MEI

Mathematics in Education and Industry

MEI STRUCTURED MATHEMATICS

STATISTICS 2, S2

Practice Paper S2-B

Additional materials: Answer booklet/paper
Graph paper
MEI Examination formulae and tables (MF2)

TIME 1 hour 30 minutes

INSTRUCTIONS

- Write your Name on each sheet of paper used or the front of the booklet used.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.

INFORMATION

- 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.
- Final answers should be given to a degree of accuracy appropriate to the context.
- The total number of marks for this paper is **72.**
- **You are reminded of the need for clear presentation in your answers.**

1 **(a)** Over many years a college has measured the height of the intake to year 12. The data indicate that the distribution of heights may be taken to be Normal with mean height 175 cm and standard deviation 8 cm.

In September 2004, each of a random sample of 50 students from all those entering year 12 was measured and found to have mean height 177 cm.

The college assumes that this sample comes from a Normal distribution with standard deviation 8 cm. It is desired to test whether this set of data could have come from a Normal distribution with a mean of 175 cm.

(i) State suitable hypotheses for this test. [2]

(ii) Carry out the test at the 5% significance level, stating your conclusions carefully. [6]

(b) A large company provides a cafeteria for lunch. A survey was carried out to see if there is any connection between the type of employee and his or her use of the cafeteria. Employees are identified into two groups. “Office workers” are those who work at desks on managerial, administrative or financial tasks and “Factory workers” are those who work on the production side, including those who move equipment and resources, those who operate the machines, cleaners, etc.

A random sample of 50 was chosen. Of these, 28 were office workers of whom 16 used the cafeteria and 22 were factory workers, of whom 14 used the cafeteria.

This information can be displayed in a 2×2 contingency table as follows.

	Type of worker		
Cafeteria use	**Office**	**Factory**	
Use cafeteria	**16**	**14**	30
Don’t use cafeteria	**12**	**8**	20
	28	22	50

(i) Give a null and alternative hypothesis for a suitable test for independence. [2]

(ii) Calculate the 2 ×2 table of expected frequencies on the hypothesis that there is no such connection. [2]

(iii) Carry out the test at the 5% level of significance, stating your conclusions carefully. [6]

2 A hospital analyst wishes to test how maternal age and birth weight of babies are correlated for the admissions to the hospital where he works.

(i) State appropriate null and alternative hypotheses for the test. Justify the alternative hypothesis you have given. [4]

A random sample of 10 admissions was taken and the birth weight (y kg) was recorded with the age of the mother (x, in completed years) as shown in the table below.

x	28	20	30	35	32	19	32	33	22	27
y	3.25	2.94	2.43	4.05	3.11	3.13	3.46	3.23	4.12	2.5

(ii) Plot a scatter diagram of the data. [2]

(iii) Calculate the product moment correlation coefficient. [6]

(iv) Carry out the hypothesis test at the 5% level of significance. State clearly the conclusion reached. [3]

(v) An analyst in another hospital has carried out the same test on 100 admissions and obtained a correlation coefficient of –0.6568. State, giving a reason, whether the conclusion reached in (iv) is still valid. [2]

3 An experiment was conducted by a researcher to determine the mass y (in grams) of a chemical which dissolved in 100 grams of water at a temperature of x^0C. For each temperature the mass was recorded in the table below.

Temperature (x^0C)	10	20	30	40	50
Mass (y g)	61	65	71	74	80

(i) Represent the data on graph paper. (Take the x axis from 10 to 80 and the y axis from 60 to 90.) [2]

(ii) Calculate the equation of the regression line of y on x. Draw this line on your graph. [6]

(iii) Calculate an estimate of the mass of the chemical that would dissolve in the water at 44^0C. [1]

(iv) Calculate an estimate of the mass of the chemical that would dissolve in the water at 60^0C. Comment on the validity of your answer. [2]

(v) Calculate the residuals for each of the temperatures. Illustrate them on your graph. [4]

(vi) In fact, the researcher carried out further experiments at higher temperatures and obtained the following results.

