
A-LEVEL

Statistics

Statistics 3 – SS03

Mark scheme

6380
June 2014

Version/Stage: Final

Mark schemes are prepared by the Lead Assessment Writer and considered, together with the relevant questions, by a panel of subject teachers. This mark scheme includes any amendments made at the standardisation events which all associates participate in and is the scheme which was used by them in this examination. The standardisation process ensures that the mark scheme covers the students' responses to questions and that every associate understands and applies it in the same correct way. As preparation for standardisation each associate analyses a number of students' scripts: alternative answers not already covered by the mark scheme are discussed and legislated for. If, after the standardisation process, associates encounter unusual answers which have not been raised they are required to refer these to the Lead Assessment Writer.

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of students' reactions to a particular paper. Assumptions about future mark schemes on the basis of one year's document should be avoided; whilst the guiding principles of assessment remain constant, details will change, depending on the content of a particular examination paper.

Further copies of this Mark Scheme are available from aqa.org.uk

Key to mark scheme abbreviations

M	mark is for method
m or dM	mark is dependent on one or more M marks and is for method
A	mark is dependent on M or m marks and is for accuracy
B	mark is independent of M or m marks and is for method and accuracy
E	mark is for explanation
✓ or ft or F	follow through from previous incorrect result
CAO	correct answer only
CSO	correct solution only
AWFW	anything which falls within
AWRT	anything which rounds to
ACF	any correct form
AG	answer given
SC	special case
OE	or equivalent
A2,1	2 or 1 (or 0) accuracy marks
-x EE	deduct x marks for each error
NMS	no method shown
PI	possibly implied
SCA	substantially correct approach
c	candidate
sf	significant figure(s)
dp	decimal place(s)

No Method Shown

Where the question specifically requires a particular method to be used, we must usually see evidence of use of this method for any marks to be awarded.

Where the answer can be reasonably obtained without showing working and it is very unlikely that the correct answer can be obtained by using an incorrect method, we must award **full marks**. However, the obvious penalty to candidates showing no working is that incorrect answers, however close, earn **no marks**.

Where a question asks the candidate to state or write down a result, no method need be shown for full marks.

Where the permitted calculator has functions which reasonably allow the solution of the question directly, the correct answer without working earns **full marks**, unless it is given to less than the degree of accuracy accepted in the mark scheme, when it gains **no marks**.

Otherwise we require evidence of a correct method for any marks to be awarded.

Q	Solution	Marks	Total	Comments																																			
1a	Spearman's rank correlation coefficient is the appropriate measure of correlation for these data because there are no measured values given.	E1	1	E1 <u>Ranks only</u> for 400m & <u>position only</u> for cross country or ref to <u>orders given</u>																																			
1b	Ranks <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Rank 400m</th> <th>Rank cross country</th> <th> d </th> </tr> </thead> <tbody> <tr> <td>A</td> <td>3</td> <td>6 3</td> <td>3</td> </tr> <tr> <td>B</td> <td>4</td> <td>1 8</td> <td>3</td> </tr> <tr> <td>C</td> <td>7</td> <td>3 6</td> <td>4</td> </tr> <tr> <td>D</td> <td>5</td> <td>2 7</td> <td>3</td> </tr> <tr> <td>E</td> <td>1</td> <td>8 1</td> <td>7</td> </tr> <tr> <td>F</td> <td>2</td> <td>7 2</td> <td>5</td> </tr> <tr> <td>G</td> <td>8</td> <td>4 5</td> <td>4</td> </tr> <tr> <td>H</td> <td>6</td> <td>5 4</td> <td>1</td> </tr> </tbody> </table>			Rank 400m	Rank cross country	d	A	3	6 3	3	B	4	1 8	3	C	7	3 6	4	D	5	2 7	3	E	1	8 1	7	F	2	7 2	5	G	8	4 5	4	H	6	5 4	1	M1 A1
	Rank 400m	Rank cross country	d																																				
A	3	6 3	3																																				
B	4	1 8	3																																				
C	7	3 6	4																																				
D	5	2 7	3																																				
E	1	8 1	7																																				
F	2	7 2	5																																				
G	8	4 5	4																																				
H	6	5 4	1																																				
	$r_s = -0.595$ (3 sig figs)	B3	5	B1 r_s negative B2 $0.590 \leq r_s \leq 0.599$																																			
1c	$H_0: \rho_s = 0$ $H_1: \rho_s \neq 0$ 2 tail 5% test stat $ r_s = 0.595$ critical value = 0.7381 $-0.595 > -0.7381$ so no significant evidence exists to reject H_0	B1 B1 M1		Hypotheses oe Correct abs value for cv 0.738(1) Correct comparison both -ve/ +ve																																			
	This suggests that there is no correlation between rank/ position in 400m races and position in county cross country final race.	E1dep	4	Conclusion correct in context																																			
1d	H_0 accepted in error as H_0 actually untrue Conclusion made that there is no correlation between rank/ position in 400m races and position in county cross country final race when, in reality, there is a correlation between them.	B1 E1		Correct explanation of Type II error In context																																			
			2																																				

