OCR
Oxford Cambridge and RSA

Thursday 08 October 2020 – Afternoon

A Level Further Mathematics B (MEI)

Y432/01 Statistics Minor

Time allowed: 1 hour 15 minutes

8313381763

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for Further Mathematics B (MEI)
- a scientific or graphical calculator

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined pages at the end of the Printed Answer Booklet. The question numbers must be clearly shown.
- Fill in the boxes on the front of the Printed Answer Booklet.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working. Marks might be given for using a correct method, even if your answer is wrong.
- Give your final answers to a degree of accuracy that is appropriate to the context.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **8** pages.

ADVICE
- Read each question carefully before you start your answer.

© OCR 2020 [M/508/5596]
DC (ST/CT) 193939/3
OCR is an exempt Charity

stemup.co.uk

Answer **all** the questions.

1 A quiz team of 4 students is to be selected from a group of 7 girls and 5 boys. The team is selected at random from the students in the group. The number of girls in the team is denoted by the random variable X.

(a) Show that $P(X = 4) = \frac{7}{99}$. **[1]**

Table 1 shows the probability distribution of X.

r	0	1	2	3	4
$P(X = r)$	$\frac{1}{99}$	$\frac{14}{99}$	$\frac{42}{99}$	$\frac{35}{99}$	$\frac{7}{99}$

Table 1

(b) Find each of the following.

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

It is decided that the quiz team must have at least 1 girl and at least 1 boy, but the team is still otherwise selected at random.

(c) Explain whether $E(X)$ would be smaller than, equal to or larger than the value which you found in part **(b)**. **[2]**

© OCR 2020

stemup.co.uk

2 On computer monitor screens there are often one or more tiny dots which are permanently dark and do not display any of the image. Such dots are known as 'dead pixels'. Dead pixels occur on screens randomly and independently of each other.

A company manufactures three types of monitor, Types A, B and C. For a monitor of Type A, the screen has a total of 2 304 000 pixels. For this type of monitor, the probability of a randomly chosen pixel being dead is 1 in 500 000. Let X represent the number of dead pixels on a monitor screen of this type.

(a) Explain why you could use either a binomial distribution or a Poisson distribution to model the distribution of X. **[3]**

(b) Use a Poisson distribution to calculate estimates of each of the following probabilities.

- $\mathrm{P}(X = 4)$
- $\mathrm{P}(X > 4)$ **[3]**

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

For a monitor of Type B, the probability of a randomly chosen pixel being dead is also 1 in 500 000. The screen of a monitor of Type B has a total of n pixels. Use a binomial distribution to find the least value of n for which the probability of finding at least 1 dead pixel is greater than 0.99. Give your answer in millions correct to 3 significant figures. **[3]**

For a monitor of Type C, the number of dead pixels on the screen is modelled by a Poisson distribution with mean λ.

(d) Given that the probability of finding at least one dead pixel is 0.8, find λ. **[2]**

© OCR 2020

stemup.co.uk

3 In this question you must show detailed reasoning.

In a survey into pet ownership, one of the questions was 'Do you own either a cat or a dog (or both)?'. A total of 121 people took part in the survey and you should assume that they form a random sample of people in a particular town. The results, classified by the age of the person being surveyed, are shown in Table 3.

		Ownership of cat or dog	
		Does own	Does not own
Age	Over 45 years	38	29
	Under 45 years	23	31

Table 3

Carry out a test at the 10% significance level to investigate whether, for people in this town, there is any association between age and ownership of a cat or dog. **[8]**

4 Cards are drawn at random from a standard pack of 52 cards, one at a time, until one of the 4 aces is drawn. After each card is drawn, it is replaced in the pack before the next one is drawn. The random variable X represents the number of draws required to draw the first ace.

(a) State fully the distribution of X. **[1]**

(b) Find $P(X = 10)$. **[2]**

(c) Find each of the following.

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

A further k aces are added to the full pack and the process described above is repeated. The random variable Y represents the number of draws required to draw the first ace.

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

Given that $P(Y = 2) = \frac{8}{81}$, find the two possible values of k. **[5]**

© OCR 2020

stemup.co.uk

5 A student is investigating immunisation. He wonders if there is any relationship between the percentage of young children who have been given measles vaccine and the percentage who have been given BCG vaccine in various countries.

