OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MATHEMATICS **4733**

Probability & Statistics 2

Thursday **15 JUNE 2006** Afternoon 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
List of Formulae (MF1)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- 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 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.
- The total number of marks for this paper is 72.
- 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 3 printed pages and 1 blank page.

 [Turn over

1 Calculate the variance of the continuous random variable with probability density function given by

$$f(x) = \begin{cases} \frac{3}{37}x^2 & 3 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$ [6]

2 **(i)** The random variable R has the distribution $B(6, p)$. A random observation of R is found to be 6. Carry out a 5% significance test of the null hypothesis H_0: $p = 0.45$ against the alternative hypothesis H_1: $p \neq 0.45$, showing all necessary details of your calculation. [4]

(ii) The random variable S has the distribution $B(n, p)$. H_0 and H_1 are as in part **(i)**. A random observation of S is found to be 1. Use tables to find the largest value of n for which H_0 is not rejected. Show the values of any relevant probabilities. [3]

3 The continuous random variable T has mean μ and standard deviation σ. It is known that $P(T < 140) = 0.01$ and $P(T < 300) = 0.8$.

(i) Assuming that T is normally distributed, calculate the values of μ and σ. [6]

In fact, T represents the time, in minutes, taken by a randomly chosen runner in a public marathon, in which about 10% of runners took longer than 400 minutes.

(ii) State with a reason whether the mean of T would be higher than, equal to, or lower than the value calculated in part **(i)**. [2]

4 **(i)** Explain briefly what is meant by a random sample. [1]

Random numbers are used to select, with replacement, a sample of size n from a population numbered 000, 001, 002, …, 799.

(ii) If $n = 6$, find the probability that exactly 4 of the selected sample have numbers less than 500. [3]

(iii) If $n = 60$, use a suitable approximation to calculate the probability that at least 40 of the selected sample have numbers less than 500. [6]

5 An airline has 300 seats available on a flight to Australia. It is known from experience that on average only 99% of those who have booked seats actually arrive to take the flight, the remaining 1% being called 'no-shows'. The airline therefore sells more than 300 seats. If more than 300 passengers then arrive, the flight is over-booked. Assume that the number of no-show passengers can be modelled by a binomial distribution.

(i) If the airline sells 303 seats, state a suitable distribution for the number of no-show passengers, and state a suitable approximation to this distribution, giving the values of any parameters. [2]

Using the distribution and approximation in part **(i)**,

(ii) show that the probability that the flight is over-booked is 0.4165, correct to 4 decimal places, [2]

(iii) find the largest number of seats that can be sold for the probability that the flight is over-booked to be less than 0.2. [5]

6 Customers arrive at a post office at a constant average rate of 0.4 per minute.

(i) State an assumption needed to model the number of customers arriving in a given time interval by a Poisson distribution. [1]

Assuming that the use of a Poisson distribution is justified,

(ii) find the probability that more than 2 customers arrive in a randomly chosen 1-minute interval, [2]

(iii) use a suitable approximation to calculate the probability that more than 55 customers arrive in a given two-hour interval, [6]

(iv) calculate the smallest time for which the probability that no customers arrive in that time is less than 0.02, giving your answer to the nearest second. [5]

7 Three independent researchers, A, B and C, carry out significance tests on the power consumption of a manufacturer's domestic heaters. The power consumption, X watts, is a normally distributed random variable with mean μ and standard deviation 60. Each researcher tests the null hypothesis H_0: $\mu = 4000$ against the alternative hypothesis H_1: $\mu > 4000$.

Researcher A uses a sample of size 50 and a significance level of 5%.

(i) Find the critical region for this test, giving your answer correct to 4 significant figures. [6]

In fact the value of μ is 4020.

(ii) Calculate the probability that Researcher A makes a Type II error. [6]

(iii) Researcher B uses a sample bigger than 50 and a significance level of 5%. Explain whether the probability that Researcher B makes a Type II error is less than, equal to, or greater than your answer to part **(ii)**. [2]

(iv) Researcher C uses a sample of size 50 and a significance level bigger than 5%. Explain whether the probability that Researcher C makes a Type II error is less than, equal to, or greater than your answer to part **(ii)**. [2]

(v) State with a reason whether it is necessary to use the Central Limit Theorem at any point in this question. [2]

