OCR
RECOGNISING ACHIEVEMENT

# Friday 21 June 2013 – Morning

## 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

### 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** It is required to select a random sample of 30 pupils from a school with 853 pupils. A student suggests the following method.
"Give each pupil sequentially a three-digit number from 001 to 853. Use a calculator to generate random three-digit numbers from 0.000 to 0.999 inclusive, multiply the answer by 853, add 1 and round off to the nearest whole number. Select the corresponding pupil, and repeat as necessary".

**(i)** Determine which pupil would be picked for each of the following calculator outputs:

$$0.103, \quad 0.104, \quad 0.105, \quad 0.106, \quad 0.107.$$ **[2]**

**(ii)** Use your answers to part **(i)** to show that this method is biased, and suggest an improvement. **[2]**

**2** The number of neutrinos that pass through a certain region in one second is a random variable with the distribution $\mathrm{Po}(5 \times 10^4)$. Use a suitable approximation to calculate the probability that the number of neutrinos passing through the region in 40 seconds is less than $1.999 \times 10^6$. **[4]**

**3** The mean of a sample of 80 independent observations of a continuous random variable $Y$ is denoted by $\overline{Y}$. It is given that $\mathrm{P}(\overline{Y} \leqslant 157.18) = 0.1$ and $\mathrm{P}(\overline{Y} \geqslant 164.76) = 0.7$.

**(i)** Calculate $\mathrm{E}(Y)$ and the standard deviation of $Y$. **[6]**

**(ii)** State

**(a)** where in your calculations you have used the Central Limit Theorem,

**(b)** why it was necessary to use the Central Limit Theorem,

**(c)** why it was possible to use the Central Limit Theorem. **[3]**

**4** The number of floods in a certain river plain is known to have a Poisson distribution. It is known that up until 10 years ago the mean number of floods per year was 0.32. During the last 10 years there were 6 floods. Test at the 1% significance level whether there is evidence of an increase in the mean number of floods per year. **[7]**

**5** Two random variables $S$ and $T$ have probability density functions given by

$$\mathrm{f}_S(x) = \begin{cases} \frac{3}{a^3}(x-a)^2 & 0 \leqslant x \leqslant a, \\ 0 & \text{otherwise}, \end{cases}$$

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

where $a$ and $c$ are constants.

**(i)** On a single diagram sketch both probability density functions. **[3]**

**(ii)** Calculate the mean of $S$, in terms of $a$. **[5]**

**(iii)** Use your diagram to explain which of $S$ or $T$ has the bigger variance. (Answers obtained by calculation will score no marks.) **[2]**

**6** The random variable $X$ denotes the yield, in kilograms per acre, of a certain crop. Under the standard treatment it is known that $E(X) = 38.4$. Under a new treatment, the yields of 50 randomly chosen regions can be summarised as

$$n = 50, \qquad \Sigma x = 1834.0, \qquad \Sigma x^2 = 70027.37.$$

Test at the 1% level whether there has been a change in the mean crop yield. **[11]**

**7** Past experience shows that 35% of the senior pupils in a large school know the regulations about bringing cars to school. The head teacher addresses this subject in an assembly, and afterwards a random sample of 120 senior pupils is selected. In this sample it is found that 50 of these pupils know the regulations. Use a suitable approximation to test, at the 10% significance level, whether there is evidence that the proportion of senior pupils who know the regulations has increased. Justify your approximation. **[11]**

**8** The random variable $R$ has the distribution $B(14, p)$. A test is carried out at the $\alpha$% significance level of the null hypothesis $H_0: p = 0.25$, against $H_1: p > 0.25$.

**(i)** Given that $\alpha$ is as close to 5 as possible, find the probability of a Type II error when the true value of $p$ is 0.4. **[4]**

**(ii)** State what happens to the probability of a Type II error as

**(a)** $p$ increases from 0.4,

**(b)** $\alpha$ increases, giving a reason. **[2]**

**9** The managers of a car breakdown recovery service are discussing whether the number of breakdowns per day can be modelled by a Poisson distribution. They agree that breakdowns occur randomly. Manager $A$ says, "it must be assumed that breakdowns occur at a constant rate throughout the day".