<p>1e(i)</p> <p>(ii)</p>	<p>PMCC $r = -0.904$ (3 sf) (from calculator)</p> <p>sc -0.90 allow M1 M1 A0 (or B2) -0.9 sc allow B1</p> <p>PMCC indicates a strong negative correlation between best time taken to run 400m and time taken to run cross country race final. This indicates that we would expect faster 400m runners to be slower at running the cross country race.</p>	<p>B3</p> <p>E1</p>	<p>4</p>	<p>$(-0.905, -0.903)$ or $r =$ $\frac{8671.488 - \frac{434.4 \times 160.07}{8}}{3.17 \times 7.08} = \frac{-20.3}{22.4}$ $= -0.904$ (3 sf) M1 (num), M1(denom), A1</p> <p>Interpretation in context</p>
--	--	---------------------	-----------------	---

Q	Solution	Marks	Total	Comments												
2a	<table border="1"> <thead> <tr> <th>Frequencies</th> <th>AP</th> <th>AV</th> </tr> </thead> <tbody> <tr> <td>Baseball</td> <td>275</td> <td>50</td> </tr> <tr> <td>Basketball</td> <td>475</td> <td>75</td> </tr> <tr> <td>Soccer</td> <td>350</td> <td>25</td> </tr> </tbody> </table>	Frequencies	AP	AV	Baseball	275	50	Basketball	475	75	Soccer	350	25	M1	2	Correct effort at % for 1 frequency(not 25) All correct
	Frequencies	AP	AV													
Baseball	275	50														
Basketball	475	75														
Soccer	350	25														
		A1														
2b	<p>H₀: Coping style is not associated with sport H₁: Coping style is associated with sport 1 tail 1%</p> <table border="1"> <thead> <tr> <th>Expected</th> <th>AP</th> <th>AV</th> </tr> </thead> <tbody> <tr> <td>Baseball</td> <td>286</td> <td>39</td> </tr> <tr> <td>Basketball</td> <td>484</td> <td>66</td> </tr> <tr> <td>Soccer</td> <td>330</td> <td>45</td> </tr> </tbody> </table>	Expected	AP	AV	Baseball	286	39	Basketball	484	66	Soccer	330	45	B1	10	Method for expected frequencies 3 or more correct All correct Can be implied by correct ts
	Expected	AP	AV													
Baseball	286	39														
Basketball	484	66														
Soccer	330	45														
		M1														
2c	$ts = \sum \frac{(O - E)^2}{E}$ $= \frac{11^2}{286} + \frac{11^2}{39} + \dots + \frac{20^2}{330} + \frac{20^2}{45}$ <p>= 15.02</p> <p>cv df = 2 1% cv = 9.21 p = 0.00055</p> <p>ts > 9.21 Reject H₀</p> <p>Sig evidence to suggest that coping strategy is associated with sport involved.</p>	M1	2	Method for ts seen or implied ts correct (14.9 -15.2) for df =2 (can be implied by cv) for cv correct or p=0.00055 Reject H ₀ Conclusion correct in context												
					A1											
2d	<p>Soccer officials are far less likely than expected to use the AV coping style (more likely than expected to use the AP coping style). Baseball officials are far more likely than expected to use the AV coping style (less likely than expected to use the AP coping style).</p> <table border="1"> <thead> <tr> <th>Expected</th> <th>AP</th> <th>AV</th> </tr> </thead> <tbody> <tr> <td>Male</td> <td>24.5</td> <td>5.5</td> </tr> <tr> <td>Female</td> <td>24.5</td> <td>5.5</td> </tr> </tbody> </table>	Expected	AP	AV	Male	24.5	5.5	Female	24.5	5.5	E1	4	In context In context Effort at expected freq seen or implied Yates used correctly – numerator seen correct ft Whole ts method correct 3.9 – 4.2			
	Expected	AP	AV													
Male	24.5	5.5														
Female	24.5	5.5														
		E1														
	$ts = \sum \frac{(O - E - 0.5)^2}{E}$ $= \frac{3^2}{24.5} + \frac{3^2}{5.5} + \frac{3^2}{24.5} + \frac{3^2}{5.5}$ <p>= 4.00</p>	M1 ml A1														

