OCR
Oxford Cambridge and RSA

AS Level Further Mathematics B (MEI)

Y412/01 Statistics a

Question Paper

Tuesday 22 May 2018 – Afternoon

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

* 7 0 6 3 5 6 9 7 9 7 *

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- 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.
- Do **not** write in the barcodes.
- 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 Over a period of time, radioactive substances decay into other substances. During this decay a Geiger counter can be used to detect the number of radioactive particles that the substance emits.

A certain radioactive source is decaying at a constant average rate of 6.1 particles per 10 seconds. The particles are emitted randomly and independently of each other.

(i) State a distribution which can be used to model the number of particles emitted by the source in a 10-second period. **[1]**

$X \sim \text{Poisson}(6.1)$

(ii) State the variance of this distribution. **[1]**

$\text{Var}X = \lambda = 6.1$

(iii) Find the probability that at least 6 particles are detected in a period of 10 seconds. **[1]**

$P(X \geq 6) = 0.5702$ (4sf)

(iv) Find the probability that at least 36 particles are detected in a period of 60 seconds. **[2]**

New $\lambda = 6.1 \times 6 = 36.6$ ×6 from 10 to 60 secs

$Y \sim Po(36.6)$. $P(Y \geq 36) = 0.5616$

(v) Another radioactive source emits particles randomly and independently at a constant average rate of 1.7 particles per 5 seconds. Find the probability that at least 10 but no more than 15 particles are detected altogether from the two sources in a period of 10 seconds. **[2]**

New $\lambda = 6.1 + (2 \times 1.7) = 9.5$. $Z \sim Po(9.5)$

$P(10 \leq Z \leq 15) = P(Z \leq 15) - P(Z \leq 9)$

$= 0.9665 - 0.5218 = 0.4447$

2 In a quiz, competitors have to match 5 landmarks to the 5 British counties which the landmarks are in. The random variable X represents the number of correct matches that a competitor gets, assuming that the competitor guesses randomly. The probability distribution of X is given in the following table.

r	0	1	2	3	4	5
$P(X = r)$	$\frac{11}{30}$	$\frac{3}{8}$	$\frac{1}{6}$	$\frac{1}{12}$	0	$\frac{1}{120}$

(i) Explain why $P(X = 4)$ must be 0. **[1]**

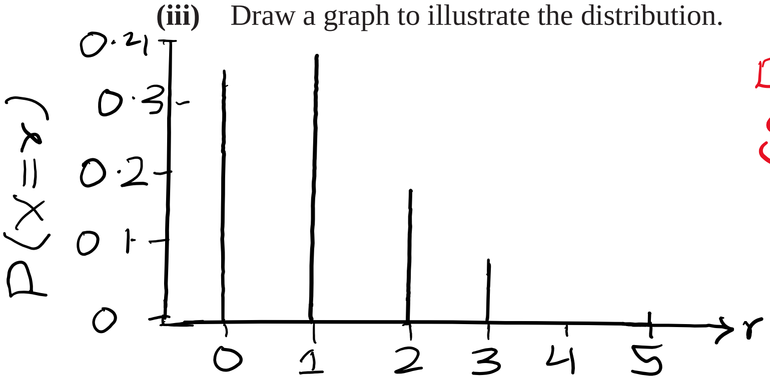
If you have matched 4 correctly then the 5th must be also matched correctly

(ii) Explain how the value $\frac{1}{120}$ for $P(X = 5)$ is calculated. **[2]**

$\frac{1}{5} \times \frac{1}{4} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{1} = \frac{1}{120}$

(You choose one from 5 options the one from 4, one from 3 etc.)

(iii) Draw a graph to illustrate the distribution. **[2]**



Draw straight line graph NOT a bar chart for discrete random variables

(iv) Find each of the following.