**(i)** Give an improved version of Manager $A$'s statement, and explain why the improvement is necessary. **[2]**

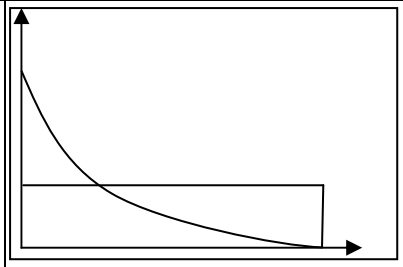
**(ii)** Explain whether you think your improved statement is likely to hold in this context. **[1]**

Assume now that the number $B$ of breakdowns per day can be modelled by the distribution $\text{Po}(\lambda)$.

**(iii)** Given that $\lambda = 9.0$ and $P(B > B_0) < 0.1$, use tables to find the smallest possible value of $B_0$, and state the corresponding value of $P(B > B_0)$. **[2]**

**(iv)** Given that $P(B = 2) = 0.0072$, show that $\lambda$ satisfies an equation of the form $\lambda = 0.12e^{k\lambda}$, for a value of $k$ to be stated. Evaluate the expression $0.12e^{k\lambda}$ for $\lambda = 8.5$ and $\lambda = 8.6$, giving your answers correct to 4 decimal places. What can be deduced about a possible value of $\lambda$? **[5]**

| Question | | | Answer | Marks | Guidance | |
|---|---|---|---|---|---|---|
| 1 | (i) | | 89, 90, 91, 91, 92 | B2<br>**2** | All correct; B2; one error (e.g. all –1), B1 | Allow 088, etc |
| | (ii) | | Not all equally likely (91 more than 90 etc)<br>Multiply by 1000 and ignore if > 853 | B1<br>B1<br>**2** | Imply different likelihood/probability<br>Or equivalent method. Not "ignore repeats". Ignore extras. | *Not* "same pupil *is* selected twice"<br>Number students, use random numbers *and ignore outside range*: B1 |
| 2 | | | $\text{Po}(2 \times 10^6)$<br>$\approx \text{N}(2 \times 10^6, 2 \times 10^6)$<br>$\Phi\left(\frac{1998999.5 - 2\times10^6}{\sqrt{2\times10^6}}\right) = \Phi(-0.70746)$<br>**= 0.2396** | M1<br>A1<br>A1<br>A1<br>**4** | N(their $40\lambda$)<br>Both parameters correct, allow √ here<br>Standardise, mean $40\lambda$, sd $\sqrt{40\lambda}$ (*not* $40\lambda$)<br>Answer, a.r.t. 0.240<br>(no cc: M1A1A0A1) | *Correct cc must be seen for this A1*<br>NB: no cc gives Φ(–0.7071), 0.23975, wrong cc gives Φ(–0.70675), 0.23986 |
| 3 | (i) | | $\frac{\mu - 157.18}{\sigma/\sqrt{80}} = 1.282$ ; $\frac{\mu - 164.76}{\sigma/\sqrt{80}} = 0.5244$<br>Solve simultaneously: $\mu$ **= 170**<br>$\sigma$ **= 89.44** | M1<br>A1<br>B1<br>B1<br>A1<br>A1<br>**6** | Standardise once with √80 or 80 and $z$, signs may be wrong, allow "1–" errors<br>Both correct ***including signs***, no cc<br>1.28(155) seen anywhere, correct to 3 SF<br>[0.524, 0.525] seen anywhere<br>$\mu$, a.r.t. 170 to 3 SF (169.98)<br>$\sigma$, in range [89, 90], *not* isw<br>*Don't* allow surds, e.g. $40\sqrt{5}$ | Allow cc, but *not* 0.1, 0.7, 0.9, 0.3 or Φ(these) [= .5398, .758, .8159, .6179]<br>$z$ may be wrong (provided it *is* $z$)<br>Ignore signs<br>Ignore signs<br>CWO×2 but allow from inaccurate $z$ if answer(s) within limits. Look out for –89.44: A0A0 |
| | (ii) | (a) | In using normal tables | B1 | Or equiv, e.g. "standardising", "dist of $\overline{Y}$" | Any reference to $\sigma/\sqrt{80}$: B0 |
| | | (b) | Parent distribution not known | B1 | Allow "it is not normal", etc | No extras |
| | | (c) | $n$ large, nothing wrong seen<br>[must be in correct order, no repeats] | B1<br>**3** | If numerical, must be of the form "$n > n_0$" or "$n \geq n_0$" with $30 \leq n_0 \leq 60$ | *Not* "≥ 80". |

