Version 1.0:0607

ASSESSMENT and
QUALIFICATIONS
ALLIANCE

General Certificate of Education

Statistics 6380

SS03 Statistics 3

Mark Scheme

2007 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.

Set and published by the Assessment and Qualifications Alliance.

The Assessment and Qualifications Alliance (AQA) is a company limited by guarantee registered in England and Wales (company number 3644723) and a registered charity (registered charity number 1073334).
Registered address: AQA, Devas Street, Manchester M15 6EX *Dr Michael Cresswell Director General*

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	or equivalent	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)

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. However, there are situations in some units where part marks would be appropriate, particularly when similar techniques are involved. Your Principal Examiner will alert you to these and details will be provided on the mark scheme.

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.

SS03

<table>
<tr><th>Q</th><th>Solution</th><th>Marks</th><th>Total</th><th>Comments</th></tr>
<tr><td>1</td><td>H_0 (pop) median/mean diff $\eta_d = 0$
H_1 (pop) median/mean diff $\eta_d \neq 0$
2 tail 5%</td><td>B1</td><td></td><td>Need ‘average’</td></tr>
<tr><td></td><td><table><tr><td>Difference (without – with)</td><td>4</td><td>2</td><td>-5</td><td>0</td><td>4</td></tr><tr><td>Rank</td><td>4½</td><td>2</td><td>-6</td><td>.</td><td>4½</td></tr></table></td><td>M1</td><td></td><td>For differences (+/- signs can be interchanged); ignore signs</td></tr>
<tr><td></td><td><table><tr><td>-6</td><td>-3</td><td>-7</td><td>-9</td><td>1</td></tr><tr><td>-7</td><td>-3</td><td>-8</td><td>-9</td><td>1</td></tr></table></td><td>M1</td><td></td><td>For ranks. Rank $1 = 1$</td></tr>
<tr><td></td><td>$T_+ = 4½ + 2 + 4½ + 1 = 12$
$T_- = 6 + 7 + \ldots. + 9 = 33$</td><td>m1</td><td></td><td>For totals</td></tr>
<tr><td></td><td>test statistic T = 12</td><td>A1</td><td></td><td>For one correct total</td></tr>
<tr><td></td><td>cv = 6 n = 9</td><td>B1</td><td></td><td>For cv = 6</td></tr>
<tr><td></td><td>T > 6</td><td>M1√</td><td></td><td>Comparison cv/ts</td></tr>
<tr><td></td><td>No significant evidence at 5% level to reject H_0. Accept H_0</td><td>A1</td><td></td><td></td></tr>
<tr><td></td><td>There is no significant evidence to suggest that customers rated differently the radio after introducing the new component</td><td>E1</td><td>9</td><td>In context (ft)</td></tr>
<tr><td></td><td>Total</td><td></td><td>9</td><td></td></tr>
<tr><td>2(a)</td><td>From calculator $r = 0.915$ (0.91456)

or $r = \dfrac{2102.57 - \left(\dfrac{135.2 \times 147.9}{10}\right)}{\sqrt{128.976} \times \sqrt{98.269}}$
$= \dfrac{102.962}{11.35 \times 9.913}$
$= 0.915$</td><td>B3</td><td>3</td><td>AWRT
B2 for 0.914 or $0.91 - 0.92$
B1 for 0.9

Alternative:
$n = 10$ $\sum x = 135.2$ $\sum y = 147.9$
$\sum x^2 = 1956.88$ $\sum y^2 = 2285.71$
$\sum xy = 2102.57$ (M1)
sub in formula (m1) (A1)</td></tr>
<tr><td>(b)</td><td>H_0 $\rho = 0$
H_1 $\rho > 0$ 1 tail 1% sig level</td><td>B1</td><td></td><td>Or words</td></tr>
<tr><td></td><td>test statistic $r = 0.915$</td><td></td><td></td><td></td></tr>
<tr><td></td><td>cv = 0.7155 n =10</td><td>B1</td><td></td><td>For cv</td></tr>
<tr><td></td><td>since ts > 0.7155</td><td>M1</td><td></td><td>For comparison ts/cv</td></tr>
<tr><td></td><td>Reject H_0</td><td>A1</td><td></td><td></td></tr>
<tr><td></td><td>Significant evidence at 1% level to suggest a positive linear association between the weight gain of mothers during pregnancy and the weight of their children at 3 years of age</td><td>E1</td><td>5</td><td>In context (ft)</td></tr>
<tr><td></td><td>Total</td><td></td><td>8</td><td></td></tr>
</table>

