OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS

Probability & Statistics 2

4733

Candidates answer on the Answer Booklet

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

Other Materials Required:
None

Tuesday 13 January 2009
Morning

Duration: 1 hour 30 minutes

4733

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 A newspaper article consists of 800 words. For each word, the probability that it is misprinted is 0.005, independently of all other words. Use a suitable approximation to find the probability that the total number of misprinted words in the article is no more than 6. Give a reason to justify your approximation. **[4]**

2 The continuous random variable Y has the distribution $N(23.0, 5.0^2)$. The mean of n observations of Y is denoted by $\overline{Y}$. It is given that $P(\overline{Y} > 23.625) = 0.0228$. Find the value of n. **[4]**

3 The number of incidents of radio interference per hour experienced by a certain listener is modelled by a random variable with distribution Po(0.42).

(i) Find the probability that the number of incidents of interference in one randomly chosen hour is

(a) 0,

(b) exactly 1.

[3]

(ii) Find the probability that the number of incidents in a randomly chosen 5-hour period is greater than 3. **[3]**

(iii) One hundred hours of listening are monitored and the numbers of 1-hour periods in which 0, 1, 2, … incidents of interference are experienced are noted. A bar chart is drawn to represent the results. Without any further calculations, sketch the shape that you would expect for the bar chart. (There is no need to use an exact numerical scale on the frequency axis.) **[2]**

4 A television company believes that the proportion of adults who watched a certain programme is 0.14. Out of a random sample of 22 adults, it is found that 2 watched the programme.

(i) Carry out a significance test, at the 10% level, to determine, on the basis of this sample, whether the television company is overestimating the proportion of adults who watched the programme. **[8]**

(ii) The sample was selected randomly. State what properties of this method of sampling are needed to justify the use of the distribution used in your test. **[2]**

5 The continuous random variables S and T have probability density functions as follows.

$$S: \quad \mathrm{f}(x) = \begin{cases} \frac{1}{4} & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

$$T: \quad \mathrm{g}(x) = \begin{cases} \frac{5}{64}x^4 & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

(i) Sketch, on the same axes, the graphs of f and g. **[3]**

(ii) Describe in everyday terms the difference between the distributions of the random variables S and T. (Answers that comment only on the shapes of the graphs will receive no credit.) **[2]**

(iii) Calculate the variance of T. **[4]**

6 The weight of a plastic box manufactured by a company is W grams, where $W \sim N(\mu, 20.25)$. A significance test of the null hypothesis $H_0 : \mu = 50.0$, against the alternative hypothesis $H_1 : \mu \neq 50.0$, is carried out at the 5% significance level, based on a sample of size n.

(i) Given that $n = 81$,

(a) find the critical region for the test, in terms of the sample mean $\overline{W}$, **[5]**

(b) find the probability that the test results in a Type II error when $\mu = 50.2$. **[5]**

(ii) State how the probability of this Type II error would change if n were greater than 81. **[1]**

7 A motorist records the time taken, T minutes, to drive a particular stretch of road on each of 64 occasions. Her results are summarised by

$$\Sigma t = 876.8, \quad \Sigma t^2 = 12\,657.28.$$

(i) Test, at the 5% significance level, whether the mean time for the motorist to drive the stretch of road is greater than 13.1 minutes. **[11]**

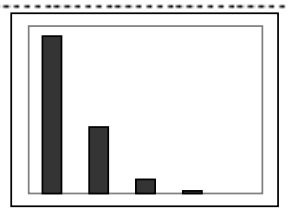
(ii) Explain whether it is necessary to use the Central Limit Theorem in your test. **[1]**

8 A sales office employs 21 representatives. Each day, for each representative, the probability that he or she achieves a sale is 0.7, independently of other representatives. The total number of representatives who achieve a sale on any one day is denoted by K.