He takes a random sample of 8 countries and finds the data for the two variables. The spreadsheet in Fig. 5.1 shows the values obtained, together with a scatter diagram which illustrates the data.

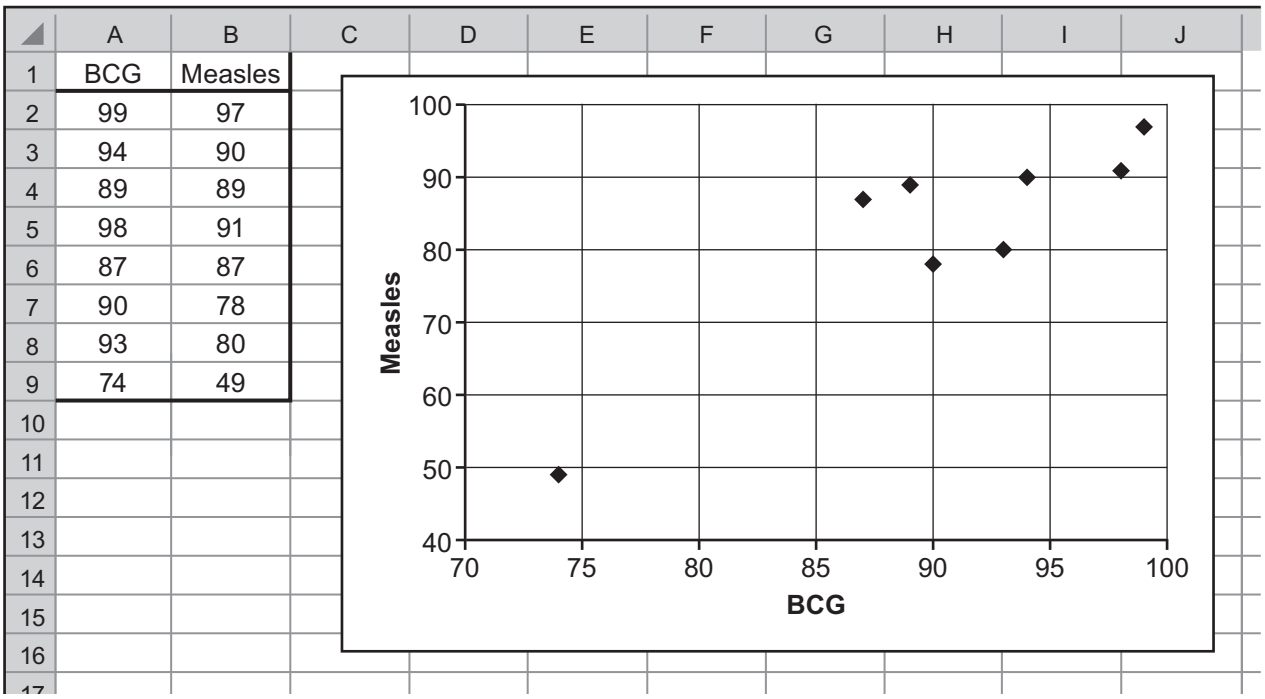


Fig. 5.1

(a) The student decides that a test based on Pearson's product moment correlation coefficient is not valid. Explain why he comes to this conclusion. **[2]**

The student carries out a test based on Spearman's rank correlation coefficient.

(b) Calculate the value of Spearman's rank correlation coefficient. **[3]**

(c) Carry out a test based on this coefficient at the 5% significance level to investigate whether there is any association between measles and BCG vaccination levels. **[5]**

© OCR 2020

stemup.co.uk

The student then decides to investigate the relationship between number of doctors per 1000 people in a country and unemployment rate in that country (unemployment rate is the percentage of the working age population who are not in paid work). He selects a random sample of 6 countries. The spreadsheet in Fig. 5.2 shows the values obtained, together with a scatter diagram which illustrates the data.

