Poisson Distribution - Woodland

University

There are on average 140 trees per hectare ($10^4\,\mathrm{m}^2$) in a wood, 30% of which are oak trees. On the assumption that the distribution of trees in the wood follows a Poisson distribution, answer the following.

Part A Less than 4 oaks

Find the probability that there are less than 4 oaks in an area of $1000\,\mathrm{m}^2$. Give your answer to 4 s.f.

Part B Most probable number of other species

Find the most probable number of trees which are not oaks in an area of $1000\,\mathrm{m}^2$.

Part C One oak in each of 3 areas

Three $500\,\mathrm{m}^2$ areas are selected at random. Find the probability that there will be exactly 1 oak in each of them. Give your answer to 3 s.f.

Part D Number of areas with exactly 6 oaks

It is known that in the wood 30 areas of $1000\,\mathrm{m}^2$ contain 6 or more oaks. Obtain an estimate of the number of these which contain exactly 6 oaks, giving your answer to the nearest integer.

Part E 4 selected areas

Four areas of $500\,\mathrm{m}^2$ are selected at random. Find the overall probability that two of the areas contain exactly 8 trees, one contains more than 8 trees and one contains fewer than 8 trees. Give your answer to 3 s.f.

Created for isaacphysics.org by Julia Riley

Hypothesis Testing: Exams

Further A

Normal Distribution

An examination board is developing a new syllabus and wants to know if the question papers are the right length. A random sample of 50 candidates was given a pre-test on a dummy paper. The times, t minutes, taken by these candidates to complete the paper can be summarised by

$$n = 50, \quad \Sigma t = 4050, \quad \Sigma t^2 = 329\,800.$$

Assume that times are normally distributed.

Part A Not completed within 90 minutes

Estimate the proportion of candidates that could not complete the paper within 90 minutes. Give your answer to 3 s.f.

Part B Hypothesis test

Test, at the 10% significance level, whether the mean time for all candidates to complete this paper is $80\,\text{minutes}$. Use a two-tailed test. Fill in the gaps below.

Let μ be the mean time to complete the test. The null and alternative hypotheses are:

$$\text{H}_0 : \mu = \boxed{} \qquad \text{H}_1 : \mu \boxed{}$$

Assuming that the null hypothesis is true, the distribution of sample means $\overline{T} \sim \text{N}(\boxed{}, \boxed{})$.

The test statistic is $\text{P}(\overline{T} \geq \boxed{}) = \boxed{}$.

For a two-tailed test at the 10% significance level, we find that the test statistic is $\boxed{}$ $\boxed{}$.

Therefore, $\boxed{}$ the null hypothesis. There $\boxed{}$ evidence to suggest that the mean time for all candidates to take this paper is not equal to $80\,\text{minutes}$.

Items:

greater than	0.250	35	0.118	**is**	81	0.116	80	0.0118	**do not reject**	**reject**
less than	**is insufficient**	0.7	$\frac{250}{7}$	$\neq 80$	0.05	0.1	> 80	$\frac{5}{7}$	< 80	

Part C Assumptions

Explain whether the assumption that times are normally distributed is necessary in answering each part. Fill in the gaps below.

In part A, the assumption that the times are normally distributed [ ] necessary. An unbiased estimate of the population mean and the population variance can be made from the sample [ ] this assumption. However, calculating a [ ] from these values requires knowledge of the distribution.

In part B, the assumption that the times are normally distributed [ ] necessary. The central limit theorem states that the sample means for samples taken from [ ] distribution will be approximately normally distributed as long as the sample size is sufficiently [ ], which is the case in this question ($n = 50$).

Items:

small **is not** **any** **large** **without** **with** **a continuous** **probability** **is** **a discrete** **a normal**

Used with permission from UCLES, A Level, June 2014, Paper 4733/01, Question 7

Gameboard:

STEM SMART Double Maths 50 - Further Statistics

Revision

Correlation - a Torsional Pendulum

Further A

An experiment was carried out to investigate the amplitude of the driven oscillations of a torsional pendulum as a function of the frequency of the driving force. The initial set of results (I) and an additional follow-up set of results (II) are shown below. (The amplitude is given in arbitrary units.)

