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

Wednesday 17 June 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.

Turn over

1 Andy, a carpenter, constructs wooden shelf units for storing CDs. The wood used for the shelves has a thickness which is Normally distributed with mean 14 mm and standard deviation 0.55 mm. Andy works to a design which allows a gap of 145 mm between the shelves, but past experience has shown that the gap is Normally distributed with mean 144 mm and standard deviation 0.9 mm. Dimensions of shelves and gaps are assumed to be independent of each other.

(i) Find the probability that a randomly chosen gap is less than 145 mm. **[3]**

(ii) Find the probability that the combined height of a gap and a shelf is more than 160 mm. **[3]**

A complete unit has 7 shelves and 6 gaps.

(iii) Find the probability that the overall height of a unit lies between 960 mm and 965 mm. Hence find the probability that at least 3 out of 4 randomly chosen units are between 960 mm and 965 mm high. **[7]**

(iv) I buy two randomly chosen CD units made by Andy. The probability that the difference in their heights is less than h mm is 0.95. Find h. **[5]**

2 Pat makes and sells fruit cakes at a local market. On her stall a sign states that the average weight of the cakes is 1 kg. A trading standards officer carries out a routine check of a random sample of 8 of Pat's cakes to ensure that they are not underweight, on average. The weights, in kg, that he records are as follows.

0.957 1.055 0.983 0.917 1.015 0.865 1.013 0.854

(i) On behalf of the trading standards officer, carry out a suitable test at a 5% level of significance, stating your hypotheses clearly. Assume that the weights of Pat's fruit cakes are Normally distributed. **[9]**

(ii) Find a 95% confidence interval for the true mean weight of Pat's fruit cakes. **[4]**

Pat's husband, Tony, is the owner of a factory which makes and supplies fruit cakes to a large supermarket chain. A large random sample of n of these cakes has mean weight $\bar{x}$ kg and variance 0.006 kg^2.

(iii) Write down, in terms of n and $\bar{x}$, a 95% confidence interval for the true mean weight of cakes produced in Tony's factory. **[3]**

(iv) What is the size of the smallest sample that should be taken if the width of the confidence interval in part **(iii)** is to be 0.025 kg at most? **[3]**

3 A company which employs 600 staff wishes to improve its image by introducing new uniforms for the staff to wear. The human resources manager would like to obtain the views of the staff. She decides to do this by means of a systematic sample of 10% of the staff.

(i) How should she go about obtaining such a sample, ensuring that all members of staff are equally likely to be selected? Explain whether this constitutes a simple random sample. **[5]**

At a later stage in the process, the choice of uniform has been reduced to two possibilities. Twelve members of staff are selected to take part in deciding which of the two uniforms to adopt. Each of the twelve assesses each uniform for comfort, appearance and practicality, giving it a total score out of 10. The scores are as follows.

Staff member	1	2	3	4	5	6	7	8	9	10	11	12
Uniform A	4.2	2.6	10.0	9.0	8.2	2.8	5.0	7.4	2.8	6.8	10.0	9.8
Uniform B	5.0	5.2	1.4	2.8	2.2	6.4	7.4	7.8	6.8	1.2	3.4	7.6

A Wilcoxon signed rank test is to be used to decide whether there is any evidence of a preference for one of the uniforms.

(ii) Explain why this test is appropriate in these circumstances and state the hypotheses that should be used. **[4]**

(iii) Carry out the test at the 5% significance level. **[8]**

4 A random variable X has probability density function $f(x) = \frac{2x}{\lambda^2}$ for $0 < x < \lambda$, where λ is a positive constant.

(i) Show that, for any value of λ, $f(x)$ is a valid probability density function. **[3]**

(ii) Find μ, the mean value of X, in terms of λ and show that $P(X < \mu)$ does not depend on λ. **[4]**

(iii) Given that $E(X^2) = \frac{\lambda^2}{2}$, find σ^2, the variance of X, in terms of λ. **[2]**

The random variable X is used to model the depth of the space left by the filling machine at the top of a jar of jam. The model gives the following probabilities for X (whatever the value of λ).

$0 < X \leqslant \mu - \sigma$	$\mu - \sigma < X \leqslant \mu$	$\mu < X \leqslant \mu + \sigma$	$\mu + \sigma < X < \lambda$
0.18573	0.25871	0.36983	0.18573

A sample of 50 random observations of X, classified in the same way, is summarised by the following frequencies.

