OCR
Oxford Cambridge and RSA

Friday 5 June 2015 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

1 2 0 4 4 2 3 5 7 6

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 inside 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.

1 A random sample of wheat seedlings is planted and their growth is measured. The table shows their average growth, y mm, at half-day intervals.

Time t days	0	0.5	1	1.5	2	2.5	3
Average growth y mm	0	7	21	33	45	56	62

(i) Draw a scatter diagram to illustrate these data. **[3]**

(ii) Calculate the equation of the regression line of y on t. **[5]**

(iii) Calculate the value of the residual for the data point at which $t = 2$. **[3]**

(iv) Use the equation of the regression line to calculate an estimate of the average growth after 5 days for wheat seedlings. Comment on the reliability of this estimate. **[2]**

It is suggested that it would be better to replace the regression line by a line which passes through the origin.

You are given that the equation of such a line is $y = at$, where $a = \dfrac{\sum yt}{\sum t^2}$.

(v) Find the equation of this line and plot the line on your scatter diagram. **[4]**

2 It was stated in 2012 that 3% of £1 coins were fakes. Throughout this question, you should assume that this is still the case.

(i) Find the probability that, in a random selection of 25 £1 coins, there is exactly one fake coin. **[2]**

A random sample of 250 £1 coins is selected.

(ii) Explain why a Poisson distribution is an appropriate approximating distribution for the number of fake coins in the sample. **[2]**

(iii) Use a Poisson distribution to find the probability that, in this sample, there are

(A) exactly 10 fake coins, **[3]**

(B) at least 10 fake coins. **[2]**

(iv) Use a suitable approximating distribution to find the probability that there are at least 50 fake coins in a sample of 2000 coins. **[5]**

It is known that 0.2% of another type of coin are fakes.

(v) A random sample of size n of these coins is taken. Using a Poisson approximating distribution, show that the probability of at most one fake coin in the sample is equal to $e^{-\lambda} + \lambda e^{-\lambda}$, where $\lambda = 0.002n$. **[2]**

(vi) Use the approximation $e^{-\lambda} + \lambda e^{-\lambda} \approx 1 - \dfrac{\lambda^2}{2}$ for small values of λ to estimate the value of n for which the probability in part **(v)** is equal to 0.995. **[3]**

3 The random variable X represents the weight in kg of a randomly selected male dog of a particular breed. X is Normally distributed with mean 30.7 and standard deviation 3.5.

(i) Find

(A) $\mathrm{P}(X < 30)$, **[3]**

(B) $P(25 < X < 35)$. **[3]**

(ii) Five of these dogs are chosen at random. Find the probability that each of them weighs at least 30 kg. **[2]**

(iii) The weights of females of the same breed of dog are Normally distributed with mean 26.8 kg. Given that 5% of female dogs of this breed weigh more than 30 kg, find the standard deviation of their weights. **[4]**

(iv) Sketch the distributions of the weights of male and female dogs of this breed on a single diagram. **[4]**

4 **(a)** As part of an investigation into smoking, a random sample of 120 students was selected. The students were asked whether they were smokers, and also whether either of their parents were smokers. The results are summarised in the table below. Test, at the 5% significance level, whether there is any association between the smoking habits of the students and their parents.

	At least one parent smokes	Neither parent smokes
Student smokes	21	27
Student does not smoke	17	55

[10]

(b) The manufacturer of a particular brand of cigarette claims that the nicotine content of these cigarettes is Normally distributed with mean 0.87 mg. A researcher suspects that the mean nicotine content of this brand is higher than the value claimed by the manufacturer. The nicotine content, x mg, is measured for a random sample of 100 cigarettes. The data are summarised as follows.

