Thursday 6 June 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767/01
- MEI Examination Formulae and Tables (MF2)

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.
- 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.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

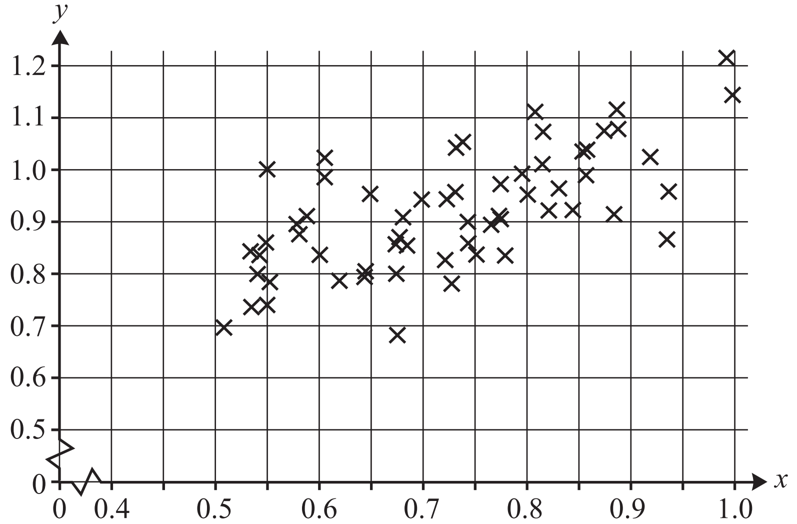
- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks 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.

Turn over

1 Salbutamol is a drug used to improve lung function. In a medical trial, a random sample of 60 people with impaired lung function was selected. The forced expiratory volume in one second (FEV1) was measured for each person, both before being given salbutamol and again after a two-week course of the drug. The variables x and y, measured in suitable units, represent FEV1 before and after the two-week course respectively. The data are illustrated in the scatter diagram below, together with the summary statistics for these data.



Summary statistics:

$$n = 60, \quad \Sigma x = 43.62, \quad \Sigma y = 55.15, \quad \Sigma x^2 = 32.68, \quad \Sigma y^2 = 51.44, \quad \Sigma xy = 40.66$$

(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to investigate whether there is positive correlation between FEV1 before and after the course. **[6]**

(iii) State the distributional assumption which is necessary for this test to be valid. State, with a reason, whether the assumption appears to be valid. **[2]**

(iv) Explain the meaning of the term 'significance level'. **[2]**

(v) Calculate the values of the summary statistics if the data point $x = 0.55$, $y = 1.00$ had been incorrectly recorded as $x = 1.00$, $y = 0.55$. **[3]**

2 Suppose that 3% of the population of a large city have red hair.

(i) A random sample of 10 people from the city is selected. Find the probability that there is at least one person with red hair in this sample. **[2]**

A random sample of 60 people from the city is selected. The random variable X represents the number of people in this sample who have red hair.

(ii) Explain why the distribution of X may be approximated by a Poisson distribution. Write down the mean of this Poisson distribution. **[3]**

(iii) Hence find

(A) $\mathrm{P}(X = 2)$, **[2]**

(B) $\mathrm{P}(X > 2)$. **[2]**

(iv) Discuss whether or not it would be appropriate to model X using a Normal approximating distribution. **[2]**

A random sample of 5000 people from the city is selected.

(v) State the exact distribution of the number of people with red hair in the sample. **[2]**

(vi) Use a suitable Normal approximating distribution to find the probability that there are at least 160 people with red hair in the sample. **[5]**

3 The scores, X, in Paper 1 of an English examination have an underlying Normal distribution with mean 76 and standard deviation 12. The scores are reported as integer marks. So, for example, a score for which $75.5 \leqslant X < 76.5$ is reported as 76 marks.

(i) Find the probability that a candidate's reported mark is 76. **[4]**

(ii) Find the probability that a candidate's reported mark is at least 80. **[3]**

(iii) Three candidates are chosen at random. Find the probability that exactly one of these three candidates' reported marks is at least 80. **[2]**

The proportion of candidates who receive an A* grade (the highest grade) must not exceed 10% but should be as close as possible to 10%.

(iv) Find the lowest reported mark that should be awarded an A* grade. **[5]**

The scores in Paper 2 of the examination have an underlying Normal distribution with mean μ and standard deviation 12.

(v) Given that 20% of candidates receive a reported mark of 50 or less, find the value of μ. **[4]**

4 An art gallery is holding an exhibition. A random sample of 150 visitors to the exhibition is selected. The visitors are asked which of four artists they prefer. Their preferences, classified according to whether the visitor is female or male, are given in the table.

