Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Tuesday 23 June 2020

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 algebraic 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 the 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

P62680A

1/1/1/

P 6 2 6 8 0 A 0 1 2 4

1 Gina receives a large number of packages from two companies, A and B. She believes that the variance of the weights of packages from company A is greater than the variance of the weights of packages from company B.

Gina takes a random sample of 7 packages from company A and an independent random sample of 10 packages from company B. Her results are summarised below

$$\bar{a} = 300 \quad S_{aa} = 145\,496 \quad \bar{b} = 233.4 \quad S_{bb} = 56\,364.4$$

[You may assume that the weights of packages from the two companies are normally distributed.]

Test Gina's belief. Use a 5% level of significance and state your hypotheses clearly.

(6)

$H_0 \quad \sigma_a^2 = \sigma_b^2$

$H_1 \quad \sigma_a^2 > \sigma_b^2$ ① ← variance of A is greater than variance of B

Degrees of freedom for A: $7-1=6$ and B: $10-1=9$

Critical value: $F_{6,9}$ (5% one-tail) $= 3.37$ ① ← from stats table

$$S_a^2 = \frac{S_{aa}}{n-1} = \frac{145496}{7-1} = 24\,249.33$$

$$S_b^2 = \frac{S_{bb}}{n-1} = \frac{56\,364.4}{10-1} = 6262.711 \quad ①$$

$$F_{test} = \frac{S_a^2}{S_b^2} = \frac{24\,249.33}{6262.711} \; ① = 3.872 \quad ①$$

F_{test} > critical value ∴ significant

There is evidence to support Gina's belief ①

Question 1 continued

(Total for Question 1 is 6 marks)

2 Jemima makes jam to sell in a local shop. The jam is sold in jars and the weight of jam in a jar is normally distributed.

Jemima takes a random sample of 8 of her jars of jam and weighs the contents of each jar, x grams. Her results are summarised as follows

$$\sum x = 3552 \qquad \sum x^2 = 1\,577\,314$$

(a) Calculate a 95% confidence interval for the mean weight of jam in a jar.

(5)

The labels on the jars state that the average contents weigh 440 grams.

(b) State, giving a reason, whether or not Jemima should be concerned about the labels on her jars of jam.

(1)

(a) The $100(1-\alpha)\%$ confidence interval is given by:

$$\left[\bar{x} - t_{n-1}\left(\frac{\alpha}{2}\right) \times \frac{s}{\sqrt{n}} \, , \, \bar{x} + t_{n-1}\left(\frac{\alpha}{2}\right) \times \frac{s}{\sqrt{n}}\right]$$

$$\bar{x} = \frac{\sum x}{n} = \frac{3552}{8} = 444 \quad (1)$$

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

$$s_x^2 = \frac{\sum x^2 - n(\sum x)^2}{n-1}$$

$$s_x^2 = \frac{1577314 - 8(444^2)}{8-1} = 32.2857\ldots \quad (1)$$

$t_7(5\%)$ 2-tailed critical value $= 2.365$ (1) ← from stats tables

lower limit: $444 - 2.365 \times \sqrt{\frac{32.2857..}{8}} = 439.25$ (1)

upper limit: $444 + 2.365 \times \sqrt{\frac{32.2857..}{8}} = 448.75$

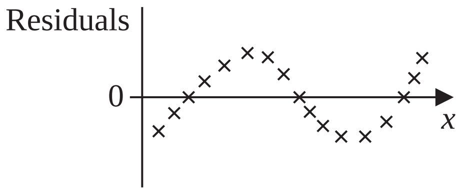
$\therefore$ confidence interval $= (439.25, 448.75)$ (1)

Question 2 continued

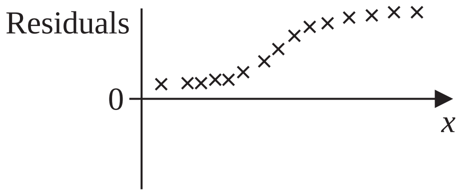
(b) 440 is in the confidence interval so the average contents label is acceptable. 1

(Total for Question 2 is 6 marks)

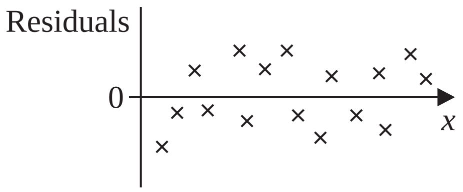
3 Below are 3 sketches from some students of the residuals from their linear regressions of y on x.



