STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Unbiased Estimates 1

Subject & topics: Maths | Statistics | Random Variables **Stage & difficulty:** Further A P2

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 unbiased estimates of the following.

$\mathrm{E}(V) =$

$\mathrm{Var}(V) =$ (3 sf)

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 unbiased estimates of the expectation and variance of X.

The expectation of X is

The variance of X is

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

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Hypothesis Testing: Animals

Subject & topics: Maths | Statistics | Hypothesis Tests **Stage & difficulty:** Further A P2

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 $\sigma^2 =$ _____ 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:

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

0.0488 | reject | is | is less than | is not | 0.407 | $\mu = 12.48$ | is equal to | $\mu \neq 13.3$ | 0.245 | $\mu > 13.3$ | is greater than | $\mu < 13.3$ | 0.951 | 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 $12.25\ \text{years}^2$ 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 means that $\text{Var}(\overline{X}) =$ _____.

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

Items:

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

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

Question deck:

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Hypothesis Testing: Motorist

Subject & topics: Maths | Statistics | Hypothesis Tests **Stage & difficulty:** Further A P2

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 $\overline{T} \sim \mathrm{N}(13.1, 10.08)$

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 $13.1\,\mathrm{minutes}$.

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

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

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

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

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

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 $13.1\,\mathrm{minutes}$.

The null and alternative hypotheses are:

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

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

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

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

The null and alternative hypotheses are:

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

Part B

Central Limit Theorem

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

- 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 the population already has 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 a normal distribution.

Question deck:
STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Confidence Intervals

Subject & topics: Maths | Statistics | Hypothesis Tests **Stage & difficulty:** Further A P3

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, giving your answer to 3 sf.

(,)

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 sf.

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 your answer to 3 sf.

(,)

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

Question deck:

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Confidence Intervals 3

Subject & topics: Maths | Statistics | Hypothesis Tests **Stage & difficulty:** Further A P3

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.

Question deck:
STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Continuous Random Variables 3

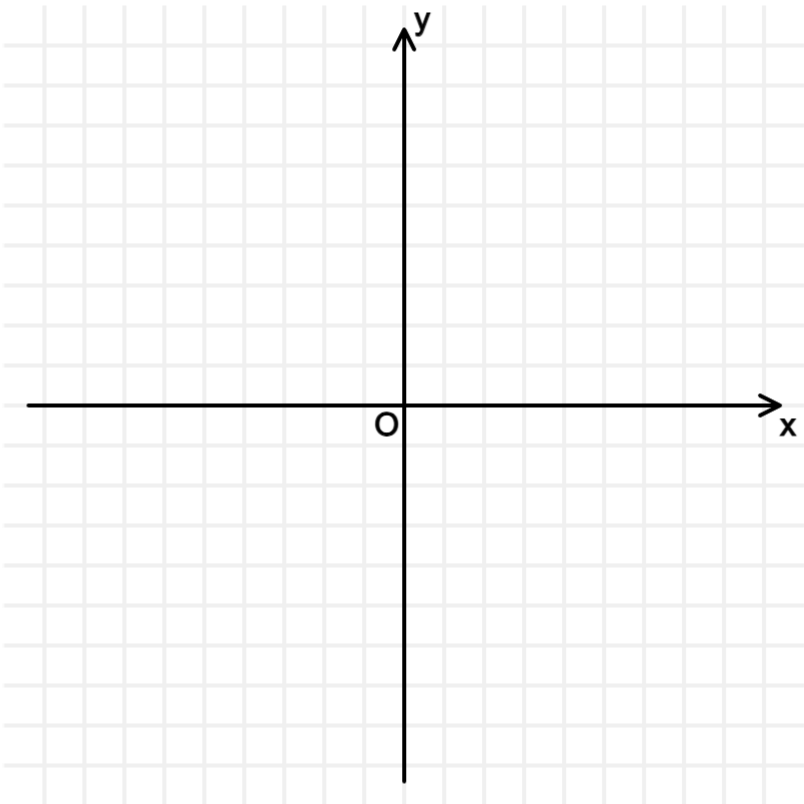
Subject & topics: Maths | Statistics | Random Variables **Stage & difficulty:** Further A P2

The continuous random variable $T \sim \mathrm{U}[5.0, 11.0]$.

Part A

Sketch the probability density function

Sketch the graph of the probability density function of T. Sketch only the nonzero parts of the probability density function.



Part B

Find $\mathrm{E}(T)$

Find the expectation of T.

Part C

Find $\mathrm{Var}(T)$

Find the variance of T.

Part D

Sample of T

A random sample of 48 observations of T is obtained. Find the probability that the mean of the sample is greater than 8.3. Give your answer to 3 significant figures.

Part E

Approximation

Explain whether or not the probability found in part D is an approximation.

- Since the distribution of T is not normal, we can't use the central limit theorem, so the distribution of sample means is only approximately normally distributed.
- Since the distribution of T is not normal, we must use the central limit theorem, which always gives an exact distribution for the distribution of sample means.
- Since the distribution of T is not normal, the central limit theorem only gives an approximate distribution for the distribution of sample means.
- Since T is a continuous random variable, the central limit theorem will give an exact distribution for the distribution of sample means.

