OCR
RECOGNISING ACHIEVEMENT

Friday 21 June 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4768/01 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768/01
- 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.
- 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.
- 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 In the past, the times for workers in a factory to complete a particular task had a known median of 7.4 minutes. Following a review, managers at the factory wish to know if the median time to complete the task has been reduced.

(i) A random sample of 12 times, in minutes, gives the following results.

6.90 7.23 6.54 7.62 7.04 7.33 6.74 6.45 7.81 7.71 7.50 6.32

Carry out an appropriate test using a 5% level of significance. **[10]**

(ii) Some time later, a much larger random sample of times gives the following results.

$$n = 80 \qquad \Sigma x = 555.20 \qquad \Sigma x^2 = 3863.9031$$

Find a 95% confidence interval for the true mean time for the task. Justify your choice of which distribution to use. **[6]**

(iii) Describe briefly one advantage and one disadvantage of having a 99% confidence interval instead of a 95% confidence interval. **[2]**

2 A company supplying cattle feed to dairy farmers claims that its new brand of feed will increase average milk yields by 10 litres per cow per week. A farmer thinks the increase will be less than this and decides to carry out a statistical investigation using a paired *t* test. A random sample of 10 dairy cows are given the new feed and then their milk yields are compared with their yields when on the old feed. The yields, in litres per week, for the 10 cows are as follows.

Cow	A	B	C	D	E	F	G	H	I	J
Old feed	144	130	132	146	137	140	140	149	138	133
New feed	148	139	138	159	138	148	146	156	147	145

(i) Why is it sensible to use a paired test? **[1]**

(ii) State the condition necessary for a paired *t* test. **[2]**

(iii) Assuming the condition stated in part **(ii)** is met, carry out the test, using a significance level of 5%, to see whether it appears that the company's claim is justified. **[10]**

(iv) Find a 95% confidence interval for the mean increase in the milk yield using the new feed. **[4]**

3 The random variable X has the following probability density function, $\mathrm{f}(x)$.

$$\mathrm{f}(x) = \begin{cases} kx(x-5)^2 & 0 \leqslant x < 5 \\ 0 & \text{elsewhere} \end{cases}$$

(i) Sketch $\mathrm{f}(x)$. **[3]**

(ii) Find, in terms of k, the cumulative distribution function, $\mathrm{F}(x)$. **[3]**

(iii) Hence show that $k = \dfrac{12}{625}$. **[2]**

The random variable X is proposed as a model for the amount of time, in minutes, lost due to stoppages during a football match. The times lost in a random sample of 60 matches are summarised in the table. The table also shows some of the corresponding expected frequencies given by the model.

Time (minutes)	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$
Observed frequency	5	15	23	11	6
Expected frequency			17.76	9.12	1.632

(iv) Find the remaining expected frequencies. **[3]**

(v) Carry out a goodness of fit test, using a significance level of 2.5%, to see if the model might be suitable in this context. **[8]**

4 A company that makes meat pies includes a "small" size in its product range. These pies consist of a pastry case and meat filling, the weights of which are independent of each other. The weight of the pastry case, C, is Normally distributed with mean 96 g and variance 21 g^2. The weight of the meat filling, M, is Normally distributed with mean 57 g and variance 14 g^2.

(i) Find the probability that, in a randomly chosen pie, the weight of the pastry case is between 90 and 100 g. **[4]**

(ii) The wrappers on the pies state that the weight is 145 g. Find the proportion of pies that are underweight. **[3]**

(iii) The pies are sold in packs of 4. Find the value of w such that, in 95% of packs, the total weight of the 4 pies in a randomly chosen pack exceeds w g. **[5]**

(iv) It is required that the weight of the meat filling in a pie should be at least 35% of the total weight. Show that this means that $0.65M - 0.35C \geqslant 0$. Hence find the probability that, in a randomly chosen pie, this requirement is met. **[6]**

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td>1</td><td>(i)</td><td></td><td>H_0: $m = 7.4$ H_1: $m < 7.4$
where m is the population median time.</td><td>B1

B1</td><td>Both. Accept hypotheses in words, but must include “population”. Do NOT allow symbols other than m unless clearly and explicitly stated to be a <u>population</u> <u>median</u>.

