OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE
MATHEMATICS
Probability & Statistics 3

4734

Candidates answer on the Answer Booklet

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

Other Materials Required:
None

Wednesday 20 January 2010
Afternoon

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 continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{2}{5} & -a \leqslant x < 0, \\ \frac{2}{5}e^{-2x} & x \geqslant 0. \end{cases}$$

Find

(i) the value of the constant a, **[3]**

(ii) $E(X)$. **[5]**

2 The amount of tomato juice, X ml, dispensed into cartons of a particular brand has a normal distribution with mean 504 and standard deviation 3. The juice is sold in packs of 4 cartons, filled independently. The total amount of juice in one pack is Y ml.

(i) Find $P(Y < 2000)$. **[4]**

The random variable V is defined as $Y - 4X$.

(ii) Find $E(V)$ and $Var(V)$. **[3]**

(iii) What is the probability that the amount of juice in a randomly chosen pack is more than 4 times the amount of juice in a randomly chosen carton? **[1]**

3 It is given that X_1 and X_2 are independent random variables with $X_1 \sim N(\mu_1, 2.47)$ and $X_2 \sim N(\mu_2, 4.23)$. Random samples of n_1 observations of X_1 and n_2 observations of X_2 are taken. The sample means are denoted by $\overline{X}_1$ and $\overline{X}_2$.

(i) State the distribution of $\overline{X}_1 - \overline{X}_2$, giving its parameters. **[3]**

For two particular samples, $n_1 = 5$, $\Sigma x_1 = 48.25$, $n_2 = 10$ and $\Sigma x_2 = 72.30$.

(ii) Test at the 2% significance level whether μ_1 differs from μ_2. **[6]**

A student stated that because of the Central Limit Theorem the sample means will have normal distributions so it is unnecessary for X_1 and X_2 to have normal distributions.

(iii) Comment on the student's statement. **[1]**

4 The continuous random variable V has (cumulative) distribution function given by

$$F(v) = \begin{cases} 0 & v < 1, \\ 1 - \dfrac{8}{(1+v)^3} & v \geqslant 1. \end{cases}$$

The random variable Y is given by $Y = \dfrac{1}{1+V}$.

(i) Show that the (cumulative) distribution function of Y is $8y^3$, over an interval to be stated, and find the probability density function of Y. **[7]**

(ii) Find $E\left(\dfrac{1}{Y^2}\right)$. **[2]**

5 Each of a random sample of 200 steel bars taken from a production line was examined and 27 were found to be faulty.

(i) Find an approximate 90% confidence interval for the proportion of faulty bars produced. **[4]**

A change in the production method was introduced which, it was claimed, would reduce the proportion of faulty bars. After the change, each of a further random sample of 100 bars was examined and 8 were found to be faulty.

(ii) Test the claim, at the 10% significance level. **[7]**

6 The deterioration of a certain drug over time was investigated as follows. The drug strength was measured in each of a random sample of 8 bottles containing the drug. These were stored for two years and the strengths were then re-measured. The original and final strengths, in suitable units, are shown in the following table.

Bottle	1	2	3	4	5	6	7	8
Original strength	8.7	9.4	9.2	8.9	9.6	8.2	9.9	8.8
Final strength	8.1	9.0	9.0	8.8	9.3	8.0	9.5	8.5

(i) Stating any required assumption, test at the 5% significance level whether the mean strength has decreased by more than 0.2 over the two years. **[9]**

(ii) Calculate a 95% confidence interval for the mean reduction in strength over the two years. **[3]**

7 A chef wished to ascertain her customers' preference for certain vegetables. She asked a random sample of 120 customers for their preferred vegetable from asparagus, broad beans and cauliflower. The responses, classified according to the gender of the customer, are shown in the table.

	Asparagus	Broad beans	Cauliflower
Female preference	31	9	25
Male preference	17	21	17

(i) Test, at the 5% significance level, whether vegetable preference and gender are independent. **[8]**

(ii) Determine whether, at the 10% significance level, the vegetables are equally preferred. **[6]**

4734 Probability & Statistics 3

1(i)	$\int_{-a}^{0} \frac{2}{5} dx + \int_{0}^{\infty} \frac{2}{5} e^{-2x} dx = 1$ $2a/5 + 1/5 = 1$ $a = 2$	M1 A1 A1 **3**	Sum of probabilities =1
(ii)	$\int_{-2}^{0} \frac{2}{5} x dx + \int_{0}^{\infty} \frac{2}{5} xe^{-2x} dx$ $\int_{-a}^{0} \frac{2}{5} x dx = -\frac{a^2}{5}$ $\int_{0}^{\infty} \frac{2}{5} xe^{-2x} dx = \left[-\frac{1}{5} xe^{-2x}\right] + \left[-\frac{1}{10} e^{-2x}\right]$ = - 0.7	M1 A1 √ M1 A1 A1 **5** **[8]**	Σ ∫ x f(x) dx √ a By parts with 1 part correct Both parts correct CAO
2(i)	4 cartons: Total, $Y \sim N(2016, 36)$ $P(Y \le 2000) = \Phi(-16/\sqrt{36})$ = 0.00383	B1B1 M1 A1 **4**	Mean and variance
(ii)	E(V) = 0 Var(V) = 36 + 16×9 = 180	B1 M1 A1 **3**	 CWO
(iii)	0.5	B1 **1** **[8]**	
3(i)	Normal distribution Mean $\mu_1 - \mu_2$; variance $2.47/n_1 + 4.23/n_2$	B1 B1B1 **3**	
(ii)	$H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \ne \mu_2$ $(9.65 - 7.23)/\sqrt{(2.47/5+4.23/10)}$ =2.527 > 2.326 Reject H_0 There is sufficient evidence at the 2% significance level that the means differ	B1 M1 B1 A1 M1 A1 **6**	 Or find critical region Numerator Compare with critical value **SR1:If no specific comparison but CV and conclusion correct B1. Same in Q5,6,7** **SR2: From CI: 2.42±$z\sigma$ M1, σ correct z = 2.326 B1, (0.193, 4.647) A1** **0 in not in CI ; reject H_0 etc M1A1 Total 6** **Conclusions not over-assertive in Q3, 5, 6**
(iii)	Any relevant comment.	B1 **1** **[10]**	e.g sample sizes too small for CLT to apply