SS03 (cont)

<table>
<tr><th>Q</th><th>Solution</th><th>Marks</th><th>Total</th><th>Comments</th></tr>
<tr><td>3(a)(i)</td><td>H_0 No association between height at one year old and income at age 50 years
H_1 An association exists between height at one year old and income at age 50 years</td><td>B1</td><td></td><td>H_0 independent
H_1 not independent</td></tr>
<tr><td></td><td>1 tail 5%</td><td></td><td></td><td></td></tr>
<tr><td></td><td><table>
<tr><th></th><th>Under £20,500</th><th>£20,500 and over</th></tr>
<tr><td>Under 75 cm</td><td>8.5</td><td>11.5</td></tr>
<tr><td>75 cm to under 80 cm</td><td>12.75</td><td>17.25</td></tr>
<tr><td>80 cm and over</td><td>12.75</td><td>17.25</td></tr>
</table></td><td>M1
m1</td><td></td><td>E method for 3 correct
For all E correct</td></tr>
<tr><td></td><td>ts = $\sum\frac{(O-E)^2}{E} =$
$\frac{5.5^2}{8.5}+\frac{5.5^2}{11.5}+\frac{0.75^2}{12.75}+\frac{0.75^2}{17.25}+\frac{4.75^2}{12.75}+\frac{4.75^2}{17.25}$</td><td>M1</td><td></td><td>ts sum with correct denominators</td></tr>
<tr><td></td><td>$= 3.56+2.63+0.044+0.033+1.77+1.31$
$= 9.34$</td><td>A1</td><td></td><td>For ts in range 9.10 – 9.50</td></tr>
<tr><td></td><td>df = 2 5% cv = 5.991</td><td>B1</td><td></td><td>For cv</td></tr>
<tr><td></td><td>ts > 5.991</td><td>m1√</td><td></td><td>For comparison ts/cv</td></tr>
<tr><td></td><td>Reject H_0
Significant evidence to suggest an association exists between height at one year old and income at age 50 years</td><td>A1</td><td>8</td><td>For reject H_0</td></tr>
<tr><td>(ii)</td><td>Those babies with a low height, under 75 cm, at age one year appear more likely to achieve a lower income at age 50 years and those babies with heights 80 cm and over at age one year appear more likely to achieve an income of £20,500 and over</td><td>E1
E1</td><td>2</td><td>Must have attempted χ^2
Indication of sources of association in context</td></tr>
</table>

SS03 (cont)

Q	Solution	Marks	Total	Comments
3(b)	H_0 Presence of parent is **independent** of occurrence of emergency medical treatment H_1 Presence of parent is **not independent** of occurrence of emergency medical treatment 1 tail 5%	B1		Or as in (a)
	<table><tr><th></th><th>Travelled</th><th>Did not travel</th></tr><tr><td>**Required**</td><td>12.490</td><td>5.510</td></tr><tr><td>**Not required**</td><td>89.510</td><td>39.490</td></tr></table>	M1		For E values method
	$\text{ts} = \sum \frac{(\lvert O-E\rvert - 0.5)^2}{E}$	M1		For ts
	$= \frac{1.990^2}{12.490} + \frac{1.990^2}{5.510} + \frac{1.990^2}{89.510} + \frac{1.990^2}{39.490}$	m1		For Yates' corr
	$= 0.317 + 0.719 + 0.044 + 0.100$ $= 1.18$	A1		For ts 1.00 – 1.30
	df = 1 5% cv =3.84	B1		For cv
	ts < 3.84	M1√		For comparison ts/cv
	Accept H_0 No significant evidence to doubt that presence of parent is independent of occurrence of emergency medical treatment	A1	8	In context
	Total		**18**	

SS03 (cont)

<table>
<tr><th>Q</th><th>Solution</th><th>Marks</th><th>Total</th><th>Comments</th></tr>
<tr><td>4(a)</td><td>H_0 (pop) median $\eta = 14$
H_1 (pop) median $\eta > 14$ } average

