A-LEVEL

Statistics

Statistics 3 – SS03

Mark scheme

6380

June 2014

Version/Stage: Final

Mark schemes are prepared by the Lead Assessment Writer 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 associates 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 associate understands and applies it in the same correct way. As preparation for standardisation each associate 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, associates encounter unusual answers which have not been raised they are required to refer these to the Lead Assessment Writer.

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

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
1a	Spearman's rank correlation coefficient is the appropriate measure of correlation for these data because there are no measured values given.	E1	**1**	E1 Ranks only for 400m & position only for cross country or ref to orders given
1b	Ranks			
		M1 A1		M1 for ranks attempt cross country A1 all correct (can be reversed) can be implied by *d*
	m1dep for $\sum d^2 = 134$ $r_s = 1 - \frac{6\times134}{8\times63} = -0.595$ M1 A1			
	$r_s = -0.595$ (3 sig figs)	B3	**5**	B1 r_s negative B2 $0.590 \le \lvert r_s \rvert \le 0.599$
1c	H_0: $\rho_s = 0$ H_1: $\rho_s \neq 0$ 2 tail 5%	B1		Hypotheses oe
	test stat $\lvert r_s \rvert = 0.595$ |critical value| = 0.7381 –0.595 > –0.7381 so no significant evidence exists to reject H_0	B1 M1		Correct abs value for cv 0.738(1) Correct comparison both –ve/ +ve
	This suggests that there is no correlation between rank/ position in 400m races and position in county cross country final race.	E1dep	**4**	Conclusion correct in context
1d	H_0 accepted in error as H_0 actually untrue Conclusion made that there is no correlation between rank/ position in 400m races and position in county cross country final race when, in reality, there is a correlation between them.	B1 E1	**2**	Correct explanation of Type II error In context

	Rank 400m	Rank cross country	|d|
A	3	6 3	3
B	4	1 8	3
C	7	3 6	4
D	5	2 7	3
E	1	8 1	7
F	2	7 2	5
G	8	4 5	4
H	6	5 4	1

1e(i)	PMCC $r = -0.904$ (3 sf) (from calculator)	B3		
				(–0.905, –0.903)
	sc			or $r =$ $\dfrac{8671.488 - \dfrac{434.4 \times 160.07}{8}}{3.17 \times 7.08} = \dfrac{-20.3}{22.4}$
(ii)	–0.90 allow M1 M1 A0 (or B2)			
	–0.9 sc allow B1			$= -0.904$ (3 sf)
				M1 (num), M1(denom), A1
	PMCC indicates a strong negative correlation between best time taken to run 400m and time taken to run cross country race final. This indicates that we would expect faster 400m runners to be slower at running the cross country race.	E1	**4**	Interpretation in context

Q	Solution	Marks	Total	Comments
2a				
	Frequencies / **AP** / **AV** **Baseball** 275 50 **Basketball** 475 75 **Soccer** 350 25	M1 A1	**2**	Correct effort at % for 1 frequency(not 25) All correct
2b	H_0: Coping style is not associated with sport H_1: Coping style is associated with sport 1 tail 1%	B1		
	Expected / **AP** / **AV** **Baseball** 286 39 **Basketball** 484 66 **Soccer** 330 45	M1 M1 A1		Method for expected frequencies 3 or more correct All correct Can be implied by correct ts
	$ts = \sum \frac{(O-E)^2}{E}$ $= \frac{11^2}{286} + \frac{11^2}{39} + ... + \frac{20^2}{330} + \frac{20^2}{45}$	M1 A1		Method for ts seen or implied ts correct (14.9 -15.2)
	= 15.02			
	cv df = 2 1% cv = 9.21 p = 0.00055	B1 B1		for df =2 (can be implied by cv) for cv correct or *p=0.00055*
	ts > 9.21 Reject H_o	A1dep E1dep		Reject H_o Conclusion correct in context
	Sig evidence to suggest that coping strategy is associated with sport involved.		**10**	
2c				
	Soccer officials are far less likely than expected to use the AV coping style (more likely than expected to use the AP coping style). Baseball officials are far more likely than expected to use the AV coping style (less likely than expected to use the AP coping style).	E1 E1	**2**	In context In context
2d				
	Expected / **AP** / **AV** **Male** 24.5 5.5 **Female** 24.5 5.5	M1		Effort at expected freq seen or implied
	$ts = \sum \frac{(\lvert O-E \rvert - 0.5)^2}{E}$ $= \frac{3^2}{24.5} + \frac{3^2}{5.5} + \frac{3^2}{24.5} + \frac{3^2}{5.5}$	M1 m1 A1		Yates used correctly – numerator seen correct ft Whole ts method correct 3.9 – 4.2
	= 4.00		**4**	

