OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE **4733/01**

MATHEMATICS

Probability & Statistics 2

THURSDAY 12 JUNE 2008 Morning

Time: 1 hour 30 minutes

Additional materials (enclosed): None

Additional materials (required):
Answer Booklet (8 pages)
List of Formulae (MF1)

INSTRUCTIONS TO CANDIDATES

- Write your name in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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**.
- **You are reminded of the need for clear presentation in your answers.**

This document consists of **4** printed pages.

[Turn over

1 The head teacher of a school asks for volunteers from among the pupils to take part in a survey on political interests.

(i) Explain why a sample consisting of all the volunteers is unlikely to give a true picture of the political interests of all pupils in the school. [2]

(ii) Describe a better method of obtaining the sample. [3]

2 The annual salaries of employees in a company have mean £30 000 and standard deviation £12 000.

(i) Assuming a normal distribution, calculate the probability that the salary of one randomly chosen employee lies between £20 000 and £24 000. [4]

(ii) The salary structure of the company is such that a small number of employees earn much higher salaries than the others. Explain what this suggests about the use of a normal distribution to model the data. [2]

3 In a factory the time, T minutes, taken by an employee to make a single item is a normally distributed random variable with mean 28.0. A new ventilation system is installed, after which the times taken to produce a random sample of 40 items are measured. The sample mean is 26.44 minutes and it is given that $\frac{\Sigma t^2}{40} - 26.44^2 = 37.05$. Test, at the 10% significance level, whether there is evidence of a change in the mean time taken to make an item. [8]

4 The random variable U has the distribution $N(\mu, \sigma^2)$, where the value of σ is known. A test is carried out of the null hypothesis $H_0 : \mu = 50$ against the alternative hypothesis $H_1 : \mu > 50$. The test is carried out at the 1% significance level and is based on a random sample of size 10.

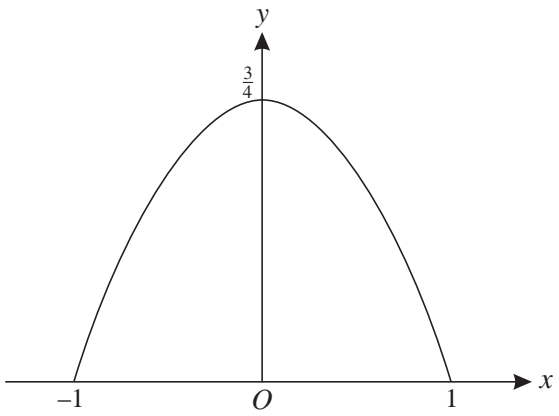
(i) The test is carried out once. The value of the sample mean is 53.0. The outcome of the test is that H_0 is not rejected. Show that $\sigma > 4.08$, correct to 3 significant figures. [4]

(ii) The test is carried out repeatedly. In each test the actual value of μ is 50. Find the probability that the first test to result in a Type I error is the fifth to be carried out. Give your answer correct to 2 significant figures. [3]

5 **(i)** A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{3}{4}(1 - x^2) & -1 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

The graph of $y = f(x)$ is shown in the diagram.



Calculate the value of $\text{Var}(X)$. [4]

(ii) A continuous random variable W has probability density function given by

$$g(x) = \begin{cases} k(9 - x^2) & -3 \leqslant x \leqslant 3, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Sketch the graph of $y = g(x)$. [1]

(b) By comparing the graphs of $y = f(x)$ and $y = g(x)$, explain how you can tell without calculation that $9k < \frac{3}{4}$. [2]

(c) State with a reason, but without calculation, whether the standard deviation of W is greater than, equal to, or less than that of X. [2]

6 **(a)** On average I receive 19 e-mails per (8-hour) working day. Assuming that a Poisson distribution is a valid model, find the probability that in one randomly chosen hour I receive either 3 or 4 e-mails. [4]

(b) **(i)** State the conditions needed to use a Poisson distribution as an approximation to a binomial distribution. [2]

(ii) 108 people each throw a pair of fair six-sided dice. Use a Poisson approximation to find the probability that at least 4 people obtain a double six. [5]

[Questions 7 and 8 are printed overleaf.]

[Turn over

7 Wendy analyses the number of 'dropped catches' in international cricket matches. She finds that the mean number of dropped catches per day is 2. In a recent 5-day match she found that there was a total of c dropped catches. She tests, at the 5% significance level, whether the mean number of dropped catches per day has increased.

(i) State conditions needed for the number of dropped catches per day to be well modelled by a Poisson distribution. [2]

Assume now that these conditions hold.

(ii) Find the probability that the test results in a Type I error.

(iii) Given that $c = 14$, carry out the test.

[10]

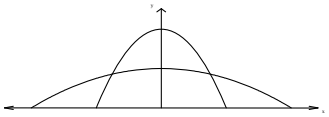
8 A company sponsors a series of concerts. Surveys show that on average 40% of audience members know the name of the sponsor. As this figure is thought to be disappointingly low, the publicity material is redesigned.

