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 25 June 2019

Morning (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.
- 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

P61184A

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

Pearson

Answer ALL questions. Write your answers in the spaces provided.

1 A machine is set to fill pots with yoghurt such that the mean weight of yoghurt in a pot is 505 grams.

To check that the machine is working properly, a random sample of 8 pots is selected. The weight of yoghurt, in grams, in each pot is as follows

508 510 500 500 498 503 508 505

Given that the weights of the yoghurt delivered by the machine follow a normal distribution with standard deviation 5.4 grams,

(a) find a 95% confidence interval for the mean weight, μ grams, of yoghurt in a pot. Give your answers to 2 decimal places.

(4)

(b) Comment on whether or not the machine is working properly, giving a reason for your answer.

(1)

(c) State the probability that a 95% confidence interval for μ will not contain μ grams.

(1)

(d) Without carrying out any further calculations, explain the changes, if any, that would need to be made in calculating the confidence interval in part (a) if the standard deviation was unknown. Give a reason for your answer.
You may assume that the weights of the yoghurt delivered by the machine still follow a normal distribution.

(2)

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

for 95% / 0 95, $z = 1.96$ ① ← from stats table for standard normal distribution

$\bar{x} = [508+510+500+500+498+503+508+505] \div 8$

$\bar{x} = 504$ ①

$CI = 504 \pm 1\ 96 \times \frac{5.4}{\sqrt{8}}$ ① → (500.258, 507 742) ①

(b) 505 is in the confidence interval therefore there is evidence that the machine is working properly ①

Question 1 continued

(c) 5% 1 ← 100% − 95%

(d) s would need to be used instead of σ

A t-value would need to be used instead of the z-value 1 for both points

Because the sample is small you cannot use the normal distribution. 1

(Total for Question 1 is 8 marks)

2 A large field of wheat is split into 8 plots of equal area. Each plot is treated with a different amount of fertiliser, f grams/m^2. The yield of wheat, w tonnes, from each plot is recorded. The results are summarised below.

$$\sum f = 28 \qquad \sum w = 303 \qquad \sum w^2 = 13\,447 \qquad S_{ff} = 42 \qquad S_{fw} = 269.5$$

(a) Calculate the product moment correlation coefficient between f and w

(2)

(b) Interpret the value of your product moment correlation coefficient.

(1)

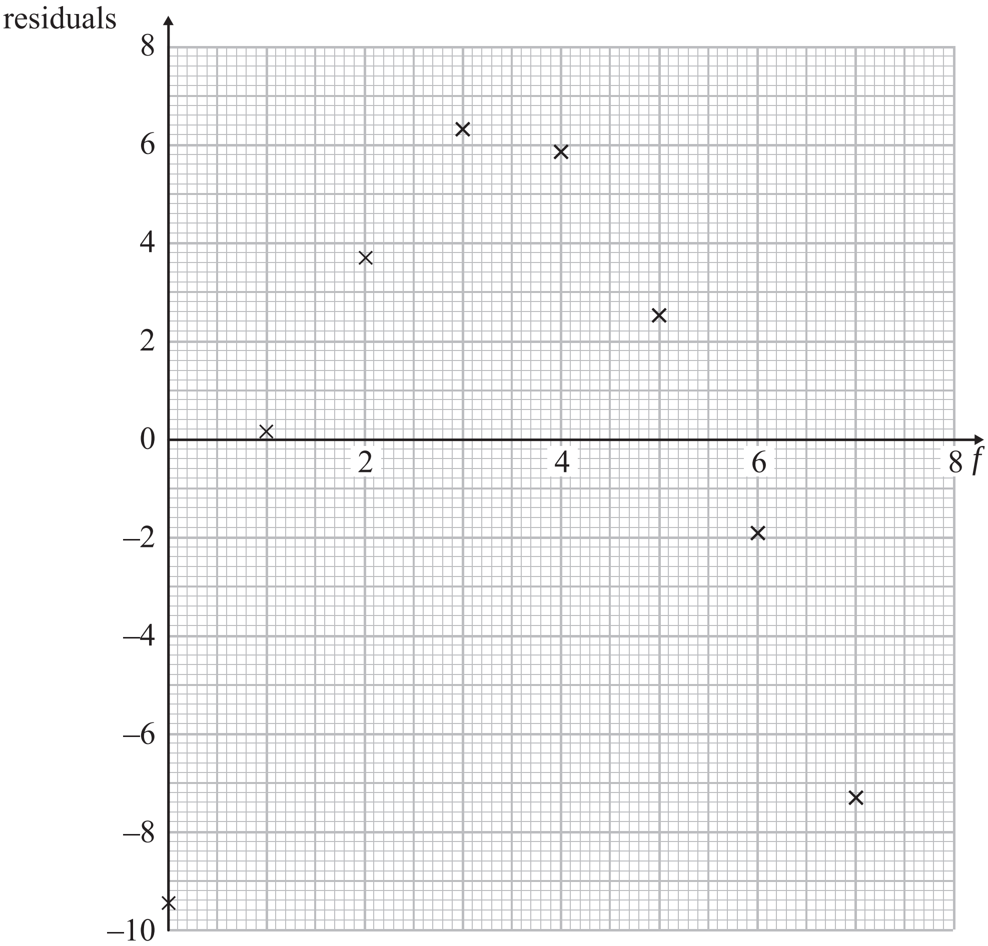
(c) Find the equation of the regression line of w on f in the form $w = a + bf$

