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

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

8313381763

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.

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.

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

1 a) Probability of 4 girls

$= \frac{7}{12} \times \frac{6}{11} \times \frac{5}{10} \times \frac{4}{9} = \frac{7}{99}$

b) $E(x) = (0 \times 1/99) + (1 \times \frac{14}{99}) + (2 \times \frac{42}{99}) + (3 \times \frac{35}{99}) + (4 \times \frac{7}{99})$

$= 0 + \frac{14}{99} + \frac{28}{33} + \frac{35}{33} + \frac{28}{99}$

$= \frac{7}{3}$

$\text{Var}(x) = (0^2 \times 1/99) + (1^2 \times \frac{14}{99}) + (2^2 \times \frac{42}{99}) + (3^2 \times \frac{35}{99}) + (4^2 \times 7/99) - (\frac{7}{3})^2$

$= (0 + \frac{14}{99} + \frac{56}{33} + \frac{35}{11} + \frac{112}{99}) - \frac{49}{9}$

$= \frac{70}{99}$

c) E(x) would be smaller than the value found in part b, because $P(x=4)$ is higher than $P(x=0)$, so other probabilities would increase in proportion.

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.

- $P(X = 4)$
- $P(X > 4)$ **[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]**

2 a) Poisson: Because n is large (2 304 000) and p is small (1/500 000) so therefore Poisson distribution is appropriate.

Binomial: There are only two possible outcomes, the pixel is either dead or not. And the probability of a pixel being dead has a fixed probability and is independent of other pixels.

make sure you put the answer in the context of the question

b) $X \sim Po(4.608)$ Remember $\lambda = np$

$P(X = 4) = 0.187310069 = 0.1873$ (4sf)

$P(X > 4) = 1 - P(X \leq 4) = 1 - 0.511734587$

$= 0.48826... = 0.4883$ (4sf)

c) $1 - \left(\frac{499\,999}{500\,000}\right)^n > 0.99$

$\therefore \quad 0.01 > \left(\frac{499\,999}{500\,000}\right)^n$

$\log 0.01 > n \log\left(\frac{499\,999}{500\,000}\right)$

$\frac{\log 0.01}{\log\left(\frac{499\,999}{500\,000}\right)} < n \quad \longrightarrow \quad 2302582.79 < n$

≈ 2.30 million

$\therefore$ n is at least 2.30 million

d) $p = 0.8, \quad n = 2302582.79$

$P(X \geq 1) = 0.8$

$\therefore P(X = 0) = 1 - P(X \geq 1) = 0.2$

$e^{-\lambda} = 0.2$

$-\lambda = \ln 0.2$

$\lambda = -\ln 0.2 = 1.6094371\ldots$

$= 1.609$ (4sf)

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	67
	Under 45 years	23	31	54
		61	60	121

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]**

USING CHI SQUARED TESTS:

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

H_1 : there is an association between age and ownership of a cat or dog

2

observed frequencies	owns	doesn't
over 45	38	29
under 45	23	31

expected frequencies	owns	doesn't
over 45	$121 \times \frac{61}{121} \times \frac{67}{121}$ $= 33.7768$	$121 \times \frac{67}{121} \times \frac{60}{121}$ $= 33.2231$
under 45	$121 \times \frac{61}{121} \times \frac{54}{121}$ $= 27.2231$	$121 \times \frac{60}{121} \times \frac{54}{121}$ $= 26.7769$

3 $\chi^2 \quad \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

$= \frac{(38-33.78)^2}{33.78} + \frac{(29-33.22)^2}{33.22} + \frac{(23-27.22)^2}{27.22} + \frac{(31-26.78)^2}{26.78}$

$= 0.5272 + 0.5361 + 0.6542 + 0.6650$

$= 2.3825$

4 $\nu = (n-1)(m-1) = (2-1)(2-1) = 1$

5 critical value at 10%, at 1 degree of freedom = 2.71

6 2.3825 < 2.71, result is not significant

The result is insignificant, there is insufficient evidence to reject Ho.

Suggests there is no association between age and owning a dog or cat.

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]**

$\frac{4}{52} = \frac{1}{13}$

4a) It is the Geometric distribution, $Geo\left(\frac{1}{13}\right)$

b) $Geo(1/13)$

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

$= 0.037428\ldots$

$= 0.03743$ (4sf)

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

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

d) $P(X=2) = \frac{8}{81}$

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

$8 = 81(1-p)p = 81(p - p^2)$

$8 = 81p - 81p$

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

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

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

$\frac{8}{9} = \frac{4+k}{52+k}$ or $\frac{1}{9} = \frac{4+k}{52+k}$

$416 + 8k = 36 + 9k$ | $52 + k = 36 + 9k$

$k = 380$ | $16 = 8k$

$2 = k$

$\therefore k = 2$ or $k = 380$

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.

