Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

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 permitted 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 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.
- There are 7 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 ▶

P69194A

©2022 Pearson Education Ltd.

Q:1/1/1/1/1/

P 6 9 1 9 4 A 0 1 2 4

Pearson

1. A researcher is investigating the number of female cubs present in litters of size 4
He believes that the number of female cubs in a litter can be modelled by B(4, 0.5)
He randomly selects 100 litters each of size 4 and records the number of female cubs.
The results are recorded in the table below.

Number of female cubs	0	1	2	3	4
Observed number of litters	10	33	33	15	9

He calculated the expected frequencies as follows

Number of female cubs	0	1	2	3	4
Expected number of litters	6.25	r	s	r	6.25

(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 number of female cubs in a litter can be modelled by B(4, 0.5)
You should clearly state your hypotheses and the critical value used.

(6)

(a) $P(X = 1) = {}^4C_1(0.5)^4 = 0.25 \Rightarrow r = 100 \times 0.25 = 25$

$P(X = 2) = {}^4C_2(0.5)^4 = 0.375 \Rightarrow s = 100 \times 0.375 = 37.5$

$r = 25, \quad s = 37.5$

(b) H_0: B(4, 0.5) is a suitable model for the number of female cubs

H_1: B(4, 0.5) is not a suitable model for the number of female cubs

O_i	10	33	33	15	9
E_i	6.25	25	37.5	25	6.25
$(O_i - E_i)^2 \div E_i$	2.25	2.56	0.54	4	1.21

All $E_i \geq 5$, so no classes need to be combined

$X^2 = \Sigma(O_i - E_i)^2 \div E_i = 2.25 + 2.56 + 0.54 + 4 + 1.21 = 10.56$

$\nu = 5 - 1 = 4$ (no parameters estimated)

Critical value $\chi^2_4(5\%) = 9.488$

$10.56 > 9.488$, so significant: reject H_0

There is evidence that the researcher's model B(4, 0.5) is not suitable for the number of female cubs in a litter.

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

Question 1 continued

(Total for Question 1 is 9 marks)

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

2. The discrete random variable X has probability distribution

x	–5	–1	0	5	b
$P(X = x)$	0.3	0.25	0.1	0.15	0.2

where b is a constant and $b > 5$

(a) Find $E(X)$ in terms of b

(1)

Given that $Var(X) = 34.26$

(b) find the value of b

(4)

(c) Find $P(X^2 < 2 - 3X)$

(4)

(a) $E(X) = -5(0.3) - 1(0.25) + 0(0.1) + 5(0.15) + 0.2b$

$E(X) = 0.2b - 1$

(b) $E(X^2) = 25(0.3) + 1(0.25) + 0(0.1) + 25(0.15) + 0.2b^2 = 11.5 + 0.2b^2$

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

$11.5 + 0.2b^2 - (0.2b - 1)^2 = 34.26$

$11.5 + 0.2b^2 - 0.04b^2 + 0.4b - 1 = 34.26$

$0.16b^2 + 0.4b - 23.76 = 0$

$b^2 + 2.5b - 148.5 = 0$

$(b - 11)(b + 13.5) = 0 \Rightarrow b = 11$ or $b = -13.5$

$b > 5$, so $b = 11$

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

Question 2 continued

(c) $X^2 < 2 - 3X \Leftrightarrow X^2 + 3X - 2 < 0$

x	-5	-1	0	5	11
x^2	25	1	0	25	121
$2 - 3x$	17	5	2	-13	-31
$x^2 < 2 - 3x$?	no	yes	yes	no	no

($x^2 + 3x - 2 = 0$ gives $x = -3.56$ or $x = 0.562$, so only $x = -1$ and $x = 0$ work)

$P(X^2 < 2 - 3X) = P(X = -1) + P(X = 0)$

$= 0.25 + 0.1$

$= 0.35$

(Total for Question 2 is 9 marks)

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

3. During the summer, mountain rescue team A receives calls for help randomly with a rate of 0.4 per day.

(a) Find the probability that during the summer, mountain rescue team A receives at least 19 calls for help in 28 randomly selected days.

