



Oxford Cambridge and RSA

Monday 20 May 2019 – Afternoon

AS Level Further Mathematics B (MEI)

Y412/01 Statistics a

Time allowed: 1 hour 15 minutes



You must have:

- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You may use:

- a scientific or graphical calculator

Model
Answers

INSTRUCTIONS

- Use black ink. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.** If additional space is required, you should use the lined page(s) at the end of the Printed Answer Booklet. The question number(s) must be clearly shown.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION

- The total number of marks for this paper is **60**.
- The marks for each question are shown in brackets [].
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is used. You should communicate your method with correct reasoning.
- The Printed Answer Booklet consists of **12** pages. The Question Paper consists of **8** pages.

Answer **all** the questions.

- 1 The discrete random variable X has probability distribution defined by

$$P(X=r) = k(r^2 + 3r) \text{ for } r = 1, 2, 3, 4, 5, \text{ where } k \text{ is a constant.}$$

- (a) Complete the table below, using the copy in the Printed Answer Booklet giving the probabilities in terms of k . [1]

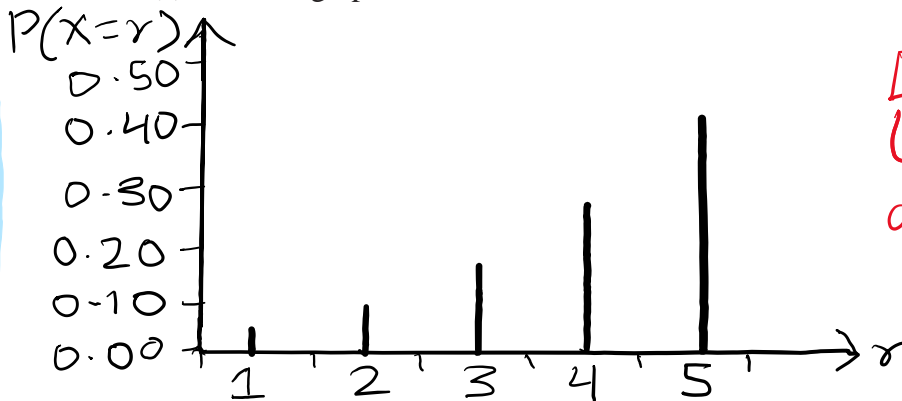
r	1	2	3	4	5
$P(X=r)$	$4k$	$10k$	$18k$	$28k$	$40k$

- (b) Show that the value of k is 0.01. [2]

$$\sum P(X=r) = 1 \text{ so } 4k + 10k + 18k + 28k + 40k = 1$$

$$100k = 1 \Rightarrow k = 0.01$$

- (c) Draw a graph to illustrate the distribution. [2]



Draw straight line graphs for discrete random variables.

- (d) Describe the shape of the distribution. [1]

Negative Skew

- (e) Find each of the following.

$$\begin{aligned} \bullet E(X) &= \sum r P(X=r) = (1 \times 0.04) + (2 \times 0.10) \\ &+ (3 \times 0.18) + (4 \times 0.28) + (5 \times 0.40) = 3.9 \end{aligned}$$

$$\begin{aligned} \bullet \text{Var}(X) \\ E(X^2) &= \sum r^2 P(X=r) = (1^2 \times 0.04) + (2^2 \times 0.10) \\ &+ (3^2 \times 0.18) + (4^2 \times 0.28) + (5^2 \times 0.40) = 16.54 \\ \text{Var} X &= E(X^2) - [E(X)]^2 = 16.54 - 3.9^2 \\ &= 1.33 \end{aligned} \quad [2]$$

3

- 2 Almost all plants of a particular species have red flowers. However on average 1 in every 1500 plants of this species have white flowers. A random sample of 2000 plants of this species is selected. The random variable X represents the number of plants in the sample that have white flowers.

- (a) Name two distributions which could be used to model the distribution of X , stating the parameters of each of these distributions. [4]

Binomial with $n=2000, p=\frac{1}{1500}$

Poisson with $\lambda=\frac{4}{3}$ ← $\lambda=np = \frac{2000}{1500} = \frac{4}{3}$

You may use either of the distributions you have named in the rest of this question.

