

ADVANCED GCE
MATHEMATICS (MEI)
Statistics 3

4768

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



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

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

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 $\text{Var}(X) = \frac{\lambda}{(\lambda + 2)(\lambda + 1)^2}$. [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]

There are no questions printed on this page.

4768 Statistics 3

Q1 (a)	$f(x) = \lambda x^c, 0 \leq x \leq 1, \lambda > 1$																																										
(i)	$\int_0^1 \lambda x^c dx = 1$ $\therefore \left[\frac{\lambda x^{c+1}}{c+1} \right]_0^1 = 1$ $\therefore \frac{\lambda}{c+1} = 1 \quad \therefore c = \lambda - 1$	M1 M1 A1	Correct integral, with limits (possibly appearing later), set equal to 1. Integration correct and limits used. c.a.o.	3																																							
(ii)	$E(X) = \int_0^1 \lambda x^2 dx$ $= \left[\frac{\lambda x^{\lambda+1}}{\lambda+1} \right]_0^1 = \frac{\lambda}{\lambda+1}$	M1 M1 A1	Correct form of integral for $E(X)$. Allow c 's expression for c . Integration correct and limits used. ft c 's c .	3																																							
(iii)	$E(X^2) = \int_0^1 \lambda x^{\lambda+1} dx$ $= \left[\frac{\lambda x^{\lambda+2}}{\lambda+2} \right]_0^1 = \frac{\lambda}{\lambda+2}$ $\text{Var}(X) = \frac{\lambda}{\lambda+2} - \left(\frac{\lambda}{\lambda+1} \right)^2 = \frac{\lambda(\lambda+1)^2 - \lambda^2(\lambda+2)}{(\lambda+2)(\lambda+1)^2}$ $= \frac{\lambda^3 + 2\lambda^2 + \lambda - \lambda^3 - 2\lambda^2}{(\lambda+2)(\lambda+1)^2} = \frac{\lambda}{(\lambda+2)(\lambda+1)^2}$	M1 A1 M1 A1	Correct form of integral for $E(X^2)$. Allow c 's expression for c . Use of $\text{Var}(X) = E(X^2) - E(X)^2$. Allow c 's $E(X^2)$ and $E(X)$. Algebra shown convincingly. Beware printed answer.	4																																							
(b)	<table border="1" style="margin-bottom: 10px;"> <thead> <tr> <th>Times</th> <th>- 32</th> <th>Rank of diff </th> </tr> </thead> <tbody> <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> </tbody> </table> <p>$W_+ = 1 + 2 + 3 + 4 + 9 = 19$</p> <p>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.</p>	Times	- 32	Rank of diff	40	8	4	20	-12	7	18	-14	8	11	-21	12	47	15	9	36	4	2	38	6	3	35	3	1	22	-10	5	14	-18	10	12	-20	11	21	-11	6	M1 M1 A1 B1 M1 A1 A1 A1	$H_0: m = 32, H_1: m < 32$, 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.	8
Times	- 32	Rank of diff																																									
40	8	4																																									
20	-12	7																																									
18	-14	8																																									
11	-21	12																																									
47	15	9																																									
36	4	2																																									
38	6	3																																									
35	3	1																																									
22	-10	5																																									
14	-18	10																																									
12	-20	11																																									
21	-11	6																																									
				18																																							

