OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE

MEI STATISTICS

Statistics 3 (Z3)

G243

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 11 June 2009
Morning

Duration: 1 hour 30 minutes

G243

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.

© OCR 2009 [R/103/0484]

Section A (47 marks)

1 For a piece of coursework in psychology, students were asked to analyse reaction times. One student found a piece of software on the internet that measures the time taken to react to a stimulus by pressing a key. She used this with a random sample of 10 friends, recording times, in milliseconds, for their non-dominant and dominant hands. The data were recorded as follows.

Friend	Non-dominant	Dominant
1	485	336
2	356	381
3	450	348
4	402	329
5	376	329
6	409	346
7	419	344
8	289	327
9	420	342
10	410	356

(i) Use a *t* test to investigate whether people have a shorter reaction time, on average, with their dominant hand. Use a 5% level of significance and state your hypotheses clearly. State the required distributional assumption for the test to be valid. **[14]**

(ii) Explain why the student should not measure the times taken with non-dominant and dominant hands in the same order for all the 10 friends. **[2]**

© OCR 2009

2 A nationally known retailer is investigating the effectiveness of its TV advertising. In the North-West, the retailer ran a major TV advertising campaign late in 2008. The sales (£ thousands) for January to March 2009 in a random sample of 10 of its stores in the North-West are as follows, listed in ascending order.

2159 2361 2570 2985 3012 5442 5756 5825 6023 6078

(i) Suggest why it might not be appropriate to compare these data with corresponding sales figures for January to March 2008. **[1]**

From historical data, the retailer believes that sales at its stores in the North-East are generally comparable with those in the North-West. However, there was no advertising campaign in the North-East.

A random sample of 8 stores in the North-East gives the following sales figures (£ thousands) for January to March 2009, listed in ascending order.

1951 2077 2193 2286 2780 2983 4912 5629

(ii) Use a suitable test, at the 5% significance level, to examine whether the median sales figure in the North-West for this period is greater than that in the North-East. What does your conclusion suggest about the effectiveness of the TV advertising campaign? **[11]**

(iii) The retailer has 30 stores in the North-East. Explain how to choose a random sample of 8 of these stores. **[4]**

3 In the course of a medical treatment, random samples of male and female patients have their temperatures, $x\,^{\circ}\mathrm{C}$, recorded. For the 36 males, the mean is $36.8\,^{\circ}\mathrm{C}$ and the standard deviation is $0.247\,^{\circ}\mathrm{C}$. For the 34 females, $\Sigma x = 1252.9$ and $\Sigma x^2 = 46\,172.85$.

(i) Calculate the mean and standard deviation for the females. **[3]**

A doctor uses these data to test whether there is a difference between the population mean temperatures of males and females undergoing this treatment.

(ii) Explain why it is appropriate to use a test based on the Normal distribution in this case. **[1]**

(iii) Carry out the test, at the 1% level of significance, stating your hypotheses. **[11]**

[Question 4 is printed overleaf.]

Section B (25 marks)

4 For employees working in a noisy environment, companies arrange for annual hearing tests in order to monitor whether hearing function decreases with years of service. Data for a random sample of 10 such employees, for one company, are given in the following table.

Employee	A	B	C	D	E	F	G	H	I	J
Years of service	12	3	27	5	2.5	4	8	9	10	14
% Hearing function	90	91	84	97	92	94	88	85	98	89

(i) Draw, on graph paper, a scatter diagram of these data. **[3]**

(ii) The product moment correlation coefficient for these data is $r = -0.5711$. Test at the 5% level of significance whether or not there is negative correlation between years of service and hearing function in the underlying population. State your hypotheses clearly. **[6]**

(iii) Explain why it might not be appropriate to carry out a test based on the product moment correlation coefficient for these data. **[2]**

(iv) Calculate Spearman's coefficient of rank correlation between years of service and hearing function. **[4]**

(v) Conduct a hypothesis test based on this new coefficient, stating your hypotheses clearly. **[4]**

(vi) Comment on the outcomes of the two tests. **[2]**

(vii) In order to focus further research, suggest two other factors that might affect hearing function, apart from years of service. **[2]**

(viii) Another company, operating on six sites, wishes to select a sample of employees for hearing tests, ensuring that each site is represented in the sample. Explain why simple random sampling would not be appropriate. Suggest a sampling method that would be appropriate. **[2]**

