OCR

RECOGNISING ACHIEVEMENT

General Certificate of Education
Advanced Subsidiary (AS) and Advanced Level

MATHEMATICS **S4**

Probability & Statistics 4

Additional materials:
Answer paper
Graph paper
List of Formulae

SPECIMEN PAPER

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

Write your name, Centre number and candidate number in the spaces provided on the answer paper.

Answer **all** the questions.

Give non-exact numerical answers correct to 3 significant figures, unless a different degree of accuracy is specified in the question or is clearly appropriate.

You are permitted to use a graphic calculator in this paper.

INFORMATION FOR CANDIDATES

The number of marks is given in brackets [ ] at the end of each question or part question.

The total number of marks for this paper is 60.

Questions carrying smaller numbers of marks are printed earlier in the paper, and questions carrying larger numbers of marks later in the paper.

You are reminded of the need for clear presentation in your answers.

This question paper consists of 4 printed pages.

1 To compare the effect of a new drug on men and women, a small-scale trial was conducted, involving 6 randomly chosen men and 4 randomly chosen women. For each person, the time, in minutes, before the drug began to take effect was recorded, with the results shown in the table below.

Men	19	32	24	33	31	28
Women	14	13	18	27		

Use a suitable non-parametric test to determine if there is evidence, at the 5% significance level, that the drug acts more quickly in women than it does in men. **[5]**

2 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where λ is a positive constant. Show that the moment generating function of X is given by

$$M_X(t) = \frac{\lambda}{\lambda - t}. \quad \textbf{[3]}$$

Use this moment generating function to find the mean and variance of X. **[4]**

3 Events A and B are such that

$$P(A) = \tfrac{1}{2}, \qquad P(B) = \tfrac{1}{3}, \qquad P(A \cup B) = \tfrac{3}{4}.$$

(i) Determine, giving your reasons clearly, whether A and B are

(a) mutually exclusive, **[1]**

(b) independent. **[3]**

(ii) Find the values of

(a) $P(A \mid B)$, **[2]**

(b) $P(A \mid B')$, where B' denotes the complement of B. **[3]**

4 A bin contains a large number of seeds of which 20% will produce a plant with a red flower and 30% will produce a plant with a yellow flower. The remaining 50% will fail to germinate. Two seeds are chosen at random from the bin and planted in a pot. The random variables R and Y denote the number of red and yellow flowers, respectively, that will be produced from the seeds. The table below shows the joint probability distribution of R and Y.

		R		
		0	1	2
Y	0	0.25	0.20	0.04
	1	0.30	0.12	0
	2	0.09	0	0

(i) Find the marginal distributions of R and Y. **[2]**

(ii) Show that $\mathrm{E}(R) = 0.4$ and find $\mathrm{E}(Y)$. **[2]**

(iii) Find $\mathrm{Cov}(R, Y)$. **[3]**

(iv) Find the distribution of Y conditional on $R = 1$, and hence state the expected value of the number of yellow flowers in a pot in which there is one red flower. **[2]**

5 The discrete random variable X denotes the score obtained in a single throw of an ordinary fair die. Show that the probability generating function of X may be expressed as

$$\frac{t(1-t^6)}{6(1-t)}. \quad \textbf{[2]}$$

Write down the probability generating function for the total score obtained when three fair dice are thrown. **[1]**

Hence show that the probability of obtaining a total score of 10 when three fair dice are thrown is $\frac{1}{8}$. **[6]**

6 Explain briefly the circumstances under which a non-parametric test of significance should be used in preference to a parametric test. **[1]**

The acidity of soil can be measured by its pH value. As a part of a Geography project a student measured the pH values of 14 randomly chosen samples of soil in a certain area, with the following results.

