Further Math with Ali Hashir

Inference Using Normal and t-Distributions

Topical Questions (CIE 9231)

Compiled from CIE 9231 Further Mathematics past papers (2015–2025), sorted by topic. Each question is numbered for this booklet, with its original paper code and session shown beneath.

About this resource: I teach CIE A Level Further Mathematics, and over the years I found that there simply weren't many well-organised, topic-wise resources available for this subject – it's a niche A Level with relatively few candidates, so it doesn't get much attention from resource makers. These topical booklets are my attempt to fix that: every past-paper question, sorted by topic, with its official mark scheme alongside it.

Help us improve these documents for future students. If you spot a mistake, email us at **vectorspacepk@gmail.com**.

What kind of mistakes might you find? These booklets are compiled by hand from hundreds of past-paper PDFs. I've checked them as carefully as I can, but I'm only human, and the odd error may have slipped through – so please help me catch what I missed. Things to look out for:

- Spelling or typing mistakes
- Wrong final answers
- A mark scheme that doesn't line up with its question (misaligned pages)
- A mark scheme that belongs to the wrong question entirely
- Pre-2020 questions that are technically out of the current syllabus (the syllabus changed in 2020, but some older questions are still included here for practice)
- Missing, extra, or wrongly-labelled marks (e.g. an M1 or A1 that shouldn't be there, or one that's missing)
- Diagrams or images that are unclear, missing, or don't match the question
- Questions placed under the wrong topic

If you spot any of these (or anything else that looks off), please let me know.

This is **Version 1** of this booklet. As readers point out errors, I'll keep correcting and improving these documents, so I'd recommend checking back on the website from time to time and using the latest version rather than relying on a saved copy.

If you think these documents brought you value, consider following me on social media – it helps more than you'd think.

Q1

9231/21, 9231/22 • May/June 2015 • Question 6

The independent random variables X and Y have distributions with the same variance σ^2. Random samples of N observations of X and 10 observations of Y are taken, and the results are summarised by

$$\Sigma x = 5, \quad \Sigma x^2 = 11, \quad \Sigma y = 10, \quad \Sigma y^2 = 160.$$

These data give a pooled estimate of 12 for σ^2. Find N. [4]

Q2

9231/21, 9231/22 • May/June 2015 • Question 7

A random sample of 8 sunflower plants is taken from the large number grown by a gardener, and the heights of the plants are measured. A 95% confidence interval for the population mean, μ metres, is calculated from the sample data as $1.17 < \mu < 2.03$. Given that the height of a sunflower plant is denoted by x metres, find the values of Σx and Σx^2 for this sample of 8 plants. [7]

Q3

9231/21, 9231/22 • May/June 2015 • Question 10

Young children at a primary school are learning to throw a ball as far as they can. The distance thrown at the beginning of the school year and the distance thrown at the end of the same school year are recorded for each child. The distance thrown, in metres, at the beginning of the year is denoted by x; the distance thrown, in metres, at the end of the year is denoted by y. For a random sample of 10 children, the results are shown in the following table.

Child	A	B	C	D	E	F	G	H	I	J
x	5.2	4.1	3.7	5.4	7.6	6.1	3.2	4.0	3.5	8.0
y	6.2	4.8	5.0	5.6	7.7	7.0	4.0	4.5	3.6	8.5

$$[\Sigma x = 50.8, \quad \Sigma x^2 = 284.16, \quad \Sigma y = 56.9, \quad \Sigma y^2 = 347.59, \quad \Sigma xy = 313.28.]$$

A particular child threw the ball a distance of 7.0 metres at the beginning of the year, but he could not throw at the end of the year because he had broken his arm. By finding the equation of an appropriate regression line, estimate the distance this child would have thrown at the end of the year. [5]

The teacher suspects that, on average, the distance thrown by a child increases between the two throws by more than 0.4 metres. Stating suitable hypotheses and assuming a normal distribution, test the teacher's suspicion at the 5% significance level. [8]

Q4

9231/23 • May/June 2015 • Question 8

A large number of long jumpers are competing in a national long jump competition. The distances, in metres, jumped by a random sample of 7 competitors are as follows.

$$6.25 \quad 7.01 \quad 5.74 \quad 6.89 \quad 7.24 \quad 5.64 \quad 6.52$$

Assuming that distances are normally distributed, test, at the 5% significance level, whether the mean distance jumped by long jumpers in this competition is greater than 6.2 metres. [7]

The distances jumped by another random sample of 8 long jumpers in this competition are recorded. Using the data from this sample of 8 long jumpers, a 95% confidence interval for the population mean, μ metres, is calculated as $5.89 < \mu < 6.75$. Find the unbiased estimates for the population mean and population variance used in this calculation. [5]

Q5

9231/21, 9231/22, 9231/23 • October/November 2015 • Question 5

A random sample of 10 observations of a normal variable X gave the following summarised data, where $\bar{x}$ is the sample mean.

