PhysicsAndMathsTutor.com

OCR
Oxford Cambridge and RSA

Accredited

AS Level Further Mathematics B (MEI)

Y412 Statistics a

Sample Question Paper

Date – Morning/Afternoon

Time allowed: 1 hour 15 minutes

OCR supplied materials:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)
- Scientific or graphical calculator

Model Answers

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.**
- Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Do **not** write in the bar codes.
- 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.

PhysicsAndMathsTutor.com

Answer **all** the questions

1 The number of failures of a machine each week at a factory is modelled by a Poisson distribution with mean 0.45. $X \sim Po(0.45)$

(i) Write down the variance of the distribution. **[1]**

$\text{Var } X = \lambda = 0.45$

(ii) Find the probability that there are exactly 2 failures in a week. **[1]**

$$P(X=2) = \frac{e^{-0.45} \times 0.45^2}{2} = 0.0646$$

(iii) State a distribution which can be used to model the number of failures in a period of 4 weeks. **[2]**

$4 \times 0.45 = 1.8$ ($\times 4$ as 4 weeks)

Poisson with $\lambda = 1.8$

(iv) Find the probability that there are at least 2 failures in a period of 4 weeks. **[2]**

$Y \sim Po(1.8)$. $P(Y \geq 2) = 1 - P(Y=0) - P(Y=1)$

$= 1 - e^{-1.8} - (e^{-1.8} \times 1.8) = 0.5372$ (4sf)

2 The discrete random variable Y is uniformly distributed over the values $\{12, 13, \ldots, 20\}$.

(i) Write down $P(Y < 15)$. **[1]**

$P(Y<15) = P(Y=12) + P(Y=13) + P(Y=14)$

$= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{1}{3}$

(ii) Two independent observations of Y are taken. Find the probability that one of these values is less than 15 and the other is greater than 15. **[3]**

$\leftarrow P(Y=15)$

$P(Y<15) = \frac{1}{3}$. $P(Y>15) = 1 - \frac{1}{3} - \frac{1}{9} = \frac{5}{9}$

$P(Y<15)P(Y>15) + P(Y>15)P(Y<15)$

$= \left(\frac{1}{3} \times \frac{5}{9}\right) + \left(\frac{1}{3} \times \frac{5}{9}\right) = \frac{10}{27}$ $\hookrightarrow$ As 2 independent observations

(iii) Find $P(Y > E(Y))$ **[2]**

$E(Y) = \frac{12+20}{2} = 16$. $P(Y>16) = \frac{4}{9}$

$\frac{1}{9} \times 4$ as [17, 18, 19, 20]

PhysicsAndMathsTutor.com

3 **In this question you must show detailed reasoning.**

A student is investigating what people think about organic food. She wishes to see if there is any difference between the opinions of females and males. She takes a random sample of 100 people and asks each of them if they think that organic food is better for their health than non-organic food. She will use the data to conduct a hypothesis test. The table below shows the opinions of these 100 people.

		Sex	
		Female	Male
Opinion on organic food	Organic better	35	18
	Not better	22	25

(i) Explain why the student should use a random sample. **[2]**

A random sample allows a proper inference about the population to be undertaken

(ii) Carry out a test at the 5% significance level to examine whether there is any association between a person's sex and their opinion on organic food. Show your calculations. **[8]**

H_0: There is no association between a person's sex and their opinion on organic food

H_1: There is an association between a person's sex and their opinion on organic food

observed

f_o	Female	Male	Total
Organic better	35	18	53
Not better	22	25	47
Total	57	43	100

expected frequency

f_e	Female	Male
Organic better	30.21	22.79
Not better	26.79	20.21

$\frac{47 \times 57}{100}$ $\frac{43 \times 47}{100}$

Contribution	Female	Male
Organic better	0.7545	1.0068
Not better	0.8564	1.1353

$$\chi^2 = \sum \text{CONT}$$
$$= 3.76 \text{ (3sf)}$$

$$\text{CONT} = \frac{(f_o - f_e)^2}{f_e}$$

PhysicsAndMathsTutor.com

n and m are no of columns + rows in table

$v = (n-1)(m-1) = (2-1)(2-1) = 1$. Critical value at $v = 1$ and $p = 5\%$ is 3.841.
As $3.76 < 3.841$, this result is not significant. Therefore, there is insufficient evidence to suggest there is an association between sex and opinion on organic food so we can accept H_0.

4 The discrete random variable X has probability distribution defined by

$$P(X = r) = k(2r - 1) \quad \text{for } r = 1, 2, 3, 4, 5, 6, \text{ where } k \text{ is a constant.}$$

(i) Complete the table in the Printed Answer Booklet giving the probabilities in terms of k.