(3)

(d) Using your equation, estimate the decrease in yield when the amount of fertiliser decreases by 0.5 grams/m^2

(1)

The residuals of the data recorded are calculated and plotted on the graph below.



(e) With reference to this graph, comment on the suitability of the model you found in part (c).

(2)

(f) Suggest how you might be able to refine your model.

(1)

Question 2 continued

(a) $PMCC = \dfrac{S_{fw}}{\sqrt{S_{ff} \times S_{ww}}}$ where $S_{xy} = \Sigma xy - \dfrac{\Sigma x \Sigma y}{n}$

$S_{ww} = \Sigma w^2 - \dfrac{(\Sigma w)^2}{8}$

$S_{ww} = 13447 - \dfrac{303^2}{8} = 1970\ 875$ 1

$PMCC = \dfrac{269\ 5}{\sqrt{42 \times 1970 \cdot 875}}$ 1 $= 0\ 9367$ 1

(b) As the amount of fertiliser increases, the yield increases 1

since PMCC is close to 1 there is a strong positive correlation.

(c) $w = a + bf$ so b is gradient, a is intercept

$b = \dfrac{S_{fw}}{S_{ff}} = \dfrac{269\ 5}{42} = 6\ 42$ 1

$a = \bar{w} - b \times \bar{f} = \dfrac{\overset{\Sigma w}{303}}{8} - 6\ 42 \times \dfrac{\overset{\Sigma f}{28}}{8} = 15\ 42$ 1

$\therefore\ w = 15\ 42 + 6\ 42f$ 1

(d) $w = [15\ 42 + 6\ 42(28)] - [15\ 42 + 6\ 42(28 - 0\ 5)]$

$w = 3\ 21$ tonnes 1

Question 2 continued

(e) Residuals appear to not be randomly scattered ①

So the model in (c) is unlikely to be suitable ①

(f) Fit a curve rather than a line ①

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 10 marks)

3 Yin grows two varieties of potato, plant *A* and plant *B*. A random sample of each variety of potato is taken and the yield, x kg, produced by each plant is measured. The following statistics are obtained from the data.

	Number of plants	$\sum x$	$\sum x^2$
A	25	194.7	1637.37
B	26	227.5	2031.19

(a) Stating your hypotheses clearly, test, at the 10% significance level, whether or not the variances of the yields of the two varieties of potato are the same.

(7)

(b) State an assumption you have made in order to carry out the test in part (a).

(1)

(a) Use F-test to check if the samples are from distributions with the same variance.

