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:
- Scientific or graphical calculator

Tuesday 22 June 2010
Afternoon

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.

3R–0A14

1 **(i)** The manager of a company that employs 250 travelling sales representatives wishes to carry out a detailed analysis of the expenses claimed by the representatives. He has an alphabetical (by surname) list of the representatives. He chooses a sample of representatives by selecting the 10th, 20th, 30th and so on. Name the type of sampling the manager is attempting to use. Describe a weakness in his method of using it, and explain how he might overcome this weakness. **[3]**

The representatives each use their own cars to drive to meetings with customers. The total distance, in miles, travelled by a representative in a month is Normally distributed with mean 2018 and standard deviation 96.

(ii) Find the probability that, in a randomly chosen month, a randomly chosen representative travels more than 2100 miles. **[3]**

(iii) Find the probability that, in a randomly chosen 3-month period, a randomly chosen representative travels less than 6000 miles. What assumption is needed here? Give a reason why it may not be realistic. **[5]**

(iv) Each month every representative submits a claim for travelling expenses plus commission. Travelling expenses are paid at the rate of 45 pence per mile. The commission is 10% of the value of sales in that month. The value, in £, of the monthly sales has the distribution $N(21\,200,\ 1100^2)$. Find the probability that a randomly chosen claim lies between £3000 and £3300. **[7]**

2 William Sealy, a biochemistry student, is doing work experience at a brewery. One of his tasks is to monitor the specific gravity of the brewing mixture during the brewing process. For one particular recipe, an initial specific gravity of 1.040 is required. A random sample of 9 measurements of the specific gravity at the start of the process gave the following results.

1.046 1.048 1.039 1.055 1.038 1.054 1.038 1.051 1.038

(i) William has to test whether the specific gravity of the mixture meets the requirement. Why might a t test be used for these data and what assumption must be made? **[3]**

(ii) Carry out the test using a significance level of 10%. **[9]**

(iii) Find a 95% confidence interval for the true mean specific gravity of the mixture and explain what is meant by a 95% confidence interval. **[6]**

3 **(a)** In order to prevent and/or control the spread of infectious diseases, the Government has various vaccination programmes. One such programme requires people to receive a booster injection at the age of 18. It is felt that the proportion of people receiving this booster could be increased and a publicity campaign is undertaken for this purpose. In order to assess the effectiveness of this campaign, health authorities across the country are asked to report the percentage of 18-year-olds receiving the booster before and after the campaign. The results for a randomly chosen sample of 9 authorities are as follows.

Authority	A	B	C	D	E	F	G	H	I
Before	76	98	88	81	86	84	83	93	80
After	82	97	93	77	83	95	91	95	89

This sample is to be tested to see whether the campaign appears to have been successful in raising the percentage receiving the booster.

(i) Explain why the use of paired data is appropriate in this context. **[1]**

(ii) Carry out an appropriate Wilcoxon signed rank test using these data, at the 5% significance level. **[10]**

(b) Benford's Law predicts the following probability distribution for the first significant digit in some large data sets.

Digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

On one particular day, the first significant digits of the stock market prices of the shares of a random sample of 200 companies gave the following results.

Digit	1	2	3	4	5	6	7	8	9
Frequency	55	34	27	16	15	17	12	15	9

Test at the 10% level of significance whether Benford's Law provides a reasonable model in the context of share prices. **[7]**

[Question 4 is printed overleaf.]

4 A random variable X has an exponential distribution with probability density function $\mathrm{f}(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$, where λ is a positive constant.

(i) Verify that $\int_0^\infty \mathrm{f}(x)\,\mathrm{d}x = 1$ and sketch $\mathrm{f}(x)$. **[5]**

(ii) In this part of the question you may use the following result.

$$\int_0^\infty x^r e^{-\lambda x}\,\mathrm{d}x = \frac{r!}{\lambda^{r+1}} \quad \text{for } r = 0,\ 1,\ 2,\ \ldots\ .$$

Derive the mean and variance of X in terms of λ. **[6]**

The random variable X is used to model the lifetime, in years, of a particular type of domestic appliance. The manufacturer of the appliance states that, based on past experience, the mean lifetime is 6 years.

