OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) **4768**

Statistics 3

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
None

Thursday 15 January 2009
Morning

Duration: 1 hour 30 minutes

4768

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that 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 graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- 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**.
- This document consists of **4** pages. Any blank pages are indicated.

1 **(a)** A continuous random variable X has probability density function

$$f(x) = \lambda x^c, \quad 0 \leqslant x \leqslant 1,$$

where c is a constant and the parameter λ is greater than 1.

(i) Find c in terms of λ. **[3]**

(ii) Find $E(X)$ in terms of λ. **[3]**

(iii) Show that $\text{Var}(X) = \dfrac{\lambda}{(\lambda+2)(\lambda+1)^2}$. **[4]**

(b) Every day, Godfrey does a puzzle from the newspaper and records the time taken in minutes. Last year, his median time was 32 minutes. His times for a random sample of 12 puzzles this year are as follows.

40 20 18 11 47 36 38 35 22 14 12 21

Use an appropriate test, with a 5% significance level, to examine whether Godfrey's times this year have decreased on the whole. **[8]**

2 A factory manufactures paperweights consisting of glass mounted on a wooden base. The volume of glass, in cm^3, in a paperweight has a Normal distribution with mean 56.5 and standard deviation 2.9. The volume of wood, in cm^3, also has a Normal distribution with mean 38.4 and standard deviation 1.1. These volumes are independent of each other. For the purpose of quality control, paperweights for testing are chosen at random from the factory's output.

(i) Find the probability that the volume of glass in a randomly chosen paperweight is less than $60\,cm^3$. **[3]**

(ii) Find the probability that the total volume of a randomly chosen paperweight is more than $100\,cm^3$. **[3]**

The glass has a mass of 3.1 grams per cm^3 and the wood has a mass of 0.8 grams per cm^3.

(iii) Find the probability that the total mass of a randomly chosen paperweight is between 200 and 220 grams. **[6]**

(iv) The factory manager introduces some modifications intended to reduce the mean mass of the paperweights to 200 grams or less. The variance is also affected but not the Normality. Subsequently, for a random sample of 10 paperweights, the sample mean mass is 205.6 grams and the sample standard deviation is 8.51 grams. Is there evidence, at the 5% level of significance, that the intended reduction of the mean mass has not been achieved? **[6]**

3 Pathology departments in hospitals routinely analyse blood specimens. Ideally the analysis should be done while the specimens are fresh to avoid any deterioration, but this is not always possible. A researcher decides to study the effect of freezing specimens for later analysis by measuring the concentrations of a particular hormone before and after freezing. He collects and divides a sample of 15 specimens. One half of each specimen is analysed immediately, the other half is frozen and analysed a month later. The concentrations of the particular hormone (in suitable units) are as follows.

Immediately	15.21	13.36	15.97	21.07	12.82	10.80	11.50	12.05
After freezing	15.96	10.65	13.38	15.00	12.11	12.65	12.48	8.49

Immediately	10.90	18.48	13.43	13.16	16.62	14.91	17.08
After freezing	9.13	15.53	11.84	8.99	16.24	14.03	16.13

A t test is to be used in order to see if, on average, there is a reduction in hormone concentration as a result of being frozen.

(i) Explain why a paired test is appropriate in this situation. **[2]**

(ii) State the hypotheses that should be used, together with any necessary assumptions. **[4]**

(iii) Carry out the test using a 1% significance level. **[7]**

(iv) A $p\%$ confidence interval for the true mean reduction in hormone concentration is found to be (0.4869, 2.8131). Determine the value of p. **[5]**

4 **(i)** Explain the meaning of 'opportunity sampling'. Give one reason why it might be used and state one disadvantage of using it. **[3]**

A market researcher is conducting an 'on-street' survey in a busy city centre, for which he needs to stop and interview 100 people. For each interview the researcher counts the number of people he has to ask until one agrees to be interviewed. The data collected are as follows.

