A LEVEL Cambridge Topical Past Papers

# FURTHER STATISTICS P4

2020 — 2025




EXAMINENT
www.examinent.com

1 - (9231/42_Winter_2024_Q3)

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)

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)

<table>
<tr><td rowspan="2">(a)</td><td>$\bar{x} = \frac{40+86+105+64+20}{150} = \frac{315}{150} = 2.1, \quad p = \frac{2.1}{5} = 0.42$</td><td>B1</td><td>Must see either 315 or 2.1.<br>AG</td></tr>
<tr><td></td><td>1</td><td></td></tr>
<tr><td rowspan="10">(b)</td><td rowspan="2"><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></td><td>B1</td><td>Calculate expected frequencies (must be seen) at least 2 correct to at least 2 decimal places.</td></tr>
<tr><td>B1</td><td>At least 4 correct to at least 2 decimal places.</td></tr>
<tr><td>Combine last two columns</td><td>M1</td><td>20, 15.50, may be implied by answer 3.90 – 3.91.</td></tr>
<tr><td>Chi-squared contributions:<br>0.4712 0.5314 1.4416 0.1521 1.3088</td><td>M1</td><td>At least 2 correct, may be implied by answer 3.90 – 3.91.</td></tr>
<tr><td>Test statistic = 3.905</td><td>A1</td><td>accept 3.90 – 3.910</td></tr>
<tr><td>$H_0$: Binomial B(5, 0.42) fits the data<br>$H_1$: Binomial B(5, 0.42) does not fit the data</td><td>B1</td><td>Allow ‘Binomial’ for ‘B(5, 0.42)’<br>Allow ‘Number of seeds that germinate can be modelled by B(5, 0.42)’</td></tr>
<tr><td>Critical value is 6.251</td><td>B1</td><td>Must come from combined columns.<br>Allow 7.779.</td></tr>
<tr><td>‘3.905’ < ‘6.251’ Accept $H_0$</td><td>M1</td><td>Reject $H_1$, not significant.</td></tr>
<tr><td>Insufficient evidence to suggest that B(5, 0.42) is a not a good fit (to the data)</td><td>A1</td><td>Correct work only, including hypotheses, level of uncertainty in language.</td></tr>
<tr><td></td><td>9</td><td></td></tr>
</table>

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:<br>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<br>$H_1$: 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/<br>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)

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)

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)

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)

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)

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)

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 $\mu$ 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]

7 - (9231/41_Winter_2020_Q1)

Kayla is investigating the lengths of the leaves of a certain type of tree found in two forests $X$ and $Y$. She chooses a random sample of 40 leaves of this type from forest $X$ and records their lengths, $x$ cm. She also records the lengths, $y$ cm, for a random sample of 60 leaves of this type from forest $Y$. Her results are summarised as follows.

$$\sum x = 242.0 \qquad \sum x^2 = 1587.0 \qquad \sum y = 373.2 \qquad \sum y^2 = 2532.6$$

Find a 90% confidence interval for the difference between the population mean lengths of leaves in forests $X$ and $Y$. [7]

8 - (9231/41_Winter_2020_Q4)

Members of the Sprints athletics club have been taking part in an intense training scheme, aimed at reducing their times taken to run 400 m. For a random sample of 9 athletes from the club, the times taken, in seconds, before and after the training scheme are given in the following table.

| Athlete | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $I$ |
|---|---|---|---|---|---|---|---|---|---|
| Time before | 48.8 | 48.2 | 50.3 | 49.6 | 49.4 | 48.9 | 47.6 | 50.3 | 48.4 |
| Time after | 47.9 | 47.8 | 49.6 | 49.1 | 49.6 | 48.9 | 47.7 | 49.1 | 48.1 |

The organiser of the training scheme claims that on average an athlete's time will be reduced by at least 0.3 seconds.

Test at the 10% significance level whether the organiser's claim is justified, stating any assumption that you make. [8]

9 - (9231/42_Winter_2020_Q1)

The heights of the members of a large sports club are normally distributed. A random sample of 11 members of the club is chosen and their heights, $x$ cm, are measured. The results are summarised as follows, where $\bar{x}$ denotes the sample mean of $x$.

$$\bar{x} = 176.2 \qquad \sum(x - \bar{x})^2 = 313.1$$

Test, at the 5% significance level, the null hypothesis that the population mean height for members of this club is equal to 172.5 cm against the alternative hypothesis that the mean differs from 172.5 cm. [5]

10 - (9231/42_Winter_2020_Q6)

Nassa is researching the lengths of a particular type of snake in two countries, $A$ and $B$.

**(a)** He takes a random sample of 10 snakes of this type from country $A$ and measures the length, $x$ m, of each snake. He then calculates a 90% confidence interval for the population mean length, $\mu$ m, for snakes of this type, assuming that snake lengths have a normal distribution. This confidence interval is $3.36 \leqslant \mu \leqslant 4.22$.

Find the sample mean and an unbiased estimate for the population variance. [4]

**(b)** Nassa also measures the lengths, $y$ m, of a random sample of 8 snakes of this type taken from country $B$. His results are summarised as follows.

$$\Sigma y = 27.86 \qquad \Sigma y^2 = 98.02$$

Nassa claims that the mean length of snakes of this type in country $B$ is less than the mean length of snakes of this type in country $A$. Nassa assumes that his sample from country $B$ also comes from a normal distribution, with the same variance as the distribution from country $A$.

Test at the 10% significance level whether there is evidence to support Nassa's claim. [8]

11 - (9231/43_Winter_2020_Q1)

Kayla is investigating the lengths of the leaves of a certain type of tree found in two forests $X$ and $Y$. She chooses a random sample of 40 leaves of this type from forest $X$ and records their lengths, $x$ cm. She also records the lengths, $y$ cm, for a random sample of 60 leaves of this type from forest $Y$. Her results are summarised as follows.

$$\Sigma x = 242.0 \qquad \Sigma x^2 = 1587.0 \qquad \Sigma y = 373.2 \qquad \Sigma y^2 = 2532.6$$

Find a 90% confidence interval for the difference between the population mean lengths of leaves in forests $X$ and $Y$. [7]

12 - (9231/43_Winter_2020_Q4)

Members of the Sprints athletics club have been taking part in an intense training scheme, aimed at reducing their times taken to run 400 m. For a random sample of 9 athletes from the club, the times taken, in seconds, before and after the training scheme are given in the following table.

| Athlete | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $I$ |
|---|---|---|---|---|---|---|---|---|---|
| Time before | 48.8 | 48.2 | 50.3 | 49.6 | 49.4 | 48.9 | 47.6 | 50.3 | 48.4 |
| Time after | 47.9 | 47.8 | 49.6 | 49.1 | 49.6 | 48.9 | 47.7 | 49.1 | 48.1 |

The organiser of the training scheme claims that on average an athlete's time will be reduced by at least 0.3 seconds.