4(i)	$G(y)=P(Y \le y)=P(1/(1+V) \le y)$ $=P(V \ge 1/y - 1)$ $= 1 - F(1/y - 1)$ $=\begin{cases} 0 & y \le 0, \\ 8y^3 & 0 < y \le 1/2, \\ 1 & y > 1/2. \end{cases}$ $g(y)=\begin{cases} 24y^2 & 0 < y \le 1/2, \\ 0 & \text{otherwise.} \end{cases}$	M1 A1 A1 A1 B1 M1 A1 **7**	 Use of F $8y^3$ obtained correctly Correct range. Condone omission of $y \le 0$ For $G'(y)$ Correct answer with range √
	$\int 24y^2/y^2\, dy$ with limits $=12$	M1 A1 **2** **[9]**	 With attempt at integration
5(i)	Use $p_s \pm z\sqrt{(p_s q_s/200)}$ z=1.645 $s = \sqrt{(0.135 \times 0.865/200)}$ (0.0952, 0.1747)	M1 B1 A1 A1 **4**	 Or /199 (0.095, 0.175) to 3DP
(ii)	H_0: $p_1 - p_2 = 0$, H_1: $p_1 - p_2 > 0$ $\dfrac{27/200 - 8/100}{\sqrt{35/300 \times 265/300 \times (200^{-1} + 100^{-1})}}$ $= 1.399$ > 1.282 Reject H_0. There is sufficient evidence at the 10% significance level that the proportion of faulty bars has reduced	B1 M1 B1 A1 A1 M1 A1 **7** **[11]**	Or equivalent Correct form. Pooled estimate of p = 35/300 Correct form of sd OR: $P(z \ge 1.399) = 0.0809 < 0.10$ **SR: No pooled estimate: B1M1B0B0 A1 for 1.514, M1A1 Max 5/7**
6(i)	Assumes that decreases have a normal distn H_0:$\mu_{O-F} = 0.2$ (or $\ge$), H_1:$\mu_{O-F} > 0.2$ O-F: 0.6 0.4 0.2 0.1 0.3 0.2 0.4 0.3 $\bar{D} = 0.3125$ $s^2 = 0.024107$ $(0.3125-0.2)/\sqrt{(0.024107/8)}$ =2.049 > 1.895 Reject H_o – there is sufficient evidence at the 5% significance level that the reduction is more than 0.2	B1 B1 M1 B1 A1 M1 A1 M1 A1 **9**	 B1 Use paired differences t-test Must have /8 OR: $P(t \ge 2.049) = 0.0398 < 0.05$ Allow M1 from $t_{14} = 1.761$ **SR: 2-sample test:B1B1M0B1A0 M1 using 1.761 A0 Max 4/9**
(ii)	$0.3125 \pm t\sqrt{(0.024107/8)}$ $t = 2.365$ (0.1827, 0.4423)	M1 B1 A1 **3** **[12]**	Allow with z but with /8 Rounding to (0.283, 0.442)

7(i)	H_0:Vegetable preference is independent of gender H_1: All alternatives E-Values 26 16.25 22.75 22 13.75 19.25 $\chi^2 = 5^2(26^{-1} + 22^{-1}) + 7.25^2(16.25^{-1} + 13.75^{-1})$ $+ 2.25^2(22.75^{-1} + 19.25^{-1})$ =9.641 9.64 > 5.991 Reject H_0 , (there is sufficient evidence at the 5% that) vegetable preference and gender are not independent	B1 M1 A1 M1 A1 A1 M1 A1 **8**	For both hypotheses At least one correct All correct Correct form of any one All correct ART 9.64 OR: $P(\geq 9.641)=0.00806 < 0.05$
(ii)	(H_0: Vegetables have equal preference H_1: All alternatives) Combining rows: 48 30 42 E-Values: 40 40 40 $\chi^2 = (8^2+10^2 + 2^2)/40$ = 4.2 4.2 < 4.605 Do not reject H_0, there is insufficient evidence at the 10% significance level of a difference in the proportion of preferred vegetables	 M1 A1 M1 A1 M1 A1 **6** **[14]**	 OR:$P(\geq 4.2) = 0.122 > 0.10$ AEF in context