Please check the examination details below before entering your candidate information

Candidate surname	Other names
MODEL SOLUTIONS	

Centre Number

Candidate Number

Pearson Edexcel Level 3 GCE

Monday 26 June 2023

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 allowed 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 8 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

P74084A

N:1/1/1/1/1/1/

P 7 4 0 8 4 A 0 1 2 0

Pearson

1. Baako is investigating the times taken by children to run a 100 m race, x seconds, and a 500 m race, y seconds. For a sample of 20 children, Baako obtains the time taken by each child to run each race.

Here are Baako's summary statistics.

$$S_{xx} = 314.55 \qquad S_{yy} = 9026 \qquad S_{xy} = 1610$$

$$\bar{x} = 19.65 \qquad \bar{y} = 108$$

(a) Calculate the product moment correlation coefficient between the times taken to run the 100 m race and the times taken to run the 500 m race.

(2)

(b) Show that the equation of the regression line of y on x can be written as

$$y = 5.12x + 7.42$$

where the gradient and y intercept are given to 3 significant figures.

(3)

The child who completed the 100 m race in 20 seconds took 104 seconds to complete the 500 m race.

(c) Find the residual for this child.

(1)

The table below shows the signs of the residuals for the 20 children in order of finishing time for the 100 m race.

Sign of residual	+	+	+	+	–	–	+	–	–	–	–	–	–	–	–	+	+	+	+	+

(d) Explain what the signs of the residuals show about the model's predictions of the 500 m race times for the children who are fastest and slowest over the 100 m race.

(1)

a) From the Formula Booklet, we know that

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$r = \frac{1610}{\sqrt{314.55 \times 9026}} \text{ (1)} = 0.9555 \text{ (1)}$$

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

b) $y = ax + b$

From the Formula Booklet, we know that

$b = \frac{S_{xy}}{S_{xx}}$, $\bar{a} = \bar{y} - b\bar{x}$

$b = \frac{1610}{314.55} = 5.118$ 1

$a = 108 - 5.118 \times 19.65 = 7.4229$ 1

$\Rightarrow y = 5.12x + 7.42$ 1

c) $512(20) + 7.42 = 104.82$

$104 - 104.82 = -5.82$ 1

d) Positive residuals show that the model underestimates the 500m race times of the fastest and slowest children. 1

(Total for Question 1 is 7 marks)

P 7 4 0 8 4 A 0 3 2 0

2. Camilo grows two types of apple, green apples and red apples.

The standard deviation of the weights of green apples is known to be 3.5 grams.

A random sample of 80 green apples has a mean weight of 128 grams.

(a) Find a 98% confidence interval for the mean weight of the population of green apples. Show your working clearly and give the confidence interval limits to 2 decimal places.
(3)

Camilo believes that the mean weight of the population of green apples is more than 10 grams greater than the mean weight of the population of red apples.

A random sample of n red apples has a mean weight of 117 grams.

The standard deviation of the weights of the red apples is known to be 4 grams.

A test of Camilo's belief is carried out at the 5% level of significance.

(b) State the null and alternative hypotheses for this test.
(1)

(c) Find the smallest value of n for which the null hypothesis will be rejected.
(6)

(d) Explain the relevance of the Central Limit Theorem in parts (a) and (c).
(1)

(e) Given that $n = 85$, state the conclusion of the hypothesis test.
(1)

a) Recall that for a confidence interval with a known standard deviation, we use the formula.

$$x \pm z \times \frac{\sigma}{\sqrt{n}}$$

From the table, $z = 2.3263$ 1

$$128 \pm 2.3263 \times \frac{3.5}{\sqrt{80}}$$ 1

So the confidence interval is

$(127.09, 128.91)$ 1

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

Question 2 continued

b) $H_0: \mu_g = \mu_r + 10$

$H_1: \mu_g > \mu_r + 10$ 1

c) From the formula booklet, we will have a test statistic

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

$\bar{G} - \bar{R} \sim N\left(10, \frac{3.5^2}{80} + \frac{4^2}{n}\right)$ 1

From the table, $z = 1.6449$. 1 (one tailed)

Our test statistic is

$$\frac{128 - 117 - 10}{\sqrt{\frac{3.5^2}{80} + \frac{4^2}{n}}} > 1.6449$$ 1 1

$$\Rightarrow \frac{1}{\sqrt{0.153125 + \frac{16}{n}}} > 1.6449$$

$$\Rightarrow \frac{1}{0.153125 + 16/n} > 2.705$$

$$\Rightarrow \frac{1}{2.705} > 0.153125 + 16/n$$

$$\Rightarrow n > 73.91$$ 1

So the smallest value of n that we reject is 74. 1

Question 2 continued

d) The sample sizes are large so CLT means that we do not know the distributions of G and R but the means $\bar{G}$ and of $\bar{G} - \bar{R}$ are approximately normally distributed. 1

e) $85 > 73.9$ so there is significant evidence to support Camilo's belief that the mean of green apples is more than 10 grams greater than red apples. 1

3. Two machines, A and B, are used to fill bottles of water. The amount of water dispensed by each machine is normally distributed.

