Tuesday 24 June 2014 – Morning

A2 GCE MATHEMATICS

4734/01 Probability & Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4734/01
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

3171009648

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found inside the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- 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.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 The independent random variables X and Y have Poisson distributions with parameters 16 and 2 respectively, and $Z = \frac{1}{2}X - Y$.

(i) Find $\mathrm{E}(Z)$ and $\mathrm{Var}(Z)$. **[3]**

(ii) State whether Z has a Poisson distribution, giving a reason for your answer. **[2]**

2 In a study of the inheritance of skin colouration in corn snakes, a researcher found 865 snakes with black and orange bodies, 320 snakes with black bodies, 335 snakes with orange bodies and 112 snakes with bodies of other colours. Theory predicts that snakes of these colours should occur in the ratios 9:3:3:1. Test, at the 5% significance level, whether these experimental results are compatible with theory. **[6]**

3 An athlete finds that her times for running 100 m are normally distributed. Before a period of intensive training, her mean time is 11.8 s. After the period of intensive training, five randomly selected times, in seconds, are as follows.

11.70 11.65 11.80 11.75 11.60

Carry out a suitable test, at the 5% significance level, to investigate whether times after the training are less, on average, than times before the training. **[7]**

4 Cola is sold in bottles and cans. The volume of cola in a bottle is normally distributed with mean 500 ml and standard deviation 10 ml. The volume of cola in a can is normally distributed with mean 330 ml and standard deviation 8 ml. Find the probability that the total volume of cola in 2 randomly selected bottles is greater than 3 times the volume of cola in a randomly selected can. **[7]**

5 The day before the 1992 General Election, an opinion poll showed that 37.6% of a random sample of 1731 voters intended to vote for the Conservative party.

(i) Calculate an approximate 99.9% confidence interval for the proportion of voters intending to vote Conservative. **[4]**

The actual proportion voting Conservative was above the upper limit of the confidence interval.

(ii) Give two possible reasons for this occurrence. **[2]**

(iii) What sample size would be required to produce a 99.9% confidence interval of width 0.05? **[3]**

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

$$f(x) = \begin{cases} k\sin x & 0 \leqslant x \leqslant \frac{1}{2}\pi, \\ k\left(2 - \frac{2x}{\pi}\right) & \frac{1}{2}\pi \leqslant x \leqslant \pi, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \dfrac{4}{4+\pi}$. **[4]**

(ii) Find $\mathrm{E}(X)$, correct to 3 significant figures, showing all necessary working. **[4]**

7 A random sample of 100 adults with a chronic disease was chosen. Each adult was randomly assigned to one of three different treatments. After six months of treatment, each adult was then assessed and classified as 'much improved', 'improved', 'slightly improved' or 'no change'. The results are summarised in Table 1.

	Treatment A	Treatment B	Treatment C
Much improved	12	16	4
Improved	13	12	6
Slightly improved	7	6	7
No change	5	3	9

Table 1

A χ^2 test, at the 5% significance level, is to be carried out.

(i) State suitable hypotheses. **[1]**

Combining the last two rows of Table 1 gives Table 2.

	Treatment A	Treatment B	Treatment C
Much improved	12	16	4
Improved	13	12	6
Slightly improved/ No change	12	9	16

Table 2

(ii) By considering the expected frequencies for Treatment C in Table 1, explain why it was necessary to combine rows. **[3]**

(iii) Show that the contribution to the χ^2 value for the cell 'slightly improved/no change, Treatment C' is 4.231, correct to 3 decimal places. **[3]**

You are given that the χ^2 test statistic is 10.51, correct to 2 decimal places.

(iv) Carry out the test. **[2]**

8 A random sample of 20 plots of land, each of equal area, was used to test whether the addition of phosphorus would increase the yield of corn. 10 plots were treated with phosphorus and 10 plots were untreated. The yields of corn, in litres, on a treated plot and on an untreated plot are denoted by X and Y respectively. You are given that

$$\Sigma x = 2112, \quad \Sigma y = 2008$$

You are also given that an unbiased estimate for the variance of treated plots is 87.96 and an unbiased estimate for the variance of untreated plots is 31.96, both correct to 4 significant figures.