| Question | | | Answer | Marks | Guidance | |
|---|---|---|---|---|---|---|
| 4 | | | $H_0: \lambda = 3.2$ (or 0.32) [Allow $\mu$]<br>$H_1: \lambda > 3.2$ (or 0.32) [Allow $\mu$]<br>$R \sim$ Po(3.2)<br>$\alpha$: P($R \geq 6$) = 0.1054<br>> 0.01<br>$\beta$: CR ≥ 9<br>and 6 < 9, with probability 0.0057<br>Do not reject $H_0$. Insufficient evidence of an increase in the number of floods. | B2<br>M1<br>A1<br>A1<br>A1<br>A1<br>M1<br>A1 ft<br>**7** | Both correct, B2. One error, e.g. wrong/ no/different symbols, or two-tail, B1<br>Stated or implied, e.g. N(3.2, 3.2)<br>[0.105, 0.106] before rounding<br>Explicit comparison with 0.01<br>CR ≥ 9 stated; allow CV = 9 if comparison ft<br>0.0057 or 0.9943 seen, and 6 compared<br>Consistent first conclusion<br>Conclusion, mentions "floods", "evidence"<br>*Not* "evidence of no increase"<br>P($R \leq 6$) = 0.9554; P($R > 6$) = 0.0446; P($R = 6$) = 0.0608: max B2 M1<br>P($R < 6$) = 0.8946 *and* compare 0.99 etc: can get full marks. Else A0A0M0A0 | But $x$, $\bar{x}$, $r$, $t$ etc: B0. E($X$), words: B1<br>E.g. $H_0$: $\lambda_0 = 3.2$, $H_1$: $\lambda_1 > 3.2$: B1<br>P(= 6) or (≤ 6) or > 6 or normal:<br>no more marks, maximum B2M1.<br><br><br>needs correct method and like-with-like comparison, but 0.01 needn't be explicit |
| 5 | (i) | |  | M1<br>A1<br>B1<br>**3** | Upwards parabola, not below $x$-axis<br>Correct place, not extending beyond limits, ignore pointed at $a$<br>Horizontal straight line, not beyond limits, $y$-intercept below curve (unless curve makes this meaningless) | *[scales/annotations not needed]*<br>Touching axes (not asymptotic)<br>Don't need vertical lines<br>i.e., 3/3 only if wholly right |
| | (ii) | | $\int_0^a \frac{3}{a^3} x(x-a)^2 \,\mathrm{d}x$<br>$= \int_0^a \frac{3}{a^3}(x^3 - 2ax^2 + a^2x)\,\mathrm{d}x$<br>$= \left[\frac{3}{a^3}\left(\frac{x^4}{4} - \frac{2ax^3}{3} + \frac{a^2x^2}{2}\right)\right]_0^a$<br>$= \frac{a}{4}$ | M1<br>M1<br>A1<br>B1<br>A1<br>**5** | Attempt this integral, correct limits seen somewhere<br>Method for $\int xf(x)$, e.g. multiply out or parts, independent of first M1<br>Correct form for integration, e.g. multiplied out correctly, or correct first stage of parts<br>Correct indefinite integral<br>$\frac{a}{4}$ or exact equivalent (e.g. 0.25$a$) only | Multiplication: needs 3 terms<br>E.g. $\frac{3}{a^3}x\frac{(x-a)^3}{3} - \int \frac{3}{a^3}\frac{(x-a)^3}{3}\mathrm{d}x$<br>E.g. $\frac{3}{a^3}x\frac{(x-a)^3}{3} - \frac{3}{a^3}\frac{(x-a)^4}{12}$<br>Limits not seen anywhere: can get M0M1A0B1A0 |

