Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Friday 16 October 2020

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 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 ▶

©2020 Pearson Education Ltd.

Pearson

1. The number of customers entering Jeff's supermarket each morning follows a Poisson distribution.

Past information shows that customers enter at an average rate of 2 every 5 minutes.

Using this information,

(a) (i) find the probability that exactly 26 customers enter Jeff's supermarket during a randomly selected 1-hour period one morning,

(2)

(ii) find the probability that at least 21 customers enter Jeff's supermarket during a randomly selected 1-hour period one morning.

(2)

A rival supermarket is opened nearby. Following its opening, the number of customers entering Jeff's supermarket over a randomly selected 40-minute period is found to be 10

0.05

(b) Test, at the 5% significance level, whether or not there is evidence of a decrease in the rate of customers entering Jeff's supermarket. State your hypotheses clearly.

(4)

A further randomly selected 20-minute period is observed and the hypothesis test is repeated. Given that the true rate of customers entering Jeff's supermarket is now 1 every 5 minutes,

(c) calculate the probability of a Type II error.

(5)

(a) $\lambda = 2$ every 5 mins

λ : Time (mins)

$\times 12$ 2 : 5 $\times 12$

24 : 60

$\therefore \lambda = 24$ every 1-hour period

Let X = no. of customers

$\therefore X \sim Po(24)$

(i) $P(X=26) = 0.071912\ldots \approx 0.0719$ (4dp)

$\boxed{\therefore P(X=26) = 0.0719}$

(ii) $P(X \geq 21) = 1 - P(X \leq 20) = 1 - 0.24263\ldots = 0.75736\ldots \approx 0.7574$ (4dp)

$\boxed{\therefore P(X \geq 21) = 0.7574}$

(b) λ : Time (mins)

$\times 8$ 2 : 5 $\times 8$

16 : 40

$\therefore \lambda = 16$ every 40 mins

$\therefore Y \sim Po(16)$

P 6 2 6 7 7 A 0 2 2 8

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
DO NOT WRITE IN THIS AREA

Question 1 continued

(b) Continued

$H_0: \lambda = 16$
$H_1: \lambda < 16$

$P(Y \leq 10 \mid Y \sim Po(16)) = 0.077396... \approx \boxed{0.0774 > 0.05}$ (4dp)

Given that

$\therefore$ There is not enough evidence to suggest a decrease in the rate of customers entering Jeff's supermarket.
$\therefore$ Accept H_0. $\therefore \lambda = 16$ every 40 mins.

(c.) $\lambda = 1$ every 5 mins

λ : Time (mins)
$\times 4$ 1 : 5 $\times 4$
4 : 20

λ : Time (mins)
$\times 4$ 2 : 5 $\times 4$
8 : 20

$\therefore \lambda = 4$ every 20 mins
$\therefore W \sim Po(4)$
True Distribution

$\therefore \lambda = 8$ every 20 mins
$\therefore Z \sim Po(8)$

Using $Z \sim Po(8)$ to find critical region:
$P(Z \leq 2) = 0.01375... < 0.05$
$P(Z \leq 3) = 0.04238... < 0.05$
$P(Z \leq 4) = 0.09963... > 0.05$
$\therefore$ Critical Region is $Z \leq 3$.

Type II Error requires the probabilities not in the critical region.
$\therefore H_0$ is not rejected when $Z \geq 4$.

$P(W \geq 4 \mid W \sim Po(4)) = 1 - P(X \leq 3) = 1 - 0.43347... = 0.56652...$

$\therefore P(\text{Type II Error}) = 0.5665$ (4dp)

Question 1 continued

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 13 marks)

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

2. The discrete random variables W, X and Y are distributed as follows

$$W \sim B(10, 0.4) \qquad X \sim Po(4) \qquad Y \sim Po(3)$$

(a) Explain whether or not Po(4) would be a good approximation to B(10, 0.4)

(1)

(b) State the assumption required for $X + Y$ to be distributed as Po(7)

(1)

Given the assumption in part (b) holds,

(c) find $P(X + Y < \mathrm{Var}(W))$

(2)

(a.) For poisson distribution to be a good approximation into binomial distribution, n must be large and p must be small.

$\therefore$ Po (4) $\approx$ B(10, 0.4) is not a good approximation.

(b.) X and Y must be independent for X + Y to be distributed as Po (7).

