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

7769553021

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

You may use:
- a scientific or graphical calculator

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.

© OCR 2019 [M/508/5596]
DC (KS/CB) 173306/4

OCR is an exempt Charity

stemup.co.uk

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 $P(X = 3)$. **[2]**

(b) Find each of the following.

- $E(X)$
- $Var(X)$ **[2]**

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]**

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

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]**

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

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

(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]**

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 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]**

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]**

© OCR 2019

stemup.co.uk

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 $10\,m^2$ can be modelled by a Poisson distribution. The table below shows the numbers of coins that she finds in randomly chosen areas of $10\,m^2$ 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]**

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	$\geqslant 5$	5		0.0015
8				

Fig. 4

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

- C3
- C7
- D3

[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]**

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

For the rest of this question, you should assume that the number of coins that Zara finds in an area of $10\,m^2$ 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 $10\,m^2$ 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 $10\,m^2$. **[2]**

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

© OCR 2019

stemup.co.uk

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.

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

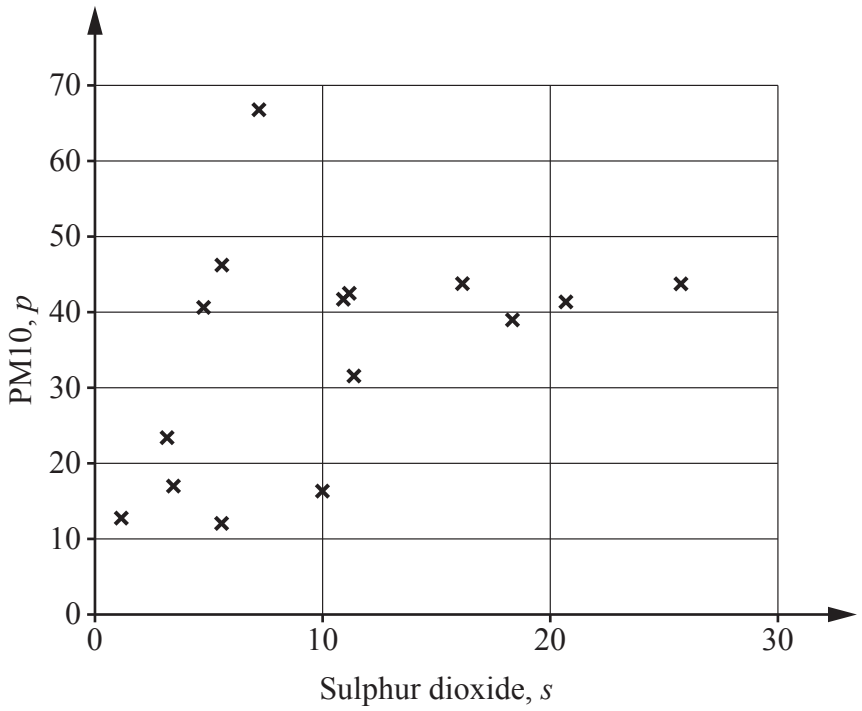


Fig. 5.1

(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]**

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

(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]**

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

© OCR 2019

stemup.co.uk

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.

