



**General Certificate of Education (A-level)**  
**January 2012**

**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](http://aqa.org.uk)

Copyright © 2011 AQA and its licensors. All rights reserved.

**Copyright**

AQA retains the copyright on all its publications. However, registered schools/colleges for AQA are permitted to copy material from this booklet for their own internal use, with the following important exception: AQA cannot give permission to schools/colleges to photocopy any material that is acknowledged to a third party even for internal use within the centre.

Set and published by the Assessment and Qualifications Alliance.

## 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)		M1	6	Attempt at ranks 14 correct (can be reversed)
		M1		
		A1		

**SS03 (cont)**

Q	Solution	Marks	Total	Comments
2 (a)	$H_0: \mu, \eta = \text{£}81,050$ $H_1: \mu, \eta < \text{£}81,050$ 1 tail 5%  diffs -17530 -16450 -9050 -22600 rank 8 7 3 9  diffs 1150 -14550 5550 -12850 rank 1 6 2 5  diffs -11950 rank 4  $T_+ = 1 + 2 = 3$ $T_- = 8 + 7 + 3 + 9 + 6 + 5 + 4 = 42$  Test stat $T = 3$ $n = 9$ $cv = 8$ $T < 8$  Reject $H_0$ .  There is significant evidence to suggest that average gross annual salary for consultants and medical specialists in the UK was greater than that for those working in France during 2003	B1  M1 m1  m1 A1  B1 m1  E1	8  1	consistent or equivalent in words  For differences Ranks  Total of ranks One correct  For cv Correct ts identified for cv comparison  In context
(b)	Distribution of <u>differences</u> in gross annual salary for consultants and medical specialists is symmetrical	E1	1	
	<b>Total</b>		<b>9</b>	

**SS03 (cont)**

Q	Solution	Marks	Total	Comments
3(a)(i)	Sign test	B1	1	
(ii)	No measured data just a decision reduced/increased or no change	E1	1	
(b)	$H_0: \eta = 0$ $H_1: \eta < 0$ 1 tail 5%  - + - - - . - - + - ts 7-, 2+  Binomial model B ( 9, 0.5) $P(\geq 7-) = P(\leq 2+) = 0.0898 > 0.05$ for one tail test  Accept $H_0$ .  There is insufficient evidence, at the 5% level, to indicate that the doctor's belief is supported.	B1  M1 M1 A1  M1 m1  E1	1        7	Allow $H_1: \eta > 0$ if signs consistent  For signs Excluding 'no change'  Using B ( 9, 0.5) Comparison correct prob with 0.05  Condone $n=10$ ts 7-, 3+ M1M1A1 B( 10, 0.5) used M1 0.1719 > 0.05 m1 E1 ft
(c)	Sample clearly not random . There is no control group.	E1	1	Or other relevant reason
	<b>Total</b>		<b>10</b>	

SS03 (cont)

