Version 1.0:0608

ASSESSMENT and QUALIFICATIONS ALLIANCE

General Certificate of Education

Statistics 6380

SS03 Statistics 3

Mark Scheme

2008 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

Q	Solution	Marks	Total	Comments
1(a)	H_0 pop median/mean diff $\eta_d = 0$ H_1 pop median/mean diff $\eta_d \neq 0$ 2 tail 5% (*d* is after – before)	B1		Or fully explained in words – population implied, average resistance same/changed
	diff 3 7 –2 5 -1 **rank** 4 7 –2½ 6 -1	M1		For differences (before – after) or (after – before); ignore signs
	diff 4 2 8 **rank** 5 2½ 8	M1 m1		For 8 ranks. smallest = 1 even if no differences or sign ignored For ties used correctly
	$T_+ = 3 + 7 + \ldots + 8 = 32½$ $T_- = 2½ + 1 = 3½$	m1 A1		For total attempted For one correct total
	Test stat T = 3½ *n* = 8 1 tail 5% *n* = 8 cv = 4 T < 4	B1 M1		For cv Comparison cv/ts if valid method seen allow cv one row/col out for M1
	Significant evidence at 5% level to reject H_o and conclude that the average resistance differs after the adjustment (higher)	E1	9	In context – only if ts/cv correct
(b)(i)	Wilcoxon signed rank test takes into account the magnitude of the differences not simply whether they are + or –	E1	1	
(ii)	When the data is not symmetrically distributed so Wilcoxon signed-rank cannot be carried out. Or Data given only as signs/preferences so only sign test possible – no numerical differences can be evaluated	B1 E1	2	Correct reasoning Explained well
	Total		**12**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
2(a)				
		M1		attempt at ranks (can be reversed)
		M1		for 16 correct
		A1		
	$r_s = -0.212$(3 sf from calc)	B3	6	Award B2 for –0.22 ~ –0.20, B1 for –0.2, but B0 for –0.189 (PMCC)
	Alternative			
	d = 5, 3, 1, 5, 3, 2, 3, 1, 6, 9			
	$\sum d^2 = 200$	(B1)		
	$r_s = 1 - \frac{6\times 200}{10\times 99}$	(M1)		
	$= 1 - 1.212 = -0.212$	(A1)		
(b)	H_0 Rank orders of annual road deaths and number of motor vehicles are independent.			H_0 no association
	H_1 Rank orders of annual road deaths and number of motor vehicles are not independent – there is an association	B1		H_1 some association
	2 tail 10%			
	cv = ±0.5636 n = 10 2 tail 10%	B1		for cv
	test stat $r_s = -0.212$ $r_s > -0.5636$	M1		for comparison ts/cv; needs r_s correct 2sf Allow $r_s = 0.212$, cv = 0.5636 but not if signs are different
	Accept H_0 No significant evidence at 10% level to suggest an association between rank orders of annual road deaths and number of motor vehicles for countries in the EU.	A1 E1	5	**SC** –0.189 used can earn max B1B1M1
	Total		**11**	

Country	A	B	C	D	E
x rank	1	2	3	4	5
y rank	6	5	4	9	2
Country	**F**	**G**	**H**	**I**	**J**
x rank	6	7	8	9	10
y rank	8	10	7	3	1

SS03 (cont)

Q	Solution	Marks	Total	Comments
3(a)	H_0 No association between survival and drug treatment used. H_1 Association exists between survival and drug treatment used.	B1		
	1 tail 5%			
	Steroid / **Placebo** **Died** 404.05 / 413.95 **Survived** 656.95 / 673.05	M1 m1		E method All correct (allow integers)
	ts = $\sum \frac{(\lvert O-E\rvert - 0.5)^2}{E}$ =	M1		ts correct denominators
	$\frac{7.55^2}{404.05}+\frac{7.55^2}{413.95}+\frac{7.55^2}{656.95}+\frac{7.55^2}{673.05}$	m1		Attempt at Yates' correction: needs $\frac{\left(\ldots - \frac{1}{2}\right)^2}{\text{denom}}$
	0.141 + … = 0.450	A1		ts = 0.162 +… if no Yates 0.4 ~ 0.5, so A0 for 0.1512 or 0.514
	cv df = 1 5% cv = 3.841	B1		
	ts < 3.841	M1		Must have ts > 0 Or p = 0.0696
	Accept H_0 No sig evidence to suggest an association between survival and whether or not additional drug treatment is used.	A1 E1	10	
	Total		**10**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
3(b)(i)	H_0 No association between the drug used and the level of consciousness H_1 An association exists between the drug used and the level of consciousness	B1		
	1 tail 1%			
	Drug / Level — **Standard** — **New** **Unconscious** — 130 — 90 **Semi-conscious** — 90 — 115 **Fully conscious** — 30 — 45	M1 A1		For attempt to find raw frequencies 4 or more correct
	Expected frequencies **Drug / Level** — **Standard** — **New** **Unconscious** — 110 — 110 **Semi-conscious** — 102.5 — 102.5 **Fully conscious** — 37.5 — 37.5	M1 m1		For one *E* correct For all *E* correct ft if original % used
	ts = $\sum \frac{(O-E)^2}{E}$			
	= $\frac{(130-110)^2}{110} + \frac{(90-110)^2}{110} + ..$	M1		ts sum with correct denominators
	= 13.3	A1		For ts in range 13.0 ~ 13.6
	df = 2 1% cv = 9.21	B1		For cv
	ts > 9.21	M1		For comparison ts/cv Or p = 0.00128
	Reject H_0	A1	10	