4	11	20	15

(iv) Carry out a suitable test at the 5% level of significance to assess the goodness of fit of X to these data. Explain briefly how your conclusion may be affected by the choice of significance level. **[9]**

4768 Statistics 3

Q1	W ~ N(14, 0.552) G ~ N(144, 0.9^2)		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(G < 145) = $\mathrm{P}\left(Z < \frac{145-144}{0.9} = 1.1111\right)$	M1 A1	For standardising. Award once, here or elsewhere.	
	= 0·8667	A1	c.a.o.	3
(ii)	$W + G \sim \mathrm{N}(14+144=158,$	B1	Mean.	
	$\sigma^2 = 0.55^2 + 0.9^2 = 1.1125\ \ )$	B1	Variance. Accept sd (= 1.0547…).	
	P(this > 160) = $\mathrm{P}\left(Z > \frac{160-158}{1.0547} = 1\cdot 896\right) = 1 - 0\cdot 9710 = 0.0290$	A1	c.a.o.	3
(iii)	$H = W_1 + ... + W_7 + G_1 + ... + G_6 \sim \mathrm{N}(962,$	B1	Mean.	
	$\sigma^2 = 0.55^2 + ... + 0.55^2 + 0.9^2 + ... + 0.9^2 = 6.9775)$	B1	Variance. Accept sd (= 2.6415).	
	P(960 < this < 965) = $\mathrm{P}\left(\frac{960-962}{2\cdot 6415} = -0.7571 < Z < \frac{965-962}{2\cdot 6415} = 1.1357\right)$	M1	Two-sided requirement.	
	= 0·8720 – (1 – 0.7755) = 0.6475	A1	c.a.o.	
	Now want P(B(4, 0.6475) ≥ 3)	M1	Evidence of attempt to use binomial. ft c's p value.	
	= 4 × 0.6475^3 × 0.3525 + 0.6475^4	M1	Correct terms attempted. ft c's p value. Accept 1 – P(… ≤ 2)	
	= 0.38277 + 0.17577 = 0.5585	A1	c.a.o.	7
(iv)	$D = H_1 - H_2 \sim \mathrm{N}(0,$	B1	Mean. (May be implied.)	
	$6.9775 + 6.9775 = 13.955)$	B1	Variance. Accept sd (= 3.7356). Ft 2 × c's 6.9775 from (iii).	
	Want h s.t. P(−h < D < h) = 0.95	M1	Formulation of requirement as 2-sided.	
	i.e. P(D < h) = 0 975 $\therefore h = \sqrt{13.955} \times 1.96 \qquad = 7.32$	B1 A1	For 1.96. c.a.o.	5
				18

Q2				
(i)	H$_0$: $\mu = 1$ H$_1$: $\mu < 1$	B1	Both hypotheses. Hypotheses in words only must include "population".	
	where μ is the mean weight of the cakes.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\bar{X} = \ldots$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a <u>population</u> mean.	
	$\bar{x} = 0.957375 \quad s_{n-1} = 0.07314(55)$	B1	s_n = 0.06842 but do <u>NOT</u> allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{0.957375 - 1}{\frac{0.07314}{\sqrt{8}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 1 + (c's –1.895) × $\frac{0.07314}{\sqrt{8}}$ (= 0.950997) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's –1.895) × $\frac{0.07314}{\sqrt{8}}$ (= 1.006377) for comparison with 1.)	
	= –1.648(24).	A1	c.a.o. but ft from here in any case if wrong. Use of 1 – $\bar{x}$ scores M1A0, but ft.	
	Refer to t_7.	M1	No ft from here if wrong. P(t < –1.648(24)) = 0.0716.	
	Single-tailed 5% point is –1.895.	A1	Must be minus 1.895 unless absolute values are being compared. No ft from here if wrong.	
	Not significant.	A1	ft only c's test statistic.	
	Insufficient evidence to suggest that the cakes are underweight on average.	A1	ft only c's test statistic.	9
(ii)	CI is given by 0.957375 ±	M1		
	2·365	B1		
	$\times \frac{0.07314}{\sqrt{8}}$	M1		
	= 0.957375 ± 0.061156= (0.896(2), 1.018(5))	A1	c.a.o. Must be expressed as an interval. ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_7 is OK.	4
(iii)	$\bar{x} \pm 1.96 \times \sqrt{\frac{0.006}{n}}$	M1 B1 A1	Structure correct, incl. use of Normal. 1.96.	3

			All correct.	
(iv)	2 × 1.96 × $\sqrt{\frac{0.006}{n}}$ < 0.025	M1	Set up appropriate inequation. Condone an equation.	
	$n > \left(\frac{2\times1.96}{0.025}\right)^2 \times 0.006 = 147.517$	M1	Attempt to rearrange and solve.	
	So take *n* = 148	A1	c.a.o. (expressed as an integer). S.C. Allow max M1A1(c.a.o.) when the factor "2" is missing. ($n > 36.879$)	3
				19

