OCR
Oxford Cambridge and RSA

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Practice Paper – Set 2

Time allowed: 2 hours 15 minutes

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.
- 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 mark for this paper is **120**.
- 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 **16** pages. The Question Paper consists of **12** pages.

Section A (28 marks)

Answer **all** the questions.

1 On the northbound carriageway of a motorway, the number of defects requiring repair per 100 m is modelled by a Poisson distribution with mean 0.8.

(i) Write down the variance of the number of defects requiring repair per 100 m. **[1]**

(ii) Find the probability that there are at least 3 defects requiring repair in a randomly selected 100 m stretch of the northbound carriageway. **[2]**

(iii) Find the probability that the number of defects requiring repair in a randomly selected 100 m stretch of the northbound carriageway is no more than 1 standard deviation from the mean. **[2]**

(iv) Find the probability that there are fewer than 10 defects requiring repair in a randomly selected 1 km stretch of the northbound carriageway. **[2]**

On the southbound carriageway of the same motorway, the number of defects requiring repair per 100 m is modelled by a Poisson distribution with mean 1.9.

(v) Find the probability that there are no more than 3 defects requiring repair altogether in the northbound and southbound carriageways of a randomly selected 100 m stretch of the motorway. **[2]**

(vi) State an assumption necessary in order to calculate the probability in part **(v)**. **[1]**

2 In a block of holiday apartments, each apartment can be rented for any number of days between 3 and 7. Based on past records, the probability distribution of an apartment being rented for X days is as shown in the table below.

r	3	4	5	6	7
$P(X = r)$	0.18	0.23	0.15	0.07	0.37

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

(ii) Comment briefly on the shape of the distribution. **[1]**

(iii) Find each of the following.

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

The cost of renting an apartment is £320 for three days, plus an additional £80 for each extra day for which the apartment is rented.

(iv) (*A*) Find the expected total amount of money paid by somebody renting an apartment. **[2]**

(*B*) Find the standard deviation of the total amount of money paid by somebody renting an apartment. **[1]**

3 A researcher for a website which sells second-hand cars is investigating whether there is any association between the sex of a customer and their 'loyalty' to the website (i.e. whether or not they have previously bought a car on the website). The researcher decides to carry out a hypothesis test and so selects a random sample of 210 customers. The table below shows the data.

		Loyalty	
		Bought previously	New customer
Sex	Female	18	66
	Male	40	86

(i) Explain the advantage of the researcher's sample being a *random* sample of customers. **[2]**

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

Carry out the test at the 10% significance level to examine whether there is any association between sex and loyalty. **[8]**

Section B (92 marks)

Answer **all** the questions.

4 A machine in a factory produces components of a particular type. It is known that on average 4% of components are defective and that defective components occur independently of each other. An engineer checks components, one at a time, until a defective component is found.

(i) Find the probability that the engineer has to check exactly 4 components. **[2]**

(ii) Find the probability that the engineer has to check more than 20 components. **[1]**

(iii) Find the mean number of components that the engineer has to check. **[1]**

Each week the engineer has to examine 3 defective components.

(iv) Find the probability that at least 100 components have to be checked in order to find 3 defective ones. **[3]**

(v) Find the probability that in a period of 10 weeks, there are no more than 4 weeks in which at least 100 components have to be checked in order to find 3 defective ones. **[2]**

(vi) Find the expected number of weeks in a period of 10 weeks in which at least 100 components have to be checked in order to find 3 defective ones. **[2]**

5 The ideal brewing temperature for a particular type of tea is 84 °C. A manager of a café wishes to check whether, on average, employees are brewing this type of tea at the correct temperature. The temperature of the water used for brewing the tea was measured on 9 randomly chosen occasions, with the following results.

85.1	85.4	85.6	85.5	85.6	86.2	84.2	83.7	82.7

The manager uses software to produce a 95% confidence interval for the mean brewing temperature. The output from the software is shown in Fig. 5.

T Estimate of a Mean

Confidence Level 0.95

Sample

Mean	84.889
s	1.123
N	9

Result

T Estimate of a Mean

Mean	84.889
s	1.123
SE	0.3743
N	9
df	8
Interval	84.889 ± 0.8632

Fig. 5

(i) State an assumption necessary for the use of the t distribution in the construction of this confidence interval. **[1]**

(ii) State the confidence interval which the software gives in the form $a < \mu < b$. **[1]**

(iii) Explain what the symbol μ in part **(ii)** represents. **[2]**

(iv) Explain whether the confidence interval suggests that the average brewing temperature is 84 °C. **[2]**

(v) State how, using this sample, the manager could produce a wider confidence interval. **[1]**

6 A student wonders if there is any correlation between the level of carbon dioxide emissions x g/km produced by a car and its noise level y (in suitable units). She chooses a random sample of 10 cars and plots the scatter diagram shown in Fig. 6.1 using data from a government website.