4733 Statistics S2 June 2006 **MARK SCHEME: Post-standardisation**

1		$\mu = \frac{3}{37}\int_3^4 x^3 dx = \frac{3}{37}\left[\frac{x^4}{4}\right]_3^4$ [$= 3\frac{81}{148}$]	M1		Integrate $xf(x)$, limits 3 & 4 *[can be implied]* [$\frac{525}{148}$ or 3.547]
		$\frac{3}{37}\int_3^4 x^4 dx = \frac{3}{37}\left[\frac{x^5}{5}\right]_3^4$	M1 A1		Attempt to integrate $x^2f(x)$, limits 3 & 4 Correct indefinite integral, any form
		$= 12\frac{123}{185}$ or 12.665	A1		$\frac{2343}{185}$ or in range [12.6, 12.7] *[can be implied]*
		$\sigma^2 = 12\frac{123}{185} - 3\frac{81}{148}^2 = \mathbf{0.0815}$	M1 A1	6	Subtract their μ^2 Answer, in range [0.0575, 0.084]
2	(i)	Find $P(R \geq 6)$ or $P(R < 6)$ = 0.0083 or 0.9917	M1 A1		Find P(= 6) from tables/calc, OR RH critical region P(≥ 6) in range [0.008, 0.0083] or P(< 6) = 0.9917 OR CR is 6 with probability 0.0083/0.9917
		Compare with 0.025 [can be from N] [0.05 if "empty LH tail stated] Reject H_0	B1 A1√	4	Explicitly compare with 0.025 [or 0.975 if consistent] OR state that result is in critical region Correct comparison and conclusion, √ on their *p*
	(ii)	$n = 9$, $P(\leq 1) = 0.0385$ [> 0.025] $n = 10$, $P(\leq 1) = 0.0233$ [< 0.025] Therefore $n = \mathbf{9}$	M1 A1 B1	3	At least one, or $n = 8$, $P(\leq 1) = 0.0632$ Both of these probabilities seen, don't need 0.025 Answer $n = 9$ only, indep't of M1A1, *not* from P(= 1)
3	(i)	$(140 - \mu)/\sigma = -2.326$ $(300 - \mu)/\sigma = 0.842$ Solve to obtain: $\mu = \mathbf{257.49}$ $\sigma = \mathbf{50.51}$	M1 B1 A1√ M1 A1 A1	6	One standardisation equated to Φ^{-1}, allow "1–", σ^2 Both 2.33 and 0.84 at least, ignore signs Both equations completely correct, √ on their *z* Solve two simultaneous equations to find one variable μ value, in range [257, 258] σ in range [50.4, 50.55]
	(ii)	Higher as there is positive skew	B1 B1	2	"Higher" or equivalent stated Plausible reason, allow from normal calculations
4	(i)	Each element equally likely to be selected (and all selections independent) OR each possible sample equally likely	B1	1	One of these two. "Selections independent" alone is insufficient, but don't need this. An example is insufficient.
	(ii)	B(6, 5/8) ${}^6C_4 p^4 (1-p)^2$ $= \mathbf{0.32187}$	M1 M1 A1√	3	B(6, 5/8) stated or implied, allow e.g. 499/799 Correct formula, any *p* Answer, a.r.t. 0.322, can allow from wrong *p*
	(iii)	N(37.5, 225/16) $\frac{39.5-37.5}{3.75} = 0.5333$ $1 - \Phi(0.5333)$ $= \mathbf{0.297}$	B1 B1 M1 dep A1 dep M1 A1	6	Normal, mean 37.5, or 37.47 from 499/799, 499/800 14.0625 or 3.75 seen, allow 14.07/14.1 or 3.75 Standardise, wrong or no cc, *np*, *npq*, no $\sqrt{n}$ Correct cc, $\sqrt{npq}$, signs can be reversed Tables used, answer < 0.5, *p* = 5/8 Answer, a.r.t. 0.297 SR: *np* < 5: Po(*np*) stated or implied, B1
5	(i)	B(303, 0.01) ≈ Po(3.03)	B1 B1	2	B(303, 0.01) stated, allow *p* = 0.99 or 0.1 Allow Bin implied clearly by parameters Po(3.03) stated or implied, can be recovered from (ii)
	(ii)	$e^{-3.03}(1 + 3.03 + \frac{3.03^2}{2}) = 0.4165$ **AG**	M1 A1	2	Correct formula, ± 1 term or "1 – " or both Convincingly obtain 0.4165(02542) [Exact: 0.41535]
	(iii)	302 seats ⇒ μ = 3.02 $e^{-3.02}(1 + 3.02) = 0.1962$ 0.196 < 0.2 So **302** seats.	M1 M1 A1 A1 A1	5	Try smaller value of μ Formula, at least one correct term Correct number of terms for their μ 0.1962 [or 0.1947 from exact] Answer 302 only

SR: B(303, 0.99): B1B0; M0; M1 then N(298.98,2.9898) or equiv, standardise: M1A1 total 4/9
SR: $p = 0.1$: B(303, 0.1), N(30.3, 27.27) B1B0; Standardise 2 with *np* & $\sqrt{npq}$, M1A0;
N(0.1*n*, 0.09*n*); standardise with *np* & $\sqrt{npq}$; solve quadratic for $\sqrt{n}$; *n* = 339: M1M1M1A1, total 6/9
SR: B(303, 0.01) ≈ N(3.03, 2.9997): B1B0; M0A0; M1A0