$$\Sigma x = 222.8 \qquad \Sigma(x - \bar{x})^2 = 4.12$$

Find a 95% confidence interval for the population mean. [5]

Q6

9231/21, 9231/22 • May/June 2016 • Question 7

A random sample of 9 observations of a normal variable X is taken. The results are summarised as follows.

$$\Sigma x = 24.6 \qquad \Sigma x^2 = 68.5$$

Test, at the 5% significance level, whether the population mean is greater than 2.5. [8]

Q7

9231/23 • May/June 2016 • Question 8

The coach of a national athletics team carried out an investigation into the effect of high altitude training on the times of 400-metre runners. The times, in seconds, were recorded before and after a six-week period of high altitude training, for a random sample of 8 athletes. The results are given in the following table.

Athlete	A	B	C	D	E	F	G	H
Before	52.3	56.2	54.3	49.2	48.4	50.1	46.8	51.1
After	51.6	53.2	54.9	49.0	48.1	49.2	46.8	49.6

Stating any assumption that you make, test, at the 2.5% significance level, whether there is an improvement in athletes' times after high altitude training. [8]

Q8

9231/21, 9231/22, 9231/23 • October/November 2016 • Question 6

A random sample of 8 observations of a normal random variable X has mean $\bar{x}$, where

$$\bar{x} = 6.246 \quad \text{and} \quad \Sigma(x - \bar{x})^2 = 0.784.$$

Test, at the 5% significance level, whether the population mean of X is less than 6.44. [7]

Q9

9231/21, 9231/22, 9231/23 • October/November 2016 • Question 8

The amounts spent on the weekly food shopping by families in the big city P and the small town Q are to be compared. The amounts spent, in dollars, in P and Q are denoted by x and y respectively. For a random sample of 60 families in P and a random sample of 50 families in Q, the amounts are summarised as follows.

$$\Sigma x = 9600 \qquad \Sigma x^2 = 1\,560\,000 \qquad \Sigma y = 7200 \qquad \Sigma y^2 = 1\,052\,500$$

Assuming a common population variance, find

(i) a pooled estimate for the population variance, [4]

(ii) a 95% confidence interval for the difference in the population means in P and Q. [5]

Q10

9231/21, 9231/22 • May/June 2017 • Question 7

A farmer grows a particular type of fruit tree. On average, the mass of fruit produced per tree has been 6.2 kg. He has developed a new kind of soil and claims that the mean mass of fruit produced per tree when growing in this new soil has increased. A random sample of 10 trees grown in the new soil is chosen. The masses, x kg, of fruit produced are summarised as follows.

$$\Sigma x = 72.0 \qquad \Sigma x^2 = 542.0$$

Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

Q11

9231/21, 9231/22 • May/June 2017 • Question 9

Two fish farmers X and Y produce a particular type of fish. Farmer X chooses a random sample of 8 of his fish and records the masses, x kg, as follows.

$$1.2 \quad 1.4 \quad 0.8 \quad 2.1 \quad 1.8 \quad 2.6 \quad 1.5 \quad 2.0$$

Farmer Y chooses a random sample of 10 of his fish and summarises the masses, y kg, as follows.

$$\Sigma y = 20.2 \qquad \Sigma y^2 = 44.6$$

You should assume that both distributions are normal with equal variances. Test at the 10% significance level whether the mean mass of fish produced by farmer X differs from the mean mass of fish produced by farmer Y. [10]

Q12

9231/23 • May/June 2017 • Question 6

The independent variables X and Y have distributions with the same variance σ^2. Random samples of N observations of X and $2N$ observations of Y are taken, and the results are summarised by

$$\Sigma x = 4, \quad \Sigma x^2 = 10, \quad \Sigma y = 8, \quad \Sigma y^2 = 102.$$

These data give a pooled estimate of 10 for σ^2. Find N. [5]

Q13

9231/23 • May/June 2017 • Question 8

The number, x, of beech trees was counted in each of 50 randomly chosen regions of equal size in beech forests in country A. The number, y, of beech trees was counted in each of 40 randomly chosen regions of the same equal size in beech forests in country B. The results are summarised as follows.

$$\Sigma x = 1416 \qquad \Sigma x^2 = 41\,100 \qquad \Sigma y = 888 \qquad \Sigma y^2 = 20\,140$$

Find a 95% confidence interval for the difference between the mean number of beech trees in regions of this size in country A and in country B. [9]

Q14

9231/21, 9231/22, 9231/23 • October/November 2017 • Question 10

A factory produces bottles of an energy juice. Two different machines are used to fill empty bottles with the juice. The manager chooses a random sample of 50 bottles filled by machine X and a random sample of 60 bottles filled by machine Y. The volumes of juice, x and y respectively, measured in appropriate units, are summarised by

$$\sum x = 45.5, \quad \sum (x - \bar{x})^2 = 19.56, \quad \sum y = 72.3, \quad \sum (y - \bar{y})^2 = 30.25,$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the volume of juice in the bottles filled by X and Y respectively.

