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 24 June 2024

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 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

P72801A

F:1/1/1/1/1/1/

Pearson

1. Two students are experimenting with some water in a plastic bottle. The bottle is filled with water and a hole is put in the bottom of the bottle. The students record the time, t seconds, it takes for the water level to fall to each of 10 given values of the height, h cm, above the hole.

Student A models the data with an equation of the form $t = a + b\sqrt{h}$

The data is coded using $v = t - 40$ and $w = \sqrt{h}$ and the following information is obtained.

$$\sum v = 626 \quad \sum v^2 = 64678 \quad \sum w = 22.47 \quad S_{ww} = 4.52 \quad S_{vw} = -338.83$$

(a) Find the equation of the regression line of t on $\sqrt{h}$ in the form $t = a + b\sqrt{h}$

(4)

The time it takes the water level to fall to a height of 9 cm above the hole is 47 seconds.

(b) Calculate the residual for this data point.
Give your answer to 2 decimal places.

(2)

Given that the residual sum of squares (RSS) for the model of t on $\sqrt{h}$ is the same as the RSS for the model of v on w,

(c) calculate the RSS for these 10 data points.

(2)

Student B models the data with an equation of the form $t = c + dh$

The regression line of t on h is calculated and the residual sum of squares (RSS) is found to be 980 to 3 significant figures.

(d) With reference to part (c) state, giving a reason, whether Student B's model or Student A's model is the more suitable for these data.

(1)

a) From the Formula Booklet, we have that

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

$b = \dfrac{-338.83}{4.52} = -74.96$ 1

$a = \dfrac{626}{10} - \dfrac{224.7}{10} b$ 1 $\quad \bar{v} = \dfrac{\sum v}{n}$, $\bar{w} = \dfrac{\sum w}{n}$

$= 231.04$

P72801A0228

Question 1 continued

$v = t - 40$ ①

$\Rightarrow \quad t = 231.04 + 40 - 74.96\sqrt{h}$

$= 271.04 - 74.96\sqrt{h}$ ①

b) Residual $= t - (a + b\sqrt{h})$

Residual $= 47 - (271.04 - 74.96\sqrt{9})$ ①

$= 0.85$ ①

c) From the Formula Booklet, we have that

$S_{vv} = \sum v_i^2 - \frac{(\sum v_i)^2}{n}$

$S_{vv} = 64678 - \frac{626^2}{10} = 25440.4$ ①

Also, we have that

$RSS = S_{vv} - \frac{(S_{wv})^2}{S_{ww}}$

$RSS = 25440.4 - \frac{(-338.83)^2}{4.52} = 90.89$ ①

d) Student A's model is more suitable as the sum of the squares of the residuals is lower. ①

2. An estate agent asks customers to rank 7 features of a house, A, B, C, D, E, F and G, in order of importance. The responses for two randomly selected customers are in the table below.

Rank	1	2	3	4	5	6	7
Customer 1	A 1	E 6	C 3	F 7	G 2	B 4	D 5
Customer 2	E 5	F 7	C 3	G 6	A 1	D 2	B 4

(a) Calculate Spearman's rank correlation coefficient for these data.
(4)

(b) Stating your hypotheses and critical value clearly, test at the 5% level of significance, whether or not the two customers are generally in agreement.
(3)

a) Recall that Spearman's rank correlation coefficient is

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$d = (1-5, 6-7, 3-3, 7-6, 2-1, 4-2, 5-4)$ ①

$= (-4, -1, 0, 1, 1, 2, 1)$

$\sum d^2 = 16+1+0+1+1+4+1 = 24$ ①

$r_s = 1 - \frac{6(24)}{7(7^2-1)}$ ① $= 0.5714$ ①

b) $H_0: \rho_s = 0$

$H_1: \rho_s > 0$ ①

For $n = 7$, at the 5% level, compare with the table

$CV = 0.714$ ① $> 0.571 = r_s$

P 7 2 8 0 1 A 0 6 2 8

Question 2 continued

Hence, do not reject Ho, as there is insufficient evidence to conclude that the customers are in agreement. 1

(Total for Question 2 is 7 marks)

3. A factory produces bolts. The lengths of the bolts are normally distributed with mean μ mm and standard deviation 0.868 mm

A random sample of 15 of these bolts is taken and the mean length is 30.03 mm

(a) Calculate a 90% confidence interval for μ

(3)

A suitable test, at the 10% level of significance, is carried out using these 15 bolts, to see whether or not there is evidence that the variance of the length of the bolts has increased.

(b) Calculate the critical region for S^2

(3)

The manager of the factory decides that, in future, he will check each month whether the machine making the bolts is working properly. He uses a 10% level of significance to test whether or not there is evidence that

- the mean length of the bolts has changed
- the variance of the length of the bolts has increased

The next month a random sample of 15 bolts is taken.