I



II



III

For each sketch you should state, giving your reason,

(i) whether or not the sketch is feasible

and if it is feasible

(ii) whether or not the sketch suggests a linear or a non-linear relationship between y and x.

(6)

I is feasible as a residual plot but probably a non-linear relationship 1 since the residuals are not randomly scattered about zero. 1

II is impossible as a residual plot 1 since the residuals do not sum to zero 1

III is feasible as a residual plot and is probably a linear relationship 1 since the points are randomly scattered about zero. 1

Question 3 continued

(Total for Question 3 is 6 marks)

4 A biased coin has a probability p of landing on heads, where $0 < p < 1$
Simon spins the coin n times and the random variable X represents the number of heads.
Taruni spins the coin m times, $m \neq n$, and the random variable Y represents the number of heads.

Simon and Taruni want to combine their results to find unbiased estimators of p.

Simon proposes the estimator $S = \dfrac{X+Y}{m+n}$ and Taruni proposes $T = \dfrac{1}{2}\left[\dfrac{X}{n} + \dfrac{Y}{m}\right]$

(a) Show that both S and T are unbiased estimators of p.

(3)

(b) Prove that, for all values of m and n, S is the better estimator.

(4)

(a) Exactly 2 outcomes with fixed probability so X and Y follow binomial distributions

$X \sim B(n, p) \rightarrow E(X) = np$

$Y \sim B(m, p) \rightarrow E(Y) = mp$ ①

$E(S) = \dfrac{E(X+Y)}{n+m}$

$E(X+Y) = E(X) + E(Y)$

$E(S) = \dfrac{np + mp}{n+m}$

$E(S) = \dfrac{(n+m)}{n+m} p$

$E(S) = p$ ∴ S is unbiased ①

$E(T) = \dfrac{1}{2}\left[\dfrac{E(X)}{n} + \dfrac{E(Y)}{m}\right]$

$E(T) = \dfrac{1}{2}\left[\dfrac{np}{n} + \dfrac{mp}{m}\right]$

DO NOT WRITE IN THIS AREA

Question 4 continued

$E(T) = \frac{1}{2}[2p]$

$E(T) = p \quad \therefore \quad T$ is unbiased. 1

(b) Better estimator has smaller variance.

$Var(\alpha X + \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y)$

$Var(S) = \frac{np(1-p) + mp(1-p)}{(n+m)^2}$ $\quad \frac{1}{n+m} E(X) + .$

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

$Var(S) = \frac{p(1-p)}{n+m}$ 1

$Var(T) = \frac{1}{2^2}\left[\frac{np(1-p)}{n^2} + \frac{mp(1-p)}{m^2}\right]$

$Var(T) = \frac{1}{4}\left[\frac{p(1-p)}{n} + \frac{p(1-p)}{m}\right]$

$Var(T) = \frac{1}{4}\left[\frac{mp(1-p) + np(1-p)}{nm}\right]$

$Var(T) = \frac{p(1-p)(n+m)}{4nm}$ 1

Question 4 continued

$$\text{Var}(S) < \text{Var}(T)$$

$$\frac{p(1-p)}{n+m} < \frac{p(1-p)(n+m)}{4nm}$$

$$p(1-p) \times 4nm < p(1-p)(n+m)^2$$

$$4nm < (n+m)^2$$

$$4nm < n^2 + 2mn + m^2$$

$$0 < n^2 - 2mn + m^2$$

$$0 < (m-n)^2 \quad 1$$

$\therefore$ S always has the smaller variance and is the better estimator. 1

Question 4 continued

(Total for Question 4 is 7 marks)

