OCR
Oxford Cambridge and RSA

Thursday 6 June 2019 – Afternoon

A Level Further Mathematics B (MEI)

Y432/01 Statistics Minor

Time allowed: 1 hour 15 minutes

Model Answers

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You may use:
- a scientific or graphical calculator

7769553021

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 mark 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 In a game at a charity fair, a spinner is spun 4 times.
On each spin the chance that the spinner lands on a score of 5 is 0.2.
The random variable X represents the number of spins on which the spinner lands on a score of 5.

(a) Find $P(X = 3)$. $X \sim B(4, 0.2)$ **[2]**

$P(X=3) = 0.0246$ (4sf)

(b) Find each of the following.

- $E(X)$
- $\text{Var}(X)$ **[2]**

$E(x) = np = 4 \times 0.2 = 0.8$

$\text{Var } X = npq = 4 \times 0.2 \times (1-0.2)$
$= 0.64$

One game costs £1 to play and, for each spin that lands on a score of 5, the player receives 50 pence.

(c) **(i)** Find the expected total amount of money gained by a player in one game. **[2]**

Expected total $= -1 + 0.8(0.5) = -£0.6/60p$

(ii) Find the standard deviation of the total amount of money gained by a player in one game. **[1]**

$SD(x) = \sqrt{0.64} = 0.8$ so SD of total amount $= 0.8 \times 50 = 40p$

2 A market researcher wants to interview people who watched a particular television programme. Audience research data used by the broadcaster indicates that 12% of the adult population watched this programme. This figure is used to model the situation.
The researcher asks people in a shopping centre, one at a time, if they watched the programme. You should assume that these people form a random sample of the adult population.

(a) Find the probability that the fifth person the researcher asks is the first to have watched the programme. **[2]**

$P = (1-0.12)^4 \times 0.12 = 0.0720$ (4sf)

(b) Find the probability that the researcher has to ask at least 10 people in order to find one who watched the programme. **[1]**

$P(X \geq 10) = 9$ fails $= 0.88^9 = 0.3165$ (4sf)

(c) Find the probability that the twentieth person the researcher asks is the third to have watched the programme. **[3]**

2 in the first 19 so $X \sim B(19, 0.12) \Rightarrow P(X=2) = 0.28026\ldots$
Thus, $P = 0.28026\ldots \times 0.12 = 0.0336$.

(d) Find how many people the researcher would have to ask to ensure that there is a probability of at least 0.95 that at least one of them watched the programme. **[3]**

$P(X \geq 1) = 1 - P(X=0) = 1 - 0.88^n$ so

$1 - 0.88^n \geq 0.95 \Rightarrow 0.88^n \leq 0.05$

$n \geq \log_{0.88} 0.05$ so $n > 23.4\ldots \Rightarrow n = 24$

3 A company has been commissioned to make 50 very expensive titanium components.
A sample of the components needs to be tested to ensure that they are sufficiently strong. However, this is a test to destruction, so the components which are tested can no longer be used.

(a) Explain why it would not be appropriate to use a census in these circumstances. **[1]**

A census would result in all of the components being destroyed.

A manager suggests that the first 5 components to be manufactured should be tested.

(b) Explain why this would not be a sensible method of selecting the sample. **[1]**

This isn't a random method and it's not representative of all of the components.

A statistician advises the manager that the sample selected should be a random sample.

(c) Give two desirable features (other than randomness) that the sample should have. **[2]**

- Unbiased and representative of the population.

- chosen so components are selected independently

4 Zara uses a metal detector to search for coins on a beach.
She wonders if the numbers of coins that she finds in an area of $10m^2$ can be modelled by a Poisson distribution. The table below shows the numbers of coins that she finds in randomly chosen areas of $10m^2$ over a period of months.