Temperature (x^0C)	60	70	80
Mass (y g)	82	84	85

Plot these extra points on your graph. Explain whether they support your comment on the validity of your answer in (v) or not. [3]

4 A drug manufacturer claims that a certain drug cures a blood disease on average 80% of the time. To check the claim, an independent tester uses the drug on a random sample of n patients. He decides to accept the claim if k or more patients are cured.

Assume that the manufacturer's claim is true.

(i) State the distribution of X, the number of patients cured. [2]

(ii) Find the probability that the claim will be accepted when 15 individuals are tested and k is set at 10. [5]

A more extensive trial is now undertaken on a random sample of 100 patients.
The distribution of X may now be approximated by a Normal distribution.

(iii) State the parameters of this approximating distribution for X. [1]

(iv) Using this approximating distribution, estimate the probability that the claim will be rejected if k is set at 75. [5]

(v) Find the largest value of k such that the probability of the claim being rejected is no more than 1%. [5]

MEI

Mathematics in Education and Industry

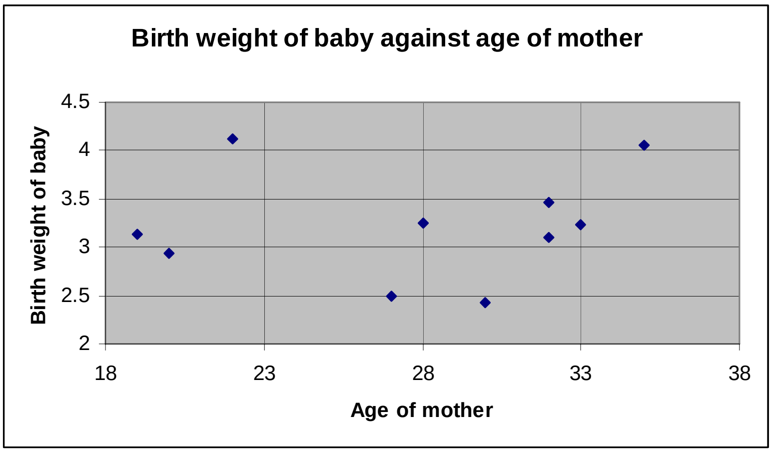
MEI STRUCTURED MATHEMATICS

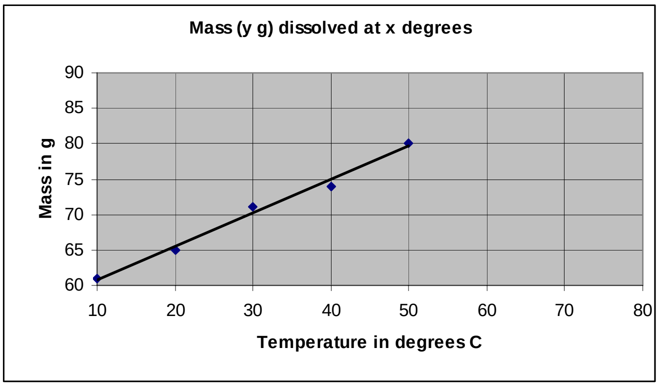
STATISTICS 2, S2

Practice Paper S2-B

MARK SCHEME

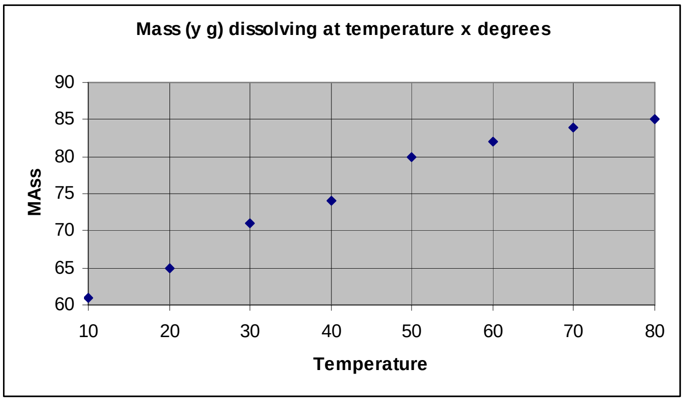
<table>
<tr><th>Qu</th><th></th><th>Answer</th><th>Mark</th><th>Comment</th></tr>
<tr><td>1</td><td>(a)(i)</td><td>H_0: $\mu = 175$
H_1: $\mu \neq 175$</td><td>B1
B1
2</td><td></td></tr>
<tr><td></td><td>(a)(ii)</td><td>Under the null hypothesis, $\bar{X} \sim N\left(175, \frac{8^2}{50}\right)$
Test statistic $= \frac{177-175}{8/\sqrt{50}} = 1.768$
For $N(0,1)$, $P(Z < 0.975) = 1.96$
i.e. 5% critical values are ± 1.96
Since 1.768 < 1.96, this is not significant
So we do not reject the null hypothesis
Conclusion: The data are consistent with a population mean of 175 cm.</td><td>B1
M1
A1
B1
B1
B1
6</td><td></td></tr>
<tr><td></td><td>(b)(i)</td><td>H_0: there is no association between the type of worker and their use of the canteen.
H_1: There is some association between the type of worker and their use of the canteen.</td><td>B1
B1
2</td><td></td></tr>
<tr><td></td><td>(b)(ii)</td><td>Expected
<table>
<tr><th></th><th colspan="2">Type of worker</th><th></th></tr>
<tr><th>Cafeteria use</th><th>Office</th><th>Factory</th><th></th></tr>
<tr><td>Use cafeteria</td><td>16.8</td><td>13.2</td><td>30</td></tr>