- E(X)
- Var(X)

[2]

$E(X) = \sum r P(X=r) = \left(0 \times \frac{11}{30}\right) + \left(1 \times \frac{3}{8}\right) + \left(2 \times \frac{1}{6}\right) + \left(3 \times \frac{1}{12}\right) + (4 \times 0) + \left(5 \times \frac{1}{120}\right)$

$= 1$

$E(X^2) = \sum r^2 P(X=r) = \left(0^2 \times \frac{11}{30}\right) + \left(1^2 \times \frac{3}{8}\right) + \left(2^2 \times \frac{1}{6}\right) + \left(3^2 \times \frac{1}{12}\right) + (4^2 \times 0) + \left(5^2 \times \frac{1}{120}\right)$

$= 2$

$\text{Var} X = E(X^2) - [E(X)]^2 = 2 - 1^2 = 1$

(v) Find P (X>E(X)). **[1]**

$P(X > 1) = 1 - P(X=1) - P(X=0) = 1 - \frac{3}{8} - \frac{11}{30} = \frac{31}{120}$

(vi) There are 12 competitors in the quiz. Assuming that they all guess randomly, find the probability that at least one of them gets all five matches correct. **[2]**

P(don't get all 5) $= 1 - \frac{1}{120} = \frac{119}{120}$

P(at least 1 gets all 5 correct) = 1 − P(no one gets any correct)

$= 1 - \left(\frac{119}{120}\right)^{12} = 0.0955$

3 Samples of water are taken from 10 randomly chosen wells in an area of a country. A researcher is investigating whether there is any relationship between the levels of dissolved oxygen, x, and the amounts of radium, y, in the water from the wells. Both quantities are measured in suitable units. The table and the scatter diagram in Fig. 3 show the values of x and y for the ten wells.

x	45.9	48.3	52.2	64.6	66.6	67.6	69.3	75.0	77.4	82.8
y	25.4	23.9	26.6	18.8	18.9	19.0	16.8	16.3	17.8	17.2

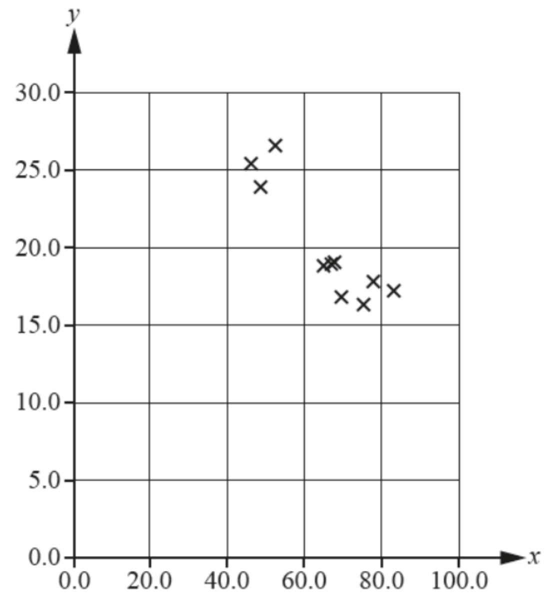


Fig. 3

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

The scatter diagram doesn't suggest a bivariate Normal distribution since the points don't appear in a roughly elliptical pattern; they appear as 2 seperate islands on the scatter diagram and thus PMCC cannot be used.

(ii) Calculate Spearman's rank correlation coefficient for these data. **[3]**

rank x	1	2	3	4	5	6	7	8	9	10
rank y	9	8	10	5	6	7	2	1	4	3
d	-8	-6	-7	-1	-1	-1	5	7	5	7
d^2	64	36	49	1	1	1	25	49	25	49

$r_x - r_y$

$\Sigma d^2 = 64 + 36 + 49 + 1 + 1 + 1 + 25 + 49 + 25 + 49$
$= 300$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{(6 \times 300)}{10(100-1)} = -0.818$$
$$\left(-\frac{9}{11}\right)$$

(iii) Using this value of Spearman's rank correlation coefficient, carry out a hypothesis test at the 1% significance level to investigate whether there is any association between *x* and *y*. **[5]**

H_0: There is no association between level of dissolved oxygen and amount of radium
H_1: There is an association between level of dissolved oxygen and amount of radium
For $n=10$ and 1% significance level, critical value $= 0.7939$.
always compare absolute values of critical value and rs
As $0.818 > 0.7939$, this result is significant so sufficient evidence to reject H_0 which would suggest there is an association between level of dissolved oxygen and amount of radium.