Samples are taken from each machine and the amount of water, x ml, dispensed in each bottle is recorded. The table shows the summary statistics for Machine A.

	Sample size	$\sum x$	$\sum x^2$
Machine A	9	2268	571 700

(a) Find a 95% confidence interval for the variance of the amount of water dispensed in each bottle by Machine A.

(4)

For Machine B, a random sample of 11 bottles is taken. The sample variance of the amount of water dispensed in bottles is 12.7 ml²

(b) Test, at the 10% level of significance, whether there is evidence that the variances of the amounts of water dispensed in bottles by the two machines are different. You should state the hypotheses and the critical value used.

(4)

a) Recall that the formula for a confidence interval for the variance is

$$\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha/2)}$$ for a $100(1-\alpha)\%$ interval.

Also, from the formula booklet, we have that

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)$$

$$\bar{x} = \frac{2268}{9} = 252$$

$$s^2 = \frac{1}{8}\left(571700 - 9(252)^2\right)$$

$\Rightarrow$ $s^2 = 20.5$ 1

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

DO NOT WRITE IN THIS AREA

Question 3 continued

From the table,

$\chi^2_8(0.025) = 17.535$

$\chi^2_8(0.975) = 2.180$ 1

$\frac{(9-1)(20.5)}{9.3527} = 9.3527$

$\frac{(9-1)(20.5)}{2.180} = 75.2293$ 1

So our confidence interval is

$(9.3527, 75.2293)$ 1

$H_0: \sigma_A^2 = \sigma_B^2$

$H_1: \sigma_A^2 \neq \sigma_B^2$ 1

From the formula booklet, we have that

$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F_{n_x-1, n_y-1}$ and here $\sigma_x^2 = \sigma_y^2$ by the null hypothesis.

$\frac{S_A^2}{S_B^2} \sim F_{8,10}$ $\frac{20.5}{12.7} = 1.614$ 1

From the table, our critical value is 3.07. 1

$1.614 < 3.07$ so we do not reject H_0 as there is insufficient evidence to suggest that the variances are different. 1

(Total for Question 3 is 8 marks)

4. The weights of eggs, E grams, follow a normal distribution, $N(60, 3^2)$

The weights of empty small boxes, S grams, follow a normal distribution, $N(24, 1.8^2)$

The weights of empty large boxes, L grams, follow a normal distribution, $N(40, 2.1^2)$

Small boxes of eggs contain 6 randomly selected eggs.

Large boxes of eggs contain 12 randomly selected eggs.

(a) Find the probability that the total weight of a randomly selected small box of 6 eggs weighs less than 387 grams.

(3)

(b) Find the probability that a randomly selected large box of 12 eggs weighs more than twice a randomly selected small box of 6 eggs.

(5)

a) Let E_i be the weight of each egg and, S and L be weights of the small and large boxes respectfully.

$W = E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + S$

$E[W] = 6E[E] + E[S] = 384$ ①

$Var(W) = 6Var(E) + Var(S) = 6 \times 3^2 + 1.8^2 = 57.24$ ①

$Pr(W < 387) = 0.6541$ ① using a calculator

b) $X = E_1 + \dots + E_{12} + L$

$E[X] = 12E[E] + E[L] = 760$

$Var(X) = 12Var(E) + Var(L) = 112.41$ ①

Let $Y = X - 2W$ ①

$E[Y] = 760 - 2(384) = -8$ ①

$Var(Y) = Var(X) + 4Var(W) = 112.41 + 4(57.24)$
$= 341.37$ ①

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 DO NOT WRITE IN THIS AREA

Question 4 continued

So $Y \sim N(-8, 341.37)$

$\Pr(Y>0) = 1 - \Pr(Y<0) = 0.3325$ 1

using a calculator.

(Total for Question 4 is 8 marks)

5. A psychologist claims to have developed a technique to improve a person's memory.

A random sample of 8 people are each given the same list of words to memorise and recall.

Each person then receives memory training from the psychologist. After the training, each person is given the same list of new words to memorise and recall.

The table shows the percentage of words recalled by each person before and after the training.