Q2	$V_G \sim N(56.5, 2.9^2)$ $V_W \sim N(38.4, 1.1^2)$		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)	$P(V_G < 60) = P(Z < \frac{60 - 56.5}{2.9} = 1.2069)$ $= 0.8862$	M1 A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	$V_T \sim N(56.5 + 38.4 = 94.9,$ $2.9^2 + 1.1^2 = 9.62)$ $P(\text{this} > 100) = P(Z > \frac{100 - 94.9}{3.1016} = 1.6443)$ $= 1 - 0.9499 = 0.0501$	B1 B1 A1	Mean. Variance. Accept sd (= 3.1016). c.a.o.	3
(iii)	$W_T \sim N(3.1 \times 56.5 + 0.8 \times 38.4 = 205.87,$ $3.1^2 \times 2.9^2 + 0.8^2 \times 1.1^2 = 81.5945)$ $P(200 < \text{this} < 220)$ $= P(\frac{200 - 205.87}{9.0330} < Z < \frac{220 - 205.87}{9.0330})$ $= P(-0.6498 < Z < 1.5643)$ $= 0.9411 - (1 - 0.7422) = 0.6833$	M1 A1 M1 A1 M1 A1	Use of "mass = density \times volume" Mean. Variance. Accept sd (= 9.0330). Formulation of requirement. c.a.o.	6
(iv)	<p>Given $\bar{x} = 205.6$ $s_{n-1} = 8.51$ $H_0: \mu = 200, H_1: \mu > 200$</p> <p>Test statistic is $\frac{205.6 - 200}{\frac{8.51}{\sqrt{10}}}$ $= 2.081.$</p> <p>Refer to t_9.</p> <p>Single-tailed 5% point is 1.833. Significant. Seems that the required reduction of the mean weight has not been achieved.</p>	M1 A1 M1 A1 A1 A1	<p>Allow alternative: $200 + (c's 1.833) \times \frac{8.51}{\sqrt{10}}$ (= 204.933) for subsequent comparison with \bar{x}. (Or $\bar{x} - (c's 1.833) \times \frac{8.51}{\sqrt{10}}$ (= 200.667) for comparison with 200.)</p> <p>c.a.o. but fit from here in any case if wrong. Use of $200 - \bar{x}$ scores M1A0, but ft.</p> <p>No ft from here if wrong. $P(t > 2.081) = 0.0336$. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.</p>	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)	$H_0: \mu_D = 0$ $H_1: \mu_D > 0$ Where μ_D is the (population) mean reduction in hormone concentration. Must assume <ul style="list-style-type: none"> • Sample is random • Normality of differences 	B1 B1 B1 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". For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow " $\bar{X} = \dots$ " or similar unless \bar{X} is clearly and explicitly stated to be a <u>population</u> mean.	4
(iii)	<p>MUST be PAIRED COMPARISON t test. Differences (reductions) (before – after) are</p> <p>–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</p> <p>$\bar{x} = 1.65$ $s_{n-1} = 2.100(3)$ ($s_{n-1}^2 = 4.4112$)</p> <p>Test statistic is $\frac{1.65 - 0}{\frac{2.100}{\sqrt{15}}}$</p> <p style="text-align: right;">= 3.043.</p> <p>Refer to t_{14}.</p> <p>Single-tailed 1% point is 2.624. Significant. Seems mean concentration of hormone has fallen.</p>	B1 M1 A1 M1 A1 A1 A1	Allow "after – before" if consistent with alternatives above. Do not allow $s_n = 2.0291$ ($s_n^2 = 4.1171$) Allow c's \bar{x} and/or s_{n-1} . Allow alternative: $0 + (c's 2.624) \times \frac{2.100}{\sqrt{15}}$ (= 1.423) for subsequent comparison with \bar{x} . (Or $\bar{x} - (c's 2.624) \times \frac{2.100}{\sqrt{15}}$ (= 0.227) for comparison with 0.) c.a.o. but ft from here in any case if wrong. Use of $0 - \bar{x}$ scores M1A0, but ft. No ft from here if wrong. $P(t > 3.043) = 0.00438$. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	7
(iv)	<p>CI is $1.65 \pm$</p> $k \times \frac{2.100}{\sqrt{15}}$ <p style="text-align: right;">= (0.4869, 2.8131)</p> <p>$\therefore k = 2.145$ By reference to t_{14} tables this is a 95% CI.</p>	M1 M1 A1 A1 A1	ft c's $\bar{x} \pm$. ft c's s_{n-1} . A correct equation in k using either end of the interval or the width of the interval. Allow ft c's \bar{x} and s_{n-1} . 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 $p = 0.25$																							
	<table border="1"> <tr> <td>Probability</td> <td>0.25</td> <td>0.1875</td> <td>0.140625</td> <td>0.105469</td> <td>0.079102</td> <td>0.059326</td> <td>0.177979</td> </tr> <tr> <td>Expected fr</td> <td>25.00</td> <td>18.75</td> <td>14.0625</td> <td>10.5469</td> <td>7.9102</td> <td>5.9326</td> <td>17.7979</td> </tr> </table>	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							
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																	
	$X^2 = 0.04 + 0.0033 + 0.6136 + 0.5706 + 1.2069 + 0.7204 + 7.8206 = 10.97(54)$ <p>(If e.g. only 2dp used for expected f's then $X^2 = 0.04 + 0.0033 + 0.6148 + 0.5690 + 1.2071 + 0.7226 + 7.8225 = 10.97(93)$)</p> <p>Refer to χ^2_6.</p> <p>Upper 10% point is 10.64. Significant. Suggests model with $p = 0.25$ does not fit.</p>	M1 M1 A1 M1 A1 M1 A1 A1 A1	Probabilities correct to 3 dp or better. $\times 100$ for expected frequencies. All correct and sum to 100. c.a.o. Allow correct df (= cells - 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 10.975) = 0.0891$. 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^2_5 . 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.	M1 A1 A1 E1	Allow correct df (= cells - 2) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 9.124) = 0.1042$. No ft from here if wrong. Correct conclusion. Comment about the effect of estimated p , consistent with conclusion in part (iii).					4																
								18																

4768 Statistics 3

General Comments

There were 291 candidates from 40 centres (January 2008: 232 from 31) for this sitting of the paper. Overall the general standard of many of the scripts seen was about as good as in recent sessions. However, the work of possibly more candidates showed considerable carelessness. For example, time and again candidates would select the wrong critical value for a hypothesis test or state the final conclusions badly. As in the past, the quality of the comments, interpretations and explanations was patchy, and usually less good than the rest of the work.

Invariably all four questions were attempted. Marks for Questions 1 and 2 were found to be somewhat higher on average than Questions 3 and 4. There was no evidence to suggest that candidates were unable to complete the paper although they may have needed to rush at the end.