(c.) For binomial distribution $Var(X) = np(1-p)$

$Var(W) = 10(0.4)(1-0.4) = 2.4$

let $X + Y = Z$. $\therefore Z \sim Po(7)$

$P(X+Y < Var(W)) = P(X+Y < 2.4)$
$= P(Z \leq 2)$
$= 0.02963...$
≈ 0.0296 (4dp)

$\boxed{\therefore P(X+Y < Var(W)) = 0.0296}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 2 continued

(Total for Question 2 is 4 marks)

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

3. Suzanne and Jon are playing a game.

They put 4 red counters and 1 blue counter in a bag.

Suzanne reaches into the bag and selects one of the counters at random. If the counter she selects is blue, she wins the game. Otherwise she puts it back in the bag and Jon selects one at random. If the counter he selects is blue, he wins the game. Otherwise he puts it back in the bag and they repeat this process until one of them selects the blue counter.

(a) Find the probability that Suzanne selects the blue counter on her 4th selection.

(2)

(b) Find the probability that the blue counter is first selected on or after Jon's third selection.

(2)

(c) Find the mean and standard deviation of the number of selections made until the blue counter is selected.

(2)

(d) Find the probability that Suzanne wins the game.

(3)

(a.) Total counters = 4 + 1 = 5

Probability of a blue counter = $\frac{1}{5} = 0.2$

Let X = no. of selection. $\therefore X \sim \text{Geo}(0.2)$ Geometric Distribution

Suzanne's 4th selection is the 7th selection overall.

$P(X=7) = (1-0.2)^6(0.2) = 0.05242\ldots \approx 0.0524$ (4dp)

$\boxed{\therefore P(X=7) = 0.0524}$

(b) Jon's 3rd selection is the 6th selection overall.

$P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.67232 = 0.32768 \approx 0.3277$ (4dp)

$\boxed{\therefore P(X \geq 6) = 0.3277}$

P 6 2 6 7 7 A 0 8 2 8

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

Question 3 continued

(c) For geometric distribution: Mean, $\mu = \frac{1}{p}$ & Variance, $\sigma^2 = \frac{1-p}{p^2}$

$\mu = \frac{1}{0.2} = 5$

$\sigma^2 = \frac{1-0.2}{0.2^2} = 20$

Standard Deviation, $\sigma = \sqrt{20} = 2\sqrt{5} \approx 4.47$ (3sf)

$\boxed{\therefore \mu = 5, \ \sigma = 2\sqrt{5}}$

(d) Geometric Series: $S_\infty = \frac{a}{1-r}$

P(Suzanne Wins) $= 0.2 + (0.8)^2(0.2) + (0.8)^4(0.2) + \ldots$

$\times 0.8^2$ $\times 0.8^2$

$\therefore a = 0.2, \ r = 0.8^2$

$S_\infty = \frac{0.2}{1-0.8^2} = \frac{5}{9}$

$\boxed{\therefore P(\text{Suzanne wins}) = \frac{5}{9}}$

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 9 marks)

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

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

x	-5	-2	3	4
$\mathrm{P}(X = x)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$

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

(3)

The discrete random variable Y is defined in terms of the discrete random variable X

When X is negative, $Y = X^2$

When X is positive, $Y = 3X - 2$

(b) Find $\mathrm{P}(Y < 9)$

(3)

(c) Find $\mathrm{E}(XY)$

(2)

(a.) $\mathrm{Var}(X) = E(X^2) - E(X)^2$

$E(X) = -5\left(\frac{1}{12}\right) - 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{4}\right) + 4\left(\frac{1}{2}\right) = 2$

$E(X^2) = (-5)^2\left(\frac{1}{12}\right) + (-2)^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{4}\right) + 4^2\left(\frac{1}{2}\right) = 13$

$\boxed{\therefore \mathrm{Var}(X) = 13 - 2^2 = 9}$

(b.)

x	-5	-2	3	4
y	25	4	7	10

$\therefore P(Y<9) = P(X=-2) + P(X=3) = \frac{1}{6} + \frac{1}{4} = \boxed{\frac{5}{12}}$

(c.) $E(XY) = (-5)(25)\left(\frac{1}{12}\right) + (-2)(4)\left(\frac{1}{6}\right) + (3)(7)\left(\frac{1}{4}\right) + (4)(10)\left(\frac{1}{2}\right) = \frac{27}{2}$

$\boxed{\therefore E(XY) = \frac{27}{2} = 13.5}$

P 6 2 6 7 7 A 0 1 2 2 8

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

Question 4 continued

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

Question 4 continued

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 8 marks)

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

5. A factory produces pins.

An engineer selects 40 independent random samples of 6 pins produced at the factory and records the number of defective pins in each sample.

