Thursday 12 June 2014 – Afternoon

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

3144250734

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.

Turn over

1 **(i)** Let X be a random variable with variance σ^2. The independent random variables X_1 and X_2 are both distributed as X. Write down the variances of $X_1 + X_2$ and $2X$; explain why they are different. **[3]**

A large company has produced an aptitude test which consists of three parts. The parts are called mathematical ability, spatial awareness and communication. The scores obtained by candidates in the three parts are continuous random variables X, Y and W which have been found to have independent Normal distributions with means and standard deviations as shown in the table.

	Mean	Standard deviation
Mathematical ability, X	30.1	5.1
Spatial awareness, Y	25.4	4.2
Communication, W	28.2	3.9

(ii) Find the probability that a randomly selected candidate obtains a score of less than 22 in the mathematical ability part of the test. **[3]**

(iii) Find the probability that a randomly selected candidate obtains a total score of at least 100 in the whole test. **[4]**

(iv) For a particular role in the company, the score $2X + Y$ is calculated. Find the score that is exceeded by only 2% of candidates. **[4]**

(v) For a different role, a candidate must achieve a score in communication which is at least 60% of the score obtained in mathematical ability. What proportion of candidates do not achieve this? **[3]**

2 **(i)** Explain what is meant by a simple random sample. **[2]**

A manufacturer produces tins of paint which nominally contain 1 litre. The quantity of paint delivered by the machine that fills the tins can be assumed to be a Normally distributed random variable.

The machine is designed to deliver an average of 1.05 litres to each tin. However, over time paint builds up in the delivery nozzle of the machine, reducing the quantity of paint delivered. Random samples of 10 tins are taken regularly from the production process. If a significance test, carried out at the 5% level, suggests that the average quantity of paint delivered is less than 1.02 litres, the machine is cleaned.

(ii) By carrying out an appropriate test, determine whether or not the sample below leads to the machine being cleaned.

0.994 1.010 1.021 1.015 1.016 1.022 1.009 1.007 1.011 1.026 **[9]**

Each time the machine has been cleaned, a random sample of 10 tins is taken to determine whether or not the average quantity of paint delivered has returned to 1.05 litres.

(iii) On one occasion after the machine has been cleaned, the quality control manager thinks that the distribution of the quantity of paint is symmetrical but not necessarily Normal. The sample on this occasion is as follows.

1.055 1.064 1.063 1.043 1.062 1.070 1.059 1.044 1.054 1.053

By carrying out an appropriate test at the 5% level of significance, determine whether or not this sample supports the conclusion that the average quantity of paint delivered is 1.05 litres. **[8]**

3 **(a)** A personal trainer believes that drinking a glass of beetroot juice an hour before exercising enables endurance tests to be completed more quickly. To test his belief he takes a random sample of 12 of his trainees and, on two occasions, asks them to carry out 100 repetitions of a particular exercise as quickly as possible. Each trainee drinks a glass of water on one occasion and a glass of beetroot juice on the other occasion.

The times in seconds taken by the trainees are given in the table.

Trainee	Water	Beetroot juice
A	75.1	72.9
B	86.2	79.9
C	77.3	71.6
D	89.1	90.2
E	67.9	68.2
F	101.5	95.2
G	82.5	76.5
H	83.3	80.2
I	102.5	99.1
J	91.3	82.2
K	92.5	90.1
L	77.2	77.9

The trainer wishes to test his belief using a paired *t* test at the 1% level of significance. Assuming any necessary assumptions are valid, carry out a test of the hypotheses $H_0: \mu_D = 0, H_1: \mu_D < 0$, where μ_D is the population mean difference in times (time with beetroot juice minus time with water). **[8]**

(b) An ornithologist believes that the number of birds landing on the bird feeding station in her garden in a given interval of time during the morning should follow a Poisson distribution. In order to test her belief, she makes the following observations in 60 randomly chosen minutes one morning.

