Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

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.
- Good luck with your examination.

Turn over

1. *Flobee* sells tomato seeds in packets, each containing 40 seeds. *Flobee* advertises that only 4% of its tomato seeds do not germinate.

Amodita is investigating the germination of *Flobee*'s tomato seeds. She plants 125 packets of *Flobee*'s tomato seeds and records the number of seeds that do not germinate in each packet.

Number of seeds that do not germinate	0	1	2	3	4	5	6 or more
Frequency	15	35	38	22	10	5	0

Amodita wants to test whether the binomial distribution B(40, 0.04) is a suitable model for these data.

The table below shows the expected frequencies, to 2 decimal places, using this model.

Number of seeds that do not germinate	0	1	2	3	4	5 or more
Expected Frequency	24.42	40.70	r	17.45	6.73	s

(a) Calculate the value of r and the value of s

(2)

(b) Stating your hypotheses clearly, carry out the test at the 5% level of significance. You should state the number of degrees of freedom, critical value and conclusion clearly.

(6)

Amodita believes that *Flobee* should use a more realistic value for the percentage of their tomato seeds that do not germinate.
She decides to test the data using a new model B(40, p)

(c) Showing your working, suggest a more realistic value for p

(2)

(a) $P(X=2) = 0.264.$ ← using $B(40, 0.04)$ on a calculator OR statistical tables

$P(X \geq 5) = 1 - P(X \leq 4)$

$= 1 - 0.978.$

$= 0.0210$ (3.s.f) ①

$r = 125 \times P(X=2) = 125 \times 0.264... = 33.07$ ①

$s = 125 \times P(X \geq 5) = 125 \times 0.0210 = 2.63$

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) We're looking at expected frequency distributions, so use a Chi-Squared test for significance:

H_0: B(40, 0·04) is a suitable model. ← State null and alternative hypothesise.

H_1: B(40, 0·04) is not a suitable model. (1)

Cells are combined when expected frequency is less than 5, so combine last 2 cells. (1)

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E} \quad \Big| \quad O = \text{observed}, \quad E = \text{expected}$$

$$X^2 = \frac{(15-24\cdot42)^2}{24\cdot42} + \frac{(35-40\cdot7)^2}{40\cdot7} + \frac{(38-33\cdot07)^2}{33\cdot07} + \frac{(22-17\cdot45)^2}{17\cdot45} + \frac{(15-(6\cdot73+2\cdot63))^2}{6\cdot73+2\cdot63}$$ (1)

$X^2 = 9\cdot752$ (1)

Degrees of freedom (v) = 5 − 1 = 4 (1)

(5 = no. of columns; 1 = no. of constraints ≥ 1)

v = 4, significance level = 5% ⇒ critical value = 9·488 (from stat tables)

There is significant evidence to reject H_0 because 9·752 > 9·488, therefore Amodita's model is not supported. (1)

$X^2 > X_4^2(5\%)$ ∴ reject H_0

(c) $$p = \frac{0\times15 + 1\times35 + 2\times38 + 3\times22 + 4\times10 + 5\times5}{125\times40}$$ (1)

← uses the existing data

$p = 0\cdot0484$ (1)

Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)

2. Rowan and Alex are both check-in assistants for the same airline.
The number of passengers, R, checked in by Rowan during a 30-minute period can be modelled by a Poisson distribution with mean 28

(a) Calculate $P(R \geqslant 23)$

(1)

The number of passengers, A, checked in by Alex during a 30-minute period can be modelled by a Poisson distribution with mean 16, where R and A are independent.
A randomly selected 30-minute period is chosen.

(b) Calculate the probability that exactly 42 passengers in total are checked in by Rowan and Alex.

(2)

The company manager is investigating the rate at which passengers are checked in.
He randomly selects 150 non-overlapping 60-minute periods and records the total number of passengers checked in by Rowan and Alex, in each of these 60-minute periods.