5.67 5.73 6.64 6.76 6.10 5.41 5.80 6.52 5.16 5.10 6.71 5.89 5.68 5.37

Use a suitable non-parametric test to test whether the average pH value for soil in this area is 6.24. Use a 10% level of significance. **[4]**

Some time later, the pH values of soil samples taken at exactly the same locations as before were again measured. It was found that, for 3 of the 14 locations, the new pH value was higher than the previous value, while for the other 11 locations the new value was lower. Test, at the 5% significance level, whether there is evidence that the average pH value of soil in this area is lower than previously. **[4]**

7 The continuous random variable X has a uniform distribution on the interval $0 \leq x \leq a$, where the value of the parameter a is unknown. Three independent observations, X_1, X_2, X_3, of X are taken.

(i) An estimator θ is defined by $\theta = \frac{2}{3}(X_1 + X_2 + X_3)$. Show that θ is an unbiased estimator of a, and find $\mathrm{Var}(\theta)$ in terms of a. **[6]**

(ii) Another estimator ϕ is based on the greatest of the three values X_1, X_2, X_3. Denoting the greatest of the three values by the variable G, use the fact that, for any value x between 0 and a,

$$G < x \Leftrightarrow (X_1 < x \text{ and } X_2 < x \text{ and } X_3 < x)$$

to write down the cumulative distribution function of G, and hence to obtain the probability density function of G. **[2]**

Hence show that, if $\phi = \frac{4}{3}G$, then ϕ is an unbiased estimator of a, and determine which of θ and ϕ is the more efficient estimator. **[4]**

1	H_0 : same distributions of times for men and women, H_1 : lower average for women	B1		For both NH and AH
	Ranks are: M 4 9 5 10 8 7 W 2 1 3 6	M1		For ranking all 10 values
	Test statistic is $1+2+3+6=12$	A1		For sum of women's ranks
	Critical Wilcoxon rank-sum value is 13	M1		For comparing with correct tabular value
	Hence reject H_0 and conclude that there is evidence to suggest that drug acts more quickly in women than it does in men	✓		
		A1	5	
2	$M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,dx$	B1		Correct integral stated
	$= \left[\frac{\lambda}{t-\lambda}e^{(t-\lambda)x}\right]_0^\infty$	M1		For correct integration method
	$= \frac{\lambda}{\lambda - t}$	A1	3	Given answer correctly shown
	$M_X(t) = 1 + \frac{t}{\lambda} + \left(\frac{t}{\lambda}\right)^2 + \ldots$	B1		Three correct terms
	$E(X) = \frac{1}{\lambda}$	B1		
	$\text{Var}(X) = \frac{2!}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$	M1		For using correct variance formula
		A1	4	Correct answer for variance
3	(i) (a) Not exclusive, since $\frac{1}{2}+\frac{1}{3} \neq \frac{3}{4}$	B1	1	Conclusion and reason both required
	(b) $P(A \cap B) = \frac{3}{4} - \frac{1}{2} - \frac{1}{3} = \frac{1}{12}$	M1		Use of correct formula
		A1		Correct value $\frac{1}{12}$
	Not independent, as $\frac{1}{2} \times \frac{1}{3} \neq \frac{1}{12}$	A1✓	3	Conclusion and reason both required
	(ii) (a) $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{4}$	M1		Use of correct formula
		A1✓	2	Follow answer to (i)(b)
	(b) $P(A \cap B') = P(A) - P(A \cap B)$	M1		Or equivalent, e.g. from Venn diagram
	$= \frac{1}{2} - \frac{1}{12} = \frac{5}{12}$	A1✓		
	$P(A \mid B') = \frac{5}{12} \div \frac{2}{3} = \frac{5}{8}$	A1✓	3	