	A	B
1	Doctors per 1000	Unemployment rate
2	3.66	6.59
3	3.20	8.66
4	3.21	10.30
5	4.13	4.98
6	3.32	7.73
7	2.78	11.26

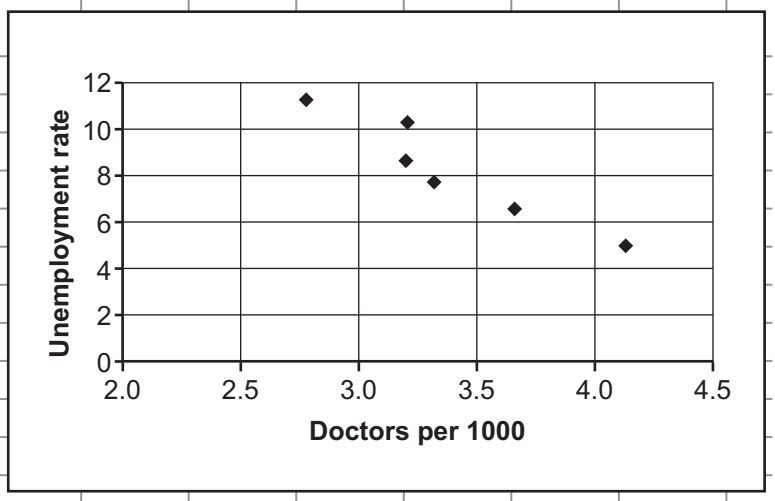


Fig. 5.2

(d) Use your calculator to write down the equation of the regression line of unemployment rate on doctors per 1000. **[2]**

(e) Use the regression line to estimate the unemployment rate for a country with 2.00 doctors per 1000. **[1]**

(f) Comment briefly on the reliability of your answer to part **(e)**. **[1]**

The student decides to add the data for another country with 3.99 doctors per 1000 and unemployment rate 11.42 to his diagram.

(g) Add this point to the scatter diagram in the Printed Answer Booklet. **[1]**

(h) Without doing any further calculations, comment on what difference, if any, including this extra data point would make to the usefulness of a regression line of unemployment rate on doctors per 1000. **[2]**

© OCR 2020

stemup.co.uk

6 **(a)** The random variable X has a uniform distribution over the values $\{1, 2, \ldots, n\}$. Show that $\text{Var}(X)$ is given by $\frac{1}{12}(n^2 - 1)$. **[3]**

(b) The random variable Y has a uniform distribution over the values $\{1, 3, 5, \ldots, 2n - 1\}$. Using the result in part **(a)** or otherwise, show that $\text{Var}(Y)$ is given by $\frac{1}{3}(n^2 - 1)$. **[2]**

(c) Given that $n = 100$, find the least value of k for which $P(\mu - k\sigma \leqslant Y \leqslant \mu + k\sigma) = 1$, where the mean and standard deviation of Y are represented by μ and σ respectively. **[4]**

END OF QUESTION PAPER

© OCR 2020

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 2020

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) $P(X = 4) = P(\text{all 4 chosen are girls})$

$= \frac{7}{12} \times \frac{6}{11} \times \frac{5}{10} \times \frac{4}{9}$

$= \frac{840}{11880} = \frac{7}{99}$ as required

1 (b) $E(X) = \frac{0\times1 + 1\times14 + 2\times42 + 3\times35 + 4\times7}{99} = \frac{231}{99} = \frac{7}{3} = 2.33$

$E(X^2) = \frac{0 + 14 + 168 + 315 + 112}{99} = \frac{609}{99}$

$\text{Var}(X) = \frac{609}{99} - (\frac{7}{3})^2 = \frac{70}{99} = 0.707$ (3 s.f.)

1 (c) Now $X = 0$ and $X = 4$ are impossible, so $P(X = 0) = 1/99$ and $P(X = 4) = 7/99$ are removed and the other probabilities increase in proportion (each divided by $91/99$).

$P(X = 4)$ is much larger than $P(X = 0)$, so more probability is removed from the top end than the bottom end.

So $E(X)$ would be smaller than $7/3$.

(Check: $E(X) = \frac{1\times14 + 2\times42 + 3\times35}{91} = \frac{203}{91} = 2.23 < 2.33$)

2 (a) Binomial: there is a fixed number of trials, $n = 2304000$ pixels, each pixel is either dead or not dead, dead pixels occur independently, and each has the same constant probability

Question number

$p = 1/500000$ of being dead. So $X \sim B(2304000, 1/500000)$.

Poisson: n is very large and p is very small, so the binomial can be approximated by a Poisson distribution with mean np.

2 (b) $\lambda = np = 2304000 \times \frac{1}{500000} = 4.608$

$X \approx Po(4.608)$

$P(X = 4) = e^{-4.608} \times \frac{4.608^4}{4!} = 0.187$ (3 s.f.)

