OCR
RECOGNISING ACHIEVEMENT

ADVANCED GCE

MATHEMATICS (MEI) 4767

Statistics 2

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

Friday 18 June 2010
Afternoon

Duration: 1 hour 30 minutes

4767

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 Two celebrities judge a talent contest. Each celebrity gives a score out of 20 to each of a random sample of 8 contestants. The scores, x and y, given by the celebrities to each contestant are shown below.

Contestant	A	B	C	D	E	F	G	H
x	6	17	9	20	13	15	11	14
y	6	13	10	11	9	7	12	15

(i) Calculate the value of Spearman's rank correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 5% significance level to determine whether there is positive association between the scores allocated by the two celebrities. **[6]**

(iii) State the distributional assumption required for a test based on the product moment correlation coefficient. Sketch a scatter diagram of the scores above, and discuss whether it appears that the assumption is likely to be valid. **[5]**

2 A radioactive source is decaying at a mean rate of 3.4 counts per 5 seconds.

(i) State conditions for a Poisson distribution to be a suitable model for the rate of decay of the source. **[2]**

You may assume that a Poisson distribution with a mean rate of 3.4 counts per 5 seconds is a suitable model.

(ii) State the variance of this Poisson distribution. **[1]**

(iii) Find the probability of

(A) exactly 3 counts in a 5-second period,

(B) at least 3 counts in a 5-second period. **[5]**

(iv) Find the probability of exactly 40 counts in a period of 60 seconds. **[3]**

(v) Use a suitable approximating distribution to find the probability of at least 40 counts in a period of 60 seconds. **[5]**

(vi) The background radiation rate also, independently, follows a Poisson distribution and produces a mean count of 1.4 per 5 seconds. Find the probability that the radiation source together with the background radiation give a total count of at least 8 in a 5-second period. **[3]**

3 In a men's cycling time trial, the times are modelled by the random variable X minutes which is Normally distributed with mean 63 and standard deviation 5.2.

(i) Find

(A) $P(X < 65)$,

(B) $P(60 < X < 65)$. **[6]**

(ii) Find the probability that 5 riders selected at random all record times between 60 and 65 minutes. **[2]**

(iii) A competitor aims to be in the fastest 5% of entrants (i.e. those with the lowest times). Find the maximum time that he can take. **[3]**

It is suggested that holding the time trial on a new course may result in lower times. To investigate this, a random sample of 15 competitors is selected. These 15 competitors do the time trial on the new course. The mean time taken by these riders is 61.7 minutes. You may assume that times are Normally distributed and the standard deviation is still 5.2 minutes. A hypothesis test is carried out to investigate whether times on the new course are lower.

(iv) Write down suitable null and alternative hypotheses for the test. Carry out the test at the 5% significance level. **[8]**

4 In a survey a random sample of 63 runners is selected. The category of runner and the type of running are classified as follows.

		Category of runner			Row totals
		Junior	Senior	Veteran	
Type of running	Track	9	8	2	19
	Road	4	8	12	24
	Both	4	10	6	20
Column totals		17	26	20	63

(i) Carry out a test at the 5% significance level to examine whether there is any association between category of runner and the type of running. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. **[12]**

(ii) For each category of runner, comment briefly on how the type of running compares with what would be expected if there were no association. **[6]**

GCE

Mathematics (MEI)

Advanced GCE **4767**

Statistics 2

Mark Scheme for June 2010

Oxford Cambridge and RSA Examinations

Question 1

(i)

x	6	17	9	20	13	15	11	14
y	6	13	10	11	9	7	12	15
Rank x	8	2	7	1	5	3	6	4
Rank y	8	2	5	4	6	7	3	1
d	0	0	2	-3	-1	-4	3	3
d^2	0	0	4	9	1	16	9	9

(i)	(table above)	M1 for attempt at ranking (allow all ranks reversed)	
		M1 for d^2	
	$\Sigma d^2 = 48$	A1 CAO for Σd^2	
	$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 48}{8 \times 63}$	M1 for method for r_s	
	$= 0.429$ (to 3 s.f.) [*allow* 0.43 to 2 s.f.]	A1 f.t. for $\lvert r_s \rvert < 1$ NB No ranking scores zero	**5**
(ii)	H_0: no association between X and Y in the population	B1 for H_0	
	H_1: some positive association between X and Y in the population	B1 for H_1	
		B1 for population SOI	**3**
		NB H_0 H_1 not ito ρ	
	One tail test critical value at 5% level is 0.6429	B1 for ± 0.6429	
	Since 0.429 < 0.6429, there is insufficient evidence to reject H_0,	M1 for sensible comparison with c.v., provided that $\lvert r_s \rvert < 1$	
	i.e. conclude that there is not enough evidence to show positive association between the two judges' scores.	A1 for conclusion in context f.t. their r_s and sensible cv	**3**
(iii)	A bivariate Normal distribution is required.	B1	
	Scatter diagram.	G1 labelled axes G1 correct points	
	Suitable discussion	E1 E1	**5**
		TOTAL	**16**