4	(i) R: $\begin{matrix}0 & 1 & 2\\ 0.64 & 0.32 & 0.04\end{matrix}$	M1		Adding rows (or columns)
	Y: $\begin{matrix}0 & 1 & 2\\ 0.49 & 0.42 & 0.09\end{matrix}$	A1	2	Both distributions correct
	(ii) $E(R) = 0\times0.64 + 1\times0.32 + 2\times0.04 = 0.4$	M1		Correct process for either mean
	$E(Y) = 0\times0.49 + 1\times0.42 + 2\times0.09 = 0.6$	A1	2	Both means correct
	(iii) $E(RY) = 0.12$	B1		
	$\text{Cov}(R, Y) = 0.12 - 0.4\times0.6$	M1		
	$= -0.12$	A1✓	3	
	(iv) Conditional distribution is: $\begin{matrix}0 & 1 & 2\\ \frac{5}{8} & \frac{3}{8} & 0\end{matrix}$	B1		
	Expected value is $\frac{3}{8}$	B1	2	
5	PGF is $\frac{1}{6}(t + t^2 + t^3 + t^4 + t^5 + t^6)$	B1		
	i.e. $\dfrac{t(1-t^6)}{6(1-t)}$	B1	2	Use of GP sum to deduce given answer
	$\dfrac{t^3(1-t^6)^3}{216(1-t)^3}$	B1	1	For cube of answer (i), in any form
	$\frac{1}{216}t^3(1-3t^6+\ldots)(1+3t+6t^2+\ldots)$	M1		For using both binomial expansions
		A1		Both correct at least as far as shown
	Terms in t^7 from product of brackets are:	M1		Picking out relevant term(s)
	$-3t^6\times3t$	A1		
	$1\times36t^7$	A1		
	Probability $= \frac{1}{216}(36-9) = \frac{1}{8}$	A1	6	Given answer correctly shown
6	A non-parametric test is needed when there is no information (or reasonable assumption) available about an underlying distribution	B1	1	
	Deviations from NH value 6.24 are: $-0.57\ \ -0.51\ \ 0.40\ \ 0.52\ \ -0.14\ \ -0.83\ \ -0.44$ $0.28\ \ -1.08\ \ -1.14\ \ 0.47\ \ -0.35\ \ -0.56\ \ -0.87$	M1		For calculating signed differences from 6.24
	Signed ranks are: $\begin{matrix}-10 & -7 & 4 & 8 & -1 & -11 & -5\\ 2 & -13 & -14 & 6 & -3 & -9 & -12\end{matrix}$	A1		For calculating correct signed ranks
	Test statistic is $2+4+6+8=20<25$	M1		For calculating T and comparing
	Conclude that there is evidence to suggest that the average pH value is not 6.24	A1	4	For correct conclusion based on correct work
	H_0: same average pH as before; H_1: lower value	B1		For both hypotheses
	Under H_0, tail probability for 3 or fewer out of 14 is	M1		For use of relevant binomial distribution
	$\left(\frac{1}{2}\right)^{14}\times(1+14+91+364) = 0.0287 < 0.05$	A1		For correct value tail probability 0.0287
	Hence reject H_0 and accept that average pH is lower	A1	4	For correct conclusion based on correct work

7	(i) $E(X) = \frac{1}{2}a$	B1		May be implied
	$E(\theta) = \frac{2}{3} \times 3E(X)$	M1		
	$= a$, as required	A1		For conclusion correctly shown and stated
	$\text{Var}(X) = \int_0^a \frac{x^2}{a}\,dx - \left(\frac{1}{2}a\right)^2$	M1		
	$= \frac{1}{12}a^2$	A1		
	$\text{Var}(\theta) = \frac{4}{9} \times \frac{3}{12}a^2 = \frac{1}{9}a^2$	A1✓	6	
	(ii) $P(G < x) = \left(\frac{x}{a}\right)^3$	B1		
	Hence pdf is $\frac{3x^2}{a^3}$	B1	2	
	$E(\phi) = \frac{4}{3}\int_0^a \frac{3x^3}{a^3}\,dx$	M1		
	$= a$, as required	A1		For conclusion correctly shown and stated
	$\text{Var}(\phi) = \frac{16}{9}\int_0^a \frac{3x^4}{a^3}\,dx - a^2$	M1		
	$= \frac{1}{15}a^2 < \frac{1}{9}a^2$			
	Hence ϕ is more efficient	A1	4	Correct $\text{Var}(\phi)$ and conclusion