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

Wednesday 17 June 2009
Morning

Duration: 1 hour 30 minutes

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 random variable H has the distribution $N(\mu, \sigma^2)$. It is given that $P(H < 105.0) = 0.2420$ and $P(H > 110.0) = 0.6915$. Find the values of μ and σ, giving your answers to a suitable degree of accuracy. **[6]**

2 The random variable D has the distribution Po(20). Using an appropriate approximation, which should be justified, calculate $P(D \geqslant 25)$. **[6]**

3 An electronics company is developing a new sound system. The company claims that 60% of potential buyers think that the system would be good value for money. In a random sample of 12 potential buyers, 4 thought that it would be good value for money. Test, at the 5% significance level, whether the proportion claimed by the company is too high. **[7]**

4 A survey is to be carried out to draw conclusions about the proportion p of residents of a town who support the building of a new supermarket. It is proposed to carry out the survey by interviewing a large number of people in the high street of the town, which attracts a large number of tourists.

(i) Give two different reasons why this proposed method is inappropriate. **[2]**

(ii) Suggest a good method of carrying out the survey. **[3]**

(iii) State two statistical properties of your survey method that would enable reliable conclusions about p to be drawn. **[2]**

5 In a large region of derelict land, bricks are found scattered in the earth.

(i) State two conditions needed for the number of bricks per cubic metre to be modelled by a Poisson distribution. **[2]**

Assume now that the number of bricks in 1 cubic metre of earth can be modelled by the distribution Po(3).

(ii) Find the probability that the number of bricks in 4 cubic metres of earth is between 8 and 14 inclusive. **[3]**

(iii) Find the size of the largest volume of earth for which the probability that no bricks are found is at least 0.4. **[4]**

6 The continuous random variable R has the distribution $N(\mu, \sigma^2)$. The results of 100 observations of R are summarised by

$$\Sigma r = 3360.0, \quad \Sigma r^2 = 115\,782.84.$$

(i) Calculate an unbiased estimate of μ and an unbiased estimate of σ^2. **[4]**

(ii) The mean of 9 observations of R is denoted by $\overline{R}$. Calculate an estimate of $P(\overline{R} > 32.0)$. **[4]**

(iii) Explain whether you need to use the Central Limit Theorem in your answer to part **(ii)**. **[2]**

7 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{2}{9}x(3-x) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the variance of X. **[5]**

(ii) Show that the probability that a single observation of X lies between 0.0 and 0.5 is $\frac{2}{27}$. **[2]**

(iii) 108 observations of X are obtained. Using a suitable approximation, find the probability that at least 10 of the observations lie between 0.0 and 0.5. **[6]**

(iv) The mean of 108 observations of X is denoted by $\overline{X}$. Write down the approximate distribution of $\overline{X}$, giving the value(s) of any parameter(s). **[3]**

8 In a large company the time taken for an employee to carry out a certain task is a normally distributed random variable with mean 78.0 s and unknown variance. A new training scheme is introduced and after its introduction the times taken by a random sample of 120 employees are recorded. The mean time for the sample is 76.4 s and an unbiased estimate of the population variance is 68.9 s^2.

(i) Test, at the 1% significance level, whether the mean time taken for the task has changed. **[7]**

(ii) It is required to redesign the test so that the probability of making a Type I error is less than 0.01 when the sample mean is 77.0 s. Calculate an estimate of the smallest sample size needed, and explain why your answer is only an estimate. **[4]**