	A	B
1	BCG	Measles
2	99	97
3	94	90
4	89	89
5	98	91
6	87	87
7	90	78
8	93	80
9	74	49

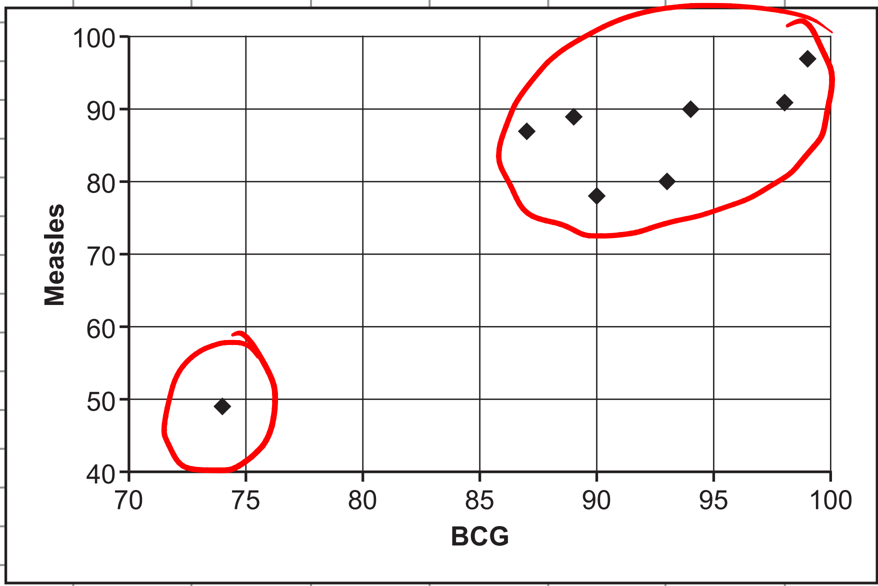

Fig. 5.1

(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]**

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
8		
9		
10		
11		
12		
13		

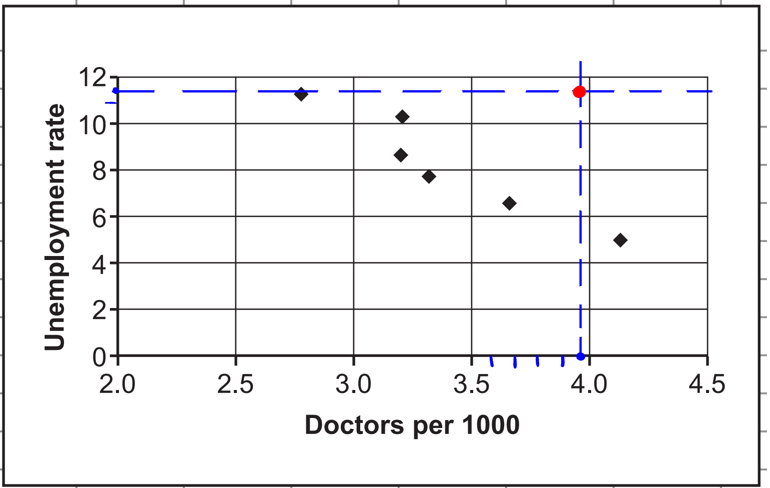

Fig. 5.2

(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]**

a) PMCC is not valid as there are two islands so there isn't an elliptical shape

b) Calculate r_s : $1 - \dfrac{6\sum d_i^2}{n(n^2-1)}$

	A	B
1	BCG	Ranking
2	99	1
3	94	3
4	89	6
5	98	2
6	87	7
7	90	5
8	93	4
9	74	8

	A	B
1	Ranking	Measles
2	1	97
3	3	90
4	4	89
5	2	91
6	5	87
7	7	78
8	6	80
9	8	49

	A	B	
1	BCG	Measles	
2	1	1	0
3	3	3	0
4	6	4	2
5	2	2	0
6	7	5	2
7	5	7	-2
8	4	6	-2
9	8	8	0

$\sum d_i^2 = 0^2+0^2+2^2+0^2+2^2+(-2)^2+(-2)^2+0^2$

$= 16$

$\therefore 1 - \dfrac{6(16)}{8(8^2-1)}$

$= 0.8095238095$

$r_s = 0.8095$ (4sf)

c) test at 5% significance

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

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

$r_s = 0.8095$ (4sf)

at 5% sig level, the CV $(n=8)$ is 0.7381

(this is a two-tailed test)

$0.8095 > 0.7381$, significant

There is sufficient evidence to reject H_0, would suggest there is some association between measles and BCG vaccination levels in the population

d) $y = a + bx \rightarrow y = 24.4899... - 4.7990...x$

$y = 24.49 - 4.80x$

e) $y = 24.49 - 4.80(2)$

$= 14.89$

f) this answer is not reliable as it is extrapolation

g) marked

h) The fit would be worse, as the regression line may not be valid anymore due to this new point, which may be an outlier

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]**

a) $E(x) = \frac{n+1}{2}$

$E(x^2) = \frac{1}{n}(1^2 + 2^2 \ldots n^2)$

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

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

$\text{Var}(x) = E(x^2) - (E(x))^2$

$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$

$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$

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

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

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

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

b) $\text{Var}(x) = 2^2(\text{Var } x) \rightarrow Y = 2X - 1$

$\therefore 4 \times \frac{1}{12}(n^2-1) = \frac{1}{3}(n^2-1)$ $\therefore$ as required

c) $E(x) = \lambda = n = 100$

$\therefore \text{Var}(x) = \frac{1}{3}(100^2 - 1) = 3333$

$\sigma = \sqrt{3333} = 57.7321401 = 57.73$ (4sf)

$100 - 57.73k = 1$

$k = 1.7148\ldots \approx k = 1.715$ (4sf)

END OF QUESTION PAPER