$H_0 \quad \sigma_A^2 = \sigma_B^2$ ← same variance

$H_1 \quad \sigma_A^2 \neq \sigma_B^2$ ① ← different variance

significance level = 10% = 0.1

$F_{n_x-1,\, n_y-1} \sim \dfrac{S_A^2}{S_B^2}$

from stats. tables at 0.1 sig level

critical value at $F_{n_A-1,\, n_B-1} = F_{24,25}(0.1)$ is 1.96 ①

$$S_A^2 = \frac{\Sigma x_A^2 - \frac{(\Sigma x_A)^2}{n_A}}{n_A - 1} = \frac{1637.37 - \frac{(194.7)^2}{25}}{25-1} = 5.0436$$ ① ①

$$S_B^2 = \frac{\Sigma x_B^2 - \frac{(\Sigma x_B)^2}{n_B}}{n_B - 1} = \frac{2031.19 - \frac{(227.5)^2}{26}}{26-1} = 1.6226$$ ①

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

Question 3 continued

$F_{test} = \frac{S_A^2}{S_B^2} = \frac{5.0436}{1.6226} = 3.108$ ①

F_{test} > critical value ∴ Reject H_0.

There is evidence to suggest that the variances are different ①

(b) The yields are normally distributed ①

(Total for Question 3 is 8 marks)

4 The continuous random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x \leqslant 0 \\ k\left(x^3 - \frac{3}{8}x^4\right) & 0 < x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

where k is a constant.

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

(1)

(b) Showing your working clearly, use calculus to find

(i) E(X)

(ii) the mode of X

(6)

(c) Describe, giving a reason, the skewness of the distribution of X

(1)

(a) $k\left(2^3 - \frac{3}{8} \times 2^4\right) = 1$ 1 ← at upper bound $(x = 2)$ the c.d.f must be equal to 1, since it is a sum of all probabilities

$k(2) = 1$

$k = \frac{1}{2}$ 1

(b) $f(x) = \frac{d}{dx}\left[F(x)\right]$

convert to p.d.f

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

(i) $E(X) = \int_0^2 x f(x)\, dx$

$E(X) = \int_0^2 \frac{1}{2}x\left[3x^2 - \frac{3}{2}x^3\right] dx$

DO NOT WRITE IN THIS AREA

Question 4 continued

$$E(X) = \frac{1}{2}\int_0^2 \left[3x^3 - \frac{3}{2}x^4\right] dx \quad ①$$

$$E(X) = \frac{1}{2}\left[\frac{3}{4}x^4 - \frac{3}{10}x^5\right]_0^2$$

$$E(X) = \frac{1}{2}\left[\frac{3}{4}(2)^4 - \frac{3}{10}(2)\right]$$

$$E(X) = \frac{6}{5} \quad ①$$

(ii) mode = x value at max. of the pdf

$$f(x) = \frac{3}{2}x^2 - \frac{3}{4}x^3$$

$$f'(x) = 3x - \frac{9}{4}x^2 = 0 \quad ①$$

$$x\left(3 - \frac{9}{4}x\right) = 0 \quad ①$$

$x = 0$ and $3 - \frac{9}{4}x = 0$

this is clearly a min. since $f(0) = 0$.

$$3 = \frac{9}{4}x$$

$$\frac{4}{3} = x$$

$$\therefore \text{mode} = \frac{4}{3} \quad ①$$

(c) The mode is greater than the mean, which implies it is negative skew ①

Question 4 continued

Question 4 continued

(Total for Question 4 is 8 marks)

5 Alexa believes that students are equally likely to achieve the same percentage score on each of two tests, paper I and paper II. She randomly selects 8 students and gives them each paper I and paper II. The percentage scores for each paper are recorded.

The following paired data are collected.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Paper I (%)	70	70	84	80	64	65	65	90
Paper II (%)	64	76	72	74	68	64	58	76

Test, at the 1% significance level, whether or not there is evidence to support Alexa's belief. State your hypotheses clearly and show your working.