(i) You may assume that the population variance estimates are sufficiently similar for the assumption of common variance to be made. What other assumption needs to be made for a t-test to be valid? **[1]**

(ii) Carry out a suitable t-test at the 1% significance level, to test whether the use of phosphorus increases the yield of corn. **[9]**

9 A rectangle of area A m^2 has a perimeter of 20 m and each of the two shorter sides are of length X m, where X is uniformly distributed between 0 and 2.

(i) Write down an expression for A in terms of X, and hence show that $A = 25 - (X - 5)^2$. **[3]**

(ii) Write down the probability density function of X. **[1]**

(iii) Show that the cumulative distribution function of A is

$$\mathrm{F}(a) = \begin{cases} 0 & a < 0, \\ \frac{1}{2}\left(5 - \sqrt{25 - a}\right) & 0 \leqslant a \leqslant 16, \\ 1 & a > 16. \end{cases}$$

[5]

(iv) Find the probability density function of A. **[2]**

END OF QUESTION PAPER

Question			Answer	Marks	Guidance	
1	(i)		$E(Z) = 6$ $Var(Z) = ¼(16) + 2$ $= 6$	B1 M1 A1 **[3]**		
	(ii)		No Difference between Poisson distributions is not Poisson, or Z may be fractional or negative.	B1 B1 **[2]**	Unless accompanied by a spurious reason. SC Allow B1 for 'no, you cannot subtract Poisson distributions'.	eg ft incorrect (i). Allow $Z \neq X+Y$
2			H_0: The data can be modelled by the theory H_1: The data can't be modelled by the theory. Expected values 918, 306, 306, 102 $TS = \frac{(865-918)^2}{918} + \ldots$ $= 7.43$ $TS < 7.815$, do not reject H_0 There is insufficient evidence to conclude that the data can't be modelled by the theory	B1 B1 M1 A1 M1 A1 **[6]**	For both Can be implied by 7.43 ft TS ft TS	Allow compatible. $p>0.05$ do not reject H_0 $p=0.05939$ and conclusion
3			H_0: $\mu = 11.8$, H_1: $\mu < 11.8$ $\bar{x} = 11.7$ $\hat{\sigma}^2 = 0.00625$ $TS = \frac{11.7-11.8}{\sqrt{\frac{0.00625}{5}}}$ $= -2.828$ $TS < -2.132$, reject H_0 There is sufficient evidence that the intensive	B1 B1 B1 M1 A1 M1 A1	NOT eg μ_0, μ_1 or $\bar{d} = \pm 0.1$ Allow $\frac{1}{160}$ Allow reversed if consistent. Allow -2.83 ft TS ft TS Contextualised, not over-assertive.	 And for following marks. OR $p<0.05$, reject H_0. Must be t, not z. $p=0.0237$ and conclusion.

Question			Answer	Marks	Guidance	
			training has improved the athlete's performance	**[7]**		
4			Consider variable $B_1 + B_2 - 3C$ Mean = 10 (or -10) Variance = 776 $\frac{0-10}{\sqrt{776}}$ = –0.359 $\Phi(0.359)$ = 0.640	M1 B1 A1 M1 A1 M1 A1 **[7]**	or use $2\times10^2+3^2\times8^2$ Allow from 2B – 3C. Allow 1000-990 Allow reversed Allow 0.359 Must be correct tail. Answer must be >0.5 Allow 0.64	
5	**(i)**		$s = \sqrt{\frac{0.376\times0.624}{1731}}$ $p_s \pm zs$ $z = 3.291$ (0.338, 0.414)	B1 M1 B1 A1 **[4]**	 s must be correct structure. Allow 3.29	Allow from % throughout.
	(ii)		2 from e.g. One in a thousand CI does not contain popn. proportion. Some of the voters lied. The CI is approx.(because a discrete distn. has been approx. by a cs. one.) or estimate. A continuity correction has not been applied. The popn. var. is estimated from the sample. The distn. of P_s is only approx. normal.	B1,B1 **[2]**	The sample was unrepresentative B2 Voters not independent. Voters may have changed their minds. Some voters forgot to vote. Sample biased.	Sample not random B0. Small sample B0