OCR
RECOGNISING ACHIEVEMENT

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations, is given to all schools that receive assessment material and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1PB.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2009

G243 Statistics 3

<table>
<tr><td>Q1</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>H_0: μ=0
H_1: μ>0 (or μ<0 if dominant - non dominant)

where μ represents the population mean difference

<table>
<tr><th>non</th><th>dom</th><th>diff</th></tr>
<tr><td>485</td><td>336</td><td>149</td></tr>
<tr><td>356</td><td>381</td><td>-25</td></tr>
<tr><td>450</td><td>348</td><td>102</td></tr>
<tr><td>402</td><td>329</td><td>73</td></tr>
<tr><td>376</td><td>329</td><td>47</td></tr>
<tr><td>409</td><td>346</td><td>63</td></tr>
<tr><td>419</td><td>344</td><td>75</td></tr>
<tr><td>289</td><td>327</td><td>-38</td></tr>
<tr><td>420</td><td>342</td><td>78</td></tr>
<tr><td>410</td><td>356</td><td>54</td></tr>
<tr><td></td><td>total</td><td>578</td></tr>
<tr><td></td><td>mean</td><td>57.8</td></tr>
<tr><td></td><td>sample SD</td><td>55.17</td></tr>
</table>
Test statistic is $t = \dfrac{\bar{d}-0}{s/\sqrt{n}} = \dfrac{57.8}{55.17/\sqrt{10}} = 3.31$

Critical region t_9>1.833

Since 3.31>1.833 H_0 is rejected,
there is sufficient evidence to suggest that reaction times for the dominant hand are faster, on average, than for the non-dominant hand.

Assume Normality;
of (population of) differences</td><td>B1
B1

B1

M1

awrt A1

awrt A1

M1
A1awrt

M1

A1

A1

A1

B1
B1</td><td>Condone absence of “population” if correct notation “μ” has been used, but do NOT accept $\bar{X}=0$ (or $\bar{X}=\bar{Y}$) or similar unless $\bar{X}$ and $\bar{Y}$ are clearly and explicitly stated to be <u>population</u> means. Accept hypothesis explained in words, provided “population” appears.

Follow-through incorrect value of test statistic

fr t_9 (No follow-through from here if wrong)
for 1.833 (No follow-through from here if wrong)</td></tr>
<tr><td>(ii)</td><td>e.g. possible “consistent order” or learning effect which cannot be disentangled in the analysis and should therefore be randomised.</td><td>E2
(E1,E1)</td><td></td></tr>
<tr><td></td><td></td><td>16</td><td></td></tr>
</table>

Q2

(i) There may have been changes in products, or changes in the economy.
Any sensible reason. — B1

(ii)

			1951	1
2159	3		2077	2
2361	6		2193	4
2570	7		2286	5
2985	10		2780	8
3012	11		2983	9
5442	13		4912	12
5756	15		5629	14
5825	16			
6023	17			
6078	18			
	116			55

H_0: Medians for both groups are the same
H_1: Median for the NW group > median for the NE group — B1

Use of joint ranking — M1
All ranks correct — A1
Use of rank sums — M1
Smaller rank sum 55 — A1
Use of $W_{8,10}$ — M1
Critical point 56 — A1
$55 \leq 56$, so reject H_0 — M1 A1
Median for NW does seem to be greater, — A1
so the advertising did seem to have an impact — E1

(iii)
Number the stores 01 to 30 — B1
Choose a random starting position in the table — B1
Select 2 digit numbers (between 01 – 30) and the corresponding stores — B1
Ignore repeats. — B1

Alternative for (ii) using Mann-Whitney method:

		0	1951
2159		0	2077
2361		1	2193
2570		1	2286
2985		3	2780
3012		3	2983
5442		5	4912
5756		6	5629
5825			
6023			
6078			
statistic		19	
MW CV(8,10) 5%	20		

Mark scheme as above.

