OCR
RECOGNISING ACHIEVEMENT

Friday 18 January 2013 – Afternoon

A2 GCE MATHEMATICS

4733/01 Probability & Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4733/01
- List of Formulae (MF1)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4738630113

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. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- 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 **12** 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 recycled. Please contact OCR Copyright should you wish to re-use this document.

1 A random variable has the distribution B(n, p). It is required to test $H_0: p = \frac{2}{3}$ against $H_1: p < \frac{2}{3}$ at a significance level as close to 1% as possible, using a sample of size $n = 8$, 9 or 10. Use tables to find which value of n gives such a test, stating the critical region for the test and the corresponding significance level. **[4]**

2 A random variable C has the distribution $N(\mu, \sigma^2)$. A random sample of 10 observations of C is obtained, and the results are summarised as

$$n = 10, \Sigma c = 380, \Sigma c^2 = 14602.$$

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

(ii) Hence calculate an estimate of the probability that $C > 40$. **[2]**

3 A factory produces 9000 music DVDs each day. A random sample of 100 such DVDs is obtained.

(i) Explain how to obtain this sample using random numbers. **[3]**

(ii) Given that 24% of the DVDs produced by the factory are classical, use a suitable approximation to find the probability that, in the sample of 100 DVDs, fewer than 20 are classical. **[5]**

4 A continuous random variable X has probability density function

$$f(x) = \begin{cases} kx & 0 \leqslant x \leqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are constants.

(i) State what the letter x represents. **[1]**

(ii) Find k in terms of a. **[2]**

(iii) Find Var(X) in terms of a. **[6]**

5 In a mine, a deposit of the substance *pitchblende* emits radioactive particles. The number of particles emitted has a Poisson distribution with mean 70 particles per second. The warning level is reached if the total number of particles emitted in one minute is more than 4350.

(i) A one-minute period is chosen at random. Use a suitable approximation to show that the probability that the warning level is reached during this period is 0.01, correct to 2 decimal places. You should calculate the answer correct to 4 decimal places. **[5]**

(ii) Use a suitable approximation to find the probability that in 30 one-minute periods the warning level is reached on at least 4 occasions. (You should use the given rounded value of 0.01 from part **(i)** in your calculation.) **[3]**

6 Gordon is a cricketer. Over a long period he knows that his population mean score, in number of runs per innings, is 28, and the population standard deviation is 12. In a new season he adopts a different batting style and he finds that in 30 innings using this style his mean score is 28.98.

(i) Stating a necessary assumption, test at the 5% significance level whether his population mean score has increased. **[8]**

(ii) Explain whether it was necessary to use the Central Limit Theorem in part **(i)**. **[2]**

7 The continuous random variable X has the distribution $\mathrm{N}(\mu, \sigma^2)$. The mean of a random sample of n observations of X is denoted by $\overline{X}$. It is given that $\mathrm{P}(\overline{X} < 35.0) = 0.9772$ and $\mathrm{P}(\overline{X} < 20.0) = 0.1587$.

(i) Obtain a formula for σ in terms of n. **[5]**

Two students are discussing this question. Aidan says "If you were told another probability, for instance $\mathrm{P}(\overline{X} > 32) = 0.1$, you could work out the value of σ." Binya says, "No, the value of $\mathrm{P}(\overline{X} > 32)$ is fixed by the information you know already."

(ii) State which of Aidan and Binya is right. If you think that Aidan is right, calculate the value of σ given that $\mathrm{P}(\overline{X} > 32) = 0.1$. If you think that Binya is right, calculate the value of $\mathrm{P}(\overline{X} > 32)$. **[4]**

8 In a large city the number of traffic lights that fail in one day of 24 hours is denoted by Y. It may be assumed that failures occur randomly.

(i) Explain what the statement "failures occur randomly" means. **[1]**

(ii) State, in context, two different conditions that must be satisfied if Y is to be modelled by a Poisson distribution, and for each condition explain whether you think it is likely to be met in this context. **[4]**

(iii) For this part you may assume that Y is well modelled by the distribution $\mathrm{Po}(\lambda)$. It is given that $\mathrm{P}(Y = 7) = \mathrm{P}(Y = 8)$. Use an algebraic method to calculate the value of λ and hence calculate the corresponding value of $\mathrm{P}(Y = 7)$. **[5]**

9 The random variable A has the distribution $\mathrm{B}(30, p)$. A test is carried out of the hypotheses H_0: $p = 0.6$ against H_1: $p < 0.6$. The critical region is $A \leqslant 13$.