(c) Using a Poisson approximation, find the probability that for at least 25 of these 60-minute periods Rowan and Alex check in a total of fewer than 80 passengers.

(4)

On a particular day, Alex complains to the manager that the check-in system is working slower than normal. To see if the complaint is valid the manager takes a random 90-minute period and finds that the total number of people **Rowan** checks in is 67

(d) Test, at the 5% level of significance, whether or not there is evidence that the system is working slower than normal. You should state your hypotheses and conclusion clearly and show your working.

(4)

(a) $R \sim Po(\lambda = 28)$

$P(R \geq 23) = 1 - P(R \leq 22)$ ← put in calculator or use $P(X = x) = \frac{e^{-\lambda} \times \lambda^{x}}{x!}$

$= 1 - 0.148$

$= 0.852$ (1)

(b) $A \sim Po(\lambda = 16)$ and $R \sim Po(\lambda = 28)$ then let.

Y = total calls in 30 min) $\Rightarrow Y \sim Po(\lambda = 44)$ (1)

↳ $\lambda_A + \lambda_B = 16 + 28$

$P(Y = 42) = 0.0587$ (1)

DO NOT WRITE IN THIS AREA

Question 2 continued

(c) Using $Y \sim Po(\lambda = 44)$: ← this is 30 min periods

For $2 \times 30 = 60$ minutes, $Y_2 \sim P(\lambda = 2 \times 44 = 88)$

P(fewer than 80 passengers checked in) = $P(X < 80)$
↳ in one 60-minute period

$= 0.183$ (1)

can approximate as n is large, P is small →
Using a binomial model $X \sim B(150, 0.183)$, then mean value for 150 periods $= np = 150 \times 0.183 = 27.48$. (1)

$T \sim Po(27.48)$ ← use new mean $\lambda = 27.48$ for T = 'number of days on which fewer than 80 passengers are checked in'.

$P(T \geq 25) = 1 - P(T \leq 24)$ (1)

$= 1 - 0.2922$

$= 0.708$ (1)

(d) $90 = 3 \times 30$ mins $\Rightarrow \lambda_{Rowan} = 3 \times 28 = 84$

$H_0: \lambda = 84$ ← system is working and Rowan checks-in expected number of passengers.
(1)
$H_1: \lambda < 84$ ← Rowan checks-in fewer passengers than expected.

$J \sim Po(\lambda = 84)$ (1)

$P(J \leq 67) = 0.0325$ (1)
↑ mean is 84 and $67 < 84$, so test for 'more extreme' values ($\leq$).

$0.0325 < 0.05$ significance so reject H_0. There is evidence at a 5% level of significance that the system is working slower than usual. (1)

Question 2 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 11 marks)

3. The discrete random variable X has probability distribution

x	-3	-2	-1	0	2	5
$\mathrm{P}(X = x)$	0.3	0.15	0.1	0.15	0.1	0.2

(a) Find $\mathrm{E}(X)$

(1)

Given that $\mathrm{Var}(X) = 8.79$

(b) find $\mathrm{E}(X^2)$

(2)

The discrete random variable Y has probability distribution

y	-2	-1	0	1	2
$\mathrm{P}(Y = y)$	$3a$	a	b	a	c

where a, b and c are constants.

For the random variable Y

$$\mathrm{P}(Y \leqslant 0) = 0.75 \quad \text{and} \quad \mathrm{E}(Y^2 + 3) = 5$$

(c) Find the value of a, the value of b and the value of c

(5)

The random variable $W = Y - X$ where Y and X are independent.