(i) Find a 90% confidence interval for the difference between the mean volume of juice in bottles filled by machine X and the mean volume of juice in bottles filled by machine Y. [7]

A test at the α% significance level does not provide evidence that there is any difference in the means of the volume of juice in bottles filled by machine X and the volume of juice in bottles filled by machine Y.

(ii) Find the set of possible values of α. [6]

Q15

9231/21, 9231/22 • May/June 2018 • Question 7

A large number of athletes are taking part in a competition. The masses, in kg, of a random sample of 7 athletes are as follows.

98.1 105.0 92.2 89.8 99.9 95.4 101.2

Assuming that masses are normally distributed, test, at the 10% significance level, whether the mean mass of athletes in this competition is equal to 94 kg. [7]

Q16

9231/21, 9231/22 • May/June 2018 • Question 10

The times taken to run 400 metres by students at two large colleges P and Q are being compared. There is no evidence that the population variances are equal. The time taken by a student at college P and the time taken by a student at college Q are denoted by x seconds and y seconds respectively. A random sample of 50 students from college P and a random sample of 60 students from college Q give the following summarised data.

$$\Sigma x = 2620 \qquad \Sigma x^2 = 138\,200 \qquad \Sigma y = 3060 \qquad \Sigma y^2 = 157\,000$$

(i) Using a 10% significance level, test whether, on average, students from college P take longer to run 400 metres than students from college Q. [9]

(ii) Find a 90% confidence interval for the difference in the mean times taken to run 400 metres by students from colleges P and Q. [3]

Q17

9231/23 • May/June 2018 • Question 10

During the summer months, all members of a large swimming club take part in intensive training. The times taken to swim 50 metres at the beginning of the summer and at the end of the summer are recorded for each member of the club. The time taken, in seconds, at the beginning of the summer is denoted by x and the time taken at the end of the summer is denoted by y. For a random sample of 9 members the results are shown in the following table.

Member	A	B	C	D	E	F	G	H	I
x	38.5	40.2	32.3	35.1	36.2	41.4	32.0	38.2	38.2
y	37.4	38.1	31.6	34.7	34.2	38.6	31.8	36.3	36.8

The swimming coach believes that, on average, the time taken by a swimmer to swim 50 metres will decrease by more than one second as a result of the intensive training.

(i) Stating suitable hypotheses and assuming a normal distribution, test the coach's belief at the 10% significance level. [8]

(ii) Find a 95% confidence interval for the population mean time taken to swim 50 metres after the intensive training, assuming a normal distribution. [4]

Q18

9231/21, 9231/23 • October/November 2018 • Question 8

The weekly salaries of employees at two large electronics companies, A and B, are being compared. The weekly salaries of an employee from company A and an employee from company B are denoted by $\$x$ and $\$y$ respectively. A random sample of 50 employees from company A and a random sample of 40 employees from company B give the following summarised data.

$$\Sigma x = 5120 \qquad \Sigma x^2 = 531\,000 \qquad \Sigma y = 3760 \qquad \Sigma y^2 = 375\,135$$

(i) The population mean salaries of employees from companies A and B are denoted by $\$\mu_A$ and $\$\mu_B$ respectively. Using a 5% significance level, test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

(ii) State, with a reason, whether any assumptions about the distributions of employees' salaries are needed for the test in part (i). [1]

Q19

9231/21, 9231/23 • October/November 2018 • Question 9

There are a large number of students at a particular college. The heights, in metres, of a random sample of 8 students are as follows.

1.75 1.72 1.62 1.70 1.82 1.75 1.68 1.84

You may assume that heights of students are normally distributed.

(i) Test, at the 5% significance level, whether the population mean height of students at this college is greater than 1.70 metres. [7]

(ii) Find a 95% confidence interval for the population mean height of students at this college. [3]

Q20

9231/22 • October/November 2018 • Question 6

The heights, in metres, of a random sample of 8 trees of a particular type are as follows.

14.2 11.3 10.8 8.4 12.8 11.5 12.1 9.2

Assuming that heights of trees of this type are normally distributed, calculate a 95% confidence interval for the mean height of trees of this type. [6]

Q21

9231/21, 9231/22 • May/June 2019 • Question 8

A large number of runners are attending a summer training camp. A random sample of 6 runners is chosen and their times to run 1500 m at the beginning of the camp and at the end of the camp are recorded. Their times, in minutes, are shown in the following table.

