OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS 4734

Probability & Statistics 3

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- List of Formulae (MF1)

Other Materials Required:
- Scientific or graphical calculator

Thursday 24 June 2010
Morning

Duration: 1 hour 30 minutes

* 4 7 3 4 *

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- 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 graphical calculator in this paper.

INFORMATION FOR CANDIDATES

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

1 The numbers of minor flaws that occur on reels of copper wire and reels of steel wire have Poisson distributions with means 0.21 per metre and 0.24 per metre respectively. One length of 5 m is cut from each reel.

(i) Calculate the probability that the total number of flaws on these two lengths of wire is at least 2. **[4]**

(ii) State one assumption needed in the calculation. **[1]**

2 A coffee machine used in a bar is claimed by the manager to dispense 170 ml of coffee per cup on average. A customer believes that the average amount of coffee dispensed is less than 170 ml. She measures the amount of coffee in 6 randomly chosen cups. The results, in ml, are as follows.

167 171 164 169 168 166

Assuming a relevant normal distribution, test the manager's claim at the 5% significance level. **[7]**

3 The developers of a shopping mall sponsored a study of the shopping habits of its users. Each of a random sample of 100 users was asked whether their weekend shopping was mainly on Saturday or mainly on Sunday. The results, classified according to whether the user lived in the city or the country, are shown in the table.

	City dweller	Country dweller
Saturday shopper	23	19
Sunday shopper	42	16

(i) Test, at the 10% significance level, whether there is an association between the area in which shoppers live and the day on which they shop at the weekend. **[8]**

(ii) State, with a reason, whether the conclusion of the test would be different at the 3% significance level. **[1]**

4 Part of an ecological study involved measuring the heights of trees in a young forest. In order to obtain an estimate of the mean height of all the trees in the forest, a random sample of 70 trees was selected and their heights measured. These heights, x metres, are summarised by $\Sigma x = 246.6$ and $\Sigma x^2 = 1183.65$. The mean height of all trees in the forest is denoted by μ metres.

(i) Calculate a symmetric 90% confidence interval for μ. **[5]**

(ii) A student was asked to interpret the interval and said,

"If 100 independent 90% confidence intervals were calculated then 90 of them would contain μ."

Explain briefly what is wrong with this statement. **[1]**

(iii) Four independent 90% confidence intervals for μ are obtained. Calculate the probability that at least three of the intervals contain μ. **[2]**

5 A random variable X is believed to have (cumulative) distribution function given by

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-x^2} & x \geqslant 0. \end{cases}$$

In order to test this, a random sample of 150 observations of X were taken, and their values are summarised in the following grouped frequency table.

Values	$0 \leqslant x < 0.5$	$0.5 \leqslant x < 1$	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$x \geqslant 2$
Frequency	41	50	32	23	4

The expected frequencies, correct to 1 decimal place, corresponding to the above distribution, are 33.2, 61.6 and 39.4 respectively for the first 3 cells.

(i) Find the expected frequencies for the last 2 cells. **[4]**

(ii) Carry out a goodness of fit test at the $2\frac{1}{2}\%$ significance level. **[6]**

6 It has been suggested that people who suffer anxiety when they are about to undergo surgery can have their anxiety reduced by listening to relaxation tapes. A study was carried out on 18 experimental subjects who listened to relaxation tapes, and 13 control subjects who listened to neutral tapes. After listening to the tapes, the subjects were given a test which produced an anxiety score, X. Higher scores indicated higher anxiety. The results are summarised in the table.

	Sample size	$\bar{x}$	$\Sigma(x - \bar{x})^2$
Experimental subjects	18	32.16	1923.56
Control subjects	13	38.21	1147.58

(i) Use a two-sample t-test, at the 5% significance level, to test whether anxiety is reduced by listening to relaxation tapes. State two necessary assumptions for the validity of your test. **[10]**

(ii) State why a test using a normal distribution would not have been appropriate. **[1]**

7 The employees of a certain company have masses which are normally distributed. Female employees have a mean of 66.7 kg and standard deviation 9.3 kg, and male employees have a mean of 78.3 kg and standard deviation 8.5 kg. It may be assumed that all employees' masses are independent. On the ground floor 6 women and 9 men enter the empty staff lift for which it is stated that the maximum load is 1150 kg.

(i) Calculate the probability that the maximum load is exceeded. **[6]**

At the first floor all 15 passengers leave and 6 women, 8 men and an unknown employee enter.