- (b) Calculate each of the following. Using $X \sim Po\left(\frac{4}{3}\right)$

• $P(X=2) = 0.2343$

• $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.8494 = 0.1506$

[3]

- (c) A random sample of 20000 plants of this species is selected. Calculate the probability that there are at least 10 plants in the sample that have white flowers.

$X \sim B\left(20000, \frac{1}{1500}\right)$

$P(X \geq 10) = 1 - 0.1448 = 0.8552$

[2]

- 3 A fair 8-sided dice has faces labelled 10, 20, 30, ..., 80.

- (a) State the distribution of the score when the dice is rolled once. [2]

Uniform distribution over [10, 20, ..., 80]

- (b) Write down the probability that, when the dice is rolled once, the score is at least 40. [1]

$P(X \geq 40) = \frac{5}{8}$ ← 40, 50, 60, 70, 80

- (c) The dice is rolled three times.

- (i) Find the variance of the total score obtained.

Variance for 1 roll = $\frac{1}{12}(8^2 - 1) \times 10^2 = 525$

Variance for 3 rolls = $3 \times 525 = 1575$

- (ii) Find the probability that on one of the rolls the score is less than 30, on another it is between 30 and 50 inclusive and on the other it is greater than 50. [3]

$P(X < 30) = \frac{2}{8}$. $P(30 \leq X \leq 50) = \frac{3}{8}$. $P(X > 50) = \frac{3}{8}$

As 3 rolls, $P = \frac{2}{8} \times \frac{3}{8} \times \frac{3}{8} \times 6 = \frac{27}{128} = 0.2109$

- 4 A student is investigating correlations between various personality traits, two of which are conscientiousness and openness to new experiences. She selects a random sample of 10 students at her university and uses standard tests to measure their conscientiousness and their openness. The product moment correlation coefficient between these two variables for the 10 students is 0.476.
- (a) Assuming that the underlying population has a bivariate Normal distribution, carry out a hypothesis test at the 10% significance level to investigate whether there is any correlation between openness and conscientiousness in students. [5]

$H_0: \rho = 0$ and $H_1: \rho \neq 0$ where ρ is the population correlation coefficient between openness and conscientiousness. Critical value at 10% 2 tail and $n=10$ is 0.5494. As $0.476 < 0.5494$, this result is not significant so insufficient evidence to reject H_0 , which would suggest there is a correlation between openness and conscientiousness.

Table 4.1 below shows the values of the product moment correlation coefficients between 5 different personality traits for a much larger sample of students. Those correlations that are significant at the 5% level are denoted by a * after the value of the correlation.

	Neuroticism	Extroversion	Openness	Agreeableness	Conscientiousness
Neuroticism	1				
Extroversion	-0.296*	1			
Openness	-0.044	0.405*	1		
Agreeableness	-0.190*	0.061	0.042	1	
Conscientiousness	-0.485*	0.145	0.235*	0.112	1

Table 4.1

The student analyses these factors for effect size.

Guidelines often used when considering effect size are given in Table 4.2 below.

Product moment correlation coefficient	Effect size
0.1	Small
0.3	Medium
0.5	Large

Table 4.2

- (b) The student notes that, despite the result of the test in part (a), the correlation between openness and conscientiousness is significant at the 5% level with this second sample.

Comment briefly on why this may be the case.

[1]

For the second sample, the sample size was much larger. This would mean a smaller critical value which could explain the significant result with a lower test statistic.

- (c) The student intends to summarise her findings about relationships between these factors, including effect sizes, in a report.

Use the information in Tables 4.1 and 4.2 to identify two summary points the student could make.

[2]

- No correlations have large effect size
- 5 out of 10 correlations involving Neuroticism are found to be significant at a 5% level

- 5 A researcher is investigating births of females and males in a particular species of animal which very often produces litters of 7 offspring. The table shows some data about the number of females per litter in 200 litters of 7 offspring. The researcher thinks that a binomial distribution $B(7, p)$ may be an appropriate model for these data.