Q3				
(i)	For a systematic sample			
	• she needs a list of all staff	E1		
	• with no cycles in the list.	E1		
	All staff equally likely to be chosen if she			
	• chooses a random start between 1 and 10	E1		
	• then chooses every 10th.	E1		
	Not simple random sampling since not all samples are possible.	E1		5
(ii)	Nothing is known about the background population ..	E1	Any reference to unknown distribution or “non-parametric” situation.	
	... of differences between the scores.	E1	Any reference to pairing/differences.	
	H_0: $m = 0$ H_1: $m \neq 0$	B1	Both hypotheses. Hypotheses in words only must include “population”.	
	where m is the population median difference for the scores.	B1	For adequate verbal definition.	4
(iii)				

Diff	−0.8	−2.6	8.6	6.2	6.0	−3.6	−2.4	−0.4	−4.0	5.6	6.6	2.2
Rank	2	5	12	10	9	6	4	1	7	8	11	3

		M1	For differences. ZERO in this section if differences not used.	
		M1	For ranks.	
		A1	ft from here if ranks wrong.	
	$W_{-} = 1 + 2 + 4 + 5 + 6 + 7 = 25$	B1	(or $W_{+} = 3 + 8 + 9 + 10 + 11 + 12 = 53$)	
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 12$.	M1	No ft from here if wrong.	
	Lower (or upper if 53 used) 2½% tail is 13 (or 65 if 53 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
	Result is not significant.	A1	ft only c’s test statistic.	
	No evidence to suggest a preference for one of the uniforms.	A1	ft only c’s test statistic.	8
				17

Q4	$\mathrm{f}(x)=\frac{2x}{\lambda^2}$ for $0<x<\lambda$, $\lambda>0$			
(i)	f(x) > 0 for all x in the domain.	E1		
	$\int_0^\lambda \frac{2x}{\lambda^2}\,\mathrm{d}x=\left[\frac{x^2}{\lambda^2}\right]_0^\lambda=\frac{\lambda^2}{\lambda^2}=1$	M1	Correct integral with limits.	
		A1	Shown equal to 1.	3
(ii)	$\mu=\int_0^\lambda \frac{2x^2}{\lambda^2}\,\mathrm{d}x=\left[\frac{2x^3/3}{\lambda^2}\right]_0^\lambda=\frac{2\lambda}{3}$	M1	Correct integral with limits.	
		A1	c.a.o.	
	$\mathrm{P}(X<\mu)=\int_0^\mu \frac{2x}{\lambda^2}\,\mathrm{d}x=\left[\frac{x^2}{\lambda^2}\right]_0^\mu$	M1	Correct integral with limits.	
	$=\frac{\mu^2}{\lambda^2}=\frac{4\lambda^2/9}{\lambda^2}=\frac{4}{9}$			
	which is independent of λ.	A1	Answer plus comment. ft c's μ provided the answer does not involve λ.	4
(iii)	Given $\mathrm{E}(X^2)=\frac{\lambda^2}{2}$			
	$\sigma^2=\frac{\lambda^2}{2}-\frac{4\lambda^2}{9}=\frac{\lambda^2}{18}$	M1	Use of Var(X) = E(X^2) – E(X)^2.	
		A1	c.a.o.	2
(iv)				

Probability	0.18573	0.25871	0.36983	0.18573
Expected f	9.2865	12.9355	18.4915	9.2865

		M1	Probs × 50 for expected frequencies.	
		A1	All correct.	
	X^2 = 3.0094 + 0.2896 + 0.1231 + 3.5152	M1	Calculation of X^2.	
	= 6.937(3)	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. P(X^2 > 6.937) = 0.0739.	
	Upper 5% point is 7.815.	A1	No ft from here if wrong.	
	Not significant.	A1	ft only c's test statistic.	
	Suggests model fits the data for these jars.	A1	ft only c's test statistic.	
	But with a 10% significance level (cv = 6.251) a different conclusion would be reached.	E1	Any valid comment which recognises that the test statistic is close to the critical values.	9
				18

4768 Statistics 3

General Comments

There were 371 candidates from 77 centres (June 2008: 348 from 72) for this sitting of the paper. The overall standard of the scripts seen suggested reasonable understanding of most, but by no means all, of the content of this module. However, Question 3 parts (i) (systematic sampling) and (ii) (reasons for the use of the Wilcoxon test) were conspicuously badly answered, with even good candidates appearing to have little, if any, feel for either of the two issues.

In a number of places it seemed that candidates had not read the question carefully before starting to answer it. Also, candidates continue to show poor regard for clear and accurate notation in their work, and for the need for accurate computation. On a number of occasions the work contained glaring errors of a kind that one simply would not normally expect to see at this level. Furthermore, despite the remarks made in last June's report concerning the quality of the language used in the conclusions to hypothesis tests, the deterioration in respect of this has continued.

Invariably all four questions were attempted. With very few exceptions there was no evidence to suggest that candidates found themselves unable to complete the paper in the available time.

Comments on Individual Questions

1) **Combinations of Normal distributions. CD units.**

   (i) In this part, just about all candidates managed to make a good start to the question and the paper as a whole.

