Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

Monday 23 June 2025

Afternoon (Time: 1 hour 30 minutes) | Paper reference **9FM0/4B**

Further Mathematics

Advanced

PAPER 4B: Further Statistics 2

You must have:
Mathematical Formulae and Statistical Tables (Green), calculator

Total Marks

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 7 questions in this question paper. The total mark for this paper is 75.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ▶

P75690A

©2025 Pearson Education Ltd.
Y:1/1/1/

P 7 5 6 9 0 A 0 1 2 4

Pearson

1. A medical researcher is exploring the relationship between the weight, x kg, and the head circumference, y cm, of newborn babies. A random sample of 15 newborn babies is taken and the data are summarised by the following statistics

$$\sum x = 50.46 \quad \sum x^2 = 171.828 \quad \sum y = 518.9 \quad \sum y^2 = 18004.47 \quad S_{xy} = 7.8284$$

(a) Show that, to 3 decimal places, $S_{xx} = 2.081$ and $S_{yy} = 53.989$

(3)

(b) Find the equation of the regression line of y on x, giving your answer in the form $y = a + bx$

(3)

(c) Find the residual sum of squares (RSS) for these data.

(1)

One of these 15 babies had a birth weight of 3.26 kg and a head circumference of 36.8 cm

(d) Find the residual for this baby.

(2)

The researcher claims that this baby could be an outlier.

(e) Using your answers to part (c) and part (d), give a reason that might support this claim.

(1)

(a) $S_{xx} = 171.828 - \frac{50.46^2}{15} = 2.08056 = 2.081$ (3 d.p.)

$S_{yy} = 18004.47 - \frac{518.9^2}{15} = 53.989333... = 53.989$ (3 d.p.)

(b) $b = \frac{S_{xy}}{S_{xx}} = \frac{7.8284}{2.08056} = 3.762640...$

$a = \bar{y} - b\bar{x} = \frac{518.9}{15} - 3.762640... \times \frac{50.46}{15}$

$= 34.593333... - 12.657522... = 21.935810...$

$y = 21.9 + 3.76x$

(c) $RSS = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 53.989333... - \frac{7.8284^2}{2.08056}$

$= 24.533875... = 24.5$

(d) $\hat{y} = 21.935810... + 3.762640... \times 3.26 = 34.202018...$

Residual $= y - \hat{y} = 36.8 - 34.202018... = 2.597981... = 2.60$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 1 continued

(e) Residual^2 = 2.597981...^{2} = 6.749... ≈ 6.75

$\frac{6.75}{24.5}$ = 0.275..., so this one baby's residual squared makes up about 27.5% of the RSS, a large proportion for 1 of the 15 babies, so this point could be an outlier.

Question 1 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 1 continued

(Total for Question 1 is 10 marks)

2. Two lecturers are marking a large batch of students' essays, to place them in rank order.

A random sample of 9 students is taken and their essays are marked by both lecturers.

The marks awarded are given in the table below.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Lecturer 1	10	16	5	17	12	7	14	11	8
Lecturer 2	17	22	19	24	21	15	23	18	20

(a) Calculate Spearman's rank correlation coefficient for these data.

(4)

(b) Carry out a suitable test to assess whether the lecturers agree about the rank order of the essays. You should use a 5% level of significance and state your hypotheses and critical value clearly.

(3)

Student	A	B	C	D	E	F	G	H	I
Rank 1	6	2	9	1	4	8	3	5	7
Rank 2	8	3	6	1	4	9	2	7	5
d	−2	−1	3	0	0	−1	1	−2	2
d^2	4	1	9	0	0	1	1	4	4

(a) $\Sigma d^2 = 4 + 1 + 9 + 0 + 0 + 1 + 1 + 4 + 4 = 24$

$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 24}{9 \times (9^2-1)} = 1 - \frac{144}{720}$

$r_s = 0.8$

(b) H_0: $\rho = 0$ H_1: $\rho > 0$

Critical value ($n = 9$, 5% one-tail): 0.6000

$0.8 > 0.6$, so significant: reject H_0

There is evidence that the lecturers agree about the rank order of the essays.

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 7 marks)

3. A scientist investigated the effects of nitrogen-fixing bacteria on cereal crop yield, x kg ha^{-1}

The data come from independent random samples from normal distributions.

The results are summarised in the table below.

	n	$\bar{x}$	s_x
Without nitrogen-fixing bacteria	6	1648	143.8
With nitrogen-fixing bacteria	5	1910	64.0

(a) Test whether or not the variances of the cereal crop yields are the same. You should use a 10% level of significance and state your hypotheses and critical value clearly.

