



General Certificate of Education

Statistics 6380

SS03 Statistics 3

Mark Scheme

2005 examination – June series

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

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of candidates' 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.

Key to mark scheme and abbreviations used in marking

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	MC	mis-copy
CAO	correct answer only	MR	mis-read
CSO	correct solution only	RA	required accuracy
AWFW	anything which falls within	FW	further work
AWRT	anything which rounds to	ISW	ignore subsequent work
ACF	any correct form	FIW	from incorrect work
AG	answer given	BOD	given benefit of doubt
SC	special case	WR	work replaced by candidate
OE	OE	FB	formulae book
A2,1	2 or 1 (or 0) accuracy marks	NOS	not on scheme
-x EE	deduct x marks for each error	G	graph
NMS	no method shown	c	candidate
PI	possibly implied	sf	significant figure(s)
SCA	substantially correct approach	dp	decimal place(s)

Application of Mark Scheme

No method shown:

Correct answer without working

mark as in scheme

Incorrect answer without working

zero marks unless specified otherwise

More than one method / choice of solution:

2 or more complete attempts, neither/none crossed out

mark both/all fully and award the mean mark rounded down

1 complete and 1 partial attempt, neither crossed out

award credit for the complete solution only

Crossed out work

do not mark unless it has not been replaced

Alternative solution using a correct or partially correct method

award method and accuracy marks as appropriate

SS03

Q	Solution	Marks	Total	Comments																																																						
<p>1(a)</p> <p>reverse d 6 13 2 12 3 10.5 1 10.5 4 9 8 8 7 7 10 6 9 5 11 4 5 3 12 2 13 1</p>	<table border="1"> <tr> <td>Team</td> <td>Bo</td> <td>Mil</td> <td>Ka</td> <td>Se</td> <td>Ca</td> </tr> <tr> <td>x rank</td> <td>8</td> <td>12</td> <td>11</td> <td>13</td> <td>10</td> </tr> <tr> <td>y rank</td> <td>1</td> <td>2</td> <td>3½</td> <td>3½</td> <td>5</td> </tr> <tr> <td>Team</td> <td>Tex</td> <td>Cle</td> <td>NY</td> <td>Mi</td> <td>Oa</td> </tr> <tr> <td>x rank</td> <td>6</td> <td>7</td> <td>4</td> <td>5</td> <td>3</td> </tr> <tr> <td>y rank</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>10</td> </tr> <tr> <td>Team</td> <td>Ch</td> <td>Bal</td> <td>To</td> <td></td> <td></td> </tr> <tr> <td>x rank</td> <td>9</td> <td>2</td> <td>1</td> <td></td> <td></td> </tr> <tr> <td>y rank</td> <td>11</td> <td>12</td> <td>13</td> <td></td> <td></td> </tr> </table> <p>$r_s = -0.787$ (3 sf from calc) sc -0.78/9 no method M1M1A1B1 subtract 1 mark if not negative</p>	Team	Bo	Mil	Ka	Se	Ca	x rank	8	12	11	13	10	y rank	1	2	3½	3½	5	Team	Tex	Cle	NY	Mi	Oa	x rank	6	7	4	5	3	y rank	6	7	8	9	10	Team	Ch	Bal	To			x rank	9	2	1			y rank	11	12	13			<p>M1 M1 A1 B3</p>	<p>6</p>	<p>Attempt at ranks Tied ranks Reverse ranks acceptable Inconsistent ranks gain M1, 1, 1</p> <p>Alternative: $d = 7, 10, 7\frac{1}{2}, 9\frac{1}{2}, 5, 0, 0, 4, 4, 7, 2, 10, 12$ $\sum d^2 = 649\frac{1}{2}$ B1 $r_s = 1 - \frac{6 \times 649.5}{13 \times 168} = -0.784$ M1, A1</p>
Team	Bo	Mil	Ka	Se	Ca																																																					
x rank	8	12	11	13	10																																																					
y rank	1	2	3½	3½	5																																																					
Team	Tex	Cle	NY	Mi	Oa																																																					
x rank	6	7	4	5	3																																																					
y rank	6	7	8	9	10																																																					
Team	Ch	Bal	To																																																							
x rank	9	2	1																																																							
y rank	11	12	13																																																							
<p>(b)</p>	<p>H_0 Rank orders number of home runs and batting averages are independent. H_1 Rank orders of home runs and batting averages are not independent. 2 tail 1% cv = ± 0.6978 test stat $r_s = -0.787$ $r_s < cv$ or $r_s > cv$</p> <p>Reject H_0 Significant evidence at 1% level to suggest an association between rank orders of number of home runs and batting average. [Results suggests that teams that score more home runs tend to have lower batting averages.]</p>	<p>B1 B1 M1 A1 E1</p>	<p>5</p>	<p>Generous</p> <p>For cv – ignore sign (range allowed B0M1) For comparison ts/cv $r_s = -0.787$ or -0.784 can be implied +/- cv/ts comparison M0 A1E1 poss</p> <p>Ft – not if contradictory in context – need not be correct but ft in context</p>																																																						
Total			11																																																							
<p>2</p>	<p>H_0 pop median, $\eta = 42$ H_1 pop median, $\eta > 42$ 1 tail 10% Signs - + + + + - + - + - + + n = 12 ts = 8+ / 4 - Binomial model B (12, 0.5) $P(\geq 8+) = P(\leq 4-) = 0.1938 > 0.10$ for one tail test Accept H_0 . There is insufficient evidence, at the 10% level, to suggest that the median is greater than 42.</p>	<p>B1 M1 A1 M1 M1 A1 E1</p>	<p>7</p>	<p>Clear η required or use of 'population'</p> <p>Signs or differences for 8+ and 4-</p> <p>Binomial model used and probability attempted and seen Comparison of Binomial probability 0.1938 (or awrt 0.19 with 0.10) Alternative – cr identified as {9,10,11,12} with prob 0.0730 used Interpretation in context – need not be correct but ft in context</p>																																																						
Total			7																																																							

