A LEVEL Cambridge Topical Past Papers

FURTHER STATISTICS P4

2020 — 2025

1 - (9231/42_Winter_2024_Q3) ANSWER

Rosie sows 5 seeds in each of 150 plant pots. The number of seeds that germinate is recorded for each pot. The results are summarised in the following table.

Number of seeds that germinate	0	1	2	3	4	5
Number of pots	12	40	43	35	16	4

Rosie suggests that the number of seeds that germinate follows the binomial distribution $B(5, p)$.

(a) Use Rosie's results to show that $p = 0.42$. [1]

(b) Carry out a goodness of fit test, at the 10% significance level, to test whether the distribution $B(5, 0.42)$ is a good fit for the data. [9]

2 - (9231/43_Winter_2024_Q3) ANSWER

A statistician believes that the number of telephone calls received by an advice centre in a 10-minute interval can be modelled by the Poisson distribution $Po(1.9)$. The number of calls received in a randomly chosen 10-minute interval was recorded on each of 100 days. The results are summarised in the table, together with some of the expected frequencies corresponding to the distribution $Po(1.9)$.

Number of calls	0	1	2	3	4	5	6 or more
Observed frequency	10	18	35	21	11	4	1
Expected frequency	14.957	28.418	26.997				1.322

(a) Complete the table. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to determine whether the statistician's belief is reasonable. [6]

1 - (9231/42_Winter_2024_Q3) QUESTION

(a)	$\bar{x} = \dfrac{40+86+105+64+20}{150} = \dfrac{315}{150} = 2.1, \qquad p = \dfrac{2.1}{5} = 0.42$	**B1**	Must see either 315 or 2.1. AG
		1	
(b)	<table><tr><td>Number of seeds that germinate</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>Number of pots</td><td>12</td><td>40</td><td>43</td><td>35</td><td>16</td><td>4</td></tr><tr><td>Expected frequency</td><td>9.846</td><td>35.6475</td><td>51.627</td><td>37.3845</td><td>13.536</td><td>1.9605</td></tr></table>	**B1** **B1**	Calculate expected frequencies (must be seen) at least 2 correct to at least 2 decimal places. At least 4 correct to at least 2 decimal places.
	Combine last two columns	**M1**	20, 15.50, may be implied by answer 3.90 – 3.91.
	Chi-squared contributions: 0.4712 0.5314 1.4416 0.1521 1.3088	**M1**	At least 2 correct, may be implied by answer 3.90 – 3.91.
	Test statistic = 3.905	**A1**	accept 3.90 – 3.910
	H_0: Binomial B(5, 0.42) fits the data H_1: Binomial B(5, 0.42) does not fit the data	**B1**	Allow 'Binomial' for 'B(5, 0.42)' Allow 'Number of seeds that germinate can be modelled by B(5, 0.42)'
	Critical value is 6.251	**B1**	Must come from combined columns. Allow 7.779.
	'3.905' < '6.251' Accept H_0	**M1**	Reject H_1, not significant.
	Insufficient evidence to suggest that B(5, 0.42) is a not a good fit (to the data)	**A1**	Correct work only, **including hypotheses**, level of uncertainty in language.
		9	

2 - (9231/43_Winter_2024_Q3)

<table>
<tr><td rowspan="3">(a)</td><td rowspan="2">17.098 8.122 3.086</td><td>B1</td><td>One correct.</td></tr>
<tr><td>B1</td><td>All correct.</td></tr>
<tr><td></td><td>2</td><td></td></tr>
<tr><td rowspan="7">(b)</td><td><table>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4 or more</td></tr>
<tr><td>10</td><td>18</td><td>35</td><td>21</td><td>16</td></tr>
<tr><td>14.957</td><td>28.418</td><td>26.997</td><td>17.098</td><td>12.53</td></tr>
</table></td><td>M1</td><td>Last two or three columns combined.</td></tr>
<tr><td>Contributions to test statistic are:
1.6428 3.8192 2.3724 0.8905 0.9609(7)</td><td>M1</td><td>May be implied by awrt 9.69</td></tr>
<tr><td>Test statistic is 9.69</td><td>A1</td><td>9.686</td></tr>
<tr><td>H_0: Po(1.9) is a good fit for the data
H_1: Po(1.9) is not a good fit for the data</td><td>B1</td><td></td></tr>
<tr><td>Critical value is 7.779, compare ‘9.69’ > 7.779 reject H_0</td><td>M1</td><td>4 degrees of freedom</td></tr>
<tr><td>Sufficient evidence to suggest that Po(1.9) is a not a good fit for the data/
Sufficient evidence to reject/not support the statistician’s claim</td><td>A1</td><td>Correct work only, including hypotheses, in context, level of uncertainty in language.</td></tr>
<tr><td></td><td>6</td><td></td></tr>
</table>