Number of females	Number of litters
0	20
1	45
2	44
3	37
4	33
5	14
6	5
7	2

- (a) Use the data to show that a suitable estimate for the value of p is 0.35.

[2]

$$\bar{x} = \frac{\sum np}{\sum n} = \frac{(0 \times 20) + (1 \times 45) + (2 \times 44) + \dots + (7 \times 2)}{200}$$

$$= 2.45.$$

$$\bar{x} = np \text{ so } 2.45 = 7p \Rightarrow p = 0.35$$

as required.

(b) In this question you must show detailed reasoning.

Complete the copy of the table of expected frequencies and contributions for a chi-squared test in the Printed Answer Booklet. (Note that the classes 5, 6 and 7 have been combined as >5). [4]

Number of females	Probability	Expected frequency	Chi-squared contribution
0	0.04902	9.8045	10.6022
1	0.18477	36.9553	1.7513
2	0.29848	59.6970	4.1274
3	0.26787	53.5742	5.1275
4	0.14424	28.8476	0.5977
>5	0.05561	11.1215	8.7744

$$0.18477 \times 200 = 36.9553$$

$$\rightarrow \frac{(37 - 53.5742)^2}{53.5742} = 5.127$$

(c) Complete the test at the 5% significance level. [6]

H_0 : Binomial model is suitable

H_1 : Binomial model is not suitable

$$\chi^2 = \sum \text{contributions} = 10.6022 + 1.7513 + 4.1274 + 5.1275 +$$

$$0.5977 + 8.7744 = 30.98$$

Critical value at $p = 5\%$ and $v = 5 - 1 = 4$ is 9.488. $30.98 > 9.488$ so this result is significant so there is sufficient evidence to suggest that the binomial model is not suitable.

7

Fig. 5 shows the probability distribution $B(7, 0.35)$ together with the relative frequencies of the observed data (the numbers of litters each divided by 200).

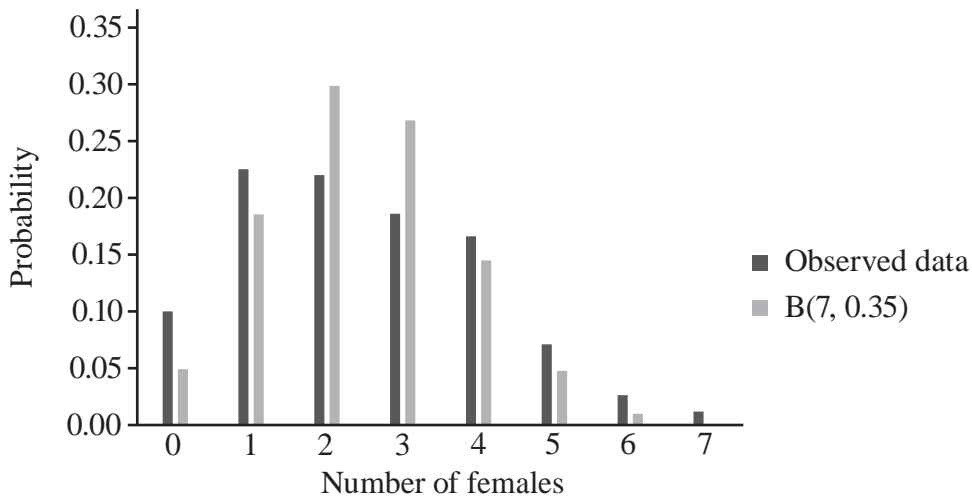


Fig. 5

(d) Comment on the result of the test completed in part (c) by considering Fig. 5.

[1]

- More litters than expected have very few females.
- The observed data shows more positive skew than the binomial model.
- The diagram shows that the observed data has more variation than the binomial distribution would suggest.

- 6 A meteorologist is investigating the relationship between altitude x metres and mean annual temperature y °C in an American state. She selects 12 locations at various altitudes and then stations a remote monitoring device at each of them to measure the temperature over the course of a year. Fig. 6 illustrates the data which she obtains.