Q	Solution	Marks	Total	Comments																										
4(a)	$H_0$ Samples from identical populations $H_1$ Samples not from identical populations 5% sig level	B1		or symbols/words ref to population average																										
	Ranks	M1		Attempt at ranks																										
	<table><tr><th>Diet A</th><th>Diet B</th><th>Diet C</th></tr><tr><td>1     20</td><td>2     19</td><td>12    9</td></tr><tr><td>3     18</td><td>4½   16½</td><td>15    6</td></tr><tr><td>4½   16½</td><td>8     13</td><td>17    4</td></tr><tr><td>6     15</td><td>9     12</td><td>19    2</td></tr><tr><td>7     14</td><td>11    10</td><td>20    1</td></tr><tr><td>10    11</td><td>13    8</td><td></td></tr><tr><td>16    5</td><td>14    7</td><td></td></tr><tr><td>18    3</td><td></td><td></td></tr></table>	Diet A	Diet B	Diet C	1     20	2     19	12    9	3     18	4½   16½	15    6	4½   16½	8     13	17    4	6     15	9     12	19    2	7     14	11    10	20    1	10    11	13    8		16    5	14    7		18    3			m1	14 or more correct (can be reversed)
	Diet A	Diet B	Diet C																											
	1     20	2     19	12    9																											
	3     18	4½   16½	15    6																											
	4½   16½	8     13	17    4																											
	6     15	9     12	19    2																											
	7     14	11    10	20    1																											
	10    11	13    8																												
	16    5	14    7																												
	18    3																													
	$T_A=65\frac{1}{2}(102\frac{1}{2})T_B=61\frac{1}{2}(85\frac{1}{2})T_C=83(22)$ $n_A = 8$ $n_B = 7$ $n_C = 5$	m1		Totals attempted																										
$\sum_{i=1}^m \frac{T_i^2}{n_i} = \frac{65\frac{1}{2}^2}{8} + \frac{61\frac{1}{2}^2}{7} + \frac{83^2}{5}$ $= 2454.4$	m1		$\sum_{i=1}^m \frac{T_i^2}{n_i}$ attempt																											
$H = \frac{12}{20 \times 21} \times 2454.4 - (3 \times 21)$ $= 7.12(\text{ or } 7.13)$	m1		H attempt																											
	A1		7.0 – 7.2																											
Critical value from $\chi^2_2 = 5.99$ $H > 5.99$	B1 M1																													
Sig evidence to reject $H_0$ and conclude that samples are not from identical populations	A1																													
There is significant evidence of a difference between <u>at least two</u> of the diets in terms of average percentage reduction in body weight for overweight men. (Diet C is most effective).	E1	11																												
(b)(i) There may be concern that diet C caused problems that led to the men not completing the diet.	E1	1																												
(ii) She may wish to check that the illness was not caused or made worse by the diet	E1	1																												
	<b>Total</b>	<b>13</b>																												

**SS03 (cont)**

Q	Solution				Marks	Total	Comments
<b>5(a)(i)</b>		<b>Accident in 09</b>	<b>No accident in 09</b>	<b>Tot</b>			
	<b>17-18 years</b>	26	<b>174</b>	<b>200</b>	M1		Some sensible effort
	<b>19-50 years</b>	<b>48</b>	<b>652</b>	<b>700</b>	A1	2	3 cell frequencies correctly placed
	<b>51 years +</b>	<b>12</b>	<b>288</b>	300			
	<b>Total</b>	86	<b>1114</b>	1200			
<b>(ii)</b>	<p>H<sub>0</sub> Involvement in a car accident is independent of age  H<sub>1</sub> Involvement in a car accident is not independent of age  1 tail 1%</p>				B1		
	<b>Expected</b>	<b>Acc</b>	<b>No acc</b>				
	<b>17-18 years</b>	14.33	185.67		M1		Method for expected frequencies
	<b>19-50 years</b>	50.17	649.83				
	<b>51 years +</b>	21.5	278.5		A1		All correct
	$ts = \sum \frac{(O - E)^2}{E}$						
	$= \frac{11.67^2}{14.33} + \frac{11.67^2}{185.67} + \dots + \frac{9.5^2}{21.5} + \frac{9.5^2}{278.5}$				m1 m1		Numerator correct Denominator correct
	= 14.85				A1		ts correct ( 13.0 -16.0)
	cv df = 2    1%    cv = 9.21				B1		
	ts > 9.21				M1		
<b>(iii)</b>	Reject H <sub>0</sub> .				A1		
	Sig evidence to suggest that involvement in a car accident is not independent of age				E1ft	10	
	17-18 year olds are far <b>more</b> likely to have had an accident.				E1		
	Those aged 51 years and over are far <b>less</b> likely to have had an accident				E1	2	