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Percentage of words recalled before training	24	33	33	39	30	38	32	34
Percentage of words recalled after training	28	30	37	41	32	44	35	34

(a) State why a paired t-test is suitable for these data.
(1)

(b) State an assumption that needs to be made in order to carry out a paired t-test in this case.
(1)

(c) Test, at the 5% level of significance, whether or not there is evidence of an increase in the percentage of words recalled after receiving the psychologist's training. State your hypotheses, test statistic and critical value used for this test.
(7)

a) The before and after results are not independent of each other. 1

b) The difference in scores are normally distributed. 1

c) Make a new row of the differences.

d: 4 -3 4 2 2 6 3 0 1

$\bar{d} = \pm 2.25,$

Using the Formula booklet,

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

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

$$S^2 = \frac{1}{7}\left[(4-2.25)^2 + \ldots + (0-2.25)^2\right]$$

$$= 7.46429 \quad (1)$$

$H_0: \mu_d = 0$

$H_1: \mu_d > 0 \quad (1)$

Recall that our test statistic is

$$\frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t_{n-1}$$

So our test statistic is

$$\frac{\pm 2.25 - 0}{\sqrt{\frac{7.46}{8}}} = \pm 2.3019 \quad (1) \quad (1)$$

Using the t-table and $8 - 1 = 7$ degrees of freedom,

our critical value is ± 1.895 (1)

$2.3019 > 1.895$

So, we reject Ho as there is sufficient evidence to support the psycologists claim that there was an increase in the percentage of words recalled. (1)

(Total for Question 5 is 9 marks)

6. The continuous random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x < 0 \\ k(x - ax^2) & 0 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$$

The values of a and k are positive constants such that $P(X < 2) = \frac{2}{3}$

(a) Find the exact value of the median of X

(6)

(b) Find the probability density function of X

(2)

(c) Hence, deduce the value of the mode of X, giving a reason for your answer.

(2)

a) $F(2) = 2/3$

$\Rightarrow k(2 - 4a) = 2/3 \quad (1)$

$F(4) = 1$

$\Rightarrow k(4 - 16a) = 1$ ①

$\Rightarrow k = \frac{1}{4 - 16a}$ sub this into (1)

$\Rightarrow \frac{2 - 4a}{4 - 16a} = \frac{2}{3}$ ①

$\Rightarrow 3 - 6a = 4 - 16a$

$\Rightarrow 10a = 1$

$\Rightarrow a = \frac{1}{10} \Rightarrow k = \frac{5}{12}$ ①

For the median, we want the value of m such that $F(m) = 1/2$

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

Question 6 continued

$F(m) = 1/12$

$\Rightarrow \frac{5}{12}\left(m - \frac{1}{10}m^2\right) = \frac{1}{2}$ (1)

$\Rightarrow \frac{5m}{12} - \frac{1}{24}m^2 = \frac{1}{2}$

$\Rightarrow 10m - m^2 = 12$

$\Rightarrow m^2 - 10m + 12 = 0$ (1)

$\Rightarrow m = 5 \pm \sqrt{13}$ Using a calculator.

reject $5+\sqrt{13}$ because it is greater than 4.

$\Rightarrow m = 5 - \sqrt{13}$ (1)

b) Recall that $f(x) = \frac{d}{dx}(F(x))$

$f(x) = \frac{d}{dx}\left[\frac{5}{12}\left(x - \frac{1}{10}x^2\right)\right]$ (1)

$= \frac{5}{12}\left(1 - \frac{1}{5}x\right)$

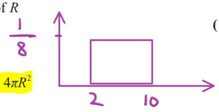
So $f(x) = \begin{cases} \frac{5}{12}\left(1 - \frac{1}{5}x\right) & 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$ (1)

b) The mode is when $X = 0$ (1) because $f(x)$ is a decreasing function. (1)

(Total for Question 6 is 10 marks)

7. The random variable R has a continuous uniform distribution over the interval [2, 10]

(a) Write down the probability density function f(r) of R



(1)

A sphere of radius R cm is formed.

The surface area of the sphere, S cm^2, is given by $S = 4\pi R^2$

(b) Show that $\mathrm{E}(S) = \dfrac{496\pi}{3}$

(4)

The volume of the sphere, V cm^3, is given by $V = \dfrac{4}{3}\pi R^3$

(c) Find, using algebraic integration, the expected value of V

(4)

a) $f(r) = \begin{cases} 1/8 & 2 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$ (1)

b) Recall that $E[g(x)] = \int g(x) f(x)\,dx$

$E[S] = E[4\pi r^2] = \int_2^{10} 4\pi r^2 \times \frac{1}{8}\,dr$ (1)

$= \frac{\pi}{2}\left[\frac{r^3}{3}\right]_2^{10}$ (1)

$= \frac{\pi}{2}\left[\frac{10^3}{3} - \frac{2^3}{3}\right]$ (1)

$= \frac{496\pi}{3}$ (1)

c) $E[V] = E[\tfrac{4}{3}\pi r^3] = \int_2^{10} \frac{4}{3}\pi r^3 \times \frac{1}{8}\,dr$ (1)

