Version 1.0

General Certificate of Education (A-level)
June 2013

Statistics SS03

(Specification 6380)

Statistics 3

Final

Mark Scheme

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 events which all examiners 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 examiner understands and applies it in the same correct way. As preparation for standardisation each examiner 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, examiners encounter unusual answers which have not been raised 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 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

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.

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
1(a)	H_0 : **Population** median time = 32 H_1 : **Population** median time < 32 1 tail test 10% level	B1		must mention population
	signs + – – – – + + + – – – –	M1		for signs or signed differences
	test stat = 8 –/ 4+	A1		for test stat 8 or 4
	Bin (12, 0.5) model	M1		for use of Bin model
	$P(\leq 4+) = 0.194 > 0.10$	M1		for comparison ts, 0.193-0.194, and 10%
	Accept H_0. No significant evidence to suggest median time to complete crossword has decreased.	A1	6	ts/cv correct Alt method Use of cr {0, 1 ,2 , 3} or {9, 10, 11, 12} with prob 0.073 used.
	Total		**6**	

<table>
<tr><th>Q</th><th>Solution</th><th>Marks</th><th>Total</th><th>Comments</th></tr>
<tr><td>2(a)</td><td>H_0 : Population average difference = 0
H_1 : Population average difference $\neq 0$
2 tail test 5 % level</td><td>B1</td><td></td><td>May refer to mean/median</td></tr>
<tr><td></td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th></tr>
<tr><td>diff</td><td>+3.3</td><td>.</td><td>–0.8</td><td>–0.9</td><td>+0.3</td></tr>
<tr><td>rank</td><td>9</td><td></td><td>5</td><td>6</td><td>2</td></tr>
<tr><th></th><th>F</th><th>G</th><th>H</th><th>I</th><th>J</th></tr>
<tr><td>diff</td><td>+0.7</td><td>+0.4</td><td>+1.7</td><td>–0.1</td><td>+1.1</td></tr>
<tr><td>rank</td><td>4</td><td>3</td><td>8</td><td>1</td><td>7</td></tr>
</table>
</td><td>M1
m1

m1
A1</td><td></td><td>For differences
Ranks: smallest rank 1

Total of ranks (any)
One correct</td></tr>
<tr><td></td><td>$T_+ = 9 + 2 + 4 + 3 + 8 + 7 = 33$
$T_- = 5 + 6 + 1 = 12$
test stat $T = 12$
critical value = 6</td><td>B1</td><td></td><td>For cv</td></tr>
<tr><td></td><td>test stat > 6</td><td>m1</td><td></td><td>ft (must be positive ts)
‘correct’ T with cv comparison
(smaller T / smaller cv
larger T / larger cv)</td></tr>
<tr><td></td><td>Accept H_0</td><td>A1</td><td></td><td>ts/cv correct</td></tr>
<tr><td></td><td>There is no significant evidence of a difference in average percentage of total expenditure spent on ‘Highways’ between 2002 and 2012.</td><td>E1</td><td>9</td><td>In context – only if conclusion correct</td></tr>
<tr><td>(b)(i)</td><td><u>Differences</u> are symmetrically distributed.</td><td>B1</td><td></td><td></td></tr>
<tr><td>(ii)</td><td>A <u>paired</u> sign test</td><td>B1</td><td>2</td><td></td></tr>
<tr><td>(c)(i)</td><td>0</td><td>B1</td><td></td><td></td></tr>
<tr><td>(ii)</td><td>$1+2+3+4+5+6+7+8+9+10=55$</td><td>M1A1</td><td>3</td><td></td></tr>
<tr><td></td><td>Total</td><td></td><td>14</td><td></td></tr>
</table>

<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 : Samples are taken from identical populations

H_1 : Samples are not taken from identical populations (average emissions lower when device fitted)
1 tail 5%