1 - (9231/41_Summer_2020_Q4) ANSWER

A company has two different machines, X and Y, each of which fills empty cups with coffee. The manager is investigating the volumes of coffee, x and y, measured in appropriate units, in the cups filled by machines X and Y respectively. She chooses a random sample of 50 cups filled by machine X and a random sample of 40 cups filled by machine Y. The volumes are summarised as follows.

$$\sum x = 15.2 \qquad \sum x^2 = 5.1 \qquad \sum y = 13.4 \qquad \sum y^2 = 4.8$$

The manager claims that there is no difference between the mean volume of coffee in cups filled by machine X and the mean volume of coffee in cups filled by machine Y.

Test the manager's claim at the 10% significance level. [9]

2 - (9231/41_Summer_2020_Q5) ANSWER

A large number of children are competing in a throwing competition. The distances, in metres, thrown by a random sample of 8 children are as follows.

19.8 22.1 24.4 21.5 20.8 26.3 23.7 25.0

(a) Assuming that distances are normally distributed, test, at the 5% significance level, whether the population mean distance thrown is more than 22.0 metres. [7]

(b) Find a 95% confidence interval for the population mean distance thrown. [3]

3 - (9231/42_Summer_2020_Q4) ANSWER

A company has two different machines, X and Y, each of which fills empty cups with coffee. The manager is investigating the volumes of coffee, x and y, measured in appropriate units, in the cups filled by machines X and Y respectively. She chooses a random sample of 50 cups filled by machine X and a random sample of 40 cups filled by machine Y. The volumes are summarised as follows.

$$\sum x = 15.2 \qquad \sum x^2 = 5.1 \qquad \sum y = 13.4 \qquad \sum y^2 = 4.8$$

The manager claims that there is no difference between the mean volume of coffee in cups filled by machine X and the mean volume of coffee in cups filled by machine Y.

Test the manager's claim at the 10% significance level. [9]

4 - (9231/42_Summer_2020_Q5) ANSWER

A large number of children are competing in a throwing competition. The distances, in metres, thrown by a random sample of 8 children are as follows.

19.8 22.1 24.4 21.5 20.8 26.3 23.7 25.0

(a) Assuming that distances are normally distributed, test, at the 5% significance level, whether the population mean distance thrown is more than 22.0 metres. [7]

(b) Find a 95% confidence interval for the population mean distance thrown. [3]

5 - (9231/43_Summer_2020_Q2) ANSWER

A random sample of 40 observations of a random variable X and a random sample of 50 observations of a random variable Y are taken. The resulting values for the sample means, $\bar{x}$ and $\bar{y}$, and the unbiased estimates, s_x^2 and s_y^2, for the population variances are as follows.

$$\bar{x} = 24.4 \qquad \bar{y} = 17.2 \qquad s_x^2 = 10.2 \qquad s_y^2 = 11.1$$

Find a 90% confidence interval for the difference between the population means of X and Y. [5]

6 - (9231/43_Summer_2020_Q5) ANSWER

Students at two colleges, A and B, are competing in a computer games challenge.

(a) The time taken for a randomly chosen student from college A to complete the challenge has a normal distribution with mean μ minutes. The times taken, x minutes, are recorded for a random sample of 10 students chosen from college A. The results are summarised as follows.

$$\Sigma x = 828 \qquad \Sigma x^2 = 68622$$

A test is carried out on the data at the 5% significance level and the result supports the claim that $\mu > k$.

Find the greatest possible value of k. [4]

(b) A random sample of 8 students is chosen from college B. Their times to complete the same challenge give a sample mean of 79.8 minutes and an unbiased variance estimate of 9.966 minutes^2.

