Please check the examination details below before entering your candidate information

Candidate surname StemUp Model Solution | Other names

Pearson Edexcel Level 3 GCE

Centre Number | Candidate Number

Thursday 4 June 2020

Morning (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 ▶

P62677A

©2020 Pearson Education Ltd.

1/1/1/1/

P 6 2 6 7 7 A 0 1 2 8

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

(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)(i) $X \sim Po(24)$ (2 every 5 minutes $\Rightarrow$ 24 per hour)

$P(X = 26) = e^{-24} \times \frac{24^{26}}{26!} = 0.071912...$

$P(X = 26) = 0.0719$

(ii) $P(X \geqslant 21) = 1 - P(X \leqslant 20) = 1 - 0.24263...$

$= 0.75736...$ so $P(X \geqslant 21) = 0.757$

(b) Let Y be the number of customers in 40 minutes. Under H_0, $Y \sim Po(16)$

$H_0: \lambda = 16$ $H_1: \lambda < 16$ (λ = mean number per 40 minutes)

$P(Y \leqslant 10 \mid \lambda = 16) = 0.077396... = 0.0774$

$0.0774 > 0.05$, so not significant: do not reject H_0 (10 is not in the CR)

There is not sufficient evidence to suggest a decrease in the rate of customers entering Jeff's supermarket.

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

Question 1 continued

(c) Let T be the number of customers in 20 minutes. Under H_0, $T \sim Po(8)$

$P(T \leqslant 3) = 0.0424 < 0.05$ $P(T \leqslant 4) = 0.0996 > 0.05$

Critical region: $T \leqslant 3$ (H_0 is not rejected when $T \geqslant 4$)

True rate 1 every 5 minutes $\Rightarrow$ in 20 minutes $T \sim Po(4)$

$P(\text{Type II error}) = P(T \geqslant 4 \mid \lambda = 4) = 1 - P(T \leqslant 3)$

$= 1 - 0.43347...$

$= 0.56652...$

$P(\text{Type II error}) = 0.567$

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

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

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

(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 $\text{P}(X + Y < \text{Var}(W))$

(2)

(a) No, Po(4) would not be a good approximation: it requires n large and p small, but here n = 10 is not large and p = 0.4 is not small.

(b) X and Y must be independent.

(c) $\text{Var}(W) = np(1 - p) = 10 \times 0.4 \times 0.6 = 2.4$

$X + Y \sim \text{Po}(7)$

$\text{P}(X + Y < 2.4) = \text{P}(X + Y \leq 2)$

$= e^{-7}\left(1 + 7 + \frac{7^2}{2}\right) = 0.029636...$

$\text{P}(X + Y < 2.4) = 0.0296$

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)

Let X = number of selections until the blue counter is selected: $X \sim Geo(0.2)$

(a) Suzanne's 4th selection is the 7th selection overall

$P(X = 7) = (0.8)^6(0.2) = 0.0524288 = 0.0524$

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

$P(X \geqslant 6) = (1 - 0.2)^5 = 0.8^5 = 0.32768 = 0.328$

(c) Mean $= \frac{1}{p} = \frac{1}{0.2} = 5$

Standard deviation $= \sqrt{\left(\frac{1-p}{p^2}\right)} = \sqrt{\left(\frac{0.8}{0.2^2}\right)} = \sqrt{20} = 4.472\ldots = 4.47$

(d) Suzanne wins if the blue counter is first selected on selection 1, 3, 5, ...

$P(\text{Suzanne wins}) = 0.2 + (0.8)^2(0.2) + (0.8)^4(0.2) + \ldots$

Infinite geometric series with $a = 0.2$, $r = 0.8^2 = 0.64$

$P(\text{Suzanne wins}) = \frac{0.2}{1 - 0.8^2} = \frac{0.2}{0.36} = \frac{5}{9}$

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

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

(a) Find 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 $P(Y < 9)$

(3)

(c) Find $E(XY)$

(2)

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

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

$Var(X) = E(X^2) - [E(X)]^2 = 13 - 2^2 = 9$

(b) X negative: $Y = X^2$; X positive: $Y = 3X - 2$

x	−5	−2	3	4
y	25	4	7	10
P	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$

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

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

$= -\frac{125}{12} - \frac{4}{3} + \frac{21}{4} + 20$

$E(XY) = 13.5$

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

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

(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) Total defective = $0 \times 19 + 1 \times 11 + 2 \times 7 + 3 \times 2 + 4 \times 0 + 5 \times 1 + 6 \times 0 = 36$

