MEI

Mathematics in Education and Industry

MEI STRUCTURED MATHEMATICS

STATISTICS 2, S2

Practice Paper S2-A

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 At the end of a management training course, a mental thinking test is given to the trainees. They are assessed for the number of correct answers and also for the time taken to complete the test.

The following results were obtained for a random sample of 12 trainees taken from the whole population of trainees.

Trainee	A	B	C	D	E	F	G	H	I	J	K	L
Number correct	4	6	7	9	10	11	13	14	16	17	18	20
Time taken	112	140	115	130	125	128	135	137	142	150	145	148

(i) Calculate Spearman's rank correlation coefficient for the data. [6]

(ii) Explain what the sign of your correlation coefficient indicates about the data. [1]

(iii) Carry out a test, at the 1% level of significance, of whether or not there is any association between the time taken to complete the test and the number of correct answers. [5]

(iv) The result for one trainee seems out of line with the others, indicating that it might be a "rogue" data point. Identify this trainee. Calculate a new Spearman's rank correlation coefficient eliminating this item and working with the remaining 11 items. [4]

(v) In what circumstances might such an item of data be rejected from a random sample? [2]

2 A student makes occasional spelling errors when typing documents.

(i) Explain why it might be appropriate to model the number of errors in a document by a Poisson distribution. [2]

On average, the student makes one spelling error per 50 words.

(ii) Find the probability that in a document of 200 words

(*A*) there are no errors, [2]

(*B*) there are more than 5 errors. [4]

On one particular day the student types up his coursework task of 2000 words.

(iii) State the number of errors that might be expected in 2000 words. [1]

(iv) Using a suitable approximating distribution, show that there is a probability of just under 0.9 that the number of errors lies between 30 and 50. [9]

3 **(a)** A manager was assessing the tendency of his staff to take time off work when sick. He took a random sample of 40 men and 40 women from all levels of employment in the firm and noted whether each person had been off sick with illness during the last winter.

The results of his survey are shown in the table.

	Took time off work with illness last winter	**Did not take time off work last winter for illness**	
Men	13	27	**40**
Women	21	19	**40**
	34	**46**	**80**

(i) Assuming independence of sex and sickness, calculate the expected frequencies. [2]

(ii) State the null and alternative hypotheses for a suitable test for independence. [2]

(iii) Carry out the test at the 1% level of significance. [6]

(b) In a different survey, the manager took a random sample of 10 employees and noted the number of days they had been off sick over the last 5 years, together with their current salary. He believes that the higher the salary, the fewer number of days are taken off for sickness. He decides to carry out a hypothesis test to see if there is any such association.

The data are given in the table below.

Current salary, x, (to the nearest £1000)	22	35	44	31	30	41	37	33	27	33
Number of days off sick, y	35	28	15	22	30	18	15	20	27	33

Summary statistics for this data set are as follows.

$n = 10$ $\Sigma x = 333$ $\Sigma y = 243$ $\Sigma x^2 = 11\,463$ $\Sigma y^2 = 6385$ $\Sigma xy = 7763$

$S_{xx} = 374.1$ $S_{yy} = 480.1$ $S_{xy} = -328.9$

(i) Calculate the product moment correlation coefficient for the data. [2]

(ii) State appropriate null and alternative hypotheses for the test. Carry out the test at the 5% significance level, stating your conclusions carefully. [6]

4 Joe regularly plays golf. The distances that he hits a golf ball down a fairway with each of three clubs may be modelled by independent Normal distributions. You may assume that the distances are also independent of each other.
With Club 1 the mean distance is 225 metres and standard deviation 23.6 metres.

(i) Find the probability that the distance he hits the ball on any one shot with Club 1 is less than 200 metres. [2]

With Club 2, the mean distance is 140 metres, while 18% of the balls travel less than 115 metres.

(ii) Find the standard deviation of the distribution of distances for Club 2. [3]

With Club 3, the distribution of distances has mean 200 metres and standard deviation 32.4 metres.

(iii) Find the probability that when he hits the ball with Club 3 the distance is within the range 190 metres to 205 metres. [4]

Joe buys Club 4. It is designed to enable him to hit the ball further than with Club 3, but with the same degree of control. It may therefore be assumed that the distribution of distances using Club 4 can also be modelled by a Normal distribution with standard deviation 32.4 metres.

(iv) Joe hits 9 balls with Club 4 and measures the distances, in metres, to the nearest metre. The mean of these distances is 212.6 metres.

Perform a hypothesis test at the 10% significance level to test whether there is evidence that the mean distance with Club 4 is, in fact, greater than 200 metres, the mean distance with Club 3.