<tr><td>Don’t use cafeteria</td><td>11.2</td><td>8.8</td><td>20</td></tr>
<tr><td></td><td>28</td><td>22</td><td>50</td></tr>
</table></td><td>B2
2</td><td>B1 for one error</td></tr>
<tr><td></td><td>(b)
(iii)</td><td>$X^2 = \frac{(16-16.8)^2}{16.8} + \frac{(14-13.2)^2}{13.2} + \frac{(12-11.2)^2}{11.2} + \frac{(8-8.8)^2}{8.8}$
$= 0.03809 + 0.04848 + 0.05714 + 0.07272$
$= 0.21645$
At the 5% significance level, the critical value is 3.841.
Since 0.21645 < 3.841 the test statistic is not significant, so we do not reject the null hypothesis.
Conclusion: The data are consistent with there being no association between the types of work and the use of the cafeteria.</td><td>M1
A1
M1
A1
E1
E1
6</td><td></td></tr>
<tr><td>2</td><td>(i)</td><td>Let the underlying population correlation coefficient between maternal age and birth weight of babies be p.
H_0: $\rho = 0$
H_1: $\rho \neq 0$
We cannot rule out direct inverse correlation, so we should use a two-tailed test.</td><td>B1
B1
B1
B1
4</td><td></td></tr>
</table>

	(ii)		B2	**2**	B1 one point plotted in error
	(iii)	$\sum x = 278 \quad \sum y = 32.22 \quad \sum x^2 = 8020$ $\sum y^2 = 106.6114 \quad \sum xy = 898.89$ $S_{xx} = 8020 - \frac{278^2}{10} = 291.6$ $S_{yy} = 106.6114 - \frac{32.22^2}{10} = 2.79856$ $S_{xy} = 898.89 - \frac{278 \times 32.22}{10} = 3.174$ $\Rightarrow r = \frac{3.174}{\sqrt{291.6 \times 2.79856}} = 0.1111$	B2 B2 M1 A1	**6**	Values of sums (possibly by implication) B1 1 error Values of *S* B1 one error
	(iv)	For a sample of size 10, the critical values are ±0.6319 so the test statistic is not significant, so we do not reject the null hypothesis. Conclusion: The data are consistent with there being no correlation between maternal age and birth weight of baby.	B2 B1	**3**	
	(v)	The value of –0.6568 for a sample of size 100 is very highly significant and very strongly suggests an inverse/negative correlation between maternal age and birth weight of baby. (Note that the tables, page 43 only goes to n = 60 and the value here is 0.2545; for n = 100 this value would be numerically smaller.) This does not invalidate the result from the first hospital, since the populations from which the samples are taken could be different. Also, one or both of the samples might not be representative of the population from which it was taken, just by pure chance in the sampling. However, other things being equal, it makes sense to take more notice of the larger sample.	B1 B1	**2**	