Use a 2-sample test at the 5% significance level to test whether the mean time for students at college B to complete the challenge is the same as the mean time for students at college A to complete the challenge. You should assume that the two distributions are normal and have the same population variance. [7]

1 - (9231/41_Summer_2020_Q4) QUESTION

	H_0: $\mu_x = \mu_y$ H_1: $\mu_x \neq \mu_y$	**B1**
	$s_x^2 = \frac{1}{49}\left(5.1 - \frac{15.2^2}{50}\right) = 0.0097796$: $s_y^2 = \frac{1}{39}\left(4.8 - \frac{13.4^2}{40}\right) = 0.007974$	**M1A1**
	$s^2 = \frac{0.00977959}{50} + \frac{0.007974}{40} = 0.0003949$	**M1A1**
	$z = \frac{0.304 - 0.335}{\sqrt{0.0003949}} = (-)1.56$	**M1A1**
	Compare with 1.645	**M1**
	Accept H_0: insufficient evidence to reject manager's claim	**A1**
		9

2 - (9231/41_Summer_2020_Q5) QUESTION

(a)	$\sum x = 183.6,\ \sum x^2 = 4249.08,\ \ \bar{x} = 22.95$	**B1**
	$s^2 = \frac{1}{7}\left(4249.08 - \frac{183.6^2}{8}\right) = 5.066$	**M1**
	H_0: $\mu = 22.0$, H_1: $\mu > 22.0$	**B1**
	$t = \frac{22.95 - 22.0}{\sqrt{\frac{s^2}{8}}} = 1.194$	**M1A1**
	Compare t with correct tabular value 1.895	**M1**
	Accept H_0: mean distance thrown is not more than 22.0 m	**A1**
		7
(b)	$22.95 \pm t\sqrt{\frac{s^2}{8}}$	**M1**
	With t = 2.365	**B1**
	[21.1, 24.8]	**A1**
		3

3 - (9231/42_Summer_2020_Q4) QUESTION

	H_0: $\mu_x = \mu_y$ H_1: $\mu_x \neq \mu_y$	**B1**
	$s_x^2 = \frac{1}{49}\left(5.1 - \frac{15.2^2}{50}\right) = 0.0097796$: $s_y^2 = \frac{1}{39}\left(4.8 - \frac{13.4^2}{40}\right) = 0.007974$	**M1A1**
	$s^2 = \frac{0.00977959}{50} + \frac{0.007974}{40} = 0.0003949$	**M1A1**
	$z = \frac{0.304 - 0.335}{\sqrt{0.0003949}} = (-)1.56$	**M1A1**
	Compare with 1.645	**M1**
	Accept H_0: insufficient evidence to reject manager's claim	**A1**
		9

4 - (9231/42_Summer_2020_Q5) QUESTION

(a)	$\sum x = 183.6,\ \sum x^2 = 4249.08,\ \ \bar{x} = 22.95$	**B1**
	$s^2 = \frac{1}{7}\left(4249.08 - \frac{183.6^2}{8}\right) = 5.066$	**M1**
	H_0: $\mu = 22.0$, H_1: $\mu > 22.0$	**B1**
	$t = \frac{22.95 - 22.0}{\sqrt{\frac{s^2}{8}}} = 1.194$	**M1A1**
	Compare t with correct tabular value 1.895	**M1**
	Accept H_0: mean distance thrown is not more than 22.0 m	**A1**
		7
(b)	$22.95 \pm t\sqrt{\frac{s^2}{8}}$	**M1**
	With $t = 2.365$	**B1**
	[21.1, 24.8]	**A1**
		3

5 - (9231/43_Summer_2020_Q2) QUESTION

	$s^2 = \frac{10.2}{40} + \frac{11.1}{50} = 0.477$	**M1A1**
	CI = $(24.4 - 17.2) \pm zs$	**M1**
	= $(24.4 - 17.2) \pm 1.645\sqrt{0.477}$	**A1**
	= [6.06, 8.34]	**A1**
		5

6 - (9231/43_Summer_2020_Q5)

(a)	$\overline{x} = \frac{828}{10} = 82.8 \quad s^2 = \frac{1}{9}\left(68622 - \frac{828^2}{10}\right) = 7.0667$	**B1**
	$\frac{82.8 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ where $t = 1.833$ (**M1** if equality, **A1** for inequality and correct t value)	**M1 A1**
	$k \leqslant 81.259 \quad k \leqslant 81.3$	**A1**
		4
(b)	H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$	**B1**
	Pooled variance = $\frac{9 \times 7.0667 + 7 \times 9.966}{10 + 8 - 2} = s_p{}^2$	**M1**
	= 8.335	**A1**
	$t = \frac{82.8 - 79.8}{s_p\sqrt{\frac{1}{10} + \frac{1}{8}}} = 2.19$	**M1A1**
	Compare with 2.12 ($t(16, 0.975)$) and reject H_0	**M1**
	Population means are not the same	**A1**
		7