(iv) Explain the meaning of the term 'significance level' in the context of the test carried out in part **(iii)**. **[2]**

The significance level is the probability of rejecting H_0 when it is true. In this case, the 1% significance level means that if there is no association between x and y, about 1 sample in 100 would lead to the false result that there is an association

4 The probability that an expert darts player hits the bullseye on any throw is 0.12, independently of any other throw. The player throws darts at the bullseye until she hits it.

(i) Find the probability that the player has to throw exactly six darts. **[2]**

$X \sim Geo(0.12)$. $P(X=6) = P(5 \text{ misses}) \times P(1 \text{ hit})$
$= 0.88^5 \times 0.12 = 0.0633$

(ii) Find the probability that the player has to throw more than six darts. **[1]**

$P(X>6) = P(6 \text{ misses}) = 0.88^6$
$= 0.464$ (3sf)

(iii) (*A*) Find the mean number of darts that the player has to throw. **[1]**

$E(X) = \frac{1}{p} = \frac{1}{0.12} = 8.33$

(*B*) Find the variance of the number of darts that the player has to throw. **[1]**

$\mathrm{Var}\,X = \frac{1-p}{p^2} = \frac{1-0.12}{0.12^2} = 61.1$

The player continues to throw more darts at the bullseye after she has hit it for the first time.

(iv) Find the probability that the player hits the bullseye at least twice in the first ten throws. **[2]**

$X \sim B(10, 0.12)$. $P(X \geq 2) = 1 - P(X \leq 1)$

$= 1 - 0.6583 = 0.3417$

As the player continues after hitting it once, the distribution switches from Geometric to Binomial.

(v) Find the probability that the player hits the bullseye for the second time on the tenth throw. **[2]**

As the first bullseye can be thrown in any of the first 9 throws;

$P = 9 \times 0.88^8 \times 0.12^2 = 0.0466$

5 A random sample of workers for a large company were asked whether they are smokers, ex-smokers or have never smoked. The responses were classified by the type of worker: Managerial, Production line or Administrative.

Fig. 5 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	E	F
1	**Observed frequencies**					
2		**Smoker**	**Ex-smoker**	**Never smoked**	**Totals**	
3	**Managerial**	2	10	5	17	
4	**Production line**	18	15	21	54	
5	**Administrative**	13	6	14	33	
6	**Totals**	33	31	40	104	
7						
8		**Expected frequencies**				
9		5.3942	5.0673	6.5385		
10		17.1346		20.7692		
11		10.4712	9.8365	12.6923		
12						
13		**Contributions to the test statistic**				
14		2.1358	4.8017	0.3620		
15		0.0437		0.0026		
16			1.4964	0.1347		
17				**Test statistic**	9.66	
18						

Fig. 5

(i) (*A*) State the sample size. **[1]**

Sample size = 104

(*B*) State the null and alternative hypotheses for a test to investigate whether there is any association between type of worker and smoking status. **[1]**

H_0: No association between smoking status and type of worker
H_1: Some association between smoking status and type of worker

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

- C10 $= \dfrac{31 \times 54}{104} = 16.10$
- C15 $= \dfrac{(15-16.10)^2}{16.10} = 0.0746$
- B16 $= \dfrac{(13-10.4712)^2}{10.4712} = 0.6107$

[4]

(iii) Complete the hypothesis test at the 10% level of significance. **[4]**

Test statistic $= 9.66$ (χ^2)
Degrees of freedom $= (3-1)(3-1) = 4$ and $p = 10\%$ so critical value $= 7.779$. As $9.66 > 7.779$, this result is significant so sufficient evidence to reject H_0, which would suggest there is some association between smoking status and type of worker.

