OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS

4733

Probability & Statistics 2

QUESTION PAPER

Candidates answer on the printed answer book.

OCR supplied materials:
- Printed answer book 4733
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Friday 14 January 2011
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the printed answer book and the question paper.

- The question paper will be found in the centre of the printed answer book.
- Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- 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.

INFORMATION FOR CANDIDATES

This information is the same on the printed answer book and the question paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the question paper.
- **You are reminded of the need for clear presentation in your answers.**
- The total number of marks for this paper is **72**.
- The printed answer book consists of **16** pages. The question paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this question paper for marking; it should be retained in the centre or destroyed.

1 A random sample of nine observations of a random variable is obtained. The results are summarised as

$$\Sigma x = 468, \quad \Sigma x^2 = 24\,820.$$

Calculate unbiased estimates of the population mean and variance. **[4]**

2 The random variable H has the distribution $N(\mu, 5^2)$. The mean of a sample of n observations of H is denoted by $\overline{H}$. It is given that $P(\overline{H} > 53.28) = 0.0250$ and $P(\overline{H} < 51.65) = 0.0968$, both correct to 4 decimal places. Find the values of μ and n. **[6]**

3 The probability that a randomly chosen PPhone has a faulty casing is 0.0228. A random sample of 200 PPhones is obtained. Use a suitable approximation to find the probability that the number of PPhones in the sample with a faulty casing is 2 or fewer. Justify your approximation. **[6]**

4 The continuous random variable X has mean μ and standard deviation 45. A significance test is to be carried out of the null hypothesis H_0: $\mu = 230$ against the alternative hypothesis H_1: $\mu \neq 230$, at the 1% significance level. A random sample of size 50 is obtained, and the sample mean is found to be 213.4.

(i) Carry out the test. **[5]**

(ii) Explain whether it is necessary to use the Central Limit Theorem in your test. **[2]**

5 A temporary job is advertised annually. The number of applicants for the job is a random variable which is known from many years' experience to have a distribution Po(12). In 2010 there were 19 applicants for the job. Test, at the 10% significance level, whether there is evidence of an increase in the mean number of applicants for the job. **[7]**

6 The number of randomly occurring events in a given time interval is denoted by R. In order that R is well modelled by a Poisson distribution, it is necessary that events occur independently.

(i) Let R represent the number of customers dining at a restaurant on a randomly chosen weekday lunchtime. Explain what the condition 'events occur independently' means in this context, and give a reason why it would probably not hold in this context. **[2]**

Let D represent the number of tables booked at the restaurant on a randomly chosen day. Assume that D can be well modelled by the distribution Po(7).

(ii) Find $P(D < 5)$. **[2]**

(iii) Use a suitable approximation to find the probability that, in five randomly chosen days, the total number of tables booked is greater than 40. **[6]**

7 Two continuous random variables S and T have probability density functions f_S and f_T given respectively by

$$f_S(x) = \begin{cases} \dfrac{a}{x^2} & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise,} \end{cases}$$

$$f_T(x) = \begin{cases} b & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise,} \end{cases}$$

where a and b are constants.

(i) Sketch on the same axes the graphs of $y = f_S(x)$ and $y = f_T(x)$. **[3]**

(ii) Find the value of a. **[3]**

(iii) Find $E(S)$. **[3]**

(iv) A student gave the following description of the distribution of T: "The probability that T occurs is constant". Give an improved description, in everyday terms. **[1]**

8 A company has 3600 employees, of whom 22.5% live more than 30 miles from their workplace. A random sample of 40 employees is obtained.

(i) Use a suitable approximation, which should be justified, to find the probability that more than 5 of the employees in the sample live more than 30 miles from their workplace. **[8]**

(ii) Describe how to use random numbers to select a sample of 40 from a population of 3600 employees. **[3]**

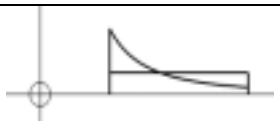
9 A pharmaceutical company is developing a new drug to treat a certain disease. The company will continue to develop the drug if the proportion p of those who have the disease and show a substantial improvement after treatment is greater than 0.7. The company carries out a test, at the 5% significance level, on a random sample of 14 patients who suffer from the disease.