| Question | | | Answer | Marks | Guidance | |
|---|---|---|---|---|---|---|
| 5 | (iii) | | $S$ is concentrated more towards 0<br>Therefore $T$ has bigger variance | M1<br>A1<br>**2** | Reason that shows understanding of PDF<br>Correct conclusion | *Not*, e.g., "$T$ is constant" |
| 6 | | | $H_0 : \mu = 38.4$ [Allow E($X$) both times]<br>$H_1 : \mu \neq 38.4$ | B2 | Both correct: B2. One error e.g. no or different symbols, one-tail etc, B1 | But $\bar{x}$, $x$, $t$ etc B0.<br>E.g. $H_0$: $\mu_0 = 38.4$, $H_1$: $\mu_1 \neq 38.4$: B1 |
| | | | $\hat{\mu} = \bar{x} = 36.68$ | B1 | 36.68 seen anywhere | $H_0$: $\mu = 36.68$, $H_1$: $\mu \neq 36.68$: B0B0B1<br>*See below and exemplars* |
| | | | $\hat{\sigma}^2 = \frac{50}{49}\left(\frac{70027.37}{50} - 36.68^2\right) = 56.25$ | M1<br>M1<br>A1 | Use biased variance formula [55.125]<br>Multiply by 50/49<br>56.25 | Single formula: M2 or M0. If M0, a divisor of 49 seen anywhere gets M1<br>Allow rounded if clearly correct |
| | | | $\alpha$: $z = \frac{36.68 - 38.4}{\sqrt{56.25/50}} = -1.62$<br>$> -2.576$ [*or* $0.0525 > .005$] | M1<br>A1<br>A1ft | Standardise using $\sqrt{50}$ or 50<br>$z$, a.r.t. –1.62 or $p = 0.0525$<br>Compare $-z$ with $-2.576$ or $+z$ with 2.576 | If 50 missing, no more marks<br>$p$ in range [0.052, 0.053]<br>Ft on $z$. Or $p$ explicitly with 0.005 |
| | | | $\beta$: CV is $38.4 - 2.576\sqrt{\frac{56.25}{50}} = 35.6677$<br>$36.68 > 35.6677$ | M1<br>A1<br>A1ft | CV $38.4 - z\sigma/\sqrt{50}$, ignore 38.4 + anything<br>A.r.t 35.7<br>CV ft and correct comparison | $36.68 + z\sigma/\sqrt{50}$: M1A0A0, M0A0<br><br>Ft on wrong $z$ or on $\sqrt{}$ only |
| | | | Do not reject $H_0$.<br>Insufficient evidence of a change in crop yield | M1<br>A1ft | Correct first conclusion, needs correct method & comparison if seen<br>Contextualised, "evidence" somewhere<br>*Not* "evidence of no change" | Like-with-like, needs $\mu$ and $\bar{x}$ right way round, needs 50<br>Ft on wrong TS and/or CV |
| | | | | | Biased variance [55.125; –1.638 or 0.0508] can get B2B1 M1M0A0 M1A0A1M1A1 (max 8)<br>$\sigma^2$ used [–1.529 or 0.0632, or –0.l2162 or 0.4144]: B2B1 M1M1A1 M1A0A1M1A1 (max 10)<br>No $\sqrt{50}$ [–0.2293 or 0.4092]: B2B1 M1M1A1M0 (max 6)<br>$H_0/H_1$ in terms of 36.68: can get last 4 marks *only* if (36.68 – 38.4) seen, and not (38.4 – 36.68) | |
| | | | | **11** | | |