Runner	A	B	C	D	E	F
Time at beginning of camp	3.82	3.62	3.55	3.71	3.75	3.92
Time at end of camp	3.72	3.55	3.52	3.68	3.54	3.73

The organiser of the training camp claims that a runner's time will improve by more than 0.05 minutes between the beginning and end of the camp. Assuming that differences in time over the two runs are normally distributed, test at the 10% significance level whether the organiser's claim is justified. [8]

Q22

9231/23 • May/June 2019 • Question 9

A farmer grows large amounts of a certain crop. On average, the yield per plant has been 0.75 kg. The farmer has improved the soil in which the crop grows, and she claims that the yield per plant has increased. A random sample of 10 plants grown in the improved soil is chosen. The yields, x kg, are summarised as follows.

$$\Sigma x = 7.85 \qquad \Sigma x^2 = 6.19$$

(i) Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

(ii) Find a 95% confidence interval for the population mean yield for plants grown in the new soil. [3]

Q23

9231/21, 9231/22, 9231/23 • October/November 2019 • Question 6

A random sample of 9 members is taken from the large number of members of a sports club, and their heights are measured. The heights of all the members of the club are assumed to be normally distributed. A 95% confidence interval for the population mean height, μ metres, is calculated from the data as $1.65 \leq \mu \leq 1.85$.

(i) Find an unbiased estimate for the population variance. [3]

(ii) Denoting the height of a member of the club by x metres, find Σx^2 for this sample of 9 members. [4]

Q24

9231/21, 9231/22, 9231/23 • October/November 2019 • Question 8

A random sample of 8 elephants from region A is taken and their weights, x tonnes, are recorded. (1 tonne = 1000 kg.) The results are summarised as follows.

$$\Sigma x = 32.4 \qquad \Sigma x^2 = 131.82$$

A random sample of 10 elephants from region B is taken. Their weights give a sample mean of 3.78 tonnes and an unbiased variance estimate of 0.1555 tonnes^2. The distributions of the weights of elephants in regions A and B are both assumed to be normal with the same population variance. Test at the 10% significance level whether the mean weight of elephants in region A is the same as the mean weight of elephants in region B. [9]

Q25

9231/41, 9231/42 • May/June 2020 • Question 4

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.

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

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]

Q26

9231/41, 9231/42 • May/June 2020 • Question 5

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]

Q27

9231/43 • May/June 2020 • Question 2

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.

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

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

Q28

9231/43 • May/June 2020 • Question 5

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.

$$\sum x = 828 \quad \sum x^2 = 68622$$

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]

Q29

9231/41, 9231/43 • October/November 2020 • Question 1

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]

Q30

9231/41, 9231/43 • October/November 2020 • Question 4

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]

Q31

9231/42 • October/November 2020 • Question 1

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]

Q32

9231/42 • October/November 2020 • Question 6

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, μ 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.

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

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]

Q33

9231/41, 9231/42 • May/June 2021 • Question 1

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

$$8.26 \quad 7.78 \quad 7.92 \quad 8.04 \quad 8.27 \quad 7.95 \quad 8.34$$

The population mean of X is μ.

(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]

Q34

9231/41, 9231/42 • May/June 2021 • Question 4

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.

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

The population mean lengths of leaves from birch trees in regions A and B are μ_A cm and μ_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]

Q35

9231/43 • May/June 2021 • Question 1

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.

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

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

Q36

9231/43 • May/June 2021 • Question 3

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.

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

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]

Q37

9231/41, 9231/43 • October/November 2021 • Question 1

The times taken for students at a college to run 200 m have a normal distribution with mean μ 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.

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

(a) Find a 90% confidence interval for μ. [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]

Q38

9231/41, 9231/43 • October/November 2021 • Question 4

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]

Q39

9231/42 • October/November 2021 • Question 1

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.

$$\sum x = 752 \qquad \sum x^2 = 14320 \qquad \sum y = 1548 \qquad \sum y^2 = 40200$$

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]

Q40

9231/42 • October/November 2021 • Question 6

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.

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

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]

Q41

9231/41, 9231/42 • May/June 2022 • Question 1

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]

Q42

9231/41, 9231/42 • May/June 2022 • Question 5

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]

Q43

9231/43 • May/June 2022 • Question 1

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

$$\bar{x} = 33.8 \qquad \sum (x - \bar{x})^2 = 94.5$$

Find a 95% confidence interval for μ. [4]

Q44

9231/43 • May/June 2022 • Question 6

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.

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

The population means of the volumes of liquid in the bottles filled by machines A and B are denoted by μ_A and μ_B.

(a) Test at the 2% significance level whether there is any difference between μ_A and μ_B. [8]

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

Q45

9231/41, 9231/43 • October/November 2022 • Question 1

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]

Q46

9231/41, 9231/43 • October/November 2022 • Question 6

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 is greater than the 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

The diameters of the pipes produced by each machine are assumed to be normally distributed with equal population variances.

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

Q47

9231/42 • October/November 2022 • Question 1

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, μ m, of players in this club is calculated. It is assumed that heights are normally distributed. The confidence interval is $1.78 \leq \mu \leq 2.02$.