(i) After the publicity material has been redesigned, a random sample of 12 audience members is obtained, and it is found that 9 members of this sample know the name of the sponsor. Test, at the 5% significance level, whether there is evidence that the proportion of audience members who know the name of the sponsor has increased. [7]

(ii) A more detailed 5% hypothesis test is carried out, based on a random sample of size 400. This test produces significant evidence that the proportion of audience members knowing the name of the sponsor has increased. Using an appropriate approximation, calculate the smallest possible number of audience members in the sample of 400 who know the name of the sponsor. [7]

4733 Probability & Statistics 2

General: Conclusions to hypothesis tests must acknowledge uncertainty. Thus "time is unchanged" is A0. Similarly, "Significant evidence that time is unchanged" is also A0.

1	**(i)**	Biased in favour of those with strong political interest	B2	**2**	"Biased", "unrepresentative", "not indept" or equiv [but *not* "not random"] stated, with sensible reason. [SR: partial answer, B1]
	(ii)	Obtain list of all pupils Allocate numbers sequentially Choose using random numbers	B1 B1 B1	**3**	List, can be implied; number serially or randomly, not just "number pupils" Select consistently with method of numbering, not just "select randomly" [SR: systematic: List B1, every n^{th} B1, random start B1] [SR: names in a hat: B2]
2	**(i)**	$\Phi\left(\frac{24-30}{12}\right) - \Phi\left(\frac{20-30}{12}\right)$ $= \Phi(-0.5) - \Phi(-0.833)$ $= (1 - 0.6915) - (1 - 0.7976) =$ **0.1061**	M1 A1 M1 A1	**4**	Standardise one, allow $\sqrt{12}$, 12^2, $\sqrt{n}$ Both standardisations correct, allow cc here Correct handling of tails [0.3085 – 0.2024] Answer, a.r.t. 0.106, c.a.o.
	(ii)	Not symmetrical (skewed) Therefore inappropriate	M1 A1	**2**	Any comment implying not symmetric Conclude "not good model" [Partial answer: B1]
3		$H_0 : \mu = 28$ $H_1 : \mu \neq 28$ $\sigma^2 = 37.05 \times 40/39$ [= 38]	B2 M1		Both hypotheses correctly stated; one error, allow wrong or no letter, but not x or t or $\bar{x}$, B1 Multiply 37.05 or $\sqrt{37.05}$ by $n/(n-1)$ or $\sqrt{[n/(n-1)]}$
	α	$z = \frac{26.44 - 28}{\sqrt{38/40}} = -1.601$ Compare –1.645, or 0.0547 with 0.05	M1 A1 B1		Standardise with $\sqrt{n}$, allow $\sqrt{}$ errors, cc, + Correct z, a.r.t –1.60, or $p \in$ [0.0547, 0.0548] Explicit comparison of z with –1.645 or p with 0.05
	β	Critical value $28 - z\sigma/\sqrt{n}$ [= 26.397] $z = 1.645$ Compare 26.44 with 26.40	M1 B1 A1√		Allow "±", √ errors, cc, ignore other tail $z = 1.645$ in CV expression, and compare 26.44 CV, √ on their z, rounding to 3 SF correct
		Do not reject H_0 [can be implied] Insufficient evidence that time taken has changed.	M1 A1√	**8**	Needs $\sqrt{n}$, correct method & comparison, *not* $\mu = 26.44$ Conclusion interpreted in context, √ on z,
4	**(i)**	$\frac{53-50}{\sigma/\sqrt{10}} < 2.326$ $\sigma >$ **4.08** **AG** [Allow ≥]	M1 A1 B1 A1	**4**	Standardise with 10 or $\sqrt{10}$ and Φ^{-1} Both sides same sign, $\sqrt{10}$, don't worry about < 2.326 or 2.33 seen Convincingly obtain $\sigma > 4.08$ to 3 SF, one other step [SR: Substitution: standardise & substitute 4.08 M1; 0.0101 A1; 4.07 or 4.075 tried, M1; full justification A1]
	(ii)	P(Type I) = 0.01 used, e.g. Geo(0.01) $0.99^4 \times 0.01$ = **0.0096**	M1 M1 A1	**3**	Not enough merely to state $p = 0.01$ $p^4 \times q$ Answer, a.r.t. 0.0096
5	**(i)**	$\int_{-1}^{1} \frac{3}{4}(x^2 - x^4)dx = \frac{3}{4}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1}$ [= 1/5] $1/5 - 0^2$ = **1/5**	M1 A1 B1 A1	**4**	Attempt $\int_{-1}^{1} x^2 f(x)dx$ Correct indefinite integral Mean 0 clearly indicated Answer 1/5 or a.r.t. 0.200, don't need $\mu = 0$
	(ii)	(a)  (b) Areas equal, more spread out, so g_{max} lower (c) W greater as more spread out	B1 M1 A1 B1dep depB1	**5**	Correct graph, don't need f(x) as well. Don't allow if graph goes further below axis than "pips". Don't worry too much about exact shape Mention areas or total probability Convincing argument, not just "flatter" W greater… …with convincing reason