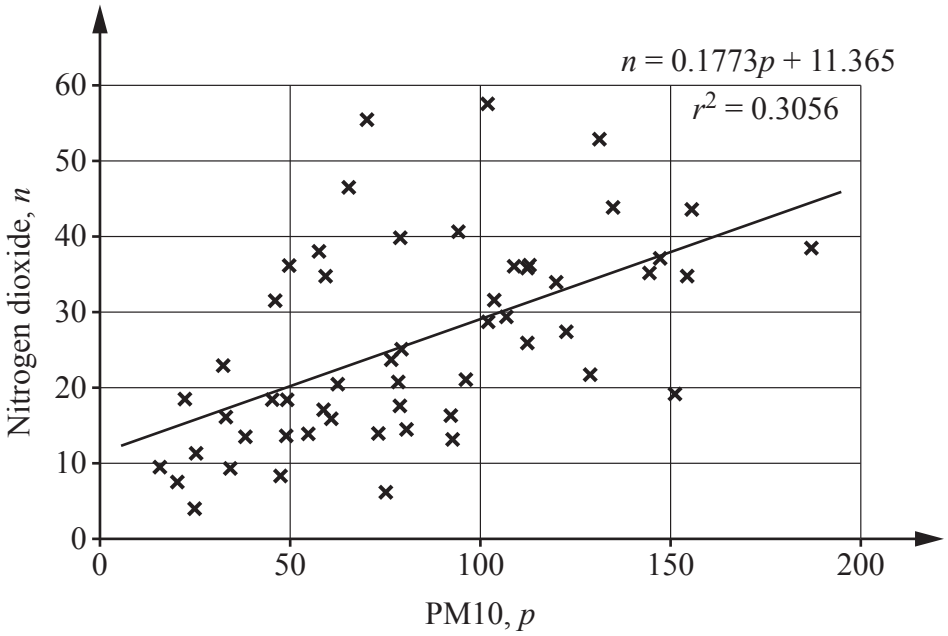


Fig. 5.2

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

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

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

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

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

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

END OF QUESTION PAPER

© OCR 2019

stemup.co.uk

BLANK PAGE

© OCR 2019

stemup.co.uk

BLANK PAGE

© OCR 2019

stemup.co.uk

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.

© OCR 2019

stemup.co.uk

Surname	StemUp Model Solution
Answer booklet	Write the question number in the left-hand margin at the start of each answer.

Question number	
1 (a)	$X \sim B(4, 0.2)$
	$P(X = 3) = {}^4C_3 \times 0.2^3 \times 0.8$
	$= 4 \times 0.008 \times 0.8$
	$= 0.0256$
1 (b)	$E(X) = np = 4 \times 0.2 = 0.8$
	$Var(X) = np(1 - p) = 4 \times 0.2 \times 0.8 = 0.64$
1 (c)(i)	Gain G (in £) $= 0.5X - 1$
	$E(G) = 0.5\ E(X) - 1 = 0.5 \times 0.8 - 1$
	$= 0.4 - 1 = -0.60$
	Expected gain is −£0.60, i.e. an expected loss of 60p per game
1 (c)(ii)	$Var(G) = 0.5^2\ Var(X) = 0.25 \times 0.64 = 0.16$
	$SD(G) = \sqrt{0.16} = 0.4$, so the standard deviation is £0.40 (40p)
2 (a)	X = number asked up to and including the first who watched
	$X \sim Geo(0.12)$
	$P(X = 5) = 0.88^4 \times 0.12$
	$= 0.0720$ (3 s.f.)
2 (b)	$P(X \geq 10) = P(\text{first 9 people did not watch}) = 0.88^9$
	$= 0.3165$ (4 d.p.)
2 (c)	Need exactly 2 watchers in the first 19, then the 20th watched

Question number	
	Number of watchers in first 19 ~ B(19, 0.12)
	P(exactly 2 in 19) = $^{19}C_2 \times 0.12^2 \times 0.88^{17} = 0.28026...$
	$P = 0.28026... \times 0.12$
	$= 0.0336$ (3 s.f.)
2 (d)	P(at least one of n watched) = $1 - 0.88^n \geq 0.95$
	$0.88^n \leq 0.05$
	$n \log 0.88 \leq \log 0.05$
	$n \geqslant \frac{\log 0.05}{\log 0.88} = 23.43...$ (inequality flips as $\log 0.88 < 0$)
	Check: $0.88^{23} = 0.0529 > 0.05$, $0.88^{24} = 0.0465 < 0.05$
	$\therefore$ the researcher must ask at least 24 people
3 (a)	A census tests every component, and testing destroys them,
	so all 50 components would be destroyed and none could be used.
3 (b)	It is not a random method, so the sample may not be representative:
	the production process might change as more components are made.
3 (c)	· The sample should be representative of all 50 components (unbiased).
	· The components should be selected independently of each other.
4 (a)	Sample variance = $1.4212^2 = 2.0198 \approx 2.02$
	For a Poisson distribution mean = variance; here the variance
	(2.02) is very close to the mean (1.98), so Poisson may be suitable.
4 (b)	Using Po(1.98) with total frequency 100:
	C3: $P(X = 1) = e^{-1.98} \times 1.98 = 0.273377$

Question number

$C3 = 100 \times 0.273377 = 27.3377$

$C7$: $P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.949133 = 0.050867$

$C7 = 100 \times 0.050867 = 5.0867$

(check: $100 - (13.8069 + 27.3377 + 27.0643 + 17.8625 + 8.8419) = 5.0867$)

$D3 = \dfrac{(28 - 27.3377)^2}{27.3377} = 0.0160$

4 (c) Otherwise some expected frequencies would be too small (less than 5), so the chi-squared test would not be valid.