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

Q48

9231/42 • October/November 2022 • Question 3

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 μ_x kg and μ_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]

Q49

9231/41, 9231/42 • May/June 2023 • Question 1

The lengths of the leaves of a particular type of tree are normally distributed with mean μ cm. The lengths, x cm, of a random sample of 12 leaves of this type are recorded. The results are summarised as follows.

$$\sum x = 91.2 \qquad \sum x^2 = 695.8$$

Find a 95% confidence interval for μ. [4]

Q50

9231/41, 9231/42 • May/June 2023 • Question 2

The children at two large schools, P and Q, are all given the same puzzle to solve. A random sample of size 10 is taken from the children at school P. Their individual times to complete the puzzle give a sample mean of 9.12 minutes and an unbiased variance estimate of 2.16 minutes^2. A random sample of size 12 is taken from the children at school Q. Their individual times, x minutes, to complete the puzzle are summarised by

$$\sum x = 99.6 \qquad \sum (x - \overline{x})^2 = 21.5,$$

where $\overline{x}$ is the sample mean. Times to complete the puzzle are assumed to be normally distributed with the same population variance.

Test at the 5% significance level whether the population mean time taken to complete the puzzle by children at school P is greater than the population mean time taken to complete the puzzle by children at school Q. [8]

Q51

9231/43 • May/June 2023 • Question 2

Shane is studying the lengths of the tails of male red kangaroos. He takes a random sample of 14 male red kangaroos and measures the length of the tail, x m, for each kangaroo. He then calculates a 90% confidence interval for the population mean tail length, μ m, of male red kangaroos. He assumes that the tail lengths are normally distributed and finds that $1.11 \leqslant \mu \leqslant 1.14$.

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

Q52

9231/43 • May/June 2023 • Question 4

An inspector is checking the lengths of metal rods produced by two machines, X and Y. These rods should be of the same length, but the inspector suspects that those made by machine X are shorter, on average, than those made by machine Y. The inspector chooses a random sample of 80 rods made by machine X and a random sample of 60 rods made by machine Y. The lengths of these rods are x cm and y cm respectively. Her results are summarised as follows.

$$\sum x = 164.0 \qquad \sum x^2 = 338.1 \qquad \sum y = 124.8 \qquad \sum y^2 = 261.1$$

(a) Test at the 10% significance level whether the data supports the inspector's suspicion. [8]

(b) Give a reason why it is not necessary to make any assumption about the distributions of the lengths of the rods. [1]

Q53

9231/41, 9231/43 • October/November 2023 • Question 1

Maya is an athlete who competes in 1500-metre races. Last summer her practice run times had mean 4.22 minutes. Over the winter she has done some intense training to try to improve her times. A random sample of 10 of her practice run times, x minutes, this summer are summarised as follows.

$$\sum x = 42.05 \qquad \sum x^2 = 176.83$$

Maya's new practice run times are normally distributed. She believes that on average her times have improved as a result of her training.

Test, at the 5% significance level, whether Maya's belief is supported by the data. [6]

Q54

9231/41, 9231/43 • October/November 2023 • Question 3

Scientists are studying the effects of exercise on LDL blood cholesterol levels. Over a three-month period, a large group of people exercised for 20 minutes each day. For a randomly chosen sample of 10 of these people, the LDL blood cholesterol levels were measured at the beginning and the end of the three-month period. The results, measured in suitable units, are as follows.

Person	A	B	C	D	E	F	G	H	I	J
Beginning	72	84	120	90	102	135	64	75	80	88
End	64	76	105	92	105	115	67	75	75	84

(a) Test, at the 2.5% significance level, whether there is evidence that the population mean LDL blood cholesterol level has reduced by more than 2 units after the three-month period. [7]

(b) State any assumption that you have made in part (a). [1]

Q55

9231/42 • October/November 2023 • Question 1

A factory produces small bottles of natural spring water. Two different machines, X and Y, are used to fill empty bottles with the water. A quality control engineer checks the volumes of water in the bottles filled by each of the machines. He chooses a random sample of 60 bottles filled by machine X and a random sample of 75 bottles filled by machine Y. The volumes of water, x and y respectively, in millilitres, are summarised as follows.

$$\sum x = 6345 \qquad \sum (x - \bar{x})^2 = 243.8 \qquad \sum y = 7614 \qquad \sum (y - \bar{y})^2 = 384.9$$

$\bar{x}$ and $\bar{y}$ are the sample means of the volume of water in the bottles filled by machines X and Y respectively.

Find a 95% confidence interval for the difference between the mean volume of water in bottles filled by machine X and the mean volume of water in bottles filled by machine Y. [6]

Q56

9231/42 • October/November 2023 • Question 5 (parts b, c)

A company is deciding which of two machines, X and Y, can make a certain type of electrical component more quickly. The times taken, in minutes, to make one component of this type are recorded for a random sample of 8 components made by machine X and a random sample of 9 components made by machine Y. These times are as follows.