$P(X > 4) = 1 - P(X \leq 4) = 1 - 0.5117$

$= 0.488$ (3 s.f.)

2 (c) Let $Y \sim B(n, \frac{1}{500000})$

$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (\frac{499999}{500000})^n > 0.99$

$\Rightarrow (\frac{499999}{500000})^n < 0.01$

$\Rightarrow n \ln(\frac{499999}{500000}) < \ln 0.01$

$\Rightarrow n > \frac{\ln 0.01}{\ln(499999/500000)}$ (inequality reverses as $\ln(499999/500000) < 0$)

$\Rightarrow n > 2302582.79...$

Least $n = 2302583$, i.e. 2.30 million pixels (3 s.f.)

2 (d) $P(\text{at least one}) = 0.8 \Rightarrow P(\text{none}) = 0.2$

$e^{-m} = 0.2$

$m = -\ln 0.2 = \ln 5 = 1.61$ (3 s.f.)

Question number

3

H_0: no association between age and ownership of a cat or dog

H_1: some association between age and ownership of a cat or dog

Row totals: Over 45 = 67, Under 45 = 54. Column totals: 61, 60. Total 121

Expected frequencies, $E = \frac{\text{row total} \times \text{column total}}{121}$:

	Does own	Does not own
Over 45	$\frac{67 \times 61}{121} = 33.78$	$\frac{67 \times 60}{121} = 33.22$
Under 45	$\frac{54 \times 61}{121} = 27.22$	$\frac{54 \times 60}{121} = 26.78$

Contributions $\frac{(O - E)^2}{E}$:

	Does own	Does not own
Over 45	0.5280	0.5368
Under 45	0.6551	0.6661

$X^2 = 0.5280 + 0.5368 + 0.6551 + 0.6661 = 2.386$

Degrees of freedom $= (2 - 1)(2 - 1) = 1$, so refer to χ^2_1

Critical value at the 10% level = 2.706

$2.386 < 2.706$, so the result is not significant: do not reject H_0

There is insufficient evidence to suggest that there is any association between age and ownership of a cat or dog for people in this town.

4 (a)

$X \sim \text{Geo}(\frac{1}{13})$ $(p = 4/52 = 1/13)$

Question number

4 (b) $P(X = 10) = \left(\frac{12}{13}\right)^9 \times \frac{1}{13}$

$= 0.0374$ (3 s.f.)

4 (c) $E(X) = \frac{1}{p} = 13$

$Var(X) = \frac{1-p}{p^2} = \frac{12/13}{1/169} = 156$

4 (d) $P(ace) = p = \frac{k+4}{k+52}$

$P(Y = 2) = (1-p)p = \frac{8}{81}$

$\Rightarrow 81p - 81p^2 = 8$

$\Rightarrow 81p^2 - 81p + 8 = 0$

$\Rightarrow (9p - 1)(9p - 8) = 0$

$\Rightarrow p = \frac{1}{9}$ or $p = \frac{8}{9}$

$\frac{k+4}{k+52} = \frac{1}{9} \Rightarrow 9k + 36 = k + 52 \Rightarrow 8k = 16 \Rightarrow k = 2$

$\frac{k+4}{k+52} = \frac{8}{9} \Rightarrow 9k + 36 = 8k + 416 \Rightarrow k = 380$

So $k = 2$ or $k = 380$

5 (a) The scatter diagram is not roughly elliptical, because of the outlier at (74, 49), so the data do not appear to come from a bivariate Normal distribution. So a PMCC test is not valid.

Question number

5 (b)

BCG	99	94	89	98	87	90	93	74
Measles	97	90	89	91	87	78	80	49
Rank BCG	1	3	6	2	7	5	4	8
Rank Meas.	1	3	4	2	5	7	6	8
d	0	0	2	0	2	−2	−2	0
d^2	0	0	4	0	4	4	4	0

$\Sigma d^2 = 16, \quad n = 8$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 16}{8 \times 63} = 1 - \frac{96}{504}$$

$$= \frac{17}{21} = 0.810 \text{ (3 s.f.)}$$

5 (c)

H_0: there is no association between measles and BCG vaccination levels in the population

H_1: there is some association between measles and BCG vaccination levels in the population

Two-tailed test, $n = 8$, 5% level: critical value = 0.7381

$0.8095 > 0.7381$, so the result is significant: reject H_0

There is sufficient evidence to suggest that there is some association between measles and BCG vaccination levels.