Test at the 10% significance level whether the organiser's claim is justified, stating any assumption that you make. [8]

13 - (9231/41_Summer_2021_Q1)

A random sample of 7 observations of a variable $X$ are as follows.

8.26 7.78 7.92 8.04 8.27 7.95 8.34

The population mean of $X$ is $\mu$.

**(a)** Test, at the 10% significance level, the null hypothesis $\mu = 8.22$ against the alternative hypothesis $\mu < 8.22$. [6]

**(b)** State an assumption necessary for the test in part **(a)** to be valid. [1]

14 - (9231/41_Summer_2021_Q4)

A scientist is investigating the lengths of the leaves of birch trees in different regions. He takes a random sample of 50 leaves from birch trees in region $A$ and a random sample of 60 leaves from birch trees in region $B$. He records their lengths in cm, $x$ and $y$, respectively. His 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 $A$ and $B$ are $\mu_A$ cm and $\mu_B$ cm respectively.

Carry out a test at the 5% significance level to test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

15 - (9231/42_Summer_2021_Q1)

A random sample of 7 observations of a variable $X$ are as follows.

8.26 7.78 7.92 8.04 8.27 7.95 8.34

The population mean of $X$ is $\mu$.

**(a)** Test, at the 10% significance level, the null hypothesis $\mu = 8.22$ against the alternative hypothesis $\mu < 8.22$. [6]

**(b)** State an assumption necessary for the test in part **(a)** to be valid. [1]

16 - (9231/42_Summer_2021_Q4)

A scientist is investigating the lengths of the leaves of birch trees in different regions. He takes a random sample of 50 leaves from birch trees in region $A$ and a random sample of 60 leaves from birch trees in region $B$. He records their lengths in cm, $x$ and $y$, respectively. His 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 $A$ and $B$ are $\mu_A$ cm and $\mu_B$ cm respectively.

Carry out a test at the 5% significance level to test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

17 - (9231/43_Summer_2021_Q1)

Farmer $A$ grows apples of a certain variety. Each tree produces 14.8 kg of apples, on average, per year. Farmer $B$ grows apples of the same variety and claims that his apple trees produce a higher mass of apples per year than Farmer $A$'s trees. The masses of apples from Farmer $B$'s trees may be assumed to be normally distributed.

A random sample of 10 trees from Farmer $B$ is chosen. The masses, $x$ kg, of apples produced in a year are summarised as follows.

$$\sum x = 152.0 \qquad \sum x^2 = 2313.0$$

Test, at the 5% significance level, whether Farmer $B$'s claim is justified. [6]

18 - (9231/43_Summer_2021_Q3)

The heights, $x$ m, of a random sample of 50 adult males from country $A$ were recorded. The heights, $y$ m, of a random sample of 40 adult males from country $B$ were also recorded. The results are summarised as follows.

$$\sum x = 89.0 \qquad \sum x^2 = 159.4 \qquad \sum y = 67.2 \qquad \sum y^2 = 113.1$$

Find a 95% confidence interval for the difference between the mean heights of adult males from country $A$ and adult males from country $B$. [8]

19 - (9231/41_Winter_2021_Q1)

The times taken for students at a college to run 200 m have a normal distribution with mean $\mu$ s. The times, $x$ s, are recorded for a random sample of 10 students from the college. The results are summarised as follows, where $\bar{x}$ is the sample mean.

$$\bar{x} = 25.6 \qquad \Sigma(x-\bar{x})^2 = 78.5$$

**(a)** Find a 90% confidence interval for $\mu$. [4]

A test of the null hypothesis $\mu = k$ is carried out on this sample, using a 10% significance level. The test does not support the alternative hypothesis $\mu < k$.

**(b)** Find the greatest possible value of $k$. [3]

20 - (9231/41_Winter_2021_Q4)

Manet has developed a new training course to help athletes improve their time taken to run 800 m. Manet claims that his course will decrease an athlete's time by more than 2 s on average. For a random sample of 10 athletes the times taken, in seconds, before and after the course are given in the following table.

| Athlete | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| Before | 150 | 146 | 131 | 135 | 126 | 142 | 130 | 129 | 137 | 134 |
| After | 145 | 138 | 129 | 135 | 122 | 135 | 132 | 128 | 127 | 137 |

Use a $t$-test, at the 5% significance level, to test whether Manet's claim is justified, stating any assumption that you make. [8]

21 - (9231/42_Winter_2021_Q1)

The number, $x$, of pine trees was counted in each of 40 randomly chosen regions of equal size in country $A$. The number, $y$, of pine trees was counted in each of 60 randomly chosen regions of the same equal size in country $B$. The results are summarised as follows.

$$\Sigma x = 752 \qquad \Sigma x^2 = 14\,320 \qquad \Sigma y = 1548 \qquad \Sigma y^2 = 40\,200$$

Find a 95% confidence interval for the difference between the mean number of pine trees in regions of this size in countries $A$ and $B$. [7]

22 - (9231/42_Winter_2021_Q6)

A scientist is investigating the masses of a particular type of fish found in lakes $A$ and $B$. He chooses a random sample of 10 fish of this type from lake $A$ and records their masses, $x$ kg, as follows.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2.1 | 1.8 | 0.9 | 3.0 | 2.4 | 2.6 | 1.8 | 2.2 | 1.9 | 2.5 |

The scientist also chooses a random sample of 12 fish of this type from lake $B$, but he only has a summary of their masses, $y$ kg, as follows.

$$\Sigma y = 24.48 \qquad \Sigma y^2 = 53.75$$

Test at the 10% significance level whether the mean mass of fish of this type in lake $A$ is greater than the mean mass of fish of this type in lake $B$. You should state any assumptions that you need to make for the test to be valid. [10]

23 - (9231/43_Winter_2021_Q1)

The times taken for students at a college to run 200 m have a normal distribution with mean $\mu$ s. The times, $x$ s, are recorded for a random sample of 10 students from the college. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 25.6 \qquad \Sigma(x-\overline{x})^2 = 78.5$$

**(a)** Find a 90% confidence interval for $\mu$. [4]

$\mu = k$ is carried out on this sample, using a 10% significance level. The test does not support the alternative hypothesis $\mu < k$.