Number of birds	0	1	2	3	4	5	6	$\geqslant 7$
Frequency	2	5	10	17	14	7	4	1

Given that the data in the table have a mean value of 3.3, use a goodness of fit test, with a significance level of 5%, to investigate whether the ornithologist is justified in her belief. **[11]**

Question 4 begins on page 4

4 The probability density function of a random variable X is given by

$$f(x) = \begin{cases} kx & 0 \leqslant x \leqslant a, \\ k(2a-x) & a < x \leqslant 2a, \\ 0 & \text{otherwise,} \end{cases}$$

where a and k are positive constants.

(i) Sketch f(x). Hence explain why E(X) = a. **[3]**

(ii) Show that $k = \dfrac{1}{a^2}$. **[3]**

(iii) Find Var(X) in terms of a. **[4]**

In order to estimate the value of a, a random sample of size 50 is taken from the distribution. It is found that the sample mean and standard deviation are $\bar{x} = 1.92$ and $s = 0.8352$.

(iv) Construct a symmetrical 95% confidence interval for a. Give one reason why the answer is only approximate. **[5]**

(v) A non-statistician states that the probability that a lies in the interval found in part **(iv)** is 0.95. Comment on this statement. **[2]**

END OF QUESTION PAPER

Question			Answer	Marks	Guidance
1	(i)		$\text{Var}(X_1 + X_2) = 2\sigma^2$ $\text{Var}(2X) = 4\sigma^2$ $X_1 + X_2$ means two independent values from X are added together. $2X$ means that one value from X is multiplied by 2.	B1 B1 E1 **[3]**	Allow 2Var(X) and 4Var(X) Any comment explaining why $X_1 + X_2$ is different from $2X$
1	(ii)		$\text{P}(X < 22) = \text{P}\left(Z < \frac{22 - 30.1}{5.1}\right)$ $= \text{P}(Z < -1.5882)$ $= 0.0561$	M1 A1 A1 **[3]**	For standardising. Award once, here or elsewhere Correct z value cao
1	(iii)		$X + Y + W \sim \text{N}(83.7, 58.86)$ $\text{P}(X + Y + W > 100) = \text{P}\left(Z > \frac{100 - 83.7}{\sqrt{58.86}}\right)$ $\text{P}(Z > 2.1246) = 0.0168$	B1 B1 M1 A1 **[4]**	Mean Variance (or sd = 7.67) Correct set up cao
1	(iv)		$2X \sim \text{N}(60.2, 104.04)$ $\rightarrow 2X + Y \sim \text{N}(85.6, 121.68)$ $\text{P}(2X + Y > b) = 0.02$ $\rightarrow \frac{b - 85.6}{\sqrt{121.68}} = 2.054$ $\rightarrow b = 108.26$ Score exceeded by 2% is 108.3	 B1 B1 M1 A1 **[4]**	 Variance 2.054 seen Correct set up cao

Question			Answer	Marks	Guidance
1	(v)		$\mathrm{P}(W < 0.6X) = \mathrm{P}(W - 0.6X < 0)$	M1	Either way round
			$W - 0.6X \sim \mathrm{N}(0.14, 24.5736)$	B1	Mean and variance
			$\mathrm{P}(W - 0.6X < 0) = \mathrm{P}(Z < -2.0455)$		
			=0.0204	A1 **[3]**	Cao. Allow convincing recovery
2	(i)		A simple random sample is one where every sample of the same size has an equal probability of being selected.	E2,1,0 **[2]**	Allow E1 for every item has the same probability of being selected
2	(ii)		$\mathrm{H}_0 : \mu = 1.02 \quad \mathrm{H}_1 : \mu < 1.02$	B1	Both hypotheses. Hypotheses in words only must include "population". Do NOT allow "$\bar{X} = ...$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean.
			Where μ is the population mean volume	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.
			$\bar{x} = 1.0131 \quad s = 0.009146$	B1	Do not allow $s_n = 0.00868$here or in construction of test statistic
			Test statistic is $\dfrac{1.0131 - 1.02}{0.009146/\sqrt{10}}$	M1*	For method, allow candidates $\bar{x}, s$. Allow confidence interval approach.
			=-2.3857 value between –2.38 and –2.39	A1	cao
			Refer to t_9	M1	For t_9 No FT from here
			5% point is ± 1.833	A1	For matching 1.833 seen. No FT from here
			$-2.3857 < -1.833$ reject H_0	M1 dep	For rejection. Must compare test statistic with matching 1.833 unless absolute values are being compared. No ft from here if wrong.
			Conclude mean appears to be below 1.02 and so machine will be cleaned.	E1 dep **[9]**	Needs context