$$\Sigma x = 88.20 \qquad \Sigma x^2 = 78.68$$

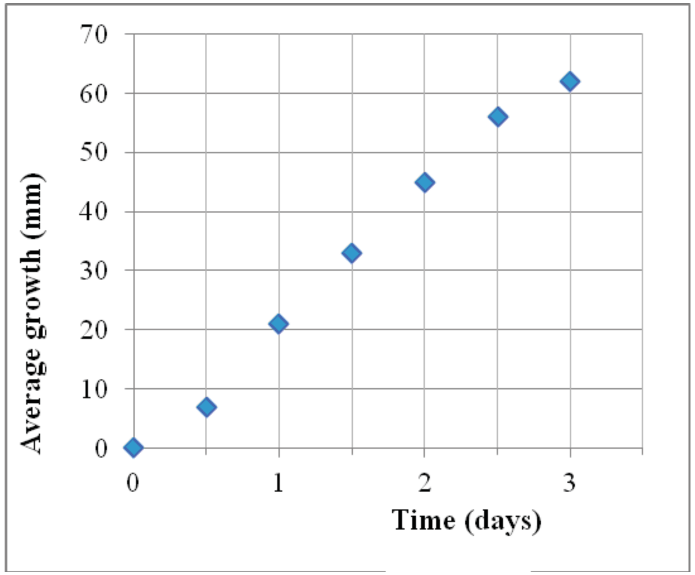
Carry out a test at the 1% significance level to investigate the researcher's belief. **[10]**

END OF QUESTION PAPER

1 (i)

1 (ii)

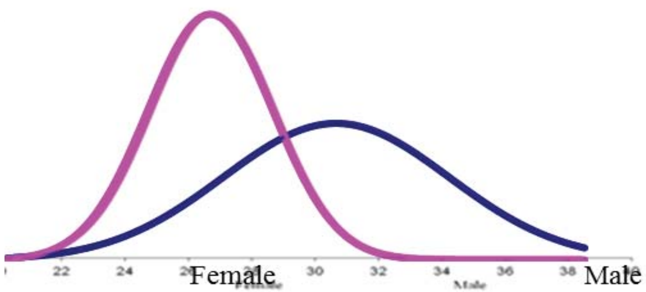
(answer space continued on next page)

Question			Answer	Marks	Guidance	
1	(i)			G1*	Both axes labeled (allow t and y) with indication of scale	Allow axes interchanged Condone x for t
				G1dep*	for values of time BOD if (0,0) not clearly visible	(evenly spaced)
				G1dep*	for values of average growth BOD if (0,0) not clearly visible. BOD if confusion arises from points plotted for part (v).	visually correct SC1 for points having the correct distribution and G0* awarded. Line through origin should appear but this is rewarded in part **(v)**
				[3]		
1	(ii)		$\bar{t} = 1.5, \quad \bar{y} = 32$	B1	For $\bar{t}$ and $\bar{y}$ seen or implied by final answer.	Seen either in calculating b or in forming the equation of the line.
			$b = \frac{S_{yt}}{S_{tt}} = \frac{490-(224\times10.5/7)}{22.75-10.5^2/7} = \frac{154}{7} = 22$ OR $b = \frac{490/7-(32\times1.5)}{22.75/7-1.5^2} = \frac{22}{1} = 22$	M1	For attempt at gradient (b)	Correct structure needed. See additional notes. FT their $\bar{t}$ and $\bar{y}$ for M1
				A1	For 22 cao	
			hence least squares regression line is: $y - \bar{y} = b(t - \bar{t})$	M1	For equation of line	With their $b > 0$, $\bar{t}$ and $\bar{y}$

Question			Answer	Marks	Guidance	
			$\Rightarrow \quad y - 32 = 22(t - 1.5)$ $\Rightarrow \quad y = 22t - 1$	A1 **[5]**	CAO	A0 for $y = 22x - 1$
1	**(iii)**		$t = 2 \Rightarrow$ predicted average growth $= (22 \times 2) - 1 = 43$ Residual $= 45 - 43$ $= 2$	B1 M1 A1 **[3]**	for prediction for subtraction (either way) FT	FT their equation 45 – their prediction
1	**(iv)**		$(22 \times 5) - 1 = 109$ Likely to be **unreliable as extrapolation** (oe)	B1 B1 **[2]**	Estimate calculated using equation	FT their equation
1	**(v)**		$a = \dfrac{490}{22.75} = 21.538\ldots = 21.5$ (3 s.f.) Equation is $y = 21.5t$ Line plotted on diagram	M1 A1 A1 A1 **[4]**	Allow $y = 21.54t$ CAO For line correctly plotted CAO A0 if axes not scaled or $a \neq 21.5$ to 3 sf	Allow $y = (280/13)t$ Through (0,0) and between (3, 64) and (3,65)
2	**(i)**		P(Exactly one) $= \binom{25}{1} \times 0.03^1 \times 0.97^{24}$ $= 0.361$	M1 A1 **[2]**	Binomial calculation with correct structure Allow 0.3611 and 0.36 www A0 for 0.3612	$25 \times p \times (1 - p)^{24}$
2	**(ii)**		n is large p is small.	B1 B1 **[2]**	n large or sample is large p is small, or $np \approx np(1 - p)$ B0 for the “probability” is small unless “probability” is correctly defined.	or $n > 30$ or $np < 10$