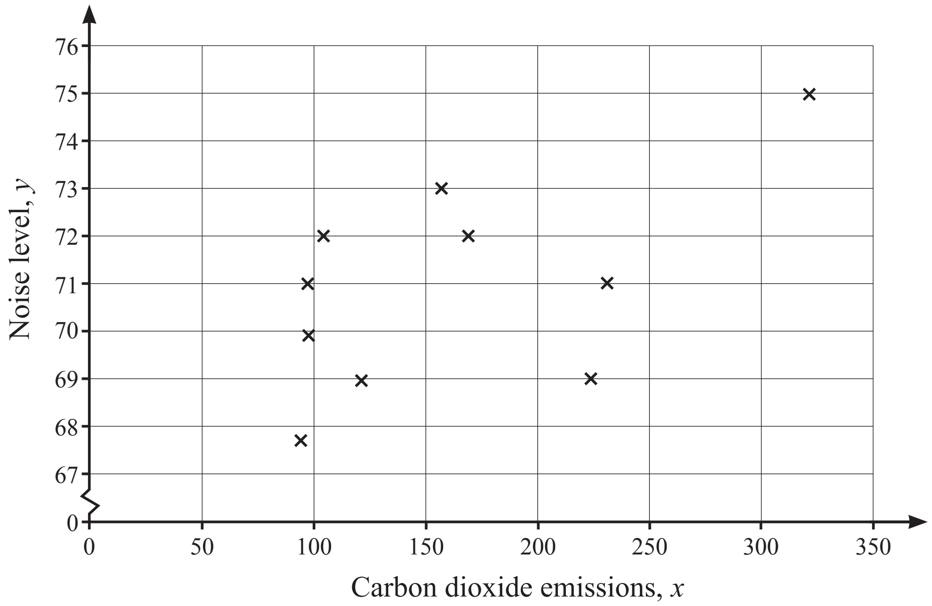

Fig. 6.1

(i) Explain why it may be appropriate to carry out a test based on the product moment correlation coefficient. **[2]**

Fig. 6.2 shows part of a spreadsheet used to analyse the data. Some rows of the spreadsheet have been deliberately omitted.

	A	B	C	D	E	F
1		***x***	***y***	***x*²**	***y*²**	***xy***
2		169	72	28561	5184	12168
3						
4						
9						
10		321	75	103041	5625	24075
11		94	67.7	8836	4583.29	6363.8
12	**Sum**	1615	709.6	312647	50395.3	115430
13						

Fig. 6.2

(ii) Calculate the product moment correlation coefficient. **[4]**

(iii) Carry out a hypothesis test at the 5% significance level to investigate whether there is any correlation between carbon dioxide emissions and noise level. **[5]**

(iv) A subsequent investigation, involving data for over 5000 cars, resulted in a lower estimate for the product moment correlation coefficient between carbon dioxide emissions and noise level. Explain why using a much larger sample provides a much more reliable result. **[2]**

The student then investigates fuel consumption of cars within towns and outside of towns. The 'urban' (within towns) fuel consumption of a car is denoted by u and the 'extra-urban' (outside of towns) fuel consumption is denoted by v. Both u and v are measured in miles per gallon (mpg). Fig. 6.3 shows two scatter diagrams for the two measures of fuel consumption for the same random sample of 24 cars. Also shown are the equations of the regression lines for v on u and for u on v together with the value of r^2.

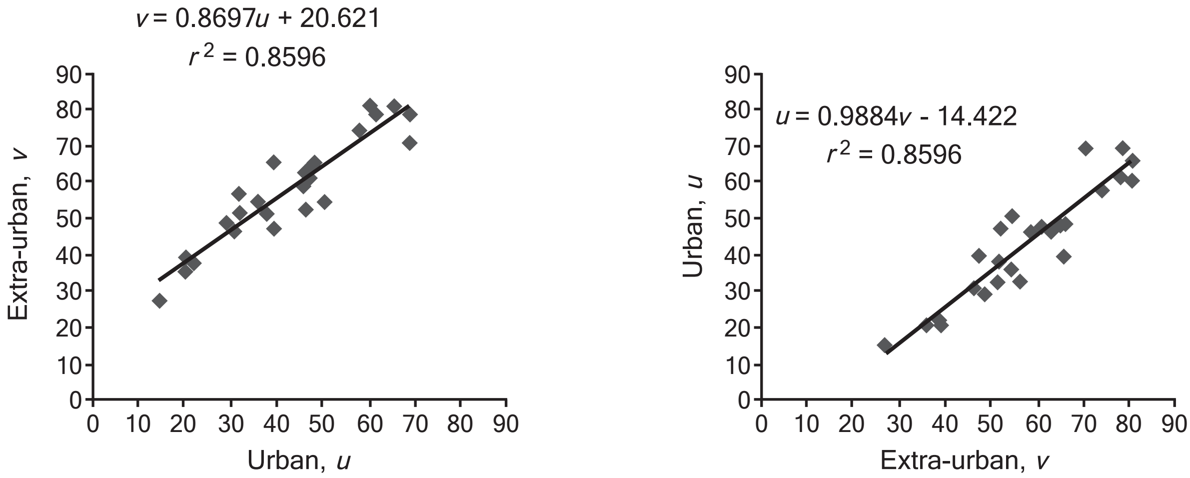

Fig. 6.3

(v) Using the appropriate equation, estimate the extra-urban fuel consumption of a car whose urban fuel consumption is 55 mpg. **[2]**