(2)

The leader of mountain rescue team A randomly selects 250 summer days from the last few years.
She records the number of calls for help received on each of these days.

(b) Using a Poisson approximation, estimate the probability of the leader finding at least 20 of these days when more than 1 call for help was received by mountain rescue team A.

(4)

Mountain rescue team A believes that the number of calls for help per day is lower in the winter than in the summer. The number of calls for help received in 42 randomly selected winter days is 8

(c) Use a suitable test, at the 5% level of significance, to assess whether or not there is evidence that the number of calls for help per day is lower in the winter than in the summer. State your hypotheses clearly.

(4)

During the summer, mountain rescue team B receives calls for help randomly with a rate of 0.2 per day, independently of calls to mountain rescue team A.

The random variable C is the total number of calls for help received by mountain rescue teams A and B during a period of n days in the summer.
On a Monday in the summer, mountain rescue teams A and B each receive a call for help.

Given that over the next n days $P(C = 0) < 0.001$

(d) calculate the minimum value of n

(3)

(e) Write down an assumption that needs to be made for the model to be appropriate.

(1)

(a) W = number of calls in 28 days, $W \sim Po(0.4 \times 28) = Po(11.2)$

$P(W \geqslant 19) = 1 - P(W \leqslant 18) = 1 - 0.97922... = 0.020776... \approx 0.0208$

(b) S = number of calls in a day, $S \sim Po(0.4)$

$P(S > 1) = 1 - P(S = 0) - P(S = 1) = 1 - e^{-0.4}(1 + 0.4) = 0.061551...$

D = number of days with more than 1 call, $D \sim B(250, 0.061551...)$

n large, p small, so $D \approx Po(250 \times 0.061551...) = Po(15.3879...)$

$P(D \geqslant 20) \approx 1 - P(D \leqslant 19) = 1 - 0.85248... = 0.14751...$

$= 0.148$

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

Question 3 continued

(c) λ = mean number of calls in 42 winter days

H_0: $\lambda = 16.8$ H_1: $\lambda < 16.8$

Under H_0, $U \sim Po(0.4 \times 42) = Po(16.8)$

$P(U \leqslant 8) = 0.014106... = 0.0141$

$0.0141 < 0.05$, so significant: reject H_0

There is sufficient evidence at the 5% level that the number of calls for help per day is lower in winter than in summer.

(d) Calls occur independently, so the calls on the Monday do not affect the next n days. $C \sim Po(0.4n + 0.2n) = Po(0.6n)$

$P(C = 0) = e^{-0.6n} < 0.001$

$-0.6n < \ln 0.001 \Rightarrow n > (\ln 1000) \div 0.6 = 11.51...$

Minimum value of $n = 12$

(e) Calls for help occur independently of each other and at a constant rate (the rate of calls per day is constant through the summer).

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

Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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 3 continued

(Total for Question 3 is 14 marks)

4. In a game a spinner is spun repeatedly. When the spinner is spun, the probability of it landing on blue is 0.11

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

(i) for the first time on the 6th spin,

(2)

(ii) for the first time before the 6th spin,

(2)

(iii) exactly 4 times during the first 6 spins,

(2)

(iv) for the 4th time on or before the 6th spin.

(4)

Zac and Izana play the game. They take turns to spin the spinner. The winner is the first one to have the spinner land on blue. Izana spins the spinner first.

(b) Show that the probability of Zac winning is 0.471 to 3 significant figures.

(3)

(a)(i) W = spin on which the first blue occurs, $W \sim Geo(0.11)$

$P(W = 6) = (0.89)^5(0.11) = 0.061424... = 0.0614$

(ii) $P(W < 6) = P(W \leq 5) = 1 - (0.89)^5$

$= 0.44159... = 0.442$

(iii) X = number of blues in 6 spins, $X \sim B(6, 0.11)$

$P(X = 4) = {}^6C_4(0.11)^4(0.89)^2 = 0.0017395... = 0.00174$

(iv) Y = spin on which the 4th blue occurs, $Y \sim NB(4, 0.11)$

$P(Y \leq 6) = P(Y = 4) + P(Y = 5) + P(Y = 6)$

$= (0.11)^4 + {}^4C_3(0.11)^4(0.89) + {}^5C_3(0.11)^4(0.89)^2$

$= 0.00014641 + 0.00052122 + 0.00115971$

$= 0.0018273... = 0.00183$

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
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 4 continued

(b) Izana spins 1st, 3rd, 5th, ... and Zac spins 2nd, 4th, 6th, ...

