Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Specimen Paper

Paper Reference **8FM0/23**

Further Mathematics

Advanced Subsidiary

23: Further Statistics 1

You must have:
Mathematical Formulae and Statistical Tables, calculator

Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. 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.
- 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 ▶

S60742A

1/

S 6 0 7 4 2 A 0 1 1 2

Pearson

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

1. In a survey, 200 people aged 25 and older were randomly selected and asked how much time they spend on social media each day. The table below shows a summary of the results.

		Less than 1 hour	1 hour or more
Age	**25 to 54**	60	74
	55 and older	32	34

Noah carries out a test, at the 5% level of significance, to see if there is an association between age and time spent on social media. He uses the hypotheses

H_0: There is no association between age and time spent on social media.
H_1: There is an association between age and time spent on social media.

He calculates $\sum \frac{(O-E)^2}{E} = 0.245$ for this information.

(a) State the conclusion of the test. Justify your answer.

(2)

(b) Explain why having a large number of age groups may cause a problem when carrying out the hypothesis test.

(1)

Jade decides to take the same information and subdivide the age groups.

She then uses the information in the partially complete table below to carry out a test, at the 5% level of significance, of the same hypotheses.

		Observed	
		Less than 1 hour	**1 hour or more**
Age	**25 to 34**	18	28
	35 to 44	20	25
	45 to 54	22	21
	55 to 64	25	15
	65 and older	7	19

	$\frac{(O-E)^2}{E}$	
	Less than 1 hour	**1 hour or more**
25 to 34	0.47	0.40
35 to 44	0.02	0.02
45 to 54	0.25	0.21
55 to 64	2.37	2.02
65 and older		

(c) Complete Jade's hypothesis test. State clearly the degrees of freedom and the critical value used in the test.

(4)

(d) State, giving a reason, which of the conclusions in part (a) and part (c) you believe to be the more reliable.

(1)

Question 1 continued

(a) ★ For contingency tables the formula for DoF is:

$$DoF = (\#\text{of rows} - 1)(\#\text{of columns} - 1)$$

$DoF = (2-1)(2-1) = 1$ DoF

Get critical value from tables:

$\chi^2_1(5\%) = 3.841$ M1

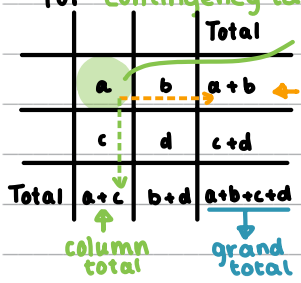
Given test statistic $\rightarrow \chi^2 = 0.245$

$0.245 < 3.841$ ∴ insufficient evidence to reject H_0. Not enough evidence to suggest an association between age and time spent on social media. A1

(b) Some of the expected frequencies may be smaller than 5 B1

(c) We need to calculate the expected frequencies for "older than 65"

For contingency tables we calculate the expected frequency like this:



to get the E_i of this, we calculate $E_i = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$

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

			row total
65 and older	7	19	26
column total	92	108	200

< 1h and older than 65

$E_i = \frac{26 \times 92}{200} = 11.96$

>1h and older than 65

$E_i = \frac{108 \times 26}{200} = 14.04$

Calculate $\frac{(O-E)^2}{E}$:

$\frac{(7-11.96)^2}{11.96} = 2.05699...$

$\frac{(19-14.04)^2}{14.04} = 1.7523...$

Use $\chi^2 = \sum \frac{(O-E)^2}{E}$ to get our test statistic

$\chi^2 = 0.47 + 0.40 + 0.02 + 0.02 + 0.25 + 0.21 + 2.37 + 2.02 + 2.06 + 1.75$ M1

$= 9.57$ test statistic A1

★ For contingency tables the formula for DoF is:

$\text{DoF} = (\#\text{of rows} - 1)(\#\text{of columns} - 1)$

$\text{DoF} = (5-1)(2-1) = 5$ DoF

Get critical value from tables:

$\chi^2_5(5\%) = 9.488$ B1

$9.57 > 9.488$ $\therefore$ reject H_0. There is an association between *age and* time spent on social media. A1

(d) Part (c) uses more information, so it's probably more reliable. B1

Question 1 continued

(Total for Question 1 is 8 marks)

2. The discrete random variable X has probability distribution given by

x	2	3	6	11
$P(X = x)$	$\frac{1}{50}$	a	$\frac{1}{25}$	b

The discrete random variable $Y = X^2$

Given that $E(Y) = 50.3$

(a) find the value of a and the value of b

(3)

(b) Find $P(9 - Y > 0)$

(2)

Independent observations $X_1, X_2, X_3, \ldots, X_{120}$ of X are taken.

The random variable T represents the total number of these 120 observations that are even.

(c) Find

(i) $E(T)$

(ii) $\text{Var}(T)$

(2)

(d) Find, using a suitable approximation, $P(T > 10)$

(3)

(a)

x	2	3	6	11
$Y = x^2$	4	9	36	121
$P(X=x)$	$\frac{1}{50}$	a	$\frac{1}{25}$	b

Formula for Mean

$E(X) = \Sigma x P(X=x)$ mean, μ

$E(Y) = 4 \times \frac{1}{50} + 9 \times a + 36 \times \frac{1}{25} + 121b = 50.3$ M1

$9a + 121b = 48.38$ Eq1

We know that Σprobabilities $= 1$

$1 = \frac{1}{50} + a + \frac{1}{25} + b$

$a + b = 0.94$ Eq2

Solve simultaneously Eq1 and Eq2:

$9a + 121b = 48.38$

$a + b = 0.94 \quad |\times 9| \quad 9a + 9b = 8.46$ M1

$112b = 40.32 \longrightarrow b = 0.36, a = 0.58$ A1

(b) $P(9-Y>0) \rightarrow P(Y<9)$:

$Y=x^2$	4	9	36	121
$P(X=x)$	$\frac{1}{50}$	0.58	$\frac{1}{25}$	0.36

only $Y=4$ is <9

$\therefore P(Y<9)=\frac{1}{50}$ M1 A1

(c)

x	2	3	6	11
$P(X=x)$	$\frac{1}{50}$	0.58	$\frac{1}{25}$	0.36

the only even results are $x=2$ and $x=6$.

i. $E(T)=120\times(\frac{1}{50}+\frac{1}{25})$

$=7.2$ B1

ii. The distribution is basically Binomial: $T\sim B(120, 0.06)$ ($\frac{1}{50}+\frac{1}{25}$)

$\therefore$ Use Formula for Var. of Binomial:

$Var(X)=np(1-p)$

Substitute:

$Var(T)=120\times0.06\times(1-0.06)$

$=6.77$ B1

(d) Use $T\sim B(120, 0.06)$

$E(T)=7.2=\lambda \quad \therefore T\sim Po(7.2)$ M1

$P(T>10)=1-P(T\leq10)$ M1

$=1-0.88667$

$=0.113$ A1

(Total for Question 2 is 10 marks)

3. A hotel has 30 rooms. The manager models the number of empty rooms each Friday night using a binomial distribution, B(30, 0.08)

The manager recorded the number of empty rooms in the hotel each Friday night over a period of 80 weeks.

Number of empty rooms	0	1	2	3	4	5	6 or more
Frequency	14	18	22	11	10	5	0

The table below shows the expected frequencies using the manager's model.

Number of empty rooms	0	1	2	3	4	5 or more
Expected frequency	6.56	17.11	r	17.50	10.27	s

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

(2)

(b) Stating your hypotheses clearly, test the manager's model at the 5% level of significance.

(6)

(c) Suggest an improved model for the number of empty rooms in the hotel each Friday night.

(2)

(a) $X \rightarrow$ # of empty rooms on a Friday night

Use $X \sim B(30, 0.08)$

$r = 80 \times P(X = 2) = 21.57$ M1

$s = 80 - (6.56 + 17.11 + 21.57 + 17.5 + 10.24) = 6.99$ A1

(b) Hypotheses

H_0: B(30,0.08) is a suitable model

H_1: B(30,0.08) is not a suitable model B1

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

$\chi^2 = \frac{14^2}{6.56} + \frac{18^2}{17.11} + \frac{22^2}{21.57} + \frac{11^2}{17.50} + \frac{10^2}{10.27} + \frac{5^2}{6.99} - 80$ M1

$= 11.5$ A1

DoF = 6 - 1 = 5 B1

Get critical value from tables:

$\chi^2_5(5\%) = 11.070 < 11.5 \therefore$ reject H_0. B(30,0.08) is not a suitable model for empty hotel rooms. M1A1

Question 3 continued

(c) Use observed data to get a new p:

$p = \frac{0\times14 + 1\times18 + 2\times22 + 3\times11 + 4\times10 + 5\times5}{80\times30}$ M1

$= \frac{2}{30}$

$\therefore$ New Model: $B(30, \frac{2}{30})$ A1

(Total for Question 3 is 10 marks)

4. An office has a photocopier and a printer. The photocopier and the printer break down independently.

The number of breakdowns per month for the photocopier follows a Poisson distribution with mean 2.4

The number of breakdowns per month for the printer follows a Poisson distribution with mean 1.6

(a) Determine which machine is more likely to break down exactly twice in one month.

(2)

(b) Work out the probability that both machines break down at most once in one month.

(2)

In a randomly selected month there were 4 breakdowns.

(c) Find the probability that in this month the photocopier broke down more than the printer.

(4)

A repair company carried out repairs on the photocopier and the printer. Following the repairs, there were a total of 3 breakdowns in two months.

(d) Test, at the 5% level of significance, whether or not there is evidence that the rate of breakdowns has decreased following the repairs.

(4)

(a) $X \rightarrow$ #of break downs of Photocopier

$Y \rightarrow$ #of break downs of Printer

$X \sim Po(2.4)$

$Y \sim Po(1.6)$

$P(X=2) = 0.261 \longrightarrow \therefore$ the Photocopier is more likely to break down exactly twice **A1**

$P(Y=2) = 0.258$ **M1**

(b) $P(X \leq 1) \times P(Y \leq 1)$ **M1** ("and")

$= 0.3084 \times 0.5249$

$= 0.162$ **A1**

(c) Given that total breakdowns = 4:

Photocopier 4 and Printer 0 or Photocopier 3 and Printer 1

$\therefore \dfrac{P(X=4)\times P(Y=0) + P(X=3)\times P(Y=1)}{P(X+Y=4)} = \dfrac{0.125\times 0.201 + 0.209\times 0.323}{0.195}$ **M1M1M1**

$= \dfrac{297}{625}$

$= 0.475$ to 3sf **A1**

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

(d) Hypotheses

H_0: $\lambda = 4$ $\times 2 \rightarrow$ $\mu = 8$

H_1: $\lambda < 4$ $\mu < 8$ B1

(1 month) (2 months)

$X + Y = T$

$T \sim Po(8)$ M1

$P(T \leq 3) = 0.0424 < 0.05 \therefore$ Reject H_0. Sufficient evidence to suggest that the breakdown rate has decreased after the repairs. M1A1

Question 4 continued

(Total for Question 4 is 12 marks)

TOTAL FOR FURTHER STATISTICS 1 IS 40 MARKS