Adequate definition of m to include “population”.</td></tr>
<tr><td></td><td></td><td></td><td>
<table>
<tr><th>Times</th><th>−7.4</th><th>Rank of |diff|</th></tr>
<tr><td>6.90</td><td>−0.50</td><td>8</td></tr>
<tr><td>7.23</td><td>−0.17</td><td>3</td></tr>
<tr><td>6.54</td><td>−0.86</td><td>10</td></tr>
<tr><td>7.62</td><td>0.22</td><td>4</td></tr>
<tr><td>7.04</td><td>−0.36</td><td>6</td></tr>
<tr><td>7.33</td><td>−0.07</td><td>1</td></tr>
<tr><td>6.74</td><td>−0.66</td><td>9</td></tr>
<tr><td>6.45</td><td>−0.95</td><td>11</td></tr>
<tr><td>7.81</td><td>0.41</td><td>7</td></tr>
<tr><td>7.71</td><td>0.31</td><td>5</td></tr>
<tr><td>7.50</td><td>0.10</td><td>2</td></tr>
<tr><td>6.32</td><td>−1.08</td><td>12</td></tr>
</table>
</td><td>M1

M1

A1</td><td>for subtracting 7.4.

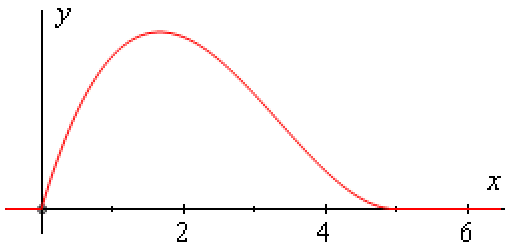
for ranking.

All correct. ft if ranks wrong.</td></tr>
<tr><td></td><td></td><td></td><td>$W_+ = 2 + 4 + 5 + 7 = 18$</td><td>B1</td><td>($W_- = 1 + 3 + 6 + 8 + 9 + 10 + 11 + 12 = 60$)</td></tr>
<tr><td></td><td></td><td></td><td>Refer to Wilcoxon single sample tables for $n = 12$.</td><td>M1</td><td>No ft from here if wrong.</td></tr>
<tr><td></td><td></td><td></td><td>Lower 5% point is 17 (or upper is 61 if 60 used).</td><td>A1</td><td>i.e. a 1-tail test. No ft from here if wrong.</td></tr>
<tr><td></td><td></td><td></td><td>Result is not significant.</td><td>A1</td><td>ft only c’s test statistic.</td></tr>
<tr><td></td><td></td><td></td><td>Insufficient evidence to suggest that the median time has been reduced.</td><td>A1

[10]</td><td>ft only c’s test statistic. Conclusion in context.</td></tr>
</table>

Question			Answer	Marks	Guidance
1	(ii)		$\bar{x} = 6.94 \quad s = 0.37$	B1	Accept $s^2 = 0.1369$. Beware use of msd (0.13518875) or rmsd (0.3676(8)). Do not allow here or below.
			CI is given by $6.94 \pm$	M1	ft c's $\bar{x} \pm$.
			1.96	B1	1.96 seen.
			$\times \dfrac{0.37}{\sqrt{80}}$	M1	ft c's s but not rmsd.
			$= 6.94 \pm 0.0811 = (6.859, 7.021)$	A1	c.a.o. Must be expressed as an interval. [rmsd gives $6.94 \pm 0.0805(7) = (6.8594(2), 7.0205(7))$]
			Normal distribution can be used because the sample size is large enough for the Central Limit Theorem to apply.	E1	CLT essential
				[6]	
1	(iii)		Advantage: A 99% confidence interval is more likely to contain the true mean.	E1	O.e.
			Disadvantage: A 99% confidence interval is less precise/wider.	E1	O.e.
				[2]	
2	(i)		A paired test would eliminate any differences between individual cattle.	E1	
				[1]	
2	(ii)		Must assume: Normality of population …	B1	
			… of <u>differences</u>.	B1	
				[2]	