5 (d)

Let x = doctors per 1000 and y = unemployment rate

$y = 24.49 - 4.80x$

5 (e)

$y = 24.49 - 4.799 \times 2.00 = 14.89$, so about 14.9%

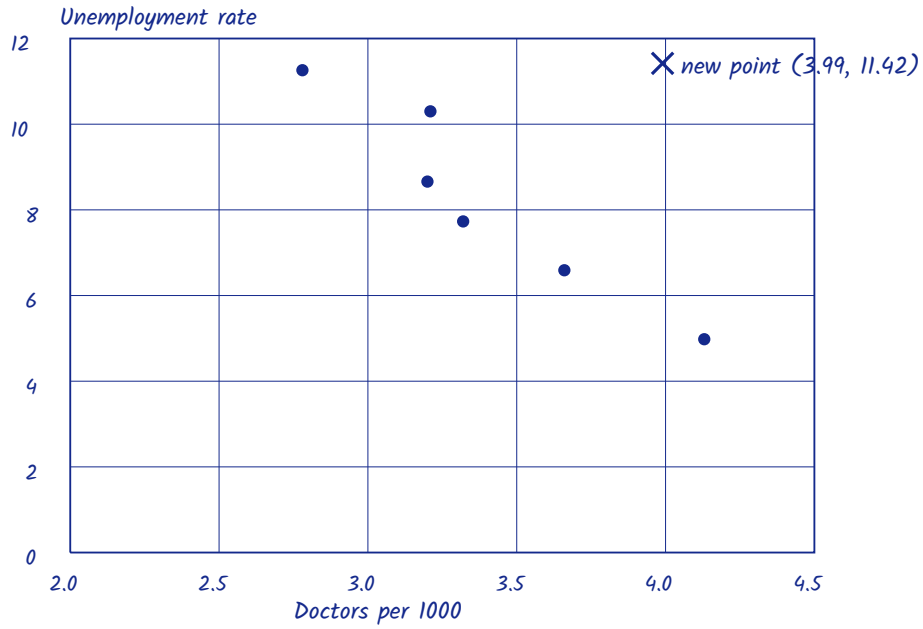
5 (f)

Not reliable: $x = 2.00$ is outside the range of the data

Question number

(2.78 to 4.13), so this is extrapolation.

5 (g)



5 (h)

The new point lies well away from the pattern of the other points (it is a possible outlier), so the fit of the line would be worse (larger residuals, weaker correlation).

So the regression line would be less useful / less reliable for estimating unemployment rate from doctors per 1000.

6 (a)

$P(X = r) = \frac{1}{n}$ for $r = 1, 2, \ldots, n$

$$E(X) = \frac{1}{n}(1 + 2 + \ldots + n) = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2}$$

$$E(X^2) = \frac{1}{n}(1^2 + 2^2 + \ldots + n^2) = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6}$$

$$= \frac{(n+1)(2n+1)}{6}$$

Question number

$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$

$= \frac{2(2n^2+3n+1) - 3(n^2+2n+1)}{12}$

$= \frac{4n^2+6n+2-3n^2-6n-3}{12}$

$= \frac{1}{12}(n^2-1)$ as required

6 (b) If X is uniform on $\{1, 2, \ldots, n\}$ then $Y = 2X - 1$ is uniform on $\{1, 3, 5, \ldots, 2n-1\}$

$\text{Var}(Y) = 2^2\,\text{Var}(X) = 4 \times \frac{1}{12}(n^2-1) = \frac{1}{3}(n^2-1)$ as required

6 (c) $\mu = E(Y) = 2E(X) - 1 = 2 \times \frac{n+1}{2} - 1 = n = 100$

$\sigma^2 = \frac{1}{3}(100^2-1) = 3333$, $\sigma = \sqrt{3333} = 57.735\ldots$

Y takes values from 1 to 199, so we need $\mu - k\sigma \leq 1$ and $\mu + k\sigma \geq 199$

$100 - 57.735k \leq 1 \Rightarrow k \geq 99/57.735$

Least $k = 1.7148\ldots = 1.715$ (4 s.f.)