OCR
Oxford Cambridge and RSA

Thursday 08 October 2020 – Afternoon

AS Level Further Mathematics A

Y532/01 Statistics

Time allowed: 1 hour 15 minutes

831567552 0

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for AS Level Further Mathematics A
- 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**. You can use extra paper if you need to, but you must clearly show your candidate number, the centre number and the question numbers.
- 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 non-exact numerical answers correct to 3 significant figures unless a different degree of accuracy is specified in the question.
- The acceleration due to gravity is denoted by $g\,\mathrm{m\,s^{-2}}$. When a numerical value is needed use $g = 9.8$ unless a different value is specified in the question.
- 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 Five observations of bivariate data (x, y) are given in the table.

x	7	8	12	6	4
y	20	16	7	17	23

(a) Find the value of Pearson's product-moment correlation coefficient. **[2]**

(b) State what your answer to part **(a)** tells you about a scatter diagram representing the data. **[2]**

(c) A new variable a is defined by $a = 3x + 4$. Dee says "The value of Pearson's product-moment correlation coefficient between a and y will not be the same as the answer to part **(a)**."

State with a reason whether you agree with Dee. **[1]**

2 Every time a spinner is spun, the probability that it shows the number 4 is 0.2, independently of all other spins.

(a) A pupil spins the spinner repeatedly until it shows the number 4.

Find the mean of the number of spins required. **[2]**

(b) Calculate the probability that the number of spins required is between 3 and 10 inclusive. **[2]**

(c) Each pupil in a class of 30 spins the spinner until it shows the number 4. Out of the 30 pupils, the number of pupils who require at least 10 spins is denoted by X.

Determine the variance of X. **[4]**

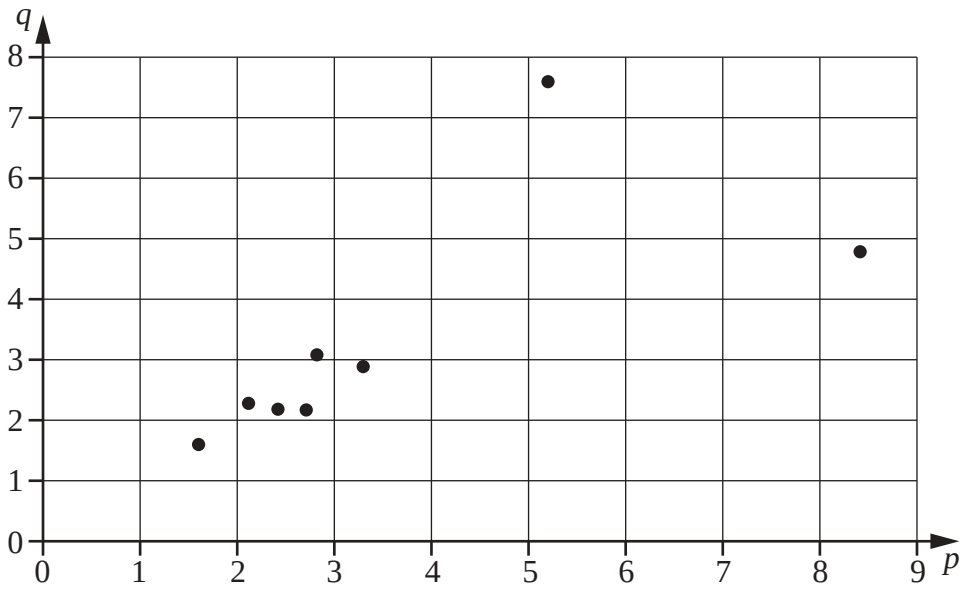
3 An investor obtains data about the profits of 8 randomly chosen investment accounts over two one-year periods.

The profit in the first year for each account is p% and the profit in the second year for each account is q%.

The results are shown in the table and in the scatter diagram.

