Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

Friday 16 June 2023

Afternoon (Time: 1 hour 30 minutes)

Paper reference **9FM0/3B**

Further Mathematics

Advanced

PAPER 3B: Further Statistics 1

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 statistical tables should be quoted in full. If a calculator is used instead of 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 7 questions in this paper. The total mark for this part of the examination 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

P74080A

N:1/1/1/1/

P 7 4 0 8 0 A 0 1 2 4

Pearson

1. The discrete random variable X has probability distribution

x	-2	-1	0	1	3
$\mathrm{P}(X=x)$	0.25	a	b	a	0.30

where a and b are probabilities.

(a) Find $\mathrm{E}(X)$ **(2)**

Given that $\mathrm{Var}(X) = 3.9$

(b) find the value of a and the value of b **(4)**

The independent random variables X_1 and X_2 each have the same distribution as X

(c) Find $\mathrm{P}\left(X_1 + X_2 > 3\right)$ **(3)**

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) $E(X) = -2(0.25) + \cancel{(-1)(a)} + \cancel{0(b)} + \cancel{1(a)} + 3(0.3)$ M1

$= -0.5 + 0.9$

$E(X) = 0.4$ A1

Question 1 continued

(b)

x	-2	-1	0	1	3
x^2	4	1	0	1	9
$P(X=x)$	0.25	a	b	a	0.30

$E(X)=0.4$ from (a)

$E(X^2)=4(0.25)+1(a)+0(b)+1(a)+9(0.3)$ M1

$= 1+a+a+2.7$

$E(X^2)=2a+3.7$

substitute into $Var(X)=3.9$ (given)

$Var(X)=E(X^2)-[E(X)]^2$ dM1

$3.9 = 2a+3.7-(0.4)^2$

$3.9-3.7+0.16=2a$

$0.2+0.16=2a$

$0.36=2a$

$a=0.18$ A1

Use fact that Σprobabilities $=1$:

$2a+b+0.55=1$

$2a+b=0.45$

Substitute $a=0.18$ into $2a+b=0.45$ and solve for b:

$2(0.18)+b=0.45$

$b=0.45-0.36$

$b=0.09$ A1

$\therefore a=0.18,\ b=0.09$ a and b

(c) We have 5 possible values for X: $-1, -2, 0, 1$ and 3

For $X_1+X_2>3$ to be true, the possible combinations are:

$X_1=1$ and $X_2=3$
$X_1=3$ and $X_2=1$
$X_1=3$ and $X_2=3$

and → multiply

M1

$\therefore P(X_1+X_2>3)=(0.18\times0.3)+(0.3\times0.18)+(0.3\times0.3)$ ($P(X=1)=0.18$, $P(X=3)=0.3$)

$=0.054+0.054+0.09$ M1

$P(X_1+X_2>3)=0.198$ A1

(Total for Question 1 is 9 marks)

2. Telephone calls arrive at a call centre randomly, at an average rate of 1.7 per minute. After the call centre was closed for a week, in a random sample of 10 minutes there were 25 calls to the call centre.

(a) Carry out a suitable test to determine whether or not there is evidence that the rate of calls arriving at the call centre has changed.
Use a 5% level of significance and state your hypotheses clearly.
(4)

Only 1.2% of the calls to the call centre last longer than 8 minutes.

One day Tiang has 70 calls.

(b) Find the probability that out of these 70 calls Tiang has more than 2 calls lasting longer than 8 minutes.
(3)

The call centre records show that 95% of days have at least one call lasting longer than 30 minutes.
On Wednesday 900 calls arrived at the call centre and none of them lasted longer than 30 minutes.

(c) Use a Poisson approximation to estimate the proportion of calls arriving at the call centre that last longer than 30 minutes.
(4)

(a) Use Poisson Distribution since we are talking about an average rate.

$X \rightarrow$ # of calls in 10min span define variable

$X \sim Po(10 \times 1.7)$ given rate per minute, multiply by 10 to get rate for 10mins

$X \sim Po(17)$ M1

Hypotheses

$H_0: \lambda = 17$ 5% significance $\rightarrow$ 0.025 in each tail.

$H_1: \lambda \neq 17$ B1

Since $25 > 17$, test upper tail.