| Question | | | Answer | Marks | Guidance | |
|---|---|---|---|---|---|---|
| 7 | | | $H_0$: $p = 0.35$ | B2 | One error (e.g. $\mu$, no symbol, 2-tailed) B1, but $\bar{x}$, $t$ etc: B0. Allow π | $H_0$: $\mu = 42$, $H_1$: $\mu > 42$: B1 only |
| | | | $H_1$: $p > 0.35$ | | | |
| | | | B(120, 0.35) | M1 | B(120, 0.35) stated or implied | |
| | | | ≈ N(42, 27.3) | M1 | N(*np*, *npq*), their attempt at 120 × 0.35 | 120 × 0.35 × 0.65 *Not* N(*np*, *nq*). |
| | | | α: $z = \frac{49.5 - 42}{\sqrt{27.3}}$ | A1ft | Standardise, with their *np* and √*npq*, right cc<br>Allow both 49.5 and 50.5 and both in CR | √50 or √120: M1M1A0A0A1M0A0 |
| | | | = 1.435 | A1 | *z* in range [1.43, 1.44] before rounding | Or *p* in range [0.075, 0.0764] |
| | | | > 1.282 [or 0.0757 < 0.1] | A1ft | Comparison with 1.282, ft on *z*/*p* or √120 | Or *p* explicit comparison with 0.1 |
| | | | β: CV = 42.5 + 1.282 × √27.3 [= 49.198] | A1ft | CV 42.5 + *z*×√27.3, ignore LH, ft on *np*, *npq* | No cc: 48.618, can get A0A1A0 |
| | | | *z* = 1.282 and compare 50 | A1 | *z* = 1.282 used in RH CV and compare 50 | |
| | | | CR ≥ 50 or ≥ 49.2 | A1ft | CV correct ft on *z*, but don't worry about ≥ | Must round up. 49 from 49.2: A1A1A0 |
| | | | Reject $H_0$. | M1 | Consistent first conclusion, needs correct method and comparison | Can give M1A1 even if comparison not explicit. Allow from exact binomial |
| | | | Significant evidence that proportion who know regulations has increased | A1ft | Contextualised, needs "who know regulations" or "pupils", and "evidence" | Ft on TS & CV<br>Or exact equivalent somewhere |
| | | | *np* > 5 [= 42] from normal attempted | M1 | From *p* = 0.35 or 5/12, don't need 42 | *or n* large or *p* close to 0.5 asserted |
| | | | *nq* = 78 > 5 and no others apart from *n* large | A1 | Need 78, or 70 from 5/12, *not npq* | *and* the other qualitative reason asserted |
| | | | SC: If B0, B(120, 5/12): | | Wrong or no cc [1.627, 0.0519 or 1.5311, 0.0629]: loses (α) first two A1A1 only | |
| | | | N(50, 29.17) M1M1 | | Exact B(120, 0.35): P(≥ 50) = 0.076824, CR ≥ 50. B2M1, M0A0A0A0, M1A1M0A0 | |
| | | | *np* > 5, *nq* = 70 > 5: M1A1 Max 4 | | *NB: If S3 difference of proportions test used, consult PE* | |
| | | | SC: P(≥ 42): B2 M1M1A0A0A1M0A0 | | | |
| | | | | **11** | | |
| **8** | **(i)** | | B(14, 0.25): Critical region ≥ 7 | M1 | Use B(14, 0.25) and find *r* for an upper tail<br>*All marks need upper tail* | e.g. CV 5 or 6 or 7, or .1117, .0383, .0103, 0.8883, 0.9617, .9897 |
| | | | | A1 | CR ≥ 7 or AR ≤ 6 stated or clearly implied | Not just "CV = 7" |
| | | | B(14, 0.4): P(≤ 6) | M1 | Find P(in AR when *p* = 0.4) *[indept of M1]* | *Not* P(≥ *r*), e.g. final answer 0.3075 |
| | | | **= 0.6925** | A1 | Answer 0.692 or 0.693 or a.r.t. 0.6925 or 0.6924 only, *not* isw [0.692452] | NB: expect CV 8 or 9 and answer 0.9825 or 0.9417: M0M0 |
| | | | | **4** | | |

| Question | | | Answer | Marks | Guidance | |
|---|---|---|---|---|---|---|
| | **(ii)** | | (a) Decreases | B1 | One correct answer & one correct reason *or* two correct answers | Allow from numerical calculation |
| | | | (b) Decreases; increased prob (Type I) ⇔ decreased prob (Type II) | B1<br>**2** | Two correct answers and one correct reason, e.g. "CR becomes larger", etc | Allow equivalent or similar reason<br>Allow from numerical calculation |
| **9** | **(i)** | | Constant *average* rate; *or* [*] same statement *plus* "breakdowns independent" | B1 | State "average" or equiv, "random" or "uniform". | No extras apart from independence (ignore "singly") |
| | | | Otherwise it means that they occur at exactly regular intervals | B1<br>**2** | Correct explanation | Can't get from [*] |
| | **(ii)** | | No because breakdowns more likely in rush hours, etc | B1<br>**1** | Any plausible reason for either "yes" or "no" that shows understanding of what the *statistical* concept means | Not "equally likely". *Not* reason for (in)dependence, unless [*], which needs *both* conditions if affirmed |
| | **(iii)** | | 13 | B1 | | |
| | | | 0.0739 | B1<br>**2** | 0.074 or a.r.t. 0.0739. Marks independent | |
| | **(iv)** | | $e^{-\lambda}\frac{\lambda^2}{2!} = 0.0072$ | M1* | Correct formula = their 0.0072 seen | |
| | | | $\lambda = \sqrt{(0.0144e^{\lambda})}$ | M1dep | Rearrange $e^{-\lambda}$ and square root, to get $\lambda = f(\lambda)$ | Allow even if left with $e^{\lambda}$ or $e^{-\lambda}$ or exact equivalent |
| | | | $= 0.12e^{\lambda/2}$ | A1 | Correctly obtain AG, with $k = 0.5$ | |
| | | | $8.5 \rightarrow 8.4126$; $8.6 \rightarrow 8.8440$ | A1 | Two correct evaluations to 4 dp at least | 4 dp explicitly required |
| | | | Therefore solution between 8.5 and 8.6 | A1<br>**5** | All completely correct and deduction stated | CWO, except allow if only 3 SF |