The mean length of these bolts is 30.06 mm and the standard deviation is 1.02 mm

(c) With reference to your answers to part (a) and part (b), state whether or not there is any evidence that the machine is **not** working properly.
Give reasons for your answer.

(2)

a) From the tables or a calculator,

$z = 1.6449$ (1)

Recall that the formula for a confidence interval of a mean with known population standard deviation is

$\bar{x} \pm z\sqrt{\frac{\sigma^2}{n}}$

$30.03 \pm 1.6449 \times \frac{0.868}{\sqrt{15}}$ (1)

$(29.661, 30.399)$ (1)

P 7 2 8 0 1 A 0 8 2 8

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

b) Recall that the critical value for the upperbound of the variance is

$$\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha)} > \sigma^2$$

From the table $\chi^2_{14}(0.1) = 21.064$ ①

So we have $\frac{14s^2}{21.064} > 0.868^2$ ①

$\Rightarrow 14s^2 > 15.87$

$\Rightarrow s^2 > 1.133$ ①

c) There is insufficient evidence that the machine is not working properly as

- 30.06 is not in the confidence interval, ①
- $1.02^2 = 1.0404$ is not in the critical region. ①

(Total for Question 3 is 8 marks)

4. The random variable G has a continuous uniform distribution over the interval $[-3, 15]$

(a) Calculate $P(G > 12)$

(1)

The random variable H has a continuous uniform distribution over the interval $[2, w]$

The random variables G and H are independent and $E(H) = 10$

(b) Show that the probability that G and H are both greater than 12 is $\frac{1}{16}$

(3)

The random variable A is the area on a coordinate grid bounded by

$$y = -3$$

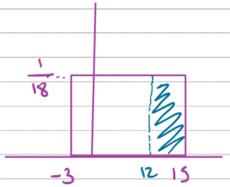
$$y = -4|x| + k$$

where k is a value from the continuous uniform distribution over the interval [5, 10]

(c) Calculate the expected value of A

(5)

a)



$Pr(G>12) = 3 \times \frac{1}{18} = \frac{1}{6}$ (1)

b) Recall that for the Uniform Distribution,

$E[X] = \frac{a+b}{2}$

$E[H] = \frac{2+w}{2} = 10 \Rightarrow w = 18$ (1)

Question 4 continued

as independent

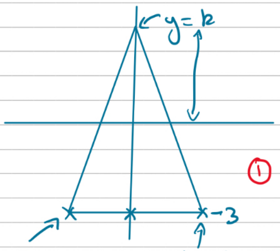
$$\Pr(G>12 \mid H>12) = \Pr(G>12)\Pr(H>12)$$

$$= \frac{1}{6} \times \frac{18-12}{18-2} \quad 1$$

$$= \frac{1}{16} \quad 1$$

c) We can plot a graph with

$y = -3$, $y = -4x + k$, $y = 4x + k$



1

$$\frac{-k-3}{4} = x_1 \qquad \frac{k+3}{4} = x_2 \quad 1$$

$$\text{Area of triangle} = \frac{1}{2}(y+3)(x_2 - x_1)$$

$$= \frac{1}{2}(k+3)\frac{(k+3)}{2}$$

$$= \frac{(k+3)^2}{4} \quad 1$$

P 7 2 8 0 1 A 0 1 1 2 8

Question 4 continued

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

$$f(k) = \frac{1}{10-5} = \frac{1}{5}$$

$$E[A] = \int_5^{10} \frac{1}{5} \times \frac{(k+3)^2}{4}\,dk \quad ①$$

$$= \left[\frac{(k+3)^3}{60}\right]_5^{10}$$

$$= \frac{13^3}{60} - \frac{8^3}{60} = \frac{337}{12} \quad ①$$

5. A continuous random variable X has probability density function

$$f(x) = \begin{cases} ax^{-2} - bx^{-3} & 2 \leqslant x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where a and b are constants.

Given that $P(X \leqslant 4) = \frac{3}{8}$

(a) use algebraic integration to show that $a = 3$
Show your working clearly.

(6)

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

(4)

a) $\int_2^{\infty} ax^{-2} - bx^{-3}\,dx = 1$ as the sum of probabilities is 1.

