OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS **2615**

Statistics 3

Friday **21 JANUARY 2005** Afternoon 1 hour 20 minutes

Additional materials:
Answer booklet
Graph paper
MEI Examination Formulae and Tables (MF12)

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** questions.
- You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The allocation 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.
- Final answers should be given to a degree of accuracy appropriate to the context.
- The total number of marks for this paper is 60.

This question paper consists of 4 printed pages.

[Turn over

1 The random variable X has the rectangular (continuous uniform) distribution over the interval $1 \leqslant x \leqslant 5$, so that its probability density function is

$$f(x) = \begin{cases} \frac{1}{4} & 1 \leqslant x \leqslant 5, \\ 0 & \text{elsewhere.} \end{cases}$$

The cumulative distribution function (cdf) of X is denoted by $F(x)$.

(i) Write down the values of the cdf of X at $x = 1$ and $x = 5$. [2]

(ii) Obtain $F(x)$, and verify that the values of $F(1)$ and $F(5)$ are correct. [3]

The random variable Y is defined by $Y = 2X^2$.

(iii) The values that Y can take are given by $y_1 \leqslant y \leqslant y_2$. Write down y_1 and y_2. [1]

(iv) Let $G(y)$ denote the cdf of Y. By considering

$$G(y) \equiv P(Y \leqslant y) = P(2X^2 \leqslant y)$$

and using the cdf of X, show that

$$G(y) = \frac{\sqrt{y}}{4\sqrt{2}} - \frac{1}{4}$$

(for $y_1 \leqslant y \leqslant y_2$). [3]

(v) Obtain the probability density function of Y. [2]

(vi) Use $G(y)$ to obtain m_Y, the median of Y. [2]

(vii) Write down m_X, the median of X. Verify that $m_Y = 2m_X^2$. [2]

2 A craftsman makes hand-made souvenirs of two types, A and B. The time taken to make a type A souvenir is a Normally distributed random variable with mean 34 minutes and standard deviation 2.6 minutes. Independently, the time taken to make a type B souvenir is a Normally distributed random variable with mean 39 minutes and standard deviation 4.0 minutes.

(i) Find the probability that it takes more than 30 minutes to make a type A souvenir. [2]

(ii) Find the probability that a type A souvenir takes more time to make than a type B souvenir. [4]

(iii) The souvenirs are packed in boxes containing 3 of each type. Find the probability that the total time to make the 6 souvenirs in such a box exceeds 210 minutes. (Assume that the souvenirs are chosen randomly and independently.) [3]

The craftsman undertakes a training course to improve his skill at making type A souvenirs. Afterwards, a random sample of 8 times taken to make type A souvenirs is as follows (in minutes).

34.9 31.8 26.1 29.9 31.4 33.3 29.1 27.9

(iv) Assuming that the underlying standard deviation has not changed, provide a two-sided 95% confidence interval for the true mean time to make a type A souvenir after the training course. Interpret this confidence interval carefully. [6]

3 Researchers at an industrial company are studying new apparatus for delivering a carefully controlled volume of a certain chemical at a particular stage in the production process. The volume delivered is specified to be 21 ml. Small variations from this can be tolerated, but the researchers need to determine whether the volume delivered is as specified on average.

The volumes delivered are carefully measured on 10 occasions and found to be as follows (in ml).

21.0 21.6 19.8 22.9 22.0 20.9 22.5 21.4 21.8 20.6

(i) State the appropriate null and alternative hypotheses for the usual t test that the researchers might use. [2]

(ii) State two conditions necessary for the correct use of the t test. [2]

(iii) Carry out the test, using a 5% significance level. [7]

(iv) Provide a two-sided 99% confidence interval for the true mean volume delivered. [4]

[Turn over

4 Psychologists are developing a new index of overall intelligence for 11-year-old children. It is assumed that the index is Normally distributed over the whole underlying population and that the standard deviation of this distribution is 12. If the index has been created correctly, its mean over the population should be 50.

(i) The index is measured for a random sample of 100 11-year-old children. It is found that the sample mean value is 47.8. Test the hypothesis that the true mean of the index is 50, against the alternative that it is not 50, at the 1% level of significance. Discuss briefly whether the assumption that the underlying distribution is Normal is necessary for this test to be valid. [8]

(ii) If the assumption of underlying Normality with mean 50 and standard deviation 12 is correct, then

- 6.68% of the population will have an index value less than 32,
- 43.32% of the population will have an index value between 32 and 50,
- 43.32% of the population will have an index value between 50 and 68,
- 6.68% of the population will have an index value greater than 68.

