Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | | | | | | Candidate Number | | | | |

Pearson Edexcel Level 3 GCE

Paper reference **8FM0/23**

Further Mathematics

Advanced Subsidiary
Further Mathematics options
23: Further Statistics 1
(Part of options B, E, F and G)

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 the 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.
- The total mark for this part of the examination is 40. There are 4 questions.
- 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

P 7 1 9 6 7 A 0 1 1 6
Pearson

1. Stuart is investigating a treatment for a disease that affects fruit trees. He has 400 fruit trees and applies the treatment to a random sample of these trees. The remainder of the trees have no treatment. He records the number of years, y, that each fruit tree remains free from this disease.

The results are summarised in the table below.

		Treatment	
		Applied	**Not applied**
Number of years free from this disease	$y < 1$	15	25
	$1 \leqslant y < 2$	35	61
	$2 \leqslant y$	124	140

The data are to be used to determine whether or not there is an association between the application of the treatment and the number of years that a fruit tree remains free from this disease.

(a) Calculate the expected frequencies for

(i) Applied and $y < 1$

(ii) Not applied and $1 \leqslant y < 2$

(2)

The value of $\sum \frac{(O - E)^2}{E}$ for the **other four** classes is 2.642 to 3 decimal places.

(b) Test, at the 5% level of significance, whether or not there is an association between the application of the treatment and the number of years a fruit tree remains free from this disease.

You should state your hypotheses, test statistic, critical value and conclusion clearly.

(5)

Question 1 continued

(a)

years w/o disease	Applied	Not applied	total
$y < 1$	(i) 15	25	40
$1 \leqslant y < 2$	35	61 (ii)	96
$2 \leqslant y$	124	140	264
total	174	226	400 grand total

For contingency tables we calculate the expected frequency like this:

			Total
	a	b	a+b (row total)
	c	d	c+d
Total	a+c (column total)	b+d	a+b+c+d (grand total)

to get the E_i of this, we calculate

$$E_i = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$$

∴ here it would be $E_i = \dfrac{(a+b)(a+c)}{a+b+c+d}$

(i) $E_i = \dfrac{40 \times 174}{400} = 17.4$ to 3sf

(ii) $E_i = \dfrac{96 \times 226}{400} = 54.2$ to 3sf M1A1

(b) Hypotheses

H_0: There is no association between application of treatment and # of years the fruit tree remains disease-free

H_1: There is an association between application of treatment and # of years the fruit tree remains disease-free

B1

We need $\sum \frac{(O-E)^2}{E_i}$ for $y<1$ Applied and $1 \leqslant y < 2$ Not applied:

$$\sum \frac{(O-E)^2}{E} = \frac{(15-17.4)^2}{17.4} + \frac{(61-54.24)^2}{54.24} + 2.642$$

$y<1$ applied, $1 \leqslant y < 2$ Not-applied, given

★ we calculated E values in (a)!

M1

$= 3.815 \rightarrow 3.82$ to 3sf A1

★ For contingency tables the formula for DoF is:

DoF = (# of rows − 1)(# of columns − 1)

DoF = (2 − 1)(3 − 1) = 2 DoF

$\chi^2_2(0.05) = 5.991$ from tables B1

5.991 > 3.82 ∴ our test statistic is smaller than our critical value ∴ no significant evidence of association between application of treatment and the # of years the tree stays disease-free. Reject H_0.

A1

Question 1 continued

Question 1 continued

(Total for Question 1 is 7 marks)

2. Xena catches fish at random, at a constant rate of 0.6 per hour.

(a) Find the probability that Xena catches exactly 4 fish in a 5-hour period.
(2)

The probability of Xena catching no fish in a period of t hours is less than 0.16

(b) Find the minimum value of t, giving your answer to one decimal place.
(3)

Independently of Xena, Zion catches fish at random with a mean rate of 0.8 per hour.

Xena and Zion try using new bait to catch fish. The number of fish caught in total by Xena and Zion after using the new bait, in a randomly selected 4-hour period, is 12

(c) Use a suitable test to determine, at the 5% level of significance, whether or not there is evidence that the rate at which fish are caught has increased after using the new bait. State your hypotheses clearly and the p-value used in your test.
(5)

We are given an average rate which indicates Poisson Distribution.