The random variable $T = 3W - 8$

(d) Calculate $\mathrm{P}(W > T)$

(4)

(a) $\mathrm{E}(X) = \sum_{i=1}^{n} x_i \times \mathrm{P}(X = x)_i$

$\mathrm{E}(X) = -3 \times 0.3 + -2 \times 0.15 + -1 \times 0.1 + 0 \times 0.15 + 2 \times 0.1 + 5 \times 0.2$

$\mathrm{E}(X) = -0.1$ (1)

(b) $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$ (1)

$\mathrm{E}(X^2) = 8.79 - (-0.1)^2 = 8.8$ (1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

(c) $E(Y^2 + 3) = 5 \Rightarrow E(Y^2) = 5 - 3 = 2$

$E(Y^2) = (-2)^2 \times 3a + (-1)^2 \times a + 0^2 \times b + 1^2 \times a + 2^2 \times c$ (1)

$2 = 12a + a + a + 4c$

$1 = 7a + 2c$ (1)

$P(Y \leq 0) = 0.75 \Rightarrow 3a + a + b = 0.75 \Rightarrow 4a + b = 0.75$

$5a + b + c = 1$ ← probabilities sum to 1

$7a + 2c = 1$
$4a + b = 0.75$ (1)
$5a + b + c = 1$ (1)

$a = 0.1$
$b = 0.35$
$c = 0.15$

(1)

(d) $P(W > T) = P(W > 3W - 8)$ ← $T = 3W - 8$

$= P(W < 4)$ (1) ← $-3W, \times -1, \div 2$

$P(W < 4) = 1 - [P(X = -3) \times P(Y = 1)$ (1) $\quad 1 - (-3) \geq 4$
$+ P(X = -3) \times P(Y = 2)$ $\quad 2 - (-3) \geq 4$
$+ P(X = -2) \times P(Y = 2)]$ $\quad 2 - (-2) \geq 4$

These are the cases in which $W < 4$ is False →

$= 1 - [0.3 \times 0.1 + 0.3 \times 0.15 + 0.15 \times 0.15]$ (1)

$= 0.9025$ (1) ← using a, b, c from (c).

P66791RA01116

Turn over

Question 3 continued

Question 3 continued

(Total for Question 3 is 12 marks)

4. Charlie carried out a survey on the main type of investment people have.
The contingency table below shows the results of a survey of a random sample of people.

		Main type of investment			
		Bonds	Cash	Stocks	
Age	25–44	a	$b-e$	e	
	45–75	c	$d-59$	59	$c+d$
		$a+c$			$a+b+c+d$

(a) Find an expression, in terms of a, b, c and d, for the difference between the observed and the expected value $(O-E)$ for the group whose main type of investment is Bonds and are aged 45–75
Express your answer as a single fraction in its simplest form.

(4)

Given that $\sum \frac{(O-E)^2}{E} = 9.62$ for this information,

(b) test, at the 5% level of significance, whether or not there is evidence of an association between the age of a person and the main type of investment they have.
You should state your hypotheses, critical value and conclusion clearly.
You may assume that no cells need to be combined.

(3)

(a) $E = \frac{(c+d)(a+c)}{a+b+c+d}$ ① ← for contingency tables, $E = \frac{\text{row total} \times \text{col total}}{\text{overall total}}$

(see annotations on table in Q4)

$O - E = c - \frac{(c+d)(a+c)}{a+b+c+d}$ ①

value observed in the table

$= \frac{ca + cb + c^2 + cd - ac - c^2 - ad - dc}{a+b+c+d}$ ①

$O - E = \frac{cb - ad}{a+b+c+d}$ ①

DO NOT WRITE IN THIS AREA

Question 4 continued

(b) H_0: there is no association between a person's age and the type of investment they have.

H_1: there is an association between a person's age and the type of investment they have. ① for H_0 AND H_1

Degrees of freedom = $(3-1)(2-1) = 2$
rows (ages) ↑ ↑ columns (types) ①

χ^2_2 5% = 5.991 ① ← from statistical tables

Reject H_0. There is evidence of an association between a person's age and the type of investment they have because 5.991 < 9.62 (the test statistic). ①

Question 4 continued

(Total for Question 4 is 7 marks)

TOTAL FOR FURTHER STATISTICS 1 IS 40 MARKS