Machine X	4.0	4.6	4.7	4.8	5.0	5.2	5.6	5.8	
Machine Y	4.5	4.9	5.1	5.3	5.4	5.7	5.9	6.3	6.4

The manager claims that on average the time taken by machine X to make one component is less than that taken by machine Y.

(b) Assuming that the times taken to produce the components by the two machines are normally distributed with equal variances, carry out a t-test at the 5% significance level to test whether the manager's claim is supported by the data. [9]

(c) In general, would you expect the conclusions from the tests in parts (a) and (b) to be the same? Give a reason for your answer. [1]

Q57

9231/41, 9231/42 • May/June 2024 • Question 1

The times taken by members of a large cycling club to complete a cross-country circuit have a normal distribution with mean μ minutes. The times taken, x minutes, are recorded for a random sample of 14 members of the club. The results are summarised as follows, where $\bar{x}$ is the sample mean.

$$\bar{x} = 42.8 \qquad \sum(x - \bar{x})^2 = 941.5$$

Find a 95% confidence interval for μ. [4]

Q58

9231/41, 9231/42 • May/June 2024 • Question 6

Jade is a swimming instructor at a sports college. She claims that, as a result of an intensive training course, the mean time taken by students to swim 50 metres has reduced by more than 1 second. She chooses a random sample of 10 students. The times taken, in seconds, before and after the training course are recorded in the table.

Student	A	B	C	D	E	F	G	H	I	J
Time before course	54.2	47.4	52.1	59.0	55.3	51.0	48.9	52.2	58.4	51.4
Time after course	50.1	46.3	52.5	58.8	51.4	48.4	49.5	48.7	58.3	51.4

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

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

Q59

9231/43 • May/June 2024 • Question 2

A rowing club has a large number of members. A random sample of 12 of these members is taken and the pulse rate, x beats per minute (bpm), of each is measured after a 30-minute training session. A 98% confidence interval for the population mean pulse rate, μ bpm, is calculated from the sample as $64.22 < \mu < 68.66$.

(a) Find the values of $\sum x$ and $\sum x^2$. [6]

(b) State an assumption that is necessary for the confidence interval to be valid. [1]

Q60

9231/43 • May/June 2024 • Question 6

Seva is investigating the lengths of the tails of adult wallabies in two regions of Australia, X and Y. He chooses a random sample of 50 adult wallabies from region X and records the lengths, x cm, of their tails. He also chooses a random sample of 40 adult wallabies from region Y and records the lengths, y cm, of their tails. His results are summarised as follows.

$$\sum x = 1080 \qquad \sum x^2 = 23480 \qquad \sum y = 940 \qquad \sum y^2 = 22220$$

It cannot be assumed that the population variances of the two distributions are the same.

(a) Find a 90% confidence interval for the difference between the population mean lengths of the tails of adult wallabies in regions X and Y. [6]

The population mean lengths of the tails of adult wallabies in regions X and Y are μ_X cm and μ_Y cm respectively.

(b) Test, at the 10% significance level, the null hypothesis $\mu_Y - \mu_X = 1.1$ against the alternative hypothesis $\mu_Y - \mu_X > 1.1$. State your conclusion in the context of the question. [4]

Q61

9231/41, 9231/43 • October/November 2024 • Question 1

Ellie is investigating the heights of two types of beech tree, A and B, in a certain region. She has chosen a random sample of 60 beech trees of type A in the region, recorded their heights, x m, and calculated unbiased estimates for the population mean and population variance as 35.6 m and $4.95\,\text{m}^2$ respectively.

Ellie also chooses a random sample of 50 beech trees of type B in the region and records their heights, y m. Her results are summarised as follows.

$$\sum y = 1654 \qquad \sum y^2 = 54850$$

Find a 95% confidence interval for the difference between the population mean heights of type A and type B beech trees in the region. [6]

Q62

9231/41, 9231/43 • October/November 2024 • Question 6

Ansal is investigating the wingspans of Monarch butterflies in two different regions, X and Y. He takes a random sample of 8 Monarch butterflies from region X and records their wingspans, x cm. His results are as follows.

$$8.2 \quad 7.0 \quad 7.3 \quad 8.8 \quad 7.8 \quad 8.5 \quad 9.2 \quad 7.4$$

Ansal also takes a random sample of 9 Monarch butterflies from region Y and records their wingspans, y cm. His results are summarised as follows.

$$\sum y = 71.10 \qquad \sum y^2 = 567.13$$

Ansal suspects that the mean wingspan of Monarch butterflies from region X is greater than the mean wingspan of Monarch butterflies from region Y. It is known that the wingspans of Monarch butterflies in regions X and Y are normally distributed with equal population variances.