$\Rightarrow \left[\frac{-a}{x} + \frac{b}{2x^2} \right]_2^{\infty} = 1$ 1

$\Rightarrow -0 + 0 + \frac{a}{2} - \frac{b}{8} = 1$

$\Rightarrow 4a - b = 8$ 1

Also, $Pr(X \leqslant 4) = 3/8$

$\Rightarrow \int_2^4 ax^{-2} - bx^{-3}\,dx = \frac{3}{8}$

$\Rightarrow \left[-\frac{a}{x} + \frac{b}{2x^2} \right]_2^4 = \frac{3}{8}$ 1

$\Rightarrow -\frac{a}{4} + \frac{b}{32} + \frac{a}{2} - \frac{b}{8} = \frac{3}{8}$ 1

Question 5 continued

$\Rightarrow -8a + b + 16a - 4b = 12$

$\Rightarrow 8a - 3b = 12$ (1)

and $4a - b = 8$

$\Rightarrow b = 4a - 8$

$\Rightarrow 8a - 3(4a - 8) = 12$

$\Rightarrow 8a - 12a + 24 = 12$

$\Rightarrow -4a = -12$

$\Rightarrow a = 3$ (1)

b) $b = 4(3) - 8 = 4$ (1)

Median

$\int_2^m 3x^{-2} - 4x^{-3}\, dx = \frac{1}{2}$

$\Rightarrow \left[-\frac{3}{x} + \frac{2}{x^2} \right]_2^m = \frac{1}{2}$ (1)

$\Rightarrow -\frac{3}{m} + \frac{2}{m^2} + \frac{3}{2} - \frac{2}{4} = \frac{1}{2}$

$\Rightarrow m^2 - 6m + 4 = 0$ (1)

$\Rightarrow m = 3 \pm \sqrt{5}$ Using a calculator.

$\Rightarrow m = 3 + \sqrt{5}$ $(3 - \sqrt{5} < 2)$ (1)

6. A researcher set up a trial to assess the effect that a food supplement has on the increase in weight of Herdwick lambs. The researcher randomly selected 8 sets of twin lambs. One of each set of twins was given the food supplement and the other had no food supplement. The gain in weight, in kg, of each lamb over the period of the trial was recorded.

	Set of twin lambs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Weight gain (kg)	With food supplement	4.1	5.3	6.0	3.6	5.9	4.2	7.1	6.4
	No food supplement	5.0	4.8	5.2	3.4	5.1	3.9	7.0	6.5

(a) State why a two sample t-test is not suitable for use with these data.

(1)

(b) Suggest 2 other factors about the lambs that the researcher may need to control when selecting the sample.

(2)

(c) State one assumption, in context, that needs to be made for a paired t-test to be valid.

(1)

For a pair of twin lambs, the random variable W represents the weight gain of the lamb given the food supplement minus the weight gain of the lamb not given the food supplement.

(d) Using the data in the table, calculate a 98% confidence interval for the mean of W
Show your working clearly.

(5)

The researcher believes that the mean of W is greater than 200 g

(e) Stating your hypotheses clearly, use your confidence interval to explain whether or not there is evidence to support the researcher's belief.

(3)

a) The two samples are not independent. 1

b) They should consider the birth weight 1 and the gender of the lambs. 1

c) We need to assume that the distribution of the difference between the weight gains are normally distributed. 1

Question 6 continued

d) We do not know the population variance, so we use a t confidence interval.

$$\bar{x} \pm t\frac{s}{\sqrt{n}}$$

Differences $= (-0.4, 0.5, 0.8, 0.2, 0.8, 0.3, 0.1, -0.1)$ 1

Recall that $\bar{x} = \frac{\Sigma x}{n}$, $s^2 = \frac{1}{n-1}\Sigma x^2 - n\bar{x}^2$

$n = 8$, $\bar{x} = \frac{1.7}{8} = 0.2125$

$\Sigma x^2 = 2.49$, $\bar{x}^2 = 0.04516$

$\Rightarrow s^2 = \frac{1}{7}(2.49 - 8(0.04516)) = 0.3041$ 1

For t, we have $8-1=7$ degrees of freedom
So from the table, we have a critical value of

$t = 2.998$ 1

$0.2125 \pm 2.998\sqrt{\frac{0.3041}{8}}$ 1

So we have

$(-0.372, 0.797)$ 1

e) $H_0: \mu_w = 0.2$
$H_1: \mu_w > 0.2$ 1

$200g = 0.2kg$, which is in our confidence interval.
1

Turn over

P 7 2 8 0 1 A 0 1 9 2 8

Question 6 continued

Hence, we do not reject H_0, as there is insufficient evidence that μ_w is greater than 0.2

1

7. Two organisations are each asked to carry out a survey to find out the proportion, p, of the population that would vote for a particular political party.

The first organisation finds that out of m people, X would vote for this particular political party.

The second organisation finds that out of n people, Y would vote for this particular political party.

An unbiased estimator, Q, of p is proposed where

$$Q = k\left(\frac{X}{m} + \frac{Y}{n}\right)$$

(a) Show that $k = \frac{1}{2}$

(2)

A second unbiased estimator, R, of p is proposed where

$$R = \frac{aX}{m} + \frac{bY}{n}$$

(b) Show that $a + b = 1$

(2)

Given that $m = 100$ and $n = 200$ and that R is a better estimator of p than Q

(c) calculate the range of possible values of a
Show your working clearly.