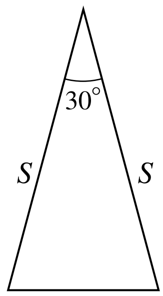
(ii) Assuming that the unknown employee is equally likely to be a woman or a man, calculate the probability that the maximum load is now exceeded. **[6]**

[Question 8 is printed overleaf.]

8 The continuous random variable S has probability density function given by

$$\mathrm{f}(s) = \begin{cases} \dfrac{8}{3s^3} & 1 \leqslant s \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

An isosceles triangle has equal sides of length S, and the angle between them is 30° (see diagram).



(i) Find the (cumulative) distribution function of the area X of the triangle, and hence show that the probability density function of X is $\dfrac{1}{3x^2}$ over an interval to be stated. **[7]**

(ii) Find the median value of X. **[3]**

1(i)	Total has Poisson distribution with mean $\lambda = 0.21\times5+0.24\times5 = 2.25$ $P(\geq2) = 1 - e^{-\lambda}(1+\lambda)$ $=0.657$	M1 A1 M1 A1 **4**	With ×5 λ or 1+λ in brackets (their λ) Or interpolation from tables
(ii)	EITHER: Each length is a random sample OR: Flaws occur independently on the reels	B1 **1** **[5]**	In context Accept randomly
2	H_0: μ =(or ≥) 170 , H_1: μ < 170 $\overline{x}$ =167.5 s^2 = 5.9 EITHER: (α) (167.5 – 170)/√(5.9/6) = - 2.52(1) Compare with – 2.015 OR: (β) 170 – t√(5.9/6) = 168.0 Compare 167.5 with CV and reject H_0 There is sufficient evidence at the 5% significance level that the machine dispenses less than 170 ml on average.	B1 B1 B1 M1 A1 M1 M1 A1 M1 A1 **[7]**	For both hypotheses; accept words SR 2-tail test: B0B1B1M1A1M1A0 Max 5/7 Standardise 167.5; + or – for M; /6 seen Explicitly Allow 2.571 Finding critical value or region. With t= 2.015 or 2.571 Explicitly. Allow correct use of | t| M0 if z used SR: B1 if no explicit comparison but conclusion "correct"
3(i)	H_0: There is no association between the area in which a shopper lives and the day they shop (H_1: All alternatives) E-Values 27.3 14. 7 37.7 20.3 χ^2= $(4.3\text{-}0.5)^2(27.3^{-1}+37.7^{-1}+14.7^{-1}+20.3^{-1})$ = 2.606 Compare with 2.706 Do not reject H_0. There is insufficient evidence of an association. SR: If H_0 association, lose 1st B1 and last M1A1	B1 M1 A1 M1 ft A1 A1 M1 A1 **8**	SR difference in proportions B1 define and evaluate p_1 and p_2 with H_0 B1 for p=0.42 M1A1 for z = ±1.827 or 1.835(no pe) M1A0 Max 5/8 At least one E value correct (M1) All correct(A1) At least one χ^2, no or wrong cc, (M1FtE) All correct (A1); 2.606 or 2.61 (A1) Or use calculator (p = 0.106) SR: B1 if no explicit comparison, as Q2 SR: If H_0 association, lose 1st B1 and last M1A1
(ii)	Conclusion the same since critical value > 2.706 (and test statistic unchanged)	B1 **1** **[9]**	OR from z=±2.17, SR