Test, at the 10% significance level, whether Ansal's suspicion is supported by the data. [8]

Q63

9231/42 • October/November 2024 • Question 1

A scientist is investigating the lengths of the leaves of a certain type of plant. The scientist assumes that the lengths of the leaves of this type of plant are normally distributed. He measures the lengths, x cm, of the leaves of a random sample of 8 plants of this type. His results are as follows.

3.5 4.2 3.8 5.2 2.9 3.7 4.1 3.2

Find a 90% confidence interval for the population mean length of leaves of this type of plant. [4]

Q64

9231/42 • October/November 2024 • Question 5

Dev owns a small company which produces bottles of juice. He uses two machines, X and Y, to fill empty bottles with juice. Dev is investigating the volumes of juice in the bottles. He chooses a random sample of 35 bottles filled by machine X and a random sample of 60 bottles filled by machine Y. The volumes of juice, x and y respectively, measured in suitable units, are summarised by

$$\sum x = 30.8, \qquad \sum x^2 = 29.0, \qquad \sum y = 62.4, \qquad \sum y^2 = 76.8.$$

Dev claims that the mean volume of juice in bottles filled by machine Y is greater than the mean volume of juice in bottles filled by machine X. A test at the $\alpha\%$ significance level suggests that there is sufficient evidence to support Dev's claim.

Find the set of possible values of α. [9]

Q65

9231/41, 9231/42 • May/June 2025 • Question 1

The manager of a hardware store is interested in whether there is a difference in the amount spent per customer on weekdays (x) compared to weekends (y). Random samples of 120 customers on weekdays and 80 customers on weekends are taken and the amount spent by each customer is recorded. The results are summarised as follows.

$$\sum x = 10470 \qquad \sum (x - \bar{x})^2 = 12283 \qquad \sum y = 6560 \qquad \sum (y - \bar{y})^2 = 13520$$

Test at the 1% significance level whether there is a difference in the mean amount spent per customer on weekdays compared to weekends. You should not assume that the population variances of the amounts spent on weekdays and weekends are equal. [7]

Q66

9231/41, 9231/42 • May/June 2025 • Question 4

A researcher is interested in whether there is a difference between two schools in students' aptitude for English. She randomly chooses 10 students from school X and 8 students of a similar age from school Y to take a written English test. The scores for the students from school X (x) and school Y (y) are summarised as follows.

$$\sum x = 612 \qquad \sum x^2 = 40104 \qquad \sum y = 444 \qquad \sum y^2 = 27460$$

You should assume that the two distributions are normal and have the same population variance.

(a) Find a 95% confidence interval for the difference in the mean scores for students from school X and students from school Y in the written English test. [6]

(b) Use the confidence interval you found in part (a) to explain why there is insufficient evidence at a 5% significance level to suggest that the English scores of students from school X and students from school Y are different. [1]

Q67

9231/43 • May/June 2025 • Question 2

A farmer is investigating whether using a new fertiliser will increase the yield of tomato plants. The farmer selects 40 tomato plants at random and gives them the new fertiliser. The crop mass, x kg, of each of these 40 plants is recorded. The farmer selects a further 60 tomato plants at random and gives them a standard fertiliser. The crop mass, y kg, of each of these 60 plants is recorded. The results are summarised as follows.

$$\sum x = 168 \qquad \sum x^2 = 720 \qquad \sum y = 228 \qquad \sum y^2 = 900$$

Find a 90% confidence interval for the difference in mean crop mass associated with each type of fertiliser. [7]

Q68

9231/43 • May/June 2025 • Question 5

A doctor is investigating the concentration of blood glucose in patients at risk of developing type 2 diabetes, where blood glucose is measured in appropriate units. The doctor claims that a particular intervention reduces the concentration by more than k units on average. A group of 8 at risk patients is selected at random and each patient follows the intervention for six months. The blood glucose concentrations before and after the intervention are given in the following table.

Patient	A	B	C	D	E	F	G	H
Before	183	165	172	165	143	176	161	153
After	164	148	164	149	134	153	155	148

(a) Use a t-test at the 5% significance level to find the range of values of k for which the result of the test is to reject the null hypothesis. [7]

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

Q69

9231/44 • May/June 2025 • Question 1

A random sample of 12 observations of a normal random variable is taken. The results give unbiased estimates for the population mean and variance as 10.24 and 0.52 respectively.

Test, at the 10% significance level, the null hypothesis that the population mean is 10.6 against the alternative hypothesis that the population mean is less than 10.6. [4]

Q70

9231/44 • May/June 2025 • Question 6

Lina and Mona are two statisticians who also write songs. The 'time' of a song is the number of minutes for which it lasts. For a random sample of 10 of her songs, Lina calculates a 95% confidence interval for the population mean time, μ minutes. This confidence interval is $2.95 \leqslant \mu \leqslant 3.13$. The times, x minutes, of Lina's songs are normally distributed.

