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:
- Scientific or graphical calculator

Tuesday 22 June 2010
Afternoon

Duration: 1 hour 30 minutes

* 4 7 3 3 *

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 **(i)** The number of inhabitants of a village who are selected for jury service in the course of a 10-year period is a random variable with the distribution Po(4.2).

(a) Find the probability that in the course of a 10-year period, at least 7 inhabitants are selected for jury service. **[2]**

(b) Find the probability that in 1 year, exactly 2 inhabitants are selected for jury service. **[3]**

(ii) Explain why the number of inhabitants of the village who contract influenza in 1 year can probably not be well modelled by a Poisson distribution. **[2]**

2 A university has a large number of students, of whom 35% are studying science subjects. A sample of 10 students is obtained by listing all the students, giving each a serial number and selecting by using random numbers.

(i) Find the probability that fewer than 3 of the sample are studying science subjects. **[3]**

(ii) It is required that, in selecting the sample, the same student is not selected twice. Explain whether this requirement invalidates your calculation in part **(i)**. **[2]**

3 Tennis balls are dropped from a standard height, and the height of bounce, H cm, is measured. H is a random variable with the distribution N(40, σ^2). It is given that P($H < 32$) = 0.2.

(i) Find the value of σ. **[3]**

(ii) 90 tennis balls are selected at random. Use an appropriate approximation to find the probability that more than 19 have $H < 32$. **[6]**

4 The proportion of commuters in a town who travel to work by train is 0.4. Following the opening of a new station car park, a random sample of 16 commuters is obtained, and 11 of these travel to work by train. Test at the 1% significance level whether there is evidence of an increase in the proportion of commuters in this town who travel to work by train. **[7]**

5 The time T seconds needed for a computer to be ready to use, from the moment it is switched on, is a normally distributed random variable with standard deviation 5 seconds. The specification of the computer says that the population mean time should be not more than 30 seconds.

(i) A test is carried out, at the 5% significance level, of whether the specification is being met, using the mean $\bar{t}$ of a random sample of 10 times.

(a) Find the critical region for the test, in terms of $\bar{t}$. **[4]**

(b) Given that the population mean time is in fact 35 seconds, find the probability that the test results in a Type II error. **[3]**

(ii) Because of system degradation and memory load, the population mean time μ seconds increases with the number of months of use, m. A formula for μ in terms of m is $\mu = 20 + 0.6m$. Use this formula to find the value of m for which the probability that the test results in rejection of the null hypothesis is 0.5. **[4]**

6 **(a)** The random variable D has the distribution Po(24). Use a suitable approximation to find $P(D > 30)$. **[5]**

(b) An experiment consists of 200 trials. For each trial, the probability that the result is a success is 0.98, independent of all other trials. The total number of successes is denoted by E.

(i) Explain why the distribution of E cannot be well approximated by a Poisson distribution. **[1]**

(ii) By considering the number of failures, use an appropriate Poisson approximation to find $\mathrm{P}(E \leqslant 194)$. **[4]**

7 A machine is designed to make paper with mean thickness 56.80 micrometres. The thicknesses, x micrometres, of a random sample of 300 sheets are summarised by

$$n = 300, \qquad \Sigma x = 17\,085.0, \qquad \Sigma x^2 = 973\,847.0.$$

Test, at the 10% significance level, whether the machine is producing paper of the designed thickness. **[11]**

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

$$\mathrm{f}(x) = \begin{cases} kx^{-a} & x \geqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are constants and a is greater than 1.

(i) Show that $k = a - 1$. **[3]**

(ii) Find the variance of X in the case $a = 4$. **[5]**

(iii) It is given that $\mathrm{P}(X < 2) = 0.9$. Find the value of a, correct to 3 significant figures. **[4]**