(4)

A suitable test to determine whether there is an increase in mean cereal crop yield has a p-value of 0.002

(b) (i) Explain the importance of the conclusion of the test in part (a) to this test.

(ii) Comment on the effectiveness of nitrogen-fixing bacteria on cereal crop yields.

(2)

(a) H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 \neq \sigma_2^2$

(1 = without bacteria, 2 = with bacteria)

$F = \frac{s_1^2}{s_2^2} = \frac{143.8^2}{64.0^2} = 5.04844...$

Critical value (10% two-tailed): $F_{5,4}(0.05) = 6.26$

5.05 < 6.26, so not significant: do not reject H_0

There is insufficient evidence of a difference in the variances of the cereal crop yields.

(b)(i) The t-test for a difference in means requires the two populations to have equal variances (so that a pooled estimate of σ^2 can be used). Part (a) shows the variances can be assumed equal, so the test is valid.

(b)(ii) $p = 0.002 < 0.05$ (very small), so there is strong evidence that the nitrogen-fixing bacteria increase the mean cereal crop yield (but they do not appear to affect the variance).

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 3 continued

(Total for Question 3 is 6 marks)

4. The weights of apples Sam grows are normally distributed with a mean of 131 g and a standard deviation of 6 g

A supermarket will only buy apples with weights in the range 125 g to 137 g

(a) Find the proportion of apples Sam grows that the supermarket will buy.

(1)

Some of Sam's apple trees are treated with the aim of reducing the variability in weight of apples, without significantly affecting the mean weight. The weights of apples are still normally distributed.

A random sample of 25 apples was taken from treated trees and the weight, x grams, of each apple was recorded. The results are summarised by the following statistics

$$n = 25 \quad \sum x = 3250 \quad \sum x^2 = 422\,862$$

(b) Use a 5% level of significance to carry out a suitable test to determine whether there is evidence that the variance of the weights of apples from treated trees has decreased.
You should clearly state your test statistic, hypotheses and critical value.

(6)

(c) Calculate a 95% confidence interval for the mean weight of apples produced by Sam's trees after treatment.
You should clearly state the formula and numerical expression you have used.

(3)

(d) (i) State, giving a reason, whether the aims of the treatment have been met.

(ii) Give an estimate of the proportion of apples the supermarket will buy from Sam's trees that have been treated.

(3)

(a) $W \sim N(131, 6^2)$: $P(125 < W < 137) = P(-1 < Z < 1) = 0.68268...$

Proportion = 0.683

(b) $H_0: \sigma^2 = 6^2$ $\quad H_1: \sigma^2 < 6^2$

$s^2 = \frac{422862 - 25 \times (3250 \div 25)^2}{24} = \frac{362}{24} = 15.0833...$

Test statistic $\frac{(n-1)s^2}{\sigma^2} = \frac{24 \times 15.0833...}{6^2} = 10.0555... = 10.1$

Critical value (5%, lower tail): $\chi^2_{24}(0.95) = 13.848$

$10.1 < 13.848$, so significant: reject H_0

There is evidence that the variance of the weights of apples from treated trees has decreased (the treatment works).

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 4 continued

(c) $\bar{x} = \frac{3250}{25} = 130$, $s = \sqrt{15.0833...} = 3.883726...$

95% CI: $\bar{x} \pm t_{24}(0.025) \times \frac{s}{\sqrt{n}}$

$= 130 \pm 2.064 \times \frac{3.883726...}{\sqrt{25}} = 130 \pm 1.603202...$

$= (128.396..., 131.603...)$

95% CI: (128, 132) grams

(d)(i) Yes, the aims have been met: 131 lies inside the confidence interval, so the mean weight has not significantly changed, and part (b) shows the variance has decreased.

(d)(ii) Treated weights: $W \sim N(131, 3.883726...^2)$

$P(125 < W < 137) = P(|Z| < \frac{6}{3.883726...}) = P(-1.5449 < Z < 1.5449)$

$= 0.87763... \approx 0.878$

Estimated proportion bought ≈ 0.878 (about 88%)

Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 4 continued

(Total for Question 4 is 13 marks)

5. The random variable $X \sim \mathrm{U}[1, 4]$

(a) Find

(i) $\mathrm{P}(1.8 < X < 3.2)$

(ii) $\mathrm{P}(X > 3.2 \mid X > 1.8)$

(3)

A random sample of 10 observations of X is taken.

The random variable M represents the maximum value of these 10 observations.