SS03 (cont)

Q	Solution	Marks	Total	Comments																																	
3(a)(i)	$H_0 \mu_{\text{difference}} = 0$ $H_1 \mu_{\text{difference}} < 0$ 2 tail 5%	B1		$H_0 \mu_{\text{difference}} = 0$ or $\eta_1 = \eta_2$ $H_1 \mu_{\text{difference}} < 0$ or $\eta_1 < \eta_2$ etc or in words																																	
	<table border="1"> <thead> <tr> <th>Carpet Type</th> <th>Difference</th> <th>Rank</th> </tr> <tr> <td></td> <td>1 – 2</td> <td>- +</td> </tr> </thead> <tbody> <tr><td>A</td><td>- 4</td><td>2</td></tr> <tr><td>B</td><td>- 14</td><td>7</td></tr> <tr><td>C</td><td>- 15</td><td>8</td></tr> <tr><td>D</td><td>- 11</td><td>6</td></tr> <tr><td>E</td><td>- 45</td><td>9</td></tr> <tr><td>F</td><td>- 7</td><td>5</td></tr> <tr><td>G</td><td>+ 6</td><td>4</td></tr> <tr><td>H</td><td>- 5</td><td>3</td></tr> <tr><td>I</td><td>+ 3</td><td>1</td></tr> </tbody> </table>	Carpet Type	Difference	Rank		1 – 2	- +	A	- 4	2	B	- 14	7	C	- 15	8	D	- 11	6	E	- 45	9	F	- 7	5	G	+ 6	4	H	- 5	3	I	+ 3	1	M1		For differences – ignore signs (1 – 2 or 2 – 1)
	Carpet Type	Difference	Rank																																		
		1 – 2	- +																																		
	A	- 4	2																																		
	B	- 14	7																																		
	C	- 15	8																																		
	D	- 11	6																																		
	E	- 45	9																																		
	F	- 7	5																																		
G	+ 6	4																																			
H	- 5	3																																			
I	+ 3	1																																			
		m1		For ranks (1 = smallest diff)																																	
	Rank totals $T_- = 40$ $T_+ = 5$	m1		For total of + / - ranks																																	
	Test stat $T = 5$	A1		(even if ranked incorrectly)																																	
	critical value = 6	B1		For cv (range allowed B0M1)																																	
	$T < cv$																																				
	Reject H_0																																				
	There is significant evidence to suggest that there is a difference in the average fixing times for the two adhesives.	M1 A1 E1	9	For comparison ts/cv must be in context and refer to average																																	
(ii)	Cost of adhesive, size of carpet Long term reliability Method of application etc	E1	1	Disallow ‘carpets the same’ allow ‘surface carpet laid on’ any relevant comment																																	
(b)	By using each adhesive on each type of carpet, experimental error is reduced and any difference in fixing time detected is due to adhesive used.	B1		For idea of reducing experimental error																																	
		E1	2	For clear explanation in context disallow ‘more accurate’ unless in explanation																																	
	Total		12																																		