$P(X \geq 25) = 1 - P(X \leq 24)$ A1

$= 0.0406 > 0.025$ $\therefore$ 25 calls do not fall in the critical region A1

insufficient evidence to reject H_0, not enough evidence to suggest that the rate of calls has changed

(b) Use binomial distribution since we have fixed # of trials

$Y \rightarrow$ # of calls lasting longer than 8 minutes

$Y \sim B(70, 0.012)$ M1

$P(Y > 2) = 1 - P(Y \leq 2)$ M1

$= 1 - 0.947725$

$= 0.0523$ 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

Question 2 continued

(c) $C \rightarrow$ # of calls out of 900 that last longer than 30 mins

$C \sim B(900, p)$ use poisson approximation $\rightarrow C \approx\sim Po(900p)$ M1

$P(C=0) = \dfrac{e^{-900p} \times \cancel{(-900p)^0}^{\,1}}{\cancel{0!}^{\,1}}$ M1

Formula for Poisson Distribution:

$P(X=x) = \dfrac{e^{-\lambda} \times \lambda^x}{x!}$

$0.05 = e^{-900p}$

$\ln 0.05 = \ln e^{-900p}$ apply ln to both sides

$\ln 0.05 = -900p$ M1 use ln rule: $\ln e^x = x$

$\dfrac{\ln 0.05}{-900} = p$

$p = 0.003328 \rightarrow 0.00333$ A1

Question 2 continued

Question 2 continued

(Total for Question 2 is 11 marks)

3. In a class experiment, each day for 170 days, a child is chosen at random and spins a large cardboard coin 5 times and the number of heads is recorded.
The results are summarised in the following table.

Number of heads	0	1	2	3	4	5
Frequency	3	10	45	62	38	12

Marcus believes that a B(5, 0.5) distribution can be used to model these data and he calculates expected frequencies, to 2 decimal places, as follows

Number of heads	0	1	2	3	4	5
Expected frequency	r	26.56	s	s	26.56	r

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

(3)

(b) Carry out a suitable test, at the 5% level of significance, to determine whether or not the B(5, 0.5) distribution is a good model for these data.
You should state clearly your hypotheses, the test statistic and the critical value used.

(6)

Nima believes that a better model for these data would be B(5, p)

(c) Find a suitable estimate for p

(1)

To test her model, Nima uses this value of p, to calculate expected frequencies as follows

Number of heads	0	1	2	3	4	5
Expected frequency	2.07	14.65	41.44	58.63	41.47	11.74

The test statistic for Nima's test is 1.62 (to 3 significant figures)

(d) State,
(i) giving your reasons, the degrees of freedom
(ii) the critical value
that Nima should use for a test at the 5% significance level.

(3)

(e) With reference to Marcus' and Nima's test results, comment on
(i) the probability of the coin landing on heads,
(ii) the independence of the spins of the coin.
Give reasons for your answers.

(2)

(a) $X \sim B(5, 0.5)$

$P(X=0) = P(X=5) = 0.03125$ | $\times 170$ | $= r = 5.31$ M1A1A1

$P(X=2) = P(X=3) = 0.3125$ | $\times 170$ | $= s = 53.1$

Question 3 continued

(b) Hypotheses

H_0: $B(5, 0.5)$ is a good model

H_1: $B(5, 0.5)$ is not a good model B1

# of heads	0	1	2	3	4	5
observed frequency	3	10	45	62	38	12
expected frequency	5.31	26.56	53.1	53.1	26.56	5.31
$\frac{O_i^2}{E_i}$	$\frac{3^2}{5.31}$	$\frac{10^2}{26.56}$	$\frac{45^2}{53.1}$	$\frac{62^2}{53.1}$	$\frac{38^2}{26.56}$	$\frac{12^2}{5.31}$

Use formula for test statistic: M1

$$\chi^2 = \sum \frac{O_i^2}{E_i} - N$$

Substitute:

$$\chi^2 = \frac{3^2}{5.31} + \frac{10^2}{26.56} + \frac{45^2}{53.1} + \frac{62^2}{53.1} + \frac{38^2}{26.56} + \frac{12^2}{5.31} - 170 \quad \text{A1}$$

$\chi^2 = 27.4$ test statistic

dof: $6 - 1 = 5$ dof. B1

# of columns

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

$24.7 > 11.070$ $\therefore$ significant result. Sufficient evidence to reject H_0.