The cumulative distribution function of M, $\mathrm{F}(y)$, is given by

$$\mathrm{F}(y) = \begin{cases} 0 & y < 1 \\ \left(\dfrac{y-1}{3}\right)^{10} & 1 \leqslant y \leqslant 4 \\ 1 & y > 4 \end{cases}$$

(b) Find the probability that the maximum value in the sample is greater than 3.75

(1)

(c) (i) Sketch the probability density function of M for $1 \leqslant y \leqslant 4$

(ii) Write down the mode of M

(3)

(d) Use algebraic integration to find the exact value of $\mathrm{E}(M)$

(4)

(a)(i) $P(1.8 < X < 3.2) = \frac{3.2 - 1.8}{4 - 1} = \frac{1.4}{3} = \frac{7}{15}$

(a)(ii) $P(X > 3.2 \mid X > 1.8) = \frac{P(X > 3.2)}{P(X > 1.8)} = \frac{0.8 \div 3}{2.2 \div 3} = \frac{0.8}{2.2} = \frac{4}{11}$

(b) $P(M > 3.75) = 1 - F(3.75) = 1 - \left(\frac{3.75 - 1}{3}\right)^{10}$

$= 1 - 0.418903... = 0.581096... = 0.581$

(c)(i) $f(y) = \frac{dF(y)}{dy} = 10 \times \left(\frac{y-1}{3}\right)^9 \times \frac{1}{3} = \frac{10}{3}\left(\frac{y-1}{3}\right)^9, \quad 1 \leq y \leq 4$

$f(1) = 0, \quad f(4) = \frac{10}{3}$

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

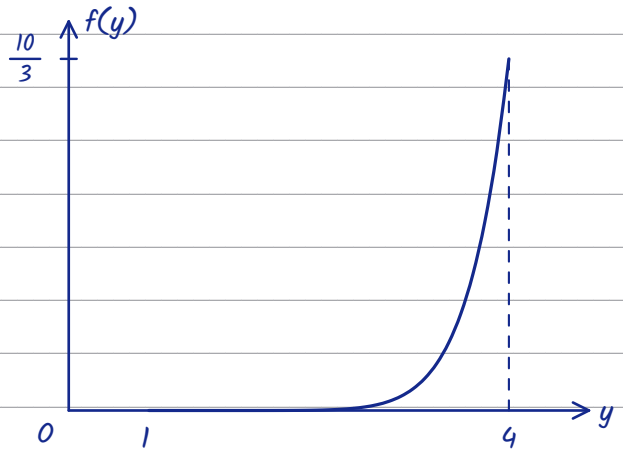
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 5 continued

(c)(i) Sketch:

$f(y)$ increasing from
$f(1) = 0$ to $f(4) = \frac{10}{3}$

(zero outside $[1, 4]$)



(c)(ii) Mode of $M = 4$

(d) $E(M) = \int_1^4 y\, f(y)\, dy = \int_1^4 y \times \frac{10}{3}\left(\frac{y-1}{3}\right)^9 dy$

By parts: $u = y$, $\frac{dv}{dy} = \frac{10}{3}\left(\frac{y-1}{3}\right)^9 \Rightarrow \frac{du}{dy} = 1$, $v = \left(\frac{y-1}{3}\right)^{10}$

$E(M) = \left[y\left(\frac{y-1}{3}\right)^{10}\right]_1^4 - \int_1^4 \left(\frac{y-1}{3}\right)^{10} dy$

$= \left[y\left(\frac{y-1}{3}\right)^{10}\right]_1^4 - \left[\frac{3}{11}\left(\frac{y-1}{3}\right)^{11}\right]_1^4$

$= (4 \times 1^{10} - 1 \times 0) - \left(\frac{3}{11} \times 1^{11} - 0\right)$

$= 4 - \frac{3}{11} = \frac{41}{11}$

$E(M) = \frac{41}{11}$

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 5 continued

(Total for Question 5 is 11 marks)

6. Xiang owns a takeaway restaurant and wants to employ a delivery company. Xiang uses two companies, A and B, for one week each and records the delivery rate, x minutes per km, for each delivery.

Company A made 61 deliveries with $\bar{x}_A = 3.7$ and $s_A^2 = 1.64$

Company B made 51 deliveries with $\sum x_B = 142.8$ and $\sum x_B^2 = 461.34$

(a) Find unbiased estimates for the mean and variance of the delivery rate for company B.

(3)

Xiang wants to carry out a two-sample t-test to determine whether there is a difference in the mean delivery rate for these companies.

(b) State any assumptions Xiang must make about the delivery rates of the two companies.

(1)

(c) Using a 5% level of significance, carry out Xiang's test. You should clearly state the hypotheses, test statistic and critical value you use.

(7)

Xiang noticed that a greater proportion of company A's deliveries were in the town, with lots of traffic, whereas a greater proportion of company B's deliveries were rural, with less traffic.

Xiang decides to carry out a further test. He selects a random sample of 10 customers who regularly place an order at the same time each week. He asks company A to deliver these orders one week and company B the next week. There were no roadworks or other incidents affecting travel time in either week.