Question		Answer	Marks	Guidance
2	(iii)	H_0 : median of population volumes = 1.05 H_1 : median of population volumes ≠ 1.05	B1	Both Allow location parameter or mean (if explanation is given)

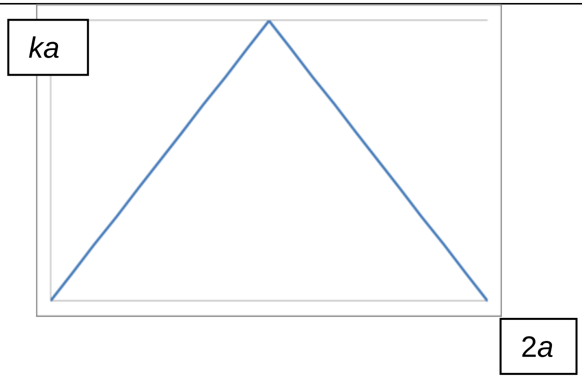
Difference	0.005	0.014	0.013	-0.007	0.012	0.020	0.009	-0.006	0.004	0.003
Rank	3	9	8	5	7	10	6	4	2	1

Answer	Marks	Guidance
	M1	For differences from 1.05. ZERO from this point if differences not used.
	M1	For ranks.
	A1	ft from here if ranks wrong.
$W_- = 4 + 5 = 9$	B1	Or $W_+ = 1+2+3+6+7+8+9+10 = 46$
Refer to tables of Wilcoxon paired (/single sample) statistic for *n* = 10	M1	No FT from here if wrong.
Critical value = 8. (47 for upper)	A1	No FT from here if wrong (i.e. 1 tail)
Cannot reject H_0 and conclude median volume appears to be 1.05	E1	Must have context.
	[8]	

Question			Answer	Marks	Guidance
3	(a)		$\mathrm{H}_0 : \mu_D = 0 \qquad \mathrm{H}_1 : \mu_D < 0$		
			Where μ_D is the population mean of the differences.		
			MUST BE PAIRED COMPARISON *t* test.	M1	
			Differences are –2.2, –6.3, –5.7, 1.1, 0.3, –6.3, –6.0, –3.1, –3.4, –9.1, –2.4, 0.7		
			$\overline{\mathrm{D}} = -3.533 \qquad s_D = 3.225$	A1	Do not allow $s_n = 3.088$
			Test statistic is $\dfrac{-3.533 - 0}{3.225/\sqrt{12}}$	M1*	For method, allow their $\overline{D}, s$. Allow confidence interval approach.
			$= -3.79$	A1	cao
			Refer to t_{11}	M1	No FT from here if wrong
			Single-tailed 1% point is –2.718	A1	No FT from here if wrong
			–3.79 < –2.718 so reject H_0	A1dep	FT their –3.79 if relevant M1 earned
			Conclude mean time appears reduced	A1dep	In context. FT their –3.79 if relevant M1 earned
				[8]	