(7)

Let d be the difference between scores in Paper I and Paper II

$H_0 : \mu_d = 0$ ← there is no difference in means

$H_1 : \mu_d \neq 0$ 1 ← there is a difference in means

Student	A	B	C	D	E	F	G	H
d	6	-6	12	6	-4	1	7	14

1 ← Paper I - Paper II

$$\bar{d} = \frac{\Sigma d}{n}$$

$$\bar{d} = \frac{6-6+12+6-4+1+7+14}{8}$$

$$\bar{d} = \frac{36}{8}$$

$$\bar{d} = 4\ 5$$

$$S_d = \sqrt{\frac{\Sigma d^2 - n\bar{d}}{n-1}}$$

$$S_d = \sqrt{\frac{[6^2+(-6)^2+12^2+6^2+(-4)^2+1^2+7^2+14^2]-8(4\ 5)}{8-1}}$$

DO NOT WRITE IN THIS AREA

Question 5 continued

$S_d = \sqrt{50.285}$

$S_d = 7.09$ ①

$t = \dfrac{\bar{d} - \mu_d}{\frac{S_d}{\sqrt{n}}}$

$t = \dfrac{4.5 - 0}{\frac{7.09}{\sqrt{8}}}$ ①

$t = \pm 1.7948$ ①

Degrees of freedom $= n - 1 = 8 - 1 = 7$

Critical value for $t_7 = \pm 3.499$ ①

$1.7948 < 3.499$ $\therefore$ Accept H_0

There is insufficient evidence that Alexa's belief is correct ①

(Total for Question 5 is 7 marks)

6 A company manufactures bolts. The diameter of the bolts follows a normal distribution with a mean diameter of 5 mm.

Stan believes that the mean diameter of the bolts is less than 5 mm. He takes a random sample of 10 bolts and measures their diameters. He calculates some statistics but spills ink on his work before completing them. The only information he has left is as follows

4.5 4.5 5.5 4.8 4.9 4.7 5
$X \sim$ N(5
$\sum x = 48.4$
$\bar{x} =$
99% confidence interval for the variance is = (0.01712, 0.23280)

Stating your hypotheses clearly, test, at the 5% level of significance, whether or not Stan's belief is supported.

(9)

$H_0 \quad \mu = 5$

$H_1 : \mu < 5$ 1

$\bar{x} = \frac{\Sigma x}{n} = \frac{48.4}{10} = 4.84$ 1

Using a χ^2 distribution, confidence interval for variance uses χ^2 values of 1.735, 23.589. 1

$\frac{9s^2}{1.735} = 0.2328$ 1 $\leftarrow$ $\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \leq \sigma^2$

$s^2 = \frac{0.2328 \times 1.735}{9}$

$s^2 = 0.0449$ 1

Degrees of freedom $= n - 1 = 10 - 1 = 9$

Critical Value : $t_9 = -1.833$ 1 $\leftarrow$ from stats tables

DO NOT WRITE IN THIS AREA

Question 6 continued

test statistic $= \pm \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

test statistic $= \pm \dfrac{4.84 - 5}{\frac{\sqrt{0.0449}}{\sqrt{10}}}$ (1)

test statistic $= -2.39$ (1)

$-2.39 < -1.833$ $\therefore$ Reject H_0.

There is sufficient evidence to support Stan's belief (1)

(Total for Question 6 is 9 marks)

7 A manufacturer makes two versions of a toy. One version is made out of wood and the other is made out of plastic.

The weights, W kg, of the wooden toys are normally distributed with mean 2.5 kg and standard deviation 0.7 kg. The weights, X kg, of the plastic toys are normally distributed with mean 1.27 kg and standard deviation 0.4 kg. The random variables W and X are independent.

(a) Find the probability that the weight of a randomly chosen wooden toy is more than double the weight of a randomly chosen plastic toy.

(6)

The manufacturer packs n of these wooden toys and $2n$ of these plastic toys into the same container. The maximum weight the container can hold is 252 kg.

The probability of the contents of this container being overweight is 0.2119 to 4 decimal places.

(b) Calculate the value of n.

(8)

(a) $W \sim N(2.5, 0.7^2)$ and $X \sim N(1.27, 0.4^2)$

$T = W - 2X$ ← T is the 'more than' amount.