(7)

a) Q is unbiased, so $E[Q] = p$

$E[Q] = E\left[k\left(\frac{X}{m} + \frac{Y}{n}\right)\right] = p$

$\Rightarrow \quad k\left[\frac{E[X]}{m} + \frac{E[Y]}{n}\right] = p$

As X and Y are binomially distributed,

$E[X] = mp$, $E[Y] = np$

$\Rightarrow \quad k\left[\frac{mp}{m} + \frac{np}{n}\right] = p$ 1

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

P 7 2 8 0 1 A 0 2 2 2 8

Question 7 continued

$\Rightarrow \quad 2kp = p$

$\Rightarrow \quad k = \frac{1}{2}$ (1)

b) $E[R] = \frac{aE[X]}{m} + \frac{bE[Y]}{n} = p$

$\Rightarrow \quad \frac{amp}{m} + \frac{bnp}{n} = p$ (1)

$\Rightarrow \quad ap + bp = p$

$\Rightarrow \quad a + b = 1$ (1)

c) As Q and R are unbiased, if R is a better estimator, then it has a lower variance.

$Var(Q) = Var\left(k\left(\frac{X}{m} + \frac{Y}{n}\right)\right)$

$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$ if X, Y independent

$= \frac{k^2}{m^2} Var(X) + \frac{k^2}{n^2} Var(Y)$

$A \sim Bin(n, p)$
$Var(A) = np(1-p)$

$= \frac{k^2}{m^2} mp(1-p) + \frac{k^2}{n^2} np(1-p)$

$= \frac{k^2 p(1-p)}{m} + \frac{k^2 p(1-p)}{n}$ (1)

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

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

Question 7 continued

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

$$\mathrm{Var}(R) < \mathrm{Var}(Q) \Rightarrow \frac{a^2}{m} + \frac{b^2}{n} < \frac{k^2}{m} + \frac{k^2}{n} \quad (1)$$

Plug in $k = 1/2$, $m = 100$, $n = 200$, $b = 1 - a$

$$\Rightarrow \frac{a^2}{100} + \frac{(1-a)^2}{200} < \frac{1}{4}\left(\frac{1}{100} + \frac{1}{200}\right) \quad (1)$$

$$\Rightarrow \frac{a^2}{100} + \frac{(1-a)^2}{200} < \frac{3}{800}$$

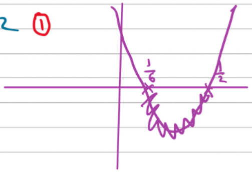
$$\Rightarrow 8a^2 + 4(1-a)^2 < 3$$

$$\Rightarrow 8a^2 + 4a^2 - 8a + 4 < 3$$

$$\Rightarrow 12a^2 - 8a + 1 < 0 \quad (1)$$

The critical values are $1/6$ and $1/2$ (1)

So $1/6 < a < 1/2$ (1)



8. A company packs chickpeas into small bags and large bags.

The weight of a small bag of chickpeas is normally distributed with mean 500 g and standard deviation 5 g

A random sample of 3 small bags of chickpeas is taken.

(a) Find the probability that the total weight of these 3 bags of chickpeas is between 1490 g and 1530 g

(3)

The weight of a large bag of chickpeas is normally distributed with mean 1020 g and standard deviation 20 g

One large bag and one small bag of chickpeas are chosen at random.

(b) Calculate the probability that the weight of the large bag of chickpeas is at least 30 g more than twice the weight of the small bag of chickpeas.
Show your working clearly.

(6)

a) Let $T = S_1 + S_2 + S_3$ $\quad E[S_i] = 500$

$\text{Var}(S_i) = 25$

$E[T] = E[S_1 + S_2 + S_3]$

$= E[S_1] + E[S_2] + E[S_3] = 1500$ 1

$\text{Var}(T) = \text{Var}(S_1 + S_2 + S_3)$

$= \text{Var}(S_1) + \text{Var}(S_2) + \text{Var}(S_3) = 75$ 1

So $T \sim N(1500, 75)$

$\Pr(1490 < T < 1530) = 0.8756$ 1 using a calculator.

b) $S \sim N(500, 25)$, $L \sim N(1020, 400)$

$\Pr(L \geqslant 2S + 30)$

$\Rightarrow \Pr(L - 2S \geqslant 30)$

Introduce a new variable $X = L - 2S$ 1

Question 8 continued

$E[X] = E[L] - 2E[S]$

$= 1020 - 2(500) = 20$ ①

$Var(X) = Var(L) + 4\,Var(S)$ ① $Var(aX+bY) = a^2 Var(X) + b^2 Var(Y)$

$= 400 + 4(25) = 500$ ①

$X \sim N(20, 500)$ ①

$Pr(X \geq 30) = 0.327$ using a calculator.
①