4 (d) H_0: the Poisson model is a good fit for the number of coins found

H_1: the Poisson model is not a good fit

$X^2 = 0.0472 + 0.0160 + 0.3184 + 0.8352 + 0.1517 + 0.0015$

$= 1.370$

Degrees of freedom $= 6 - 1 - 1 = 4$ (mean estimated from data)

Critical value of χ^2_4 at 5% level $= 9.488$

$1.370 < 9.488$, so do not reject H_0

There is insufficient evidence to suggest that the Poisson model is not a good fit for the number of coins found per 10 m².

4 (e) Total items $T \sim Po(1.98 + 0.42) = Po(2.4)$ (independent Poissons)

$P(T = 3) = e^{-2.4} \times \dfrac{2.4^3}{3!}$

$= 0.2090$ (4 d.p.)

4 (f) 100 m² = 10 × 10 m², so total items $\sim Po(10 \times 2.4) = Po(24)$

$P(\geq 30) = 1 - P(\leq 29)$

$= 1 - 0.8679$

Question number

= 0.1321 (4 d.p.)

5 (a) The scatter diagram is roughly elliptical in shape,

which suggests the data may come from a bivariate Normal distribution,

so a test based on the pmcc may be valid.

5 (b) $S_{sp} = 6009.1 - \frac{155.4 \times 518.9}{15} = 633.296$

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

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

$r = \frac{S_{sp}}{\sqrt{S_{ss}\, S_{pp}}} = \frac{633.296}{\sqrt{712.756 \times 3320.019}}$

$r = 0.4117$ (4 d.p.)

5 (c) H_0: $\rho = 0$

H_1: $\rho > 0$ (one-tailed test)

where ρ is the population correlation coefficient between s and p

$n = 15$, 5% one-tailed: critical value = 0.4409

$0.4117 < 0.4409$, so do not reject H_0

There is insufficient evidence at the 5% level to suggest that there

is positive correlation between sulphur dioxide and PM10 amounts.

5 (d) A random sample is unbiased, so it allows proper inference

about the population; the hypothesis test is only valid if

the sample is random.

Question number	
5 (e)	$n = 0.1773 \times 150 + 11.365 = 37.96$
	Predicted $n \approx 38$
5 (f)	$p = 150$ is within the range of the data (interpolation), but
	$r^2 = 0.3056$ is low and the points are widely scattered about the line,
	so the prediction is not very reliable.
6 (a)	n odd: X takes the $n + 1$ values $n, n + 1, \ldots, 2n$ (an even number of values)
	$\frac{3}{2}n$ is exactly halfway and is not an integer, so half the values are below it
	$P(X < \frac{3}{2}n) = \frac{1}{2}$
6 (b)	X can take $n + 1$ equally likely values: $n, n + 1, \ldots, 2n$
	n even, so $\frac{3}{2}n$ is an integer; values below it are $n, \ldots, \frac{3}{2}n - 1$
	Number of these $= (\frac{3}{2}n - 1) - n + 1 = \frac{1}{2}n$
	$P(X < \frac{3}{2}n) = \frac{\frac{1}{2}n}{n + 1} = \frac{n}{2n + 2}$
6 (c)	X is uniform over $n + 1$ consecutive integers, so
	$\mathrm{Var}(X) = \frac{1}{12}((n + 1)^2 - 1) = \frac{1}{12}(n^2 + 2n)$
	$\mathrm{Var}(Y) = 6\,\mathrm{Var}(X)$ (independent values)
	$= \frac{6}{12}(n^2 + 2n)$

Question number

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