$= \frac{\pi}{6}\left[\frac{r^4}{4}\right]_2^{10}$ (1)

Question 7 continued

$$= \frac{\pi}{6}\left[\frac{10^4}{4} - \frac{2^4}{4}\right] \quad (1)$$

$$= 416\pi \quad (1)$$

(Total for Question 7 is 9 marks)

8. A bag contains a large number of marbles of which an unknown proportion, p, is yellow.

Three random samples of size n are taken, and the number of yellow marbles in each sample, Y_1, Y_2 and Y_3, is recorded. Binomial

Two estimators $\hat{p}_1$ and $\hat{p}_2$ are proposed to estimate the value of p

$$\hat{p}_1 = \frac{Y_1 + 3Y_2 - 2Y_3}{2n}$$

$$\hat{p}_2 = \frac{2Y_1 + 3Y_2 + Y_3}{6n}$$

(a) Show that $\hat{p}_1$ and $\hat{p}_2$ are both unbiased estimators of p **(3)**

(b) Find the variance of $\hat{p}_1$ **(2)**

The variance of $\hat{p}_2$ is $\dfrac{7p(1-p)}{18n}$

(c) State, giving a reason, which is the better estimator. **(2)**

The estimator $\hat{p}_3 = \dfrac{Y_1 + aY_2 + 3Y_3}{bn}$ where a and b are positive integers.

(d) Find the pair of values of a and b such that $\hat{p}_3$ is a better unbiased estimator of p than both $\hat{p}_1$ and $\hat{p}_2$
You must show all stages of your working. **(5)**

a) $Y \sim Bin(n, p) \Rightarrow E[Y] = np$ (1)

$E[Y_i] = E[Y]$ for all $i = 1, 2, 3$

$E[\hat{p}_1] = \dfrac{np + 3np - 2np}{2n} = p$ (1)

$E[\hat{p}_2] = \dfrac{2np + 3np + np}{6n} = p$

Hence, $\hat{p}_1$ and $\hat{p}_2$ are unbiased estimators of p. (1)

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 8 continued

b) $\mathrm{Var}(Y) = np(1-p)$ from the Formula Booklet

$$\mathrm{Var}(\hat{p}_1) = \mathrm{Var}\left(\frac{Y_1}{2n}\right) + \mathrm{Var}\left(\frac{3Y_2}{2n}\right) + \mathrm{Var}\left(\frac{Y_3}{n}\right) \quad 1$$

$$= \frac{1}{4n^2}\mathrm{Var}(Y) + \frac{9}{4n^2}\mathrm{Var}(Y) + \frac{1}{n^2}\mathrm{Var}(Y)$$

$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$

$$= \frac{7}{2n^2}\mathrm{Var}(Y)$$

$$= \frac{7}{2n^2}np(1-p)$$

$$= \frac{7p(1-p)}{2n} \quad 1$$

c) $\frac{7p(1-p)}{2n} > \frac{7p(1-p)}{18n}$

as $\frac{7}{2} > \frac{7}{18}$

So $\mathrm{Var}(\hat{p}_1) > \mathrm{Var}(\hat{p}_2)$ 1

Also, both estimators are unbiased.

So $\hat{p}_2$ is the better estimator. 1

d) $\hat{p}_3$ is unbiased so $E[\hat{p}_3] = p$

$$E[\hat{p}_3] = \frac{np + anp + 3np}{bn} = \frac{4+a}{b}p$$

Question 8 continued

$$\Rightarrow \frac{4+a}{b} = 1 \Rightarrow 4+a = b \quad (1) \quad ①$$

$$Var(\hat{p}_3) = Var\left(\frac{Y_1}{bn}\right) + Var\left(\frac{aY_2}{bn}\right) + Var\left(\frac{3Y_3}{bn}\right)$$

$$= \frac{1}{b^2n^2} Var(Y) + \frac{a^2}{b^2n^2} Var(Y) + \frac{9}{b^2n^2} Var(Y)$$

$$= \frac{a^2+10}{b^2n^2} Var(Y)$$

$$= \frac{p(1-p)}{b^2n}(a^2+10)$$

For $\hat{p}_3$ to be a better estimator, we have

$$\frac{p(1-p)}{b^2n}(a^2+10) < \frac{7p(1-p)}{18n}$$

$$\Rightarrow \frac{a^2+10}{b^2} < \frac{7}{18} \quad ①$$

$$\Rightarrow \frac{a^2+10}{(a+4)^2} < \frac{7}{18}$$ by subbing in (1)

$$\Rightarrow 11a^2 - 56a + 68 < 0 \quad ①$$ as $a \in \mathbb{N}$

$$\Rightarrow 2 < a < \frac{34}{11} \quad ① \Rightarrow a = 3 \Rightarrow b = 7 \quad ①$$

using a calculator

(Total for Question 8 is 12 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