**(b)** Find the greatest possible value of $k$. [3]

24 - (9231/43_Winter_2021_Q4)

Manet has developed a new training course to help athletes improve their time taken to run 800 m. Manet claims that his course will decrease an athlete's time by more than 2 s on average. For a random sample of 10 athletes the times taken, in seconds, before and after the course are given in the following table.

| Athlete | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $I$ | $J$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Before | 150 | 146 | 131 | 135 | 126 | 142 | 130 | 129 | 137 | 134 |
| After | 145 | 138 | 129 | 135 | 122 | 135 | 132 | 128 | 127 | 137 |

Use a $t$-test, at the 5% significance level, to test whether Manet's claim is justified, stating any assumption that you make. [8]

25 - (9231/41_Summer_2022_Q1)

A manager is investigating the times taken by employees to complete a particular task as a result of the introduction of new technology. He claims that the mean time taken to complete the task is reduced by more than 0.4 minutes. He chooses a random sample of 10 employees. The times taken, in minutes, before and after the introduction of the new technology are recorded in the table.

| Employee | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| Time before new technology | 10.2 | 9.8 | 12.4 | 11.6 | 10.8 | 11.2 | 14.6 | 10.6 | 12.3 | 11.0 |
| Time after new technology | 9.6 | 8.5 | 12.4 | 10.9 | 10.2 | 10.6 | 12.8 | 10.8 | 12.5 | 10.6 |

**(a)** Test at the 10% significance level whether the manager's claim is justified. [7]

**(b)** State an assumption that is necessary for this test to be valid. [1]

26 - (9231/41_Summer_2022_Q5)

Raman is researching the heights of male giraffes in a particular region. Raman assumes that the heights of male giraffes in this region are normally distributed. He takes a random sample of 8 male giraffes from the region and measures the height, in metres, of each giraffe. These heights are as follows.

5.2 5.8 4.9 6.1 5.5 5.9 5.4 5.6

**(a)** Find a 90% confidence interval for the population mean height of male giraffes in this region. [5]

Raman claims that the population mean height of male giraffes in the region is less than 5.9 metres.

**(b)** Test at the 2.5% significance level whether this sample provides sufficient evidence to support Raman's claim. [4]

27 - (9231/42_Summer_2022_Q1)

A manager is investigating the times taken by employees to complete a particular task as a result of the introduction of new technology. He claims that the mean time taken to complete the task is reduced by more than 0.4 minutes. He chooses a random sample of 10 employees. The times taken, in minutes, before and after the introduction of the new technology are recorded in the table.

| Employee | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| Time before new technology | 10.2 | 9.8 | 12.4 | 11.6 | 10.8 | 11.2 | 14.6 | 10.6 | 12.3 | 11.0 |
| Time after new technology | 9.6 | 8.5 | 12.4 | 10.9 | 10.2 | 10.6 | 12.8 | 10.8 | 12.5 | 10.6 |

**(a)** Test at the 10% significance level whether the manager's claim is justified. [7]

**(b)** State an assumption that is necessary for this test to be valid. [1]

28 - (9231/42_Summer_2022_Q5)

Raman is researching the heights of male giraffes in a particular region. Raman assumes that the heights of male giraffes in this region are normally distributed. He takes a random sample of 8 male giraffes from the region and measures the height, in metres, of each giraffe. These heights are as follows.

5.2 5.8 4.9 6.1 5.5 5.9 5.4 5.6

**(a)** Find a 90% confidence interval for the population mean height of male giraffes in this region. [5]

Raman claims that the population mean height of male giraffes in the region is less than 5.9 metres.

**(b)** Test at the 2.5% significance level whether this sample provides sufficient evidence to support Raman's claim. [4]

29 - (9231/43_Summer_2022_Q1)

The times taken by members of a large quiz club to complete a challenge have a normal distribution with mean $\mu$ minutes. The times, $x$ minutes, are recorded for a random sample of 8 members of the club. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 33.8 \qquad \Sigma(x-\overline{x})^2 = 94.5$$

Find a 95% confidence interval for $\mu$. [4]

30 - (9231/43_Summer_2022_Q6)

A company has two machines, $A$ and $B$, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines $A$ and $B$ are denoted by $x$ and $y$ respectively. A scientist at the company takes a random sample of 40 bottles filled by machine $A$ and a random sample of 50 bottles filled by machine $B$. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines $A$ and $B$ are denoted by $\mu_A$ and $\mu_B$.

**(a)** Test at the 2% significance level whether there is any difference between $\mu_A$ and $\mu_B$. [8]

**(b)** Find the set of values of $\alpha$ for which there would be evidence at the $\alpha$% significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

31 - (9231/41_Winter_2022_Q1)

Jasmine is researching the heights of pine trees in forests in two regions $A$ and $B$. She chooses a random sample of 50 pine trees in region $A$ and records their heights, $x$ m. She also chooses a random sample of 60 pine trees in region $B$ and records their heights, $y$ m. Her results are summarised as follows.

$$\sum x = 1625 \qquad \sum x^2 = 53200 \qquad \sum y = 1854 \qquad \sum y^2 = 57900$$

Find a 95% confidence interval for the difference between the population mean heights of pine trees in regions $A$ and $B$. [7]

32 - (9231/41_Winter_2022_Q6)

A company manufactures copper pipes. The pipes are produced by two different machines, $A$ and $B$. An inspector claims that the mean diameter of the pipes produced by machine $A$
mean diameter of the pipes produced by machine $B$. He takes a random sample of 12 pipes produced by machine $A$ and measures their diameters, $x$ cm. His results are summarised as follows.

$$\sum x = 6.24 \qquad \sum x^2 = 3.26$$

He also takes a random sample of 10 pipes produced by machine $B$ and measures their diameters in cm. His results are as follows.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.48 | 0.53 | 0.47 | 0.54 | 0.54 | 0.55 | 0.46 | 0.55 | 0.50 | 0.48 |

equal population variances.

Test at the 2.5% significance level whether the data supports the inspector's claim. [9]

33 - (9231/42_Winter_2022_Q1)

A basketball club has a large number of players. The heights, $x$ m, of a random sample of 10 of these players are measured. A 90% confidence interval for the population mean height, $\mu$ m, of players in this club is calculated. It is assumed that heights are normally distributed. The confidence interval is $1.78 \leqslant \mu \leqslant 2.02$.

Find the values of $\sum x$ and $\sum x^2$ for this sample. [6]

34 - (9231/42_Winter_2022_Q3)

A scientist is investigating the masses of birds of a certain species in country $X$ and country $Y$. She takes a random sample of 50 birds of this species from country $X$ and a random sample of 80 birds of this species from country $Y$. She records their masses in kg, $x$ and $y$, respectively. Her results are summarised as follows.

$$\sum x = 75.5 \qquad \sum x^2 = 115.2 \qquad \sum y = 116.8 \qquad \sum y^2 = 172.6$$

The population mean masses of these birds in countries $X$ and $Y$ are $\mu_x$ kg and $\mu_y$ kg respectively.