6	(a)	Po(2.375) $e^{-2.375}\left(\frac{2.375^3}{3!}+\frac{2.375^4}{4!}\right)$ [= 0.2079 + 0.1233] = **0.3310**	M1 M1 A1 A1	**4**	Po(19/8) stated or implied One correct Poisson formula, *not* tables Complete correct expression, including addition Answer, a.r.t. 0.331 [SR: Po(2) or Po(2.4) and tables, M1]
	(b)	(i) *n* large OR $n > 50$ *p* small OR $np < 5$	B1 B1	**2**	Or equivalent [Allow ≤ and ≥ throughout] Or equivalent, e.g. $np \approx npq$, or $p < 0.1$ [Treat "$np < 5$, $npq < 5$" as single wrong statement]
		(ii) B(108, $\frac{1}{36}$) ≈ Po(3) 1 – P(≤ 3) = 1 – 0.6472 = **0.3528**	M1 M1 A1 M1 A1	**5**	Correct binomial distribution stated or implied Po(*np*), √ on their *n*, *p* Po(3) Use Po tables, "1 –", *or* correct formula, ± 1 term, e.g. 0.1847; a.r.t. 0.353, allow from exact Binomial
7	(i)	Dropped catches must occur independently of one another and at constant average rate	B1 B1	**2**	"independently", in context, allow "random" "Constant average rate", in context ["Singly" doesn't gain B1]
Marks can be awarded in either part	(ii)	Use: "Reject H_0 when correct" Po(10) P(≥ 16) = 1 – P(≤ 15) = 1 – 0.9513 Probability 0.0487	M1 M1 M1 A1		Find P(≥ *r*) where $r > \lambda$, e.g. P(≥ 6) from Po(2) Po(10) stated or implied [can be recovered in (iii)] Seek biggest prob < 0.05, e.g. 0.0835 or 0.0166, allow 0.0293 but no other LH tail Answer in range [0.0487, 0.0488], cwd, cwo
	(iii)	$H_0 : \lambda = 10$ or 2 [or μ] $H_1 : \lambda > 10$ or 2 [or μ] α: P(≥ 14) = 1 – 0.8645 = 0.1355 > 0.05	B2 A1 B1		Hypotheses fully correct, allow λ or μ [SR: one error, B1, but *r* or *R* or *x* or $\bar{x}$: B0] $p \in$ [0.135, 0.136] from Po(10) Compare explicitly with 0.05 or 0.0487
		β: Critical region $r \geq 16$, $p = 0.0487$ Compare $r = 14$	A1√ B1√		√ on answer from (ii)
		Do not reject H_0 [can be implied] Insufficient evidence of an increase in the number of dropped catches	M1 A1√	**10**	Method correct, √ on *p*, must be upper tail and "≥" Conclusion interpreted in context [SR: P(≤ 14) = 0.9165 < 0.95: (B2 M1) A0 B1 M0A0; same for P(> 14) or P(= 14)] [SR: N(10,10): (ii) 0.05 M0. (iii) (B2) M1 A0 B1 M0A0]
8	(i)	$H_0 : p = 0.4$ or $\mu = 4.8$ $H_1 : p > 0.4$ or $\mu > 4.8$ B(12, 0.4) P(≥ 9) = 1 – 0.9847 = 0.0153 < 0.05 Reject H_0 [can be implied] Significant evidence of increase in proportion of audience members who know sponsor's name	B2 M1 A1 B1√ M1 A1√	**7**	Both fully correct, B2. [SR: one error, B1, but *x* or *R* or *r* or $\bar{x}$: B0] B(12, 0.4) stated or implied, e.g. 0.9972 or 0.9847 Or: CR is ≥ 9 and $p \in$ [0.015, 0.0153] Explicitly compare with 0.05, or 9 with ≥ 9, √ on < Reject H_0, √ on probability, must be "≥" Conclusion interpreted in context [SR: P(≤ 9) or P(= 9) or P(> 9): (B2 M1) A0 B1 M0A0] [SR: N(4.8, 2.88): (B2) M1 A0 B0 M0A0]
	(ii)	N(160, 96) $\frac{(x-0.5)-160}{\sqrt{96}} = 1.645$ Solve to find *x* [= 176.6] Minimum value is **177**	B1 B1 M1 A1 B1 M1 A1	**7**	Normal, mean 160 Variance (or SD) 96 [96/400: B2M0] Standardise unknown with *np* and √*npq* or *npq,* & equate to Φ^{-1}; √96 and signs correct, ignore cc RHS = 1.645 Solve [implied by 177 or 176.6 or 176.1] 177 only, from 176.6, CWO [cc error: 6 ex 7]