Version 1.0

General Certificate of Education (A-level)
June 2011

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 candidates' 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 candidates' 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 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.

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.

SS03

Q	Solution	Marks	Total	Comments
1(a)	H_0 Samples are taken from identical populations H_1 Samples are not taken from identical populations (or population average substance level differs)	B1		Or equivalent hypotheses referring to population medians. Allow pop median A/B or η_A and η_B Must have 'average'; disallow mean
	2 tail 5%			
	Ranks A 1 4 5 6 8 10.5 14 15 (15) (12) (11) (10) (8) (5.5) (2) (1) B 2 3 7 9 10.5 12 13 (14) (13) (9) (7) (5.5) (4) (3)	M1 m1		Attempt at ranks as one group - either way For ties
	$T_A = 1 + 4 + \ldots + 15 = 63.5$ (or 64.5) $T_B = 2 + 3 + \ldots + 13 = 56.5$ (or 55.5)	M1		Totals (dep on ranks - any)
	$U_A = 63.5 - \frac{8\times9}{2} = 27.5$ $U_B = 56.5 - \frac{7\times8}{2} = 28.5$ Test stat U = 27.5	m1A1		U calculated - either correct
	cv = 13	B1		For cv
	U >13	A1		Correct U chosen for comparison with relevant cv (27.5/13 or 28.5/43)
	Accept H_o	A1		Only if ts/cv correct
	No significant evidence at the 5% level to suggest that there is any difference in the average level of the substance for drugs A and B.	E1	10	In context (can ft)
(b)	Type II error is to accept H_0 when actually H_0 is not true.	B1		Allow 'my conclusion was wrong'
	This would mean that the conclusion to the test in part (a) that there is no significant difference in the average level of the substance for drugs A and B is incorrect and there is actually a difference between the two.	E1	2	For context 2 marks only if in context and correct
	Total		**12**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
2(a)(i)	H_0 μ, η = 10.8			
	H_1 μ, η > 10.8 1 tail 5%	B1		Or equivalent in words, eg 'average time spent on study' or 'population average'
	diffs 6.65 3.85 1.5 0.8 –3.7 rank 10 6 2 1 5	M1		For differences from 10.8
	diffs 4.35 5.4 –3.2 –4.05 –2.2 rank 8 9 4 7 3	m1		Ranks as one group dep on differences (allow either way) SC1 for sign test
	T_+ = 10 + 6 + 2 + 1 + 8 + 9 = 36 T_- = 5 + 4 + 7 + 3 = 19 Test stat T = 19	m1 A1		Total of any ranks dep on diffs One correct
	n = 10 cv = 11 T > 11	B1 m1		For cv Comparison lower (plausible) T (not –ve) and cv. Can ft or 44/36
	Accept H_0	A1		
	There is no significant evidence to suggest that average time spent per week of term has increased from 10.8 hours.	E1	9	In context
(a)(ii)	Conclusions cannot be generalised to whole population. Students at the college concerned may not represent a random sample of all such students in the country. Study patterns may vary at different times of the year.	E1	1	For any one point clearly explained (not 'may have lied', 'not correctly recorded')
(b)(i)	Wilcoxon signed-rank takes takes into account the magnitude of the ranks of the differences whereas the sign test only considers the sign of those differences. **or** Wilcoxon signed-rank is more likely to detect a difference if one exists. **or** More powerful.	E1	1	Or 'magnitude of differences' (not 'takes data/size of data into account')