(i) State the probability that H_0 is rejected when $p = 0.6$. **[1]**

(ii) Find the probability that a Type II error occurs when $p = 0.5$. **[2]**

(iii) It is known that on average $p = 0.5$ on one day in five, and on other days the value of p is 0.6. On each day two tests are carried out. If the result of the first test is that H_0 is rejected, the value of p is adjusted if necessary, to ensure that $p = 0.6$ for the rest of the day. Otherwise the value of p remains the same as for the first test. Calculate the probability that the result of the second test is to reject H_0. **[5]**

Question			Answer	Marks	Guidance	
1			$n = 9$ CR is ≤ 2 **0.0083**	B1 M1A1 A1 **[4]**	Stated explicitly 2 seen but not ≤: M1A0. Allow "P(≤ 2)" Or more SF. "n = 9, CR ≥ 3", 0.0083 seen: B1M1A0A1	CR must be stated explicitly for A1 SR: ≤ 3 with 0.0424: (B1)M1A0 SR: If 0, give B1 for at least 3 of 0.0083, 0.0113, 0.0026, 0.0197, 0.0034 seen
2	(i)		$\hat{\mu} = \bar{x} = 38$ $\frac{\Sigma x^2}{10} - 38^2$ [= 16.2] ×10/9 to get **18**	B1 M1 M1 A1 **[4]**	38 stated separately Use of $\Sigma x^2/n - \bar{x}^2$ Multiply by 10/9 18 or a.r.t. 18.0 only	Correct single formula: M2 If single formula, divisor of 9 seen anywhere gets second M1
2	(ii)		$\Phi\left(\frac{40-38}{\sqrt{18}}\right) = \Phi(0.4714) = \mathbf{0.3187}$	M1 A1 **[2]**	Standardise with their μ and σ, allow cc, √ errors Answer, a.r.t. 0.319	√10 used: M0. Allow a.r.t. 0.311 [0.3106] from 16.2
3	(i)		Allocate 4-digit number to each DVD; Select using random numbers Ignore random numbers outside range	B1 B1 B1 **[3]**	"DVD" & "4 digits/1 to 9000/sequentially" etc must be mentioned *somewhere* Mention random numbers Unbiased method, mention of "outside range" or "repeats"	*Not* allocate "random" numbers, unless subsequently sorted If "pick random numbers in range 1 to 9000", must mention repeats
3	(ii)		B(100, 0.24) ≈ N(24, 18.24) $\Phi\left(\frac{19.5-24}{\sqrt{18.24}}\right) = \Phi(-1.0537)$ = **0.1461**	M1 A1 M1 A1 A1 **[5]**	N(attempt at np) Both parameters correct Standardise with their np and $\sqrt{npq}$ or npq Both cc correct and $\sqrt{npq}$ used Answer, a.r.t. 0.146	Allow 18.24/100 A1 but then M0A0 Allow cc/√ errors.

Question			Answer	Marks	Guidance	
4	(i)		Values taken by X	B1 **[1]**	This answer only	Not "values taken by f"
4	(ii)		$\int_0^a kx\,dx = 1 \Rightarrow k = \frac{2}{a^2}$	M1 A1 **[2]**	Use definite integral and equate to 1, Correctly obtain $2/a^2$	*Or* clear argument from triangle area
4	(iii)		$\int_0^a kx^2\,dx = \left[k\frac{x^3}{3}\right]_0^a = \frac{2}{3}a$ $\int_0^a kx^3\,dx = \left[k\frac{x^4}{4}\right]_0^a = \frac{a^2}{2}$ $\frac{a^2}{2} - \left(\frac{2}{3}a\right)^2 = \frac{1}{18}a^2$	M1 B1 A1√ M1* depM1 A1 **[6]**	Attempt to integrate xf(x), limits 0 and a Correct indefinite integral seen Correct mean *or* correct E(X^2) [= $a^2/2$], √ on k Attempt to integrate x^2f(x), limits 0, a Subtract their μ^2 Correct final answer, ae exact f, no k now	 either here or for x^2f(x) Can be in terms of k Or decimal, $0.056a^2$ or better
5	(i)		Po(4200) ≈ N(4200, 4200) $1 - \Phi\left(\frac{4350.5 - 4200}{\sqrt{4200}}\right)$ $= 1 - \Phi(2.322)$ = **0.010**(1)	M1 M1 M1 A1 A1 **[5]**	Po(60λ) stated or implied N(60λ, 60λ) Standardise with their 60λ and $\sqrt{60\lambda}$ or 60λ 4350.5 explicitly seen and $\sqrt{60\lambda}$ not wrong Answer, allow a.r.t. 0.010	 Allow wrong or no cc, or no √ $\sqrt{60\lambda}$ needn't be explicit Allow [0.0103, 0.0106] from no CC, but *not* 0.0105 from wrong CC
5	(ii)		B(30, 0.010(1)) ≈ Po(0.30(3)) 1 – 0.9997 = **0.0003** *or:* $1 - (q^{30} + 30q^{29}p + 435q^{28}p^2 + 4060q^{27}q^3)$ = 1 – (.7397 + .2242 + .0328 + .0031) = 1 – .999777 = 0.0002226	M1 A1 A1 **[3]**	B(30, their (*i*)) stated or implied Po(0.3) or 0.303 etc Final answer a.r.t. 0.0003 Exact binomial: 1 – (3,4 or 5 terms) (M1)M1 Answer a.r.t. 0.0002: A1 Normal (0.3, 0.297) (M1)M1 Answer 0 (4 dp) (z = 5.87) A1	 [0.30→0.000266. 0.303→0.000276. 0.309→0.000297] Needs clear ^nC_r or right answer No mention of dist: assume exact