Total number of pins = $40 \times 6 = 240$

$p = \frac{36}{240} = 0.15$

(b) H_0: a binomial distribution is a suitable model for the number of defective pins

H_1: a binomial distribution is not a suitable model

Using $X \sim B(6, 0.15)$, expected frequency = $40 \times P(X = x)$:

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

Question 5 continued

x	0	1	2	3	4	5	6
E_i	15.09	15.97	7.047	1.658	0.219	0.015	0.000

Expected frequencies must be > 5, so combine the cells for $x \geq 2$:

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

x	0	1	≥ 2
O_i	19	11	10
E_i	15.09	15.97	8.94

This leaves 3 cells. 2 is subtracted as there are 2 restrictions: the totals agree and the proportion p was estimated from the data.

Degrees of freedom $\nu = 3 - 2 = 1$

Critical value $\chi^2_1(10\%) = 2.706$

Test statistic 2.689 < 2.706, so it is not in the critical region:

insufficient evidence to reject H_0

The data are consistent with the engineer's binomial model.

(c) New total defective = $1 \times 11 + 2 \times 6 + 3 \times 3 + 4 \times 1 = 11 + 12 + 9 + 4 = 36$,

so the proportion of defective pins is still 0.15 and the expected frequencies are unchanged.

The cells for $x \geq 2$ are still combined: $O(x \geq 2) = 6 + 3 + 1 = 10$ as before,

and $O(0) = 19$, $O(1) = 11$ are unchanged.

So there is no change to the value of the test statistic (still 2.689).

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

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 13 marks)

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

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) $P(X = 3) = 0$ (there is no term in t^3)

(b)(i) $G_X(t) = \frac{1}{64}(a^2 + 2abt^2 + b^2t^4)$

Coefficient of t^4: $\frac{b^2}{64} = \frac{25}{64} \Rightarrow b^2 = 25$

$b = 5$ (reject $b = -5$ since $b > 0$)

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

$a + 5 = \pm 8 \Rightarrow a = 3$ (reject $a = -13$ since $a > 0$)

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

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)(ii) $E(X) = G'_X(1)$

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

$G'_X(1) = \frac{2}{64}(3 + 5)(10) = \frac{160}{64} = 2.5$

$E(X) = 2.5$

(c) $G_Y(t) = E(t^{3X+2}) = t^2E((t^3)^X) = t^2G_X(t^3)$

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

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, $\overline{S}$, is calculated.

Assuming the die is fair and using a suitable approximation,

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

(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 $\overline{S} < 3.1$ or $\overline{S} > 3.9$, 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 over $\{1, 2, 3, 4, 5, 6\}$

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

$Var(S) = \frac{6^2 - 1}{12} = \frac{35}{12}$

By the CLT, $\overline{S} \sim N(3.5, \sigma^2)$ approximately, where

$\sigma^2 = Var(\overline{S}) = \frac{35}{12} \div 45 = \frac{7}{108} = 0.0648...$

$P(\overline{S} < k) = 0.05 \Rightarrow \frac{k - 3.5}{\sqrt{(7 \div 108)}} = -1.6449$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 7 continued

$k = 3.5 - 1.6449 \times 0.25458... = 3.08122...$

$k = 3.08$

(b) The sample size $n = 45$ is large, so the CLT applies.

The CLT states that the sample mean $\bar{S}$ is approximately normally distributed

(even though S is discrete uniform), so $\bar{S} \sim N(3.5, \frac{7}{108})$ can be used.

(c) True distribution: $\bar{S} \sim N(4, \frac{3}{45})$ i.e. $N(4, 0.0667)$

Power $= P(\bar{S} < 3.1) + P(\bar{S} > 3.9)$

$= P(Z < -3.4856...) + P(Z > -0.3872...)$

$= 0.000245... + 0.650732... = 0.65097...$

Power $= 0.651$

(d) Increasing the number of rolls n decreases $\text{Var}(\bar{S}) = \frac{\sigma^2}{n}$, so

$P(\bar{S} > 3.9)$ increases (and $P(\bar{S} < 3.1)$ remains negligible).

So the power of the test would increase.

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

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