(ii)	If a direction/preference only was given then there would be no numerical data available to find the differences in the data that need to be used for the Wilcoxon signed-rank test. An example would be if students only had to state whether they were studying more hours, less hours or the same hours this year as last year. **or** If data to be analysed was very asymmetrical. An example could be that the times for study were found to be skew.	B1 E1	2	For one valid situation - a direction/preference or asymmetrical … … explained clearly in context
	Total		**13**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
3(a)(i)	<table><tr><th></th><th>None</th><th>1 or 2</th><th>> 2</th></tr><tr><td>**1**</td><td>5.74</td><td>7.40</td><td>2.87</td></tr><tr><td>**2–4**</td><td>18.64</td><td>24.04</td><td>9.32</td></tr><tr><td>**> 4**</td><td>13.62</td><td>17.57</td><td>6.81</td></tr></table>	M1 A1 A1	 3	Method for expected frequencies (16 × 38 / 106 etc); ft incorrect totals 4 correct; allow slight dp inaccuracy All correct to 1 dp
(ii)	There is an expected frequency that is below 5.	E1	1	Must refer to expected frequency
(iii)	"More than 2 falls" pooled with "1 or 2 falls" to make 2 categories of fall: "None" or "One or more"	E1	1	Allow '1' pooled with '2–4' to make 2 categories of medications taken: "1 to 4" and "More than 4"
(iv)	H_0 (Number of) falls is independent of number of medications taken H_1 (Number of) falls is not independent of number of medications taken	B1		OE eg H_0 No association H_1 Association
	1 tail 1%			Alternative if pooled rows:
	<table><tr><th>Obs</th><th>None</th><th>1 or more</th></tr><tr><td>**1**</td><td>10</td><td>6</td></tr><tr><td>**2–4**</td><td>21</td><td>31</td></tr><tr><td>**More 4**</td><td>7</td><td>31</td></tr></table>	M1		<table><tr><th>Obs</th><th>None</th><th>1 or 2</th><th>More 2</th></tr><tr><td>**1 – 4**</td><td>31</td><td>31</td><td>6</td></tr><tr><td>**More 4**</td><td>7</td><td>18</td><td>13</td></tr></table> M1
	<table><tr><th>Exp</th><th>None</th><th>1 or more</th></tr><tr><td>**1**</td><td>5.74</td><td>10.26</td></tr><tr><td>**2–4**</td><td>18.64</td><td>33.56</td></tr><tr><td>**More 4**</td><td>13.62</td><td>24.38</td></tr></table>	m1		For 3E correct <table><tr><th>Exp</th><th>None</th><th>1 or 2</th><th>More 2</th></tr><tr><td>**1 – 4**</td><td>24.38</td><td>31.43</td><td>12.19</td></tr><tr><td>**More 4**</td><td>13.62</td><td>17.57</td><td>6.81</td></tr></table> m1
	ts = $\sum \frac{(O-E)^2}{E}$ = $\frac{4.26^2}{5.74}+\frac{4.26^2}{10.26}+\frac{2.36^2}{18.64}+\frac{2.36^2}{33.36}+\frac{6.62^2}{13.62}+\frac{6.62^2}{24.38}$	m1		For ts; Yates used M0
	ts = 10.4 (10.0–11.0) (SC4 ts = 10.4 NMS if (i) ok)	A1		ts = 13.80 (13.5–14.5) m1A1 (SC4 ts = 13.8 NMS if (i) ok)
	df = 2 1% cv = 9.210 ts > 9.210	B1		
	Reject H_0			
	Sig evidence to suggest that the number of falls is not independent of number of medications taken.	E1	7	E1 only if ts/cv correct and must make sense Allow B1E0 for 4.605, 5.991, 7.378, 10.597