Question			Answer	Marks	Guidance	
6	(i)		H_0: $\mu = 28.0$ H_1: $\mu > 28.0$ α: $\frac{28.98-28}{12/\sqrt{30}} = 0.4473$ [$p = 0.3274$] $z < 1.645$, or $p > 0.05$ OR: CC: $28.98 - \frac{1}{60} \to 0.4397$, $p = 0.33$	B2 M1 A1 A1	One error, e.g. p, or μ_0, μ_1, or 2-tail: B1. Standardise with √30, allow √ errors, cc Correct value of z or p: z = art 0.447 or p in range [0.327, 0.328] Compare z (incl 30) with 1.645, or p with 0.05, or with 0.95 if correct tail	But $\bar{x}$ etc: B0 CC is CORRECT here *Not* –0.447 but can be recovered if 0.327 used. Not 0.455/0.3246 Needs μ and $\bar{x}$ right way round
			β: $28 + 1.645 \times 12/\sqrt{30}$ $= 31.6$ $28.98 < 31.6$	M1 A1√ A1	$28 + z\times12/\sqrt{30}$, allow √ errors, cc Correct CV, √ on z (only) Explicitly compare 28.98	Ignore 28 – …, do not allow 28.98 – ...
			γ: Totals used: $\frac{869.4-840}{12\sqrt{30}} = 0.4473$		Same scheme	NB: If totals used, allow ANY plausible CC or none
			Do not reject H_0. Insufficient evidence of an increase in mean score SD unchanged, *or* random sample/indept	M1 A1 B1 **[8]**	Consistent first conclusion Contextualised, "evidence" or exact equivalent somewhere One of these seen, nothing irrelevant	Needs correct method & comparison, 30 used, μ and $\bar{x}$ right way round "Evidence" in either part of conclusion
6	(ii)		Yes because population not stated to be normal	B2 **[2]**	Partial answer: B1 "Yes as parent distribution not normal" (i.e., "stated to be" omitted): B2 SR: "No as assumed normal" if in (i): B1	"Yes, because n large": B1 "Yes, as not normal and n large": B1 "Yes as not normal, but can be used as n large": B2
7	(i)		$\frac{\mu-20}{\sigma/\sqrt{n}} = 1.0$; $\frac{35-\mu}{\sigma/\sqrt{n}} = 2.0$ Solve to get $\sigma = 5\sqrt{n}$	M1 A1 B1 M1 A1 **[5]**	Standardise either 20 or 35, equate to Φ^{-1} Both equations completely correct Both correct z-values seen (to 3 SF at least) Correctly obtain $\sigma = k\sqrt{n}$ or $\sigma^2 = kn$ $\sigma = 5\sqrt{n}$ or $\sqrt{25n}$ only.	With √n or n and z, allow "1 –", cc Including signs, but can have wrong z Independent of previous marks Allow √ errors, ALLOW from not Φ^{-1} [only mark from 0.7998 & 0.8358]
7	(ii)		Binya is right $\mu = 25$ $1 - \Phi\left(\frac{32-\mu}{5}\right) = 1 - \Phi(1.4)$ $= 1 - 0.9192$ $=$ **0.0808**	B1 B1 M1 A1 **[4]**	Binya stated $\mu = 25$ following no wrong working Standardise with their $\sigma/\sqrt{n}$ and their numerical μ Answer, a.r.t. 0.081, CWO.	"Aidan" used: max B0B1M0 But allow if √n omitted or wrong NB: use of 1.282 probably implies "Aidan"