$E(T) = 2.5 - 2(1.27)$ ① ← $E(\alpha X - \beta Y) = \alpha E(X) - \beta E(Y)$

$= -0.04$ ①

$Var(T) = 0.7^2 + 2^2 \times 0.4^2$ ①

$= 1.13$ ① $Var(\alpha X - \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y)$

$T \sim N(-0.04, 1.13)$

$P(T > 0) = P\left(Z > \frac{0-(-0.04)}{\sqrt{1.13}}\right)$

$= P(Z > 0.0376)$ ①

$= 0.0484$ ①

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

Question 7 continued

(b) $B = nW + 2nX$ ①

$E(B) = n \times 2.5 + 2n \times 1.17 = 5.04n$ ①

$Var(B) = n \times 0.7^2 + 2n \times 0.4^2 = 0.81n$ ①

$P(B > 252) = 1 - P(B < 252)$

$0.2119 = 1 - P(B < 252)$

$P(B < 252) = 0.7881$

$\therefore z = 0.8$ ← from standard normal stats table

$\dfrac{252 - 5.04n}{\sqrt{0.81n}}$ ① $= 0.8$ ①

$252 - 5.04n = 0.8 \times 0.9\sqrt{n}$

$5.04n + 0.72\sqrt{n} - 252 = 0$

$\sqrt{n} = 7$ and -7.14 ①

↑ ignore - square root of an integer cannot be negative

$n = 7^2$ ①

$n = 49$ ①

Question 7 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 7 continued

(Total for Question 7 is 14 marks)

8 Nine athletes, A, B, C, D, E, F, G, H and I, competed in both the 100 m sprint and the long jump. After the two events the positions of each athlete were recorded and Spearman's rank correlation coefficient was calculated and found to be 0.85

(a) Stating your hypotheses clearly, test whether or not there is evidence to suggest that the higher an athlete's position is in the 100 m sprint, the higher their position is in the long jump. Use a 5% level of significance.

(4)

The piece of paper the positions were recorded on was mislaid. Although some of the athletes agreed their positions, there was some disagreement between athletes B, C and D over their long jump results.

The table shows the results that are agreed to be correct.

Athlete	A	B	C	D	E	F	G	H	I
Position in 100 m sprint	4	6	7	9	2	8	3	1	5
Position in long jump	5				4	9	3	1	2

Given that there were no tied ranks,

(b) find the correct positions of athletes B, C and D in the long jump. You must show your working clearly and give reasons for your answers.

(5)

(c) Without recalculating the coefficient, explain how Spearman's rank correlation coefficient would change if athlete H was disqualified from both the 100 m sprint and the long jump.

(2)

(a) $H_0: \rho_s = 0$ $H_1: \rho_s > 0$ ①

At 5% = 0.05 significance level, critical value = 0.6 ①

from stats tables using sample level = 9

$r_s = 0.85$ and $0.85 > 0.6$, so in the critical region. ①

There is evidence to suggest that there is a relationship between the position in the 100m sprint and the position in the long jump ①

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

Question 8 continued

(b) Let d_n be the difference between scores for each athlete.

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

$0.85 = 1 - \frac{6\Sigma d^2}{9(9^2-1)}$ (1)

$1 - 0.85 = \frac{6\Sigma d^2}{720}$

$\Sigma d^2 = 18$ (1)

Σd^2 in table $= (4-5)^2 + (2-4)^2 + (8-9)^2 + (5-2)^2$

$= 15$

$18 - 15 = 3$ so $\Sigma d^2 = 3$ for the missing values (1)

So B, C, D must all have $d = \pm 1$

B must be position $6+1=7$ since 5 is taken by A (1)

By the same logic, C is in $7-1=6$ and D is in $9-1=8$

$\therefore$ $B \to 7$
$C \to 6$
$D \to 8$ (1)

Question 8 continued

(c) The Σd^2 will not change but the value of n will decrease ① so Spearman's Rank Correlation will decrease ①

(Total for Question 8 is 11 marks)

TOTAL FOR PAPER IS 75 MARKS