SS03 (cont)

Q	Solution	Marks	Total	Comments
3(b)(i)	<table><tr><th></th><th>No falls</th><th>At least one</th></tr><tr><td>0</td><td>1680</td><td>660</td></tr><tr><td>1</td><td>448</td><td>180</td></tr><tr><td>2</td><td>392</td><td>192</td></tr><tr><td>3+</td><td>280</td><td>168</td></tr></table>	M1 m1 A1	 3	Method for frequencies eg 0.60 × 2800 4 correct all correct
(ii)	H_0 (Number of) falls is independent of number of chronic diseases suffered H_1 (Number of) falls is not independent of number of chronic diseases suffered	B1		OE: H_0 No association H_1 Association
	ts = $\sum \frac{(O-E)^2}{E}$ = 18.4			
	df = 3 1% cv = 11.345	B1		For 11.345
	ts > 11.345	M1		Comparison with 18.4 Allow M1 for 6.251, 7.815, 9.348, 12.838
	Reject H_0			
	Significant evidence to suggest that the number of falls suffered is not independent of number of diseases suffered	E1	4	Conclusion in context
(iii)	Must have attempted (i) and (ii) to gain marks in (iii)			E0,0 if conclusion in (ii) is Accept H_0
	Women who do not suffer from any chronic diseases are less likely than expected to have a fall.	E1		
	Women who suffer from 3 or more chronic diseases are more likely than expected to have a fall. [Women who suffer from 2 chronic diseases are more likely than expected to have a fall.]	E1	2	Or equivalent – no need to refer to expected frequencies but disallow comment referring simply to a comparion of observed frequencies
	Total		**21**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
4(a)	$r = 0.895$	B3	3	SC2 0.89/0.90/0.894 SC1 0.9 Allow M1 summations M1 correct use of S_{xx} S_{xy} S_{yy}
(b)	H_0 $\rho = 0$ H_1 $\rho > 0$	B1 B1		OE in words H_0 pop PMCC = 0 or H_0 no association between BMR and BMI H_1 correct direction B1
	ts $r = 0.895$ $n = 10$ cv = 0.5494 $r > 0.5494$ reject H_0	 B1 M1		 CAO for cv ft provided $-1 < r < +1$
	Significant evidence that there is a positive correlation between BMR and BMI. Men with a higher BMR tend to have a higher BMI.	A1	5	For Reject H_0 ts/cv correct
(c)	Ranks for BMR 10 9 8 7 6 5 4 2½ 2½ 1 or 1 2 3 4 5 6 7 8½ 8½ 10	 M1		 Ranks (can be reversed)
		m1		Ties
		A1		For d 2 1 2 0 3 0 3 ½ ½ 3 $\sum d^2 = 4 + 1 + \ldots + 9 = 36½$
	SRCC $r_s = 0.778(11909)$ ignore sign	B2	5	SRCC = $1 - \frac{6 \times 36\frac{1}{2}}{10 \times 99} = 0.779$
	NMS SC4 $r = 0.78$ SC2 $r = 0.8$ SC0 $r = 0.5636$ SC3 $r = 0.770/0.769$ with ranks SC2 $r = 0.770/0.769$ no ranks			Reversed ranks $\sum d^2 = 292.5$ $r = -0.773$ M1A1 must be 0.779/–0.773
(d)	There is a significant positive correlation between BMR and BMI and there is strong positive rank correlation between BMR and level of daily physical activity.	E1 (no ft)		<u>Both</u> results put together
	Men who have a high BMI tend to have a high BMR as do men who have a high level of daily physical activity.	E1	2	Interpretation in context (not just repeat of conclusion)
(e)(i)	BMR and BMI measurements are normally (or bivariate normal) distributed	B1	1	Mention of normal distribution or linear relationship seen
(ii)	Ranks only available for level of daily physical activity so SRCC is the only correlation coefficient that can be evaluate **or** No actual values given for DPA	E1	1	Clearly in context
	Total		**17**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
5(a)	The purpose is to ensure that each participant has the same opportunity to be assigned to any one of the three methods so that each method group should be roughly equivalent. Therefore any difference observed between method groups can be linked to the effect of the method, not due to a characteristic of the individuals in the group.	E1 E1	2	'Student effect' eliminated 'bias eliminated' More likely to detect any difference as groups more equivalent E0 'more accurate'
(b)	H_0 Samples from identical populations H_1 Samples not from identical populations	B1		Or hypotheses referring to difference between at least 2 population averages (not mean)
	Ranks Prog: 5, 13; 10, 8; 6, 12; 1, 17; 13, 5; 9, 9 Comp: 17, 1; 16, 2; 14, 4; 12, 6; 7, 11 Cont: 15, 3; 11, 7; 8, 10; 4, 14; 3, 15; 2, 16	M1		For ranks as one group (can be reversed)
	Totals of ranks: $T_{prog} = 44/64$ $T_{comp} = 66/24$ $T_{cont} = 43/65$ $n_{prog} = 6$ $n_{comp} = 5$ $n_{cont} = 6$	m1 A1		Totals ft but dep on ranks 2 totals correct
	$\sum_{i=1}^{m} \frac{T_i^2}{n_i} = \frac{44^2}{6} + \frac{66^2}{5} + \frac{43^2}{6} = 1502.03$ 322.7 +..... m1 implied or 682.7 +.....	m1 m1		Numerators correct Denominators correct
	$H = \frac{12}{17 \times 18} \times 1502.03 - (3 \times 18)$	m1		H formula correctly used (need a $\sum_{i=1}^{m} \frac{T_i^2}{n_i}$)
	$= 4.90$	A1		4.80–5.00
	Critical value from $\chi_2^2 = 5.99$	B1		For cv (correct cv only)
	$H < 5.99$ Accept H_0 No reason to doubt that samples are from identical poulations. No significant difference in average scores in test for the 3 methods.	E1	10	Conclusion correct in context (must have ts/cv both correct)
	Total		**12**	
	TOTAL		**75**	