Initial results I

Frequency (Hz)	0.28	0.36	0.43	0.49	0.51	0.56
Amplitude	1.0	1.4	1.9	3.1	6.2	5.2

Follow-up results II

Frequency (Hz)	0.52	0.53	0.54	0.64	0.87
Amplitude	15.2	12.5	7.9	1.6	0.5

Part A Initial results I - Spearman's rank correlation coefficient

Find Spearman's rank correlation coefficient for the initial set of results I, giving your answer to 3 s.f.

Part B Initial results I - significance of the rank correlation coefficient

Test at the 5% significance level whether there is a correlation between the driving frequency and the amplitude of the oscillations for the initial results I, using Spearman's rank correlation coefficient.

Find the appropriate critical value for Spearman's rank correlation coefficient, giving your answer to 3 s.f.

What do you conclude about whether there is a correlation at the 5% significance level between the driving frequency and the amplitude of the oscillations?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Spearman's rank correlation coefficient is [ ] the calculated value; there is therefore [ ] evidence to [ ] the null hypothesis at the 5% level and [ ] evidence for a correlation at this level.

Items:

significant **equal to** **less than** **reject** **accept** **greater than** **no** **sufficient** **negative** **insufficient** **positive**

Part C Initial results I - Pearson's product moment correlation coefficient

Find Pearson's product moment correlation coefficient for the initial set of results I, giving your answer to 3 s.f.

Part D Initial results I - significance of the product moment correlation coefficient

Test at the 5% significance level whether there is a correlation between the driving frequency and the amplitude of the oscillations for the initial results I, using Pearson's product moment correlation coefficient.

Find the appropriate critical value for Pearson's product moment correlation coefficient, giving your answer to 3 s.f.

What do you conclude about whether there is a correlation at the 5% significance level between the driving frequency and the amplitude of the oscillations?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Pearson's product moment correlation coefficient is [ ] the calculated value; there is therefore sufficient evidence to [ ] the null hypothesis at the 5% level and [ ] evidence for a correlation at this level.

Items:

accept | **greater than** | **no** | **positive** | **negative** | **significant** | **reject** | **equal to** | **less than**

Part E Results I and II - Pearson's product moment correlation coefficient

Now consider the initial and follow-up results together (11 pairs of values in all). Find Pearson's product moment correlation coefficient for this data set consisting of the 11 pairs of values, giving your answer to 3 s.f.

Part F Results I and II - significance of the product moment correlation coefficient

Test at the 5% significance level whether there is a correlation between the driving frequency and the amplitude of the whole data set, using Pearson's product moment correlation coefficient.

Find the appropriate critical value for Pearson's product moment correlation coefficient, giving your answer to 3 s.f.

What do you conclude now about whether there is a correlation at the 5% significance level between the driving frequency and the amplitude of the oscillations?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Pearson's product moment correlation coefficient is [ ] the calculated value; there is therefore [ ] evidence for a correlation between the driving frequency and the amplitude of the oscillations.

Items:

greater than **equal to** **sufficient** **no significant** **less than**

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 50 - Further Statistics Revision

Continuous Random Variables 7

Further A

The probability density function $f(x)$ for a random variable X is given by

$$f(x) = \begin{cases} \alpha & 0 \leq x \leq \frac{b}{2}, \\ 2\alpha & \frac{b}{2} < x \leq b, \\ 0 & \text{otherwise}, \end{cases}$$

where α and b are constants.

Part A Find α

Find an expression for α in terms of b.

The following symbols may be useful: `alpha, b`

Part B Find $\mathrm{E}(X)$

Find the expectation of X in terms of b.

The following symbols may be useful: `b`

Part C Find $\mathrm{E}(X^2)$

Find the expectation of X^2 in terms of b.

The following symbols may be useful: b

Part D Standard deviation

Find the standard deviation of X in terms of b.

Created for isaacphysics.org by Julia Riley

Gameboard:
STEM SMART Double Maths 50 - Further Statistics Revision

Chi-Squared Tests: Hair Colour

Further A

In an investigation into a possible association between hair colour and height, a random sample of 200 adult men was taken, and the data shown in the table was obtained.