Question			Answer	Marks	Guidance	
2	(iii)	(*A*)	Mean = 250 × 0.03 = 7.5 P(exactly 10) = $e^{-7.5}\dfrac{7.5^{10}}{10!}$ Or from tables = 0.8622 – 0.7764 = 0.0858	B1 M1 A1 **[3]**	For mean (SOI) For Poisson probability calculation Allow 0.08583 or 0.086www	 Or using P($X \le 10$) - P($X \le 9$) with Poisson tables
2	(iii)	(*B*)	P(At least 10) = 1 – P($X \le 9$) = 1 – 0.7764 = 0.2236	M1 A1 **[2]**	For **using** 1 - P($X \le 9$) CAO	 Allow 0.224 www
2	(iv)		Mean 2000 × 0.03 = 60 Variance = 2000 × 0.03 × 0.97 = 58.2 Using Normal approx. to the binomial, $X \sim N(60, 58.2)$ $P(X \geq 50) = P\left(Z \geq \dfrac{49.5-60}{\sqrt{58.2}}\right)$ = P(Z > -1.376) = Φ(1.376) = 0.9157 (allow 0.9156 and 0.916)	B1 B1 B1 M1 A1 **[5]**	Normal approximation **used** For parameters (soi) For correct continuity correction For probability using correct structure. CAO **(**Do not FT wrong or omitted CC)	Award full credit for use of Normal approximation to Poisson distribution N(60, 60) N(60, 60) leads to P($Z > -1.356$) = 0.9125 (or 0.913) Allow 0.9124 (or 0.912)
2	(v)		Using a Poisson approximation to the binomial the mean λ = ***np*** (= 0.002n) or λ = **n × 0.002** (= 0.002*n*). P(At most one fake coin) = P(zero or one fake coins) = $e^{-\lambda}\dfrac{\lambda^0}{0!}+e^{-\lambda}\dfrac{\lambda^1}{1!}=e^{-\lambda}+\lambda e^{-\lambda}$ **AG**	B1 B1 **[2]**	For evidence of using *np* from binomial distribution (to give $\lambda = 0.002n$) Evidence of using P(X = 0) + P(X = 1) with $\lambda = 0.002n$ ***NB* ANSWER GIVEN**	Need to **see use of *n*×*p*** or obtaining 0.002*n* from B(*n*, 0.002)

Question			Answer	Marks	Guidance	
2	(vi)		$1-\frac{\lambda^2}{2}=0.995$ $\lambda^2 = 0.01$ so $\lambda = 0.1$ $n = 50$	M1 A1 A1 **[3]**	For equation in λ or **equivalent** equation in n For λ SOI or for $n^2 = 2500$ CAO	
3	(i)	(*A*)	P(X < 30) $= \mathrm{P}\left(Z < \frac{30-30.7}{3.5}\right)$ = P(Z < -0.20) = Φ(-0.20) = 1 – Φ(0.20) = (1 – 0.5793) = 0.4207	M1 M1 A1 **[3]**	For standardising For correct structure CAO	Penalise erroneous continuity corrections and wrong sd. Condone numerator reversed. **1 –** Φ(positive *z*) Allow 0.421 www
3	(i)	(*B*)	P(25 < X < 35) $= \mathrm{P}\left(\frac{25-30.7}{3.5} < Z < \frac{35-30.7}{3.5}\right)$ = P(-1.629 < X < 1.229) = Φ(1.229) – Φ(-1.629) = 0.8904 – (1 – 0.9483) = 0.8904 – 0.0517 = 0.8387	M1 M1 A1 **[3]**	Correctly standardising both. For correct structure Use of differences column required	Penalise erroneous continuity corrections and wrong sd. Condone both numerators reversed. Φ(1.23) – Φ(-1.63) leads to 0.8907 – 0.0516 =0.8391 Only allow 0.839 if 0.8387 is seen.