Test, at the 5% significance level, the null hypothesis $\mu_x = \mu_y$ against the alternative hypothesis $\mu_x > \mu_y$. State your conclusion in the context of the question. [8]

1 - (9231/41_Summer_2020_Q4)

| | | |
|---|---|---|
| | $H_0$: $\mu_x = \mu_y$ $H_1$: $\mu_x \neq \mu_y$ | **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)

| | | |
|---|---|---|
| (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)

| | | |
|---|---|---|
| | H0: $\mu_x = \mu_y$ H1: $\mu_x \neq \mu_y$ | **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 H0: insufficient evidence to reject manager's claim | **A1** |
| | | **9** |

4 - (9231/42_Summer_2020_Q5)

| | | |
|---|---|---|
| (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)

| | | |
|---|---|---|
| | $s^2 = \frac{10.2}{40} + \frac{11.1}{50} = 0.477$ | **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) | $\bar{x} = \frac{828}{10} = 82.8 \quad s^2 = \frac{1}{9}\left(68622 - \frac{828^2}{10}\right) = 7.0667$ | **B1** |
| | $\frac{82.8 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ where $t = 1.833$<br>(**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** |

7 - (9231/41_Winter_2020_Q1)

| | | | |
|---|---|---|---|
| | $s_x{}^2 = \frac{1}{39}\left(1587.0 - \frac{242.0^2}{40}\right) = 3.15128 \quad \left(= \frac{1229}{390}\right)$<br>$s_y{}^2 = \frac{1}{59}\left(2532.6 - \frac{373.2^2}{60}\right) = 3.58129 \quad \left(= \frac{26412}{7375}\right)$ | **M1 A1** | Either unsimplified<br>Both correct |
| | $s^2 = \frac{3.15128}{40} + \frac{3.58129}{60} = 0.13847$ | **M1 A1** | Pooled variance is M0A0<br>or $s = 0.3721$ |
| | $6.05 - 6.22 \pm zs$ | **M1** | FT their *s*, must be a *z* value |
| | $= 0.17 \pm 1.645\sqrt{0.13847}$ | **A1** | With 1.645 |
| | = [− 0.442, 0.782] or [− 0.782, 0.442] | **A1** | |
| | | **7** | |

8 - (9231/41_Winter_2020_Q4)

| | | | |
|---|---|---|---|
| | Assume (population) differences are normally distributed | B1 | |
| | H$_0$: $\mu_X - \mu_Y = 0.3$ H$_1$: $\mu_X - \mu_Y > 0.3$ | B1 | |
| | Diff: $0.9\ 0.4\ 0.7\ 0.5 - 0.2\ 0 - 0.1\ 1.2\ 0.3$ | M1 | Signed differences |
| | $\sum d = 3.7,\ \sum d^2 = 3.29$<br>$\bar{d} = 0.411,\ s^2 = \frac{1}{8}\left(3.29 - \frac{3.7^2}{9}\right) = 0.2211$ | M1 | $s = 0.4702$ |
| | $t = \frac{0.411 - 0.3}{\sqrt{\frac{0.2211}{9}}}$<br>$= 0.708$ or $0.709$ | M1 A1 | |
| | ‘0.708’ < 1.397 | M1 | Compare *their* 0.708 with 1.397 |
| | Accept H$_0$<br>Insufficient evidence to support claim | A1 FT | In context, except possibly hypotheses<br>Level of uncertainty in language used. No contradictions |
| | | 8 | |

9 - (9231/42_Winter_2020_Q1)

| | | | |
|---|---|---|---|
| | $s^2 = \dfrac{313.1}{10} = 31.3$ | **B1** | Can be implied |
| | $t = \dfrac{176.2 - 172.5}{s/\sqrt{11}}$ | **M1** | |
| | = 2.19 | **A1** | |
| | '2.19' < 2.228 | **M1** | Comparison of their *t* value with 2.228 |
| | accept $H_0$:<br>insufficient evidence to reject mean height is 172.5<br>OR<br>sufficient evidence to accept mean height is 172.5 | **A1 FT** | Correct conclusion in context, level of uncertainty in language used. No contradictions.<br>CWO<br>Do not accept: 'mean height is 172.5'<br>Do not accept use of symbols that are undefined e.g. insufficient evidence to reject $\mu = 172.5$ |
| | | **5** | |

10 - (9231/42_Winter_2020_Q6)

| | | | |
|---|---|---|---|
| (a) | $\bar{x} = \frac{1}{2}(4.22 + 3.36) = 3.79$ | **B1** | |
| | $4.22 - 3.36 = \frac{2ts}{\sqrt{10}}$ | **M1** | Using $z$ value implies M0 |
| | With $t = 1.833$ | **A1** | |
| | ($s = 0.7418$) variance = 0.55(0) | **A1** | |
| | | **4** | |
| (b) | $H_0$: $\mu_A = \mu_B$ $H_1$: $\mu_A > \mu_B$ | **B1** | Not $\bar{x}$ |
| | $s^2 = \frac{1}{7}\left(98.02 - \frac{27.86^2}{8}\right) = 0.142(507)$ | **M1** | $= \frac{1}{7}(98.02 - 97.02)$ |
| | Pooled estimate = $\frac{9 \times 0.550316 + 7 \times 0.142507}{16}$ | **M1** | |
| | $= 0.372$ | **A1** | |
| | $t = \frac{\frac{27.86}{8} - 3.79}{\sqrt{0.372}\sqrt{\frac{1}{10} + \frac{1}{8}}} = (-)1.063$ | **M1A1** | |
| | '1.063' < 1.337 oe | **M1** | Compare *their* value with 1.337 |
| | Accept $H_0$<br>Nassa's claim is not supported | **A1ft** | Correct conclusion in context, level of uncertainty in language used. No contradictions. |
| | | **8** | |

11 - (9231/43_Winter_2020_Q1)

| | | | |
|---|---|---|---|
| | $s_x^{\ 2}=\frac{1}{39}\left(1587.0-\frac{242.0^2}{40}\right)=3.15128 \quad \left(=\frac{1229}{390}\right)$<br>$s_y^{\ 2}=\frac{1}{59}\left(2532.6-\frac{373.2^2}{60}\right)=3.58129 \quad \left(=\frac{26412}{7375}\right)$ | **M1 A1** | Either unsimplified<br>Both correct |
| | $s^2=\frac{3.15128}{40}+\frac{3.58129}{60}=0.13847$ | **M1 A1** | Pooled variance is M0A0<br>or $s=0.3721$ |
| | $6.05-6.22\pm zs$ | **M1** | FT their $s$, must be a $z$ value |
| | $=\ 0.17\pm1.645\sqrt{0.13847}$ | **A1** | With 1.645 |
| | = [− 0.442, 0.782] or [− 0.782, 0.442] | **A1** | |
| | | **7** | |

12 - (9231/43_Winter_2020_Q4)

| | | | |
|---|---|---|---|
| | Assume (population) differences are normally distributed | **B1** | |
| | $H_0$: $\mu_X - \mu_Y = 0.3$  $H_1$: $\mu_X - \mu_Y > 0.3$ | **B1** | |
| | Diff: $0.9\ 0.4\ 0.7\ 0.5 -0.2\ 0 -0.1\ 1.2\ 0.3$ | **M1** | Signed differences |
| | $\sum d = 3.7,\ \sum d^2 = 3.29$<br>$\bar{d} = 0.411,\ s^2 = \frac{1}{8}\left(3.29 - \frac{3.7^2}{9}\right) = 0.2211$ | **M1** | $s = 0.4702$ |
| | $t = \frac{0.411 - 0.3}{\sqrt{\frac{0.2211}{9}}}$<br>$= 0.708$ or $0.709$ | **M1 A1** | |
| | '0.708' < 1.397 | **M1** | Compare *their* 0.708 with 1.397 |
| | Accept $H_0$<br>Insufficient evidence to support claim | **A1 FT** | In context, except possibly hypotheses<br>Level of uncertainty in language used. No contradictions |
| | | **8** | |