16

Q3			
(i)	$\frac{\sum x}{n} = \frac{1252.9}{34}$ $\sqrt{\frac{1}{n-1}(\sum x^2 - \frac{(\sum x)^2}{n})}$ $= \sqrt{\frac{1}{33}(46172.85 - \frac{1252.9^2}{34})}$ $= \sqrt{0.105606} = 0.325\, awrt$	B1 M1 A1	 S_{xx} or better
(ii)	Both samples are large or Central Limit Theorem.	B1	
(iii)	$\frac{\bar{x}_m - \bar{x}_f}{\sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}} = \frac{-0.05}{\sqrt{\frac{0.247^2}{36} + \frac{0.325^2}{34}}} = -0.7216$ $= -0.72$ awrt H_0: $\mu_m = \mu_f$ H_1: $\mu_m \neq \mu_f$ μ_m = population mean temperature for males μ_f = population mean temperature for females Critical value +/- 2.5758 ensuring that they compare like with like Since -0.7216 > -2.5758 there is no evidence to suggest rejecting H_0 We can accept that the two samples come from populations which have the same mean.	M1 M1 (m) M1 (f) M1 (all) A1CAO B1 B1 B1 B1 M1 E1	numerator denominator Structure Condone absence of "population" if correct notation "μ" has been used, but do NOT accept $\bar{X} = \bar{Y}$ or similar unless $\bar{X}$ and $\bar{Y}$ are clearly and explicitly stated to be <u>population</u> means. Accept hypothesis explained in words, provided "population" appears. No FT if critical value wrong
		15	

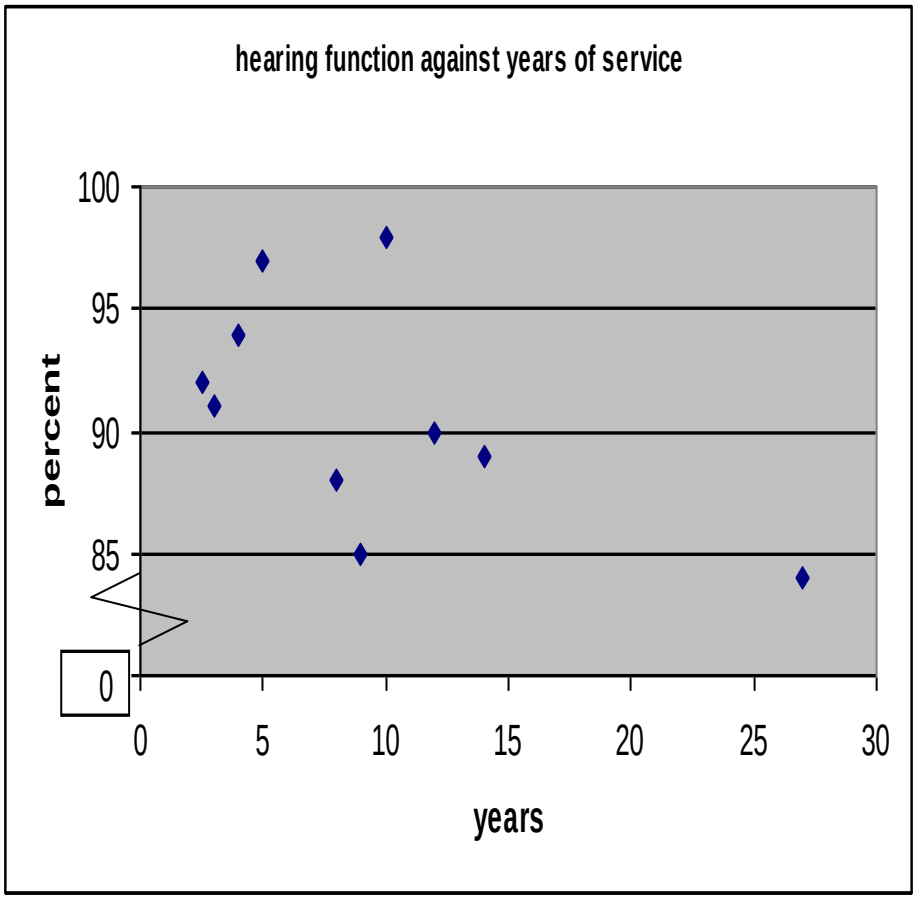
Q4			
(i)	See graph on page 9	G3	G1 for labelled axes G1 for "break" in vertical axis or full linear scale G1 for correct points
(ii)	H_0: $\rho=0$ H_1: $\rho<0$ where ρ is the population coefficient Critical Region < -0.5494 (one tail) Since -0.5711 < -0.5494 reject H_0 Hence this data would show that there is evidence of negative correlation between hearing function and years service.	 B1 B1 B1 M1 A1 E1	 No FT if critical value wrong
(iii)	The outlier (27 years) tends to suggest that these data are not bivariate Normal.	B1 B1	

(iv)