		Hair colour			
		Dark	**Fair**	**Red**	**Total**
Height	Less than 165 cm	16	7	11	34
	165 cm to 180 cm	46	39	9	94
	More than 180 cm	33	35	4	72
Total		95	81	24	200

It is proposed to carry out a χ^2 test for independence between hair colour and height.

Part A Expected frequencies

Calculate the expected frequencies. Fill in the gaps below.

		Hair colour		
		Dark	**Fair**	**Red**
Height	Less than 165 cm			
	165 cm to 180 cm			
	More than 180 cm			

Items:

2.52 | 2.84 | 4.08 | 4.23 | 7.6 | 7.76 | 8.64 | 11.28 | 13.77 | 16.15 | 21.62 | 29.16 | 31.11 | 34.2 | 38.07 | 43.54 | 44.65 | 45.26

Part B Combining cells

Explain why some combining of rows or columns should be carried out. Fill in the gaps below.

Since one of the expected frequencies is [ ], we should combine cells until they are [ ].
We could combine rows or columns to do so. In this case we will combine rows [ ].

Items:

1 **and** 3 | **greater than or equal to** 5 | **less than** 10 | **less than** 5 | **greater than or equal to** 10 | 2 **and** 3 | 1 **and** 2

Part C χ^2 test

Carry out the test, combining suitable rows and using a 5% significance level. Fill in the gaps below.

The null and alternative hypotheses are

H_0 : Hair colour and height are [ ].

H_1 : Hair colour and height are [ ].

We find the value of the χ^2 statistic to be [ ].

The appropriate critical value $\chi^2_{\text{crit}} =$ [ ].

We see that χ^2 [ ] χ^2_{crit}. Therefore we [ ] the null hypothesis. There [ ] evidence to suggest that hair colour and height are dependent.

Items:

7.598	5.787	0.989	12.592	5.991	7.378	**is**	$<$	19.294	6.424	$>$	0.161	**reject**

9.488	4.605	**dependent**	**do not reject**	**independent**	$=$	**is insufficient**

Used with permission from UCLES, A Level, January 2002 S3, Paper 2643, Question 4

Gameboard:

STEM SMART Double Maths 50 - Further Statistics

Revision

Chi-squared Tests: Pension

Further A

A survey is conducted of customers using a Post Office. Starting when the Post Office opens in the morning, a count is made of the number of customers up to and including the first person to collect their old age pension. This is repeated on each of a total of 40 different days and the results are summarised below.

Number of customers	1	2	3	4	5	6	7	8	9	10	≥ 11
Frequency	23	5	3	3	2	1	0	1	0	2	0

It is thought that this distribution may be modelled by a geometric distribution with parameter p, where p is the probability that a customer is collecting their old age pension.

Part A Mean

Calculate the mean using the data above.

Part B Value of p

Calculate the value of p from the data above, giving your answer as a fraction.

The following symbols may be useful: p

Part C χ^2 statistic

It is intended to test, at the 5% significance level, the goodness of fit of the model to the data.

Using the proposed geometric model, calculate the χ^2 statistic for these data. Give your answer to 4 s.f.

Part D Hypothesis test

Carry out the test at the 5% significance level. Fill in the gaps below.

The null and alternative hypotheses are

H_0 : The data [ ] the proposed model.

H_1 : The data [ ] the proposed model.

The appropriate critical value $\chi^2_{\text{crit}} =$ [ ].

We see that χ^2 [ ] χ^2_{crit}. Therefore we [ ] the null hypothesis. There [ ] evidence to suggest that the data do not fit the proposed model.

Items:

7.378	9.488	**is**	9.348	4.605	$=$	**reject**	$<$	**do not reject**	7.815	$>$	5.991
do not fit	**is insufficient**	**fit**	6.251								

Adapted with permission from UCLES, A Level, January 2003 S3, Paper 2643, Question 5

Gameboard:

STEM SMART Double Maths 50 - Further Statistics

Revision

t-tests: Milk

Further A

A farmer decided to test the effectiveness of a food additive for cows. She chose a random sample of 12 of her cows and recorded their milk yields over a period of one week. She then included the additive in the cows' food for a period of several weeks and recorded their milk yields again in the last of these weeks. The increases in milk yields, $x\,\text{litres}$, for the 12 cows are summarised by $\Sigma x = 15.4$ and $\Sigma x^2 = 63.88$.