		Artist preferred			
		Monet	Renoir	Degas	Cézanne
Sex	Male	8	25	18	19
	Female	18	35	10	17

(i) Carry out a test at the 10% significance level to examine whether there is any association between artist preferred and sex of visitor. Your working should include a table showing the contributions of each cell to the test statistic. **[12]**

(ii) For each artist, comment briefly on how the preferences of each sex compare with what would be expected if there were no association. **[6]**

Question			Answer	Marks	Guidance	
1	(i)		**EITHER:**			
			$S_{xy} = \sum xy - \frac{1}{n}\sum x \sum y = 40.66 - \frac{1}{60} \times 43.62 \times 55.15$ $= 0.56595$	M1*	For method for S_{xy}	
			$S_{xx} = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2 = 32.68 - \frac{1}{60} \times 43.62^2$ $= 0.96826$	M1*	For method for at least one of S_{xx} or S_{yy}	
			$S_{yy} = \sum y^2 - \frac{1}{n}\left(\sum y\right)^2 = 51.44 - \frac{1}{60} \times 55.15^2$ $= 0.74796$	A1	For at least one of S_{xy}, S_{xx} or S_{yy} (to 2 sf) Note Allow 0.57322 for S_{xy} and 0.76634 for S_{yy} from rounding mean of y to 0.919.	
			$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{0.56595}{\sqrt{0.96826 \times 0.74796}} = 0.665$	M1 dep* A1 **[5]**	For structure of r For answer rounding to 0.66 or 0.67	
			OR:			
			$\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 40.66/60 - (43.62/60 \times 55.15/60)$ $= 0.0094325$	M1*	For method for cov (x,y)	
			$\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(0.96826/60)} = \sqrt{0.016137\ldots} = 0.1270$	M1*	For method for at least one msd or rmsd	
			$\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(0.74796/60)} = \sqrt{0.012466} = 0.1117$	A1	For at least one of cov (x,y), msd or rmsd correct (to 2 sf)	
			$r = \frac{\text{cov}(x,y)}{\text{rmsd}(x)\text{rmsd}(y)} = \frac{0.0094325}{0.1270 \times 0.1117} = 0.665$	M1 dep* A1 **[5]**	For structure of r For answer rounding to0.66 or 0.67 Methods mixed – max M0M1A1M0A0	

Question			Answer	Marks	Guidance	
1	(ii)		H$_0$: $\rho = 0$ H$_1$: $\rho > 0$ (one-tailed test)	B1	For H$_0$, H$_1$ in symbols. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population correlation coefficient.	
			where ρ is the population correlation coefficient	B1	For defining ρ. Condone omission of "population" if correct notation ρ is used, but if ρ is defined as the **sample** correlation coefficient then award **B0**. Allow "ρ is the pmcc".	
			For $n = 60$, 5% critical value = 0.2144	B1	For critical value	
			Since 0.665 > 0.2144, the result is significant.	M1	For sensible comparison leading to a conclusion provided that $\|r\| < 1$. The comparison can be in the form of a diagram as long as it is clear and unambiguous. Sensible comparison: e.g. 0.665 > 0.2144 is 'sensible' whereas 0.665 > - 0.2144 is 'not sensible'. Reversed inequality sign e.g. 0.665 < 0.2144 etc. gets max M1 A0.	
			Thus we have sufficient evidence to reject H$_0$	A1	For reject H$_0$ o.e. FT their r and critical value from 5% 1-tail column.	
			There is sufficient evidence at the 5% level to **suggest** that there is **positive** correlation between FEV1 before and after the two-week course.	E1	For correct, **non-assertive** conclusion in context (allow 'x and y' for context). E0 if H$_0$ and H$_1$ not stated, reversed or mention a value other than zero for ρ in H$_0$.	
				[6]		