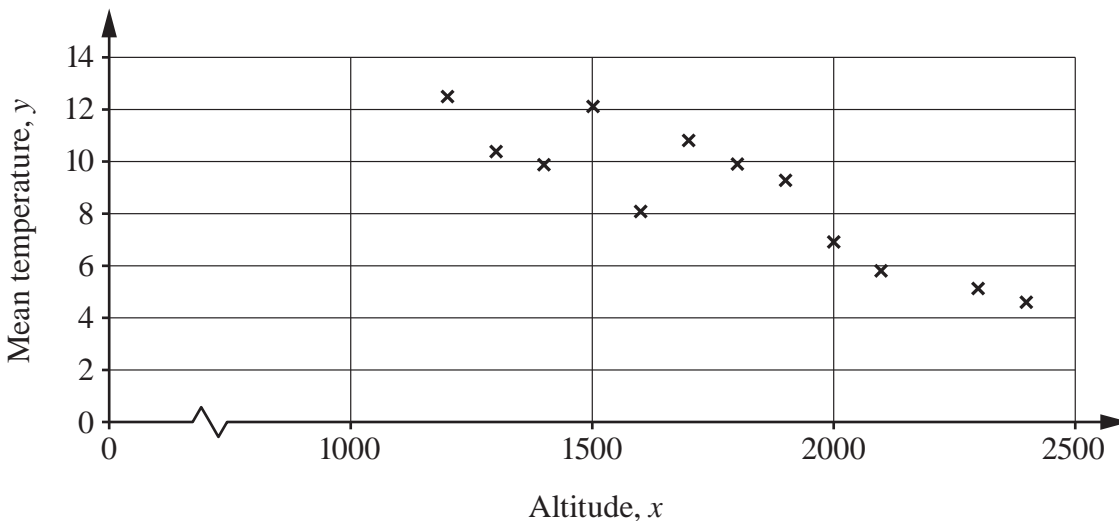


Fig. 6

- (a) Explain why it would not be appropriate to carry out a hypothesis test for correlation based on the product moment correlation coefficient. [2]

PMCC is only valid for random variables and altitude are not random variables but controlled.

- (b) Explain why altitude has been plotted on the horizontal axis in Fig. 6. [1]

Altitude is the independent variable

Summary statistics for x and y are as follows.

$$\Sigma x = 21200 \quad \Sigma y = 105.4 \quad \Sigma x^2 = 39100000 \quad \Sigma y^2 = 1004 \quad \Sigma xy = 176090$$

- (c) Calculate the equation of the regression line of y on x . [4]

$$S_{xy} = 176090 - \frac{21200 \times 105.4}{12} = -\frac{30350}{3}$$

$$S_{xx} = 39100000 - \frac{21200^2}{12} = 1646666.66\dots$$

$$b = \frac{S_{xy}}{S_{xx}} = -0.00614 \quad \bar{x} = \frac{21200}{12} = 1766.6\dots$$

$$\bar{y} = \frac{105.4}{12} = 8.783\dots$$

$$y - 8.783 = -0.00614(x - 1766.6)$$

$$\Rightarrow y = -0.00614x + 19.637$$

(d) Use the equation of the regression line to predict the values of the mean annual temperature at each of the following altitudes.

• 2000 metres $= -0.00614(2000) + 19.637 = 7.35^{\circ}\text{C}$

• 3000 metres $= -0.00614(3000) + 19.637 = 1.21^{\circ}\text{C}$

[2]

(e) Comment on the reliability of your predictions in part (d).

[2]

2000m is more reliable as interpolation and 3000m prediction is less reliable as extrapolation.

(f) Calculate the value of the residual for the data point (1600, 8.1).

[2]

$$\begin{aligned} \text{Residual} &= 8.1 - (-0.00614(1600) + 19.637) \\ &= -1.71 \text{ (3sf)} \end{aligned}$$

END OF QUESTION PAPER

OCR

Oxford Cambridge and RSA

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.