In the random sample of 100 11-year-old children, it is found that the numbers in these categories are 12, 48, 36 and 4 respectively. Use this information to test whether the assumption is reasonable, at the 10% significance level. [7]

Mark Scheme

2615 Mathematics

January 2005

Mark Scheme

MAXIMUM MARK	**60**

Q1	$X \sim U(1, 5)$, $f(x) = ¼$, $1 \le x \le 5$. $Y = 2X^2$.			
(i)	$F(1) = 0$ $F(5) = 1$	B1 B1		2
(ii)	$F(x) = \int_1^x f(t)\mathrm{d}t$ $= \left[\frac{t}{4}\right]_1^x = \frac{x}{4} - \frac{1}{4}$ Verifications that $F(1) = 0$ and $F(5) = 1$.	M1 A1 B1	Definition of cdf, including limits, possibly implied later. Some valid method must be seen. [for $1 \le x \le 5$ – do not insist on this.] For <u>both</u>. Must convince. If part (ii) precedes part (i) then there must be included some comment which "verifies" or reconciles the values of F(1) and F(5) between the two parts.	3
(iii)	Range of values of Y is (2, 50).	B1	For <u>both</u>. [i.e. $y_1 = 2$, $y_2 = 50$]	1
(iv)	Cdf of Y is $G(y) \equiv P(Y \le y)$ $= P(2X^2 \le y)$ $= P\left(X \le \sqrt{\frac{y}{2}}\right)$ $= F\left(\sqrt{\frac{y}{2}}\right)$ $= \frac{\sqrt{y}}{4\sqrt{2}} - \frac{1}{4}$	M1 M1 A1	<u>Note</u> Strictly speaking, there should be some consideration of a component $P\left(-\sqrt{\frac{y}{2}} \le X \le 1\right)$ which is zero. IGNORE absence of, or any errors in, this. For use of F(…). This is independent of the previous mark. BEWARE PRINTED ANSWER. Depends only on the previous M mark.	3
(v)	P.d.f. of Y is $\frac{\mathrm{d}}{\mathrm{d}y}G(y)$ $= \frac{1}{4\sqrt{2}} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{8\sqrt{2}\sqrt{y}}$	M1 A1	 [for $2 \le y \le 50$ – do not insist on this.]	2
(vi)	Median of Y, m_y, is given by $\frac{1}{2} = G(m_y)$ $= \frac{\sqrt{m_y}}{4\sqrt{2}} - \frac{1}{4}$ $\Rightarrow \sqrt{m_y} = 3\sqrt{2} \quad \Rightarrow m_y = 18$	M1 A1		2
(vii)	Median of X, m_x, is 3 and we have $2 \times 3^2 = 18$ (= m_y).	B1 B1	Allow ft from c's $F(x)$ and c's m_y.	2
				15

Q2	$A \sim N(34, \sigma = 2{\cdot}6)$ $B \sim N(39, \sigma = 4)$		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)	$P(A > 30) = P(Z > \frac{30-34}{2\cdot6} = -1{\cdot}538(46))$ $= 0{\cdot}9380$	M1 A1	For standardising. Award once, here or elsewhere. Accept 0·9381.	 2
(ii)	Want P(A > B) i.e. $P(A - B > 0)$ $A - B \sim N(-5,$ $2{\cdot}6^2 + 4^2 = 22{\cdot}76)$ $P(\text{this} > 0) = P(Z > \frac{0-(-5)}{4\cdot7707} = 1{\cdot}048)$ $= 1 - 0{\cdot}8526 = 0{\cdot}1474$	M1 B1 B1 A1	Allow $B - A$ provided subsequent work is consistent. Mean. Variance. Accept sd = $\sqrt{22{\cdot}76} = 4{\cdot}7707$. c.a.o.	 4
(iii)	$A_1 + A_2 + A_3 + B_1 + B_2 + B_3 \sim N(219,$ $68{\cdot}28)$ $P(\text{this} > 210) = P(Z > \frac{210-219}{8\cdot2632} = -1{\cdot}089)$ $= 0{\cdot}8620$	B1 B1 A1	Mean. Variance. Accept sd = $\sqrt{68{\cdot}28} = 8{\cdot}2632$. For any use of $3A + 3B$ allow B1 for mean = 219 but then B0A0. c.a.o.	 3
(iv)	$\bar{x} = 30{\cdot}55$; $\sigma = 2{\cdot}6$ is given. CI is given by $30{\cdot}55 \pm$ $1{\cdot}96$ $\times \frac{2\cdot6}{\sqrt{8}}$ $= 30{\cdot}55 \pm 1{\cdot}80(17) = (28{\cdot}74(8), 32{\cdot}35(2))$ 95% of all intervals obtained in this way in repeated sampling would "capture" the true mean time to make a type A souvenir after the training.	 M1 B1 M1 A1 E2	 ft c's $\bar{x}$ ±. c.a.o. Must be expressed as an interval. (E0, E1, E2)	 6
				15