(iv) Discuss briefly what the data suggest about smoking status for different types of workers. You should make a comment for each type of worker. **[3]**

For managerial workers, large contributions shows there are fewer smokers and more ex-smokers than expected.
For production line, small contributions show the numbers are as expected.
For administrative workers, large contributions suggest there are fewer ex-smokers than expected.

6 A researcher is investigating various bodily characteristics of frogs of various species. She collects data on length, x mm, and head width, y mm, of a random sample of 14 frogs of a particular species. A scatter diagram of the data is shown in Fig. 6, together with the equation of the regression line of y on x and also the value of r^2.

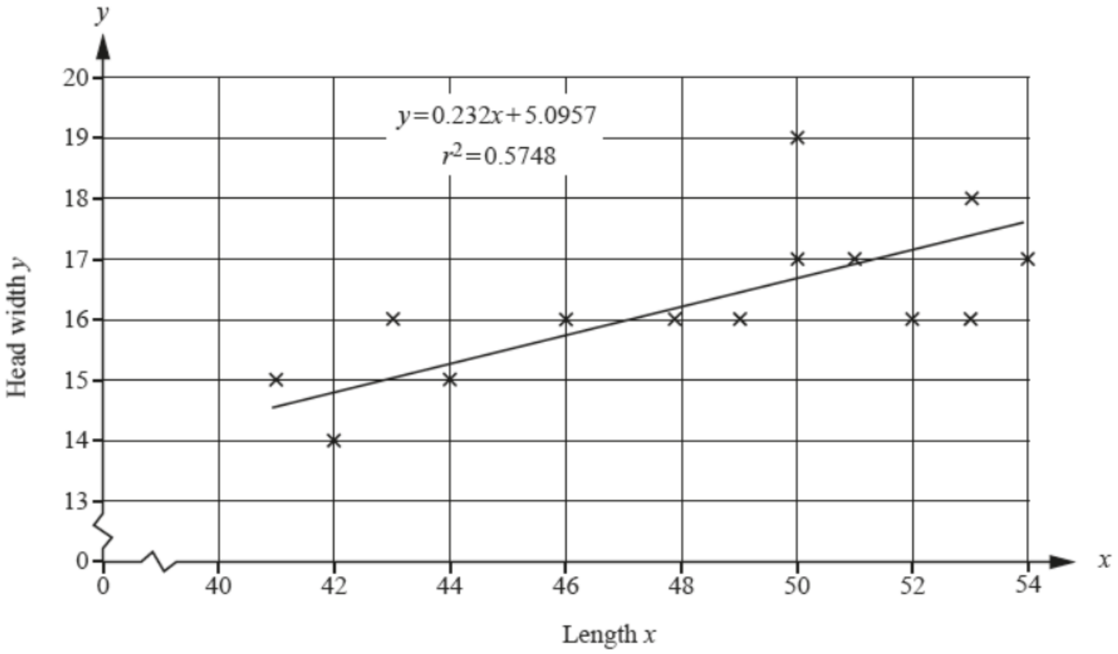


Fig. 6

(i) (*A*) Use the equation of the regression line to estimate the mean head width for frogs of each of the following lengths.

- 45 mm

$y=0.232(45)+5.0957=15.5357$

- 60 mm

$y=0.232(60)+5.0957=19.0157$

[2]

(*B*) Comment briefly on each of the estimates in part **(i)**(*A*).

$x=45$ is interpolated so it is a fairly accurate estimate. $x=60$ is extrapolated so less accurate estimate

[2]

(ii) Explain how the mean length of frogs with head width 16 mm should be estimated. **[2]**

You need to work out the equation of the regression line of x on y and substitute $y=16$ into the equation

(iii) Calculate the value of the product moment correlation coefficient. **[1]**

$PMCC=\sqrt{0.5748}=0.7582$

(iv) In the light of the information in the scatter diagram, comment on the goodness of fit of the regression line. **[2]**

As r^2 is relatively large and the points lie fairly close to a straight line, the fit is moderately good.

END OF QUESTION PAPER

BLANK PAGE