(i) Using a suitable approximation (which should be justified), find $P(K \geqslant 16)$. **[7]**

(ii) Using a suitable approximation (which should be justified), find the probability that the mean of 36 observations of K is less than or equal to 14.0. **[7]**

4733 Probability & Statistics 2

1		$U \sim$ B(800, 0.005) ≈ Po(4) P($U \le 6$) = **0.8893** $n > 50$/large, $np < 5$/p small	B1 M1 A1 B1 **4**	Po(np) stated or implied Tables or formula ± 1 term, e.g. 0.7851, 0 .9489, 0.1107, *not* 1– Answer 0.889 or a.r.t. 0.8893 Both conditions
2		$\frac{23.625-23}{5/\sqrt{n}} = 2$ $\sqrt{n} = 16$ n = **256**	M1 A1 M1 A1 **4**	Standardise with $\sqrt{n}$, allow $\sqrt{}^2$ errors Equate to 2 or a.r.t. 2.00, signs correct Solve for $\sqrt{n}$, needs Φ^{-1}, *not* from /n 256 only, allow from wrong signs
3	**(i)**	(a) $e^{-0.42}$ = **0.657** (b) $0.42\ e^{-0.42}$ = **0.276**	M1 A1 A1 **3**	Correct formula for R = 0 or 1 P(0), a.r.t. 0.657 P(1), a.r.t. 0.276
	(ii)	Po(2.1): 1 – P(≤ 3) = 1 – 0.8386 = **0.1614**	M1 M1 A1 **3**	Po(2.1) stated or implied Tables or formula, e.g. 0.8386 or 0.6496 or 0.9379 or complement; Answer, in range [0.161, 0.162]
	(iii)		B2 **2**	At least 3 separate bars, all decreasing *Allow histogram. Allow convex* P(0) < P(1) but otherwise OK: B1 Curve: B1 *[no hint of normal allowed]*
4	**(i)**	$H_0 : p = 0.14$ $H_1 : p < 0.14$ B(22, 0.14) P(≤ 2) = $.86^{22}$ + (22×$.86^{21}$×.14) + (231×$.86^{20}$×$.14^2$) = **0.3877** > 0.1 Do not reject H_0. Insufficient evidence that company overestimates viewing proportion	B2 M1 A1 A1 B1 M1 A1 **8**	Both correct. 1 error, B1, but x or r or $\bar{x}$ etc: 0 B(22, 0.14) stated or implied, e.g. N(3.08, 2.6488) or Po(3.08) Correct formula for 2 or 3 terms, *or* P(≤ 0) = 0.036 and CR Correct answer, a.r.t. 0.388, *or* CR is = 0 Explicitly compare 0.1 or CR with 2, OK from Po but *not* from N Correct comparison type and conclusion, needs binomial, at least 2 terms, *not* from P(< 2) Contextualised, some acknowledgement of uncertainty [SR: Normal: B2 M1 A0 B0 M0] [SR: 2-tailed, or $p > 0.14$, P(≥ 2): B1M1A2B0M1A1]
	(ii)	Selected independently Each adult equally likely to be chosen	B1 B1 **2**	Independent *selection* Choice of sample elements equally likely (no credit if not focussed on selection) [Only "All samples of size n equally likely": B1 only unless related to Binomial conditions]
5	**(i)**		B1 B1 B1 **3**	Horizontal straight line Symmetrical U-shaped curve Both correct, including relationship between the two and not extending beyond [–2, 2], curve through (0,0)
	(ii)	S is equally likely to take any value T is more likely at extremities	B2 **2**	Correct statement about both distributions, √ on their graph [Correct for one only, or partial description: B1] *Not* "probability of S is constant", etc.
	(iii)	$\frac{5}{64}\int_{-2}^{2} x^6 dx = \frac{5}{64}\left[\frac{x^7}{7}\right]_{-2}^{2} \left[= \frac{20}{7}\right]$ -0^2 = $\frac{20}{7}$	M1 A1 B1 A1 **4**	Integrate x^2g(x), limits –2, 2 Correct indefinite integral [= $5x^7/448$] 0 or 0^2 subtracted or E(X) = 0 seen, *not* $\int x^2 f(x)dx - \int xf(x)dx$ Answer $\frac{20}{7}$ or $2\frac{6}{7}$ or a.r.t. 2.86, don't need 0