Marcus' model is not a good fit. A1

(c) $p = \frac{0\times3 + 1\times10 + 2\times45 + 3\times62 + 4\times38 + 5\times12}{5\times170} = 0.58588...$

$\hookrightarrow p = 0.586$ to 3sf B1

(d) i. Use pooling, since $E(0) < 5$. We calculated p from observed values $\therefore$ subtract 2 degrees of freedom.

dof = 5 columns − 2 constraints = 3 dof M1A1

ii. $\chi^2_3(5\%) = 7.815$ critical value B1

(e) i. Nima's model ($B(5, 0.58)$) is a good fit since $\underbrace{1.62}_{\text{her test stat.}} < \underbrace{7.815}_{\text{critical value}}$ $\therefore$ her H_0 is not rejected.

The calculated $p \approx 0.6$ suggests the dice is biased B1

ii. Nima's test was a good model and it was Binomial $\therefore$ we can assume spins were independent. B1

Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

(Total for Question 3 is 15 marks)

4. There are 32 students in a class.

Each student rolls a fair die repeatedly, stopping when their total number of sixes is 4
Each student records the total number of times they rolled the die.

Estimate the probability that the mean number of rolls for the class is less than 27.2

(6)

"stopping when their total number of sixes is 4" → Negative Binomial. We are looking for

# of trials until a certain # of successes is reached (4)

$X \rightarrow$ # of rolls it takes to get 4 sixes   define variable

$X \sim NegB(4, \frac{1}{6})$ M1

"mean number of rolls is less than 27.2" → mention of "mean" indicates Central Limit Theorem

$\bar{X} \approx \sim N(\mu, \frac{\sigma^2}{N})$

Calculate μ and σ^2:

$\mu = \frac{4}{\frac{1}{6}} = 4 \times 6 = 24 \Rightarrow \mu = 24$ A1

★ Formulae for Negative Binomial:

$\mu = \frac{r}{p}$ $\sigma^2 = \frac{r(1-p)}{p^2}$

$\sigma^2 = \frac{4 \times (1-\frac{1}{6})}{(\frac{1}{6})^2} = \frac{4 \times \frac{5}{6}}{\frac{1}{36}} = \cancel{36}^{6} \times 4 \times \frac{5}{\cancel{6}} = 120$ $\sigma^2 = 120$ A1

$\therefore \bar{X} \approx \sim N(24, \sqrt{\frac{120}{32}}^2)$ M1M1

in your calculator enter $\sqrt{\frac{120}{32}} = \sigma$

$P(\bar{X} < 27.2) = 0.95087$

$= 0.951$ A1

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

Question 4 continued

(Total for Question 4 is 6 marks)

5. A machine fills cartons with juice.

The amount of juice in a carton is normally distributed with mean μ ml and standard deviation 8 ml. σ

A manager wants to test whether or not the amount of juice in the cartons, X ml, is less than 330 ml. The manager takes a random sample of 25 cartons of juice and calculates the mean amount of juice $\bar{x}$ ml.

(a) Using a 5% level of significance, find the critical region of $\bar{X}$ for this test.
State your hypotheses clearly.
(4)

The Director is concerned about the machine filling the cartons with more than 330 ml of juice as well as less than 330 ml of juice. The Director takes a sample of 55 cartons, records the mean amount of juice $\bar{y}$ ml and uses a test with a critical region of N

$$\{\bar{Y} < 328\} \cup \{\bar{Y} > 332\}$$

(b) Find P(Type I error) for the Director's test.
(2)

When $\mu = 325$ ml

(c) find P(Type II error) for the test in part (a)
(2)

(a) Hypotheses $X \to$ amount of juice in a carton define variable

$H_0: \mu = 330$ $X \sim N(330, 8^2)$

$H_1: \mu < 330$ B1 $\bar{X} \approx \sim N(330, (\frac{8}{\sqrt{25}})^2)$ we are told he takes a sample $\therefore$ use sample mean

M1

critical value

$P(\bar{X} < C) = 0.05$ M1

$C = \text{InvN}(0.05)$ with parameters $\mu = 330$, $\sigma = \frac{8}{5}$

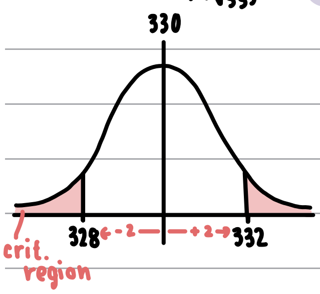
$C = 327.368$

$\therefore$ Critical Region: $\bar{X} < 327$ A1

(b) Type I Error is the probability of falsefully rejecting H_0. (H_0 is true but we reject it)

It is just the actual significance of the test.