Comments on Individual Questions

- 1) Continuous random variables; Wilcoxon single sample test; times for Godfrey to complete his daily puzzle.
 - (a) Most candidates made a decent attempt at this half of the question and many fully correct solutions were seen. Those relatively few candidates who were not successful were usually struggling from the outset.
 - (i) It was pleasing to see the correct integral set up, including limits, and equated to 1. Almost always the correct expression for c was found.
 - (ii) Again the work seen here was usually correct and competent.
 - (iii) In this part it was very pleasing to find that candidates were able to obtain the given expression for the variance in a thorough, careful manner with sufficient working shown for it to be “convincing”.
 - (b) The Wilcoxon test was found by the vast majority to be straightforward and easily achieved. Only a handful of candidates tried to test the median using a t test.
- 2) Combinations of Normal distributions; t test for a population mean; paperweights.
 - (i) This part was answered correctly by almost everyone.
 - (ii) Except for a very occasional problem with the variance, this part, too, was almost always correct.
 - (iii) In this part most candidates coped very well, finding the variance of the total mass correctly and going on to interpret the requirement correctly. However, some (but not as many as in the past) did forget to square all parts fully when working out the variance of the mass. Also there were a number of candidates who interpreted $P(200 < X < 220)$ as $P(X < 220)$ and $P(X > 200)$, which were then combined in a variety of ways.

- (iv) Although the hypotheses for this test were not requested, many candidates followed good practice and stated them nonetheless. On this occasion it was particularly helpful to be able to see what they thought the question was asking. Much of the time the alternative hypothesis seen was " $H_1: \mu < 200$ ". However the work that followed was either mostly correct (up to deciding to "reject H_0 ") but giving the wrong conclusion in context or it broke down for some other reason (e.g. looking up the wrong critical value).

As for the conclusion in context, candidates might reasonably have reflected that a sample mean as high as 205.6 could not be taken to suggest that the intended reduction had been achieved. In other words, in a one-tail test significance at one end of a distribution does not imply significance at the other end. Even when the sense of the conclusion was correct, it was often flawed in the way it was expressed (e.g. being too assertive or omitting "mean").

- 3) Paired t test for the population mean reduction in hormone concentration; confidence interval for the true population mean.

Throughout this question marks were lost through carelessness and/or a lack of thoroughness.

- (i) This part was very poorly answered. There were only a very small handful of candidates who managed to say anything meaningful at all about why a paired test was being used. Usually the answer given was "because the data occur in pairs", with no thought as to why that was so.
- (ii) Answers to this part of the question were also disappointing. The hypotheses were badly expressed; words like "mean" and "population" were frequently missing. There seemed to be a widespread reluctance to use " μ " or " μ_D " for the population mean reduction/difference.

Similarly, the assumptions were poorly stated. For a paired test it is the population of differences that needs to be Normally distributed, not the "before" and "after" measurements, and, in order to avoid misunderstanding, candidates must be prepared to make clear which population they are referring to. Also, very many candidates overlooked the need for a random sample as one of the assumptions.

- (iii) There was much careless work at the start of this part of the question with candidates making mistakes finding the differences in concentration and/or the mean and standard deviation of the sample of differences. Then, as for the test in Question 2, there were errors finding the correct critical value from the tables and in expressing the final conclusion.
- (iv) Most candidates showed that they were familiar with the structure of a confidence interval. However, time after time, having found the correct percentage point and even related it to the correct entry in the tables for t_{14} they left their answers as " $p = 5\%$ ", forgetting that the confidence interval would normally be described as "95%".

- 4) Sampling; Chi-squared test of goodness of fit of a given model; numbers of people asked to take part in a survey.
- (i) Most candidates could give a passable explanation of “opportunity sampling” and could suggest a disadvantage of using it. Far fewer were able to provide a convincing explanation of why one might end up using it; the tendency was to focus on “when” rather than “why”.
 - (ii) It was very disappointing to find that hardly any of the candidates appeared able to recognise, let alone write down the sum of, a geometric progression. A very common alternative approach was to substitute $(1 - p)$ for q , expand the binomial expressions and collect all the terms. Almost all attempts at this fell apart fairly quickly as candidates seemed unable to manage the expansions (beyond about $(1 - p)^2$) and the use of Pascal’s triangle. A third possible approach was to replace p by $(1 - q)$ and see all the resulting terms in q cancel out. This worked well and easily for the relatively few who tried it.
 - (iii) For the most part the expected frequencies and the value of the test statistic were calculated correctly, and only occasional inconsistencies in rounding results were noticed. Usually, but not always, the correct number of degrees of freedom and critical value followed. The final conclusion of the test was not as carefully expressed as it should have been. Some candidates thought they were testing a binomial model, others thought that they were testing “ $p = 0.25$ ” rather than a model in which $p = 0.25$.
 - (iv) Although many candidates came to the correct conclusion for the revised test, there were also many who did not think to adjust the critical value to allow for the loss of one degree of freedom. Hardly any candidates explained the change in outcome, i.e. the improved fit of the model, as a consequence of estimating a parameter from the data.