$P(\text{Zac wins}) = 0.89 \times 0.11 + (0.89)^3 \times 0.11 + (0.89)^5 \times 0.11 + \ldots$

Geometric series with $a = 0.89 \times 0.11 = 0.0979$, $r = (0.89)^2 = 0.7921$

$$P(\text{Zac wins}) = \frac{a}{1-r} = \frac{0.89 \times 0.11}{1-(0.89)^2} = \frac{0.0979}{0.2079}$$

$= 0.470899\ldots = 0.471$ (3 s.f.)

Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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 4 continued

(Total for Question 4 is 13 marks)

5. A random sample of 150 observations is taken from a geometric distribution with parameter 0.3

Estimate the probability that the mean of the sample is less than 3.45

(5)

$X \sim Geo(0.3)$: $E(X) = \frac{1}{0.3} = \frac{10}{3}$

$Var(X) = \frac{1 - 0.3}{0.3^2} = \frac{0.7}{0.09} = \frac{70}{9}$

$n = 150$ is large, so by the Central Limit Theorem

$\bar{X} \approx N\left(\frac{10}{3}, \frac{70/9}{150}\right) = N\left(\frac{10}{3}, \frac{7}{135}\right)$

$P(\bar{X} < 3.45) = P\left(Z < \frac{3.45 - 10/3}{\sqrt{(7/135)}}\right)$

$= P(Z < 0.5123...)$

$= 0.69579...$

$= 0.696$

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 5 marks)

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

6. The discrete random variable V has probability distribution

v	2	3	4
$P(V = v)$	$\frac{9}{25}$	$\frac{12}{25}$	$\frac{4}{25}$

(a) Show that the probability generating function of V is

$$G_V(t) = t^2\left(\frac{2}{5}t + \frac{3}{5}\right)^2$$

(2)

The discrete random variable W has probability generating function

$$G_W(t) = t\left(\frac{2}{5}t + \frac{3}{5}\right)^5$$

(b) Use calculus to find

(i) $E(W)$

(2)

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

(4)

Given that V and W are independent,

(c) find the probability generating function of $X = V + W$ in its simplest form.

(2)

The discrete random variable $Y = 2X + 3$

(d) Find the probability generating function of Y

(2)

(e) Find $P(Y = 15)$

(2)

(a) $G_V(t) = E(t^V) = \frac{9}{25}t^2 + \frac{12}{25}t^3 + \frac{4}{25}t^4$

$= t^2\left(\frac{9}{25} + \frac{12}{25}t + \frac{4}{25}t^2\right)$

$= t^2\left(\frac{2}{5}t + \frac{3}{5}\right)^2$

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
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 6 continued

(b)(i) $G_W'(t) = \left(\frac{2}{5}t + \frac{3}{5}\right)^5 + t \times 5 \times \frac{2}{5}\left(\frac{2}{5}t + \frac{3}{5}\right)^4$

$= \left(\frac{2}{5}t + \frac{3}{5}\right)^5 + 2t\left(\frac{2}{5}t + \frac{3}{5}\right)^4$

$E(W) = G_W'(1) = 1^5 + 2(1)^4 = 3$

(ii) $G_W''(t) = 2\left(\frac{2}{5}t + \frac{3}{5}\right)^4 + 2\left(\frac{2}{5}t + \frac{3}{5}\right)^4 + \frac{16}{5}t\left(\frac{2}{5}t + \frac{3}{5}\right)^3$