Question			Answer	Marks	Guidance	
3	(ii)		P(all 5 weigh at least 30kg) $= 0.5793^5$ $= 0.0652$	M1 A1 **[2]**	Allow FT $(1 - \text{their (i)}(A))^5$ or [their $P(X \geq 30)]^5$ FT only $(1 - \text{their (i)}(A))^5$	Allow 0.06524, allow 0.065 www
3	(iii)		$P(\text{weight} > 30) = 0.05$ $P(Z > \frac{30-26.8}{\sigma}) = 0.05$ $\Phi^{-1}(0.95) = 1.645$ $\frac{30-26.8}{\sigma} = 1.645$ $\sigma = \frac{30-26.8}{1.645} = 1.945$ kg	B1 M1* M1dep* A1 **[4]**	For 1.645. B0 for 1 – 1.645 or 0.1645 For equation as seen or equivalent, with their $z > 1$. Rearranging for σ CAO	NOTE use of -1.645 allowed only if numerator reversed. Condone use of spurious c.c. if already penalised in parts (i)(*A*) or (i)(*B*). See additional guidance notes. Allow $\sigma = 1.95$ www
3	(iv)			G1 G1 G1 G1	For two Normal shapes including attempt at asymptotic behaviour with horizontal axis at each of the four ends For means, shown explicitly or by scale **on a single diagram** For lower max height for Male For visibly greater width for	Penalise clear asymmetry If shown explicitly, the positions must be consistent with horizontal scale if present. If not labelled, assume the larger mean represents Male If not labelled, assume the

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th colspan="2">Guidance</th></tr>
<tr><td></td><td></td><td></td><td></td><td>[4]</td><td>Male</td><td>larger mean represents Male</td></tr>
<tr><td>4</td><td>(a)</td><td></td><td>H_0: no association between student smoking and parent smoking
H_1: some association between student smoking and parent smoking</td><td>B1</td><td>Correct hypotheses in context

NB if H_0 H_1 reversed do not award first B1 or final B1dep*</td><td>Allow hypotheses in terms of independence, in context.
Do not allow "relationship" or "correlation" for "association"</td></tr>
<tr><td></td><td></td><td></td><td>

Expected frequency	Parent smokes	Parent does not smoke
Student smokes	15.2	32.8
Student does not smoke	22.8	49.2

</td><td>B1

B1</td><td>For at least one row/column of expected values correct May be implied by correct contributions or correct X^2

All correct</td><td></td></tr>
<tr><td></td><td></td><td></td><td>

Contribution	Parent smokes	Parent does not smoke
Student smokes	2.213	1.026
Student does not smoke	1.475	0.684

</td><td>M1

A1</td><td>For valid attempt at $(O-E)^2/E$

All correct (to 3 d.p.)</td><td>NB These M1A1 marks cannot be implied by a correct final value of X^2</td></tr>
<tr><td></td><td></td><td></td><td>$X^2 = 5.398$</td><td>B1</td><td>Allow awrt 5.40</td><td>Do not penalise use of Yates correction, giving $X^2 = 4.51$</td></tr>
<tr><td></td><td></td><td></td><td>Refer to χ_1^2
Critical value at 5% level = 3.841</td><td>B1
B1</td><td>For 1 degree of freedom
CAO for cv.</td><td>
p value = 0.02016</td></tr>
</table>

Question			Answer	Marks	Guidance	
					No further marks from here if wrong or omitted, unless *p*-value used instead.	
			Result is significant	B1*		For significant oe FT their test statistic
			There is sufficient evidence to **suggest** that there is **association between student smoking and parent smoking.**	B1dep*	NB if H_0 H_1 reverse do not award first B1 or final B1dep*	For **non-assertive** conclusion in context Allow conclusion in terms of independence FT their test statistic. Do not allow "relationship" or "correlation" for "association".
				[10]		
4	(b)		$\bar{x} = 88.2/100 = 0.882$	B1	For 0.882 seen.	
			$s = \sqrt{\dfrac{78.68-(88.2)^2/100}{99}} = \sqrt{\dfrac{0.8876}{99}}$ $= \sqrt{0.0089657}$	M1	For correctly structured calculation for the sample standard deviation or variance	
			= 0.0947 (allow 0.095 www)	A1	Allow A1 for s^2 = 0.0089657	or 0.00897 (allow 0.009**0**)
			H_0: $\mu = 0.87$; H_1 : $\mu > 0.87$	B1	For both correct	Hypotheses in words only must refer to population.
			Where μ denotes the **mean nicotine** content (of cigarettes of this brand in the population)	B1	For definition of μ in context.	Do not allow other symbols unless clearly defined as population mean.