**SS03 (cont)**

Q	Solution	Marks	Total	Comments												
5 cont. (b)(i)	<table><tr><td>Expected</td><td>£0-£2000</td><td>£2001-£4000</td><td>Over £4000</td></tr><tr><td>17-30 years</td><td>31.40</td><td>15.70</td><td>6.90</td></tr><tr><td>31 years +</td><td>18.60</td><td>9.30</td><td>4.10</td></tr></table>	Expected	£0-£2000	£2001-£4000	Over £4000	17-30 years	31.40	15.70	6.90	31 years +	18.60	9.30	4.10	M1	2	For 3 correct
	Expected	£0-£2000	£2001-£4000	Over £4000												
	17-30 years	31.40	15.70	6.90												
	31 years +	18.60	9.30	4.10												
		A1	All correct to 1 dp These marks may be gained in part (ii) Allow already pooled													
	(ii)	<p>Pooled expected values</p> <table><tr><td>Expected</td><td>£0-2000</td><td>Over £2000</td></tr><tr><td>17-30 years</td><td>31.40</td><td>22.60</td></tr><tr><td>31 years +</td><td>18.60</td><td>13.40</td></tr></table>	Expected	£0-2000	Over £2000	17-30 years	31.40	22.60	31 years +	18.60	13.40	M1		Last 2 columns pooled		
		Expected	£0-2000	Over £2000												
		17-30 years	31.40	22.60												
		31 years +	18.60	13.40												
<p>H<sub>0</sub> Size of claim is independent of age H<sub>1</sub> Size of claim is not independent of age 1 tail 1%</p>		B1														
$ts = \sum \frac{( O - E  - 0.5)^2}{E}$		M1		ts												
$= \frac{4.9^2}{31.4} + \frac{4.9^2}{22.6} + \frac{4.9^2}{18.6} + \frac{4.9^2}{13.4}$		m1		Yates used												
$= 4.91$		A1		4.7 – 5.2												
<p>cv df = 1    1%    cv = 6.635</p>																
<p>ts &lt; 6.635 Accept H<sub>0</sub></p>		B1		<b>Alt</b> for non pooling <b>sc 5</b> B1 hypotheses M1 test stat m0 no Yates A1 ft 6.06 ( 5.9 – 6.2) B1 ft df=2    cv = 9.210 E1 ft												
<p>No sig evidence to suggest that size of claim is associated with age</p>	E1	7	Pooled but no Yates <b>sc 6</b> M1B1M1m0A1B1M1													
	<b>Total</b>		<b>23</b>													

SS03 (cont)

Q	Solution	Marks	Total	Comments																														
6	<p><math>H_0</math> Samples from identical populations <math>H_1</math> Samples not from identical populations</p> <p>5% sig level 2 tail</p> <p>Ranks</p> <table><thead><tr><th colspan="2">Men who have taken caffeine</th><th colspan="2">Men who have not taken caffeine</th></tr></thead><tbody><tr><td>2</td><td>13</td><td>1</td><td>14</td></tr><tr><td>6</td><td>9</td><td>3</td><td>12</td></tr><tr><td>8</td><td>7</td><td>4</td><td>11</td></tr><tr><td>10½</td><td>4½</td><td>5</td><td>10</td></tr><tr><td>12</td><td>3</td><td>7</td><td>8</td></tr><tr><td>13</td><td>2</td><td>9</td><td>6</td></tr><tr><td>14</td><td>1</td><td>10½</td><td>4½</td></tr></tbody></table> <p><math>T_{Caf} = 65\frac{1}{2}(39\frac{1}{2})</math> <math>T_{nocaf} = 39\frac{1}{2}(65\frac{1}{2})</math> <math>n_{Caf} = 7</math> <math>n_{nocaf} = 7</math></p> <p><math>U_{Caf} = 65\frac{1}{2} - \frac{7 \times 8}{2} = 37\frac{1}{2}</math> <math>U_C = 39\frac{1}{2} - \frac{7 \times 8}{2} = 11\frac{1}{2}</math> <math>U = 11\frac{1}{2}</math></p> <p>cv = 9 for n= 7, m = 7 2 tail 5%</p> <p><math>U &gt; 9</math></p> <p>Accept <math>H_0</math></p> <p>No significant evidence of any difference between average RER for the men who took caffeine and those who did not.</p>	Men who have taken caffeine		Men who have not taken caffeine		2	13	1	14	6	9	3	12	8	7	4	11	10½	4½	5	10	12	3	7	8	13	2	9	6	14	1	10½	4½	B1  
Men who have taken caffeine		Men who have not taken caffeine																																
2	13	1	14																															
6	9	3	12																															
8	7	4	11																															
10½	4½	5	10																															
12	3	7	8																															
13	2	9	6																															
14	1	10½	4½																															