(a) $X \rightarrow$ # of fish in 5-hour period   define variable

in 1 hour $\rightarrow$ 0.6 fish $\therefore 0.6 \times \underset{\text{5 hours}}{5} = 3$ fish

$X \sim Po(3)$ M1

$P(X=4) = 0.168$ to 3sf   A1

(b) $T \rightarrow$ # of fish in t hours   define variable

$T \sim Po(0.6t)$

$P(T=0) < 0.16$ Formula for Poisson Distribution: $P(X=x) = \frac{e^{-\lambda} \times \lambda^x}{x!}$

Sub. into formula for poisson:

$\frac{e^{-0.6t} \times \cancel{(0.6t)^0}^{\,1}}{\cancel{0!}^{\,1}} < 0.16$ M1

$e^{-0.6t} < 0.16$ apply ln to both sides

$\ln e^{-0.6t} < \ln 0.16$ log law: $\ln e^x = x$

$-0.6t < \ln 0.16$ dM1

$t > \frac{\ln 0.16}{-0.6}$ flip inequality sign when you divide by negative

$t > 3.054 \Rightarrow$ $t = 3.1$ to 1dp. minimum value

A1

DO NOT WRITE IN THIS AREA

Question 2 continued

(c) Hypotheses

$Z \rightarrow$ fish Zion caught in 4h

$X \rightarrow$ fish Xena caught in 4h

$Z + X = Y \rightarrow$ fish both of them caught

define variables

$Z \sim Po(0.8 \times 4) = Po(3.2)$

$X \sim Po(0.6 \times 4) = Po(2.4)$

we can add these two since they are independent

$\therefore Y \sim Po(3.2 + 2.4) \rightarrow Y \sim Po(5.6)$ distribution for total # of fish

Hypotheses B1

$H_0: \lambda = 5.6$

$H_1: \lambda > 5.6$ B1

$\rightarrow$ they caught 12 fish in 4h

$P(Y \geq 12) = 1 - P(Y \leq 11)$ M1

$= 1 - 0.9875$

$= 0.01248 < 0.05 \therefore$ falls in critical region A1

Sufficient evidence to reject H_0, average rate of fish caught has increased A1

Question 2 continued

Question 2 continued

(Total for Question 2 is 10 marks)

3. In a game, a coin is spun 5 times and the number of heads obtained is recorded. Tao suggests playing the game 20 times and carrying out a chi-squared test to investigate whether the coin might be biased.

(a) Explain why playing the game only 20 times may cause problems when carrying out the test.

(1)

Chris decides to play the game 500 times. The results are as follows

Number of heads	0	1	2	3	4	5
Observed frequency	2	27	93	181	146	51

Chris decides to test whether or not the data can be modelled by a binomial distribution, with the probability of a head on each spin being 0.6

She calculates the expected frequencies, to 2 decimal places, as follows

Number of heads	0	1	2	3	4	5
Expected frequency	5.12	38.40	115.20	172.80	129.60	38.88

(b) State the number of degrees of freedom in Chris' test, giving a reason for your answer.

(1)

(c) Carry out the test at the 5% level of significance.
You should state your hypotheses, test statistic, critical value and conclusion clearly.

(5)

(d) Showing your working, find an alternative model which would better fit Chris' data.

(2)

(a) Sample size is too small B1

(b) DoF = Columns - 1

= 6 - 1 = 5

5 DoF as we have 6 columns and no pooling as all $E_i > 5$ B1

Question 3 continued

(c) Hypotheses

H_0: $B(5, 0.6)$ is a suitable model

H_1: $B(5, 0.6)$ is not a suitable model B1

Use $\chi^2 = \sum \frac{O_i^2}{E_i} - N$ to get our test statistic

$\chi^2 = \frac{2^2}{5.12} + \frac{27^2}{38.4} + \frac{93^2}{115.2} + \frac{181^2}{172.8} + \frac{146^2}{129.6} + \frac{51^2}{38.88} - 500$ M1

$= 15.8063 \rightarrow 15.8$ to 3sf test statistic A1

$\chi^2_5(5\%) = 11.070$ critical value from tables B1

$15.8 > 11.070 \therefore$ sufficient evidence to reject H_0

$B(5, 0.6)$ is not a suitable model for the data A1

(d) Calculate new p value:

$p = \frac{(0\times2)+(1\times27)+(2\times93)+(3\times181)+(4\times146)+(5\times51)}{5\times500} = \frac{3.19}{5} = 0.638$ M1

$\therefore B(5, 0.638)$ A1

new p-value

Question 3 continued

Question 3 continued

(Total for Question 3 is 9 marks)

4. The discrete random variable X has the following probability distribution

x	0	2	3	6
$P(X=x)$	p	0.25	q	0.4

(a) Find in terms of q

(i) $E(X)$

(ii) $E(X^2)$

(2)

Given that $Var(X) = 3.66$

(b) show that $q = 0.3$

(3)

In a game, the score is given by the discrete random variable X

Given that games are independent,

(c) calculate the probability that after the 4th game has been played, the total score is exactly 20

(3)

A round consists of 4 games plus 2 bonus games. The bonus games are only played if after the 4th game has been played the total score is exactly 20

A prize of £10 is awarded if 6 games are played in a round **and** the total score for the round is at least 27

Bobby plays 3 rounds.

(d) Find the probability that Bobby wins at least £10

(6)

Formulae for Mean and Variance

$E(X) = \sum x P(X=x)$ mean, μ

$E(X^2) = \sum x^2 P(X=x)$ mean of squares

$Var(X) = E(X^2) - [E(X)]^2$ variance, σ^2

(a)

x	0	2	3	6
x^2	0	4	9	36
$P(X=x)$	p	0.25	q	0.4

i. $E(X) = 0 \times p + 2(0.25) + 3q + 6(0.4) = 3q + 2.9$ B1

ii. $E(X^2) = 0 \times p + 4(0.25) + 9q + 36(0.4) = 9q + 15.4$ B1

DO NOT WRITE IN THIS AREA

Question 4 continued

(b) We know that $Var(X) = E(X^2) - [E(X)]^2$ and that $Var(X) = 3.66$

$3.66 = (15.4 + 9q) - (2.9 + 3q)^2$ M1

$3.66 = 15.4 + 9q - 8.41 - 17.4q + 9q^2$ Rearrange

$9q^2 + 8.4q - 3.33 = 0$ quadratic.

Use Quadratic Formula: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$q = \frac{-8.4 \pm \sqrt{8.4^2 - 4(9)(-3.33)}}{2(9)} \Rightarrow q = 0.3, -\frac{37}{30}$ M1

Since $q > 0$ as it's a probability, $q = 0.3$ shown A1

(c) $P(\underbrace{X_1 + X_2 + X_3 + X_4}_{\text{scores}} = 20) = P(6,6,6,2 \text{ or } 6,6,2,6 \text{ or } 6,2,6,6 \text{ or } 2,6,6,6)$ M1

$= 4 \times (\underbrace{0.25}_{P(X=2)} \times \underbrace{(0.4)^3}_{P(X=6)})$ M1

$= 0.064$ A1

(d) For the total to be ≥ 27 after 6 rounds, rounds 5+6 have to add up to 7 (since for 6 rounds to even be played, rounds 1-4 must add up to 20)

$P(X_5 + X_6 \geq 7) = P(6,6 \text{ or } 2,3 \text{ or } 3,2 \text{ or } 6,3 \text{ or } 3,6)$ M1

$= (0.4)^2 + 2(0.25 \times 0.4) + 2(0.4 \times 0.3)$ M1

$= 0.6$

$\therefore$ For score to be $\geq 27 \Rightarrow P(X_1 + X_2 + X_3 + X_4 = 20) \times P(X_5 + X_6 \geq 27)$

$= 0.064 \times 0.6$

$P(\text{score} \geq 27) = 0.0384$ M1

For him to win at least 10£, he needs to score ≥ 27 at least once.

$S \rightarrow$ # of times he scores ≥ 27

$S \sim B(3, 0.0384)$ dM1

$P(S \geq 1) = 1 - P(S = 0)$ M1

$= 0.1108$

$\hookrightarrow 0.111$ to 3sf A1

Question 4 continued

(Total for Question 4 is 14 marks)

TOTAL FOR FURTHER STATISTICS 1 IS 40 MARKS