Q	Solution	Marks	Total	Comments																																				
3a	So that any influence of the order of taking drugs does not affect the outcome of the investigation.	B1 E1	2	Reduction of experimental error In context																																				
3b	<p>H_0: Population mean/median hours relief difference = 0</p> <p>H_1: Population mean/median hours relief difference $\neq 0$</p> <p>2 tail test 2 % level</p> <p>Differences $B - A$</p> <table border="1"> <tr> <td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td> </tr> <tr> <td>1.5</td><td>2.1</td><td>0.2</td><td>-0.2</td><td>2.6</td><td>-0.1</td><td>-0.6</td><td>2.5</td><td>2</td><td>1.2</td><td>3</td><td>3.9</td> </tr> </table> <p>Ranks</p> <table border="1"> <tr> <td>6</td><td>8</td><td>2½</td><td>2½</td><td>10</td><td>1</td><td>4</td><td>9</td><td>7</td><td>5</td><td>11</td><td>12</td> </tr> </table> <p>$T_+ = 6 + 8 + 2\frac{1}{2} + 10 + 9 + 7 + 5 + 11 + 12 = 70.5$</p> <p>$T_- = 2\frac{1}{2} + 1 + 4 = 7.5$</p> <p>test stat $T = 7.5$</p> <p>critical value = 10</p> <p>test stat < 10</p> <p>Reject H_0</p> <p>There is significant evidence of a difference between the average number of hours of relief from pain gained using Drug A and Drug B. Allow 1 tail conclusion Drug B better/longer relief</p>	1		2	3	4	5	6	7	8	9	10	11	12	1.5	2.1	0.2	-0.2	2.6	-0.1	-0.6	2.5	2	1.2	3	3.9	6	8	2½	2½	10	1	4	9	7	5	11	12	B1 B1 M1 m1dep m1 dep m1dep A1 B1 m1dep A1dep E1dep	10
1	2	3	4	5	6	7	8	9	10	11	12																													
1.5	2.1	0.2	-0.2	2.6	-0.1	-0.6	2.5	2	1.2	3	3.9																													
6	8	2½	2½	10	1	4	9	7	5	11	12																													
3c	Conclusion based on experiment in which adults self selected to take part. These adults might not be representative of all adult arthritis sufferers	E1	1	Or mention of volunteers/not selected at random																																				

Q	Solution	Marks	Total	Comments																
4	<p>H_0: Samples from identical populations H_1: Samples not from identical populations</p> <p>2 tail 5% sig level</p> <p>Ranks</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 5%;">A</td> <td style="width: 5%;">4</td> <td style="width: 5%;">7</td> <td style="width: 5%;">9</td> <td style="width: 5%;">10</td> <td style="width: 5%;">11</td> <td style="width: 5%;">12</td> <td style="width: 20%;">$T_A = 53$</td> </tr> <tr> <td>B</td> <td>1</td> <td>2</td> <td>3</td> <td>5</td> <td>6</td> <td>8</td> <td>$T_B = 25$</td> </tr> </table> <p>$U_A = 53 - \frac{6 \times 7}{2} = 32$</p> <p>$U_B = 25 - \frac{6 \times 7}{2} = 4$ $ts = 4$</p> <p>$n = 6, m = 6$ $cv = 5$</p> <p>$ts < 5$</p> <p>Significant evidence to reject H_0 and conclude that there is a difference in the average marks in the Statistics module exam for the two schools</p>	A	4	7	9	10	11	12	$T_A = 53$	B	1	2	3	5	6	8	$T_B = 25$	<p>B1</p> <p>M1</p> <p>A1</p> <p>m1 dep</p> <p>A1</p> <p>B1</p> <p>M1</p> <p>A1dep</p> <p>E1dep</p>	<p>9</p>	<p>$H_0: \eta_A = \eta_B$ or ref to pop median $H_1: \eta_A \neq \eta_B$</p> <p>Ranks separated and totalled effort One total correct U method</p> <p>One U correct $cv = 5$ [or 31 upper tail] only</p> <p>Consistent comparison ts/correct tail cv or ts identified & compared with correct tail cv</p> <p>A1 dep ts and cv correct In context dep previous A1</p>
A	4	7	9	10	11	12	$T_A = 53$													
B	1	2	3	5	6	8	$T_B = 25$													