Q	Solution	Marks	Total	Comments
3a	So that any influence of the order of taking drugs does not affect the outcome of the investigation.	B1 E1	**2**	Reduction of experimental error In context
3b	H_0: Population mean/median hours relief difference = 0 H_1: Population mean/median hours relief difference ≠ 0 2 tail test 2 % level	B1		H_0: η_A , μ_A = η_B , μ_B H_1: η_A , μ_A ≠ η_B , μ_B
	Differences *B* – *A* 1 2 3 4 5 6 7 8 9 10 11 12 1.5 2.1 0.2 –0.2 2.6 –0.1 –0.6 2.5 2 1.2 3 3.9	M1		For differences – ignore signs Can be implied by correct ranks
	Ranks 6 8 2½ 2½ 10 1 4 9 7 5 11 12	m1dep m1 dep		For any ranks For correct ranks (smallest abs diff = rank1). Ignore ties. m1,1 Both dep differences
	T_+ = 6 + 8 + 2½ + 10 + 9 + 7 + 5 + 11 + 12 = 70.5 T_- = 2½ +1+ 4 = 7.5 test stat T = 7.5	m1dep A1		Effort at total of ranks + and/or - May not see 70.5. m1 dep ranks Either total correct
	critical value = 10 test stat < 10 Reject H_o	B1 m1dep A1dep		cv = 10 correct correct lower tail ts used/identified. Correct conclusion dep ts and cv correct
	There is significant evidence of a difference between the average number of hours of relief from pain gained using Drug A and Drug B. Allow 1 tail conclusion Drug B better/longer relief	E1dep	**10**	Conclusion correct in context dep previous A1
3c	Conclusion based on experiment in which adults self selected to take part. These adults might not be representative of **all** adult arthritis sufferers	E1	**1**	Or mention of volunteers/not selected at random

Q	Solution	Marks	Total	Comments
4	H_0: Samples from identical populations H_1: Samples not from identical populations 2 tail 5% sig level Ranks A 4 7 9 10 11 12 $T_A = 53$ B 1 2 3 5 6 8 $T_B = 25$ $U_A = 53 - \frac{6\times7}{2} = 32$ $U_B = 25 - \frac{6\times7}{2} = 4$ ts = 4 n =6, m = 6 cv =5 ts < 5 Significant evidence to reject H_0 and conclude that there is a difference in the average marks in the Statistics module exam for the two schools	B1 M1 A1 m1 dep A1 B1 M1 A1dep E1dep	**9**	H_0: $\eta_A = \eta_B$ or ref to pop median H_1: $\eta_A \neq \eta_B$ Ranks separated and totalled effort One total correct U method One U correct cv = 5 [or 31 upper tail] only Consistent comparison ts/correct tail cv or ts identified & compared with correct tail cv A1 dep ts and cv correct In context dep previous A1

Q	Solution	Marks	Total	Comments
5a(i)	H_0: Managers have no particular preference for either new or old company structure H_1: Managers prefer new company structure 1 tail 5% Use of 17+ and 8 or 13 – Use of B(25, 0.5) or B(30, 0.5) $P(X \geq 17) = 0.0539$ or 0.2923 $p > 0.05$ (5%) Or use of cr with probs Accept H_0 No significant evidence to suggest that managers/they prefer new company structure	B1 M1 M1 M1 A1dep E1dep	**6**	H_0: p =0.5 H_1: p > or < 0.5 For identifying ts Either correct Bin prob seen Comparison Bin prob and 5% dep correct Bin prob compared 5% In context dep previous A1
(ii)	There are no measurements to use – simply a judgement of prefer or not/no opinion on new structure. Wilcoxon requires symmetrically distributed measurements for preferences not just a prefer/not situation.	E1	**1**	W S-R can't be used if only preferences given, +/- given W S-R needs values

Q	Solution	Marks	Total	Comments
5b(i)	H_0: Samples from identical populations H_1: Samples not from identical populations 5% sig level	B1		H_0: $\eta_{U40} = \eta_{40\text{-}55} = \eta_{55+}$ or ref to pop medians H_1:at least 2 population medians differ oe Allow 1 pop median is different Allow ref to median occup stress if H_1 includes 'at least 2'
	Ranks **Under 40**: 12 6; 10 8; 7 11; 5 13; 4 14; 1 17 **40-55**: 17 1; 16 2; 15 3; 14 4; 8 10 **Over 55**: 13 5; 11 7; 9 9; 6 12; 3 15; 2 16	M1		Ranks effort
	$T_{under40}$= 39 69 $T_{40\text{-}55}$=70 20 T_{over55}=44 64 $n_{under40}$ = 6 $n_{40\text{-}55}$ = 5 n_{over40} = 6	m1 A1		m1 dep ranks used Ranks totalled. At least 1 correct
	$\sum_{i=1}^{m} \frac{T_i^2}{n_i} = \frac{39^2}{6} + \frac{70^2}{5} + \frac{44^2}{6} = 1556.17$	m1 m1		Denominators correct Numerators correct and terms added ft
	$H = \frac{12}{17\times18} \times 1556.17 - (3 \times 18) = 7.03$	M1 A1		*H* method correct A1 6.9 -7.2
	Critical value from χ^2_2 = 5.991	B1		cv correct
	H > 5.991 Significant evidence to reject H_0	A1dep	**10**	dep ts and cv correct
5b(ii)	At least 2 groups' average scores differ.	B1		A difference between at least 2 groups exists B1 Can be implied in (i)
	The '40-55 years' age group are significantly more stressed than the 'under 40 years' age group.	E1	**2**	Difference identified [40-55 most or under 40 least] in context E1 (full explanation in context gets B1 E1)