13 - (9231/41_Summer_2021_Q1)

| | | | |
|---|---|---|---|
| (a) | $\bar{x} = \frac{56.56}{7} = 8.08$ | **B1** | Accept unsimplified. |
| | $s^2 = \frac{1}{6}\left(457.275 - \frac{56.56^2}{7}\right) = 0.045033 \left[= \frac{1351}{30000}\right]$ | **M1** | Accept unsimplified. |
| | $t = \frac{8.08 - 8.22}{\sqrt{\frac{s^2}{7}}} = -1.745$ | **M1 A1** | 1.74 – 1.75 |
| | Tabular value 1.440<br>1.745 > 1.440 , so reject $H_0$ | **M1** | Comparison with 1.440 and correct FT conclusion. |
| | There is insufficient evidence to support the hypothesis that the (population) mean is 8.22 OR sufficient evidence to suggest mean is less than 8.22 OR sufficient evidence to suggest $\mu < 8.22$ OE | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | | **6** | |
| (b) | Underlying distribution is normal/ population is normal | **B1** | $X$ is normal distributed. |
| | | **1** | |

14 - (9231/41_Summer_2021_Q4)

| | | | |
|---|---|---|---|
| | $s_x^{\,2} = \frac{1}{49}\left(1596 - \frac{282^2}{50}\right) = 0.11265\ \left[= \frac{138}{1225}\right]$ | **M1** | One correct, unsimplified. |
| | $s_y^{\,2} = \frac{1}{59}\left(1808 - \frac{328^2}{60}\right) = 0.25311\ \left[= \frac{224}{885}\right]$ | **A1** | Both correct to 3sf. |
| | $s^2 = \frac{0.11265}{50} + \frac{0.25312}{60} = 0.006471\ldots$ | **M1 A1** | May be implied. |
| | $z = \dfrac{\frac{282}{50} - \frac{328}{60}}{\text{their } s}$ | **M1** | |
| | [$z$] = 2.155 | **A1** | Tolerance 2.15 – 2.16 |
| | Comparison with tabular value 1.96 OE<br>2.155 > 1.96<br>Reject $H_0$ | **M1** | Comparison with 1.96 OE and conclusion. |
| | Sufficient evidence to reject population means are equal OR sufficient evidence to accept population means are not equal OR sufficient evidence to suggest $\mu_A \neq \mu_B$ OE | **A1** | Correct conclusion following correct work. Level of uncertainty in language is used. |
| | | **8** | |

15 - (9231/42_Summer_2021_Q1)

| | | | |
|---|---|---|---|
| (a) | $\overline{x} = \frac{56.56}{7} = 8.08$ | **B1** | Accept unsimplified. |
| | $s^2 = \frac{1}{6}\left(457.275 - \frac{56.56^2}{7}\right) = 0.045033 \left[= \frac{1351}{30000}\right]$ | **M1** | Accept unsimplified. |
| | $t = \frac{8.08 - 8.22}{\sqrt{\frac{s^2}{7}}} = -1.745$ | **M1 A1** | 1.74 – 1.75 |
| | Tabular value 1.440<br>1.745 > 1.440 , so reject $H_0$ | **M1** | Comparison with 1.440 and correct FT conclusion. |
| | There is insufficient evidence to support the hypothesis that the (population) mean is 8.22 OR sufficient evidence to suggest mean is less than 8.22 OR sufficient evidence to suggest $\mu < 8.22$ OE | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | | **6** | |
| (b) | Underlying distribution is normal/ population is normal | **B1** | *X* is normal distributed. |
| | | **1** | |

16 - (9231/42_Summer_2021_Q4)

| | | | |
|---|---|---|---|
| | $s_x{}^2 = \frac{1}{49}\left(1596 - \frac{282^2}{50}\right) = 0.11265\ \left[= \frac{138}{1225}\right]$ | **M1** | One correct, unsimplified. |
| | $s_y{}^2 = \frac{1}{59}\left(1808 - \frac{328^2}{60}\right) = 0.25311\ \left[= \frac{224}{885}\right]$ | **A1** | Both correct to 3sf. |
| | $s^2 = \frac{0.11265}{50} + \frac{0.25312}{60} = 0.006471\ldots$ | **M1 A1** | May be implied. |
| | $z = \dfrac{\frac{282}{50} - \frac{328}{60}}{\text{their } s}$ | **M1** | |
| | $[z] = 2.155$ | **A1** | Tolerance 2.15 – 2.16 |
| | Comparison with tabular value 1.96 OE<br>2.155 > 1.96<br>Reject $H_0$ | **M1** | Comparison with 1.96 OE and conclusion. |
| | Sufficient evidence to reject population means are equal OR sufficient evidence to accept population means are not equal OR sufficient evidence to suggest $\mu_A \neq \mu_B$ OE | **A1** | Correct conclusion following correct work. Level of uncertainty in language is used. |
| | | **8** | |

17 - (9231/43_Summer_2021_Q1)

| | | | |
|---|---|---|---|
| | $s^2 = \frac{1}{9}\left(2313 - \frac{152^2}{10}\right) = 0.28889$ | **M1** | $\frac{13}{45}$, accept unsimplified. |
| | $H_0: \mu = 14.8$  $H_1: \mu > 14.8$ | **B1** | If $\mu$ not used, 'population mean' required.<br>Must see 14.8, must be = and >. |
| | $[t =]\dfrac{\frac{152}{10} - 14.8}{s/\sqrt{10}}$ | **M1** | Using unbiased estimate.<br>*p*-value is 0.0215. |
| | $[t =]\,2.35$ | **A1** | Rounds to 2.35. |
| | Tabular value = 1.833<br>'2.35' > 1.833<br>Reject $H_0$ | **M1** | Comparison with 1.833 and correct FT conclusion. |
| | Sufficient evidence to accept Farmer B's claim oe | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used.<br>Allow $\pm 1$ difference in third significant figure of their *t* for this mark. |
| | | **6** | |

18 - (9231/43_Summer_2021_Q3)

| | | | |
|---|---|---|---|
| | $\bar{x} = \frac{89}{50} = 1.78 \quad \bar{y} = \frac{67.2}{40} = 1.68$ | **B1** | Implied by 0.1 in the CI formula. |
| | $\sigma_x^{\ 2} = \frac{1}{49}\left(159.4 - \frac{89^2}{50}\right) = 0.02 \quad \left(= \frac{1}{50}\right)$ | **M1** | One variance correct unsimplified. |
| | $\sigma_y^{\ 2} = \frac{1}{39}\left(113.1 - \frac{67.2^2}{40}\right) = 0.0052308 \quad \left(= \frac{17}{3250}\right)$ | **A1** | Both correctly calculated. |
| | $\sigma^2 = \frac{0.02}{50} + \frac{0.0052308}{40} = \left(\frac{69}{130000}\right)$ | **M1** | Unsimplified expression must be seen, 40 and 50 in correct places. |
| | [ $\sigma^2$ =] 0.0005308 | **A1** | May be implied by correct final answer. |
| | CI : 1.78 – 1.68 $\pm z \times \sqrt{0.0005308}$ | **M1** | Correct form for CI with a $z$-value. |
| | with $z$ = 1.96 | **A1** | |
| | 0.10 $\pm$ 0.0452 or [0.0548, 0.145] | **A1** | Allow either form. |
| | | **8** | SC Assuming equal variances award B1M0A0, M1A1 (for use of pooled variance formula, gives 0.0135), M1A1A0. |