SS03 (cont)

Q	Solution	Marks	Total	Comments											
4(a)(i)	H ₀ Returns status is independent of when order was placed H ₁ Returns status is not independent of when order was placed 1 tail 5%	B1		Only H ₀ sufficient – must be clear on independence/no association Condone use of return/order in H ₀ even if not worded completely correctly											
	<table border="1"> <thead> <tr> <th></th> <th>Prom</th> <th>Sale</th> <th>Stand</th> </tr> </thead> <tbody> <tr> <th>No items returned</th> <td>49.2</td> <td>32.8</td> <td>82</td> </tr> <tr> <th>Some items returned</th> <td>10.8</td> <td>7.2</td> <td>18</td> </tr> </tbody> </table>		Prom	Sale	Stand	No items returned	49.2	32.8	82	Some items returned	10.8	7.2	18	M1 m1	
	Prom	Sale	Stand												
No items returned	49.2	32.8	82												
Some items returned	10.8	7.2	18												
	$ts = \sum \frac{(O - E)^2}{E}$ $= \frac{5.8^2}{49.2} + \frac{4.8^2}{32.8} + \frac{1^2}{82} + \frac{5.8^2}{10.8} + \frac{4.8^2}{7.2} + \frac{1^2}{18}$ $= 7.77$	m1 A1		For all E correct ft arithmetic error ts sum with correct denominators disallow Yates' correction For ts in range 7.70 - 7.80 awfw											
	cv df = 2 5% cv = 5.991	B1		For cv (4.605, 9.210, 7.378 B0m1)											
	ts > 5.991	m1		For comparison ts/cv											
	Reject H ₀ Sig evidence to suggest returns status is not independent of when order was placed	A1	8	No context required Can imply B1 for H ₀											
(ii)	Orders placed during a Promotion are much more likely to result in no items being returned whereas orders placed during the Sale are far less likely to result in no items being returned.	E1,1	2	E1 if inconsistent/inadequate Or greater likelihood of some returns for orders placed during the Sale etc. Must mention promotion and sale for E2. E1 generous if in context											
	Total		10												

SS03 (cont)

Q	Solution			Marks	Total	Comments
4(b)(i)		First time	Repeat	B1	2	For raw data numbers correct For contingency table headings, data all correct (E values bracketed)
	No items returned	20 (17.64)	43 (45.36)			
	Some items returned	8 (10.36)	29 (26.64)			
(ii)	<p>H₀ Returns status is independent of type of customer H₁ Returns status is not independent of type of customer 1 tail 5%</p>			B1		Only H ₀ sufficient – must be clear on independence/no association Condone use of order/customer in H ₀ even if not worded completely correctly
	$ts = \sum \frac{(O - E - 0.5)^2}{E} =$ $\frac{1.86^2}{17.64} + \frac{1.86^2}{45.36} + \frac{1.86^2}{10.36} + \frac{1.86^2}{26.64}$ $= 0.736$			M1		For E values method
	$cv \quad df = 1 \quad 5\% \quad cv = 3.841$ $ts < 3.84$			M1 m1		For ts for Yates' corr attempt include: O-E-0.5 or (O-E) ² - 0.5
	<p>Accept H₀ No sig evidence to doubt that returns status is independent of type of customer. New customers are no more/less likely to return items.</p>			A1		For ts 0.730 -0.750 awft Condone if not 3 sf
				B1 m1		For cv CAO For comparison ts/cv ft Allow A1 if H ₀ fully correct and stated earlier In context
Total					10	