$\bar{Y} \approx \sim N(330, (\frac{8}{\sqrt{55}})^2)$ M1



By symmetry, we see that $P(\bar{Y} < 328) = P(\bar{Y} > 332)$.

$\therefore P(\text{Type I}) = 2 \times P(\bar{Y} < 328)$

$= 0.063732 \to$ 0.0637 to 3sf type I error

A1

P 7 4 0 8 0 A 0 1 4 2 4

Question 5 continued

(c) Type II Error is the probability of wrongfully accepting H_0. (H_0 is false but we accept it)

test in part(a) → $\bar{X} \sim N(330, (\frac{8}{\sqrt{25}})^2)$

given correct value

$\therefore$ P(Type II Error) = $P(\bar{X} > 327.368 \mid \mu = 325)$ = 0.0694233 M1

crit. region found in (a)

↳ 0.0694 to 3sf type II error

A1

Question 5 continued

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 8 marks)

6. The discrete random variable X has probability generating function

$$G_X(t) = \frac{t^2}{(3-2t)^2}$$

(a) Specify the distribution of X

(2)

A fair die is rolled repeatedly.

(b) Describe an outcome that could be modelled by the random variable X

(1)

(c) Use calculus and $G_X(t)$ to find

(i) $E(X)$

(ii) $Var(X)$

(7)

The discrete random variable Y has probability generating function

$$G_Y(t) = \frac{t^{10}}{(3-2t^3)^2}$$

(d) Find the exact value of $P(Y = 19)$

(3)

(a) Negative Binomial, as NegB(r, p) has probability generating function $\left[\frac{pt}{1-t(1-p)}\right]^r$. M1

$$\left[\frac{pt}{1-t(1-p)}\right]^r \longrightarrow \frac{t^2}{(3-2t)^2}$$

By observation → $r=2$

Compare denominators: $(1-t(1-p))^2 = (3-2t)^2 \underset{\text{factorize}}{=} (\tfrac{1}{3}(1-\underbrace{\tfrac{2}{3}}_{1-p}t))^2$

$$1-p=\frac{2}{3} \rightarrow p=\frac{1}{3}$$

∴ distribution NegB$(2, \frac{1}{3})$ A1

(b) # of rolls to achieve a $\underbrace{5 \text{ or } 6}_{2 \text{ outcomes, so } p = 2\times\frac{1}{6}=\frac{1}{3}!}$ twice B1

P 7 4 0 8 0 A 0 1 8 2 4

DO NOT WRITE IN THIS AREA

Question 6 continued

(c) $Var(X) = G_X''(1) + G_X'(1) - [G_X'(1)]^2$

$E(X) = G_X'(1)$

$G_X(t) = t^2(3-2t)^{-2}$

i. Differentiate using product rule and chain rule

$u = t^2 \xrightarrow{\frac{du}{dt}} u' = 2t$

$v = (3-2t)^{-2} \xrightarrow{\frac{dv}{dt}} v' = -2(3-2t)^{-3} \times -2 = 4(3-2t)^{-3}$

substitute:

$G_X'(t) = uv' + vu'$

$= 4t^2(3-2t)^{-3} + 2t(3-2t)^{-2}$ M1A1

$G_X'(1) = 4(1)^2(3-2(1))^{-3} + 2(1)(3-2(1))^{-2}$

$= 4(3-2)^{-3} + 2(3-2)^{-2}$

$= 4 + 2$

$E(X) = G_X'(1) = 6$ A1

★Product Rule:

$\frac{d}{dx}(uv) = uv' + u'v$

★chain rule: multiply by the derivative of the bracket!