1	(i)(a)	$1 - P(\leq 6) = 1 - 0.8675$ $= \mathbf{0.1325}$	M1 A1	**2**	$1 - .9361$ or $1 - .8786$ or $1 - .8558$: M1. .9721: M0 Or 0.132 or 0.133
	(b)	Po(0.42) $e^{-0.42}\ \frac{0.42^2}{2!}$ $= \mathbf{0.05795}$	M1 M1 A1	**3**	Po(0.42) stated or implied Correct formula, any numerical λ Answer, art 0.058. Interpolation in tables: M1B2
	(ii)	E.g. "Contagious so incidences do not occur independently", or "more cases in winter so not at constant average rate"	B2	**2**	Contextualised reason, referred to conditions: B2. No marks for mere learnt phrases or spurious reasons, e.g. not just "independently, singly and constant average rate". See notes.
2	(i)	B(10, 0.35) P(< 3) $= \mathbf{0.2616}$	M1 M1 A1	**3**	B(10, 0.35) stated or implied Tables used, e.g. 0.5138 or 0.3373, or formula ± 1 term Answer 0.2616 or better or 0.262 only
	(ii)	Binomial requires being chosen independently, which this is not, but unimportant as population is large	B2	**2**	Focus on "Without replacement" negating independence condition. It doesn't negate "constant probability" condition but can allow B1 if "selected". See notes
3	(i)	$\left(\frac{32-40}{\sigma}\right) = \Phi^{-1}(0.2) = -0.842$ $\sigma = 9.5[06]$	M1 B1 A1	**3**	Standardise and equate to Φ^{-1}, allow "1 –" errors, σ^2, cc 0.842 seen Answer, 9.5 or in range [9.50, 9.51], c.w.o.
	(ii)	B(90, 0.2) $\approx$ N(18, 14.4) $1 - \Phi\left(\frac{19.5-18}{\sqrt{14.4}}\right) = 1 - \Phi(0.3953)$ $= 1 - 0.6537$ $= \mathbf{0.3463}$	B1 M1 A1 M1 A1 A1	**6**	B(90, 0.2) stated or implied N, their *np* … … variance their *npq*, allow √ errors Standardise with *np* and *npq*, allow √, cc errors, e.g. .396, .448, .458, .486, .472; $\sqrt{npq}$ and cc correct Answer, a.r.t. 0.346 [NB: 0.3491 from Po: 1/6]
4	(α)	$H_0 : p = 0.4$, $H_1 : p > 0.4$ $R \sim$ B(16, 0.4): $P(R \geq 11) = 0.0191$ > 0.01	B1 B1 M1 A1 A1		Fully correct, B2. Allow π. p omitted or μ used in both, or > wrong: B1 only. x or $\bar{x}$ or 6.4 etc: B0 B(16, 0.4) stated or implied, allow N(6.4, 3.84) Allow for $P(\leq 10) = 0.9808$, and < 0.99, or $z = 2.092$ or $p = 0.018$, but *not* $P(\leq 11) = 0.9951$ or $P(= 11) = 0.0143$ Explicit comp with .01, or z<2.326, *not* from ≤11 or =11
	(β)	CR $R \geq 12$ and $11 < 12$ Probability 0.0049	A1 A1		Must be clear that it's ≥ 12 and not ≤ 11 Needs to be seen, allow 0.9951 here, or $p = .0047$ from N
		Do not reject H_0. Insufficient evidence that proportion of commuters who travel by train has increased	M1 A1 FT	**7**	Needs like-with-like, $P(R \geq 11)$ or CR $R \geq 12$ Conclusion correct on their p or CR, contextualised, not too assertive, e.g. "evidence that" needed. Normal, $z = 2.34$, "reject" [no cc] can get 6/7
5	(i)	(a) $30 + 1.645 \times \frac{5}{\sqrt{10}}$ $= 32.6$ Therefore critical region is $\bar{t} > 32.6$	M1 B1 A1 A1 FT	**4**	$30 + 5z/\sqrt{10}$, allow ± but not just –, allow √ errors $z = 1.645$ seen, allow – Critical value, art 32.6 " > c" or "≥ c", FT on c provided > 30, can't be recovered. Withhold if not clear which is CR
		(b) $P(\bar{t} < 32.6 \mid \mu = 35)$ $\frac{32.6-35}{5/\sqrt{10}}$ [= –1.5178] **0.0645**	M1* dep*M1 A1	**3**	Need their c, final answer < 0.5 and $\mu = 35$ at least, but allow answer > 0.5 if consistent with their (i) Standardise their CV with 35 and √10 or 10 Answer in range [0.064, 0.065], or 0.115 from 1.96 in (a)
	(ii)	$(32.6 - \mu) = 0$ $\mu = 32.6$ $20 + 0.6m = 32.6$ $m =$ **21**	M1 A1 FT M1 A1	**4**	Standardise c with μ, equate to Φ^{-1}, can be implied by: μ = their c Equate and solve for m, allow from 30 or 35 Answer, a.r.t. 21, c.a.o. MR: 0.05: M1 A0 M1, 16.7 A1 FT Ignore variance throughout (ii)