Number of defective pins	0	1	2	3	4	5	6
Observed frequency	19	11	7	2	0	1	0

(a) Show that the proportion of defective pins in the 40 samples is 0.15

(2)

The engineer suggests that the number of defective pins in a sample of 6 can be modelled using a binomial distribution. Using the information from the sample above, a test is to be carried out at the 10% significance level, to see whether the data are consistent with the engineer's suggested model.

The value of the test statistic for this test is 2.689

0.1

(b) Justifying the degrees of freedom used, carry out the test, at the 10% significance level, to see whether the data are consistent with the engineer's suggested model. State your hypotheses clearly.

(8)

The engineer later discovers that the previously recorded information was incorrect. The data should have been as follows.

Number of defective pins	0	1	2	3	4	5	6
Observed frequency	19	11	6	3	1	0	0

(c) Describe the effect this would have on the value of the test statistic that should be used for the hypothesis test.
Give reasons for your answer.

(3)

(a.) $p = \dfrac{0(19) + 1(11) + 2(7) + 3(2) + 4(0) + 5(1) + 6(0)}{6 \times 40} = \dfrac{3}{20} = 0.15$

$\boxed{\therefore \text{Proportion of Defective Pins} = 0.15}$

P 6 2 6 7 7 A 0 1 6 2 8

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

Question 5 continued

(b.) Let X = no. of defective pins in a sample of 6.

$\therefore X \sim B(6, 0.15)$

x	0	1	2	3	4	5	6	≥ 2
O_i	19	11	7	2	0	1	0	10
$P(X=x)$	0.377...	0.399...	0.176...	0.041...	0.0054...	0.000387...	0.00...	0.223...
$E_i = 40 \times P(X=x)$	15.08...	15.97...	7.04...	1.65...	0.218...	0.0154...	0.00...	8.94...

All are less than 5. Require $40 \times P(X \geq k) > 5$.

$\therefore$ Combine with $x = 2$.

$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.7764... = 0.2235...$

$\therefore 40 \times P(X \geq 2) = 0.2235... \times 40 = 8.94...$

$\therefore$ Degrees of Freedom, $\nu = 3 - 2 = 1$

H_0: Binomial Distribution is a suitable model.

H_1: Binomial Distribution is not a suitable model.

Critical Value: $\chi^2_1(10\%) = 2.705$

Test Statistic = 2.689

$2.689 < 2.705$

$\therefore$ Test Statistic is not in the critical region.

$\therefore$ Not enough evidence to reject H_0. Accept H_0.

$\therefore$ Data is consistent with the engineer's suggest model, binomial distribution.

(c.) Proportion of Defective pins remains the same as 0.15.

The cells for $X \geq 2$ are still combined in the test.

$\therefore$ There is no change to the value of the test statistic.

Question 5 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 5 continued

(Total for Question 5 is 13 marks)

6. A discrete random variable X has probability generating function given by

$$G_X(t) = \frac{1}{64}(a + bt^2)^2$$

where a and b are positive constants.

(a) Write down the value of $P(X = 3)$

(1)

Given that $P(X = 4) = \frac{25}{64}$

(b) (i) find $P(X = 2)$

(7)

(ii) find $E(X)$

(3)

The random variable $Y = 3X + 2$

(c) Find the probability generating function of Y

(2)

(a.) $P(X=3) = 0$

(b.) $P(X=4) = \frac{25}{64}$

(i.) Coefficient of $t^4 = \frac{1}{64}b^2$

$\frac{1}{64}b^2 = \frac{25}{64}$

$b^2 = 25$

$b = \pm\sqrt{25} = \pm 5$

As $b>0$, $b \neq -5$. $\therefore b = 5$

$G_X(1) = 1 \Rightarrow \frac{1}{64}(a + 5(1^2))^2 = 1$

$(a+5)^2 = 64$

$a + 5 = \pm 8$

$a = 8 - 5 = 3$ $\qquad$ $a = -8 - 5 = -13$

As $a>0$, $a \neq -13$. $\therefore a = 3$.

