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

MATHEMATICS S3

Probability & Statistics 3

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 A test of the null hypothesis of independence of two characteristics is required for the following contingency table.

	Characteristic 1			Totals
Characteristic 2	12	9	19	40
	28	61	71	160
Totals	40	70	90	200

(i) Find the expected frequency corresponding to the cell with observed frequency 12. **[1]**

(ii) Given that the value of the χ^2 test statistic is 4.80, correct to 3 significant figures, carry out the test at the 10% significance level. **[3]**

2 The 'customer services' section in a department store deals both with enquiries by telephone and also with enquiries in person by shoppers in the store. Telephone enquiries occur at random times at an average rate of 4 per half hour. Shoppers in the store arrive to make enquiries at random times at an average rate of 5 per half hour. Assuming that the two types of enquiries occur independently of each other, find the probability that a total of between 10 and 20 enquiries (inclusive) have to be dealt with in a randomly chosen half hour period. **[5]**

3 A supermarket sells 2 kg bags of new potatoes, in which the potatoes have been selected to be all roughly the same size. The potatoes used to fill the bags may be assumed to be randomly chosen items from a population in which the mass, in grams, of an individual potato is normally distributed with mean 90 and standard deviation 4.

(i) Show that the probability that the total mass of 21 of these potatoes exceeds 2 kg is very small. **[4]**

(ii) Find the probability that the total mass of 22 of these potatoes exceeds 2 kg. **[2]**

(iii) The machine filling the bags delivers potatoes one by one until a total mass of at least 2 kg is reached. Show that the bags are almost certain to contain either 22 or 23 potatoes. **[2]**

4 A random sample of six observations of the random variable X gave the following values:

$$3.2, \quad 1.8, \quad 4.0, \quad -2.1, \quad 6.1, \quad 1.7.$$

$$[\Sigma x = 14.7, \quad \Sigma x^2 = 73.99.]$$

The population mean of X is μ. Calculate a 90% confidence interval for μ,

(i) assuming that X has a normal distribution with variance 8.5,

(ii) assuming instead that X has a normal distribution with unknown variance.

[9]

5 State what distributional assumptions are necessary for it to be valid to use

(i) the two-sample t-test,

(ii) the paired-sample t-test,

to test for a difference in population means. **[3]**

Two different types of nylon fibre were tested for the amount of stretching under tension. Ten random samples of each fibre, of the same length and diameter, were stretched by applying a standard load. For Fibre 1 the increases in length, x mm, were as follows.

12.84 14.26 13.23 14.75 15.13 14.15 13.37 12.96 15.02 14.38

$$[\Sigma x = 140.09,\ \Sigma x^2 = 1969.0513.]$$

For Fibre 2 the increases in length, y mm, were as follows.

14.27 13.25 14.17 13.11 14.92 12.12 14.21 13.68 15.14 14.81

$$[\Sigma y = 139.68,\ \Sigma y^2 = 1958.9794.]$$

Assuming that any necessary conditions for the validity of your test hold, test whether the mean increase in length of the two types of fibre is different. Use a 10% significance level. **[7]**

6 Six hens are observed over a period of 20 days and the number of eggs laid each day is summarised in the following table.

Number of eggs	3	4	5	6
Number of days	2	2	10	6

Show that the mean number of eggs per day is 5. **[2]**

It may be assumed that a hen never lays more than one egg in any day. State one other assumption that needs to be made in order to consider a binomial model, with $n = 6$, for the total number of eggs laid in a day. **[1]**

Calculate the expected frequencies using a binomial model for the above data and carry out a χ^2 goodness of fit test, using a 10% significance level. **[9]**

7 The continuous random variable X has a triangular distribution with probability density function given by

$$f(x) = \begin{cases} 1+x & -1 \le x \le 0, \\ 1-x & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Show that, for $0 \le a \le 1$,

$$P(|X| \le a) = 2a - a^2 .$$ **[3]**

The random variable Y is given by $Y = X^2$. Express $P(Y \le y)$ in terms of y, for $0 \le y \le 1$, and hence show that the probability density function of Y is given by

$$g(y) = \frac{1}{\sqrt{y}} - 1, \quad \text{for } 0 < y \le 1 .$$ **[3]**

Use the density function of Y to find $E(Y)$, and show how the value of $E(Y)$ may also be obtained directly using the probability density function of X. **[4]**