6	(a)	N(24, 24) $1-\Phi\left(\frac{30.5-24}{\sqrt{24}}\right)=1-\Phi(1.327)$ $=$ **0.0923**	B1 B1 M1 A1 A1	**5**	Normal, mean 24 stated or implied Variance or SD equal to mean Standardise 30 with λ and √λ, allow cc or √ errors, e.g. .131 or .1103 ; 30.5 and √λ correct Answer in range [0.092, 0.0925]
	(b)(i)	p or np [= 196] is too large	B1	**1**	Correct reason, no wrong reason, don't worry about 5 or 15
	(ii)	Consider $(200 - E)$ $(200 - E) \sim$ Po(4) P(≥ 6) [= 1 – 0.7851] $=$ **0.2149**	M1 M1 M1 A1	**4**	Consider complement Po(200×0.02) Poisson tables used, correct tail, e.g. 0.3712 or 0.1107 Answer a.r.t. 0.215 only
7		$H_0 : \mu = 56.8$ $H_1 : \mu \neq 56.8$ $\bar{x} = 17085/300 = 56.95$ $\frac{300}{299}\left(\frac{973847}{300}-56.95^2\right)$ $= 2.8637\ldots$	B2 B1 M1 M1 A1		Both correct One error: B1, but *not* $\bar{x}$, etc 56.95 or 57.0 seen or implied Biased [2.8541] : M1M0A0 Unbiased estimate method, allow if ÷ 299 seen anywhere Estimate, a.r.t. 2.86 [not 2.85]
	(α)	$z=\frac{56.95-56.8}{\sqrt{2.8637/300}}=1.535$ 1.535 < 1.645 or 0.0624 > 0.05	M1 A1 A1		Standardise with √300, allow √ errors, cc $z \in$ [1.53, 1.54] or $p \in$ [0.062, 0.063], *not* – 1.535 Compare explicitly z with 1.645 or p with 0.05, or $2p > 0.1$, not from $\mu = 56.95$
	(β)	CV $56.8 \pm 1.645 \times \sqrt{\frac{2.8637}{300}}$ 56.96 > 56.95	M1 A1 A1 FT		56.8 + $z\sigma/\sqrt{300}$, needn't have ±, allow √ errors $z = 1.645$ $c = 56.96$, FT on z, and compare 56.95 [$c_L = 56.64$]
		Do not reject H_0; insufficient evidence that mean thickness is wrong	M1 A1 FT	**11**	Consistent first conclusion, needs 300, correct method and comparison Conclusion stated in context, not too assertive, e.g. "evidence that" needed
8	(i)	$\int_1^\infty kx^{-a}\,dx=\left[k\frac{x^{-a+1}}{-a+1}\right]_1^\infty$ Correctly obtain $k = a - 1$ **AG**	M1 B1 A1	**3**	Integrate f(x), limits 1 and ∞ (at some stage) Correct indefinite integral Correctly obtain given answer, don't need to see treatment of ∞ but mustn't be wrong. Not k^{-a+1}
	(ii)	$\int_1^\infty 3x^{-3}dx=\left[3\frac{x^{-2}}{-2}\right]_1^\infty=1\frac{1}{2}$ $\int_1^\infty 3x^{-2}dx=\left[3\frac{x^{-1}}{-1}\right]_1^\infty-(1\frac{1}{2})^2$ Answer ¾	M1 M1 A1 M1 A1	**5**	Integrate xf(x), limits 1 and ∞ (at some stage) [x^4 is *not* MR] Integrate x^2f(x), correct limits Either $\mu = 1\frac{1}{2}$ or E(X^2) = 3 stated or implied, allow k, $k/2$ Subtract their numerical μ^2, allow letter if subs later Final answer ¾ or 0.75 only, cwo, e.g. not from $\mu = -1\frac{1}{2}$. [SR: Limits 0, 1: can get (i) B1, (ii) M1M1M1]
	(iii)	$\int_1^2 (a-1)x^{-a}dx=\left[-x^{-a+1}\right]_1^2=0.9$ $1-\frac{1}{2^{a-1}}=0.9,\ 2^{a-1}=10$ $a = 4.322$	M1* dep*M1 M1 indept A1	**4**	Equate ∫f(x)dx, one limit 2, to 0.9 or 0.1. [Normal: 0 ex 4] Solve equation of this form to get 2^{a-1} = number Use logs or equivalent to solve 2^{a-1} = number Answer, a.r.t. 4.32. T&I: (M1M1) B2 or B0

Specimen Verbal Answers

1

α	"Cases of infection must occur randomly, independently, singly and at constant average rate"	B0
β	Above + "but it is contagious"	B1
γ	Above + "but not independent as it is contagious"	B2
δ	"Not independent as it is contagious"	B2
ε	"Not constant average rate", or "not independent"	B0
λ	"Not constant average rate because contagious" *[needs more]*	B1
ζ	"Not constant average rate because more likely at certain times of year"	B2
μ	Probabilities changes because of different susceptibilities	B0
ν	Not constant average rate because of different susceptibilities	B2
η	Correct but with unjustified or wrong extra assertion *[scattergun]*	B1
θ	More than one correct assertion, all justified	B2
π	Valid reason (e.g. "contagious") but not referred to conditions	B1

[Focus is on explaining why the required assumptions might not apply. No credit for regurgitating learnt phrases, such as "events must occur randomly, independently, singly and at constant average rate, even if contextualised.]

2 Don't need either "yes" or "no".

α	"No it doesn't invalidate the calculation" *[no reason]*	B0
β	"Binomial requires not chosen twice" *[false]*	B0
γ	"Probability has to be constant but here the probabilities change"	B0
δ	Same but "probability <u>of being chosen</u>" *[false, but allow B1]*	B1
ε	"Needs to be independently chosen but probabilities change" *[confusion]*	B0
ζ	"Needs to be independent but one choice affects another" *[correct]*	B2
η	"The sample is large so it makes little difference" *[false]*	B0
θ	"The population is large so it makes little difference" *[true]*	B2
λ	Both correct and wrong reasons (scattergun approach)	B1

[Focus is on modelling conditions for binomial: On every choice of a member of the sample, each member of the population is equally likely to be chosen; and each choice is independent of all other choices.
Recall that in fact even without replacement the probability that any one person is chosen is the same for each choice. Also, the binomial "independence" condition <u>does</u> require the possibility of the same person being chosen twice.]