4(i)	s^2= (1183.65-246.6^2/70)/69 Use $\bar{x} \pm zs/\surd(70)$ $s/\surd(70)$ 1.645 (3.10, 3.94)	M1 M1 A1 A1 A1 **5**	AEF Allow without ft or with s^2; with 70 Their s A0 if interval not indicated
(ii)	Change 90 to around 90	B1 **1**	Or equivalent
(iii)	$4(0.9)^3(0.1) + 0.9^4$ =0.9477	M1 A1 **2** **[8]**	Use of bino with p=0.9 or 0.1 and 4 and Correct terms considered. art 0.948
5(i)	$e^{-2.25} - e^{-4}$ × 150 = 13.1 Last: 150 – sum=2.7	M1 A1 A1 A1 ft **4**	Or find last entry using F(x) Or 2.7 if found first Or 13.1 any accuracy
(ii)	(H_0: Data fits the model, H_1: Data does not fit) Combine last two cells $\chi^2 = 7.8^2/33.2 + 11.6^2/61.6 + 7.4^2/39.4 + 11.2^2/15.8$ = 13.3(46) Compare with 9.348 (or 11.14) , reject H_0 **(**There is sufficient evidence at the $2\frac{1}{2}$ % significance level that) the model is not a good fit	B1 M1*****Dep A1 A1 M1 A1 ft Dep* **6** **[10]**	At least two correct All correct In range 13.2 to 13.5 SR: If last 2 cells are not combined B0M1A1A1(for 13. 5) M1A1 If no explicit comparison B1 if conclusion follows
6(i)	Anxiety scores; have normal distributions; common variance; independent samples H_0: $\mu_E = \mu_C$, H_1: $\mu_E < \mu_C$ s^2 = (1923.56+1147.58)/29 (= 105.9) (t) = (32.16 – 38.21)/$\surd$[105.9(18^{-1}+13^{-1})] = - 1.615 t_{crit} = - 1.699 Compare -1.615 with -1.699 and do not reject H_0 There is insufficient evidence at the 5% significance level to show that anxiety is reduced by listening to relaxation tapes	B2 B1 B1 M1 A1 A1 B1 M1 A1 ft **10**	Context + 2 valid points B2 Context + 1VP, no context +2VP B1 Not in words Allow 1 error; eg s^2= 1923.56/(17or18) + 47.5/(12or13) All correct Or **+** Or **+;** accept art ±1.70 Or **+ , +.** M0 if t not ±1.699,±2.045 In context, not over-assertive OR Find CV or CR: B2B1B1; C= or ≥ st , t = ±1.699 or ±2.015 M1A1 t= ±1.699 B1; G= 6.11(2) A1; 6.112> 6.05 and reject H_0 etcM1A1
(ii)	Sample sizes are too small (to appeal to CLT)	B1 **1** **[11]**	

7(i)	Use $\sum F + \sum M \sim N(\mu, \sigma^2)$ $\mu = 1104.9$ $\sigma^2 = 6\times9.3^2 + 9\times8.5^2$ $= 1169.2$ $P(> 1150) = 1 - \Phi([1150 - 1104.9]/\sqrt{1169.2})$ $= 0.0937$	M1 A1 M1 A1 M1 A1 **6**	Sum of indep normal variables is normal Standardise, correct tail. M0 $\sigma/\sqrt{15}$ Accept .094
(ii)	If unknown M, prob $\frac{1}{2}$, 6F and 9M as before. If unknown W, prob $\frac{1}{2}$, 7W and 8M Having N(1093.3,1183.4) $P(> 1150) = 1 - \Phi(1.648) = 0.0497$ $P = \frac{1}{2}\times 0.0936 + \frac{1}{2}\times 0.0497$ $= 0.07165$	M1 B1 B1 A1 M1 A1 **6** **[12]**	Considering two cases Mean and variance Use of $\frac{1}{2}$ ART 0.072
8(i)	$X = \frac{1}{4}S^2$ $F(s) = \int_1^s \frac{8}{3s^3}\,ds = \left[-\frac{4}{3s^2}\right]_1^s$ $= \frac{4}{3}(1 - 1/s^2)$ $G(x) = P(X \le x) = P(S \le 2\sqrt{x})$ $= F(2\sqrt{x})$ $= \frac{4}{3} - \frac{1}{3x}$ $g(x) = \begin{cases} \frac{1}{3x^2} & \frac{1}{4} \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$	B1 M1 A1 M1 A1 ft M1 B1 **7**	 Ignore range here SR: B1 for G(x)=F($2\sqrt{x}$) without justification and with correct result ft F For G′ (a) For range
(ii)	EITHER: $G(m) = \frac{1}{2}$ $\Rightarrow \frac{4}{3} - \frac{1}{3x} = \frac{1}{2}$ $\Rightarrow m = \frac{2}{5}$ OR: $\int_{1/4}^{m} \frac{1}{3x^2}\,dx = \frac{1}{2}$ $\Rightarrow \left[-\frac{1}{3x}\right]_{1/4}^{m} = \frac{1}{2}$ $\Rightarrow m = \frac{2}{5}$	M1 A1 ft A1 M1 A1 A1 **3** **[10]**	ft G(x) in **(i)** CAO Allow wrong $\frac{1}{4}$ Allow wrong $\frac{1}{4}$ CAO