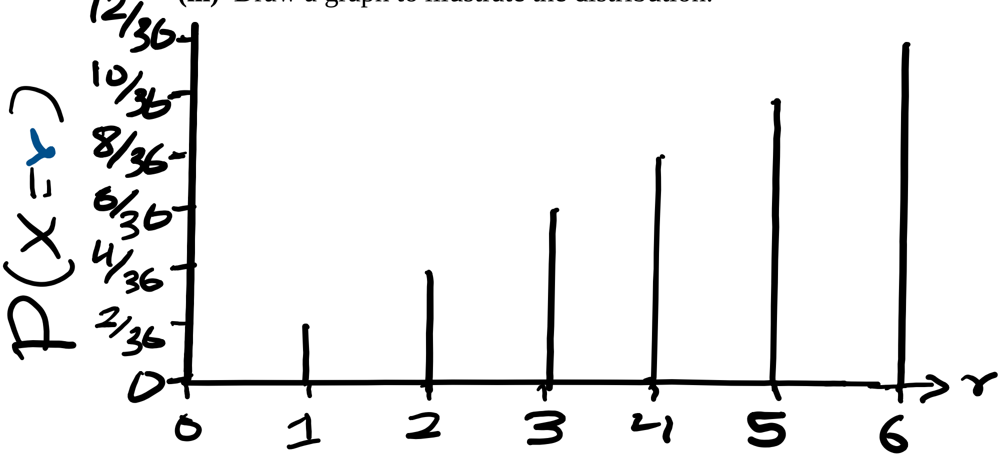
r	1	2	3	4	5	6
$P(X = r)$	k	$3k$	$5k$	$7k$	$9k$	$11k$

[1]

(ii) Show that the value of k is $\frac{1}{36}$. **[2]**

$\sum P(X=r) = 1$ so $k + 3k + 5k + 7k + 9k + 11k = 1$
$36k = 1 \Rightarrow k = \frac{1}{36}$

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



Always draw straight line graphs for discrete random variables

(iv) **In this question you must show detailed reasoning.**

Find
- $E(X)$
- $\text{Var}(X)$. **[5]**

$E(X) = \sum r P(X=r)$
$= \left(1 \times \frac{1}{36}\right) + \left(2 \times \frac{3}{36}\right) + \left(3 \times \frac{5}{36}\right) + \left(4 \times \frac{7}{36}\right) + \left(5 \times \frac{9}{36}\right) + \left(6 \times \frac{11}{36}\right) = \frac{161}{36} / 4.47$

PhysicsAndMathsTutor.com

$$E(X^2) = \sum r^2 P(X=r) = \left(1^2 \times \frac{1}{36}\right) + \left(2^2 \times \frac{3}{36}\right) + \left(3^2 \times \frac{5}{36}\right) + \left(4^2 \times \frac{7}{36}\right) + \left(5^2 \times \frac{9}{36}\right) + \left(6^2 \times \frac{11}{36}\right)$$

$$= \frac{791}{36}$$

$$\operatorname{Var} X = E(X^2) - [E(X)]^2 = \frac{791}{36} - \left(\frac{161}{36}\right)^2 = \frac{2555}{1296}$$

$(=1.97)$

A game consists of a player throwing two fair dice. The score is the maximum of the two values showing on the dice.

(v) Show that the probability of a score of 3 is $\frac{5}{36}$. **[2]**

There are 5 combinations that give you a score of 3: (1, 3), (2, 3), (3, 3), (3, 2) and (3, 1). Overall, there are 36 combinations so $p = \frac{5}{36}$.

(vi) Show that the probability distribution for the score in the game is the same as the probability distribution of the random variable X. **[3]**

Score	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

You can get a score of 1 in 1 way, score of 2 in 3 ways, score of 3 in 5 ways etc. mirroring the distribution of X

(vii) The game is played three times.

Find

- the mean of the total of the three scores.
- the variance of the total of the three scores. **[3]**

$$E(X_1 + X_2 + X_3) = \frac{161}{36} + \frac{161}{36} + \frac{161}{36} = \frac{483}{36} \quad (=13.42)$$

$$\operatorname{Var}(X_1 + X_2 + X_3) = \frac{2555}{1296} + \frac{2555}{1296} + \frac{2555}{1296}$$

$$= 5.91 \text{ (3sf)}$$

PhysicsAndMathsTutor.com

5 In a recent report, it was stated that 40% of working people have a degree. For the whole of this question, you should assume that this is true.

A researcher wishes to interview a working person who has a degree. He asks working people at random whether they have a degree and counts the number of people he has to ask until he finds one with a degree.