Account	A	B	C	D	E	F	G	H
p	1.6	2.1	2.4	2.7	2.8	3.3	5.2	8.4
q	1.6	2.3	2.2	2.2	3.1	2.9	7.6	4.8

$n = 8$ $\quad \Sigma p = 28.5$ $\quad \Sigma q = 26.7$ $\quad \Sigma p^2 = 136.35$ $\quad \Sigma q^2 = 116.35$ $\quad \Sigma pq = 116.70$



(a) State which, if either, of the variables p and q is independent. **[1]**

(b) Calculate the equation of the regression line of q on p. **[3]**

(c) **(i)** Use the regression line to estimate the value of q for an investment account for which $p = 2.5$. **[1]**

(ii) Give two reasons why this estimate could be considered reliable. **[2]**

(d) Comment on the reliability of using the regression line to predict the value of q when $p = 7.0$. **[2]**

4 After a holiday organised for a group, the company organising the holiday obtained scores out of 10 for six different aspects of the holiday. The company obtained responses from 100 couples and 100 single travellers. The total scores for each of the aspects are given in the following table.

Aspect	Couples	Single travellers
Organisation	884	867
Travel	710	633
Food	692	675
Leader	898	898
Included visits	561	736
Optional visits	683	712

Fred wishes to test whether there is significant positive correlation between the scores given by the two categories.

(a) Explain why it is probably not appropriate to use Pearson's product-moment correlation coefficient. **[1]**

(b) Carry out an appropriate test at the 1% level. **[7]**

(c) Explain what is meant by the statement that the test carried out in part **(b)** is a non-parametric test. **[1]**

5 At a cinema there are three film sessions each Saturday, "early", "middle" and "late". The numbers of the audience, in different age groups, at the three showings on a randomly chosen Saturday are given in **Table 1**.

Observed frequencies		Session		
		Early	Middle	Late
Age group	< 25	24	20	40
	25 to 60	4	2	10
	> 60	28	22	10

Table 1

The cinema manager carries out a test of whether there is any association between age group and session attended.

(a) Show that it is necessary to combine cells in order to carry out the test. **[2]**

It is decided to combine the second and third rows of the table. Some of the expected frequencies for the table with rows combined, and the corresponding contributions to the χ^2 test statistic, are shown in the following incomplete tables.

Expected frequencies		Session		
		Early	Middle	Late
Age group	< 25	29.4	23.1	
	$\geqslant 25$	26.6	20.9	

Table 2

Contribution to χ^2		Session		
		Early	Middle	Late
Age group	< 25	0.9918	0.4160	
	$\geqslant 25$	1.0962	0.4598	

Table 3

(b) In the Printed Answer Booklet, complete both tables. **[3]**

(c) Carry out the test at the 5% significance level. **[5]**

(d) Use the figures in your completed **Table 3** to comment on the numbers of the audience in different age groups. **[2]**

6 A statistician investigates the number, F, of signal failures per week on a railway network.

(a) The statistician assumes that signal failures occur randomly.

Explain what this statement means. **[1]**

(b) State two further assumptions needed for F to be well modelled by a Poisson distribution. **[2]**

In a random sample of 50 weeks, the statistician finds that the mean number of failures per week is 1.61, with standard deviation 1.28.

(c) Explain whether this suggests that F is likely to be well modelled by a Poisson distribution. **[2]**

Assume first that $F \sim \text{Po}(1.61)$.

(d) Write down an exact expression for $\text{P}(F = 0)$. **[1]**

(e) Complete the table in the Printed Answer Booklet to show the probabilities of different values of F, correct to three significant figures. **[2]**

Value of F	0	1	$\geqslant 2$
Probability	0.200		

After further investigation, the statistician decides to use a different model for the distribution of F. In this model it is now assumed that $\text{P}(F = 0)$ is still 0.200, but that if one failure occurs, there is an increased probability that further failures occur.

(f) Explain the effect of this assumption on the value of $\text{P}(F = 1)$. **[2]**

7 A bag contains $2m$ yellow and m green counters. Three counters are chosen at random, without replacement. The probability that exactly two of the three counters are yellow is $\frac{28}{55}$.

Determine the value of m. **[7]**

END OF QUESTION PAPER

BLANK PAGE