	A	B	C	D	E	F	G	H	I	J	
Years Service, x	12	3	27	5	2.5	4	8	9	10	14	
% Hearing function, y	90	91	84	97	92	94	88	85	98	89	
rank x	8	2	10	4	1	3	5	6	7	9	
rank y	5	6	1	9	7	8	3	2	10	4	
d^2	9	16	81	25	36	25	4	16	9	25	246

An attempt at ranking Complete $\Sigma d^2 = 246$ R= 1 – (6 x 246) / (10 x 99) = 1 – 1.4909 = -0.4909 awrt	M1 A1 B1 A1	
(v)	H_0: there is no association between years of service and hearing function H_1: there is negative association Critical value -0.5636 < -0.4909 There is insufficient evidence to reject H_0 So there would appear to be no association between hearing function and years of service	 B1 B1 M1 E1	 No FT if critical value wrong No marks except initial B1 if $\lvert r_s \rvert > 1$
(vi)	Discussion of different outcomes, different hypotheses, different assumptions	E1, E1	
(vii)	Age, sex, discos, shooting, disease or other sensible (quantifiable) suggestion	E1, E1	
(viii)	Simple random sampling does not guarantee that all the sites are represented Stratified sampling	 E1 E1	
		25	

Graph for question 4 (i)



G243 Statistics 3

Q1 (i) Only a few candidates gave their hypotheses in symbols. The use of the customary statistical notation in terms of μ for a population mean is likely to lead to correct statements of hypotheses, and is to be commended. Nevertheless, candidates may state their hypotheses in words without any symbols at all, but there is an absolute insistence that the word "population" appears. Many candidates went on to perform a fully correct hypothesis test. However some incorrectly performed a two sample t test. Others found the test statistic correctly but then quoted a wrong critical value (usually 2.262), thus losing several marks. Most candidates knew that normality was a required assumption, but none mentioned 'normality of the population of differences'.

(ii) Many candidates were able to explain why the student should not measure times taken in the same order for all 10 friends due to the learning effect which cannot be disentangled in the analysis.

Q2 (i) Many candidates were able to give suitable reasons, such as changes in products or in the economy.
(ii) Many candidates produced fully correct solutions. There were few candidates who did not try to rank the data other than those who thought that a t test was required. Others made an attempt at a paired sample Wilcoxon test, despite the sample sizes being different! A few candidates thought that as the test statistic was less than the critical value, they should accept rather than reject the null hypothesis.
(iii) Very few candidates were able to describe clearly how to use random number tables to choose a random sample of 8 stores. Hardly any mentioned that repeats should be ignored, and none mentioned the need for a random starting position in the table. Putting names into a hat and selecting 8 of them was as popular a method as the use of random number tables.

Q3 (i) Almost all candidates found the mean, but only a small number found the standard deviation correctly, despite this being part of the Specification for Statistics 1 (G241). A number found the variance but forgot to take the square root and others just found the value of S_{xx}.
(ii) Most candidates correctly explained that the sample is large.
(iii) The majority of candidates made a reasonable attempt at the test statistic, but often made errors such as using s instead of s^2 for either males, or females, or both, in calculating the pooled standard deviation. The hypotheses were often correctly given, either in symbols, or in words, and unlike in question 1, most candidates did mention 'population' when giving hypotheses in words. The critical value was often correct, and most candidates who got this far then completed the test correctly.

Q4 (i) Most candidates completed the scatter diagram correctly.
(ii) Only a few candidates gave their hypotheses correctly in terms of ρ, and hardly any actually defined ρ. Candidates were more successful in carrying out the test, although trying to compare -0.5711 with a positive critical value was a fairly common error.
(iii) Many candidates noted either that the scatter diagram was not elliptical, or that there was an outlier. However many also discussed the 'linearity' of the scatter diagram, which is not really relevant to the question.
(iv) Many candidates found Spearman's rank correlation coefficient correctly, although a number did not rank the results.
(v) Most candidates did not define their hypotheses correctly, describing correlation rather than association. A number however went on to correctly complete the test.
(vi) Candidates often correctly mentioned that the two tests had different results, but they rarely discussed the reasons for this.
(vii) Most candidates suggested sensible factors such as age and exposure to loud music.

(viii) Some candidates explained why simple random sampling would not be appropriate and rather more were able to suggest stratified sampling as an appropriate method. Systematic sampling and cluster sampling were popular wrong answers.