Question			Answer	Marks	Guidance	
	(iii)		$z\sqrt{\frac{0.376 \times 0.624}{n}}$	M1	Allow incorrect structure if same as (i)	
			$= 0.025$ with z=3.29(1)	A1	SC 4065 with no working B2	
			$n = 4066$	A1	Allow 4070	Must be integer.
				[3]		
6	(i)		$\int_0^{\frac{1}{2}\pi} k\sin x\,dx + \int_{\frac{1}{2}\pi}^{\pi} k(2 - \frac{2x}{\pi})dx = 1$	M1		
			$-k\cos x \Big\vert_0^{\frac{1}{2}\pi} + k\left[2x - \frac{x^2}{\pi}\right]_{\frac{1}{2}\pi}^{\pi} = 1$	B1	Both integrals correct, ignore limits.	
				M1	Substitute limits and attempt to simplify	
			$= \frac{4}{4+\pi}$ AG	A1		
				[4]		
	(ii)		$k\int_0^{\frac{1}{2}\pi} x\sin x\,dx + k\int_{\frac{1}{2}\pi}^{\pi} x(2 - \frac{2x}{\pi})dx$	M1		
			$k[-x\cos x + \sin x]_0^{\frac{1}{2}\pi} + k\left[x^2 - \frac{2x^3}{3\pi}\right]_{\frac{1}{2}\pi}^{\pi}$	M1	Correct method for both integrals	Allow 1 error for M1.
				A1	Both integrals correct, ignore limits.	
			=1.48 (3sf)	A1		
				[4]		
7	(i)		H_0: no assoc between level of improvement and treatment H_1: there is an assoc between level of improvement and treatment.	B1	oe	
				[1]		
	(ii)		$\frac{26 \times 17}{100}$	M1		
			= 4.42	A1		
			Expected value for NC, tr C , < 5	A1		
				[3]		

Question		Answer	Marks	Guidance	
	(iii)	$\frac{26\times37}{100}$ (= 9.62) $\frac{(16-\text{"9.62"})^2}{\text{"9.62"}}$ = 4.231 AG	M1 M1 A1 **[3]**		
	(iv)	10.51 > 9.488, reject H_0 There is sufficient evidence that there is an assoc between level of improvement and treatment.	M1 A1 **[2]**	Contextualised.	$p<0.05$ and reject H_0 p=0.03266 and conclusion.
8	(i)	Each popn (of yields) should be N dist.	B1 **[1]**	Allow X and Y. Allow 'data'.	NOT increase.NOT it NOT sample NOT just 'Normally distributed'
	(ii)	H_0: $\mu_x = \mu_y$, H_1 : $\mu_x > \mu_y$ $S_p^2 = \frac{9\times87.96+9\times31.96}{18}$ = 59.96 $\frac{211.2-200.8}{\sqrt{59.96\times(\frac{1}{10}+\frac{1}{10})}}$ = 3.00 CV = 2.552 "3.00">"2.552", reject H_0 There is evidence that the phosphorus treatment has increased the yield.	B1 M1 A1 M1,A1 A1 B1 M1 A1 **[9]**	If in words, must have population. Allow $\frac{1}{9}+\frac{1}{9}$ and/or incorrect mean for M1. Allow 3 Follow through both TS, CV for this mark Follow through TS, but not CV Contextualised, not over-assertive..	Allow $\frac{87.96}{10}+\frac{31.96}{10}$ for M1A1 p=0.00382 B1 $p<0.01$ reject H_0 M1 Correct p and satisfactory conc. A1.
9	(i)	$A = X(10 - X)$ Use CTS $A = 25 - (X - 5)^2$ AG	B1 M1 A1 **[3]**	from base x height or quadratic formula	Allow verification.
	(ii)	$f_x(x) = ½$	B1	Ignore range.	

Question			Answer	Marks	Guidance	
				[1]		
	(iii)		$F_X(x) = ½x$ $(F_A(a) =)\ P(A \le a) = P[X(10 - X) \le a]$ $= F_X(5 - \sqrt{(25 - a)})$ $= \frac{1}{2}(5 - \sqrt{25 - a})$ AG $0 \le A \le 16$ AG explained.	B1 M1 A1 A1 B1 **[5]**	Only if (ii) correct. Fully justified. eg $x=2 \to a=16$	 X(or x)$\ge 5+\sqrt{(25-a)}$ is impossible. $F_A(16)=1$ is not enough.
	(iv)		$f_A(a) = \frac{1}{4}(25 - a)^{-\frac{1}{2}}$	M1,A1 **[2]**	M1 for attempt at differentiation.	