Question deck:

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Continuous Random Variables 4

Subject & topics: Maths | Statistics | Random Variables **Stage & difficulty:** Further A P2

The cumulative distribution function of the continuous random variable, X, is given by

$$F(x) = \begin{cases} 0 & x < 1, \\ \frac{2x-2}{x+3} & 1 \leq x \leq 5, \\ 1 & x > 5. \end{cases}$$

Given that $Y = 2X - 3$, find the probability density function of Y.

The following symbols may be useful: y

Adapted with permission from UCLES, A Level, January 2018, Paper 4734/01, Question 3

Question deck:

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Continuous Random Variables 5

Subject & topics: Maths | Statistics | Random Variables **Stage & difficulty:** Further A P2

The continuous random variable X has probability density function

$$f(x) = \begin{cases} k\cos x & 0 \leqslant x < \frac{\pi}{4}, \\ k\sin x & \frac{\pi}{4} \leqslant x \leqslant \frac{\pi}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Part A

Value of k

Find the exact value of k.

Part B

$\mathrm{P}(X \leq 1)$

Find the value of $\mathrm{P}(X \leq 1)$. Give your answer to 3 sf.

Part C

Upper quartile

Find the upper quartile of X. Give your answer to 3 sf.

Question deck:

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Unbiased Estimates 2

Subject & topics: Maths | Statistics | Random Variables **Stage & difficulty:** Further A P3

The continuous random variable Z has probability density function

$$f(z) = \begin{cases} \frac{4z^3}{k^4} & 0 \leq z \leq k \\ 0 & \text{otherwise} \end{cases}$$

where k is a parameter whose value is to be estimated.

Part A
$\mathrm{E}(Z)$

Find $\mathrm{E}(Z)$.

The following symbols may be useful: k

Part B
Unbiased estimator bZ

Hence find b such that bZ is an unbiased estimator of k.

The following symbols may be useful: b

Part C

$\mathrm{Var}(bZ)$

Using the value of b you found earlier, find $\mathrm{Var}(bZ)$.

The following symbols may be useful: k

Part D

Unbiased estimator cY

The parameter k can also be estimated by making observations of a random variable X which has mean $\frac{1}{2}k$ and variance $\frac{1}{12}k^2$. Let $Y = X_1 + X_2 + X_3$ where X_1, X_2 and X_3 are independent observations of X.

cY is also an unbiased estimator of k. Find the value of c.

The following symbols may be useful: c

Part E

Which estimator

For the values of b and c found earlier, determine which of bZ and cY is the more efficient estimator of k and give a justification for that choice.

We know from earlier that $\mathrm{Var}(bZ) =$ ___.
We find that $\mathrm{Var}(cY) =$ ___.

Hence ___ is the more efficient estimator because its variance is ___.

Items:

$\frac{k^2}{9}$ $\frac{k^2}{24}$ $\frac{k^2}{6}$ $\frac{k^2}{27}$ bZ $\frac{4k^2}{9}$ cY $\frac{k^2}{108}$ $\frac{k^2}{4}$ greater smaller $\frac{9k^2}{16}$

Question deck:

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

STEM SMART Double Maths 35 - Estimation & Continuous Random Variables

Hypothesis Testing: Birch Leaves

Subject & topics: Maths | Statistics | Hypothesis Tests **Stage & difficulty:** Further A P3

A scientist is investigating the lengths of the leaves of birch trees in different regions. They take a random sample of 50 leaves from birch trees in region 1 and a random sample of 60 leaves from birch trees in region 2. They record the lengths in cm, x and y, respectively. Their results are summarised as follows.

$$\sum x = 282 \qquad \sum x^2 = 1596 \qquad \sum y = 328 \qquad \sum y^2 = 1808$$

The population mean lengths of leaves from birch trees in regions 1 and 2 are μ_x cm and μ_y cm respectively.

Carry out a test at the 5% significance level to determine whether there is a difference in the mean lengths of the leaves of birch trees between the two regions.

The null and alternative hypotheses are:

$$\mathrm{H}_0 : \mu_x \ \boxed{}\ \mu_y \qquad \mathrm{H}_1 : \mu_x \ \boxed{}\ \mu_y$$

Let $\overline{X}$ and $\overline{Y}$ be the sample means for the lengths of leaves from birch trees in regions 1 and 2 respectively. We then find that $\overline{X} - \overline{Y} \sim \mathrm{N}(\boxed{}, \boxed{})$.

The test statistic, $\mathrm{P}(\overline{X} - \overline{Y} > \boxed{}) = \boxed{}$.

The test statistic is $\boxed{}$ than $\boxed{}$.

Therefore we $\boxed{}$ H_0 at the 5% level. There $\boxed{}$ evidence to suggest that the lengths of leaves of birch trees is different between the two regions.

Items:

do not reject | $=$ | 0.1 | reject | 0.0804 | $\frac{13}{75}$ | 0.0187 | 0.025 | $>$ | 0.0833 | 0.00694 | 0 | 0.0156

0.0318 | is insufficient | is | 0.981 | $\frac{13}{15}$ | $\frac{93}{50}$ | 0.00647 | $\neq$ | $<$ | 0.00197 | 0.05 | greater | less