5

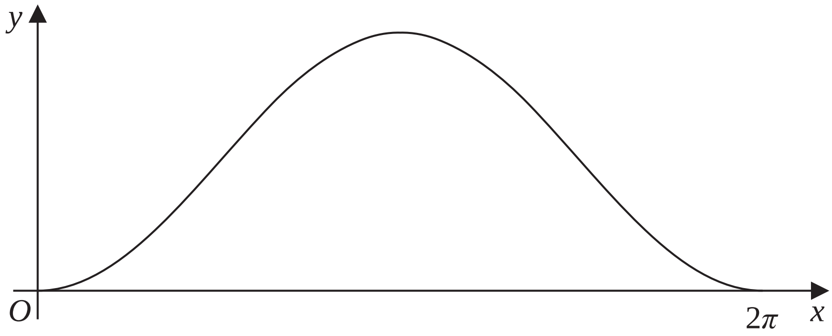


Figure 1

The random variable X has probability density function $\mathrm{f}(x)$ and Figure 1 shows a sketch of $\mathrm{f}(x)$ where

$$\mathrm{f}(x) = \begin{cases} k(1 - \cos x) & 0 \leqslant x \leqslant 2\pi \\ 0 & \text{otherwise} \end{cases}$$

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

(3)

The random variable $Y \sim \mathrm{N}(\mu, \sigma^2)$ and $\mathrm{E}(Y) = \mathrm{E}(X)$

The probability density function of Y is $\mathrm{g}(y)$, where

$$\mathrm{g}(y) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \qquad -\infty < y < \infty$$

Given that $\mathrm{g}(\mu) = \mathrm{f}(\mu)$

(b) find the exact value of σ

(3)

(c) Calculate the error in using $\mathrm{P}\left(\frac{\pi}{2} < Y < \frac{3\pi}{2}\right)$ as an approximation to $\mathrm{P}\left(\frac{\pi}{2} < X < \frac{3\pi}{2}\right)$

(4)

(a) $\int f(x)\,dx = 1$ because probabilities always sum to 1.

$k\int_0^{2\pi} (1 - \cos x)\,dx = k\left[x - \sin x\right]_0^{2\pi}$ ①

$1 = k[2\pi - 0] - [0 - 0]$ ①

$\frac{1}{2\pi} = k$ ①

P 6 2 6 8 0 A 0 1 2 2 4

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

(b) E(X) is the maximum of f(x) Since f(x) is symmetrical about π.

$E(X) = \mu = \pi$

$f(\mu) = \frac{1}{2\pi}(1-\cos\pi) = \frac{1}{\pi}$ ①

$g(\mu) = f(\mu)$

$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu-\mu}{\sigma}\right)^2} = \frac{1}{\pi}$ ① ← exponent disappears
exponent = 0
$e^0 = 1$

$\frac{1}{\sigma\sqrt{2\pi}} = \frac{1}{\pi}$

$\pi = \sigma\sqrt{2\pi}$

$\sigma = \frac{\pi}{\sqrt{2\pi}} \times \frac{\sqrt{2\pi}}{\sqrt{2\pi}}$

$\sigma = \sqrt{\frac{\pi}{2}}$ ①

(c) $P\left(\frac{\pi}{2} < Y < \frac{3\pi}{2}\right) = P\left(Y < \frac{3\pi}{2}\right) - P\left(Y < \frac{\pi}{2}\right)$

$= 0.7899$ ① ← use calculator or stats tables

$P\left(\frac{\pi}{2} < X < \frac{3\pi}{2}\right) = \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} f(x)\,dx$

Question 5 continued

$$\frac{1}{2\pi}\left[x - \sin x\right]_{\frac{\pi}{2}}^{\frac{3\pi}{2}} = \frac{1}{2\pi}\left[\frac{3\pi}{2} - \sin\frac{3\pi}{2} - \frac{\pi}{2} + \sin\frac{\pi}{2}\right] \text{①}$$

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

$$= \frac{2+\pi}{2\pi} \text{①}$$

$$= 0\ 81830$$

Error is $0\ 81830 - 0\ 7899 = 0.0284$ ①

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 10 marks)

6 A new employee, Kim, joins an existing employee, Jiang, to work in the quality control department of a company producing steel rods.
Each day a random sample of rods is taken, their lengths measured and a 95% confidence interval for the mean length of the rods, in metres, is calculated. It is assumed that the lengths of the rods produced are normally distributed.

Kim took a random sample of 25 rods and used the *t* distribution to obtain a 95% confidence interval of (1.193, 1.367) for the mean length of the rods.
Jiang commented that this interval was a little wider than usual and explained that they usually assume that the standard deviation does not change and can be taken as 0.175 metres.