(a) Find the values of $\sum x$ and $\sum x^2$ for the 10 songs in Lina's sample. [5]

Mona's songs have times, y minutes, that are normally distributed. The times for a random sample of 8 of Mona's songs are summarised as follows.

$$\sum y = 24.8 \qquad \sum y^2 = 76.98$$

Mona claims that the population mean time of her songs is greater than the population mean time of Lina's songs.

(b) Assuming that the two distributions have the same population variance, test at the 5% significance level whether there is evidence to support Mona's claim. [8]

Q71

9231/41, 9231/43 • October/November 2025 • Question 1

A group of 10 school children are asked to estimate the size of an angle $\theta°$ in a given acute angled triangle. These estimates, in degrees, are as follows.

84 85 77 85 84 87 86 88 83 85

(a) Stating any assumptions you make, calculate a 95% confidence interval for θ. [5]

(b) Give a reason why the assumptions made in part (a) may not be appropriate in this case. [1]

Q72

9231/41, 9231/43 • October/November 2025 • Question 3

A random sample of 10 newborn baby boys is taken and their masses in kg are recorded. From this sample, the population standard deviation of all newborn baby boys is estimated as 0.6 kg. A random sample of 5 newborn baby girls is taken and their masses in kg are recorded as follows.

3.9 3.1 2.9 3.1 3.6

It is assumed that the masses of newborn baby boys and girls have the same population standard deviation, σ kg.

By pooling the two samples, calculate an estimate of σ. [4]

Q73

9231/41, 9231/43 • October/November 2025 • Question 6

Nine athletes in a club have a new coach. The coach adopts a new training programme which he believes will reduce the race times of these athletes. Each athlete completes a 1500m time trial before and after completing the new training programme. Their times, in seconds (s), are recorded.

Athlete	A	B	C	D	E	F	G	H	I
Time before training (s)	250	251	252	267	276	291	310	320	335
Time after training (s)	245	251	253	261	275	293	302	313	320

(a) Carry out a paired t-test at the 5% significance level to test the coach's belief. [7]

Further research suggests that the effects of the training programme tend to reduce the times of the slower athletes by more than those of the faster athletes.

(b) Suggest a reason why the paired t-test used in part (a) may not have been an appropriate test in this case. [1]

(c) Suggest a suitable alternative test that could have been used instead of a paired t-test. [1]

Q74

9231/42 • October/November 2025 • Question 2

A factory produces packets of biscuits. The total mass of biscuits in a packet has a normal distribution with mean μ. A random sample of 12 packets is taken and the mass of the contents of each packet, x g, is recorded. The results are summarised as follows.

$$\sum x = 2390 \qquad \sum x^2 = 476117$$

(a) Find a 99% confidence interval for μ. [4]

A test of the null hypothesis $\mu = k$ is carried out on this sample using a 5% significance level. The test does not support the alternative hypothesis $\mu < k$.

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

Q75

9231/42 • October/November 2025 • Question 5

An engineer is comparing the tensile strengths of steel rods made from two machines, A and B. The engineer randomly selects 8 rods from machine A and 6 rods from machine B. The tensile strengths, in appropriate units, are given in the following table.

Machine A	402	403	415	412	409	407	406	410
Machine B	401	398	395	397	410	405		

You should assume that the two distributions are normal and have the same population variance.

Use a t-test at the 5% significance level to test whether there is any difference in the mean tensile strengths of steel rods from the two machines. [9]

Q76

9231/44 • October/November 2025 • Question 1

Local residents are concerned about the speed of cars travelling through their village. They record the speed, in $\mathrm{km\,h^{-1}}$, of a random sample of 13 cars travelling through their village. The recorded speeds are as follows.

40 53 59 42 43 48 62 67 46 82 66 45 70

(a) Construct a 90% confidence interval for the population mean speed of cars passing through the village. [5]

(b) State an assumption that is necessary for the confidence interval found in 1(a) to be valid. [1]

Q77

9231/44 • October/November 2025 • Question 3

Two different types of juice extractor, machine X and machine Y, are being compared. The manufacturer claims that machine Y extracts more juice per orange on average than machine X.

A random sample of 20 oranges is selected and each carefully measured. Oranges of a similar size and mass are then paired and numbered from 1 to 10. One orange from each pair is randomly allocated to machine X with the other allocated to machine Y. The amount of juice, in ml, extracted from each orange is recorded in the following table.

pair	1	2	3	4	5	6	7	8	9	10
machine X	65	73	58	61	72	79	64	65	69	71
machine Y	68	72	64	63	75	82	63	63	72	74

(a) Use a t-test to test the manufacturer's claim at the 1% significance level. [7]

(b) State an assumption required for the test in 3(a) to be valid. [1]

The manufacturer notices that the amount of juice extracted from the oranges in pair 4 were recorded incorrectly. In fact, the amount of juice extracted from machine X was 60 ml and from machine Y was 62 ml.

(c) Explain, with justification, whether the conclusion of the test in 3(a) remains the same. [2]

Answers available in the companion mark scheme booklet.