4733 Probability & Statistics 2

1	$\frac{105.0-\mu}{\sigma} = -0.7; \frac{110.0-\mu}{\sigma} = -0.5$ Solve: $\sigma = 25$ $\mu = 122.5$	M1 A1 B1 M1 A1 A1 **6**	Standardise once, equate to Φ^{-1}, allow σ^2 Both correct including signs & σ, no cc (continuity correction), allow wrong z Both correct z-values. "1 –" errors: M1A0B1 Get either μ or σ by solving simultaneously σ a.r.t. 25.0 $\mu = 122.5 \pm 0.3$ or 123 if clearly correct, allow from σ^2 but *not* from $\sigma = -25$.
2	Po(20) ≈ N(20, 20) Normal approx. valid as $\lambda > 15$ $1-\Phi\left(\frac{24.5-20}{\sqrt{20}}\right) = 1 - \Phi(1.006)$ $= 1 - 0.8427$ **= 0.1573**	M1 A1 B1 M1 A1 A1 **6**	Normal stated or implied (20, 20) or (20, √20) or (20, 20^2), can be implied "Valid as $\lambda > 15$", *or* "valid as λ large" Standardise 25, allow wrong or no cc, √20 errors $1.0 < z \leq 1.01$ Final answer, art 0.157
3	$H_0 : p = 0.6$, $H_1 : p < 0.6$ where p is proportion in population who believe it's good value $R \sim B(12, 0.6)$ α: $P(R \leq 4) = 0.0573$ > 0.05	B2 M1 A1 B1	Both, B2. Allow π, % One error, B1, except x or $\bar{x}$ or r or R: 0 B(12, 0.6) stated or implied, e.g. N(7.2, 2.88) *Not* P(< 4) or P(≥ 4) or P(= 4) Must be using P(≤ 4), or P(> 4) < 0.95 and binomial
	β: CR is ≤ 3 and 4 > 3 $p = 0.0153$	B1 A1	Must be using CR; explicit comparison needed
	Do not reject H_0. Insufficient evidence that the proportion who believe it's good value for money is less than 0.6	M1 A1 **7**	Correct conclusion, needs B(12,0.6) and ≤ 4 Contextualised, some indication of uncertainty [**SR**: N(7.2, …) or Po(7.2): poss B2 M1A0] [**SR**: P(< 4) or P(= 4) or P(≥ 4): B2 M1A0]
4 (i)	Eg "not all are residents"; "only those in street asked"	B1 B1 **2**	One valid relevant reason A definitely different valid relevant reason *Not* "not a random sample", *not* "takes too long"
(ii)	Obtain list of whole population Number it sequentially Select using random numbers [Ignore method of making contact]	B1 B1 B1 **3**	"Everyone" or "all houses" must be implied *Not* "number it with random numbers" unless then "arrange in order of random numbers" **SR**: "Take a random sample": B1 **SR**: Systematic: B1 B0, B1 if start randomly chosen
(iii)	Two of: α: Members of population equally likely to be chosen β: Chosen independently/randomly γ: Large sample (e.g. > 30)	B1 B1 **2**	One reason. NB : If "independent", must be "chosen" independently, not "views are independent" Another reason. Allow "fixed sample size" but not both that and "large sample". Allow "houses"

5	(i)	Bricks scattered at constant average rate & independently of one another	B1 B1 **2**	B1 for each of 2 different reasons, in context. (Treat "randomly" ≡ "singly" ≡ "independently")
	(ii)	Po(12) P(≤ 14) – P(≤ 7) [= .7720 – .0895] [or P(8) + P(9) + … + P(14)] **= 0.6825**	B1 M1 A1 **3**	Po(12) stated or implied Allow one out at either end or both, eg 0.617, or wrong column, but *not* from Po(3) nor, eg, .9105 – .7720 Answer in range [0.682, 0.683]
	(iii)	$e^{-\lambda} = 0.4$ $\lambda = -\ln(0.4)$ = 0.9163 Volume = 0.9163 ÷ 3 = **0.305**	B1 M1 A1 M1 **4**	This equation, aef, can be implied by, eg 0.9 Take ln, or 0.91 by T & I λ art 0.916 or 0.92, can be implied Divide their λ value by 3 [**SR**: Tables, eg 0.9÷3: B1 M0 A0 M1]
6	(i)	33.6 $\frac{115782.84}{100} - 33.6^2$ [= 28.8684] $\times\frac{100}{99}$ **= 29.16**	B1 M1 M1 A1 **4**	33.6 clearly stated [not recoverable later] Correct formula used for biased estimate $\times\frac{100}{99}$, M's independent. Eg $\frac{\Sigma r^2}{99}[-33.6^2]$ **SR** B1 variance in range [29.1, 29.2]
	(ii)	$\bar{R} \sim N(33.6, 29.16/9)$ $= N(33.6, 1.8^2)$ $1-\Phi\left(\frac{32-33.6}{\sqrt{3.24}}\right)$ [= Φ(0.8889)] **= 0.8130**	M1 A1 M1 A1 **4**	Normal, their μ, stated or implied Variance [their (i)]÷9 [*not* ÷100] Standardise & use Φ, 9 used, answer > 0.5, allow √ errors, allow cc 0.05 but *not* 0.5 Answer, art 0.813
	(iii)	No, distribution of R is normal so that of $\bar{R}$ is normal	B2 **2**	Must be saying this. Eg "9 is not large enough": B0. Both: B1 max, unless saying that n is irrelevant.
7	(i)	$\frac{2}{9}\int_0^3 x^3(3-x)dx = \frac{2}{9}\left[\frac{3x^4}{4} - \frac{x^5}{5}\right]_0^3$ [= 2.7] – $(1\frac{1}{2})^2$ $= \frac{9}{20}$ or **0.45**	M1 A1 B1 M1 A1 **5**	Integrate $x^2f(x)$ from 0 to 3 [*not* for μ] Correct indefinite integral Mean is 1½, soi [not recoverable later] Subtract their μ^2 Answer art 0.450
	(ii)	$\frac{2}{9}\int_0^{0.5} x(3-x)dx = \frac{2}{9}\left[\frac{3x^2}{2} - \frac{x^3}{3}\right]_0^{0.5}$ $= \frac{2}{27}$ AG	M1 A1 **2**	Integrate f(x) between 0, 0.5, must be seen somewhere Correctly obtain given answer $\frac{2}{27}$, decimals other than 0.5 not allowed, 1 more line needed (eg [ ] = ⅓)
	(iii)	B(108, $\frac{2}{27}$) ≈ N(8, 7.4074) $1-\Phi\left(\frac{9.5-8}{\sqrt{7.4074}}\right)$ = 1 – Φ(0.5511) **= 0.291**	B1 M1 A1 M1 A1 A1 **6**	B(108, $\frac{2}{27}$) seen or implied, eg Po(8) Normal, mean 8 … … variance (or SD) 200/27 or art 7.41 Standardise 10, allow √ errors, wrong or no cc, needs to be using B(108,…) Correct √ and cc Final answer, art 0.291