(iii) Let $\overline{X}$ denote the mean lifetime, in years, of a random sample of 50 appliances. Write down an approximate distribution for $\overline{X}$. **[4]**

(iv) A random sample of 50 appliances is found to have a mean lifetime of 7.8 years. Does this cast any doubt on the model? **[3]**

Q1	D ~ N(2018, $\sigma = 96$)		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)	Systematic Sampling. It lacks any element of randomness. Choose a random starting point in the range 1 – 10.	B1 E1 E1	 May be implied by the next mark. Allow reasonable alternatives e.g. "the list may contain cycles." Beware proposals for a different sampling method.	[3]
(ii)	P(D > 2100) = $\mathrm{P}\left(Z > \frac{2100-2018}{96} = 0.8542\right)$ = 1 − 0.8034 = 0.1966	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	[3]
(iii)	$D_1 + D_2 + D_3 \sim$ N(6054, $\sigma^2 = 96^2 + 96^2 + 96^2 = 27648$) P(this < 6000) = $\mathrm{P}\left(Z < \frac{6000-6054}{166.277} = -0.3248\right)$ = 1 − 0.6273 = 0.3727 Must assume that the months are independent. This is unlikely to be realistic since e.g. consecutive months may not be independent.	B1 B1 A1 E1 E1	Mean. Variance. Accept sd (= 166.277). c.a.o. Reference to independence of months. Any sensible comment.	[5]
(iv)	Claim ~ N(2018 × 0.45 + 21200 × 0.10 = 3028.10, $96^2 \times 0.45^2 + 1100^2 \times 0.10^2 = 13966.24$ P(3000 < this < 3300) $= \mathrm{P}\left(\frac{3000-3028.1}{118.18} < Z < \frac{3300-3028.1}{118.18}\right)$ $= \mathrm{P}(-0\cdot 2378 < Z < 2.3008)$ = 0.9893 − (1 − 0.5940) = 0.5833	M1 A1 M1 A1 M1 A1 A1	Mean. c.a.o. Variance. Accept sd (= 118.18). c.a.o. Formulation of requirement: a two-sided inequality. Ft c's parameters. c.a.o.	[7]
			Total	**[18]**

Q2				
(i)	A t test might be used because • sample is small • population variance is unknown Must assume background population is Normal.	 B1 B1 B1		[3]
(ii)	H_0: $\mu = 1.040$ H_1: $\mu \neq 1.040$	B1	Both hypotheses. Hypotheses in words only must include "population". Do NOT allow " $\overline{X} = ...$ " or similar unless $\overline{X}$ is clearly and explicitly stated to be a <u>population</u> mean.	
	where μ is the mean specific gravity of the mixture.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.	
	$\bar{x} = 1.0452$ $\quad s_{n-1} = 0.007155$	B1	$s_n = 0.006746$ but do <u>NOT</u> allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{1.0452 - 1.040}{\frac{0.007155}{\sqrt{9}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 1.040 + (c's 1.860) $\times \frac{0.007155}{\sqrt{9}}$ (= 1.0444) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 860) $\times \frac{0.007155}{\sqrt{9}}$ (= 1.0407) for comparison with 1.040.)	
	= 2.189(60).	A1	c.a.o. but ft from here in any case if wrong. Use of $1.040 - \bar{x}$ scores M1A0, but ft.	
	Refer to t_8.	M1	No ft from here if wrong. P(t > 2.1896) = 0.05996.	
	Double-tailed 10% point is 1.860.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Seems mean specific gravity in the mixture does not meet the requirement.	A1	ft only c's test statistic.	[9]
(iii)	CI is given by 1.0452 ± 2·306 $\times \frac{0.007155}{\sqrt{9}}$	 M1 B1 M1		
	= 1.0452 ± 0.0055= (1.039(7), 1.050(7))	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_8 is OK.	
	In repeated sampling, 95% of confidence intervals constructed in this way will contain the true population mean.	E2	E2, 1, 0.	[6]
			Total	**[18]**

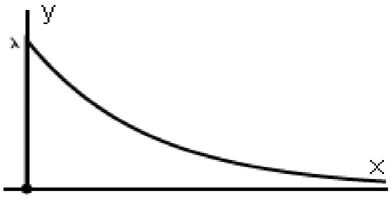
Q3			
(a) (i)	Use paired data in order to eliminate differences between authorities.	B1	[1]
(ii)	H_0: $m = 0$ H_1: $m > 0$ where m is the population median difference.	B1 Both. Accept hypotheses in words. B1 Adequate definition of m to include "population".	