SS03 (cont)

Q	Solution	Marks	Total	Comments
(b)(ii)	Sig evidence to suggest an association exists between drug used and level of consciousness – patients given the new drug are **far less** likely to be unconscious 30 minutes after their operation was completed (and vice versa)	E1 E1	2	Sensible correct interpretation in context. Sources of association identified correctly Can award E1 E0 if accept H0 in (b)(i) **SC** Working with percentages throughout part (b) can earn last 4 method marks and 1 E mark, max 5/12 Expected Frequencies 44 44 41 41 15 15 ts = 5.32
	Total		**22**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
4	H_0 Samples are taken from identical populations	B1		or H_0 $\eta_{VLow} = \eta_{Low} = \eta_{Noclaim}$
	H_1 Samples are not taken from identical populations – population average nicotine levels differ	B1		H_1 at least two of $\eta_{VLow}, \eta_{Low}, \eta_{Noclaim}$ do differ
	5% 1 tail			
	Ranks **Very Low Tar** / **Low Tar** / **No Claim Made** 1 / 3 / 6 2 / 7 / 12 4 / 10 / 14 5 / 11 / 15 8 / 13 / 16 9	M1 m1		Ranks (either way) At least 10 correct
	$T_{V\,Low} = 29$ $T_{Low} = 44$ $T_{No\ claim} = 63$ Or (73) (41) (22) $n_{V\,Low} = 6$ $n_{Low} = 5$ $n_{No\ claim} = 5$	m1 A1		Totals (of ranks) any one correct
	$\sum_{i=1}^{m} \frac{T_i^2}{n_i} = \frac{29^2}{6} + \frac{44^2}{5} + \frac{63^2}{5} = 1321.17$	m1		
	$H = \frac{12}{16\times17} \times 1321.17 - (3 \times 17) = 7.29$	A1		test stat $H =$ 7.0 ~ 7.5 $\frac{12}{N(N+1)} \sum_{i=1}^{m} \frac{T_i^2}{n_i} - 3(N+1)$
	Critical value from $\chi_2^2 = 5.991$ 5%	B1		
	$H > 5.991$	M1		Comparison; needs ts > 0
	Sig evidence to reject H_0 and conclude that samples are not from identical populations.	A1		
	Significant evidence at the 5% level to suggest that the population average nicotine level differs for the three categories of king-size cigarettes.	E1		Difference in context
	It appears that those king-size cigarettes that have no claim made about tar levels have a significantly higher average nicotine level than those claimed to have 'Very Low Tar'.	E1	13	Mention of 'at least two' or a sig difference between nicotine levels of king-size cigarettes for which no claim made and those claimed to have 'Very Low Tar'. Can award E1E0 if candidate accepts H_0
	Total		**13**	

SS03 (cont)

Q	Solution	Marks	Total	Comments
5(a)	H_0 Samples are taken from identical populations H_1 Samples are not taken from identical populations (males aged under 30 years have lower average LDL) 1 tail 5%	B1		Hypotheses referring to population averages also acceptable
	Under 30 ranks / **Over 50 ranks** 1 / 6 2 / 9 3 / 11 4 / 12 5 / 13 7 / 14 8 / 15 10 / 16	M1		Attempt at successful separation of age groups
		M1		Attempt at Mann–Whitney - ranks as one group (either way)
	$T_G = 1 + 2 + \ldots + 10 = 40$ $T_R = 6 + 9 + \ldots + 16 = 96$	M1		Attempt at total ranks
	$U_G = 40 - \frac{8\times9}{2} = 4$ $U_R = 96 - \frac{8\times9}{2} = 60$	M1		for U formula correct or alternate method see ranks total $- \frac{8\times9}{2}$
	Test stat $U = 4$	A1		
	cv = 16 $n = 8$ $m = 8$ 1 tail 5% (> 0)	B1		
	$U = 4 < 16$	M1		correct/relevant cv used
	Reject H_0	A1		
	Significant evidence at the 5% level to suggest that the average LDL level is lower for males aged under 30 years.	E1	10	In context

SS03 (cont)

Q	Solution	Marks	Total	Comments
5(b)	$H_0\ \eta = 223$ $H_1\ \eta < 223$ 1 tail 10%	B1		Or equivalent in words
	Signs – + – – – – – – +	M1		signs
	2^+ / 7^-	A1		test stat correct and identified
	Binomial (9, 0.5) model	M1		Binomial model used to attempt probability (or critical region)
	$P(\geq 7^-) = P(\leq 2^+) = 0.0898 < 0.10$ for one tail test	M1		Comparison of Binomial probability with 0.10 (or cr with ts)
	Reject H_0	A1		
	There is sufficient evidence, at the 10% level, to suggest that the median LDL level is greater for males aged 35 to 64 years living in the USA than that for those living in China.	E1	7	Interpretation in context
			17	
	TOTAL		**75**	