times

Adapted with permission from CIE November 2006, S2 Question 3.

Gameboard:

STEM SMART Double Maths 42 - The Central Limit Theorem & Estimation

Confidence Intervals

Further A

P P P

Part A Fish food

Packets of fish food have weights that are distributed with standard deviation $2.3\,\mathrm{g}$. A random sample of 200 packets is taken. The mean weight of this sample is found to be $99.2\,\mathrm{g}$. Calculate a 99% confidence interval for the population mean weight.

Give the lower endpoint of the interval to 3 s.f.

Give the upper endpoint of the interval to 3 s.f.

Part B Fitness trial

The result of a fitness trial is a random variable X which is normally distributed with mean μ and standard deviation 2.4. A researcher uses the results from a random sample of 90 trials to calculate a 98% confidence interval for μ.

What is the width of this interval? Give your answer to 3 s.f.

Part C Gown hire

The manager of a gown hire shop wishes to estimate the proportion of gowns damaged by customers. They take a random sample of 120 gowns and finds that 33 of them are damaged. Find a 95% confidence interval for the true proportion of gowns that are being damaged when hired from this shop.

Give the lower endpoint of the interval to 3 s.f.

Give the upper endpoint of the interval to 3 s.f.

Adapted with permission from CIE 2002, 2006 Paper 7 S2.

Gameboard:
STEM SMART Double Maths 42 - The Central Limit Theorem & Estimation

Confidence Intervals 2

Further A

The random variable $X \sim \mathrm{N}(\mu, 3^2)$. A random sample of 9 observtions of X produced the following values:

$$6,\ 2,\ 3,\ 6,\ 8,\ 11,\ 12,\ 5,\ 10$$

Part A Confidence interval

Find a 90% confidence interval for μ.

Give the lower bound of the confidence interval to 4 s.f.

Give the upper bound of the confidence interval to 4 s.f.

Part B Meaning of a confidence interval

Explain what is meant by a 90% confidence interval in this context.

- There is a 90% probability that the confidence interval will contain the value of μ.
- There is a 90% chance that $\bar{x}$ will lie inside the interval.
- There is a 90% chance that μ will lie inside the interval.

Used with permission from UCLES, A Level, June 2017 Paper 4734/01, Question 1

Gameboard:

STEM SMART Double Maths 42 - The Central Limit Theorem & Estimation

Confidence Intervals 3

Further A

A 95% confidence interval for the mean μ of a certain population, based on a sample of size 35, is $(6.0061,\ 7.9939)$.

Find the minimum sample size such that the width of a 95% confidence interval for μ is less than 1. Give your answer to 3 s.f.

Used with permission from UCLES, A Level, June 2018, Paper 4734/01, Question 2