$P(X=2) =$ Coefficient of $t^2 = \frac{1}{64}(2ab) = \frac{1}{64} \times 2 \times 3 \times 5 = \frac{15}{32}$

$\boxed{\therefore P(X=2) = \frac{15}{32}}$

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

Question 6 continued

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

$G_X(t) = \frac{1}{64}(3 + 5t^2)^2$

$G'(t) = \frac{1}{64}(2)(3+5t^2)(10t) = \frac{5}{16}t(3+5t^2)$

$G'(1) = \frac{5}{16}(1)(3 + 5(1)^2) = \frac{5}{2} = 2.5$

$\boxed{\therefore E(X) = 2.5}$

(c.) $Y = 3X + 2$

$G_Y(t) = t^2 G_X(t^3) = \frac{t^2}{64}\left(3 + 5(t^3)^2\right)^2$

$\boxed{\therefore G_Y(t) = \frac{t^2}{64}(3+5t^6)^2}$

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

Question 6 continued

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 13 marks)

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

7. A six-sided die has sides labelled 1, 2, 3, 4, 5 and 6

The random variable S represents the score when the die is rolled.

Alicia rolls the die 45 times and the mean score, $\bar{S}$, is calculated.

Assuming the die is fair and using a suitable approximation,

(a) find, to 3 significant figures, the value of k such that $P(\bar{S} < k) = 0.05$ **(8)**

(b) Explain the relevance of the Central Limit Theorem in part (a). **(2)**

Alicia considers the following hypotheses:

H_0: The die is fair

H_1: The die is not fair

If $\bar{S} < 3.1$ or $\bar{S} > 3.9$, then H_0 will be rejected.

Given that the true distribution of S has mean 4 and variance 3

(c) find the power of this test. **(3)**

(d) Describe what would happen to the power of this test if Alicia were to increase the number of rolls of the die.
Give a reason for your answer. **(2)**

(a.) S has a discrete uniform distribution.

$E(S) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3.5$

$E(S^2) = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) = \frac{91}{6} = 15.1\dot{6}$

$Var(S) = E(S^2) - E(S)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$

$Var(\bar{S}) = \frac{\left(\frac{35}{12}\right)}{45} = \frac{7}{108}$

Normal Distribution

$\therefore \bar{S} \sim N(3.5, 0.06481) \rightarrow \mu = 3.5, \sigma = \sqrt{\frac{7}{108}}$

$P(\bar{S} < k) = 0.05 \rightarrow \frac{k-3.5}{\sqrt{\frac{7}{108}}} = -1.6449$

$z = 1.6449$ for $p = 0.05$.

$\therefore k = \left(-1.6449 \times \sqrt{\frac{7}{108}}\right) + 3.5 = 3.081\ldots \approx 3.08$ (3sf)

$\boxed{\therefore k = 3.08}$

P 6 2 6 7 7 A 0 2 4 2 8

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

(b.) CLT applies since the sample size is large.
CLT states that the sample mean, $\bar{S}$, is approximately normally distributed.

(c.) $\mu = 4,\ \sigma^2 = 3 \rightarrow \therefore \sigma = \sqrt{\frac{3}{45}} = \frac{1}{\sqrt{15}}$

$\therefore$ True $\bar{S} \sim N\left(4, \left(\frac{1}{\sqrt{15}}\right)^2\right)$

Power $= P(\bar{S} < 3.1) + P(\bar{S} > 3.9)$
$= 1 - P(3.1 < \bar{S} < 3.9)$
$= 1 - 0.34902\ldots$
$= 0.65097\ldots$
≈ 0.6510 (4dp)

$\boxed{\therefore \text{Power} = 0.6510}$

(d.) Increasing sample size would decrease the variance of $\bar{S}$, thus decreasing σ.
This leads to an increase in $P(\bar{S} > 3.9)$ & the decrease in $P(\bar{S} < 3.1)$ would be negligible.
$\therefore$ Power would increase.

P 6 2 6 7 7 A 0 2 5 2 8

Turn over

Question 7 continued

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

Question 7 continued

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

Question 7 continued

(Total for Question 7 is 15 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
P 6 2 6 7 7 A 0 2 8 2 8