6	**(i)**	$50.0 \pm 1.96\sqrt{\frac{20.25}{81}} = 50.0 \pm 0.98$ $= 49.02, 50.98$ $\overline{W} < 49.02$ and $\overline{W} > 50.98$	M1 B1 A1A1 A1√ **5**	$50.0 \pm z\sqrt{(1.96/81)}$, allow one sign only, allow √ errors $z = 1.96$ in equation (*not* just stated) Both critical values, min 4 SF at some stage (if both 3SF, A1) CR, allow ≤ / ≥, don't need $\overline{W}$, √ on their CVs, can't recover [Ans 50 ± 0.98: A1 only] [SR: 1 tail, M1B0A0; 50.8225 or 49. 1775: A1]
	(ii)	$\frac{50.98-50.2}{0.5} = 1.56$ $\frac{49.02-50.2}{0.5} = -2.36$ $\Phi(1.56) - \Phi(-2.36) =$ **0.9315**	M1 A1 A1 M1 A1 **5**	Standardise one limit with same SD as in (i) A.r.t. 1.56, allow – ⎫ Can allow √ here A.r.t. –2.36, allow + ⎭ if very unfair Correct handling of tails for Type II error Answer in range [0.931, 0.932] [SR 1-tail M1; –1.245 or 2.045 A1; 0.893 or 0.9795 A1]
	(iii)	It would get smaller	B1 **1**	No reason needed, but withhold if definitely wrong reason seen. Allow from 1-tail
7	**(i)**	$\hat{\mu} = \bar{t} = 13.7$ $\frac{12657.28}{64} - 13.7^2$ [= 10.08]; $\times \frac{64}{63}$ = **10.24** $H_0 : \mu = 13.1$, $H_1 : \mu > 13.1$ $\frac{13.7-13.1}{\sqrt{10.24/64}} = 1.5$ or $p = 0.0668$ $1.5 < 1.645$ or $0.0668 > 0.05$ Do not reject H_0. Insufficient evidence that time taken on average is greater than 13.1 min	B1 M1 M1 A1 B2 M1 A1 B1 M1 A1 **11**	13.7 stated Correct formula for biased estimate $\times \frac{64}{63}$ used, or equivalent, can come in later Variance or SD 10.24 or 10.2 Both correct. [SR: One error, B1, but x or t or $\bar{x}$ or $\bar{t}$, 0] Standardise, or find CV, with √64 or 64 z = a.r.t. 1.50, or $p = 0.0668$, or CV 13.758 [√ on z] Compare z & 1.645, or p & 0.05 (must be correct tail), or $z = 1.645$ & 13 with CV Correct comparison & conclusion, needs 64, *not* $\mu = 13.7$ Contextualised, some acknowledgement of uncertainty [13.1 – 13.7: (6), M1 A0 B1 M0]
	(ii)	Yes, not told that dist is normal	B1 **1**	Equivalent statement, *not* "n is large", don't need "yes"
8	**(i)**	N(14.7, 4.41) Valid because $np = 14.7 > 5$; $nq = 6.3 > 5$ $1 - \Phi\left(\frac{15.5-14.7}{\sqrt{4.41}}\right) = 1 - \Phi(0.381)$ $= 1 - 0.6484$ = **0.3516**	M1 A1 B1 B1 M1 A1 A1 **7**	Normal, attempt at np Both parameters correct Check $np > 5$; ⎫ If both asserted but not both nq or $npq > 5$ ⎭ 14.7 and 6.3 seen: B1 only [Allow "n large, p close to ½"] Standardise, answer < 0.5, no √n z, a.r.t. 0.381 Answer in range [0.351, 0.352] [Exact: M0]
	(ii)	$\bar{K} \sim N(14.7, 4.41/36)$ [= N(14.7, 0.35^2)] Valid by Central Limit Theorem as 36 is large $\Phi\left(\frac{14.0+\frac{1}{72}-14.7}{\sqrt{4.41/36}}\right) = \Phi(-1.96)$ = **0.025**	M1 A1√ B1 M1 A1 A1 A1 **7**	Normal, their np from (i) Their variance/36 Refer to CLT or large n (= 36, *not* 21), or "$K \sim N$ so $\bar{K} \sim N$", *not* same as (i), *not* $np > 5$, $nq > 5$ for $\bar{K}$ Standardise 14.0 with 36 or √36 cc included, allow 0.5 here, e.g. 14.5 – 14.7 $z = -1.96$ or –2.00 or –2.04, allow + if answer < 0.5 0.025 or 0.0228 [0.284 loses last 2] [Po(25.2) etc: probably 0]
	OR:	B(756, 0.7) ≈ N(529.2, 158.76) $\Phi\left(\frac{504.5-529.2}{\sqrt{158.76}}\right) = \Phi(-1.96)$ = **0.025**	M1M1A1 B1 M1 A1 A1	×36; N(529.6,…); 158.76 CLT as above, or $np > 5$, $nq > 5$, can be asserted here Standardise 14×36 cc correct and √npq 0.025 or 0.0228