(a) Test, at the 10% level of significance, whether or not Kim's sample suggests that the standard deviation is different from 0.175 metres. State your hypotheses clearly.

(9)

Using Kim's sample and the normal distribution with a standard deviation of 0.175 metres,

(b) find a 95% confidence interval for the mean length of the rods.

(3)

(a) From confidence interval : $\bar{x} = \frac{1.193 + 1.367}{2} = 1.28$ ①

CI is given by $\bar{x} \pm t_{n-1}\left(\frac{\alpha}{2}\right) \times \frac{s}{\sqrt{n}}$

$n = 25$, $100(1-\alpha) = 95 \Rightarrow \alpha = \frac{1}{20}$

$t_{25-1}\left(\frac{1/20}{2}\right) = 2.064$ ← from stats tables

$1.367 = 1.28 - 2.064 \times \frac{s}{\sqrt{25}}$ ①

$s = 5 \times \frac{1.367 - 1.28}{2.064}$

$s = 0.2108$ ①

$H_0 : \sigma = 0.175$ $\quad H_1 : \sigma \neq 0.175$ ①

Significance level (2-tailed) $= 10\% \div 2 = 5\% = 0.05$

Degrees of freedom $= n - 1 = 25 - 1 = 24$

Question 6 continued

$$\chi^2_{24} = \frac{(n-1)s^2}{\sigma^2} = \frac{24 \times (0\ 2108)^2}{(0\ 175)^2} \ (1) = 34\ 81 \ (1)$$

$\uparrow$ test stat

$\chi^2_{24}(0.05)$ critical region is :

$\chi^2_{24} < 13.848 \qquad \chi^2_{24} > 36\ 415$ (1)

34.81 is not in the critical region so is not significant

$\therefore$ Insufficient evidence that $\sigma \neq 0\ 175$ (1)

(b) $CI = \bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$

For 95% CI, $z = 1\ 96$ (1) $\leftarrow$ from stats tables

$CI = 1\ 28 \pm 1\ 96 \times \frac{0.175}{\sqrt{25}}$ (1)

$CI = (1\ 21, 1.35)$ (1)

(Total for Question 6 is 12 marks)

7 Fence panels come in two sizes, large and small. The lengths of the large panels are normally distributed with mean 198 cm and standard deviation 5 cm. The lengths of the small panels are normally distributed with mean 74 cm and standard deviation 3 cm.

(a) Find the probability that the total length of a random sample of 3 large panels is greater than the total length of a random sample of 8 small panels.

(6)

One large panel and one small panel are selected at random.

(b) Find the probability that the length of the large panel is more than $\frac{8}{3}$ times the length of the small panel.

(5)

Rosa needs 1000 cm of fencing. The large panels cost £80 each and the small panels cost £30 each. Rosa's plan is to buy 5 large panels and measure the total length. If the total length is less than 1000 cm she will then buy one small panel as well.

(c) Calculate whether or not the expected cost of Rosa's plan is cheaper than simply buying 14 small panels.

(6)

(a) Let L = large panels, S = small panels

$X = L_1 + L_2 + L_3$ ①

$X \sim N(n\mu_1, n\sigma_1^2)$

formula for combining variables with identical normal distribution

$X \sim N(3 \times 198, 3 \times 5^2)$

$X \sim N(594, 75)$ ①

$Y = S_1 + .. + S_8$

$Y \sim N(8 \times 74, 8 \times 3^2)$

$Y \sim N(592, 72)$ ①

74 - 72

$3 \times 5^2 + 8 \times (-3)^2$

$P(X > Y) = P(D > 0)$ ① where $D \sim N(2, 147)$ ①

difference

$= 0.566$ ①

stats tables or calculator.