Q3				
(i)	$H_0 : \mu = 21$ $H_1 : \mu \neq 21$ Where μ is the (population) mean volume delivered.	 B1 B1	For both. 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. Hypotheses in words only must include "population".	2
(ii)	Underlying population is Normal. Sample is random.	B1 B1		2
(iii)	$\bar{x} = 21 \cdot 45$, $s_{n-1} = 0{\cdot}9192$ $(s_{n-1}{}^2 = 0{\cdot}845)$ Test statistic is $\dfrac{21\cdot45-21}{\left(\dfrac{0\cdot9192}{\sqrt{10}}\right)}$ $= 1{\cdot}548(0)$ Refer to t_9. Double-tailed 5% point is 2·262. Not significant. Seems mean volume delivered is as specified.	B1 M1 A1 M1 A1 E1 E1	Allow $s_n = 0{\cdot}8721$ $(s_n{}^2 = 0{\cdot}7605)$ only if correctly used in sequel. Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 21 ± (c's 2·262) × $\frac{0\cdot9192}{\sqrt{10}}$ (=20·342, 21·658) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ ± (c's 2·262) × $\frac{0\cdot9192}{\sqrt{10}}$ (=20·792, 22·108) for comparison with 21.) c.a.o. but ft from here in any case if wrong. Use of $21 - \bar{x}$ scores M1A0, but ft. No ft from here if wrong. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic. Special case: (t_{10} and 2·228) can score 1 of these last 2 marks if either form of conclusion is given.	7
(iv)	CI is given by 21·45 ± 3·250 $\times \dfrac{0\cdot9192}{\sqrt{10}}$ = 21·45 ± 0·94(47) = (20·50(53), 22·39(47))	 M1 B1 M1 A1	 ft c's $\bar{x}$ ±. ft c's s_{n-1}. c.a.o. Must be expressed as an interval. ZERO if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_9 is OK.	 4
				15

Q4				
(i)	Test of $H_0 : \mu = 50$ against $H_1 : \mu \neq 50$. $n = 100$, $\bar{x} = 47 \cdot 8$, $\sigma = 12$.			
	Test statistic is $\dfrac{47\cdot 8 - 50}{\left(\dfrac{12}{\sqrt{100}}\right)}$	M1	Allow alternative: $50 \pm$ (c's 2·576) × $\frac{12}{\sqrt{100}}$ (=46·909, 53·091) for subsequent comparison with $\bar{x}$. (Or $\bar{x} \pm$ (c's 2·576) × $\frac{12}{\sqrt{100}}$ (=44·709, 50·891) for comparison with 50.)	
	$= -1\cdot 833$	A1	c.a.o. but ft from here in any case if wrong. Use of $50 - \bar{x}$ scores M1A0, but ft.	
	Refer to N(0, 1).	M1	No ft from here if wrong.	
	Double-tailed 1% point is 2·576.	A1	No ft from here if wrong. $\Phi(-1\cdot 833) = 0\cdot 0334$.	
	Not significant.	E1		
	Seems mean could be taken as 50.	E1		
	With such a large sample, it will not matter if the underlying distribution is not Normal.	E2	(E0, E1, E2)	8
(ii)	Obs: 12, 48, 36, 4 Exp: 6·68, 43·32, 43·32, 6·68	B1	For correct table of observed and expected frequencies. Award this mark but not the next A1 if the expected frequencies are rounded.	
	$X^2 = 4\cdot 2369 + 0\cdot 5056 + 1\cdot 2369 + 1\cdot 0752$	M1	Combining cells based on observed frequencies gets M0, but based (correctly) on incorrect expected frequencies allow M1 and ft for A1.	
	$= 7\cdot 05(46)$	A1	(See above.)	
	Refer to χ_3^2.	M1	ft if df ≤ 4.	
	Upper 10% point is 6·251	A1	No ft if not 10% point of candidate's χ^2.	
	Significant.	E1		
	Seems underlying distribution is not Normal.	E1		7
				15

The table in part (ii):

Obs	12	48	36	4
Exp	6·68	43·32	43·32	6·68

Examiner’s Report