Part A Estimate the population variance

Calculate an unbiased estimate of the population variance for the increase in milk yield of one of the farmer's cows, s_D. Give your answer to 3 d.p.

Part B Null and alternative hypotheses

Let the mean increase in milk yield be μ_D. State the null and alternative hypotheses for a t-test to test, at the 1% significance level, whether there has been an increase in the mean milk yield.

H_0 : ☐ H_1 : ☐

Items:

$\mu_D < 0$ $\mu_D > 0$ $\mu_D = 0$ $\mu \neq 0$

Part C Carrying out the test

Carry out the test, at the 1% significance level. Fill in the gaps below and state your conclusion.

The sample mean, $\overline{x}$, is [ ].

The test statistic is given by the expression $t = \dfrac{\overline{x} - \mu_D}{\frac{s_D}{\sqrt{n}}}$. The value of the test statistic is [ ].

The number of degrees of freedom is [ ]. For a one-tailed test at the 1% significance level, the critical value is [ ].

The value of the test statistic is [ ] the critical value. Therefore, [ ] the null hypothesis. There [ ] significant evidence that the additive has increased the average milk yield.

Items:

13	11	**less than**	3.055	**greater than**	2.220	**is not**	7.690	12	1.283	**do not reject**
reject	**is**	2.681	2.718	1.4						

Part D A necessary assumption

State an assumption needed to carry out your test.

In order to carry out the test, it has to be assumed that the increases in the milk yields of the cows have a [ ] distribution.

Items:

uniform	**binomial**	**normal**	χ^2	**Poisson**	**geometric**

Adapted with permission from UCLES, A Level, June 2001 S3, Paper 2643, Question 4

Gameboard:

STEM SMART Double Maths 50 - Further Statistics Revision

t-tests: Oak Leaves

Further A

The leaves from oak trees growing in two different areas A and B are being measured. The lengths, in cm, of a random sample of 7 oak leaves from area A are:

$$6.2, \quad 8.3, \quad 7.8, \quad 9.3, \quad 10.2, \quad 8.4, \quad 7.2$$

Part A Confidence interval

Assuming that the distribution is normal, find a 95% confidence interval for the mean length of oak leaves from area A.

Give the lower confidence limit.

Give the upper confidence limit.

Part B Assumptions

The lengths, in cm, of a random sample of 5 oak leaves from area B are:

$$5.9, \qquad 7.4, \qquad 6.8, \qquad 8.2, \qquad 8.7$$

State two assumptions needed to carry out a suitable t-test on the difference between the mean lengths of oak leaves from areas A and B.

- [ ] The length of oak leaves from area B follow a uniform distribution.
- [ ] The lengths of oak leaves from areas A and B have equal means.
- [ ] The length of oak leaves from area B follow a normal distribution.
- [ ] The lengths of oak leaves from areas A and B have equal variances.
- [ ] The length of oak leaves from area B follow a geometric distribution.

Part C Hypothesis test

Test, at the 5% significance level, whether the mean length of oak leaves from area A is greater than the mean length of oak leaves from area B.

The null and alternative hypotheses are:

$$\mathrm{H}_0 : \mu_a \ \square \ \mu_b \qquad \mathrm{H}_1 : \mu_a \ \square \ \mu_b$$

Calculating the difference as $\overline{a} - \overline{b}$, the test statistic, $t =$ $\square$. The critical value is $t_{\mathrm{crit}} =$ $\square$.

Comparing these, we find that t $\square$ t_{crit}.

Therefore we $\square$ H_0 at the 5% level. There $\square$ evidence to suggest that the mean length of oak leaves from area A is greater than the mean length of oak leaves from area B.

Items:

reject	1.137	**is insufficient**	1.102	$=$	$>$	1.782	1.140	1.208	2.228	1.812	**is**	$<$
0.889	$\neq$	2.179	**do not reject**									