(i) Find he probability that he has to ask 5 people. **[2]**

$P = 0.6^4 \times 0.4 = 0.05184$ (first 4 don't have degree and 5th does)

(ii) Find the mean number of people the researcher has to ask. **[1]**

$X \sim Geo(0.4)$. $E(X) = \frac{1}{p} = \frac{1}{0.4} = 2.5$

Subsequently, the researcher decides to take a random sample from the population of working people.

(iii) A random sample of 5 working people is chosen. What is the probability that at least one of them has a degree? **[2]**

$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.6^5$
$= 0.9222$

(iv) How large a random sample of working people would the researcher need to take to ensure that the probability that at least one person has a degree is 0.99 or more? **[3]**

$P(X \geq 1) = 1 - 0.6^n$. $1 - 0.6^n = 0.99$
so $0.6^n = 0.01 \Rightarrow n = \log_{0.6} 0.01$
$= 9.015....$
so 10 people are needed

6 A motorist decides to check the fuel consumption, y miles per gallon, of her car at particular speeds, x mph, on flat roads. She carries out the check on a suitable stretch of motorway. Fig.6 shows her results.

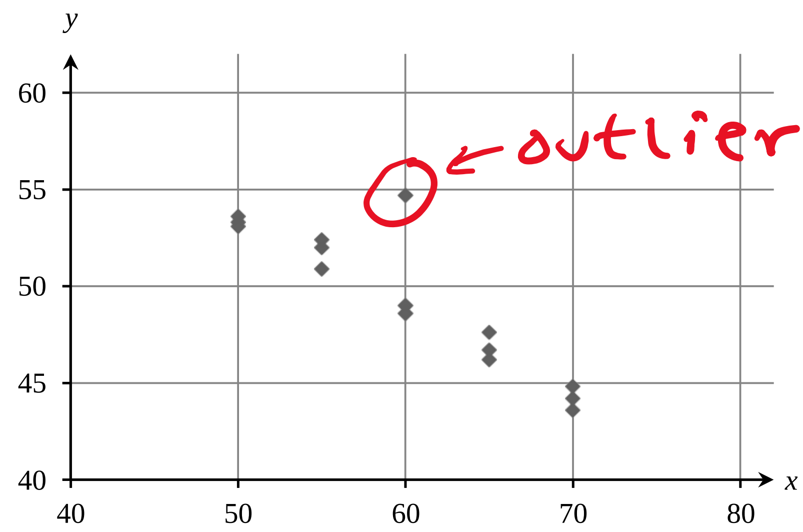


Fig.6

PhysicsAndMathsTutor.com

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

PMCC is only valid for random on random data but as speeds are controlled, its not a random variable so a PMCC test wouldn't be valid

(ii) (*A*) One of the results is an outlier. Circle the outlier on the copy of Fig. 6 in the Printed Answer Booklet. **[1]**

(*B*) Suggest one possible reason for the outlier in part **(ii)** (*A*) not being used in any analysis. **[1]**

It is not representative of the rest of the data

The motorist decides to remove this item of data from any analysis. The table below shows part of a spreadsheet that was used to analyse the 14 remaining data items (with the outlier removed). Some rows of the spreadsheet have been deliberately omitted.

	Data item	x	y	x^2	y^2	xy
	1	50	53.6	2500	2872.96	2680
	2	50	53.3	2500	2840.89	2665
	13	70	44.8	4900	2007.04	3136
	14	70	44.2	4900	1953.64	3094
Sum		840	686	51150	33779.7	40812

(iii) Calculate the equation of the regression line of y on x. **[4]**

$\bar{x} = \frac{840}{14} = 60$. $\bar{y} = \frac{686}{14} = 49$.

$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 40812 - \frac{840 \times 686}{14}$

$= -348$

$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 51150 - \frac{840^2}{14} = 750$

$b = \frac{S_{xy}}{S_{xx}} = \frac{-348}{750} = -0.464$

$y - \bar{y} = b(x - \bar{x})$ so $y - 49 = -0.464(x - 60)$

$y - 49 = -0.464x + 27.84$

$\Rightarrow y = -0.464x + 76.84$

PhysicsAndMathsTutor.com

(iv) Use the equation of the regression line to predict the fuel consumption of the car at

(A) 58 mph,

$y = -0.464(58) + 76.84 = 49.9$

(*B*) 30 mph.

$y = -0.464(30) + 76.84 = 62.9$

[2]

(v) Comment on the reliability of your predictions in part **(iv)**. **[2]**

The answer for (A) is reliable as interpolated and points lie close to line. The answer to (B) is less reliable as it is extrapolated.

END OF QUESTION PAPER