Q	Solution	Marks	Total	Comments
5a(i)	<p>H_0: Managers have no particular preference for either new or old company structure H_1: Managers prefer new company structure</p> <p>1 tail 5%</p> <p>Use of 17+ and 8 or 13 – Use of B(25, 0.5) or B(30, 0.5) $P(X \geq 17) = 0.0539$ or 0.2923 $p > 0.05$ (5%) Or use of cr with probs Accept H_0 No significant evidence to suggest that managers/they prefer new company structure</p>	<p>B1</p> <p>M1</p> <p>M1</p> <p>M1</p> <p>A1dep</p> <p>E1dep</p>	<p>6</p>	<p>$H_0: p = 0.5$ $H_1: p > \text{or} < 0.5$</p> <p>For identifying ts</p> <p>Either correct Bin prob seen Comparison Bin prob and 5%</p> <p>dep correct Bin prob compared 5%</p> <p>In context dep previous A1</p>
(ii)	<p>There are no measurements to use – simply a judgement of prefer or not/no opinion on new structure. Wilcoxon requires symmetrically distributed measurements for preferences not just a prefer/not situation.</p>	<p>E1</p>	<p>1</p>	<p>W S-R can't be used if only preferences given, +/- given W S-R needs values</p>

Q	Solution	Marks	Total	Comments																					
5b(i)	<p>H_0: Samples from identical populations H_1: Samples not from identical populations 5% sig level</p> <p>Ranks</p> <table border="1"> <thead> <tr> <th>Under 40</th> <th>40-55</th> <th>Over 55</th> </tr> </thead> <tbody> <tr> <td>12 6</td> <td>17 1</td> <td>13 5</td> </tr> <tr> <td>10 8</td> <td>16 2</td> <td>11 7</td> </tr> <tr> <td>7 11</td> <td>15 3</td> <td>9 9</td> </tr> <tr> <td>5 13</td> <td>14 4</td> <td>6 12</td> </tr> <tr> <td>4 14</td> <td>8 10</td> <td>3 15</td> </tr> <tr> <td>1 17</td> <td></td> <td>2 16</td> </tr> </tbody> </table> <p>$T_{under40}=39$ 69 $T_{40-55}=70$ 20 $T_{over55}=44$ 64 $n_{under40}=6$ $n_{40-55}=5$ $n_{over40}=6$</p> $\sum_{i=1}^m \frac{T_i^2}{n_i} = \frac{39^2}{6} + \frac{70^2}{5} + \frac{44^2}{6} = 1556.17$ <p>$H = \frac{12}{17 \times 18} \times 1556.17 - (3 \times 18) = 7.03$ Critical value from $\chi_2^2 = 5.991$ $H > 5.991$ Significant evidence to reject H_0</p>	Under 40	40-55	Over 55	12 6	17 1	13 5	10 8	16 2	11 7	7 11	15 3	9 9	5 13	14 4	6 12	4 14	8 10	3 15	1 17		2 16	<p>B1</p> <p>M1</p> <p>m1 A1</p> <p>m1 m1</p> <p>M1 A1</p> <p>B1</p> <p>A1 dep</p>	<p>10</p> <p>2</p>	<p>$H_0: \eta_{U40} = \eta_{40-55} = \eta_{55+}$ or ref to pop medians H_1: at least 2 population medians differ oe Allow 1 pop median is different Allow ref to median occup stress if H_1 includes 'at least 2'</p> <p>Ranks effort</p> <p>m1 dep ranks used</p> <p>Ranks totalled. At least 1 correct</p> <p>Denominators correct Numerators correct and terms added ft</p> <p>H method correct A1 6.9 -7.2 cv correct</p> <p>dep ts and cv correct</p> <p>A difference between at least 2 groups exists B1 Can be implied in (i)</p> <p>Difference identified [40-55 most or under 40 least] in context E1 (full explanation in context gets B1 E1)</p>
Under 40	40-55	Over 55																							
12 6	17 1	13 5																							
10 8	16 2	11 7																							
7 11	15 3	9 9																							
5 13	14 4	6 12																							
4 14	8 10	3 15																							
1 17		2 16																							
5b(ii)	<p>At least 2 groups' average scores differ.</p> <p>The '40-55 years' age group are significantly more stressed than the 'under 40 years' age group.</p>	<p>B1</p> <p>E1</p>																							