OCR
Oxford Cambridge and RSA

Friday 17 May 2024 – Afternoon

AS Level Further Mathematics A

Y532/01 Statistics

Time allowed: 1 hour 15 minutes

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for AS Level Further Mathematics A
- a scientific or graphical calculator

QP

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 $g\,\text{m}\,\text{s}^{-2}$. 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.

1 The random variable W can take values 1, 2 or 3 and has a discrete uniform distribution.

(a) Write down the value of $\mathrm{E}(2W)$. **[1]**

(b) Find the value of $\mathrm{Var}(2W)$. **[2]**

(c) Determine the value of the constant k for which $\mathrm{E}(2W+k) = \mathrm{Var}(2W+k)$. **[2]**

The random variable S has the probability distribution shown in the following table.

s	2	3	4	5	6
$\mathrm{P}(S = s)$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{2}{9}$

(d) Calculate $\mathrm{Var}(S)$. **[3]**

2 For a random sample of 160 employees of a large company, the principal method of transport for getting to work, arranged according to grade of employee, is shown in the table.

Grade	Walk or cycle	Private motorised transport	Public transport
A	9	13	6
B	16	43	41
C	11	8	13

A test is carried out at the 5% significance level of whether there is association between grade of employee and method of transport.

(a) State appropriate hypotheses for the test. **[1]**

The contributions to the test statistic are shown in the following table, correct to 3 decimal places.

Grade	Walk or cycle	Private motorised transport	Public transport
A	1.157	0.289	1.929
B	1.878	0.225	0.327
C	2.006	1.800	0.083

(b) Show how the value 0.225 is obtained. **[3]**

(c) Complete the test, stating the conclusion. **[3]**

(d) Which combination of grade of employee and method of transport most strongly suggests association? Justify your answer. **[1]**

3 The ages, x years, and the reaction time, t seconds, in an experiment carried out on a sample of 15 volunteers are summarised as follows.

$n = 15 \qquad \Sigma x = 762 \qquad \Sigma t = 8.7 \qquad \Sigma x^2 = 44\,204 \qquad \Sigma t^2 = 5.65 \qquad \Sigma xt = 490.1$

(a) Calculate the value of the product moment correlation coefficient between x and t. **[2]**

(b) Calculate the equation of the line of regression of t on x. Give your answer in the form $t = a + bx$ where a and b are constants to be determined. **[2]**

(c) Explain the relevance of the quantity $\Sigma(t - a - bx)^2$ to your answer to part **(b)**. **[1]**

(d) Estimate the reaction time, in seconds, for a volunteer aged 42. **[1]**

It is subsequently decided to measure the reaction time in tenths of a second rather than in seconds (so, for example, a time of 0.6 seconds would now be recorded as 6).

(e) (i) State what effect, if any, this change would have on your answer to part **(a)**. **[1]**

(ii) State what effect, if any, this change would have on your answer to part **(b)**. **[1]**

It is known that the sample of 15 volunteers consisted almost entirely of students and retired people.

(f) Using this information, and the value of the product moment correlation coefficient, comment on the reliability of your estimate in part **(d)**. **[3]**

4 In this question you must show detailed reasoning.

The random variables X and Y denote the number of telephone calls and the number of e-mails, respectively, received by a company in a randomly chosen one-minute period in a working day.

For any one-minute period, the following assumptions can be made.

- Telephone calls are received randomly and independently of one another.
- E-mails are received randomly and independently of one another.
- The average rate at which telephone calls are received is constant.
- The average rate at which e-mails are received is constant.
- Telephone calls and e-mails are received independently of one another.

It is known that $\mathrm{E}(X) = 3$.

(a) Find the probability that 4 telephone calls are received in a randomly chosen one-minute period. **[1]**

(b) A sample of 10 independent observations of X is obtained.

Find the expected number of these 10 observations that are in the interval $2 < X < 8$. **[3]**

It is also known that

$$\mathrm{P}(X + Y = 4) = \tfrac{27}{8}\mathrm{P}(X = 2) \times \mathrm{P}(Y = 2).$$

(c) Determine the possible values of $\mathrm{E}(Y)$. **[7]**

(d) Explain where in your solution to part **(c)** you have used the assumption that telephone calls and e-mails are received independently of one another. **[1]**

5 In a fashion competition, two judges gave marks to a large number of contestants.

The value of Spearman's rank correlation coefficient, r_s, between the marks given to 7 randomly chosen contestants is $\frac{27}{28}$.

(a) An excerpt from the table of critical values of r_s is shown below.

Critical values of Spearman's rank correlation coefficient

	1-tail test	5%	2.5%	1%	0.5%
	2-tail test	10%	5%	2%	1%
n	6	0.8286	0.8857	0.9429	1.0000
	7	0.7143	0.7857	0.8929	0.9286
	8	0.6429	0.7381	0.8333	0.8810

Test whether there is evidence, at the 1% significance level, that the judges agree with each another. **[4]**

The marks given by the two judges to the 7 randomly chosen contestants were as follows, where x is an integer.

Contestant	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Judge 1	64	65	67	78	79	80	86
Judge 2	61	63	78	80	81	90	x

(b) Use the value $r_s = \frac{27}{28}$ to determine the range of possible values of x. **[4]**

(c) Give a reason why it might be preferable to use the product moment correlation coefficient rather than Spearman's rank correlation coefficient in this context. **[1]**

6 Anika walks along a street that contains parked cars. The number of cars that Anika passes, up to and including the first car that is white, is denoted by X.

(a) State **two** assumptions needed for X to be well modelled by a geometric distribution. **[2]**

Assume now that X can be well modelled by the distribution Geo(p), where $0 < p < 1$.

(b) For $p = 0.1$, find $\mathrm{P}(X > 6)$. **[2]**

The number of cars that Anika passes, up to **but not including** the first car that is white, is denoted by Y.

(c) For a general value of p, determine a simplified expression for $\mathrm{E}(Y) \div \mathrm{Var}(Y)$, in terms of p. **[3]**

Ben walks along a different street that also contains parked cars. The number of cars that Ben passes, up to and including the first white car **on which the last digit of the number plate is even** is denoted by Z.

It may be assumed that Z can be well modelled by the distribution Geo$(\frac{1}{2}p)$, where p is the parameter of the distribution of X.

It is given that $\mathrm{P}(Z = 3) = k\mathrm{P}(X = 3)$, where k is a positive constant.

(d) Determine the range of possible values of k. **[5]**

END OF QUESTION PAPER