2 tail 10%</td><td>B1</td><td></td><td>Not mean</td></tr>
<tr><td></td><td>Signs: − + + + + + + − +</td><td>M1</td><td></td><td>Signs
SC2: Wilcoxon signed-rank</td></tr>
<tr><td></td><td>$n = 9$
test stat $= 7^+ / 2^-$</td><td>A1</td><td></td><td>test stat correct</td></tr>
<tr><td></td><td>Model B(9, 0.5)</td><td>M1</td><td></td><td>Bin model seen to be used
($n = 9$, $p = 0.5$ column)</td></tr>
<tr><td></td><td>$P(\le 2^-) = P(\ge 7^+) = 0.0898 < 0.10$</td><td>M1</td><td></td><td>Comparison of correct B(9, 0.5) probability with 0.05 or 0.10
Or use of identified cv
cr $[7, 8, 9]$ see 0.0898</td></tr>
<tr><td></td><td>Reject H_0
Significant evidence at 10% level to doubt H_0</td><td>A1</td><td></td><td></td></tr>
<tr><td></td><td>There is significant evidence to suggest that the median cocaine use has increased since 2000</td><td>E1</td><td>7</td><td></td></tr>
<tr><td>(b)(i)</td><td><table><tr><th>Town</th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th></tr><tr><td>x rank</td><td>9</td><td>1</td><td>5</td><td>4</td><td>2</td></tr><tr><td>y rank</td><td>1</td><td>8</td><td>5</td><td>6</td><td>9</td></tr></table></td><td>M1
M1</td><td></td><td>Attempt at ranks
x or y ranks correct
(reverse order OK)</td></tr>
<tr><td></td><td><table><tr><th>Town</th><th>F</th><th>G</th><th>H</th><th>I</th><th></th></tr><tr><td>x rank</td><td>6</td><td>3</td><td>8</td><td>7</td><td></td></tr><tr><td>y rank</td><td>3</td><td>7</td><td>2</td><td>4</td><td></td></tr></table></td><td>A1</td><td></td><td>All correct, consistent
Rank all as one M1M1 only</td></tr>
<tr><td></td><td>$r_s = -0.967$ (3 sf from calc)

(-0.966 to -0.967 B3)
(-0.96 to -0.97 B2)</td><td>B3</td><td>6</td><td>Alternative:
$d = 8, 7, 0, 2, 7, 3, 4, 6, 3$
$\sum d^2 = 236$ M1
$r_s = 1 - \frac{6 \times 236}{9 \times 80} = -0.967$ m1A1

SC -0.96 to -0.97 M1M1A1A1
SC $+0.967$ 4
SC 0.96 to 0.97 3</td></tr>
<tr><td>(ii)</td><td>The estimated cocaine use in a town is higher when it is easier to buy cocaine in the town</td><td>E1</td><td>1</td><td>Must have some sensible answer in (i); comparative required</td></tr>
<tr><td></td><td>Total</td><td></td><td>14</td><td></td></tr>
</table>

SS03 (cont)

<table>
<tr><th>Q</th><th>Solution</th><th>Marks</th><th>Total</th><th>Comments</th></tr>
<tr><td>5(a)</td><td>H_0 Samples are taken from identical populations
H_1 Samples are not taken from identical populations – population average scores differ</td><td>B1</td><td></td><td>Hypotheses referring to population averages also acceptable or fully explained in words</td></tr>
<tr><td></td><td><table>
<tr><th>Morning</th><th>Afternoon</th></tr>
<tr><td>44</td><td>46</td></tr>
<tr><td>53</td><td>47</td></tr>
<tr><td>54</td><td>51</td></tr>
<tr><td>56</td><td>58</td></tr>
<tr><td>63</td><td>59</td></tr>
<tr><td>63</td><td>61</td></tr>
<tr><td>65</td><td>62</td></tr>
<tr><td>72</td><td>67</td></tr>
<tr><td>74</td><td>68</td></tr>
<tr><td>81</td><td></td></tr>
</table></td><td>M1</td><td></td><td>Separation of am/pm</td></tr>
<tr><td></td><td>Ranks:
Afternoon 2 3 4 8 9 10 11 15 16
Morning 1 5 6 7 12½ 12½ 14 17 18 19</td><td>M1
A1</td><td></td><td>Or reversed</td></tr>
<tr><td></td><td>$T_A = 2 + 3 + \ldots + 16 = 78$
$T_M = 1 + 5 + \ldots + 19 = 112$</td><td>m1ft</td><td></td><td>102
88 Or alt method directly to U</td></tr>
<tr><td></td><td>$U_A = 78 - \frac{9 \times 10}{2} = 33$
$U_M = 112 - \frac{10 \times 11}{2} = 57$</td><td>m1</td><td></td><td></td></tr>
<tr><td></td><td>Test stat U = 33</td><td>A1</td><td></td><td>$U = 33$ or $U = 57$</td></tr>
<tr><td></td><td>cv = 21 n = 9 , m = 10</td><td>B1</td><td></td><td></td></tr>
<tr><td></td><td>$U = 33 > 21$</td><td>M1</td><td></td><td>Comparison U/cv; not if $U < 0$</td></tr>
<tr><td></td><td>Accept H_0
No significant evidence at the 5% level to suggest that there is any difference in average test scores between students taking the test in a morning or afternoon session</td><td>E1</td><td>10</td><td>In context</td></tr>
<tr><td>(b)(i)</td><td>In matched pairs design, individual differences are minimised since the same person is tested each time and therefore any difference which may exist between the two groups is more likely to be identified</td><td>B1
B1</td><td>2</td><td>Reduce experimental error; avoid bias</td></tr>
<tr><td>(ii)</td><td>She kept the students apart during the day of the test</td><td>E1</td><td></td><td>Any 2</td></tr>
<tr><td></td><td>She chose students of similar ability and randomly assigned them to a morning or an afternoon session</td><td>E1</td><td>2</td><td></td></tr>
<tr><td></td><td>Total</td><td></td><td>14</td><td></td></tr>
</table>