Question			**Answer**	**Marks**	**Guidance**	
1	**(iii)**		The underlying population must have a bivariate Normal distribution. Yes, since the scatter diagram appears to have a roughly elliptical shape.	B1 E1 **[2]**	Condone “bivariate Normal distribution”, “underlying bivariate Normal distribution”, but **do not allow** “the **data** have a bivariate Normal distribution” Condone ‘oval’ or suitable diagram	
1	**(iv)**		The significance level is the probability of rejecting the null hypothesis when in fact it is true.	E1* E1dep* **[2]**	For “probability of rejecting H_0” or “probability of a significant result”. For “when H_0 is true”	
1	**(v)**		$\sum x = 43.62 + 0.45 = 44.07$ $\sum y = 55.15 - 0.45 = 54.70$ $\sum xy = 40.66$ $\sum x^2 = 32.68 + 1 - 0.55^2 = 33.3775$ $\sum y^2 = 51.44 - 1 + 0.55^2 = 50.7425$	B1 B1 B1 **[3]**	For $\sum x$ or $\sum y$ or $\sum xy$ For $\sum x^2$ or $\sum y^2$ (to 2 dp) For all correct (ignore n)	
2	**(i)**		P(At least one has red hair) = $1 - 0.97^{10}$ = 0.263	M1 A1 **[2]**	M1 for $1 - 0.97^{10}$ Allow 0.26	
2	**(ii)**		(Because X is binomially distributed), n is large and p is small. Mean = 1.8	E1 E1 B1 **[3]**	Allow “sample is large” for n is large Allow “$np < 10$” or “mean ≈ variance” for “p is small” Do not allow “the probability is small”	

Question			Answer	Marks	Guidance	
2	**(iii)**	(*A*)	$\mathrm{P}(X=2) = e^{-1.8}\dfrac{1.8^2}{2!} = 0.2678$ OR = 0.7306 – 0.4628 = 0.2678	M1 A1 **[2]**	For calculation for $\mathrm{P}(X = 2)$ FT their mean. Allow answer to 3sf.	
2	**(iii)**	(*B*)	$\mathrm{P}(X > 2) = 1 - \mathrm{P}(X \le 2) = 1 - 0.7306$ = 0.2694	M1 A1 **[2]**	$1 - \mathrm{P}(X \le 2)$ used. e.g.$1 - \mathrm{P}(X \le 2) = 1 - 0.4628$ gets M0 CAO	
2	**(iv)**		The mean ($np = 1.8$) is too small It is not appropriate to use a Normal approximation	E1* E1dep* **[2]**	For "mean is too small" or "mean < 10" For "not appropriate". Do not allow "*p* is too small".	
2	**(v)**		Binomial(5000, 0.03)	B1* B1dep* **[2]**	For binomial, or B(,) For parameters	
2	**(vi)**		Mean 5000 × 0.03 = 150 Variance = 5000 × 0.03 × 0.97 = 145.5 Using Normal approx. to the binomial, $X \sim \mathrm{N}(150, 145.5)$ $\mathrm{P}(X \ge 160) = \mathrm{P}\left(Z \ge \dfrac{159.5-150}{\sqrt{145.5}}\right)$ $= \mathrm{P}(Z > 0.7876) = 1 - \Phi(0.7876) = 1 - 0.7846$ = 0.215 (to 3 sig.fig.)	B1 B1 B1 M1 A1 **[5]**	For mean (soi) For variance (soi) For continuity corr. For probability using correct tail and structure (condone omission of/incorrect c.c.) CAO, (Do not FT wrong or omitted CC) Allow 0.2155. Do not allow 0.216	

Question			Answer	Marks	Guidance	
3	(i)		$\text{P}(Y=76) = \text{P}\left(\frac{75.5-76}{12} \le Z \le \frac{76.5-76}{12}\right)$	B1	For one correct continuity correction used	
			$= \text{P}(-0.04166\ldots < Z < 0.04166\ldots)$ $= \Phi(0.04166\ldots) - (1 - \Phi(0.04166\ldots))$	M1	For standardizing	
			$= 2 \times \Phi(0.04166\ldots) - 1$ $= 2 \times 0.5167 - 1$	M1	For correctly structured probability calculation.	
			$= 0.0334$	A1	CAO inc use of diff tables. Allow 0.0330 – 0.0340 www.	
				[4]		
3	(ii)		$\text{P}(Y \ge 80) = \text{P}\left(Z \ge \frac{79.5-76}{12}\right)$	B1	For correct cc used	
			$= \text{P}(Z > 0.2917) = 1 - \Phi(0.2917)$	M1	For correct structure	
			$= 1 - 0.6148 = 0.3852 = 0.385$ to 3 sig fig	A1	CAO do not allow 0.386	
				[3]		
3	(iii)		$3 \times 0.3852 \times 0.6148^2 = 0.4368$	M1	$3 \times$ their $p \times (1 -$ their $p)^2$	
				A1	FT their p. Allow 2sf if working seen.	
				[2]		