Question			Answer	Marks	Guidance
2	(iii)		H_0: $\mu_D = 10$ H_1: $\mu_D < 10$	B1	Both. Accept alternatives e.g. $\mu_D > -10$ for H_1, or $\mu_A - \mu_B$ etc provided adequately defined. Hypotheses in words only must include "population". Do NOT allow " $\overline{X} = \ldots$ " or similar unless $\overline{X}$ is clearly and explicitly stated to be a <u>population</u> mean.
			Where μ_D is the (population) mean increase/difference in milk yield. <u>MUST</u> be PAIRED COMPARISON *t* test.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.
			Differences (increases) (after – before) are: 4 9 6 13 1 8 6 7 9 12	M1	Allow "before – after" if consistent with alternatives for hypotheses above.
			$\bar{x} = 7.5$ $s_{n-1} = 3.566(8)$ $(s_{n-1}{}^2 = 12.722(2))$	A1	Do not allow $s_n = 3.3837$ $(s_n{}^2 = 11.45)$.
			Test statistic is $\dfrac{7.5 - 10}{\frac{3.5668}{\sqrt{10}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow reversed numerator compared with 2.2164 Allow alternative: 10 – (c's 1.833) × $\dfrac{3.5668}{\sqrt{10}}$ (= 7.933) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ + (c's 1.833) × $\dfrac{3.5668}{\sqrt{10}}$ (= 9.567) for comparison with 10.)
			= –2.2164.	A1	c.a.o. but ft from here in any case if wrong. Use of 10 – $\bar{x}$ scores M1A0, but ft.
			Refer to t_9.	M1	No ft from here if wrong.
			Single-tailed 5% point is –1.833.	A1	Must be minus 1.833 unless absolute values are being compared. No ft from here if wrong. P(t < –2.2164) = 0.0269.
			Significant.	A1	ft only c's test statistic.
			Sufficient evidence to suggest that the mean milk yield has not increased by 10 litres (per cow per week).	A1	ft only c's test statistic. Conclusion in context to include "on average" o.e. Accept "Sufficient evidence to suggest that the **company's** claim is not justified." o.e.
				[10]	

Question			Answer	Marks	Guidance
2	(iv)		CI is given by $7.5 \pm$ $2{\cdot}262$ $\times \dfrac{3.5668}{\sqrt{10}}$ $= 7.5 \pm 2.5514 = (4.948,\ 10.052)$	M1 B1 M1 A1 **[4]**	ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_9 is OK. Allow c's $\bar{x}$. 2.262 seen. Allow c's s_{n-1}. c.a.o. Must be expressed as an interval.
3	(i)			G1 G1 G1 **[3]**	Curve, through the origin and in the first quadrant only. A single maximum; curve returns to $y = 0$; nothing to the right of $x = 5$. No t.pt at $x = 0$; t.pt. at $x = 5$; (5, 0) labelled (p.i. by an indicated scale).
3	(ii)		$\mathrm{F}(x) = k\int_0^x t(t-5)^2\,\mathrm{d}t$ $= k\left[\dfrac{t^4}{4} - \dfrac{10t^3}{3} + \dfrac{25t^2}{2}\right]_0^x$ $= k\left(\dfrac{x^4}{4} - \dfrac{10x^3}{3} + \dfrac{25x^2}{2}\right)$	M1 M1 A1 **[3]**	Correct integral for F(x) with limits (which may appear later). Correctly integrated. Limits used correctly to obtain expression. Condone absence of "–0". Do not require complete definition of F(x). Dependent on both M1's