SS03 (cont)

Q	Solution	Marks	Total	Comments																					
<p>5(a) B1 Only if μ</p>	<p>H₀ Samples from identical populations H₁ Samples not from identical populations 5% sig level</p>	B1 B1		<p>Or H₀ $\eta_A = \eta_B = \eta_C$ H₁ at least two of η_A, η_B, η_C do differ B1,0 if no ‘population’ used H₀ No difference H₁ Difference..... In context used B1,0(B1 generous)</p>																					
	<p>Ranks</p> <table border="1"> <thead> <tr> <th>Fertiliser A</th> <th>Fertiliser B</th> <th>Fertiliser C</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>9</td> <td>3</td> </tr> <tr> <td>2</td> <td>11</td> <td>4</td> </tr> <tr> <td>5</td> <td>13</td> <td>7</td> </tr> <tr> <td>6</td> <td>15</td> <td>8</td> </tr> <tr> <td>10</td> <td>16</td> <td>12</td> </tr> <tr> <td></td> <td>17</td> <td>14</td> </tr> </tbody> </table>	Fertiliser A	Fertiliser B	Fertiliser C	1	9	3	2	11	4	5	13	7	6	15	8	10	16	12		17	14	M1 A2		<p>A1 ≥ 10 A2 all correct</p>
	Fertiliser A	Fertiliser B	Fertiliser C																						
	1	9	3																						
	2	11	4																						
	5	13	7																						
	6	15	8																						
	10	16	12																						
		17	14																						
	<p>$T_A = 24$ $T_B = 81$ $T_C = 48$ $n_A = 5$ $n_B = 6$ $n_C = 6$</p>	m1 A1		<p>Totals Any one correct</p>																					
	$\sum_{i=1}^m \frac{T_i^2}{n_i} = \frac{24^2}{5} + \frac{81^2}{6} + \frac{48^2}{5} = 1592.7$	m1																							
	$H = \frac{12}{17 \times 18} \times 1592.7 - (3 \times 18) = 8.46$	m1 A1		<p>ft for test stat H with previous result substituted AWFW 8.40 – 8.50</p>																					
	<p>Critical value from $\chi^2_2 = 5.99$ H > 5.99</p>	B1 M1		<p>For cv (range allowed B0M1)</p>																					
<p>Sig evidence to reject H₀ and conclude that samples are not from identical populations</p>	A1	13	<p>Can imply B1,B0 at start of solution</p>																						
<p>(b)(i) There is significant evidence that at least two of the median yields (from plants fed with Fertilisers A, B or C) do differ. Medians 26 , 42, 28½</p>	E1 E1	2	<p>Ft for E1 Difference in context Mention of ‘at least two’</p>																						
<p>(ii) It would appear that those plants that were fed Fertiliser B produced a significantly higher yield on average.</p>	B1	1	<p>Identification of B</p>																						
	Total		16																						

SS03 (cont)

Q	Solution	Marks	Total	Comments
6(a)	H ₀ Samples are taken from identical populations H ₁ Samples are not taken from identical populations 2 tail 5%	B1		Or H ₀ $\eta_H = \eta_C$ H ₁ $\eta_H \neq \eta_C$ need both or population average used – need both
	$U_H = 66 - \frac{7 \times 8}{2} = 38$	M1		For attempt at U
	$U_C = 70 - \frac{9 \times 10}{2} = 25$	A1		For either correct
	U = 25 cv = 13 for n = 7, m = 9 5% U > 13 Accept H ₀ No significant evidence of any difference between the two techniques.	B1 B1 M1		For correct U and consistent cv For cv (range allowed B0M1) For comparison ts/cv
		A1		
				7
(b)	A Type II error would be to conclude that H ₀ is true, that is there is no difference between the two techniques, when in fact H ₀ is not true and the techniques do differ	B1		Concept of Type II correct
		E1		In context
			2	
	Total		9	
	Total		75	