No. of people asked	1	2	3	4	5	6	7 or more
Frequency	26	19	17	13	11	8	6

A model for these data is proposed as follows, where p (assumed constant throughout) is the probability that a person asked agrees to be interviewed, and $q = 1 - p$.

No. of people asked	1	2	3	4	5	6	7 or more
Probability	p	pq	pq^2	pq^3	pq^4	pq^5	q^6

(ii) Verify that these probabilities add to 1 whatever the value of p. **[2]**

(iii) Initially it is thought that on average 1 in 4 people asked agree to be interviewed. Test at the 10% level of significance whether it is reasonable to suppose that the model applies with $p = 0.25$. **[9]**

(iv) Later an estimate of p obtained from the data is used in the analysis. The value of the test statistic (with no combining of cells) is found to be 9.124. What is the outcome of this new test? Comment on your answer in relation to the outcome of the test in part **(iii)**. **[4]**

4768 Statistics 3

<table>
<tr><td>Q1 (a)</td><td>$f(x) = \lambda x^c,\ 0 \le x \le 1,\ \lambda > 1$</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>$\int_0^1 \lambda x^c \,dx = 1$
$\therefore \left[\dfrac{\lambda x^{c+1}}{c+1}\right]_0^1 = 1$
$\therefore \dfrac{\lambda}{c+1} = 1 \qquad \therefore c = \lambda - 1$</td><td>M1
M1
A1</td><td>Correct integral, with limits (possibly appearing later), set equal to 1.
Integration correct and limits used.
c.a.o.</td><td>3</td></tr>
<tr><td>(ii)</td><td>$E(X) = \int_0^1 \lambda x^\lambda \,dx$
$= \left[\dfrac{\lambda x^{\lambda+1}}{\lambda+1}\right]_0^1 = \dfrac{\lambda}{\lambda+1}.$</td><td>M1
M1
A1</td><td>Correct form of integral for E(X). Allow c's expression for c.
Integration correct and limits used. ft c's c.</td><td>3</td></tr>
<tr><td>(iii)</td><td>$E(X^2) = \int_0^1 \lambda x^{\lambda+1} \,dx$
$= \left[\dfrac{\lambda x^{\lambda+2}}{\lambda+2}\right]_0^1 = \dfrac{\lambda}{\lambda+2}.$
$\mathrm{Var}(X) = \dfrac{\lambda}{\lambda+2} - \left(\dfrac{\lambda}{\lambda+1}\right)^2 = \dfrac{\lambda(\lambda+1)^2 - \lambda^2(\lambda+2)}{(\lambda+2)(\lambda+1)^2}$
$= \dfrac{\lambda^3 + 2\lambda^2 + \lambda - \lambda^3 - 2\lambda^2}{(\lambda+2)(\lambda+1)^2} = \dfrac{\lambda}{(\lambda+2)(\lambda+1)^2}.$</td><td>M1
A1
M1
A1</td><td>Correct form of integral for $E(X^2)$. Allow c's expression for c.

Use of $\mathrm{Var}(X) = E(X^2) - E(X)^2$. Allow c's $E(X^2)$ and E(X).
Algebra shown convincingly. Beware printed answer.</td><td>4</td></tr>
<tr><td>(b)</td><td>
<table>
<tr><th>Times</th><th>− 32</th><th>Rank of |diff|</th></tr>
<tr><td>40</td><td>8</td><td>4</td></tr>
<tr><td>20</td><td>−12</td><td>7</td></tr>
<tr><td>18</td><td>−14</td><td>8</td></tr>
<tr><td>11</td><td>−21</td><td>12</td></tr>
<tr><td>47</td><td>15</td><td>9</td></tr>
<tr><td>36</td><td>4</td><td>2</td></tr>
<tr><td>38</td><td>6</td><td>3</td></tr>
<tr><td>35</td><td>3</td><td>1</td></tr>
<tr><td>22</td><td>−10</td><td>5</td></tr>
<tr><td>14</td><td>−18</td><td>10</td></tr>
<tr><td>12</td><td>−20</td><td>11</td></tr>
<tr><td>21</td><td>−11</td><td>6</td></tr>
</table>
$W_+ = 1 + 2 + 3 + 4 + 9 = 19$