ii. We need to differentiate again

$G_X'(t) = 4t^2(3-2t)^{-3} + 2t(3-2t)^{-2}$

$u = 4t^2 \quad u' = 8t \qquad u = 2t \quad u' = 2$

$v = (3-2t)^{-3} \quad v' = 6(3-2t)^{-4} \qquad v = (3-2t)^{-2} \quad v' = 4(3-2t)^{-3}$

$\therefore G_X''(t) = 24t^2(3-2t)^{-4} + 8t(3-2t)^{-3} + 2(3-2t)^{-2} + 8t(3-2t)^{-3}$ M1

$= 24t^2(3-2t)^{-4} + 16t(3-2t)^{-3} + 2(3-2t)^{-2}$

$G_X''(1) = 24(1)^2(\cancel{3-2(1)})^{-4} + 16(1)(\cancel{3-2(1)})^{-3} + 2(\cancel{3-2(1)})^{-2}$

$= 24 + 16 + 2$

$G_X''(1) = 42$ A1 + from i $G_X'(t) = 6$

Substitute into $Var(X) = G_X''(1) + G_X'(1) - [G_X'(1)]^2$:

$Var(X) = 42 + 6 - (6)^2$ M1

$= 48 - 36$

$Var(X) = 12$ A1

(d) $G_Y(t) = t^{10} \times \frac{1}{9}\left[1 - \frac{2}{3}t^3\right]^{-2}$ M1

we want t^{19} (P(Y=19)). t^{10} is up front, ∴ we need t^9 in the bracket.

A1 $= \frac{t^{10}}{9}\left[1 + \ldots \frac{(-2)(-3)(-4)}{3!}\left(-\frac{2}{3}t^3\right)^3 \ldots\right]$

In our pgf, $t = t^3$ ∴ we need the third term of the binomial expansion $(t^3)^3 = t^9$

$= \frac{t^{10}}{9} \times \frac{-24}{6}\left(-\frac{8}{27}t^9\right)$

$= \frac{t^{10}}{9} \times \frac{32}{27}t^9$

$= \frac{32}{243}t^{19}$

$\therefore P(Y=19) = \frac{32}{243}$ A1

Question 6 continued

Question 6 continued

(Total for Question 6 is 13 marks)

7. Each time a spinner is spun, the probability that it lands on red is 0.2

(a) Find the probability that the spinner lands on red

(i) for the 1st time on the 4th spin **(2)**

(ii) for the 3rd time on the 8th spin **(2)**

(iii) exactly 4 times during 10 spins **(2)**

Each time the spinner is spun, the probability that it lands on yellow is 0.4

In a game with this spinner, a player must choose one of two events

R is the event that the spinner lands on **red** for the **1st** time in at most 4 spins

Y is the event that the spinner lands on **yellow** for the **3rd** time in at most 7 spins

(b) Showing your calculations clearly, determine which of these events has the greater probability. **(7)**

(a) i. Use geometric distribution.

$X \sim Geo(0.2)$

$P(X=4) = 0.8^3 \times 0.2$

$= 0.1024$ M1A1

ii. Use Negative Binomial Distribution

$T \sim NegB(3, 0.2)$

$P(T=8) = \binom{7}{2}(0.2)^3 \times (0.8)^5$

$= 0.05505 \rightarrow 0.0551$ to 3sf M1A1

iii. Use binomial distribution

$F \sim B(10, 0.2)$

$P(F=4) = 0.088080 \longrightarrow 0.0881$ to 3sf M1A1

(b) For Red use geometric distribution

$R \sim Geo(0.2)$

$P(R \leq 4) = 1 - P(X > 4)$

$= 1 - 0.8^4 = 0.5904 = P(R)$ M1M1A1

For Yellow use Negative Binomial distribution

$Y \sim NegB(3, 0.4)$ M1

$P(Y \leq 7) = 1 - \left[\binom{7}{2}0.4^2 \times 0.6^5 + \binom{7}{1}0.4^1 \times 0.6^6 + \binom{7}{0}0.6^7\right]$ M1

we are subtracting the cases where we don't succeed (less than 3 successes in 7 tries)

$= 0.580096 = P(Y)$ A1

Compare Probabilities:

$0.5904 > 0.580096$

$P(R) > P(Y)$

Red has higher probability A1

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

Question 7 continued

Question 7 continued

(Total for Question 7 is 13 marks)

TOTAL FOR PAPER IS 75 MARKS