(i) Find the critical region for the test. **[3]**

(ii) Given that 12 of the 14 patients in the sample show a substantial improvement, carry out the test. **[5]**

(iii) Find the probability that the test results in a Type II error if in fact $p = 0.8$. **[3]**

1		$\hat{\mu} = \bar{x} = \frac{468}{9} = 52$ $\frac{24820}{9} - 52^2$ [= 53.78] $\hat{\sigma}^2 = \frac{9}{8} \times 53.78$ = **60.5**	B1 M1 M1 A1 **4**	52 stated Correct method for biased estimator Multiply by 9/8 [if single formula, allow M0 M1 if wrong but divisor 8 seen anywhere] Answer 60.5 or exact equivalent
2		$\frac{53.28 - \mu}{5/\sqrt{n}} = 1.96$ $\frac{\mu - 51.65}{5/\sqrt{n}} = 1.3$ $\sqrt{n} = 10$, n = **100** μ = **52.3**	M1dep A1 B1 depM1 A1 B1 **6**	Standardise with $\sqrt{n}$ once & equate to z, allow sign, square/$\sqrt{}$ errors twice, signs correct, zs may be wrong Both correct z values seen Solve to get $\sqrt{n}$ or μ, needs first M1 n = 100, not from wrong signs a.r.t. 52.3, right arithmetic needed but $\sqrt{n}$ can be omitted
3		B(200, 0.0228) Po(4.56) $e^{-4.56}(1 + 4.56 + \frac{4.56^2}{2})$ **= 0.167** n large *or* $n > 50$; p small *or* $np < 5$	M1 A1 M1 A1 A1 B1 **6**	B(200, 0.0228) stated or implied Po(4.56) stated or implied, allow 4.6 here Correct formula for P(≤ 2) ± 1 term, any λ (tables: M0) Correct formula, 4.56 needed Answer, a.r.t. 0.167 [0.16694] Both, can be merely asserted. If numbers, must be these SR interpolation: clear method M1, answer A2 MR: typically B(200, 0.228) ≈ N(45.6, 3.52): M1A1; standardise correctly, M1; state np, $nq > 5$, B1
4	(i)	*Either* $z = \frac{213.4 - 230}{45/\sqrt{50}}$ = –2.608 –2.608 < –2.576 *or* 0.0047 < 0.005	M1 A1 B1	Standardise z with $\sqrt{50}$, ignore sign or $\sqrt{}$ or squaring errors z-value, a.r.t. –2.61, or p in range [0.0044, 0.005) Correctly compare (–)2.576, signs consistent, *or p* explicitly with 0.005
	Or	CV is $230 - 2.576 \times \frac{45}{\sqrt{50}} = 213.6$ 213.4 < 213.6	M1 B1 A1	$230 - z\sigma/\sqrt{50}$, allow $\sqrt{}$ or squaring errors, allow ± but not just +; $z = 2.576$ Explicitly compare 213.4 with 213.6
		Reject H_0. Significant evidence that population mean is not 230	M1 A1 FT **5**	"Reject", FT, needs correct method and form of comparison; interpreted, acknowledge uncertainty
	(ii)	Yes, population distribution is not known to be normal	B2 **2**	*Not*, "yes, sample size is large" but ignore "*can* use it as …" SR: Both right and wrong answers: B1 α "Yes as it must be assumed normal": B1
5		H_0: $\lambda = 12$; H_1: $\lambda > 12$ *Either* : P(≥ 19) = 1 – P(≤ 18) = 1 – 0.9626 = 0.0374 < 0.1	B2 M1 A1 B1	Both correct: B2. Allow μ. One error, B1, but *not* x, r etc. Po(12) stated or implied, e.g. 0.9787 0.0374, *or* 0.9626 if compared with 0.9 Explicitly compare P(≥ 19) with 0.1, *or* P(≤ 18) with 0.9
		Or: CR is ≥ 18, p = 0.063 19 ≥ 18	A1 B1	≥ 18 and 0.063 stated Explicit comparison of CV (right-hand CR) with 19
		Reject H_0. Significant evidence of increase in mean number of applicants	M1 A1 FT **7**	"Reject" FT, needs correct method and comparison, e.g. *not* from ≤ 19 or = 19, withhold if inconsistent Interpreted in context, acknowledge uncertainty