SS03 (cont)

<table>
<tr><th>Q</th><th>Solution</th><th>Marks</th><th>Total</th><th>Comments</th></tr>
<tr><td>6</td><td>H_0 Samples are taken from identical populations</td><td>B1</td><td></td><td>or
H_0 $\eta_{Happy} = \eta_{Unhappy} = \eta_{Unmarried}$</td></tr>
<tr><td></td><td>H_1 Samples are not taken from identical populations – population average protective antibody levels differ</td><td>B1</td><td></td><td>H_1 at least two of $\eta_{Happy}, \eta_{Unhappy}, \eta_{Unmarried}$ do differ
Allow mean
B1 H_0 antibody independent of marital status
H_1 antibody not independent of marital status</td></tr>
<tr><td></td><td>Ranks:
<table>
<tr><th>Happily Married</th><th>Unhappily Married</th><th>Unmarried</th></tr>
<tr><td>7</td><td>5</td><td>1</td></tr>
<tr><td>10</td><td>6</td><td>2</td></tr>
<tr><td>12</td><td>9</td><td>3</td></tr>
<tr><td>13</td><td>11</td><td>4</td></tr>
<tr><td>14</td><td></td><td>8</td></tr>
<tr><td>15</td><td></td><td></td></tr>
</table></td><td>M1
A1</td><td></td><td>For 10</td></tr>
<tr><td></td><td>$T_{Hap} = 71$ $T_{Unhap} = 31$ $T_{Unmarr} = 18$
$n_{Hap} = 6$ $n_{Unhap} = 4$ $n_{Unmarr} = 5$</td><td>m1</td><td></td><td>Totals</td></tr>
<tr><td></td><td>$\sum_{i=1}^{m} \frac{T_i^2}{n_i} = \frac{71^2}{6} + \frac{31^2}{4} + \frac{18^2}{5} = 1145.22$
$H = \frac{12}{15 \times 16} \times 1145.22 - (3 \times 16)$</td><td>m1</td><td></td><td>Correct method
test stat:
$H = \frac{12}{N(N+1)} \sum_{i=1}^{m} \frac{T_i^2}{n_i} - 3(N+1)$</td></tr>
<tr><td></td><td>$= 9.26$</td><td>A1</td><td></td><td>$9.1 - 9.4$</td></tr>
<tr><td></td><td>Critical value from $\chi_2^2 = 5.991$</td><td>B1</td><td></td><td></td></tr>
<tr><td></td><td>$H > 5.991$</td><td>M1</td><td></td><td></td></tr>
<tr><td></td><td>Sig evidence to **reject H_0** and conclude that samples are not from identical populations</td><td>A1</td><td></td><td></td></tr>
<tr><td></td><td>Significant evidence at the 5% level to suggest that the population average level of protective antibodies differs for the three marital categories: at least two of the averages differ</td><td>E1</td><td></td><td>Difference in context
Mention of ‘at least two’ or happily married/unmarried differ</td></tr>
<tr><td></td><td>It appears that those males who are happily married have a significantly higher level of antibodies than those who are unmarried</td><td>E1</td><td>12</td><td></td></tr>
<tr><td></td><td>**Total**</td><td></td><td>**12**</td><td></td></tr>
<tr><td></td><td>**TOTAL**</td><td></td><td>**75**</td><td></td></tr>
</table>