   (ii) This part was well answered too. Only occasionally was there an issue with the wrong variance.

   (iii) The first stage of this part was answered correctly by the majority of candidates, but a substantial minority made a fundamental mistake when calculating the variance. The second stage, which required a binomial probability, was usually recognised as such but the calculation was frequently limited to the probability of just 3 out of 4, not "at least 3 out of 4".

   (iv) Fully correct answers to this part were rarely seen. Many, but by no means all, candidates gave the correct mean and variance for the difference between two CD units. However, in almost all cases the requirement was interpreted as one- rather than two-sided.

2) **The *t* distribution: test and confidence interval for a population mean. Use of a confidence interval for a population mean from a large sample to find the sample size. The weights of Pat & Tony's cakes.**

(i) Most candidates appeared to have learnt good habits when stating the hypotheses for this kind of test; weaker ones still neglected to define their symbol "μ". It was noticeable that many candidates did not seem to be making the best use of their calculators to find the mean and sample standard deviation; a range of values for these and the subsequent test statistic was seen, all the consequence of different levels of rounding at different stages of the calculations. The last stage, completion of the test, was usually well done except that the wording of the final conclusion often lacked any reference to the *average* weight of the cakes and/or was considered to be too assertive.
A noticeable minority of candidates elected to test the *difference* between the weights of Pat's cakes and the advertised average, 1 kg. This created extra, unnecessary work for them, and usually they were unable to express their hypotheses in a clear and coherent manner.

(ii) The confidence interval proved to be a straightforward task for many candidates. Among weaker candidates there was a costly tendency to use the Normal distribution.

(iii) In this part the large sample size meant that the Normal distribution should be used (Central Limit Theorem); that was not a problem for most. However, many candidates used the value of the variance given in the question as the standard deviation. This had a significant implication for part (iv).

(iv) It seemed that many candidates knew what to do here, and most remembered to include a factor of 2 for the total width of the interval. However, if the variance was used as the standard deviation (see part (iii)) then the final answer became rather unrealistic. Candidates seemed to accept this without any apparent concern. Furthermore, most candidates undertook to answer this part by setting up and attempting to solve an equation instead of an inequation, and many let themselves down by their poor facility with some fairly basic algebra.

3) **Sampling; Wilcoxon paired sample test. Employees' attitudes to new uniforms.**

(i) This part was very badly answered. Two things were quite clear: that candidates did not read the question and that they had little, if any, understanding of "systematic sampling" and "simple random sampling". This part of the specification (Sampling) continues to be conspicuously badly understood. Candidates suggested a variety of strategies, including the random ordering of the list of employees, intended to ensure that all would be equally likely to be selected, thereby guaranteeing, in their eyes, that a "simple random sample" would be obtained. Time and again candidates wrote that, in order to obtain a systematic sample of 10% of 600, one should select every 60th employee.

(ii) This was another part that was badly answered. While candidates would go on to carry out successfully the Wilcoxon test using paired data in part (iii), their responses here suggested that they could neither explain why they were doing it nor state (through their hypotheses) precisely and clearly what they would be testing. The word "mean" and/or the symbol "μ" were not uncommon. Frequently any symbol that was conveniently to hand was used to represent the median in the hypotheses ("m" would seem to be an obvious choice), and usually the word "population" was missing from the definition of it. In a number of cases the null hypothesis was given as "median of A – median of B = 0"; candidates should be aware that this is *not* equivalent to "median of (A – B) = 0".

(iii) Most candidates answered this part well, showing that they could reproduce the algorithm of the Wilcoxon test easily and reliably. Furthermore, for this test the final conclusion was usually well expressed in non-assertive terms.

4) **Continuous random variables; Chi-squared test of goodness of fit. The depth of space left in the top of a jar of jam.**

(i) Most candidates realised that they needed to show the condition that the integral of a p.d.f. over the domain equals 1. However, very few remembered to check (or even just state) that the given function was required to be non-negative in the domain, and those that did remember usually did not address the first condition.

(ii) Many candidates integrated successfully to find the mean of the distribution. Not as many managed to show satisfactorily that the probability connected with it was independent of λ; some worked out the correct probability but neglected to comment on it and some just made no attempt.

(iii) It was a little frustrating to see candidates working out by integration the value of $E(X^2)$, even though it was given to them in the question. These candidates then often seemed to think that the given expression was $\{E(X)\}^2$ which, of course, would make the variance equal to 0.

(iv) There were many good answers to this part of the question. A number of candidates believed, incorrectly, that they should combine the first two classes. There were occasional errors concerning the number of degrees of freedom and/or the critical value, and there was, for some, the usual shortcomings in the final conclusion. The last point required candidates to notice that the test statistic actually fell between critical values given in the tables and to pass comment accordingly. It was pleasing to see that very many of them did precisely that.