Ranks
<table>
<tr><td colspan="2">New device fitted</td><td colspan="2">Device not fitted</td></tr>
<tr><td>139.1</td><td>5 8</td><td>145.4</td><td>12 1</td></tr>
<tr><td>134.6</td><td>3 10</td><td>144.0</td><td>11 2</td></tr>
<tr><td>128.9</td><td>1 12</td><td>138.7</td><td>4 9</td></tr>
<tr><td>139.8</td><td>8 5</td><td>139.7</td><td>7 6</td></tr>
<tr><td>129.5</td><td>2 11</td><td>139.6</td><td>6 7</td></tr>
<tr><td>140.9</td><td>10 3</td><td>140.5</td><td>9 4</td></tr>
</table>
$T_{fitted} = 29$ 49 $T_{not} = 49$ 29
$n_{fitted} = 6$ $n_{not} = 6$

$U_{fitted} = 29 - \frac{6\times7}{2} = 8$
$U_{not} = 49 - \frac{6\times7}{2} = 28$

$U = 8$

cv = 7 for n= 6, m = 6 1 tail 5%

$U > 7$
(or cv=29 and 28 < 29)

Accept H_0

No significant evidence of a reduction in average CO_2 emissions for cars fitted with new device.</td><td>B1

M1

A1

m1

m1
A1

B1

M1

A1

E1</td><td>10</td><td>Or equivalent referring to <u>population</u> medians

Ranks as one group either way

Ranks correct

Ranks totalled or reversed

Attempt to find U
Either U correct

cv correct cao

correct comparison ft –
upper ts /29
lower ts /7

ts/cv correct

In context – only if conclusion correct</td></tr>
<tr><td>(ii)</td><td>The 12 new cars can be regarded as a random sample.
OR They were randomly selected.</td><td>B1</td><td>1</td><td>Disallow ‘random’ no context</td></tr>
<tr><td>(b)</td><td>A Type II error would be to conclude that H_0 is true, that is there is no reduction in average CO_2 emissions for cars fitted with new device, when in fact H_0 is not true and there is a reduction in average CO_2 emissions for cars fitted with new device</td><td>B1

E1</td><td>2</td><td>Type II correctly explained

In context</td></tr>
<tr><td></td><td>Total</td><td></td><td>13</td><td></td></tr>
</table>

Q	Solution	Marks	Total	Comments
4(a)				
		B1		One total correct 13/12
		M1		One frequency correct
		A1	3	All Correct
(b)	H_0 : Mouthwash type not associated with category of mouth infection H_1 : Mouthwash type is associated with category of mouth infection 1 tail 5%	B1		H_0 Indep/No Assoc H_1 Not Indep/ Assoc
		M1		Method for expected freqs
		A1		
	ts = $\sum \frac{(\lvert O-E\rvert - 0.5)^2}{E}$	M1		ts 'correct' with/without Yates
	$= \frac{2.26^2}{6.24} + \frac{2.26^2}{5.76} + \frac{2.26^2}{6.76} + \frac{2.26^2}{6.24}$	M1		Yates used correctly
	= 3.28	A1		ts correct (3.1 – 3.4)
	cv df = 1 5% cv = 3.84	B1		cv cao
	ts < 3.84 Accept H_0			No Yates used can gain M1 A1 M1 B1 ts = 4.89
	No significant evidence to suggest that mouthwash is associated with category of mouth infection	E1	8	Conclusion correct in contest and ts/cv correct
	Total		**11**	

	Minor	**Serious**	**total**
New	**9**	**3**	12
Current	**4**	**9**	**13**
Total	13	**12**	25