Question 2

(i)	Counts have a uniform average rate of occurrence All counts are independent	E1 E1	**2**
(ii)	Variance = 3.4	B1	**1**
(iii)	(A) *Either* $P(X = 3) = 0.5584 - 0.3397 = 0.2187$ *Or* $P(X = 3) = e^{-3.4}\frac{3.4^3}{3!} = 0.2186$	M1 for use of tables or calculation A1	**2**
	(B) Using tables: $P(X \geq 3) = 1 - P(X \leq 2)$ $= 1 - 0.3397$ $= 0.6603$	M1 for 1 - $P(X \leq 2)$ M1 correct use of Poisson tables A1	**3**
(iv)	$\lambda = 12 \times 3.4 = 40.8$ $P(X = 40) = e^{-40.8}\frac{40.8^{40}}{40!} = 0.0625$	B1 for mean M1 for calculation A1	**3**
(v)	Mean no. per hour = $12 \times 3.4 = 40.8$ Using Normal approx. to the Poisson, $X \sim N(40.8, 40.8)$ $P(X \geq 40) = P\left(Z > \frac{39.5 - 40.8}{\sqrt{40.8}}\right)$ $= P(Z > -0.2035) = \Phi(0.2035)$ $= 0.5806$	B1 for Normal approx. B1 for correct parameters (SOI) B1 for correct continuity corr. M1 for probability using correct tail A1 CAO (3 s.f.)	**5**
(vi)	Overall mean = 4.8 $P(X \geq 8) = 1 - P(X \leq 7)$ $= 1 - 0.8867 = 0.1133$	B1 for 4.8 M1 A1	**3**
		TOTAL	**19**

Question 3

(i)	(*A*) $P(X < 65) = P\left(Z < \frac{65-63}{5.2}\right)$ $= P(Z < 0.3846)$ $= \Phi(0.3846) = 0.6497$ (*B*) $P(60 < X < 65) = P\left(\frac{60-63}{5.2} < Z < \frac{65-63}{5.2}\right)$ $= P(-0.5769 < Z < 0.3846)$ $= \Phi(0.3846) - (1 - \Phi(0.5769))$ $= 0.6497 - (1 - 0.7181)$ $= 0.3678$	M1 for standardizing M1 for structure A1 CAO (min 3 s.f.), NB When a candidate's answers suggest that (s)he appears to have neglected to use the difference column of the Normal distribution tables penalise the first occurrence only M1 for standardizing both M1 for correct structure A1 CAO 3s.f.	**3** **3**
(ii)	P(All 5 between 60 and 65) $= 0.3678^5 = 0.00673$	M1 A1 FT (min 2sf)	**2**
(iii)	From tables $\Phi^{-1}(0.95) = 1.645$ $\frac{k-63}{5.2} = -1.645$ $x = 63 - 5.2 \times 1.645 = 54.45$ mins	B1 for ±1.645 seen M1 for correct equation in *k* A1 CAO	**3**
(iv)	H_0: $\mu = 63$ minutes; H_1: $\mu < 63$ minutes. Where μ denotes the population mean time on the new course. Test statistic $= \frac{61.7-63}{5.2/\sqrt{15}} = \frac{-1.3}{1.3426}$ $= -0.968$ 5% level 1 tailed critical value of $z = 1.645$ $-0.968 > -1.645$ so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that the new course results in lower times.	B1 for use of 63 B1 for both correct B1 for definition of μ M1 must include √15 A1 B1 for ±1.645 M1 for sensible comparison leading to a conclusion A1 FT for correct conclusion in words in context	**3** **5**
			19

Question 4

<table>
<tr><td>(i)</td><td>H_0: no association between category of runner and type of running;
H_1: some association between category of runner and type of running;</td><td>B1</td><td>1</td></tr>
<tr><td></td><td>

EXPECTED	Junior	Senior	Veteran
Track	5.13	7.84	6.03
Road	6.48	9.90	7.62
Both	5.40	8.25	6.35

</td><td>M1 A2 for expected values (to 2 dp)
(allow A1 for at least one row or column correct)</td><td></td></tr>
<tr><td></td><td>

CONTRIBUTN	Junior	Senior	Veteran
Track	2.9257	0.0032	2.6949
Road	0.9468	0.3663	2.5190
Both	0.3615	0.3694	0.0192

</td><td>M1 for valid attempt at $(O-E)^2/E$
A1 for all correct
NB These M1A1 marks cannot be implied by a correct final value of X^2</td><td></td></tr>
<tr><td></td><td>$X^2 = 10.21$</td><td>M1 for summation
A1 for X^2</td><td>7</td></tr>
<tr><td></td><td>Refer to X_4^2</td><td>B1 for 4 deg of f</td><td></td></tr>
<tr><td></td><td>Critical value at 5% level = 9.488</td><td>B1 CAO for cv</td><td></td></tr>
<tr><td></td><td>Result is significant</td><td>B1 FT their 'sensible' X^2</td><td></td></tr>
<tr><td></td><td>There is evidence to suggest that there is some association between category of runner and type of running.
NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1</td><td>E1 must be consistent with their X^2</td><td>4</td></tr>
<tr><td>(ii)</td><td>• Juniors appear be track runners more often than expected and road less often than expected.</td><td>E1 E1</td><td></td></tr>
<tr><td></td><td>• Seniors tend to be as expected ín all three categories of running.</td><td>E1 E1</td><td></td></tr>
<tr><td></td><td>• Veterans tend to be road runners more than expected and track runners less than expected.</td><td>E1 E1</td><td>6</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>18</td></tr>
</table>