$G_W''(1) = 2 + 2 + \frac{16}{5} = \frac{36}{5}$

$Var(W) = G_W''(1) + G_W'(1) - [G_W'(1)]^2$

$= \frac{36}{5} + 3 - 3^2 = \frac{6}{5}$

(c) V, W independent: $G_X(t) = G_V(t) \times G_W(t)$

$= t^2\left(\frac{2}{5}t + \frac{3}{5}\right)^2 \times t\left(\frac{2}{5}t + \frac{3}{5}\right)^5$

$= t^3\left(\frac{2}{5}t + \frac{3}{5}\right)^7$

Question 6 continued

(d) $G_Y(t) = E(t^{2X+3}) = t^3E[(t^2)^X] = t^3G_X(t^2)$

$= t^3 \times (t^2)^3\left(\frac{2}{5}t^2 + \frac{3}{5}\right)^7$

$= t^9\left(\frac{2}{5}t^2 + \frac{3}{5}\right)^7$

(e) $P(Y = 15)$ = coefficient of t^{15} in $G_Y(t)$, so we need the t^6 term of $(0.4t^2 + 0.6)^7$, i.e. the $(t^2)^3$ term:

${}^7C_3\left(\frac{2}{5}t^2\right)^3\left(\frac{3}{5}\right)^4$

$P(Y = 15) = 35 \times \frac{8}{125} \times \frac{81}{625} = \frac{22680}{78125}$

$= \frac{4536}{15625} = 0.290304$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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 6 continued

(Total for Question 6 is 14 marks)

7. A machine fills bags with flour. The weight of flour delivered by the machine into a bag, X grams, is normally distributed with mean μ grams and standard deviation 30 grams. To check if there is any change to the mean weight of flour delivered by the machine into each bag, Olaf takes a random sample of 10 bags. The weight of flour, x grams, in each bag is recorded and $\bar{x} = 1020$

(a) Test, at the 5% level of significance, $H_0 : \mu = 1000$ against $H_1 : \mu \neq 1000$

(4)

Olaf decides to alter the test so that the hypotheses are $H_0 : \mu = 1000$ and $H_1 : \mu > 1000$ but keeps the level of significance at 5%

He takes a second sample of size n and finds the critical region, $\bar{X} > c$

(b) Find an equation for c in terms of n

(2)

When the true value of μ is 1020 grams, the probability of making a Type II error is 0.0050, to 2 significant figures.

(c) Calculate the value of n and the value of c

(5)

(a) Under H_0, $\bar{X} \sim N\left(1000, \frac{30^2}{10}\right) = N(1000, 90)$

$P(\bar{X} \geqslant 1020) = P\left(Z \geqslant \frac{1020 - 1000}{\sqrt{90}}\right) = P(Z \geqslant 2.108\ldots)$

$= 0.017507\ldots = 0.0175$

Two-tailed test: $0.0175 < 0.025$, so significant: reject H_0

There is evidence that the mean weight of flour in a bag is not 1000 g (there is evidence of a change in the mean weight).

(b) Under H_0, $\bar{X} \sim N\left(1000, \frac{900}{n}\right)$

$P(\bar{X} > c) = 0.05$: $\frac{c - 1000}{30/\sqrt{n}} = 1.6449$

$c = 1000 + \frac{49.347}{\sqrt{n}}$

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
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 7 continued

(c) Type II error: accept H_0 when $\mu = 1020$, so $P(\bar{X} < c \mid \mu = 1020) = 0.0050$

$\bar{X} \sim N\left(1020, \frac{900}{n}\right)$: $\frac{c - 1020}{30/\sqrt{n}} = -2.5758$

$c = 1020 - \frac{77.274}{\sqrt{n}}$

$1000 + \frac{49.347}{\sqrt{n}} = 1020 - \frac{77.274}{\sqrt{n}}$

$\frac{126.621}{\sqrt{n}} = 20 \Rightarrow \sqrt{n} = 6.3310\ldots$

$n = 40.08\ldots$ so $n = 40$

$c = 1000 + \frac{49.347}{\sqrt{40}} = 1007.80\ldots$

$c = 1007.8$ (1010 to 3 s.f.)

Question 7 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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 7 continued

Question 7 continued

(Total for Question 7 is 11 marks)

TOTAL FOR PAPER IS 75 MARKS

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