Refer to Wilcoxon single sample tables for $n = 12$.
Lower (or upper if 59 used) 5% tail is 17 (or 61 if 59 used).
Result is not significant.
Seems that there is no evidence that Godfrey's times have decreased.</td><td>

M1
M1
A1

B1

M1
A1

A1
A1</td><td>H_0: $m = 32$, H_1: $m < 32$, where m is the population median time.

for subtracting 32.
for ranks.
ft if ranks wrong.

(or $W_- = 5 + 6 + 7 + 8 + 10 + 11 + 12 = 59$)
No ft from here if wrong.
i.e. a 1-tail test. No ft from here if wrong.
ft only c's test statistic.
ft only c's test statistic.</td><td>8</td></tr>
<tr><td></td><td></td><td></td><td></td><td>18</td></tr>
</table>

Q2	$V_G \sim \text{N}(56.5,\ 2.9^2)$ $V_W \sim \text{N}(38.4,\ 1.1^2)$		When a candidate's answers suggest that (s)he appears to have neglected to use the difference columns of the Normal distribution tables penalise the first occurrence only.	
(i)	$\text{P}(V_G < 60) = \text{P}(Z < \frac{60-56.5}{2.9} = 1.2069)$ $= 0.8862$	M1 A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	$V_T \sim \text{N}(56.5 + 38.4 = 94.9,$ $2.9^2 + 1.1^2 = 9.62)$ $\text{P}(\text{this} > 100) = \text{P}(Z > \frac{100-94.9}{3.1016} = 1.6443)$ $= 1 - 0.9499 = 0.0501$	B1 B1 A1	Mean. Variance. Accept sd (= 3.1016). c.a.o.	3
(iii)	$W_T \sim \text{N}(3.1 \times 56.5 + 0.8 \times 38.4 = 205.87,$ $3.1^2 \times 2.9^2 + 0.8^2 \times 1.1^2 = 81.5945)$ $\text{P}(200 < \text{this} < 220)$ $= \text{P}(\frac{200-205.87}{9.0330} < Z < \frac{220-205.87}{9.0330})$ $= \text{P}(-0.6498 < Z < 1.5643)$ $= 0.9411 - (1 - 0.7422) = 0.6833$	M1 A1 M1 A1 M1 A1	Use of "mass = density × volume" Mean. Variance. Accept sd (= 9.0330). Formulation of requirement. c.a.o.	6
(iv)	Given $\bar{x} = 205.6$ $s_{n-1} = 8.51$ H_0: $\mu = 200$, H_1: $\mu > 200$ Test statistic is $\frac{205.6 - 200}{\frac{8.51}{\sqrt{10}}}$ $= 2.081.$ Refer to t_9. Single-tailed 5% point is 1.833. Significant. Seems that the required reduction of the mean weight has not been achieved.	M1 A1 M1 A1 A1 A1	Allow alternative: 200 + (c's 1.833) × $\frac{8.51}{\sqrt{10}}$ (= 204.933) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 1.833) × $\frac{8.51}{\sqrt{10}}$ (= 200.667) for comparison with 200.) c.a.o. but ft from here in any case if wrong. Use of $200 - \bar{x}$ scores M1A0, but ft. No ft from here if wrong. $\text{P}(t > 2.081) = 0.0336$. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	6
				18