6	(i)	If one customer arrives, it does not change the probability that another one does so; customers probably arrive in groups of at least 2	B1 B1 **2**	Answer that shows correct understanding of "independent", in context; *not* just equivalent to "singly" Plausible reason, in context, nothing wrong, nothing that suggests "constant average rate"
	(ii)	**0.1730**	M1 A1 **2**	Correct use of tables or formula, e.g. .3007, or .4405 from Po(5) if Po(7) stated; answer 0.173, 0.1730 or better
	(iii)	Po(35) N(35, 35) $1-\Phi\left(\frac{40.5-35}{\sqrt{35}}\right) = 1-\Phi(0.9297)$ = **0.1763**	B1 M1 A1 M1 A1 A1 **6**	Po(5×7) stated or implied Normal, μ = their λ Both parameters correct, allow 35^2, $\sqrt{35}$ Standardise 40 with λ, $\sqrt{\lambda}$, allow $\sqrt{}$, cc errors Both $\sqrt{\lambda}$ and cc correct Answer, a.r.t. 0.176 [penalise 0.1765]
7	(i)		B1 B1 B1 **3**	Horizontal line above axis Concave decreasing curve above axis Both correct including approx relationship, not extending beyond [1, 3], verticals and scale not needed
	(ii)	$\int_1^3 \frac{a}{x^2}\mathrm{d}x = 1,\ \left[\frac{-a}{x}\right]_1^3 = 1;\ a = \frac{3}{2}$	M1 B1 A1 **3**	Attempt $\int f_X(x)\mathrm{d}x$, limits 1, 3 at some stage, and equate to 1 Correct indefinite integral Correctly obtain 3/2 or 1.5 or exact equivalent
	(iii)	$\int_1^3 \frac{a}{x}\mathrm{d}x = [a\ln x]_1^3$ $= \frac{3}{2}\ln 3$	M1 B1 FT A1 FT **3**	Attempt $\int x f_X(x)\mathrm{d}x$, limits 1, 3 at some stage Correct indefinite integral, FT on a Answer, any exact equivalent or a.r.t 1.65, FT on a, or $a\ln 3$
	(iv)	T is equally likely to take any value between 1 and 3	B1 **1**	Must be "values taken by T" (or "of T") or clear equivalent Any hint that they think T is an *event* gets B0. α "Same chance of occurring anywhere between 1 and 3": 0 β "For values of T between 1 and 3, T is equally likely": 0 γ "Each value of T is equally likely to occur": 1
8	(i)	B(40, 0.225) ≈ N(9, 6.975) $\frac{5.5-9}{\sqrt{6.975}} = -1.325$ **0.9074** $np = 9 > 5$ *or* n large; and $nq = 31 > 5$ *or* p close to 0.5	M1 M1 A1 M1 A1 A1 B2 **8**	B(40, 0.225) stated or implied Normal, mean 9 Variance 6.975 or SD 2.641 or 6.975 Standardise with np and $\sqrt{npq}$, allow npq, no or wrong cc CC and $\sqrt{npq}$ correct, allow from N(3600, 0.225) Answer, in range [0.907, 0.908] Full conditions B2; partial, B1 (assertions OK). Allow npq, allow from e.g. n = 3600
	(ii)	Number list sequentially and select using random numbers If # > 3600, ignore (etc)	B1 B1 B1 **3**	Number list, don't need "sequentially" Mention random numbers (*not* "select numbers randomly") Deal with issue of # > 3600, *or* "ignore repeats" α "Randomly pick numbers from 0 to 3599": (B1) B0 B1
9	(i)	B(14, 0.7) CR is **≥ 13** with probability 0.0475	M1 A1 A1 **3**	B(14, 0.7) stated or implied, e.g. N(9.8, 2.94), can be recovered CV 13, or > 12 or {13, 14}, allow = but no other inequalities Exactly correct CR, and supporting prob .0475 or .9525 seen
	(ii)	H_0: $p = 0.7$, H_1: $p > 0.7$ 12 < 13 Do not reject H_0. Insufficient evidence that proportion who show improvement is greater than 0.7	B2 B1 M1 A1 FT **5**	Both, B2. Allow π. One error, B1, but r, x etc: B0 Compare CV *from correct tail and inequality* with 12, *or* P(≥ 12) = 0.1608 and > 0.05 *or* P(< 12) = 0.8392 and < 0.95 Correct method & conclusion, requires like-with-like; CV method needs ≥ 13 or < 12; p method needs ≥ 12 or < 12 Withhold if inconsistent Contextualised, acknowledge uncertainty [SR: Normal or Po: (i) M1, (ii) B2 maximum] [0.9932 or 0.0068 probably B2 maximum]
	(iii)	B(14, 0.8) P(≤ 12) from B(14, 0.8) **0.8021**	M1 M1 A1 **3**	B(14, 0.8) stated or implied, allow from B(14, 0.75) Attempt prob of acceptance region, e.g. 0.8990, √ on (i) Answer 0.802 or a.r.t. 0.8021