Number of coins found	0	1	2	3	4	5	6	>6
Frequency	13	28	30	14	10	2	3	0

(a) Software gives the sample mean as 1.98 and the sample standard deviation as 1.4212.
Explain how these values suggest that a Poisson distribution may be an appropriate model for the numbers of coins found. **[2]**

Variance $= 1.4212^2 = 2.0198$ so as variance is close to the mean, Poisson may be appropriate

Zara decides to carry out a chi-squared test to investigate whether a Poisson distribution is an appropriate model.
Fig. 4 is a screenshot showing part of the spreadsheet used to analyse the data. Some values in the spreadsheet have been deliberately omitted.

	A	B	C	D
1	Number of coins found	Observed frequency	Expected frequency	Chi-squared contribution
2	0	13	13.8069	0.0472
3	1	28		
4	2	30	27.0643	0.3184
5	3	14	17.8625	0.8352
6	4	10	8.8419	0.1517
7	≥5	5		0.0015

Fig. 4

(b) Showing your calculations, find the missing values in each of the following cells.

- C3

$P(X=1) = 0.273377$ so $C3 = 27.3377$

- C7

$P(X \geq 5) = 0.050867$ so $C7 = 5.0867$

- D3

$D3 = \frac{(28-27.3377)^2}{27.3377} = 0.0160$ (4sf)

[4]

(c) Explain why the numbers for 5, 6 and more than 6 coins found have been combined into the single category of at least 5 coins found, as shown in the spreadsheet. **[1]**

Some of the expected frequencies would be less than 5 so the test wouldn't be valid.

(d) Complete the hypothesis test at the 5% level of significance. **[6]**

H_0: Poisson model is a good fit
H_1: Poisson model is not a good fit
Test statistic $\chi^2 = \sum$ contributions $= 1.37$
and critical value for $p = 5\%$ and $v = 6-1-1 = 4$
is 9.488. $1.37 < 9.488$ so this result
is not significant so insufficient
evidence to reject H_0 which would
suggest the Poisson model is a good fit.

For the rest of this question, you should assume that the number of coins that Zara finds in an area of 10m^2 can be modelled by a Poisson distribution with mean 1.98.
Zara also finds pieces of jewellery independently of the coins she finds. The number of pieces of jewellery that she finds per 10m^2 area is modelled by a Poisson distribution with mean 0.42.

(e) Find the probability that Zara finds a total of exactly 3 items (coins and/or jewellery) in an area of 10m^2. **[2]**

$\lambda = 1.98 + 0.42 = 2.4$ so $X \sim Po(2.4)$
$P(X=3) = 0.2090$ (4sf)

(f) Find the probability that Zara finds a total of at least 30 items (coins and/or jewellery) in an area of 100m^2. **[2]**

$\leftarrow \times 10$ to get $10\text{m}^2 \rightarrow 100\text{m}^2$

$\lambda = 10 \times 2.4 = 24$ so $Y \sim Po(24)$
$P(Y \geq 30) = 1 - P(Y \leq 29) = 1 - 0.8679$
$= 0.1321$

5 A student wants to know if there is a positive correlation between the amounts of two pollutants, sulphur dioxide and PM10 particulates, on different days in the area of London in which he lives; these amounts, measured in suitable units, are denoted by s and p respectively.
He uses a government website to obtain data for a random sample of 15 days on which the amounts of these pollutants were measured simultaneously. Fig. 5.1 is a scatter diagram showing the data.
Summary statistics for these 15 values of s and p are as follows.