Question			Answer	Marks	Guidance	
8	(i)		Failures do not occur at regular or predictable intervals	B1 [1]	*Not* equivalent of "independent". *Not* "equally likely at any moment"	Both right and wrong: B0
8	(ii)		Failures occur independently; Might not happen if a power cut … and at constant average rate; Might not happen if manipulated to change more rapidly at peak times	B1 B1 B1 B1 [4]	"Failures" needed in one reason, else B0(B3) Plausible reason Exact equivalents only Must be during one day and not week/year Allow any answers that show correct statistical understanding, however implausible	*Not* "randomly", allow "singly" only if also "independent" in this part Not "equal probability", *not* "constant rate", but allow second mark if OK. Extra wrong reason loses explanation mark
8	(iii)		$e^{-\lambda}\frac{\lambda^7}{7!} = e^{-\lambda}\frac{\lambda^8}{8!} \Rightarrow \lambda = 8$ **0.1396**	M1 A1 M1 A1 B1√ [5]	At least one correct formula Both sides correct Cancel exp and some λ Obtain $\lambda = 8$ only, CWO Answer in range [0.139, 0.14], √ on their λ	[before rounding]
9	(i)		4.81% or 0.0481	B1 [1]	One of these only, or more SF	N(18, 7.2) → 0.0468: B1
9	(ii)		P(≥ 14) = 0.7077	M1 A1 [2]	Allow M1 for answer 0.5722 or 0.8192 0.708 or 0.7077 or more SF	0.2923: 0 N(15, 7.5) → 0.78: M1A1; 0.8194 or 0.7674: M1A0
9	(iii)		Only way that p = 0.5 for second test is if Type II error on first, where 0.2 × 0.7077 = 0.14154. Therefore 0.14154 × 0.2923 + 0.85846 × 0.0481 = **0.0827**	M1 M1 M2 A1 [5]	0.2 × 0.7077 × 0.2923 [= 0.04137] Consider 1 – 0.14154 0.2 × (ii) × (1 – (ii)) + (1 – [0.2 × (ii)]) × (i) [= 0.04137 + 0.04127] Answer, a.r.t. 0.083 OR: 0.8 × 0.0481 × 0.0481 [0.00185] + 0.8 × 0.9519 × 0.0481 [0.03663] M1 + 0.2 × 0.2923 × 0.0481 [0.00281] M1 + 0.2 × 0.7077 × 0.2923 [0.04137] M1 Add up 4 terms of 3 multiplications M1 Answer 0.0827 A1	Normal: 0.1416×0.292 + 0.8584×0.0468 or 0.00175+0.03569+0.00273+0.04135 = 0.0815: full marks Any two of these three M1 Third of these three M1 This one M1 SR: No 0.8 or 0.2 but 2 products: M1 4 products: M2

APPENDIX 1

Generic mark scheme issues for S2:

1 Standardisation using the normal distribution.

(a) When *stating* parameters of normal distributions, don't worry about the difference between σ and σ^2, so allow N(9, 16) or N(9, 4^2) or N(9, 4). When *calculating* $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, the following mistakes are accuracy mistakes and not method mistakes so can generally score M1A0:

- confusion of σ with σ^2 or $\sqrt{\sigma}$
- n versus $\sqrt{n}$
- wrong or no continuity corrections.

(b) Some candidates are taught to calculate, for example, P(X > 5) from N(9, 16) by calculating instead P(X < 13). This is a correct method, though it looks very strange the first time you see it.

(c) In hypothesis tests, use of $\frac{\mu-\bar{x}}{\sigma}$ instead of $\frac{\bar{x}-\mu}{\sigma}$ is not penalised if it leads to a correct probability, but if the candidate is using a z-value in a hypothesis test, an answer of $z = -2.15$ when it ought to be 2.15 is an accuracy error and loses the relevant A1. When finding μ or σ from probabilities, some candidates are taught to use $\frac{\mu-\bar{x}}{\sigma}$ whenever $\mu > \bar{x}$; provided the signs are consistent this gains full marks.

(d) When calculating normal approximations to binomial or Poisson distributions, use of the wrong, or no, continuity correction generally loses the last two marks: A0 A0.

2 **Conclusions to hypothesis tests.** There are generally 2 marks for these.

(a) In order to gain M1, candidates must not only say the correct "Reject/do not reject H_0" but have done the whole test in essence correctly apart from numerical errors. In other words:

- they must have compared their *p* value with a critical *p* value or other "like-with-like" (e.g. *not* say 0.0234 with 1.96)
- using the correct tail (e.g. *not* –2.61 with +2.576), and
- the working should in general have accuracy errors only.