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

Question 7 continued

(b) $W = L - \frac{8}{3}S$ ① ← since L and S are normally distributed, W is also normally distributed

$\mu_W = \mu_L - \frac{8}{3}\mu_S$

$\mu_W = 198 - \frac{8}{3} \times 74 = \frac{2}{3}$ ①

$Var_W = Var_L - \frac{8}{3} \times Var_S$

$Var_W = 5^2 - \left(\frac{8}{3}\right)^2 \times 3^2 = 89$

$\therefore W \sim N\left(\frac{2}{3}, 89\right)$ ①

$P(W > 0)$ ① $= 0\ 528$ ①

(c) Let F be the length of the fence

$F = L_1 + .. + L_5$ ①

$F \sim N(5 \times 198, 5 \times 5^2)$

$F \sim N(990, 125)$ ①

$P(F < 1000) = 0\ 814$ ①

$E(\text{cost of plan}) = 0\ 814(5 \times 80 + 30) + (1 - 0\ 814)(5 \times 80)$ ①

$= £424 \cdot 43$ ①

Buying 14 small panels $= 14 \times 30 = £420$

$\therefore$ Rosa's plan is likely to be more expensive ①

Question 7 continued

Question 7 continued

(Total for Question 7 is 17 marks)

8 A circle, centre O, has radius x cm, where x is an observation from the random variable X which has a rectangular distribution on $[0, \pi]$

(a) Find the probability that the area of the circle is greater than $10\,\text{cm}^2$

(3)

(b) State, giving a reason, whether the median area of the circle is greater or less than $10\ \text{cm}^2$

(1)

The triangle OAB is drawn inside the circle with OA and OB as radii of length x cm and angle AOB x radians.

(c) Use algebraic integration to find the expected value of the area of triangle OAB. Give your answer as an exact value.

(7)

(a) rectangular = continuous uniform

$X \sim U[0, \pi]$

area $= \pi X^2$

$P(\pi X^2 > 10) = P\left(X^2 > \frac{10}{\pi}\right)$

$= P\left(X > \sqrt{\frac{10}{\pi}}\right)$ 1

$= 1 - P\left(X \leq \sqrt{\frac{10}{\pi}}\right)$

$F(x) = \int f(x)\,dx = P(X \leq x)$

$f(x) = \frac{1}{b-a} = \frac{1}{\pi}$

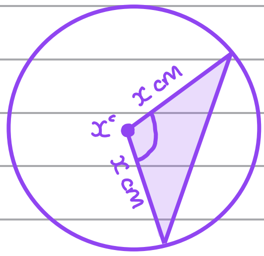
$F(x) = \int \frac{1}{\pi}\,dx = \frac{x}{\pi}$

$1 - P\left(X \leq \sqrt{\frac{10}{\pi}}\right) = 1 - \frac{\sqrt{\frac{10}{\pi}}}{\pi}$ 1 $= 0.432$ 1

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

Question 8 continued

(b) P(area > median) = 0 5 because (a) < 0 5
so the median < 10 ①

(c) 

Area of triangle $= \frac{1}{2}ab\sin c$

$= \frac{1}{2} \times x^2 \times \sin x$

$= 0.5x^2 \sin x$ ①

$E(\text{area}) = E(f(x) \times \text{area})$

$= E\left(\frac{1}{\pi} \times 0\ 5x^2 \sin x\right)$

$= \int_0^{\pi} \frac{1}{\pi} 0\ 5x^2 \sin x \; dx$

$= \frac{1}{2\pi} \int_0^{\pi} x^2 \sin x \; dx$ ①

Use integration by parts: $\int uv' = uv - \int vu'$

$u = x^2 \rightarrow u' = 2x$

$v' = \sin x \rightarrow v = -\cos x$

① $uv - \int vu' = -x^2 \cos x + \int 2x \cos x \; dx$ ①

Repeat for new integral

$u = 2x \rightarrow u' = 2$

$v' = \cos x \rightarrow v = \sin x$

Question 8 continued

② $uv - \int vu' = 2x\sin x - \int 2\sin x \, dx$

Sub ② into ① . $-x^2\cos x + 2x\sin x - \int 2\sin x \, dx$

$E(\text{area}) = \frac{1}{2\pi}\left[-x^2\cos x + 2x\sin x + 2\cos x\right]_0^\pi$ ① ①

$= \frac{1}{2\pi}\left[-\pi^2\cos\pi + 2\pi\sin\pi + 2\cos\pi - 2\right]$

$= \frac{1}{2\pi}\left[\pi^2 - 2 - 2\right]$ ①

$= \frac{\pi^2}{2\pi} - \frac{4}{2\pi}$

$= \frac{\pi}{2} - \frac{2}{\pi}$ ①

(Total for Question 8 is 11 marks)

TOTAL FOR PAPER IS 75 MARKS