Diff (After − Before)	6	−1	5	−4	−3	11	8	2	9
Rank of |diff|	6	1	5	4	3	9	7	2	8

		M1 For differences. ZERO in this section if differences not used. M1 For ranks. A1 FT from here if ranks wrong	
	$W_- = 1 + 3 + 4 = 8$ (or $= 2+5+6+7+8+9 = 37$)	B1	
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 9$.	M1 No ft from here if wrong.	
	Lower 5% point is 8 (or upper is 37 if W_+ used).	A1 i.e. a 1-tail test. No ft from here if wrong.	
	Result is significant.	A1 ft only c's test statistic.	
	Evidence suggests the percentage has been raised (on the whole).	A1 ft only c's test statistic.	[10]
(b)	H_0: Stock market prices can be modelled by Benford's Law. H_1: Stock market prices can not be modelled by Benford's Law.		

Prob	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046
Exp f	60.2	35.2	25.0	19.4	15.8	13.4	11.6	10.2	9.2
Obs f	55	34	27	16	15	17	12	15	9

		M1 Probs × 200 for expected frequencies. All correct.	
	$X^2 = 0.44917 + 0.04091 + 0.16 + 0.59588 + 0.04051$ $+ 0.96716 + 0.01379 + 2.25882 + 0.00435$ $= 4.5305(9)$	M1 Calculation of X^2. A1 c.a.o.	
	Refer to χ_8^2.	M1 Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 4.53059) = 0.80636$.	
	Upper 5% point is 13.36.	A1 No ft from here if wrong.	
	Not significant.	A1 ft only c's test statistic.	
	Suggests Benford's Law provides a reasonable model in the context of share prices.	A1 ft only c's test statistic.	[7]
		Total	[18]

Q4	$\mathrm{f}(x) = \lambda \mathrm{e}^{-\lambda x}$ for $x \geq 0$, where $\lambda > 0$.	Given $\int_0^\infty x^r \mathrm{e}^{-\lambda x} \mathrm{d}x = \dfrac{r!}{\lambda^{r+1}}$	
(i)	$\int_0^\infty \mathrm{f}(x)\mathrm{d}x = \int_0^\infty \lambda \mathrm{e}^{-\lambda x}\mathrm{d}x$ $= \left[-\mathrm{e}^{-\lambda x}\right]_0^\infty$ $= \left(0 - (-\mathrm{e}^0)\right) = 1$  	M1 Integration of f(x). M1 Use of limits or the given result. A1 Convincingly obtained (Answer given.) G1 Curve, with negative gradient, in the first quadrant only. Must intersect the y-axis. G1 (0, λ) labelled; asymptotic to x-axis.	**[5]**
(ii)	$\mathrm{E}(X) = \int_0^\infty \lambda x \mathrm{e}^{-\lambda x}\mathrm{d}x$ $= \lambda \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda}$ $\mathrm{E}(X^2) = \int_0^\infty \lambda x^2 \mathrm{e}^{-\lambda x}\mathrm{d}x$ $= \lambda \dfrac{2}{\lambda^3} = \dfrac{2}{\lambda^2}$ $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \dfrac{2}{\lambda^2} - \left(\dfrac{1}{\lambda}\right)^2 = \dfrac{1}{\lambda^2}$	M1 Correct integral. A1 c.a.o. (using given result) M1 Correct integral. A1 c.a.o. (using given result) M1 Use of $\mathrm{E}(X^2) - \mathrm{E}(X)^2$ A1	**[6]**
(iii)	$\mu = 6 \quad \therefore \lambda = \dfrac{1}{6}$ $\bar{X} \sim$ (approx) $\mathrm{N}\left(6, \dfrac{6^2}{50}\right)$	B1 Obtained λ from the mean. B1 Normal. B1 Mean. ft c's λ. B1 Variance. ft c's λ.	**[4]**
(iv)	**<u>EITHER</u>** can argue that 7.8 is more than 2 SDs from μ. ($6 + 2\sqrt{0.72} = 7.697$; <u>must</u> refer to SD ($\bar{\mathrm{X}}$), not SD(X)) i.e. outlier. $\Rightarrow$ doubt. **<u>OR</u>** formal significance test: $\dfrac{7.8 - 6}{\sqrt{0.72}} = 2.121$, refer to N(0,1), sig at (eg) 5% $\Rightarrow$ doubt.	M1 A 95% C.I would be (6.1369, 9.4631). M1 A1 M1 M1 Depends on first M, but could imply it. P($\|Z\|$ > 2.121)= 0.0339 A1	**[3]**
		Total	**[18]**

Chief Examiners' Report

In this series, as always, the Principal Examiners' reports have tried to give teachers information to help them to evaluate the work of their students in the context of the strengths and weaknesses of the overall entry.

Some weaknesses are commonly mentioned: poor recognition and use of 'technical' language and notation, failure to present methods or reasons clearly and failure to set out work clearly.