$\Sigma s = 155.4$ $\quad \Sigma p = 518.9$ $\quad \Sigma s^2 = 2322.7$ $\quad \Sigma p^2 = 21270.5$ $\quad \Sigma sp = 6009.1$

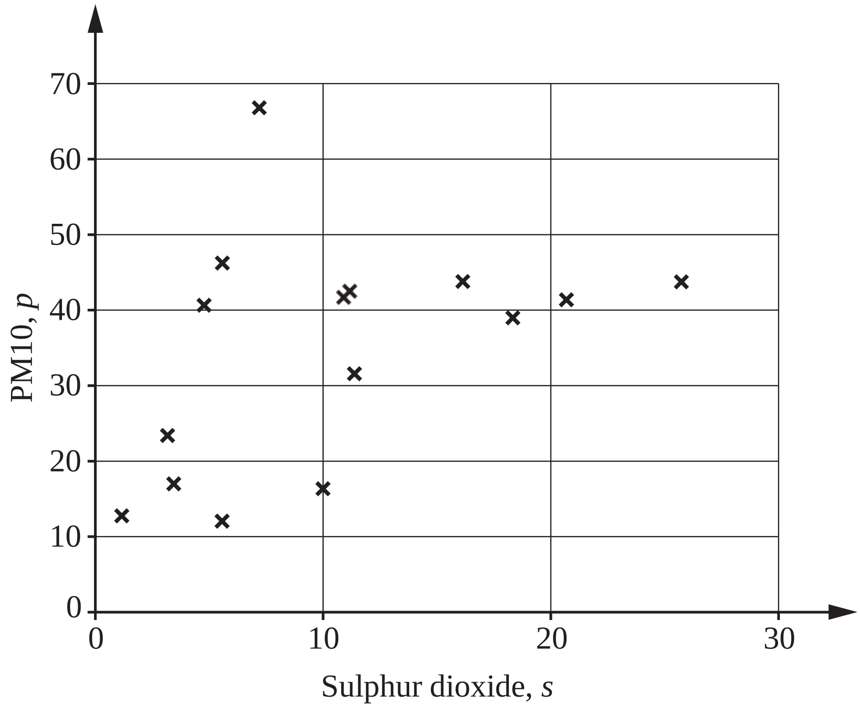

Fig. 5.1

(a) Explain why the student might come to the conclusion that a test based on Pearson's product moment correlation coefficient may be valid. **[2]**

The scatter diagram appears to be roughly elliptical which suggests the underlying distribution may be Bivariate Normal.

(b) Find the value of Pearson's product moment correlation coefficient. **[4]**

$$S_{sp} = 6009.1 - \frac{155.4 \times 518.9}{15} = 633.296$$

$$S_{pp} = 21270.5 - \frac{518.9^2}{15} = 3320.019\ldots$$

$$S_{ss} = 2322.7 - \frac{155.4^2}{15} = 712.756$$

$$r = \frac{S_{sp}}{\sqrt{S_{ss} S_{pp}}} = \frac{633.296}{\sqrt{712.756 \times 3320}} = 0.412$$

(3sf)

(c) Carry out a test at the 5% significance level to investigate whether there is positive correlation between the amounts of sulphur dioxide and PM10 particulates. **[5]**

$H_0: \rho = 0$ and $H_1: \rho > 0$ where ρ is the population correlation coefficient between s and p. Critical value for p=5%, one tail and n=15 is 0.4409 so as 0.412<0.4409, this result is not significant so insufficient evidence to reject H_0 which would suggest there is no correlation between s and p.

(d) Explain why the student made sure that the sample chosen was a random sample. **[2]**

A random sample enables proper inference about the population to be undertaken,

The student also wishes to model the relationship between the amounts of nitrogen dioxide *n* and PM10 particulates *p*.
He takes a random sample of 54 values of the two variables, both measured at the same times. Fig. 5.2 is a scatter diagram which shows the data, together with the regression line of *n* on *p*, the equation of the regression line and the value of r^2.