Thus miscalculation of *z*, comparison with 1.645 instead of 1.96, or using *n* instead of $\sqrt{n}$, or omission of a continuity correction when it is necessary, are all accuracy errors and the candidate can still gain the last M1 A1. Omission of $\sqrt{n}$ where it is necessary is a method mistake and so gets M0. In hypothesis tests using discrete distributions, use of P(≤ 12) or P(> 12) or P(= 12) when it should be P(≥ 12) is a method mistake and usually loses all the final marks in a question.

(b) The A1 mark is for interpreting the answer *in the context of the question*, and *without over-assertiveness*. Thus "The mean number of applicants has increased" is over-assertive and gets A0 (although we allow "There is sufficient evidence to reject H_0. The mean number of applicants has increased", A1), and "There is sufficient evidence that the mean has increased" is not contextualised, so that too is A0.

(c) A wrong statement such as –2.61 > –2.576 generally gets B0 for comparison but can get the subsequent M1A1. Otherwise:

(d) If there is a self-contradiction, award M1 only if "Reject/Accept H_0" is consistent with their comparison. Thus if, say, we had $z = 2.61 > z_{crit} = 2.576$:
"Reject H_0, there is insufficient evidence that the mean number of … has changed" is M1A0.
but "Do not reject H_0, there is evidence that the mean number of … has changed" is M0A0.
If they omit any mention of H_0/H_1 and just say "there is evidence that the mean number of …" etc, that's A2 or A0.

(e) We don't usually worry about differences between "Reject H_0" and "Accept H_1" etc.

APPENDIX 2

Question 6(i) specific examples – marks out of 7 (rather than 8: condition not included)

α	H_0: $\bar{x}$ = 28.0; H_1: $\bar{x}$ > 28.0 *[wrong symbol]* $z = \frac{28.98-28.0}{\sqrt{12/30}}$ = 1.550 *[wrong √]* < 1.645 Accept H_0, no increase in average score *[over-assertive, otherwise A1]*	B0B0 M1 A0 A1 M1A0 **3**	δ	H_0 = 28.0; H_1 > 28.0 *[missing symbol]* $z = \frac{28.0-28.98}{12/\sqrt{30}}$ = –0.447 *[loses 1]* > –1.645 Insufficient evidence to reject H_0. No change in average score. *[OK]*	B1 only M1 A0 A1 M1 A1 **5**
β	H_0: μ = 28.98; H_1: μ < 28.98 *[WRONG]* $z = \frac{28.98-28.0}{12/\sqrt{30}}$ = 0.447 *[allow this – BOD]* < 1.645 Accept H_0. Insufficient evidence of a change in maximum daily temperature.	B0B0 M1 A1 A1 M1 A1 **5**	ε	H_0: μ = 28.0; H_1: $\mu \neq$ 28.0 *[two-tail]* $z = \frac{28.98-28.0}{12/\sqrt{30}}$ = 0.447 < 1.96 [also if < 1.645] Accept H_0. Insufficient evidence of a change in average score.	B1B0 M1 A1 A0 M1 A1 **5**
γ	*CONTRAST:* H_0: μ = 28.98; H_1: μ < 28.98 *[WRONG]* $z = \frac{28.0-28.98}{12/\sqrt{30}}$ = –0.447 *[DON'T allow this]* > –1.645 Accept H_0. Insufficient evidence of a change in average score.	B0B0 M1 A0 A1 M0 A0 **2**	ζ	H_0: μ = 28.0; H_1: μ > 28.0 $z = \frac{28.0-28.98}{12/\sqrt{30}}$ = –0.447 *but then…* So p = 0.327 > 0.05 *[OK here]* Accept H_0. Insufficient evidence of a change in average score.	B2 M1 A1 A1 M1 A1 **7**
			η	H_0: μ = 28.0; H_1: μ > 28.0 $z = \frac{28.98-28.0}{12}$ = 0.0817 *[no √30]* < 1.645 Accept H_0. Insufficient evidence of a change in average score.	B2 M0 A0 A0 M0 A0 **2**

Question 8, specimen answers:

(i)	There is no pattern to the failures and they occur independently of one another	B0	
	Equally likely to occur at any moment in time	B0	
	Impossible to predict	B1	
(ii)	Failures occur singly, unlikely as there could be a power failure that affects all lights in an area:	B0B1	
	Failures occur independently of each other:	(B1)	
	Likely because one failure does not cause another	B1	
	Mean number of traffic light failures is constant each <u>day</u>	B0	(OK if each <u>hour</u> etc)
	Failures occur at constant average rate:	(B1)	
	Unlikely as could change with season	B0	
	Likely as each set has same probability of failing	B0	
	Likely as they run in the same mode all day	B1	