Any candidate who does not know the meaning of technical words or notation given in the specification is at a great disadvantage. This is obviously the case when this lack of knowledge prevents the candidate from completely understanding what is required but also, poor or inaccurate use of technical terms or notation can impair a candidate's attempt to comment on an answer or explain a method.

Almost all solutions should include a clear indication of the method used. The rubric for each paper advises candidates that 'an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used'. Of course, when candidates are asked to establish a *given* answer, the detail required may be much greater that when the answer is not known.

Good, clear (and compact) display of working helps a candidate produce a coherent argument and reduces the chance of 'slips'. Candidates of all levels of ability can benefit from presenting their work and ideas well and there is often an association between good layout and high quality of work. It is to be hoped that the introduction of Printed Answer Books will encourage candidates to consider more carefully their setting out of solutions.

There are three matters that have been raised about how candidates should use the Printed Answer Books (that will be scanned). The first is that they should put their answers in the correct sections; the second is that they should not try to erase writing or drawing but should cross it out – the scanning process is sensitive and copies the faint images and marks that often are left after attempts at erasure. Finally it should be noted that the use of additional answer sheets should be unusual, and that sheets of rough working should not be handed in.

Note on accuracy in Statistics modules

The Principal Examiners' reports that follow discuss the candidates' performances on the individual modules. There is one matter that should be discussed in a general way as it applies to all the statistics modules. This is in respect of arithmetical accuracy in intermediate working and in quotation of final answers. Please note that these remarks are specific to the *statistics* modules; they do not necessarily apply to other modules, where it may be natural for somewhat different criteria to be appropriate.

Most candidates are sensible in their arithmetical work, but there is some unease as to exactly what level of accuracy the examiners are expecting. There is no general answer to this! The standard rubric for all the papers sums the situation up by including "final answers should be given to a degree of accuracy appropriate to the context". Three significant figures may often be the norm for this, but this always needs to be considered in the context of the problem in hand. For example, in quoting from Normal tables, *some* evidence of interpolation is generally expected and so quotation to four decimal places will often be appropriate. But even this does not always apply – quotations of the standard critical points for significance tests such as 1.96, 1.645, 2.576 (maybe even 2.58 – but not 2.57) will commonly suffice.

Talking now in general terms, the examiners always exercise sensible discretion in cases of small variations in the degree of accuracy to which an answer is given. For example, if 3 significant figures are expected (either because of an explicit instruction or because the general context of a problem demands it) but only 2 are given, a candidate is likely to lose an Accuracy mark; but if 4 significant figures are given, there would normally be no penalty. Likewise, answers which are slightly deviant from what is expected in a very minor manner are not penalised (for example, a Normal probability given, after an attempt at interpolation, as 0.6418 whereas 0.6417 was expected). However, there are increasing numbers of cases where candidates give answers which are *grossly* over- or under-specified, such as insistence that the value of a test statistic is (say) 2.128888446667 merely because that is the value that happens to come off the candidate's calculator. **Such gross over-specification indicates a lack of appreciation of the nature of statistical work and, with effect from the January 2011 examinations, will be penalised by withholding of associated Accuracy marks.**

Candidates must however always be aware of the dangers of premature rounding if there are several steps in a calculation. If, say, a final answer is desired that is correct to 3 decimal places, this can in no way be guaranteed if only 3 decimal places are used in intermediate steps; indeed, it may not be safe to carry out the intermediate work even to 4 decimal places. The issue of over-specification may arise for the final answer but not for intermediate stages of the working.

It is worth repeating that most candidates act sensibly in all these respects, but it is hoped that this note may help those who are perhaps a little less confident in how to proceed.

4768 Statistics 3

General comments
There were 428 candidates from 83 centres (June 2009: 371 from 77) for this sitting of the paper. The overall standard of the scripts seen was very pleasing: the increase in the size of the entry seemed to correspond to an improvement in quality. There were many occasions when questions were answered completely, or almost completely, correctly. It was pleasing to note that most candidates remembered to state the hypotheses in Questions 2 and 3 despite not having been instructed to do so.
However, in those parts of Questions 1, 2 and 3 where discussion or comment was called for, very few candidates showed little, if any, statistical insight.
Yet again 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 recent reports on this unit concerning the quality of the language used in the conclusions to hypothesis tests, there remains much room for improvement in this respect.