Question			Answer	Marks	Guidance
3	(iii)		F(5) = 1 $\therefore k\left(\frac{5^4}{4}-\frac{10\times 5^3}{3}+\frac{25\times 5^2}{2}\right)=1$ $\therefore k\left(\frac{1875-5000+3750}{12}\right)=1$ $\therefore k\times\frac{625}{12}=1$ $\therefore k=\frac{12}{625}$	M1 A1 **[2]**	Substitute x = 5 and equate to 1. Expect to see evidence of at least this line of working (oe) for A1. Convincingly shown. Beware printed answer.
3	(iv)		For $0 \le x < 1$, Expected f = 60 × F(1) $=60\times\frac{12}{625}\left(\frac{1^4}{4}-\frac{10\times 1^3}{3}+\frac{25\times 1^2}{2}\right)=10.848$ For $1 \le x < 2$, Expected f = 60 – Σ(the rest) = 20.64	M1 A1 B1 **[3]**	Use of 60 × F(x) with correct k. Allow also 31.488 – frequency for $1 \le x < 2$ provided that one found using F(x). Allow either frequency found by integration. FT 31.488 – previous answer. Or allow 60 × (F(2) – F(1))

Question			Answer	Marks	Guidance
3	(v)		H_0: The model is suitable / fits the data. H_1: The model is not suitable / does not fit the data.	B1	Both hypotheses. Must be the right way round. Do not accept "data fit model" oe.
			Merge last 2 cells: Obs f = 17, Exp f = 10.752	M1	
			X^2 = 3.1525 + 1.5411 + 1.5460 + 3.6307	M1	Calculation of X^2.
			= 9.870	A1	c.a.o.
			Refer to χ_3^2.	M1	Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong.
			Upper 2.5% point is 9.348.	A1	No ft from here if wrong. $P(X^2 > 9.870) = 0.0197$.
			Significant.	A1	ft only c's test statistic.
			Sufficient evidence to suggest that the model is not suitable in this context.	A1	ft only c's test statistic. Conclusion in context. Do not accept "data do not fit model" oe.
				[8]	
4			$C \sim N(96, 21)$ $M \sim N(57, 14)$		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.
4	(i)		$P(90 < C < 100)$		
			$= P\left(\frac{90-96}{\sqrt{21}} < Z < \frac{100-96}{\sqrt{21}}\right)$	M1	For standardising. Award once, here or elsewhere. SC – candidates with consistent variances of 21² and 14² can be awarded all M and B marks
			$= P(-1.3093 < Z < 0.8729)$	A1	Either side correct. SC – 0.2857, 0.1905
			= 0.8086 – (1 – 0.9047)	A1	Both table values correct. Or 0.8086 – 0.0953 SC 0.5755 – (1 – 0.6125)
			= 0.7133	A1	c.a.o.
				[4]	
4	(ii)		Total weight $T \sim N(153, 35)$	B1	Mean.
				B1	Variance. Accept sd = 5.916… SC 637 sd = 25.239
			$P(T < 145) = P\left(Z < \frac{145-153}{\sqrt{35}} = -1.3522\right)$		
			= 1 – 0.9118 = 0.0882	A1	c.a.o.
				[3]	

Question			Answer	Marks	Guidance
4	(iii)		$T_1 + T_2 + T_3 + T_4 \sim \text{N}(612, 140)$	B1	Mean.
				B1	Variance. Accept sd = 11.832… SC = 2548 sd = 50.478
			Require w such that P(this > w) = 0.95	M1	
			$\therefore w = 612 - 1.645 \times \sqrt{140}$	B1	1.645 seen.
			= 592.5(3)	A1	c.a.o.
				[5]	
4	(iv)		Require $M \geq 0.35(M + C)$	M1	Formulate requirement.
			$\therefore 0.65M \geq 0.35C$		
			$\therefore 0.65M - 0.35C \geq 0$	A1	Convincingly shown. Beware printed answer.
			$0.65M - 0.35C \sim$	B1	Mean.
			$\text{N}((0.65 \times 57) - (0.35 \times 96) = 3.45,$	M1	For use of at least one of $0.65^2 \times \ldots$ or $0.35^2 \times \ldots$
			$(0.65^2 \times 14) + (0.35^2 \times 21) = 8.4875)$	A1	Variance. Accept sd = 2.913… SC variance = 136.83 sd = 11.70
			$\text{P(This} \geq 0) = \text{P}\left(Z \geq \frac{0 - 3.45}{\sqrt{8.4875}} = -1.1842 \right)$		
			= 0.8818	A1	c.a.o.
				[6]	