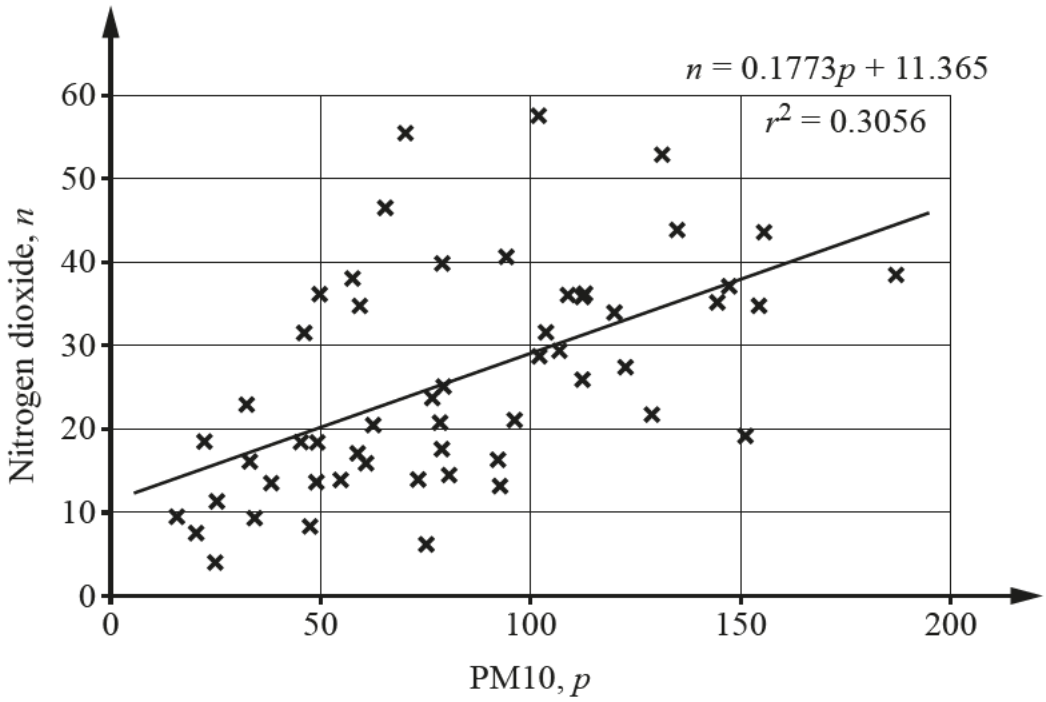

Fig. 5.2

(e) Predict the value of n for $p = 150$. **[1]**

$n = 0.1773(150) + 11.365 = 37.96 \approx 38$

(f) Discuss the reliability of your prediction in part **(e)**. **[2]**

It is interpolated but the points don't lie close to the line so not very reliable.

6 The discrete random variable X has a uniform distribution over $\{n, n+1, \ldots, 2n\}$.

(a) Given that n is odd, find $P(X < \frac{3}{2}n)$. **[1]**

$n = 1 \Rightarrow \{\underline{1}, 2\}$ — means $< \frac{3}{2}n$

$n = 3 \Rightarrow \{\underline{3}, \underline{4}, 5, 6\}$

2

$n = 5 \Rightarrow \{\underline{5}, \underline{6}, \underline{7}, 8, 9, 10\}$

Therefore $P(X < \frac{3}{2}n) = \frac{1}{2}$

(b) Given instead that n is even, find $P(X < \frac{3}{2}n)$ giving your answer as a single algebraic fraction. **[3]**

X can take $n+1$ values as $(2n - n) + 1 = n + 1$ items. Out of $n+1$ items, $\frac{1}{2}n$ are below $\frac{3}{2}n$ so $P(X < \frac{3}{2}n) = \frac{n}{2(n+1)} = \frac{n}{2n+2}$

(c) The sum of 6 independent values of X is denoted by Y.
Find Var(Y). **[3]**

$\text{Var}X = \frac{1}{12}((n+1)^2 - 1) = \frac{1}{12}(n^2 + 2n)$

$Y = X + X + X + X + X + X$ so $\text{Var}Y = 6 \times \text{Var}X$

so $\text{Var}Y = \frac{1}{2}(n^2 + 2n) = \frac{1}{2}n(n+2)$

END OF QUESTION PAPER