Question			Answer	Marks	Guidance	
			Test statistic = $\dfrac{0.882-0.87}{0.0947/\sqrt{100}}=\dfrac{0.012}{0.00947}$	M1*	including correct use of √100.	FT their s (not s = 0.87 or √0.87 or 88.2 or √88.2 or 78.68 or √78.68)
			=1.267	A1	CAO	
			Upper 1% level 1 tailed critical value of z = 2.326	B1	For 2.326 **No further marks from here if wrong**	
			1.267 < 2.326 (Not significant.)	M1dep*	For sensible comparison leading to a conclusion (even if incorrect)	
			There is **insufficient evidence to suggest** that the **mean nicotine content** of this brand is **greater** than 0.87mg.	A1	For non-assertive conclusion in words and in context. FT only candidate's test statistic	
				[10]		

<u>Additional Notes on Correct Structure in Q1(ii)</u>
Equivalent calculations for finding b are allowed. For example use of $7S_{yt}/7S_{tt}$ is allowed. However, where these are mixed we award M0. e.g. use of $7S_{yt}/S_{tt}$ would earn M0. For M1 to be awarded, the calculation must be structurally equivalent to the one provided – NOTE if it is believed that the candidate has made an error in transcription of a number (for example using 244 instead of 224) we can allow M1 BOD if the structure is otherwise correct.

<u>Additional Notes for Q3 (iii)</u>
M1* is for forming a suitable equation using their z-value but it must be reasonably clear that the value used is a z-value – for example we do not allow 0.05 or 0.95 to be treated as z-values here. The M1dep* can be awarded if the candidate correctly rearranges their equation to find σ. Hence, use of an incorrect z-value could earn max B0M1*M1dep*A0. However, if it is clear that the z-value is from the wrong tail (e.g. -1.645 used in place of +1.645) then award 0/4. In cases where -1.645 is used and the numerator of the equation is reversed allow full credit and annotate with BOD.

Additional Notes on Sensible Comparisons

In Q4 (b) Neither 1.267 > 0.05 nor 0.1026 < 2.326 are considered sensible as each compares a z-value with a probability.
For 1.267 > 2.326 leading to a conclusion, allow M1A0.

Additional Notes on Conclusions to Hypothesis Tests

The following are examples of conclusions which are considered too assertive.

There is sufficient evidence to reject H_0 and **conclude** that…
"there is a positive association between…" or
"there seems to be evidence that there is a positive association between…" or
"the mean nicotine content is greater …."
"there doesn't appear to be association between…"

Also note that final conclusions **must refer to H_1 in context** for the final mark to be given.
e.g. In Q4 (a), a conclusion just stating that "the evidence suggests that there is association" gets A0 as this does not refer to the context.

Additional Notes on Alternative Methods in Q4 (b)

Critical value method	cv = 0.87 + 2.326 × 0.0946÷ √100	gets M1* B1 (for 2.326)
	= 0.8920	gets A1 cao (replacing the A1 for 1.267)
	0.882 < 0.8920	gets M1dep* if a conclusion is made. The final A1 available as before if 2.326 used.
Probability Method	P(sample mean > 1.267) = 0.1026	gets M1*A1 B1 (the B1 for 0.1026 (allow 0.1025), from tables, replaces the B1 for 2.326).
	0.1026 > 0.01	gets M1dep* if a conclusion is made. The final A1 available as before provided that B1 for 0.1026 awarded NOTE Condone B1 0.8974 (0.8975) if compared with 0.99 at which point the final M1dep*A1 are available. B0M0A0A0 if 0.8974 obtained from P(sample mean > -1.267).