6	(i)	Customers arrive independently	B1	**1**	Valid reason in context, allow "random"
	(ii)	1 – 0.9921	M1		Poisson tables, "1 –", or correct formula ± 1 term
		= **0.0079**	A1	**2**	Answer, a.r.t. 0.008 [1 – 0.9384 = 0.0606: M1A0]
	(iii)	N(48, 48)	B1		Normal, mean 48
		$z = \frac{55.5 - 48}{\sqrt{48}}$	B1√		Variance or SD same as mean√
			M1 dep		Standardise, wrong or no cc, $\mu = \lambda$
		= 1.0825	A1		Correct cc, $\sqrt{\lambda}$
		1 – Φ(1.0825)	dep M1		Use tables, answer < 0.5
		= **0.1394**	A1	**6**	Answer in range [0.139, 0.14]
	(iv)	$e^{-\lambda} < 0.02$	M1		Correct formula for P(0), OR P(0 | λ = 4) at least
		$\lambda > -\ln 0.02$	M1		ln used OR λ = 3.9 at least by T & I
		= 3.912	A1		3.91(2) seen OR λ = 3.91 at least by T & I
		$0.4t = 3.912$: $t = 9.78$ minutes	M1		Divide λ by 0.4 or multiply by 150, any distribution
		t = 9 minutes 47 seconds	A1	**5**	587 seconds ± 1 sec *[inequalities not needed]*
7	(i)	$\frac{c - 4000}{60/\sqrt{50}} = 1.645$	M1		Standardise unknown with √50 or 50 *[ignore RHS]*
			B1		z = 1.645 or –1.645 seen
			A1√		Wholly correct eqn, √ on their z [1 – 1.645: M1B1A0]
		Solve	M1		Solve to find c
		c = 4014 [4013.958]	A1		Value of c, a.r.t. 4014
		Critical region is **> 4014**	A1√	**6**	Answer "> 4014", allow ≥, √ on their c, needs M1M1
	(ii)	Use "Type II is: accept when H_0 false"	M1dep		Standardise 4020 and 4014√ , allow 60^2, cc
		$\frac{4020 - 4014}{60/\sqrt{50}}$	depM1		With √50 or 50
			A1√		Completely correct LHS, √ on their c
		= 0.7071 [0.712 from 4013.958]	A1		z-value in range [0.707, 0.712]
		1 – Φ(0.7071)	M1		Normal tables, answer < 0.5
		= **0.240**[0.238 from 4013.958]	A1	**6**	Answer in range [0.2375, 0.2405]
	(iii)	Smaller	B1		"Smaller" stated, no invalidating reason
		Smaller cv, better test etc	B1	**2**	Plausible reason
	(iv)	Smaller	B1		"Smaller" stated, no invalidating reason
		Smaller cv, larger prob of Type I etc	B1	**2**	Plausible reason
	(v)	No, parent distribution known to be normal	B2	**2**	"No" stated, convincing reason SR: If B0, "No", reason that is not invalidating: B1

UNIVERSITY *of* CAMBRIDGE
Local Examinations Syndicate

4733 Statistics S2 June 2006 **MARK SCHEME: Post-standardisation**

Exemplar Answers

3 *All from "Increase because ..."*

α	More people run a faster time assuming that it is approximately normal		B1B0
β	Only 1 in 10 run slower so possible that none of them was chosen		B1B0
γ	For it to remain normal the normal curve should be symmetrical	best	B1B0
δ	A large proportion took much longer than the mean and to compensate 10% would have had to run in about 10 minutes which is impossible		B1B1

4

α	Not biased, selected fairly		B0
β	Where you assign each element a number and select using random numbers		B0

6 (i)

α	Events are independent of one another		B0
β	Customers arrive at a constant random rate		B0
γ	Customers are random	worst	B1
δ	Customers arrive singly		B1

7 (iii) *All from "Smaller because ..."*

α	As sample size increases the probability of wrongly accepting H_0 decreases		B1B0
β	Bigger n makes z-value more negative		B1B0
γ	Bigger sample means lower probability of inaccurate conclusion	worst	B1B1
δ	Bigger sample is more accurate		B1B1
ε	Denominator will be smaller		B1B1
ζ	Smaller critical value so less chance of wrongly accepting H_0		B1B1

(iv) *All from "Smaller because ..."*

α	Critical region will be larger making it less likely to reject wrongly		B1B0
β	Bigger significance level makes P(wrongly accept H_0) smaller		B1B0
γ	Less chance of concluding that H_0 is correct initially		B1B0
δ	The test statistic will increase, giving a larger value of $\bar{x}$		B1B0
ε	Smaller acceptance region so less likely to lie within it	worst	B1B1
ζ	Higher significance level will decrease probability of accepting H_0		B1B1
η	Bigger difference between critical value and new mean [with diagram]		B1B1
θ	A larger significance level means a more accurate answer		B1B1
ι	Smaller z-value so smaller critical value		B1B1

(v) *All from "not necessary to use CLT because ..."*

α	At any point the sample mean and variance are not greater than 4000		B0B0
β	The distribution is normal and the sample fairly large		B1B0
γ	The mean is normally distributed		B1B0