19 - (9231/41_Winter_2021_Q1)

| | | | |
|---|---|---|---|
| (a) | $s^2 = \frac{78.5}{9} = 8.7222$ | **B1** | $\frac{157}{18}$ |
| | CI: $25.6 \pm t\sqrt{\frac{s^2}{10}}$ | **M1** | Correct expression with a *t* value. |
| | With $t = 1.833$ | **A1** | With correct *t* value. |
| | $25.6 \pm 1.71(2)$ or [23.9, 27.3] | **A1** | Accept in either form, ISW. Accept inequality form. |
| | | **4** | |
| (b) | $\frac{25.6 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ | **M1** | |
| | With $t = -1.383$ | **A1** | Allow 1.383. |
| | 26.9 | **A1** | CAO |
| | | **3** | |

20 - (9231/41_Winter_2021_Q4)

| | | | |
|---|---|---|---|
| | $H_0: \mu_B - \mu_A = 2$ and $H_1: \mu_B - \mu_A > 2$ | **B1** | Or use of $\mu_d$. |
| | Differences: 5 8 2 0 4 7 −2 1 10 −3 | **M1** | Allow one error. |
| | $\sum d = 32, \quad \sum d^2 = 272$<br>$\bar{d} = 3.2,\ s^2 = \frac{1}{9}\left(272 - \frac{32^2}{10}\right) = 18.84 \quad \left(= \frac{848}{45}\right)$ | **M1** | Sample mean and variance. |
| | $t = \dfrac{3.2 - 2}{\sqrt{\dfrac{s^2}{10}}} = 0.874$ | **M1 A1** | |
| | Compare with tabular value 1.833: 0.874 < 1.833.<br>Accept $H_0$. | **M1** | Compare their value with 1.833 and conclusion. |
| | Insufficient evidence to support claim. | **A1** | Correct conclusion, in context, following correct work.<br>Level of uncertainty in language is used. |
| | Assumption: population differences are normally distributed | **B1** | |
| | | **8** | |

21 - (9231/42_Winter_2021_Q1)

| | | | |
|---|---|---|---|
| | $s_x^2 = \frac{1}{39}\left(14320 - \frac{752^2}{40}\right) = 4.677 = \frac{304}{65}$ | **M1** | One correct unsimplified. |
| | $s_y^2 = \frac{1}{59}\left(40200 - \frac{1548^2}{60}\right) = 4.434 = \frac{1308}{295}$ | **A1** | Both correct to at least 3sf. |
| | $s^2 = \frac{4.6769}{40} + \frac{4.4339}{60}$ | **M1** | |
| | $0.1908 \left(= \frac{3659}{19175}\right)$ | **A1** | |
| | CI = $\left(\frac{752}{40} - \frac{1548}{60}\right) \pm 1.96s$ | **M1A1** | Correct form with a $z$-value with 1.96 $s$ can be unsimplified. |
| | $-7.0 \pm 0.856 = [-7.86, -6.14]$ | **A1** | Accept in either form, ISW.<br>Accept [6.14, 7.86].<br>Accept inequality form. |
| | | **7** | |

22 - (9231/42_Winter_2021_Q6)

| | | | |
|---|---|---|---|
| | $H_0 : \mu_A = \mu_B$ and $H_1 : \mu_A > \mu_B$ | **B1** | |
| | $\sum x = 21.2 \quad \sum x^2 = 47.92$<br>$s_x{}^2 = \frac{1}{9}\left(47.92 - \frac{21.2^2}{10}\right) = 0.33067 \ = \frac{124}{375}$<br>$s_y{}^2 = \frac{1}{11}\left(53.75 - \frac{24.48^2}{12}\right) = 0.34644 \ = \frac{9527}{27500}$ | **M1 A1** | |
| | $s^2 = \frac{9 \times 0.33067 + 11 \times 0.34644}{10 + 12 - 2} \ = \ 0.3393$ | **M1 A1** | |
| | $t = \frac{2.12 - 2.04}{s\sqrt{\frac{1}{10} + \frac{1}{12}}} = 0.321$ | **M1 A1** | Accept 0.321 – 0.322. |
| | Critical value: 1.325<br>0.321 < 1.325 accept $H_0$. | **M1** | Compare their value with 1.325 and correct FT conclusion. |
| | Insufficient evidence that mean of *A* is greater than mean of *B* | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | EITHER: Distributions are normal and equal variances<br>OR: Distributions are normal | **B1** | Assumptions consistent with method used.<br>Accept 'population is normal'.<br>'Distribution of difference of means is normal'. |
| | **Alternative method for question 6** | | |
| | $H_0 : \mu_A = \mu_B$ and $H_1 : \mu_A > \mu_B$ | **B1** | |

| | | | |
|---|---|---|---|
| | $\sum x = 21.2 \quad \sum x^2 = 47.92$ | **M1** | For one correct unsimplified. |
| | $s_x{}^2 = \frac{1}{9}\left(47.92 - \frac{21.2^2}{10}\right) = 0.33067 \ = \frac{124}{375}$<br>$s_y{}^2 = \frac{1}{11}\left(53.75 - \frac{24.48^2}{12}\right) = 0.34644 \ = \frac{9527}{27500}$ | **A1** | For both correct. |
| | $s^2 = \frac{0.33067}{10} + \frac{0.34644}{12} = 0.061936$ | **M1 A1** | |
| | $t = \frac{2.12 - 2.04}{s} = 0.321(5)$ | **M1 A1** | Accept 0.321 – 0.322. |
| | Critical value: 1.325<br>0.321 < 1.325 accept $H_0$. | **M1** | Compare their value with 1.325 and correct FT conclusion. |
| | Insufficient evidence that mean of *A* is greater than mean of *B*. | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | EITHER: Distributions are normal and equal variances<br>OR: Distributions are normal. | **B1** | Assumptions consistent with method used.<br>Accept ‘population is normal’.<br>‘Distribution of difference of means is normal’. |
| | | **10** | |