Question			Answer	Marks	Guidance	
3	(iv)		**EITHER:** P(Score $\ge k$) = 0.1			
			$\Phi^{-1}(0.9) = 1.282$	B1	For 1.282	
			$\frac{k-76}{12} = 1.282$	M1	Allow $k - 0.5$ used for k. Positive z used.	
			$k = 76 + (1.282 \times 12) = 91.38$ or $k = 76 + 0.5 + (1.282 \times 12) = 91.88$	A1	For 91.38 or 91.88	
			$91.38 > 90.5$ or $91.88 > 91$	M1	Relevant comparison (e.g. diagram)	www
			so lowest reported mark = 92	A1		
			OR Trial and improvement method	M1	M1 for attempt to find P(Mark $\ge$ integer)	
			P(Mark $\ge$ 91) = P(Score $\ge$ 90.5) = 0.1135	A1	A1 for 0.1135	
			P(Mark $\ge$ 92) = P(Score $\ge$ 91.5) = 0.0982	A1	A1 for 0.0982	
			P(Mark $\ge$ 91) > 10% and P(Mark $\ge$ 92) < 10%	M1	M1 for comparisons	www
			so lowest reported mark = 92	A1		
				[5]		
3	(v)		$P(Y \le 50) = 0.2$			
			$P(Z \le \frac{50.5-\mu}{12}) = 0.2$	B1	For 50.5 used	
			$\frac{50.5-\mu}{12} = \Phi^{-1}(0.2) = -0.8416$	B1	For –0.8416. Condone – 0.842 Condone 0.8416 if numerator reversed.	
				M1	For structure.	
			$\mu = 50.5 + (12 \times 0.8416) = 60.6$	A1	CAO	
				[4]		

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th colspan="2">Guidance</th></tr>
<tr><td>4</td><td>(i)</td><td></td><td>H_0: no association between sex and artist preferred
H_1: some association between sex and artist preferred</td><td>B1</td><td>For both hypotheses in context</td><td></td></tr>
<tr><td></td><td></td><td></td><td><table><tr><td>EXPECTED</td><td>Monet</td><td>Renoir</td><td>Degas</td><td>Cézanne</td></tr><tr><td>Male</td><td>12.13</td><td>28</td><td>13.07</td><td>16.8</td></tr><tr><td>Female</td><td>13.87</td><td>32</td><td>14.93</td><td>19.2</td></tr></table></td><td>M1
A2</td><td>For expected values
(to 2 dp where appropriate)
(allow A1 for at least one row or column correct)</td><td></td></tr>
<tr><td></td><td></td><td></td><td><table><tr><td>CONTRIB'N</td><td>Monet</td><td>Renoir</td><td>Degas</td><td>Cézanne</td></tr><tr><td>Male</td><td>1.4081</td><td>0.3214</td><td>1.8626</td><td>0.2881</td></tr><tr><td>Female</td><td>1.2321</td><td>0.2813</td><td>1.6298</td><td>0.2521</td></tr></table></td><td>M1
A2</td><td>For valid attempt at $(O–E)^2/E$
For all correct (to 2 dp) and presented in a table or clear list. (Allow A1 for at least one row or column correct)</td><td>NB:These three marks cannot be implied by a correct final value of X^2</td></tr>
<tr><td></td><td></td><td></td><td>$X^2 = 7.28$
Refer to χ_3^2</td><td>B1
B1</td><td>Allow 7.27
for 3 deg of f</td><td>www</td></tr>
<tr><td></td><td></td><td></td><td>Critical value at 10% level = 6.251</td><td>B1</td><td>CAO for cv No FT from here if wrong or omitted, unless p-value used instead</td><td>B1 for p-value = 0.0636</td></tr>
<tr><td></td><td></td><td></td><td>Result is significant</td><td>B1</td><td>FT their X^2</td><td></td></tr>
<tr><td></td><td></td><td></td><td>There is evidence to suggest that there is some association between sex and artist preferred</td><td>E1</td><td>For correct (FT their X^2), non-assertive conclusion, in context.</td><td></td></tr>
<tr><td></td><td></td><td></td><td>NB if H_0 H_1 reversed, or ‘correlation’ mentioned, do not award first B1 or final E1</td><td>[12]</td><td></td><td></td></tr>
</table>

Question			**Answer**	**Marks**	**Guidance**	
4	**(ii)**		Monet: **More females and fewer males than expected** prefer Monet, as indicated by l**arge contribution(s)** (of 1.4081 and 1.2321).	E1* E1dep*	FT their table of contributions	NB MAX 3/6 for answers not referring to contributions (explicitly or implicitly).
			Renoir: Preferences are much **as expected**, as indicated by s**mall contributions.**	E1		
			Degas: **Fewer females and more males than expected** prefer Degas, as indicated by **large contribution(s)** (of 1.8626 and 1.6298).	E1* depE1*		
			Cézanne: Preferences are much **as expected**, as indicated by s**mall contributions**.	E1		SC1 Renoir and Cézanne have correct comments for both but without referring to contributions
				[6]		