(vi) Comment on the reliability of your estimate. **[2]**

(vii) Explain the purpose of the two separate regression lines. **[1]**

(viii) Find the coordinates of the point at which the two regression lines would meet if they were plotted on a single diagram. **[1]**

(ix) State what the point found in part **(viii)** represents. **[1]**

7 A particular brand of perfume has an ethyl alcohol base. In a 90 ml bottle there should be 62 ml of ethyl alcohol. A quality control operative selects a random sample of 40 bottles from the production line in order to check the amount of ethyl alcohol that is present. The volumes of ethyl alcohol, x ml, in the 40 bottles are summarised as follows.

$$\sum x = 2497 \qquad \sum x^2 = 155\,970$$

(i) Calculate estimates of

- the population mean,
- the population variance. **[3]**

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

Carry out a test at the 5% significance level to investigate whether the average volume of ethyl alcohol is 62 ml. **[8]**

(iii) Explain why you can carry out this test despite the distribution of the parent population being unknown and the population variance also being unknown. **[3]**

8 A machine manufactures aluminium sheets specified as being 1 mm thick. In fact the thickness of the sheets is Normally distributed with mean 1.01 mm and standard deviation 0.015 mm. You should assume that each sheet is of uniform thickness and that the thicknesses of different sheets are independent of each other.

(i) Find the probability that the total thickness of 5 sheets is less than 5.0 mm. **[3]**

(ii) Find the probability that, if a single sheet is cut into pieces and 5 of the pieces are stacked together, the total thickness of the stack is less than 5.0 mm. **[2]**

The machine also manufactures aluminium sheets specified as being 2 mm thick. The thickness of these sheets is Normally distributed with mean 2.008 mm and standard deviation 0.019 mm.

(iii) Find the probability that the total thickness of two 1 mm sheets is less than the thickness of one 2 mm sheet. **[4]**

9 The time in minutes between successive hits to a website is modelled by the random variable T with cumulative distribution function given by

$$\mathrm{F}(t) = \begin{cases} 0 & t < 0, \\ k\ln(t+1) & 0 \leqslant t \leqslant 8, \\ 1 & t > 8, \end{cases}$$

where k is a constant.

(i) (*A*) Show that the median value of T is 2. **[3]**

(*B*) Find the upper quartile of T. **[2]**

(ii) Find the probability density function of T. **[2]**

(iii) Sketch the graph of the probability density function of T. **[2]**

(iv) Find $\mathrm{P}(T - \mu > \sigma)$, where μ and σ are the mean and standard deviation of T respectively. **[7]**

10 The discrete random variable X is uniformly distributed over the values $\{0, 1, 2, \ldots, 10\}$. The continuous random variable Y is uniformly distributed over the interval $[0, 10]$.

(i) Write down the mean value of $X - Y$. **[1]**

(ii) Find the variance of $X - Y$. **[4]**

The spreadsheet in Fig. 10 is used to simulate the distributions of X and Y. Each of the 25 rows below the heading row consists of a value of X, a value of Y and the value of $X - Y$.

	A	B	C
1	***X***	***Y***	***X – Y***
2	7	9.1823	-2.1823
3	5	1.2598	3.7402
4	3	8.1959	-5.1959
5	8	2.1379	5.8621
6	8	9.7434	-1.7434
7	6	0.7244	5.2756
8	9	6.6281	2.3719
9	1	8.2534	-7.2534
10	7	7.8500	-0.8500
11	5	7.4100	-2.4100
12	4	5.4082	-1.4082
13	3	3.6765	-0.6765
14	10	9.0232	0.9768
15	7	4.0283	2.9717
16	4	2.4132	1.5868
17	5	9.1441	-4.1441
18	2	6.2759	-4.2759
19	2	7.3061	-5.3061
20	2	8.8499	-6.8499
21	4	0.3194	3.6806
22	8	0.2698	7.7302
23	4	6.6746	-2.6746
24	6	5.2149	0.7851
25	5	2.9266	2.0734
26	10	5.7717	4.2283

Fig. 10

(iii) Use the spreadsheet to estimate the following.

- $P(X - Y > 0)$
- $P(X - Y > 1)$ **[2]**

(iv) How could the estimates in part **(iii)** be improved? **[1]**

The mean of 100 values of $X - Y$ is denoted by the random variable Z.

(v) Write down $P(Z > 0)$. **[1]**

(vi) Calculate an estimate of $P(Z > 1)$. **[3]**

Two randomly selected values of X are denoted by X_1 and X_2. The random variable W represents the mean of 100 values of $X_1 - X_2$.

(vii) Calculate an estimate of $P(W > 1)$. **[3]**

END OF QUESTION PAPER

BLANK PAGE

BLANK PAGE