23 - (9231/43_Winter_2021_Q1)

| | | | |
|---|---|---|---|
| (a) | $s^2 = \frac{78.5}{9} = 8.7222$ | **B1** | $\frac{157}{18}$ |
| | CI: $25.6 \pm t\sqrt{\frac{s^2}{10}}$ | **M1** | Correct expression with a *t* value. |
| | With $t = 1.833$ | **A1** | With correct *t* value. |
| | $25.6 \pm 1.71(2)$ or [23.9, 27.3] | **A1** | Accept in either form, ISW. Accept inequality form. |
| | | **4** | |
| (b) | $\frac{25.6 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ | **M1** | |
| | With $t = -1.383$ | **A1** | Allow 1.383. |
| | 26.9 | **A1** | CAO |
| | | **3** | |

24 - (9231/43_Winter_2021_Q4)

| | | | |
|---|---|---|---|
| | $H_0 : \mu_B - \mu_A = 2$ and $H_1 : \mu_B - \mu_A > 2$ | **B1** | Or use of $\mu_d$ . |
| | Differences: 5 8 2 0 4 7 −2 1 10 −3 | **M1** | Allow one error. |
| | $\sum d = 32, \quad \sum d^2 = 272$<br>$\overline{d} = 3.2,\ s^2 = \frac{1}{9}\left(272 - \frac{32^2}{10}\right) = 18.84 \quad \left(= \frac{848}{45}\right)$ | **M1** | Sample mean and variance. |
| | $t = \frac{3.2 - 2}{\sqrt{\frac{s^2}{10}}} = 0.874$ | **M1 A1** | |
| | Compare with tabular value 1.833: 0.874 < 1.833.<br>Accept $H_0$ . | **M1** | Compare their value with 1.833 and conclusion. |
| | Insufficient evidence to support claim. | **A1** | Correct conclusion, in context, following correct work.<br>Level of uncertainty in language is used. |
| | Assumption: population differences are normally distributed | **B1** | |
| | | **8** | |

25 - (9231/41_Summer_2022_Q1)

| | | | |
|---|---|---|---|
| (a) | Differences: 0.6 1.3 0.0 0.7 0.6 0.6 1.8 –0.2 –0.2 0.4 | **M1** | Attempt at differences, allow one error. |
| | $\sum d = 5.6 \quad \sum d^2 = 6.74$<br>$s^2 = \frac{1}{9}\left(6.74 - \frac{5.6^2}{10}\right) = 0.400(4)$ | **M1** | |
| | $H_0$: $\mu_B - \mu_A = 0.4$ and $H_1$: $\mu_B - \mu_A > 0.4$ | **B1** | Allow use of $\mu_d$ |
| | $t = \frac{0.56 - 0.4}{\sqrt{\frac{0.4004}{10}}}$ (0.800) | **M1** | |
| | $t = 0.7996$ (0.800) | **A1** | |
| | Critical value = $t_{0.90}(9)$ = 1.383<br>Compare: ‘0.7996’ < 1.383<br>Accept $H_0$ | **M1** | Compare calculated value with 1.383 and correct FT conclusion. |
| | Insufficient evidence to support manager’s claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **7** | |
| (b) | Distribution of **population** differences is normal | **B1** | |
| | | **1** | |

26 - (9231/41_Summer_2022_Q5)

| | | | |
|---|---|---|---|
| (a) | $\sum x = 44.4, \quad \sum x^2 = 247.48$ | **B1** | |
| | $s^2 = \frac{1}{7}\left(247.48 - \frac{44.4^2}{8}\right) = 0.1514\left(= \frac{53}{350}\right)$ | **M1** | |
| | CI: $\frac{44.4}{8} \pm 1.895\sqrt{\frac{0.1514}{8}}$ | **M1** | Correct formula with a $t$ value. |
| | 1.895 | **B1** | 1.895 used. |
| | $5.55 \pm 0.261 = [5.29, 5.81]$ | **A1** | Either form. |
| | | **5** | |
| (b) | $H_0 : \mu = 5.9$ and $H_1 : \mu < 5.9$ | **B1** | |
| | $t = \frac{5.55 - 5.9}{\sqrt{\frac{s^2}{8}}} = -2.544$ | **M1** | |
| | Critical value $t_{0.975}(7) = 2.365$<br>Compare '– 2.544' < – 2.365, reject $H_0$. | **M1** | Compare calculated value with $-2.365$ and correct FT conclusion. |
| | Sufficient evidence to support Raman's claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **4** | |

27 - (9231/42_Summer_2022_Q1)

| | | | |
|---|---|---|---|
| (a) | Differences:<br>0.6 1.3 0.0 0.7 0.6 0.6 1.8 –0.2 –0.2 0.4 | **M1** | Attempt at differences, allow one error. |
| | $\sum d = 5.6 \quad \sum d^2 = 6.74$<br>$s^2 = \frac{1}{9}\left(6.74 - \frac{5.6^2}{10}\right) = 0.400(4)$ | **M1** | |
| | $H_0$: $\mu_B - \mu_A = 0.4$ and $H_1$: $\mu_B - \mu_A > 0.4$ | **B1** | Allow use of $\mu_d$ |
| | $t = \frac{0.56 - 0.4}{\sqrt{\frac{0.4004}{10}}}$ (0.800) | **M1** | |
| | $t = 0.7996$ (0.800) | **A1** | |
| | Critical value = $t_{0.90}(9)$ = 1.383<br>Compare: ‘0.7996’ < 1.383<br>Accept $H_0$ | **M1** | Compare calculated value with 1.383 and correct FT conclusion. |
| | Insufficient evidence to support manager’s claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **7** | |
| (b) | Distribution of **population** differences is normal | **B1** | |
| | | **1** | |

28 - (9231/42_Summer_2022_Q5)

| | | | |
|---|---|---|---|
| (a) | $\sum x = 44.4, \quad \sum x^2 = 247.48$ | **B1** | |
| | $s^2 = \frac{1}{7}\left(247.48 - \frac{44.4^2}{8}\right) = 0.1514\left(= \frac{53}{350}\right)$ | **M1** | |
| | CI: $\frac{44.4}{8} \pm 1.895\sqrt{\frac{0.1514}{8}}$ | **M1** | Correct formula with a $t$ value. |
| | 1.895 | **B1** | 1.895 used. |
| | $5.55 \pm 0.261 = [5.29, 5.81]$ | **A1** | Either form. |
| | | **5** | |
| (b) | $H_0 : \mu = 5.9$ and $H_1 : \mu < 5.9$ | **B1** | |
| | $t = \frac{5.55 - 5.9}{\sqrt{\frac{s^2}{8}}} = -2.544$ | **M1** | |
| | Critical value $t_{0.975}(7) = 2.365$<br>Compare '$-2.544$' $< -2.365$, reject $H_0$. | **M1** | Compare calculated value with $-2.365$ and correct FT conclusion. |
| | Sufficient evidence to support Raman's claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **4** | |

29 - (9231/43_Summer_2022_Q1)

| | | | |
|---|---|---|---|
| | $s^2 = \frac{94.5}{7} (= 13.5)$ | **B1** | |
| | CI: $33.8 \pm 2.365\sqrt{\frac{s^2}{8}}$ | **M1** | Correct formula with a *t* value, must be 33.8. |
| | | **B1** | 2.365 used. |
| | $33.8 \pm 3.072 = [30.7, 36.9]$ | **A1** | Either form. |
| | | **4** | |