	(iv)	$\bar{X} \sim N(1.5, \frac{1}{240})$	B1 B1√ B1√ **3**	Normal NB: *not* part (iii) Mean their μ Variance or SD (their 0.45)/108 [not (8, 50/729)]
8	(i)	$H_0 : \mu = 78.0$ $H_1 : \mu \neq 78.0$ $z = \frac{76.4 - 78.0}{\sqrt{68.9/120}} = -2.1115$ > -2.576 or $0.0173 > 0.005$	B1 B1 M1 A1 B1	Both correct, B2. One error, B1, but x or $\bar{x}$: B0. Needs $\pm(76.4 - 78)/\sqrt{(\sigma \div 120)}$, allow √ errors art –2.11, or $p = 0.0173 \pm 0.0002$ Compare z with (–)2.576, or p with 0.005
		$78 \pm z\sqrt{(68.9/120)}$ $= 76.048$ $76.4 > 76.048$	M1 A1√ B1	Needs 78 and 120, can be – only Correct CV to 3 sf, √ on z $z = 2.576$ and compare 76.4, allow from 78 ↔ 76.4
		Do not reject H_0. Insufficient evidence that the mean time has changed	M1 A1√ **7**	Correct comparison & conclusion, needs 120, "like with like", correct tail, $\bar{x}$ and μ right way round Contextualised, some indication of uncertainty
	(ii)	$\frac{1}{\sqrt{68.9/n}} > 2.576$ $\sqrt{n} > 21.38$, $n_{\min} =$ **458** Variance is estimated	M1 M1 A1 B1 **4**	IGNORE INEQUALITIES THROUGHOUT Standardise 1 with n and 2.576, allow √ errors, cc etc but *not* 2.326 Correct method to solve for $\sqrt{n}$ (*not* from n) 458 only (*not* 457), *or* 373 from 2.326, signs correct Equivalent statement, allow "should use t". In principle nothing superfluous, but "variance stays same" B1 bod

Specimen Answers

Question 4: Part (i)

α	Takes too long/too slow	B0
β	Interviewing people in the street isn't a random sample	B0
γ	Many tourists so not representative	B1
δ	Those who don't shop won't have their views considered	B1
ε	Interviewers biased as to who they ask	B1
ζ	Views influenced by views of others	B1

Part (ii)

α	Choose a random sample of the town and ask their opinion	B1
β	Choose names at random from the town's phone book	B1
γ	A random number machine determines which house numbers should be used, and every street should have the same proportion of residents interviewed	B0B0B1
δ	Visit everyone door to door and give them a questionnaire	B1B0B0
ε	Assign everyone a number and select randomly	B1B0B0
ζ	Assign everyone a number and select using random numbers	B1B0B1
η	Ditto + "ignoring numbers that don't correspond to a resident"	B1B1B1
θ	Assign each eligible person a number and pick numbers from a hat	B1B1B0
ι	Put names of all residents into a hat and pick them out [NB: postal survey is biased]	B1B1B0

Part (iii)

α	One person's view should not affect another's	B0
β	It is without bias	B0
γ	Results occur randomly	B0
δ	Should be asked if they are for or against (binomial testing)	B0
ε	It will survey a diverse group from different areas so should be representative	B0
ζ	Everyone's should be chose independently of everyone else	B1
η	The sample size must be large	B1
θ	Participants are chosen at random and independently from one another [though η & θ together would get B2]	B1 only

Question 5 (i)

α	Number of bricks must always be the same		B0
β	Results occur randomly		B0
γ	The chance of a brick being in one place is always the same		B0
δ	Events must occur independently and at constant average rate		B0
ε	They must occur independently and at constant average rate		B1 only
ζ	Bricks' locations must be random and independent	*[effectively the same]*	B1 only
η	Only one brick in any one place; bricks independent	*[effectively the same]*	B1 only