Invariably all four questions were attempted. With 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) *(Sampling; Combinations of Normal distributions. Salesmen and their expenses.)*

(i) While most, but by no means all, could name “Systematic sampling” correctly, far fewer were able to identify and fix the weakness without resorting to a completely different method, usually simple random sampling, which missed the point.

(ii) It was very rare indeed for anyone to not get this part right.

(iii) The correct probability was obtained here most of the time, but, unsurprisingly, a largish minority of candidates found the variance of $3X$ instead of $X_1 + X_2 + X_3$. One way or another most candidates identified the need for independence, but then far fewer were able to supply a convincing explanation that focused on the independence of one month from another. Seasonal variation was a popular but incorrect suggestion.

(iv) There were the inevitable errors with the variance in this part (usually neglecting to square either or both of the factors 0.45 and 0.1), but on the whole the work seen was impressive. Allowing for these errors, candidates worked through the calculation to a creditable conclusion and many scored full marks.

2) *(The t distribution: test and confidence interval for a population mean. The specific gravity of a brewing mixture.)*

(i) Some candidates gave clear and correct answers to this part but for many others their responses were woolly and vague.

(ii) There were many good answers to this part. However, in many cases there was room for improvement in all aspects of it: stating the hypotheses concisely, accurate calculation of the test statistic from the data and an appropriately worded non-assertive conclusion. In the calculation of the test statistic considerable tolerance was allowed, but candidates are expected to know how to use their calculators efficiently and they should be able to obtain an accurate result.

(iii) Most candidates knew how to calculate the required confidence interval, but as in the past there were too many candidates who, for some reason, chose to switch to the Normal distribution instead of staying with the *t* distribution that they had used for the test in part (ii). While many correct well-rehearsed explanations of the meaning of a confidence interval were seen, there was a fairly widespread lack of understanding of this concept.

3) *(Wilcoxon paired sample test. Effectiveness of a vaccination programme. Chi-squared test of goodness of fit. Benford's Law.)*

(a)(i) Hardly any candidates could explain why the use of paired data was appropriate. A common response was that there were differences between the two sets of data which the pairing would eliminate thus allowing the differences to be investigated. ("Within" instead of "between" would have resolved the matter satisfactorily.)

(ii) The hypotheses here were very poorly expressed. In many cases it seemed that candidates thought that they were testing a mean rather than a median while in many others there was no mention of a parameter at all. Furthermore, references to difference and/or population were often missing.
The calculation of the test statistic, on the other hand, was almost always correct. There were occasional issues with the conduct of the test but, by and large, apart from assertive conclusions, it was done well.

(b) The test statistic was usually obtained correctly and, as above, apart from some assertive conclusions, the test was conducted as required. Unusually, however, there were a few instances of candidates using the critical value chosen from the left-hand tail of the distribution.

4) *(Continuous random variables; the Central Limit Theorem. The lifetime of domestic appliances.)*

(i) In this part some candidates were careful to ensure that their work was complete and convincing, including stating that $e^{-\lambda x} \to 0$ as $x \to \infty$. Others were probably reliant on the stated result to keep them on the straight and narrow.
The quality of the sketches seen was very mixed: axes drawn without a ruler and not labelled; curves that were not drawn carefully and sometimes went into the second quadrant.

(ii) In this part many candidates did not seem to think carefully enough about the relationship between the given result and the integrals needed to find the mean and the variance. The consequence was that a factor of λ was often missing, resulting in a mean of $\frac{1}{\lambda^2}$ and a variance of $\frac{2\lambda-1}{\lambda^4}$, the latter having potentially disastrous consequences later on. Thankfully very few ignored the given result and attempted to integrate by parts. There was some evidence that a few candidates were able to quote the results for the exponential distribution.

(iii) This part was either answered very well or badly. There was a clear split between candidates who realised that this part was all about the Central Limit Theorem and those who appeared unfamiliar with it. This was heightened further by the apparent lack of appreciation that the variance of the sample mean was needed. Many did not realise that, since they were given that $\mu = 6$, they could find λ and hence the variance. Furthermore, with the exception of perhaps just one or two candidates, of those who made the mistake described in part (ii) above none realised that they were ending up with a negative variance.

(iv) Candidates who answered part (iii) correctly usually had little difficulty in this part. The simplest approach was to consider the usual criterion for an outlier and to set out the evidence that 7.8 was indeed an outlier. However a worrying aspect of many attempts was that candidates did not seem to recognise what they were supposed to have written down in part (iii), ie the distribution, including the variance, of the sample mean. Consequently many of the calculations seen were inappropriate. Meanwhile, attempts that did not contain hard evidence were deemed unacceptable.