Q3				
(i)	In this situation a paired test is appropriate because there are clearly differences between specimens … … which the pairing eliminates.	E1 E1		2
(ii)	H_0: $\mu_D = 0$ H_1: $\mu_D > 0$	B1	Both. Accept alternatives e.g. $\mu_D < 0$ for H_1, or $\mu_A - \mu_B$ etc provided adequately defined. Hypotheses in words only must include "population".	
	Where μ_D is the (population) mean reduction in hormone concentration.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\overline{X} = ...$" or similar unless $\overline{X}$ is clearly and explicitly stated to be a <u>population</u> mean.	
	Must assume • Sample is random • Normality of differences	B1 B1		4
(iii)	<u>MUST</u> be PAIRED COMPARISON t test. Differences (reductions) (before – after) are		Allow "after – before" if consistent with alternatives above.	
	–0.75 2.71 2.59 6.07 0.71 –1.85 –0.98 3.56 1.77 2.95 1.59 4.17 0.38 0.88 0.95			
	$\bar{x} = 1.65$ $s_{n-1} = 2.100(3)$ $(s_{n-1}{}^2 = 4.4112)$	B1	Do not allow $s_n = 2.0291$ ($s_n{}^2 = 4.1171$)	
	Test statistic is $\dfrac{1.65 - 0}{\frac{2.100}{\sqrt{15}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 0 + (c's 2.624) × $\frac{2.100}{\sqrt{15}}$ (= 1.423) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 2.624) × $\frac{2.100}{\sqrt{15}}$ (= 0.227) for comparison with 0.)	
	= 3.043.	A1	c.a.o. but ft from here in any case if wrong. Use of $0 - \bar{x}$ scores M1A0, but ft.	
	Refer to t_{14}.	M1	No ft from here if wrong. P(t > 3.043) = 0.00438.	
	Single-tailed 1% point is 2.624.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Seems mean concentration of hormone has fallen.	A1	ft only c's test statistic.	7
(iv)	CI is 1.65 ±	M1	ft c's $\bar{x}$ ±.	
	$k \times \dfrac{2.100}{\sqrt{15}}$	M1	ft c's s_{n-1}.	
	= (0.4869, 2.8131)	A1	A correct equation in k using either end of the interval or the width of the interval.	
	∴ $k = 2.145$	A1	Allow ft c's $\bar{x}$ and s_{n-1}.	
	By reference to t_{14} tables this is a 95% CI.	A1	c.a.o.	5
				18

Q4				
(i)	Sampling which selects from those that are (easily) available. Circumstances may mean that it is the only economically viable method available. Likely to be neither random nor representative.	E1 E1 E1		3
(ii)	$p + pq + pq^2 + pq^3 + pq^4 + pq^5 + q^6$ $= \frac{p(1-q^6)}{1-q} + q^6 = \frac{p(1-q^6)}{p} + q^6$ $= 1 - q^6 + q^6 = 1$	M1 A1	Use of GP formula to sum probabilities, or expand in terms of p or in terms of q. Algebra shown convincingly. Beware answer given.	2
(iii)	With $p = 0.25$			

Probability	0.25	0.1875	0.140625	0.105469	0.079102	0.059326	0.177979
Expected fr	25.00	18.75	14.0625	10.5469	7.9102	5.9326	17.7979

		M1 M1 A1	Probabilities correct to 3 dp or better. × 100 for expected frequencies. All correct and sum to 100.	
	$X^2 = 0.04 + 0.0033 + 0.6136 + 0.5706 + 1.2069 + 0.7204 + 7.8206 = 10.97(54)$	M1 A1	c.a.o.	
	(If e.g. only 2dp used for expected f's then $X^2 = 0.04 + 0.0033 + 0.6148 + 0.5690 + 1.2071 + 0.7226 + 7.8225 = 10.97(93)$)			
	Refer to χ_6^2.	M1	Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 10.975) = 0.0891$.	
	Upper 10% point is 10.64. Significant. Suggests model with $p = 0.25$ does not fit.	A1 A1 A1	No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	9
(iv)	Now with $X^2 = 9.124$ Refer to χ_5^2.	M1	Allow correct df (= cells – 2) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 9.124) = 0.1042$.	
	Upper 10% point is 9.236. Not significant. (Suggests new model does fit.) Improvement to the model is due to estimation of p from the data.	A1 A1 E1	No ft from here if wrong. Correct conclusion. Comment about the effect of estimated p, consistent with conclusion in part (iii).	4
				18