State your hypothesis and conclusions carefully. [9]

MEI

Mathematics in Education and Industry

MEI STRUCTURED MATHEMATICS

STATISTICS 2, S2

Practice Paper S2-A

MARK SCHEME

Qu		Answer	Mark	Comment
1	**(i)**	Ranks (1 = low) A B C D E F G H I J K L 1 2 3 4 5 6 7 8 9 10 11 12 1 8 2 5 3 4 6 7 9 12 10 11 d 0 -6 1 -1 2 2 1 1 0 -2 1 1 d^2 0 36 1 1 4 4 1 1 0 4 1 1 $\sum d^2 = 54$ $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 54}{12 \times 143} = 0.811$	 B1 B1 B1 B1 M1 A1 **6**	
	(ii)	The positive sign indicates that those taking longer over the test generally do better	B1 **1**	
	(iii)	H_0 : No association H_1 : There is some association For n = 12, 1% 2-tailed critical value = 0.7273 Since 0.7273 < 0.811, we reject H_0: There is some evidence of association.	B1 M1 A1 M1 E1 **5**	Statement of a hypothesis test
	(iv)	B appears to be the rogue value New ranks A C D E F G H I J K L 1 2 3 4 5 6 7 8 9 10 11 1 2 5 3 4 6 7 8 11 9 10 d 0 0 -2 1 1 0 0 0 -2 1 1 d^2 0 0 4 1 1 0 0 0 4 1 1 $\sum d^2 = 12$ $r_s = 1 - \frac{6 \times 12}{11 \times 120} = 0.945$	B1 B1 B1 M1 A1 **4**	
	(v)	If either it is an error (i.e. written down incorrectly) or if there is some other compelling reason (i.e. trainee did not understand the test)	E1 E1 **2**	One for each point.
2	**(i)**	Occurrences may be assumed to be random and independent and that the errors will occur uniformly over the whole interval.	E1 E1 **2**	One for each point
	(ii) (*A*)	P(0 errors) = $e^{-4} = 0.0183$ (or from tables Page 40)	M1 A1 **2**	
	(ii) (*B*)	From tables Page P(>5 errors) = 1 – P(≤ 5 errors) =1 – 0.7851 = 0.2149	B1 M1 A1 A1 **4**	
	(iii)	$2000 \times 0.02 = 40$	B1 **1**	

	(iv)	Approximating distribution is N(40,40) A continuity correction must be used: from the wording of the question 30 and 50 are **not** included $P(30.5 < X < 49.5) = P(z_1 < Z < z_2)$ where $z_2 = \dfrac{49.5-40}{\sqrt{40}} \approx 1.502$ (Hence $z_1 \approx -1.502$) From Normal Tables $P(Z < 1.502) = 0.9334$ $\Rightarrow P(0 < Z < 1.502) = 0.9334 - 0.5 = 0.4334$ $\Rightarrow P(-1.502 < Z < 1.502) = 0.4334 \times 2 = 0.8668$	B1 M1 A1 A1 B1 M1 A1 M1 A1 **9**	Failure to use continuity correction should lose A marks
3	**(a)(i)**	$\dfrac{34\times 40}{80} = 17$ | $\dfrac{46\times 40}{80} = 23$ $\dfrac{34\times 40}{80} = 17$ | $\dfrac{46\times 40}{80} = 23$	M1 A1 **2**	
	(ii)	H_0: No association of sex with sickness H_1: Association of sex with sickness	B1 B1 **2**	
	(iii)	$X^2 = \dfrac{(13-17)^2}{17} + \dfrac{(27-23)^2}{23} + \dfrac{(21-17)^2}{17} + \dfrac{(19-23)^2}{23}$ $= 3.27$ 1% Critical value for $v = 1$ is 6.635. Since 3.27 < 6.635 accept H_0; i.e. there is no evidence of any association between sex and sickness.	M1 A1 M1 A1 M1 A1 **6**	
	(b)(i)	$r = \dfrac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} = \dfrac{-328.9}{\sqrt{374.1 \times 480.1}} = -0.776$	M1 A1 **2**	
	(ii)	H_0: $\rho = 0$ H_1: $\rho < 0$ where is the population coefficient For n = 10, 5% 1-tailed critical value = 0.5494 Since 0.5494 < 0.776 we reject H_0; There is sufficient evidence at the 5% level to suggest a negative correlation between salary and days off.	B1 B1 B1 M1 A1 E1 **6**	

4	**(i)**	Given X_1 ~N(225, 23.6^2) $P(X < 200) = 1 - P(Z < -1.0593)$ $= 1 - 0.8552 = 0.1448$	M1 A1 **2**	
	(ii)	Given X_2 ~N(140, σ^2) $\Phi^{-1}(0.82) = 0.9154$ $\Rightarrow \sigma = \frac{25}{0.9154} = 27.3$ (3 sig figs)	M1 M1 A1 **3**	
	(iii)	Given X_3 ~N(200, 32.4^2) $P(190 < X < 205) = P(z_1 < Z < z_2)$ where $z_1 = \frac{190-200}{32.4} = -0.3086$ and $z_2 = \frac{205-200}{32.4} = 0.1543$ $\Rightarrow P(190 < X < 205) = P(-0.3086 < Z < 0.1543)$ $= (0.5613 - 0.5) + (0.6212 - 0.5) = 0.1825$	M1 A1 M1 A1 **4**	Standard-isation at least one value of z Dealing with tables
	(iv)	H_0: $\mu = 200$ H_1: $\mu > 200$ where μ is the population mean of distances using Club 4 Test statistic is $\frac{212.6-200}{\frac{32.4}{\sqrt{9}}} = 1.167$ For N(0,1), P(Z < 0.9) = 1.282 so upper single-tailed 10% point is 1.282 Since 1.167 < 1.282, the result is not significant. So accept H_0; i.e. there is no significant evidence to suggest an improved mean.	B1 B1 B1 M1 A1 M1 A1 E1 E1 **9**	