<table>
<tr><td>3</td><td>(i)</td><td>
</td><td>B2

2</td><td>(Ignore line, residuals and extra points in this part

B1 one point plotted in error</td></tr>
<tr><td></td><td>(ii)</td><td>For the data: $\sum x = 150$ $\sum y = 351$ $\sum x^2 = 5500$ $\sum xy = 11000$
$S_{xx} = 5500 - \frac{150^2}{5} = 1000$
$S_{xy} = 11000 - \frac{150 \times 351}{5} = 470$
$\Rightarrow$ Line of regression:
$y - \frac{351}{5} = \frac{470}{1000}\left(x - \frac{150}{5}\right) \Rightarrow y = 0.47x + 56.1$</td><td>B2
B1
M1
A1
B1
6</td><td>B1 if one error

Line drawn</td></tr>
<tr><td></td><td>(iii)</td><td>When $x = 44$, $y = 0.47 \times 44 + 56.1 = 76.78$</td><td>B1
1</td><td></td></tr>
<tr><td></td><td>(iv)</td><td>When $x = 60$, $y = 0.47 \times 60 + 56.1 = 84.3$
This involves extrapolation which is always dangerous!</td><td>B1
B1
2</td><td></td></tr>
<tr><td></td><td>(v)</td><td><table>
<tr><td>x</td><td>10</td><td>20</td><td>30</td><td>40</td><td>50</td></tr>
<tr><td>y</td><td>61</td><td>65</td><td>71</td><td>74</td><td>80</td></tr>
<tr><td>Fitted</td><td>60.8</td><td>65.5</td><td>70.2</td><td>74.9</td><td>79.6</td></tr>
<tr><td>Residual</td><td>0.2</td><td>–0.5</td><td>0.8</td><td>–0.9</td><td>0.4</td></tr>
</table>
On the graph the residuals are vertical lines from the plotted points to the line.</td><td>B2
B2
4</td><td>B1 one error
B1 one error</td></tr>
<tr><td></td><td>(vi)</td><td>

The new points illustrate the dangers of extrapolation as the curve levels off .
Perhaps saturation point has been reached.</td><td>B1
B1
B1
3</td><td>All extra points plotted (a new graph is not required or expected)</td></tr>
</table>

4	**(i)**	X is binomial with $p = 0.8$ i.e. $X \sim \text{Bin}(n, 0.8)$	B1 B1 2	Distribution parameters
	(ii)	When $n = 15$ $X \sim \text{Bin}(15, 0.8)$ When $k = 10$, P(Claim accepted) = $P(X \geq 10)$ = $1 - P(X \leq 9)$ = 1 − 0.0611 (From tables, page 37) = 0.9389	B1 M1 M1 A1 A1 **5**	
	(iii)	Approximating distribution is $X \sim \text{N}(80,16)$	B1 **1**	Both parameters
	(iv)	P(claim rejected) = $P(X < 74.5) = P(Z < z_1)$ where $z_1 = \frac{74.5-80}{4} = -1.375$ = 1 − 0.9155 (from tables, page 44) = 0.0845	M1 B1 M1 A1 A1 **5**	This mark for continuity correction - may be earned in (v) if not here
	(v)	We require $P(X < k - 0.5) < 0.01$ $\Rightarrow P(Z < z_1)$ where $z_1 = \frac{k-0.5-80}{4} < -2.326$ $\Rightarrow k - 80.5 < -9.304$ $\Rightarrow k < 71.196$ $\Rightarrow$ largest value of k is 71	M1 M1 M1 A1 A1 **5**	