Determine $E(\sqrt{Y})$. **[2]**

1 (i) $\dfrac{40\times 40}{200}=8$	B1	1	For correct answer; working not needed
(ii) 4.80 is greater than critical value 4.605	M1		Comparing with a tabular value
	A1		Using correct figure 4.605
There is evidence to suggest that the characteristics are not independent	A1√	3	
2 Model for all enquiries is Po(9)	B1		For any mention of a Poisson distribution
	M1		Summing two Poisson distributions
	A1		For statement of correct parameter 9
Probability $= 0.9996 - 0.5874$	M1		Subtracting relevant tabular values
$= 0.412$	A1√	5	
3 (i) $T_{21} \sim \mathrm{N}(1890, 336)$	M1		For normal distribution with correct mean
	A1		For variance 21×4^2 (both these first two marks may be implied by later working)
$\mathrm{P}(T_{21}>2000)=1-\Phi\left(\dfrac{2000-1890}{\sqrt{336}}\right)=1-\Phi(6)$	M1		For correct processes for relevant tail
Hence prob is very small since $\Phi(6)\approx 1$	A1√	4	Correct conclusion, based on suff large z
(ii) $\mathrm{P}(T_{22}>2000)=1-\Phi\left(\dfrac{2000-22\times 90}{\sqrt{22\times 16}}\right)$	M1		For relevant new normal calculation
$=1-\Phi(1.066)=0.143$	A1	2	For correct probability
(iii) $\mathrm{P}(T_{23}>2000)=1-\Phi\left(\dfrac{2000-2070}{\sqrt{368}}\right)=\Phi(3.649)\approx 1$	M1		For relevant calculation for 23 potatoes
Hence 23 potatoes is almost always enough	A1	2	For correct conclusion based on correct figs
4 (i) $\bar{x}=\dfrac{14.7}{6}=2.45$	B1		At any stage; may be implied
Interval is $2.45 \pm 1.645\times\sqrt{\dfrac{8.5}{6}} = 2.45 \pm 1.9579$	M1		For calculation of the form $\bar{x}\pm z\sqrt{\sigma^2/n}$
	A1		For relevant use of $z=1.645$
i.e. $0.49<\mu<4.41$	A1	4	For correct interval
(ii) $s^2=\dfrac{1}{5}\left(73.99-\dfrac{14.7^2}{6}\right)=7.595$	M1		For correct unsimplified expression for s^2
	A1		For correct value (may be implied)
Interval is $2.45 \pm 2.015\times\sqrt{\dfrac{7.595}{6}} = 2.45 \pm 2.267$	M1		For calculation of the form $\bar{x}\pm t\sqrt{s^2/n}$
	A1		For relevant use of $t=2.015$
i.e. $0.18<\mu<4.72$	A1√	5	For correct interval

5 **(i)** The distributions must be normal The distributions must have equal variances	B1 B1		
(ii) The differences must be normally distributed	B1	3	
$H_0: \mu_x - \mu_y = 0$; $H_1: \mu_x - \mu_y \neq 0$	B1		For both
$S_p^2 = \dfrac{\left(1969.0513 - \dfrac{140.09^2}{10}\right) + \left(1958.9794 - \dfrac{139.68^2}{10}\right)}{18}$	M1		Use of correct formula
$= 0.8033$	A1		May be implied
Test statistic $t = \dfrac{14.009 - 13.968}{\sqrt{0.8033\left(\frac{1}{10}+\frac{1}{10}\right)}} = 0.1023$	M1		Use of correct formula
	A1√		Correct value of t
This is less than the critical value 1.734	M1		Comparing to tabular value of t
There is not significant evidence of a difference in mean increase	A1√	7	
6 Mean $= \dfrac{3\times2+4\times2+5\times10+6\times6}{20} = 5$	M1		
	A1	2	Obtain given answer correctly
Assume that the probability of any hen laying an egg on any day is constant and that hens lay eggs independently of each other	B1	1	For either constant prob or independence
Distribution to be fitted is $B(6, \frac{5}{6})$	B1		May be implied
Expected frequencies are $f_i = 20 \times \binom{6}{i} \times \left(\frac{5}{6}\right)^i \times \left(\frac{1}{6}\right)^{6-i}$	M1		For any one calculation using correct method
$f_6 = 6.70,\ f_5 = 8.04,\ f_4 = 4.02$	A1		All three correct to at least 2dp
$f_3 = 1.07,\ f_2 = 0.16,\ f_1 = 0.02,\ f_0 = 0.00$	A1		For all four correct, or $f_{\le 3} = 1.25$ stated
Combining cells: ≤ 4 5 6 f_o 4 10 6 f_e 5.27 8.04 6.70	M1		Use of criterion $f_e < 5$ for combining
$\chi^2 = \dfrac{1.27^2}{5.27} + \dfrac{1.96^2}{8.04} + \dfrac{0.70^2}{6.70} = 0.857$	M1		
	A1√		Allow anything between 0.85 and 0.86 (inc)
This is less than the critical value 2.706	M1		Compare with tabular value
Hence there is a satisfactory fit	A1√	9	

7	$P(\lvert X\rvert < a) = P(-a < X < a)$	M1		For consideration of two areas, or equiv
	$= \int_{-a}^{0} (1+x)\,dx + \int_{0}^{a} (1-x)\,dx$	A1		Or equivalent trapezium areas
	$= \left[x + \frac{1}{2}x^2\right]_{-a}^{0} + \left[x - \frac{1}{2}x^2\right]_{0}^{a} = 2a - a^2$	A1	3	Given answer correctly shown
	$P(Y \le y) = P(X^2 \le y) = P(\lvert X\rvert \le \sqrt{y}) = 2\sqrt{y} - y$	B1		For correct expression
	Hence the pgf of Y is $\frac{d}{dy}(2\sqrt{y} - y) = \frac{1}{\sqrt{y}} - 1$	M1		For differentiation of previous expression
		A1	3	Given answer correctly shown
	$E(Y) = \int_0^1 y^{\frac{1}{2}} - y\,dy = \left[\frac{2}{3}y^{\frac{3}{2}} - \frac{1}{2}y^{\frac{1}{2}}\right]_0^1 = \frac{1}{6}$	M1		For correct integral
		A1		For answer $\frac{1}{6}$
	$E(X^2) = \int_{-1}^{0} (x^2 + x^3)\,dx + \int_0^1 (x^2 - x^3)\,dx$	M1		
	$= \left[\frac{1}{3}x^3 + \frac{1}{4}x^4\right]_{-1}^{0} + \left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_0^1 = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$	A1	4	
	$E(\sqrt{Y}) = \int_0^1 y^{\frac{1}{2}}\,g(y)\,dy = \int_0^1 (1 - y^{\frac{1}{2}})\,dy = \left[y - \frac{2}{3}y^{\frac{3}{2}}\right]_0^1 = \frac{1}{3}$	M1		For forming the correct integral
		A1	2	Correct answer