Question			Answer	Marks	Guidance
3	(b)		H_0 : The Poisson model fits the data H_1 : The Poisson model does not fit the data	B1	Both hypotheses. Must be the right way round. Do not accept "data fits model" or equivalent.
				M1	At least 3 probabilities to 3dp or better or 3 expected values to 3sf or better
				M1	Multiply by 60 to obtain expected values
				A1	All correct to 3sf or better
			See table below	M1	Merge first 2 and last 2 cells
			$X^2 = \frac{2.516^2}{9.516} + \frac{2.050^2}{12.050} + \frac{3.745^2}{13.255} + \frac{3.065^2}{10.935} + \frac{0.217^2}{7.217} + \frac{2.027^2}{7.027}$	M1*	Calculation of X^2
			= 0.6652+0.3488+1.0581+0.8591+0.0065+0.5847		
			= 3.522 awrt 3.52	A1	cao
			Refer to χ_4^2	M1	Allow correct df from wrongly grouped table.
			Upper 5% point is 9.49	B1	No FT from here if wrong.
			3.522 < 9.49 cannot reject H_0	A1dep	FT candidates 3.522 if relevant M1 earned.
			Poisson model appears to fit data.	A1dep	FT candidates 3.522 if relevant M1 earned. Do not accept "data fits Poisson model" or equivalent.
				[11]	

r	$P(X = r)$	Expected value	Combined
0	0.03688	2.213	9.516
1	0.12171	7.303	
2	0.20083	12.050	
3	0.22091	13.255	
4	0.18225	10.935	
5	0.12029	7.217	
6	0.06616	3.969	7.027
≥ 7	0.05097	3.058	

Question			Answer	Marks	Guidance
4	(i)			G1 G1	Shape Scales on axes
			$\mathrm{E}(X) = a$ because the distribution is symmetrical about $x = a$	B1 **[3]**	Do not allow integration method
4	(ii)		Total area = $\frac{1}{2} 2a.ka$ Or $\int_0^a kx\,dx + \int_a^{2a} k(2a - x)\,dx$ $k\left[\frac{x^2}{2}\right]_0^a + k\left[2ax - \frac{x^2}{2}\right]_a^{2a}$ $k\left(\frac{a^2}{2} - 0\right) + k\left(2a^2 - \frac{3a^2}{2}\right)$	M1	Attempting to find area of triangle or setting up the correct integrals including limits (which may appear later).
			$= ka^2$	A1	oe Correctly finding area in terms of k, a
			$ka^2 = 1$ $k = \frac{1}{a^2}$	A1 **[3]**	Equating area to 1 and convincingly obtaining result. Answer given

Question		Answer	Marks	Guidance
4	(iii)	$\text{Var}(X) = k\int_0^a x^3 \mathrm{d}x + k\int_a^{2a} (2ax^2 - x^3)\,\mathrm{d}x - a^2$	M1	Correct integral for $\mathrm{E}(X^2)$ including limits (which may appear later).
		$k\left[\frac{x^4}{4}\right]_0^a + k\left[\frac{2ax^3}{3} - \frac{x^4}{4}\right]_a^{2a} - a^2$	M1	Correctly integrated (dependent on M1 above)
		$\frac{a^2}{4} + \frac{16a^2}{3} - 4a^2 - \frac{2a^2}{3} + \frac{a^2}{4} - a^2$	M1	Using $\mathrm{E}(X^2) - (\mathrm{E}(X))^2$
		$\text{Var}(X) = \frac{a^2}{6}$	A1	cao
			[4]	
4	(iv)	$\bar{x} = 1.92 \quad s = 0.8352$		Given
			B1	1.96 seen
		Interval is $1.92 \pm 1.96\frac{0.8352}{\sqrt{50}}$	M1	Correct SE
			M1	Centred on 1.92
		$= (1.69, 2.15)$	A1	cao
		The distribution of $\bar{X}$ is *approximately* Normal (CLT) or s is only an estimate.	E1	
			[5]	
4	(v)	This statement is incorrect	E1	
		The value of a either lies within this particular interval or it does not – so the probability is either 0 or 1. If a large number of such intervals were constructed, then *a* would lie in 95% of them.	E1	A comment either about p=0 or 1, or about a large number of intervals
			[2]	