Expected	**Minor**	**Serious**
New	**6.24**	**5.76**
Current	**6.76**	**6.24**

Q	Solution	Marks	Total	Comments
5(a)	H_0 : Samples from identical populations	B1		
	H_1 : Samples not from identical populations 5% sig level			
	Ranks **I**: 16 1, 15 2, 14 3, 12 5, 10 7, 8 9 **II**: 13 4, 9 8, 6 11, 4 13, 1 16 **III**: 11 6, 7 10, 5 12, 3 14, 2 15	M1 A1		For ranks as **one** group 10 or more correct
	Totals of ranks $T_I = 75$ 27 $T_{II} = 33$ 52 $T_{III} = 28$ 57 $n_I = 6$ $n_{II} = 5$ $n_{III} = 5$	m1		Totals – can be 27 52 57
	$\sum_{i=1}^{m} \frac{T_i^2}{n_i} = \frac{75^2}{6} + \frac{33^2}{5} + \frac{28^2}{5} = 1312.1$	m1 m1		Numerators correct Denominators correct
	$H = \frac{12}{16\times17} \times 1312.1 - (3 \times 17)$	m1		H formula correctly used ft $\sum_{i=1}^{m} \frac{T_i^2}{n_i}$
	$= 6.89$	A1		(6.7 – 7.1) ts/cv correct
	Critical value from $\chi^2_2 = 5.991$ $H > 5.991$	B1		For cv cao
	Reject H_0. Significant evidence to suggest that samples are not from identical populations. Significant difference in average score for **at least 2** of the three methods involved.	E1	10	Conclusion correct in context
(b)	Use Approach I since average of ranks is highest (lowest ft) (so lowest average scores)	B1 E1	2	Approach I Reason (vice versa for reversed ranks). Allow reference to average scores.
	Total		**12**	

Q	Solution	Marks	Total	Comments
6(a)(i)	H_0 : Colour preference is independent of personality H_1 : Colour preference is not independent of personality 1 tail 5%	B1		H_0 Indep / No Assoc H_1 Not Indep / Assoc
	Exp | **R** | **Y** | **G** | **B** **Intro** | 47 | 9.4 | 18.8 | 18.8 **Extro** | 153 | 30.6 | 61.2 | 61.2	M1 M1 A1		Any one E_i correct At least 5 correct All correct SC1 integers
	ts = $\sum \frac{(O-E)^2}{E}$			
	$= \frac{11^2}{47} + \frac{1.4^2}{9.4} + \frac{5.2^2}{18.8} + \frac{7.2^2}{18.8} + \frac{11^2}{153} + \frac{1.4^2}{30.6} + \frac{5.2^2}{61.2} + \frac{7.2^2}{61.2}$	M1 M1		Numerators OK ft Denominators OK ft and added
	= 9.12	A1		(9.0-9.3)
	df = 3 5% cv = 7.815 ts > 7.815	B1		For cv or = 0.0277
	Reject H_0 Sig evidence to suggest colour preference is not independent of personality	E1	9	In context
(ii)	Introverts <u>far more likely than expected</u> to prefer blue or green (introverts far less likely than expected to choose red)	B1 E1	2	Alt Extroverts are <u>more likely than expected</u> to prefer red

Q	Solution	Marks	Total	Comments
6(b)(i)	d 0 0 1.5 0 0 1 1 1 2.5 $\sum d^2 = 11.5$	M1		Differences
	SRCC $r_s = 1 - \dfrac{6\times\sum d^2}{9\times 80} = 0.904$	M1 A1		Formula correct
	or SRCC $r_s = 0.904$ (from calc)	(B3)	3	SC1 0.9 SC2 0.90 if no method shown
(ii)	H_0 : Rank orders of personality score and happiness score are independent. H_1 : Rank orders of personality score and happiness score are not independent. 2 tail 1%	B1		Hypothesis
	cv = 0.8167 test stat r_s = 0.904	B1		cv cao
	r_s > cv	M1		comparison ft seen or implied
	Reject H_0 Significant evidence at 1% level to suggest an association (positive) between rank orders of personality score and happiness score. Students with a higher extrovert personality score tend to have a higher happiness score.	A1 E1	5	ts/cv correct in context – vice versa OK
	Total		**19**	
	TOTAL		**75**	