30 - (9231/43_Summer_2022_Q6)

| | | | |
|---|---|---|---|
| (a) | $s_x^{\ 2} = \frac{1}{39}\left(31400 - \frac{1120^2}{40}\right) = 1.0256$<br>$s_y^{\ 2} = \frac{1}{49}\left(37600 - \frac{1370^2}{50}\right) = 1.2653$ | **M1** | $\frac{40}{39}$<br>$\frac{62}{49}$ Both required |
| | $s^2 = \frac{1.0256}{40} + \frac{1.2653}{50}$ | **M1** | $\frac{2434}{47775}$<br>Condone $s^2 = \frac{1.03}{40} + \frac{1.27}{50}$ or $0.05115$. |
| | $s^2 = 0.050947$ | **A1** | May be implied by 2.66 as $z$ value. |
| | $H_0 : \mu_a - \mu_b = 0$ and $H_1 : \mu_a - \mu_b \neq 0$ | **B1** | |
| | $z = \frac{\frac{1120}{40} - \frac{1370}{50}}{s} = \frac{0.6}{\sqrt{0.050947}}$ | **M1** | |
| | $z = 2.66$ | **A1** | |
| | Critical value of $z = 2.326$<br>$2.66 > 2.326$<br>Reject $H_0$ | **M1** | Compare their z value with 2.326 and correct FT conclusion. |
| | Sufficient evidence to support difference in **population** means oe | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **8** | |

| | | | |
|---|---|---|---|
| (b) | $z=\frac{0.6-0.25}{s}$ | **M1** | Condone 1.54. |
| | $z=1.55$ | **A1** | |
| | 0.9392 – 0.9395 or 0.0605 – 0.0608 | **A1** | |
| | $\alpha \geqslant 6.05$ | **A1** | Condone 6.05 – 6.08, accept $>$ |
| | | **4** | |

31 - (9231/41_Winter_2022_Q1)

| | | | |
|---|---|---|---|
| | $\left[\bar{x}=\frac{1625}{50}=32.5\right]\quad s_x{}^2=\frac{1}{49}\left(53200-\frac{1625^2}{50}\right)=7.908$ | **M1** | $\frac{775}{98}$ |
| | $\left[\bar{y}=\frac{1854}{60}=30.9\right]\quad s_y{}^2=\frac{1}{59}\left(57900-\frac{1854^2}{60}\right)=10.3627$ | **A1** | $\frac{3057}{295}$ Both correct |
| | $s^2=\frac{7.908}{50}+\frac{10.3627}{60}\ [=0.33087]$ | **M1 A1** | |
| | CI: $(32.5-30.9)\pm 1.96s$ | **M1 A1** | Correct form with a *z*-value. Correct with 1.96. |
| | $1.6\pm 1.1274=[0.473,2.73]$ | **A1** | At least 3sf. |
| | | **7** | Pooled variance used: M1A1 M0A0 M1A1A0 max 4/7 |

32 - (9231/41_Winter_2022_Q6)

| | | | |
|---|---|---|---|
| | $s_x{}^2 = \frac{1}{11}\left(3.26 - \frac{6.24^2}{12}\right) = 0.001382$<br>$s_y{}^2 = \frac{1}{9}\left(2.6124 - \frac{5.1^2}{10}\right) = 0.001267$ | **M1** | $\frac{19}{13750}$<br>$\frac{19}{15000}$ |
| | | **A1** | Both correct. |
| | $s^2 = \frac{11 \times 0.001382 + 9 \times 0.001267}{12 + 10 - 2} = 0.001330$ | **M1 A1** | Find pooled variance. |
| | $t = \dfrac{\frac{6.24}{12} - \frac{5.1}{10}}{s\sqrt{\frac{1}{12} + \frac{1}{10}}} = 0.640$ | **M1 A1** | Use correct formula for $t$. |
| | $H_0 : \mu_a = \mu_b \quad H_1 : \mu_a > \mu_b$ | **B1** | |
| | Critical value is 2.086 ($t_{0.975}(20)$)<br>0.640 < 2.086<br>Accept $H_0$ | **M1** | Compare their calculated value with 2.086 and conclusion. |
| | There is insufficient evidence to support inspector's claim | **A1** | All correct except possibly B1, in context, level of uncertainty in language.<br>'Prove' scores A0 |
| | | **9** | Pooled variance not used: M1A1 M0A0 M0A0 B1 M1A0 max 4/9 |

33 - (9231/42_Winter_2022_Q1)

| | | | |
|---|---|---|---|
| | $2.02 = \bar{x} + 1.833\sqrt{\frac{s^2}{10}}$<br>$1.78 = \bar{x} - 1.833\sqrt{\frac{s^2}{10}}$<br>Add, $\bar{x} = \frac{2.02+1.78}{2} = 1.90$ | **M1** | Add. |
| | $\sum x = 19$ | **A1** | |
| | Subtract: $2.02 - 1.78 = 2 \times 1.833\sqrt{\frac{s^2}{10}}$ | **M1** | Allow 1.372, 1.383, 1.812 instead of 1.833. |
| | $s^2 = 0.042859$ | **A1** | May be implied.<br>$s = 0.20702$ |
| | But $s^2 = \frac{1}{9}(\sum x^2 - \frac{(\sum x)^2}{10})$ | **M1** | |
| | $\sum x^2 = 9s^2 + \frac{19^2}{10} = 36.5$ | **A1** | 36.486 |
| | | | Using 1.645: maximum M1A1 M0A0 M1A0 3/6 |
| | | **6** | |

34 - (9231/42_Winter_2022_Q3)

| | | | |
|---|---|---|---|
| | $s_x^{\ 2} = \frac{1}{49}\left(115.2 - \frac{75.5^2}{50}\right) = 0.02439$<br>$s_y^{\ 2} = \frac{1}{79}\left(172.6 - \frac{116.8^2}{80}\right) = 0.02623$ | **M1** | $\frac{239}{9800}$<br>$\frac{259}{9875}$ |
| | | **A1** | Both correct |
| | $s^2 = \frac{0.02439}{50} + \frac{0.02623}{80} \quad [= 0.00081565]$ | **M1 A1** | May be implied by 1.75 for $z$ |
| | $z = \dfrac{\frac{75.5}{50} - \frac{116.8}{80}}{s}$ | **M1** | |
| | 1.75 | **A1** | |
| | Compare with 1.645: 1.75 >1.645<br>Reject null hypothesis | **M1** | Using areas, 0.04 < 0.05. |
| | Sufficient evidence to suggest that population mean in country $X$ is greater than population mean in country $Y$ | **A1** | Correct conclusion in context, following correct work, level of uncertainty in language.<br>'Prove' gives A0. |
| | | **8** | Pooled variance: $z = 1.74$<br>M1A1 M0A0M0A0 M1A0 max 3/8 |