The delivery rates are given in the table below.

Customer	1	2	3	4	5	6	7	8	9	10
Company A	5.4	3.6	2.5	4.7	2.3	3.2	3.4	4.1	5.9	4.5
Company B	4.9	3.9	2.3	4.5	2.6	2.8	3.1	3.8	5.2	4.3

(d) Use a 5% level of significance to carry out a suitable test to determine whether there is a difference in the delivery rates of the two companies.
You should clearly state your hypotheses, test statistic and critical value.

(6)

(a) $\bar{x}_B = \frac{142.8}{51} = 2.8$

$s_B^2 = \frac{461.34 - 51 \times 2.8^2}{50} = \frac{61.5}{50} = 1.23$

(b) The delivery rates for both companies are normally distributed with equal variances.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 6 continued

(c) H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$

$$s_p^2 = \frac{60 \times 1.64 + 50 \times 1.23}{110} = \frac{159.9}{110} = 1.453636... = 1.45$$

$$t = \frac{3.7 - 2.8}{\sqrt{(1.453636... \times (1/61 + 1/51))}} = 3.934192... = 3.93$$

Critical value (5% two-tailed): $t_{110}(0.025) = 1.982$

$3.93 > 1.982$, so significant: reject H_0

There is evidence of a difference in the mean delivery rates of the two companies.

(d) Paired t-test. Let D = rate for A − rate for B

H_0: $\mu_D = 0$ H_1: $\mu_D \neq 0$

d: 0.5, −0.3, 0.2, 0.2, −0.3, 0.4, 0.3, 0.3, 0.7, 0.2

$\Sigma d = 2.2$, $\Sigma d^2 = 1.38$

$$\bar{d} = \frac{2.2}{10} = 0.22, \quad s_D^2 = \frac{1.38 - 10 \times 0.22^2}{9} = 0.099555..., \quad s_D = 0.315524...$$

$$t = \frac{0.22 - 0}{0.315524... \div \sqrt{10}} = 2.204905... = 2.20$$

Critical value (5% two-tailed): $t_9(0.025) = 2.262$

$2.20 < 2.262$, so not significant: do not reject H_0

There is insufficient evidence of a difference in the delivery rates of the two companies.

Question 6 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 6 continued

(Total for Question 6 is 17 marks)

7. A random sample X_1, X_2, X_3, X_4 and X_5 is taken from a distribution $X \sim \mathrm{N}(\mu, \sigma^2)$

Nina and Jonah each use unbiased estimators to estimate μ

Nina uses $\bar{X} = \frac{1}{5}\sum_{r=1}^{5} X_r$ but Jonah uses $Y = \frac{X_1 + X_5}{2}$

(a) Find $\mathrm{P}(\bar{X} - Y > \sigma)$

(7)

Nina and Jonah both calculated 95% confidence intervals for μ

They used their estimates and the normal distribution with the same value of σ

Their intervals were

(21.823, 38.177) and (27.029, 37.371)

(b) State, giving a reason, which of these intervals Nina calculated.

(1)

(c) Using one of these intervals, find the value of σ

(3)

(a) Let $D = \bar{X} - Y$. $E(D) = \mu - \mu = 0$

$D = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} - \frac{X_1 + X_5}{2}$

$= \frac{1}{10}[2(X_2 + X_3 + X_4) - 3(X_1 + X_5)]$

$\mathrm{Var}(D) = \frac{1}{100}(2^2 \times 3\sigma^2 + 3^2 \times 2\sigma^2) = \frac{1}{100}(12\sigma^2 + 18\sigma^2) = \frac{3}{10}\sigma^2$

$D \sim N(0, \frac{3}{10}\sigma^2)$

$P(D > \sigma) = P\left(Z > \frac{\sigma - 0}{\sigma\sqrt{(3/10)}}\right) = P(Z > 1.825741\ldots)$

$= 1 - 0.966055\ldots = 0.033944\ldots = 0.0339$

(b) Nina's interval is (27.029, 37.371), since it is the narrower interval: Nina's estimator $\bar{X}$ uses a larger sample ($n = 5$ rather than 2), so it has a smaller variance and gives a narrower confidence interval.

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Question 7 continued

(c) Width of Nina's interval: $37.371 - 27.029 = 10.342$

$10.342 = 2 \times 1.96 \times \frac{\sigma}{\sqrt{5}}$

$\sigma = \frac{10.342 \times \sqrt{5}}{2 \times 1.96} = 5.899340...$

$\sigma = 5.90$

Question 7 continued

(Total for Question 7 is 11 marks)

TOTAL FOR PAPER IS 75 MARKS

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA