

Edexcel Maths S3

Mark Scheme Pack

2002-2015

Question Number	Scheme	Marks
1.	<p>(a) Stratified</p> <p>(b) Label De-luxe rooms 1 – 20</p> <p>Using <i>random numbers</i> in range 1 – 20 select 2 rooms</p> <p>Repeat for Premier using 1 – 40 and select 4 rooms</p> <p>Repeat for Standard using 1 – 100 and select 10 rooms</p>	<p>B1 (1)</p> <p>B1</p> <p>B1 B1</p> <p>B1 (4)</p> <p>(5 marks)</p>
2.	<p>(a) $H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$</p> <p>standard error = $\sqrt{\frac{9.1^2}{100} + \frac{8.4^2}{120}} = 1.19$ (awrt)</p> <p>$\alpha = 0.01 \Rightarrow$ CR: $z < -2.5758$ or $z > 2.5758$</p> <p>$z = \frac{70.6 - 67.2}{1.19} = 2.86$ (awrt)</p> <p>Since 2.86 is in the critical range, H_0 is rejected. There is evidence of a difference in mean playing time.</p> <p>(b) Central Limit Theorem applies to enable normal distribution to be used.</p>	<p>B1 B1</p> <p>M1 A1</p> <p>B1 need both</p> <p>M1 A1</p> <p>A1ft (8)</p> <p>B1 (1)</p> <p>(9 marks)</p>
3.	<p>(a) $\bar{M} \sim N(80, \frac{2.6^2}{10})$ or $N(80, 0.676)$</p> <p>(b) $P(\bar{M} < 78.5) = P(z < \frac{78.5 - 80}{2.6/\sqrt{10}})$</p> <p style="padding-left: 40px;">$= P(z < -1.82)$</p> <p style="padding-left: 40px;">$= 0.0344$</p> <p>(c) Let W = weight of all 10 people</p> <p style="padding-left: 20px;">$W = M_1 + \dots + M_6 + F_1 + \dots + F_4$</p> <p style="padding-left: 20px;">$E(W) = (6 \times 80) + (4 \times 59) = 716$</p> <p style="padding-left: 20px;">$\text{Var}(W) = (6 \times 2.6^2) + (4 \times 1.9^2) = 55$</p> <p style="padding-left: 20px;">$P(W > 730) = P(z > \frac{730 - 716}{\sqrt{55}})$</p> <p style="padding-left: 40px;">$= P(z > 1.89)$</p> <p style="padding-left: 40px;">$= 0.0294$</p>	<p>B1 B1 (2)</p> <p>M1</p> <p>A1</p> <p>A1 (3)</p> <p>B1</p> <p>B1</p> <p>M1 A1</p> <p>A1 (5)</p> <p>(10 marks)</p>

awrt = “anything which rounds to...”

Question Number	Scheme	Marks																																	
<p>4. (a)</p>	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;"></th> <th style="width: 5%;">A</th> <th style="width: 5%;">B</th> <th style="width: 5%;">C</th> <th style="width: 5%;">D</th> <th style="width: 5%;">E</th> <th style="width: 5%;">F</th> <th style="width: 5%;">G</th> <th style="width: 5%;">H</th> <th style="width: 5%;">I</th> <th style="width: 5%;">J</th> </tr> </thead> <tbody> <tr> <td>Performance</td> <td>10</td> <td>5</td> <td>8</td> <td>3</td> <td>9</td> <td>6</td> <td>1</td> <td>4</td> <td>7</td> <td>2</td> </tr> <tr> <td>Dedication</td> <td>7</td> <td>6</td> <td>3</td> <td>5</td> <td>9</td> <td>10</td> <td>4</td> <td>2</td> <td>8</td> <td>1</td> </tr> </tbody> </table> <p>$\Sigma d^2 = 70$</p> $r_s = 1 - \frac{6 \times 70}{10 \times 99} = 0.576$		A	B	C	D	E	F	G	H	I	J	Performance	10	5	8	3	9	6	1	4	7	2	Dedication	7	6	3	5	9	10	4	2	8	1	<p>M1</p> <p>M1 A1</p> <p>M1 A1 (5)</p>
	A	B	C	D	E	F	G	H	I	J																									
Performance	10	5	8	3	9	6	1	4	7	2																									
Dedication	7	6	3	5	9	10	4	2	8	1																									
<p>(b)</p>	<p>$H_0: \rho = 0; H_1: \rho \neq 0$</p> <p>$n = 10 \Rightarrow$ critical value = 0.5636</p> <p>0.576 is in the critical region</p> <p>Evidence of correlation between performance and dedication.</p>	<p>B1 B1</p> <p>B1</p> <p>M1</p> <p>A1ft (5)</p>																																	
<p>(c)</p>	<p>Likely to be an element of judgement in grading.</p> <p>Dedication unlikely to be normally distributed.</p>	<p>B1 (1)</p> <p style="text-align: right;">(11 marks)</p>																																	
<p>5.</p>	<p>Expected Frequency Male: 50.98 27.85 39.17 Female: 57.02 31.15 48.83</p> <p>H_0: no association between gender and facility</p> <p>H_1: Association between gender and facility</p> $\sum \frac{(O - E)^2}{E} = \frac{(50.98 - 40)^2}{50.98} + \frac{(57.02 - 68)^2}{57.02} + \dots + \frac{(43.83 - 31)^2}{43.83}$ <p style="text-align: center;">= 12.7</p> <p>$\alpha = 0.05, \nu = 2 \Rightarrow$ CR: $\chi^2 > 5.991$</p> <p>Evidence of association between gender and facility</p>	<p>M1 A1 A1</p> <p>B1</p> <p>B1</p> <p>M1 A1</p> <p>A1</p> <p>B1 B1</p> <p>A1ft (11)</p> <p style="text-align: right;">(11 marks)</p>																																	

ft = follow through mark

Question Number	Scheme	Marks
6.	<p>(a) $R = 43.76$; $S = 54.68$; $T = 43.76$ using tables (OR $R = 43.75$; $S = 54.69$; $T = 43.75$ using calculator)</p> <p>(b) H_0: Binomial model with $n = 8$, $p = 0.5$ is suitable H_1: Binomial model with $n = 8$, $p = 0.5$ is not suitable Amalgamation of data $\sum \frac{(O - E)^2}{E} = 5.69 \text{ (awrt)}$ $\alpha = 0.05$, $\nu = 6 \Rightarrow \text{CR: } \chi^2 > 12.592$ Since 5.69 is not in the critical region there is no evidence to reject H_0. The binomial model with $n = 8$ and $p = 0.5$ is a suitable model.</p> <p>(c) Apart from the expected values and $\sum \frac{(O - E)^2}{E}$ being different, the degrees of freedom would have been reduced by 1 ($\nu = 5$).</p>	<p>M1 A1; B1 B1 (4)</p> <p>B1 (both) M1 M1 A1 <u>B1 B1</u> A1ft (7)</p> <p>B1 (1)</p> <p>(12 marks)</p>
7.	<p>(a) Cooling by subtracting 500 for each observation gives $\text{Mean} = 500 + \frac{22}{10} = 502.2$ $\text{Variance} = \frac{1}{9} \left\{ 288 - \frac{22^2}{10} \right\} = 26.622$</p> <p>(b) Limits are $502.2 \pm 1.6449 \times 5.0$ (493.98, 510.42) [accept (494, 510)]</p> <p>(c) 95 % confidence limits are $502.2 \pm 1.96 \times \frac{5.0}{\sqrt{10}}$ (499, 505)</p> <p>(d) $H_0: \mu = 500$ $H_1: \mu > 500$ $\alpha = 0.05 \Rightarrow \text{CR: } z > 2.3263$ $z = \frac{503.9 - 500}{5.0 / \sqrt{15}} = 1.47$ 1.47 is not in the critical region \Rightarrow no evidence to reject H_0; no evidence to suggest mean is greater than 500g</p>	<p>M1 A1 M1 A1 A1 (5)</p> <p>M1 A1 (2)</p> <p>M1 A1ft B1 (for 1.96) A1 A1 (5)</p> <p>B1 (both) B1 M1 A1 A1 ft (5)</p> <p>(17 marks)</p>

Question number	Scheme	Marks
<p>1. (a)</p> <p>(b)</p>	<p>Take a (simple) random sample from (mutually exclusive) groups of the population Sample sizes within strata in strict proportion to numbers in each strata in the population Advantage: More accurate estimate of variance of population mean Individual estimates for strata available Disadvantage: Difficult if strata are large Definition of strata problematic (may overlap)</p> <p>Non-random sampling from groups of the population Advantage: Representative sample can be achieved with small sample size Cheap (costs kept to a minimum) Administration relatively easy Disadvantage Not possible to estimate sampling errors due to lack of randomness Judgment of interviewer can affect choice of sample – bias OK Non-response not recorded Difficulties of defining controls e.g. social class</p>	<p>1g/1h B1 B1 Any one B1 Any one B1 B1 B1dep Any one (not quick) B1 Any one B1</p> <p>(4) (4)</p> <p style="text-align: center;">8</p>
<p>2. (a)</p> <p>(b)</p>	<p>$X \sim N(124, 20^2)$ or $\bar{X} \sim (124, \frac{20^2}{30})$ or assume σ^2 estimated by s^2 or CLT, vals. $\bar{x} \pm 2.5758 \frac{\sigma}{\sqrt{n}} = 124 \pm 2.5758 \frac{20}{\sqrt{30}}$ $= 124 \pm 9.405$ $= (115, 133)$</p> <p>140 is not in confidence interval Underweight apples chosen or Sample may not be representative/may be biased</p>	<p>B1,B1 B1M1A1 3 sf A1 M1 Any one A1f</p> <p>(6) (2)</p> <p style="text-align: center;">8</p>

Question number	Scheme	Marks																															
<p>3. (a) (b) (c)</p>	<p>$E(X-Y)=20-10=10$ $Var(X-Y)=5+4=9$ $X-Y \sim N(10,9)$ $P(13 < X-Y < 16) = P(X-Y < 16) - P(X-Y < 13)$ $= P(Z < \frac{16-10}{3}) - P(Z < \frac{13-10}{3})$ $= P(Z < 2) - P(Z < 1)$ $= 0.9772 - 0.8413 = 0.1359$</p>	<p>Require minus, 10 M1A1 Require plus, 9 M1A1 Implied B1f Subtract M1 Standardise M1 2&1 A1 0.1359 A1 (5)</p>																															
<p>4.</p>	<p>H_0 : Taking drug and catching a cold are independent (not associated) H_1 : Taking drug and catching a cold are not independent (associated) (not ditto)</p> <table border="1" data-bbox="284 1003 1133 1133"> <tr> <td></td> <td>Cold</td> <td>Not Cold</td> <td></td> </tr> <tr> <td>Drug</td> <td>34 (39.5)</td> <td>66 (60.5)</td> <td>100</td> </tr> <tr> <td>Dummy</td> <td>45 (39.5)</td> <td>55 (60.5)</td> <td>100</td> </tr> <tr> <td></td> <td>79</td> <td>121</td> <td>200</td> </tr> </table> <table border="1" data-bbox="284 1155 922 1375"> <tr> <td>O</td> <td>E</td> <td>$\frac{(O - E)^2}{E}$</td> </tr> <tr> <td>34</td> <td>39.5</td> <td>0.766</td> </tr> <tr> <td>66</td> <td>60.5</td> <td>0.5</td> </tr> <tr> <td>45</td> <td>39.5</td> <td>0.765</td> </tr> <tr> <td>55</td> <td>60.5</td> <td>0.5</td> </tr> </table> <p>$\sum \frac{(O - E)^2}{E} = 2.53$ (NB with Yates 2.09) attempt & add, awrt 0.766 & 0.5 twice, awrt 2.53 M1A1A1 $\nu = 1, \chi^2_1(5\%) = 3.841 > 2.53$ 1, 3.841 B1,B1 No reason to believe that the chance of catching a cold is affected by taking the new drug A1f</p>		Cold	Not Cold		Drug	34 (39.5)	66 (60.5)	100	Dummy	45 (39.5)	55 (60.5)	100		79	121	200	O	E	$\frac{(O - E)^2}{E}$	34	39.5	0.766	66	60.5	0.5	45	39.5	0.765	55	60.5	0.5	<p>B1 both B1 All totals B1 $E = \frac{RT \times CT}{GT}$ M1A1A1 (9)</p>
	Cold	Not Cold																															
Drug	34 (39.5)	66 (60.5)	100																														
Dummy	45 (39.5)	55 (60.5)	100																														
	79	121	200																														
O	E	$\frac{(O - E)^2}{E}$																															
34	39.5	0.766																															
66	60.5	0.5																															
45	39.5	0.765																															
55	60.5	0.5																															

Question number	Scheme	Marks																																								
5	<p>μ_a and μ_b are mean weight of population after and before closure respectively.</p> <p>$H_0 : \mu_b = \mu_a$</p> <p>$H_1 : \mu_b > \mu_a$</p> $z = \frac{10 - 8}{\sqrt{\frac{2.64^2}{100} + \frac{1.94^2}{120}}}$ <p style="text-align: right;">Fraction, denom Ok alone</p> $z = \frac{2}{\sqrt{0.1011}} = 6.29$ <p style="text-align: right;">awrt 6.29</p> <p>Critical region is $z \geq 1.6449$, $6.29 > 1.6449$ or in critical region or Reject H_0 (or $P(Z \geq 6.29) = 0, 0 < 0.05$ or z in critical region or Reject H_0 B1M1)</p> <p>There is evidence that closing the factory has reduced mean river pollution</p>	<p>B1</p> <p>B1B1</p> <p>M1A1</p> <p>M1A1</p> <p>A1</p> <p>1.6449 B1, M1</p> <p>A1f</p> <p>(11)</p> <p style="text-align: center;">11</p>																																								
6 (a)	<table border="1" data-bbox="277 1003 1286 1137"> <tbody> <tr> <td>A</td> <td>2</td> <td>5</td> <td>3</td> <td>7</td> <td>8</td> <td>1</td> <td>4</td> <td>6</td> <td></td> </tr> <tr> <td>B</td> <td>3</td> <td>2</td> <td>6</td> <td>5</td> <td>7</td> <td>4</td> <td>1</td> <td>8</td> <td></td> </tr> <tr> <td> d </td> <td>1</td> <td>3</td> <td>3</td> <td>2</td> <td>1</td> <td>3</td> <td>3</td> <td>2</td> <td></td> </tr> <tr> <td>d²</td> <td>1</td> <td>9</td> <td>9</td> <td>4</td> <td>1</td> <td>9</td> <td>9</td> <td>4</td> <td>46</td> </tr> </tbody> </table> $r_s = 1 - \frac{6 \times 46}{8 \times 63}$ $r_s = 0.452$	A	2	5	3	7	8	1	4	6		B	3	2	6	5	7	4	1	8		d	1	3	3	2	1	3	3	2		d ²	1	9	9	4	1	9	9	4	46	<p>d M1</p> <p>$\sum d^2$ M1A1</p> <p>M1A1f</p> <p>0.452 A1</p> <p>(6)</p>
A	2	5	3	7	8	1	4	6																																		
B	3	2	6	5	7	4	1	8																																		
d	1	3	3	2	1	3	3	2																																		
d ²	1	9	9	4	1	9	9	4	46																																	
(b)	<p>$H_0 : \rho = 0$, $H_1 : \rho \neq 0 (\rho > 0)$</p> <p>critical values are ± 0.7381 (0.6429)</p> <p>$0.452 < 0.7381$ ($0.452 < 0.6429$) or not sig or Insufficient evidence to reject H_0</p> <p>No agreement between the two judges.</p>	<p>B1B1</p> <p>0.7381(0.6429) B1</p> <p>M1</p> <p>Context A1f</p> <p>(5)</p> <p style="text-align: center;">11</p>																																								

Question number	Scheme	Marks														
7																
(a)	$\mu = 0.3 \times 50 + 0.2 \times 10 + 0.5 \times 2 = 18$ $\sigma^2 = (0.3 \times 50^2 + 0.2 \times 10^2 + 0.5 \times 2^2) - 18^2 = 448$	<p>M1A1</p> <p>M1A1</p> <p>(4)</p>														
(b)	<p>(50,50) or (50,50) without ordered pairs</p> <p>(10,2) (10,2)</p> <p>(2,10) (10,10)</p> <p>(10,10) (50,10)</p> <p>(50,10) (2,2)</p> <p>(10,50) (50,2)</p> <p>(2,2)</p> <p>(50,2)</p> <p>(2,50)</p> <p style="text-align: right;">either, -1 each missing pair</p>	<p>B2</p> <p>(2)</p>														
(c)	<table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="text-align: center;">\bar{x}</td> <td style="text-align: center;">2</td> <td style="text-align: center;">6</td> <td style="text-align: center;">10</td> <td style="text-align: center;">26</td> <td style="text-align: center;">30</td> <td style="text-align: center;">50</td> </tr> <tr> <td style="text-align: center;">$P(\bar{X} = \bar{x})$</td> <td style="text-align: center;">0.25</td> <td style="text-align: center;">0.2</td> <td style="text-align: center;">0.04</td> <td style="text-align: center;">0.3</td> <td style="text-align: center;">0.12</td> <td style="text-align: center;">0.09</td> </tr> </table>	\bar{x}	2	6	10	26	30	50	$P(\bar{X} = \bar{x})$	0.25	0.2	0.04	0.3	0.12	0.09	<p>All means, probabs multiplied, -1 each error B1 M1 A2</p> <p>(4)</p>
\bar{x}	2	6	10	26	30	50										
$P(\bar{X} = \bar{x})$	0.25	0.2	0.04	0.3	0.12	0.09										
(d)	$P(2 \leq \bar{X} < 7) = 0.25 + 0.2 = 0.45$	<p>Probabilities of 2 and 6 added, 0.45 M1 A1J</p> <p>(2)</p>														
(e)	$E(\bar{X}) = 2 \times 0.25 + 6 \times 0.2 + \dots = 18$ $\text{Var}(\bar{X}) = 2^2 \times 0.25 + 6^2 \times 0.2 + \dots - 18^2 = 224$ $\sum x^2 P(X = x) - (\text{theirs})^2, 224$ <p>So $E(\bar{X}) = 18 = \mu$ and $\text{Var}(\bar{X}) = 224 = \frac{1}{2} \sigma^2$ as required.</p>	<p>$\sum xP(X = x)$ from table, 18 M1 A1</p> <p>M1A1</p> <p>A1</p> <p>(5)</p>														

Question Number	Scheme	Marks
1a)	<p><u>Allocate a number between 1 and N (or equiv) to each pupil.</u></p> <p>Use <u>random number tables, computer or calculator</u> to select 15 <u>different</u> numbers between 1 and 120 (or equiv).</p> <p>Pupils corresponding to these numbers become the sample.</p>	<p>M1</p> <p>B1</p> <p>B1</p> <p>(3)</p>
(b)	<p>Allocate numbers 1 – 64 to girls and 1 – 56 to boys. Idea of different sets for boys and girls</p> <p>Select $\frac{64}{120} \times 15 = 8$ random numbers between 1 – 64 for girls</p> <p>Select 7 random numbers between 1 – 56 for boys.</p>	<p>M1</p> <p>attempt find no</p> <p>M1</p> <p>A1</p> <p>Both 7 and 8</p> <p>(3)</p>
2a)	<p>$H_0 : \rho = 0 ; H_1 : \rho > 0$</p> <p>$\rho$</p> <p>5% CV – PMCC <u>0.6215</u></p> <p>0.572 < 0.6215 / not in critical region / not significant</p> <p>No evidence of <u>positive</u> correlation</p> <p>Spearman <u>0.6429</u></p>	<p>both and</p> <p>B1</p> <p>B1</p> <p>M1</p> <p>A1</p> <p>B1</p> <p>B1</p> <p>(6)</p>
(b)	<p>Evidence of <u>positive</u> correlation</p> <p>No evidence to suggest that as <u>Statistics marks increased</u> <u>Geography marks increased.</u></p> <p>Evidence that students <u>ranked highly in Statistics were also ranked highly in Geography</u></p>	<p>B1</p> <p>Context and not correlation</p> <p>B1</p> <p>ranked</p> <p>(2)</p>

Question Number	Scheme	Marks
3a)	<p>$H_0 : \mu_A = \mu_B ; H_1 : \mu_B > \mu_A$</p> <p>$\mu$</p> $z = \pm \frac{249 - 251}{\sqrt{\frac{2.5^2}{10} + \frac{2.3^2}{15}}}$ <p>$= \pm 2.0227\dots$</p> <p>$CV = \pm 1.6449$</p> <p>or $P(Z \geq 2.02) = 0.0212 - 0.0217,$ or $P(Z \leq 2.02) = 0.9788 - 0.9783$</p> <p>$- 2.0227 < - 1.6449$ or $2.0227 > 1.6449$, or $0.0212 - 0.0217 < 0.05$ or $0.9788 - 0.9783 > 0.95$</p> <p>There is evidence that the <u>mean amount of coffee</u> dispensed by B <u>is greater</u> than A.</p>	<p>both and B1</p> <p>249,251 accept M1</p> <p>$\sqrt{\frac{2.5}{10} + \frac{2.3}{15}}$ for M A1</p> <p>awrt ± 2.02 A1</p> <p>B1</p> <p>M1</p> <p>comparison and consistency needed</p> <p>A1√ context (7)</p>
b)	<p>Machine B amounts are normally distributed.</p>	<p>B1 (1)</p>

Question Number	Scheme	Marks
4a)	$\bar{x} = 75.3$ $s^2 = \frac{1}{9} \left\{ 57455 - \frac{753^2}{10} \right\}$ $= 83.7\dot{8} , 83\frac{71}{90}, 83.8$	B1 M1 awrt 83.8 A1 (3)
b)	$74.8 \pm 1.96\sqrt{\frac{84.6}{100}}$ (73.0, 76.6)	1.96 B1 any z value, may use 75.3, 83.8 for M M1 A1√ on z only awrt 73.0, 76.6 A1, A1 (5)
c)	Journey times independent Sample large enough to use central limit theorem Same distribution / population	any 2 B1, B1 (2)

Question Number	Scheme	Marks
5.	<p style="text-align: center;">Never Sometimes Regularly Totals</p> <p style="text-align: center;">Males 30 132 78 240</p> <p style="text-align: center;">Females 26 143 91 260</p> <p style="text-align: center;">56 275 169 500</p> <p>H_0 : No association (independent) between gender and exercise</p> <p>H_1 : association (not independent) between gender and exercise</p> <p>Expected Values</p> <p style="text-align: center;">Never Sometimes Regularly Totals</p> <p style="text-align: center;">Males 26.88 132 81.12 240</p> <p style="text-align: center;">Females 29.12 143</p>	<p>M1 convert % to freq A1 (26, 91, 30, 132)</p> <p>A1 (143, 78)</p> <p>B1</p> <p>B1</p> <p>M1 A1 at least 3sf</p> <p>B1; B1√</p> <p>M1 A1</p> <p>A1√</p> <p style="text-align: right;">(12)</p>

EDEXCEL STATISTICS S3 (6685) – JUNE 2004

PROVISIONAL MARK SCHEME

87.88

260

56

275

169

500

$$\alpha = 0.05 \quad \nu = 2; \quad CV \chi^2 > \underline{5.991}$$

$$\sum \frac{(O - E)^2}{E} \text{ OR } \sum \frac{O^2}{E} - N = 0.9271$$

answers in range 0.90 – 0.95

Not in critical region – no evidence of association between
gender and exercise

Question Number	Scheme	Marks																				
6a)	$X \sim B(3, 1/6)$	bino 3, 1/6 B1 B1 (2)																				
b)	<table border="0" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; width: 5%;">X</th> <th style="text-align: left; width: 20%;">Prob</th> <th style="text-align: left; width: 20%;">Expected freq</th> <th style="width: 55%;"></th> </tr> </thead> <tbody> <tr> <td>0</td> <td>$\left(\frac{5}{6}\right)^3$</td> <td>144.68</td> <td>prob – must show working and use B(3,p) or may be implied by correct answer</td> </tr> <tr> <td>1</td> <td>$3 \times \left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right)$</td> <td>86.81</td> <td>expected</td> </tr> <tr> <td>2</td> <td>$3 \times \left(\frac{5}{6}\right) \left(\frac{1}{6}\right)^2$</td> <td>17.36</td> <td>awrt 145,86.8,17.4,1.15/1.16</td> </tr> <tr> <td>3</td> <td>$\left(\frac{1}{6}\right)^3$</td> <td>1.15 (1.16)</td> <td></td> </tr> </tbody> </table> <p>H_0 : Binomial model is a good fit H_1 : Binomial model is not a good fit</p> <p>both, no ditto</p> <p>Amalgamate 3 with another group</p> <p>$\alpha = 0.01 \quad v = 2 \quad ; \quad CR \chi^2 > \underline{9.210}$</p> <p>$\sum \frac{(O - E)^2}{E} \text{ OR } \sum \frac{O^2}{E} - N = 8.6894...$ answers in range 8.67 – 8.70 or</p> <p>Evidence that Binomial is a good model.</p>	X	Prob	Expected freq		0	$\left(\frac{5}{6}\right)^3$	144.68	prob – must show working and use B(3,p) or may be implied by correct answer	1	$3 \times \left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right)$	86.81	expected	2	$3 \times \left(\frac{5}{6}\right) \left(\frac{1}{6}\right)^2$	17.36	awrt 145,86.8,17.4,1.15/1.16	3	$\left(\frac{1}{6}\right)^3$	1.15 (1.16)		M1 M1 B2 (-1 ee) B1 M1 B1 ; B1√ M1 A1 A1√ (11)
X	Prob	Expected freq																				
0	$\left(\frac{5}{6}\right)^3$	144.68	prob – must show working and use B(3,p) or may be implied by correct answer																			
1	$3 \times \left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right)$	86.81	expected																			
2	$3 \times \left(\frac{5}{6}\right) \left(\frac{1}{6}\right)^2$	17.36	awrt 145,86.8,17.4,1.15/1.16																			
3	$\left(\frac{1}{6}\right)^3$	1.15 (1.16)																				

Question Number	Scheme	Marks
6.c)	<p>Estimate p Degrees of freedom reduced by 1</p> <p><u>Special case</u></p> <p>Use of B(3,0.192) in part (b)</p> <p>Expected frequencies</p> <p>131.8785 94.01242 22.339 1.769</p> <p>H_0 : Binomial model is a good fit H_1 : Binomial model is not a good fit</p> <p>Amalgamate 3 with another group</p> <p>$\alpha = 0.01$ $v = 1$; CR $\chi^2 > 6.635$ $\sum \frac{(O - E)^2}{E}$ OR $\sum \frac{O^2}{E} - N$ in range 5.45 -5.50</p> <p>Evidence that Binomial is a good model.</p>	<p>B1 B1 (2)</p> <p>M1 M1 B0 B1 M1 B1 ; B1√ M1 A1 A1√ (11)</p> <p>both, no ditto</p>

Question Number	Scheme	Marks
7a)	$E(D) = E(A) - 3E(B) + 4E(C)$ $= 20$ $\text{Var}(D) = \text{Var}(A) + 9\text{Var}(B) + 16\text{Var}(C)$ $= 341$ $P(D < 44) = P\left(z < \frac{44 - 20}{\sqrt{341}}\right)$ $= P(z < 1.30)$	M1 A1 M1 M1 A1 M1, A1√ A1
b)	$= \underline{0.9032}$ $E(X) = 20$ $\text{Var}(X) = \text{Var}(A) + 3\text{Var}(B) + 16\text{Var}(C)$ $= 287$ $P(X > 0) = P\left(z > \frac{-20}{\sqrt{287}}\right)$ $= P(z > -1.18)$ $= 0.8810$	A1 (9) B1 M1 M1 A1 M1 A1 A1 (7)

Publication

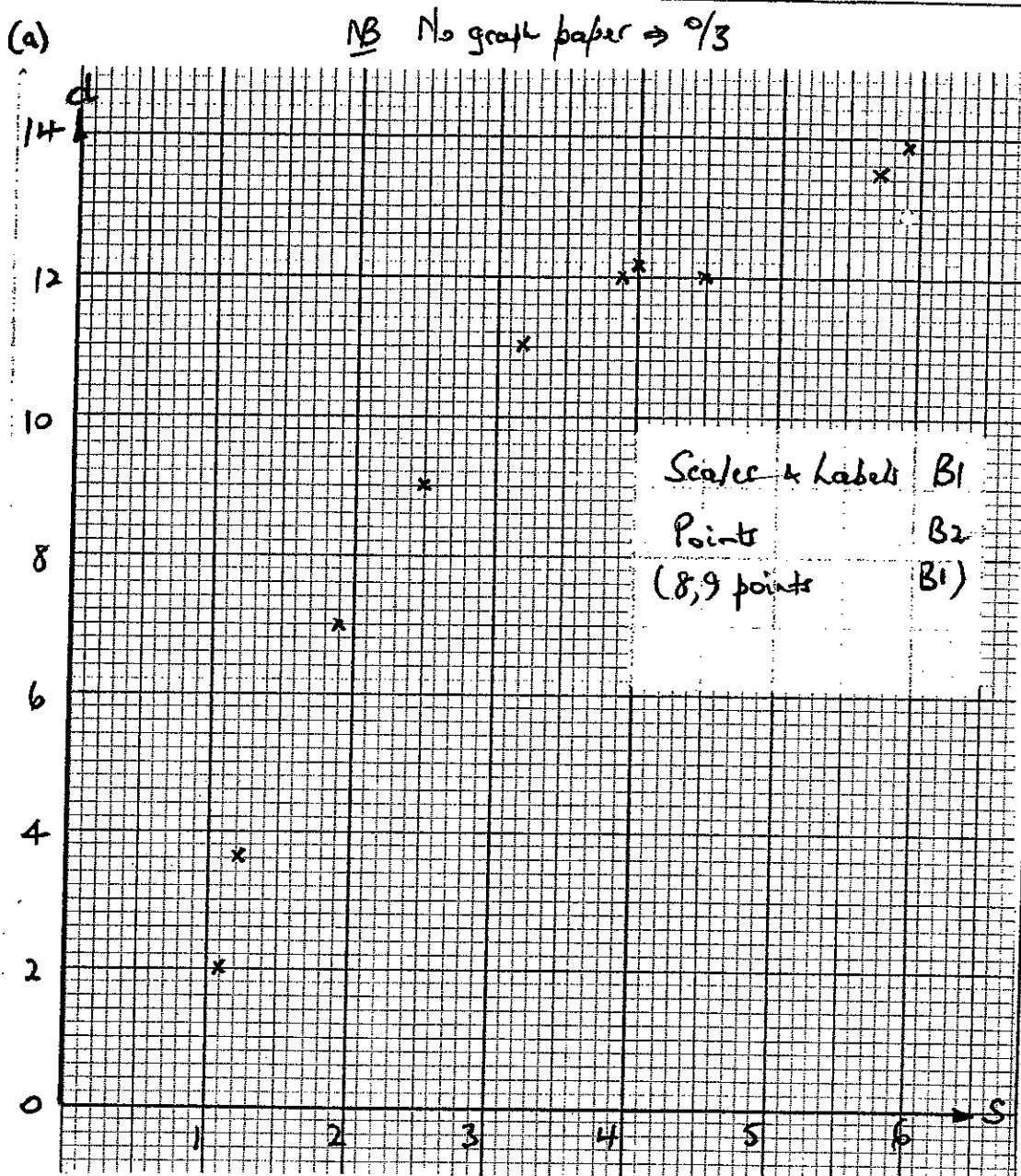
June 2005
6685 Statistics S3
Mark Scheme

FINAL

Question Number	Scheme	Marks
1.	<p>(a) Population divides into <u>mutually exclusive</u>; groups <u>distinct</u> strata</p> <p>(b) <u>Advantages</u></p> <ul style="list-style-type: none"> - enables fieldwork to be done quickly - costs kept to a minimum - administration is relatively easy <p><u>Disadvantages</u></p> <ul style="list-style-type: none"> - non-random so not possible to estimate sampling error - subject to possible interviewer bias - non-response not recorded 	<p>B1; B1 (2)</p> <p>Any ONE B1</p> <p>Any ONE B1 (2)</p>
2.	<p>$X \sim N(10, 3^2) \therefore \bar{X} \sim N(10, \frac{9}{5})$ can be implied 10; 9/5</p> <p>$P(7 \leq \bar{X} \leq 10) = P\left(\frac{7-10}{\sqrt{\frac{9}{5}}} < Z < 0\right)$ Standardising with 10 & their σ</p> <p>$= P(-2.236 < Z < 0)$ 2.236</p> <p>$= \Phi(0) - [1 - \Phi(2.24)]$</p> <p>$= \underline{0.4875}$</p>	<p>B1; B1</p> <p>M1 A1</p> <p>A1</p> <p>M1 (p < 0.5)</p> <p>A1 (6)</p>

Question Number	Scheme				Marks																									
3.	<table border="1"> <thead> <tr> <th></th> <th>No action</th> <th>Remove diseased branches</th> <th>Spray with Chemicals</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Tree died within 1 year</td> <td>10 (7)</td> <td>5 (7)</td> <td>6 (7)</td> <td>21</td> </tr> <tr> <td>Survived 1-4 years</td> <td>5 (7)</td> <td>9 (7)</td> <td>7 (7)</td> <td>21</td> </tr> <tr> <td>Survived > 4 years</td> <td>5 (6)</td> <td>6 (6)</td> <td>7 (6)</td> <td>18</td> </tr> <tr> <td>Totals</td> <td>20</td> <td>20</td> <td>20</td> <td>60</td> </tr> </tbody> </table>					No action	Remove diseased branches	Spray with Chemicals	Total	Tree died within 1 year	10 (7)	5 (7)	6 (7)	21	Survived 1-4 years	5 (7)	9 (7)	7 (7)	21	Survived > 4 years	5 (6)	6 (6)	7 (6)	18	Totals	20	20	20	60	
	No action	Remove diseased branches	Spray with Chemicals	Total																										
Tree died within 1 year	10 (7)	5 (7)	6 (7)	21																										
Survived 1-4 years	5 (7)	9 (7)	7 (7)	21																										
Survived > 4 years	5 (6)	6 (6)	7 (6)	18																										
Totals	20	20	20	60																										
$\frac{RT \times CT}{GT}$ $\frac{6 \times 7}{3 \times 6}$					M1 A1 A1																									
<p> H_0: Treatment & Survival are independent (not associated) H_1: Treatment & Survival are not independent (associated) </p>					BI both																									
$\alpha = 0.05$																														
$L = (3-1) \times (3-1) = 4$					BI																									
CR: $\chi^2 > 9.488$					BI ✓																									
$\sum \frac{(O-E)^2}{E} = \frac{9}{7} + \frac{4}{7} + \frac{1}{7} + \frac{4}{7} + \frac{4}{7} + 0 + \frac{1}{6} + 0 + \frac{1}{6}$ $= 3.47619 \dots$ <p style="text-align: right; margin-right: 50px;"> Use of $\sum \frac{(O-E)^2}{E}$ Any 2 values AWT 3.48 </p>					M1 A1 A1																									
<p>Since 3.47619... is NOT in the critical region (ie < 9.488) there is insufficient evidence to reject H_0.</p>																														
<p>There is no evidence of association between treatment and length of survival.</p>					Comparison Conclusion M1 A1 ✓ (11)																									

4



(3)

(b) linear association between s and d

B1 (1)

(c) $S_{SS} = 141.51 - \frac{33.9^2}{10} = 26.589$; $S_{dd} = 152.444$; $S_{sd} = 59.524$

B1; B1; B1 (3)

(d) $r = \frac{59.524}{\sqrt{152.444 \times 26.589}}$
 $= 0.93494\dots$

M1

AWRT 0.935

A1 (2)

	<p>(e) $H_0: \rho = 0$; $H_1: \rho > 0$</p> <p>Critical Value at 1% = 0.7155</p> <p>Reject H_0; Levels of serum & disease are positively correlated</p> <p>(f) linear correlation significant is ^{but} scatter diagram looks non-linear.</p>	<p>B1</p> <p>B1</p> <p>B1 (3)</p> <p>B1 (1)</p>																							
<p>5.</p>	<p>H_0: Poisson distribution is a suitable model both</p> <p>H_1: Poisson distribution is not a suitable model</p> $\hat{\lambda} = \frac{(0 \times 99) + (1 \times 65) + \dots + (4 \times 2)}{200} = \frac{153}{200} = 0.765$ <p>Using $P(X=x) = \frac{0.765^x e^{-0.765}}{x!}$ where X represents the number of restarts given $200 \times P(X=x)$</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;">X</th> <th style="text-align: center;">Observed Frequency</th> <th style="text-align: center;">Expected Frequency</th> <th></th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">0</td> <td style="text-align: center;">99</td> <td style="text-align: center;">93.06678...</td> <td></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">65</td> <td style="text-align: center;">71.19604...</td> <td style="text-align: center;">0, 1, 2</td> </tr> <tr> <td style="text-align: center;">2</td> <td style="text-align: center;">22</td> <td style="text-align: center;">27.23250...</td> <td></td> </tr> <tr> <td style="text-align: center;">3</td> <td style="text-align: center;">12</td> <td style="text-align: center;">6.94428...</td> <td rowspan="2" style="vertical-align: middle;">} 8.50468</td> </tr> <tr> <td style="text-align: center;">≥ 4</td> <td style="text-align: center;">2</td> <td style="text-align: center;">1.56040...</td> </tr> </tbody> </table> <p>$\chi^2 = 4 - 1 - 1 = 2$; CR: $\chi^2 > 5.991$ from Poisson</p> <p>$\chi^2 = 4 - 1 = 3$; CR: $\chi^2 > 7.815$ from Poisson (0.765)</p> <p>$\sum \frac{(O-E)^2}{E} = 5.47368...$ OR $\sum \frac{(O-E)^2}{E}$</p> <p>5.47 is not in the critical region.</p> <p>Number of computer failures per day can be modelled by a Poisson distribution</p>	X	Observed Frequency	Expected Frequency		0	99	93.06678...		1	65	71.19604...	0, 1, 2	2	22	27.23250...		3	12	6.94428...	} 8.50468	≥ 4	2	1.56040...	<p>B1</p> <p>M1 A1</p> <p>M1</p> <p>A1, A1 (-1e.2.)</p> <p>A1</p> <p>B1; B1✓</p> <p>M1</p> <p>A1</p> <p>A1✓ (12)</p>
X	Observed Frequency	Expected Frequency																							
0	99	93.06678...																							
1	65	71.19604...	0, 1, 2																						
2	22	27.23250...																							
3	12	6.94428...	} 8.50468																						
≥ 4	2	1.56040...																							

<p>6.</p>	<p>(a) Let X represent repair time $\therefore \sum x = 1435 \quad \therefore \bar{x} = \frac{1435}{5} = \underline{287}$ $\sum x^2 = 442575 \quad \therefore s^2 = \frac{1}{4} \left\{ 442575 - \frac{1435^2}{5} \right\}$ $= \underline{7682.5}$</p> <p>(b) $P(\mu - \bar{x} < 20) = 0.95$ $\therefore \frac{20}{\sigma/\sqrt{n}} = 1.96$ $\therefore n = \frac{1.96^2 \sigma^2}{20^2} = \frac{1.96^2 \times 100^2}{400} = \underline{96.04}$ $\therefore \underline{\text{Sample size } (\geq) 97 \text{ required}}$</p>	<p>BI MIAI AI (4) MIAI MI BI MIAI AI MI AI AI (6)</p>
<p>7.</p>	<p>Let $W = C_1 - C_2$ is $W = C_1 + C_2 \Rightarrow$ MIA or MI only (a) $\therefore W \sim N(0, 16)$ Normal $\therefore P(W > 6) = 2P(W > 6)$ $0; 16$ $= 2 \times P\left(Z > \frac{6-0}{\sqrt{16}}\right)$ is $W = C - L$ treat as MR $\text{Prob} = 0.4846$ $= 2 \times P(Z > 1.5)$ Standardizing, their σ $= 2 \times (1 - 0.9332) = \underline{0.1336}$</p> <p>(b) Let $W = C - L$ $\therefore W \sim N(5, 25)$ $5; 25$ $P(W > 0) = P\left(Z > \frac{0-5}{\sqrt{25}}\right)$ $= P(Z < 1)$ $= \underline{0.8413}$</p>	<p>MI AI; MI MI MI AI (6) BI; BI MIAI MI ($p > 0.5$) AI (6)</p>

$$(g) \text{ Let } W = C_1 + \dots + C_{24} + B$$

$$\therefore E(W) = 24 \times 350 + 100 = \underline{8500}$$

$$\text{Var}(W) = 24 \times 8 + 2^2 = \underline{196}$$

$$P(8510 \leq W \leq 8520) = P\left(\frac{8510 - 8500}{\sqrt{196}} \leq Z \leq \frac{8520 - 8500}{\sqrt{196}}\right)$$

$$= P(0.714 \leq Z \leq 1.428) \text{ AwRT}$$

$$= 0.9236 - 0.7611$$

$$= \underline{0.1625}$$

$$0.61 - 0.163$$

(d) All random variables are independent.

BI

BI

MI

AI/ AI/

AI (6)

BI (1)

Y.E. Stephens
13/06/05

Question Number	Scheme	Marks
1.	<p>Total in School = $(15 \times 30) + 150 = 600$</p> <p>random sample of $\frac{30}{600} \times 40$ $= \underline{2}$ from each of the 15 classes</p> <p>random sample of $\frac{150}{600} \times 40$ $= \underline{10}$ from sixth form;</p> <p>Label the boys in each class from 1 – 15 and the girls from 1 – 15. use random numbers to select 1 girl and 1 boy</p> <p>Label the boys in the sixth form from 1 – 75 and the girls from 1 – 75. use random numbers to select <u>5</u> different boys and 5 different girls.</p>	<p>B1</p> <p>(Use of $\frac{40}{\text{their } 600}$) M1 A1</p> <p>Either</p> <p>A1</p> <p>B1 B1</p> <p>B1</p> <p>(7)</p>

Question Number	Scheme	Marks
2. (a)	$E(R) = 20 + 10 = 30$	B1 (1)
(b)	$\text{Var}(R) = 4 + 0.84, = 4.84$	M1, A1 (2)
(c)	$R \sim N(30, 4.84)$ (Use of normal with their (a),(b)) $P(28.9 < R < 32.64) = P(R < 32.64) - P(R < 28.9)$ $= P\left(Z < \frac{32.64 - 30}{2.2}\right) - P\left(Z < \frac{28.9 - 30}{2.2}\right)$ Stand their σ and μ $= P(Z < 1.2) - P(Z < -0.5)$ $= 0.8849 - (1 - 0.6915)$ Correct area $= 0.8849 - 0.3085 = 0.5764$ (accept AWRT 0.576)	B1ft M1 A1, A1 M1 A1 (6)

<p>3. (a)</p>	$\hat{\mu} = \frac{82 + 98 + 140 + 110 + 90 + 125 + 150 + 130 + 70 + 110}{10}$ $= 110.5$ $\hat{\sigma}^2 = \frac{1}{9}(128153 - 10 \times 110.5^2)$ $= 672.28$	<p>M1 A1 B1 M1 A1 (5)</p>
<p>(b)</p>	<p>95% confidence limits are</p> $110.5 \pm 1.96 \times \frac{25}{\sqrt{10}}$ <p>95% conf. lim. = AWRT(95, 126)</p>	<p>(condone use of 5 instead of 25) (for 1.96) M1 B1 A1√ A1 A1 (5)</p>
<p>(c)</p>	<p>Number of intervals = $\frac{95}{100} \times 15$</p> $= 14.25$	<p>(Allow 14 or 14.3 if method is clear) M1 A1 (2)</p>
		<p>12</p>

4.

H_0 : No association between gender and acceptance
 H_1 : gender and acceptance are associated

	Accept	Not accept	Total
Males	170 (180)	110 (100)	280
Females	280 (270)	140 (150)	420
Totals	450	250	700

Expected Values

B1

M1 A1

O	E	$\frac{(O - E)^2}{E}$
170	180	0.5556
110	100	1.0000
280	270	0.3704
140	150	0.6667

$$\sum \frac{(O - E)^2}{E} = 2.59 \text{ (Yates' 2.34)}$$

(Condone use of Yates')

M1 A1

$$\nu = 1; (5\%) = 3.841$$

B1; B1

$3.841 > 2.59$. There is insufficient evidence to reject H_0

There is no association between a persons gender and their acceptance (of the offer of a flu jab.)

M1
A1√

(9)

9

5. (a)	<p>μ_b = mean mark of boys, μ_g = mean mark of girls.</p> <p>$H_0 : \mu_b = \mu_g$ $H_1 : \mu_b \neq \mu_g$</p> $z = \frac{53 - 50}{\sqrt{\frac{144}{80} + \frac{144}{80}}}$ <p>= 1.58 Critical region $z \geq 1.96$ 1.58 < 1.96 insufficient evidence to reject H_0. No diff. between mean scores of boys and girls.</p>	<p>both</p> <p>B1</p> <p>M1 A1</p> <p>A1 B1 M1 A1</p> <p>(7)</p>
(b)	<p>$H_0 : \mu_b = \mu_g$ $H_1 : \mu_b < \mu_g$</p> $z = \frac{62 - 59}{\sqrt{\frac{36}{80} + \frac{36}{80}}}$ <p>= 3.16</p> <p>Critical region $z \geq 1.6449$ (accept 1.645) 3.16 > 1.6449 sufficient evidence to reject H_0. the mean mark for boys is less than the mean mark of the girls.</p>	<p>B1</p> <p>M1</p> <p>A1 B1 A1</p> <p>(5)</p>
(c)	<p>Girls have improved more than boys or girls performed better than boys after 1 year</p>	<p>B1</p> <p>(1)</p> <p>13</p>

6. (a)	$r = 27.07,$ $s = 18.04,$ $t = 0.11$ using tables or 0.12 using totals	M1 A1 B1 B1 ft (4)
(b)	<p> H_0 : A Poisson model $Po(2)$ is a suitable model. both H_1 : A Poisson model $Po(2)$ is not a suitable model. </p> <p>Amalgamate data</p> $\sum \frac{(O - E)^2}{E} = 3.28 \text{ (awrt)}$ <p>$\nu = 6 - 1 = 5$</p> $\chi^2_5(5\%) = 11.070$ (follow through their degrees of freedom) <p>$3.25 < 11.070$ There is insufficient evidence to reject H_0, <u>$Po(2)$ is a suitable model.</u></p>	B1 M1 M1 A1 B1 B1ft A1ft (7)
(c)	<p>The expected values, and hence $\sum \frac{(O - E)^2}{E}$ would be different, and the degrees of freedom would be 1 less.</p>	B1 B1 (2)
		13

7. (a)	The variables cannot be assumed to be normally distributed	B1 (1)																																			
(b)	<table border="1" data-bbox="225 232 1075 434"> <thead> <tr> <th></th> <th>20-29</th> <th>30-39</th> <th>40-49</th> <th>50-59</th> <th>60-69</th> <th>70+</th> </tr> </thead> <tbody> <tr> <td>Rank x</td> <td>5</td> <td>6</td> <td>4</td> <td>3</td> <td>1</td> <td>2</td> </tr> <tr> <td>Rank y</td> <td>6</td> <td>5</td> <td>4</td> <td>1</td> <td>3</td> <td>2</td> </tr> <tr> <td>d</td> <td>1</td> <td>1</td> <td>0</td> <td>2</td> <td>2</td> <td>0</td> </tr> <tr> <td>d^2</td> <td>1</td> <td>1</td> <td>0</td> <td>4</td> <td>4</td> <td>0</td> </tr> </tbody> </table> <p data-bbox="225 501 1294 546">$\sum d^2 = 10$ (follow through their rankings)</p> <p data-bbox="225 562 1517 645">$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{60}{210} = 0.714$ ($\frac{5}{7}$ or awrt 0.714)</p>		20-29	30-39	40-49	50-59	60-69	70+	Rank x	5	6	4	3	1	2	Rank y	6	5	4	1	3	2	d	1	1	0	2	2	0	d^2	1	1	0	4	4	0	M1 A1 dM1 (depends on ranking attempt) A1 ft M1 A1 (6)
	20-29	30-39	40-49	50-59	60-69	70+																															
Rank x	5	6	4	3	1	2																															
Rank y	6	5	4	1	3	2																															
d	1	1	0	2	2	0																															
d^2	1	1	0	4	4	0																															
(c)	<p data-bbox="225 725 443 792">$H_0: \rho = 0$ $H_1: \rho \neq 0$ (or $\rho > 0$)</p> <p data-bbox="225 824 724 860">$n = 6 \Rightarrow 5\%$ critical value = 0.8857 (or 0.8286)</p> <p data-bbox="225 891 1102 994">0.714 < 0.8857 No evidence to reject H_0; No evidence of correlation between deaths from pneumoconiosis and lung cancer.</p>	B1 B1 B1 ✓ M1 A1 (5)																																			

June 2006
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
1 (a)	<p><u>Advantages:</u></p> <ul style="list-style-type: none"> - does not require the existence of a ^{Sampling frame} population list - <u>field work can be done quickly</u> as representative sample can be achieved with a small sample size - costs kept to a minimum (<u>cheaply</u>) - administration relatively <u>easy</u> - non-response not an issue <p><u>Disadvantages:</u></p> <ul style="list-style-type: none"> - not possible to estimate sampling errors - interviewer choice and may not be able to judge easily / <u>may lead to bias</u> - non-response not recorded - non-random process 	<p style="text-align: right;">any one B1</p> <p style="text-align: right;">any one B1</p> <p style="text-align: right;">(2)</p>
(b)	<p><u>Advantages:</u></p> <ul style="list-style-type: none"> - <u>random process</u> so possible to <u>estimate sampling errors</u> - free from <u>bias</u> <p><u>Disadvantages:</u></p> <ul style="list-style-type: none"> - not suitable when sample size is large - <u>sampling frame required</u> which <u>may not exist</u> or may be difficult to construct for a large population. 	<p style="text-align: right;">any one B1</p> <p style="text-align: right;">any one B1 (2)</p> <p style="text-align: right;">TOTAL 4</p>

NO REPETITION / OPPOSITES

Question Number	Scheme	Marks
2 (a)	$\bar{X} \sim N(90, \frac{\Sigma^2}{100}) \text{ i.e. } N_9(90, 0.25)$ <p>Application of <u>central limit theorem</u> as (sample large)</p>	M1A1 B1 (3)
3 (a)	<p>(b) $P(\bar{X} \geq 91) = 1 - P(Z < \frac{91-90}{0.5})$ stand.</p> $= 1 - P(Z < 2)$ $= 1 - 0.9772$ $= 0.0228 \quad \text{aust } 0.0228$	M1A1 A1 (3) TOTAL 6
3 (a)	<p>$H_0: \mu_A = \mu_B, H_1: \mu_A \neq \mu_B$ $M_1, M_2 \text{ OK both}$</p> $s_e = \sqrt{\frac{47^2}{70} + \frac{23^2}{90}} (= \sqrt{37.43492...})$ <p>Test statistic is $\pm \frac{198-201}{s_e} = \pm 0.4903$ <small>aust 0.99 M1A1 probab aust 0.312 B1 probab cv 0.025</small></p> $cv = (\pm) 1.96$ <p>Insufficient evidence to reject H_0, no significant difference between the mean cholesterol content of the two samples. (require correct comparison for FT) <small>content required.</small></p>	B1 M1A1 M1A1 B1 A1 ✓ (7)
(b)	<ul style="list-style-type: none"> - require 1 egg from each of 70 chickens of diet A to ensure <u>independence</u>, similarly for diet B. - no chickens in common between the two samples to ensure <u>independence</u> - not same chickens on diet A and diet B because if it were we need to do a <u>paired analysis</u>. <p style="text-align: right;">Any 1</p> <p>not same chickens on diet A and diet B because if it were we need to do a paired analysis.</p>	B1, B1 (2) TOTAL 9

4.

Rank:

Shop	Distance	Price	d	d ²
A	1	9	8	64
B	2	7	5	25
C	3	10	7	49
D	4	6	2	4
E	5	4	1	1
F	6	8	2	4
G	7	2	5	25
H	8	1	7	49
I	9	5	4	16
J	10	3	7	49

ranking

M1

Reverse ranking on price, $\sum d^2 = 44$
Hairs

286 $\sum d^2$

M1, A1

(a)

$$r_s = 1 - \frac{6 \times 286}{10(100-1)} = -0.73 \text{ or } \frac{-11}{15} \text{ or } -0.733$$

or 0.733 for $\sum d^2 = 44$

M1 A1

(5)

(b)

$H_0: \rho = 0$

$H_1: \rho < 0$

($H_1: \rho > 0$ if reverse ranking)

$cv = -0.5636$

(0.5636)

B1

B1

B1

Reject H_0 , evidence there is a significant negative correlation between the price of an ice cream and the distance from a tourist attraction.

(Ice cream gets cheaper further from the tourist attraction)

B1

(4)

(-cv from correct table required) (positive in context)

TOTAL 9

5. $M = \text{wt of male worker}$ $M \sim N(78.5, 12.6^2)$
 $F = \text{wt of female worker}$ $F \sim N(62.0, 9.8^2)$

(a) $W = M_1 + \dots + M_7 + F_1 + \dots + F_8$
 $E(W) = 7 \times 78.5 + 8 \times 62.0 = 1045.50$ ^{awrt} 1050 M1A1
 $\text{Var}(W) = 7 \times 12.6^2 + 8 \times 9.8^2 = 1879.64$ 1880 M1A1 (4)

(b) Independent: (used in Variance formula) B1 (1)

(c) $P(W > 1090) = P\left(z > \frac{1090 - 1045.5}{\sqrt{1879.64}}\right)$ M1
 $= P(z > 1.03)$ ^{awrt 1.03} A1
 $= 1 - 0.8485$ 1 - H1
 $= \underline{0.1515}$ A1 (4)
 Awrt(0.152)

9

<p>6.</p>	<p>H_0 : No association between age and colour (independent)</p> <p>H_1 : Association between age and colour (Not independent)</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>O</th> <th>E</th> <th>$\frac{(O-E)^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>12</td> <td>10.08</td> <td>0.3657...</td> </tr> <tr> <td>6</td> <td>7.92</td> <td>0.4654...</td> </tr> <tr> <td>10</td> <td>9.52</td> <td>0.0242...</td> </tr> <tr> <td>7</td> <td>7.48</td> <td>0.0308...</td> </tr> <tr> <td>6</td> <td>8.4</td> <td>0.6857...</td> </tr> <tr> <td>9</td> <td>6.6</td> <td>0.8727...</td> </tr> </tbody> </table> <p>$\sum \frac{(O-E)^2}{E} = 2.4446...$</p> <p>$\nu = (3-1)(2-1) = 2, \chi^2 = 5.991$</p> <p>Insufficient evidence to reject H_0.</p> <p>No association between age and colour</p> <p>(cv for correct h/c for ft)</p>	O	E	$\frac{(O-E)^2}{E}$	12	10.08	0.3657...	6	7.92	0.4654...	10	9.52	0.0242...	7	7.48	0.0308...	6	8.4	0.6857...	9	6.6	0.8727...	<p>BI</p> <p>BI</p> <p>MIAI</p> <p>MIAI</p> <p>MIAI</p> <p>BI BI ✓</p> <p>AI ✓ (ii)</p> <p>TOTAL 11</p>
O	E	$\frac{(O-E)^2}{E}$																					
12	10.08	0.3657...																					
6	7.92	0.4654...																					
10	9.52	0.0242...																					
7	7.48	0.0308...																					
6	8.4	0.6857...																					
9	6.6	0.8727...																					
<p>7.(a)</p>	<p>$\bar{x} = \frac{500}{10} = 50$</p> <p>$s^2 = \frac{1}{9} (25001.74 - \frac{500^2}{10}) = 0.193$</p> <p>limits are 50 ± 1.966 $= (49.02, 50.98)$</p> <p>Confidence interval is $(50 - 2.5758 \times \frac{0.5}{\sqrt{10}}, 50 + 2.5758 \times \frac{0.5}{\sqrt{10}})$ $= (49.59273, 50.40727...)$</p> <p>use of estimate in (a) in (b) AND (c) assume MISREAD.</p>	<p>MIAI</p> <p>MIAIAI (5)</p> <p>MIBI</p> <p>AIAI (4)</p> <p>MIBIAN</p> <p>AIAI (5)</p> <p>TOTAL 14</p>																					

8 (a)	$B_7(5, 0.5)$	MIAI (2)																																												
(b)	<p>$H_0: B(5, 0.5)$ is a suitable model (good fit)</p> <p>$H_1: B(5, 0.5)$ is not a suitable model (not a good fit) ✓ for $\hat{p} = 0.466$.</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: left;">No. of heads</td> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> <td style="font-size: small;">100% (freq) For Bin, 1 correct = AI All correct = AI 3st or better</td> </tr> <tr> <td style="text-align: left;">Expected</td> <td style="text-align: center;">3.125</td> <td style="text-align: center;">15.625</td> <td style="text-align: center;">31.25</td> <td style="text-align: center;">31.25</td> <td style="text-align: center;">15.625</td> <td style="text-align: center;">3.125</td> <td></td> </tr> <tr> <td style="text-align: left;">Actual</td> <td style="text-align: center;">6</td> <td style="text-align: center;">18</td> <td style="text-align: center;">29</td> <td style="text-align: center;">34</td> <td style="text-align: center;">10</td> <td style="text-align: center;">3</td> <td></td> </tr> </table> <table style="width: 100%; border-collapse: collapse;"> <tr> <td></td> <td style="text-align: center;">0</td> <td style="text-align: center;">E</td> <td style="text-align: center;">$\frac{(O-E)^2}{E}$</td> </tr> <tr> <td style="text-align: left;">0 or 1</td> <td style="text-align: center;">24</td> <td style="text-align: center;">18.75</td> <td style="text-align: center;">1.47</td> </tr> <tr> <td style="text-align: left;">2</td> <td style="text-align: center;">29</td> <td style="text-align: center;">31.25</td> <td style="text-align: center;">0.162</td> </tr> <tr> <td style="text-align: left;">3</td> <td style="text-align: center;">34</td> <td style="text-align: center;">31.25</td> <td style="text-align: center;">0.242</td> </tr> <tr> <td style="text-align: left;">4 or 5</td> <td style="text-align: center;">13</td> <td style="text-align: center;">18.75</td> <td style="text-align: center;">1.763</td> </tr> </table> <p style="margin-left: 20px;">$\sum \frac{(O-E)^2}{E} = 3.6373$</p> <p style="margin-left: 20px;">$\chi^2_{(3)}(0.10) = 6.251$</p> <p style="margin-left: 20px;">Insufficient evidence to reject H_0 $B(5, 0.5)$ is a suitable model. No evidence that coins are biased</p> <p style="margin-left: 20px; font-size: small;">Ungrouped gives a wt 5.44, $\nu=5$, $\chi^2_5 = 9.236$ grouped gives a wt 5.44, $\nu=5$, $\chi^2_5 = 9.236$</p>	No. of heads	0	1	2	3	4	5	100% (freq) For Bin, 1 correct = AI All correct = AI 3st or better	Expected	3.125	15.625	31.25	31.25	15.625	3.125		Actual	6	18	29	34	10	3			0	E	$\frac{(O-E)^2}{E}$	0 or 1	24	18.75	1.47	2	29	31.25	0.162	3	34	31.25	0.242	4 or 5	13	18.75	1.763	<p>BI ✓</p> <p>MIAIAI</p> <p>MIAI</p> <p>MIAI</p> <p>BI ✓ BI ✓</p> <p>AI ✓</p> <p>(11)</p> <p>TOTAL 13</p>
No. of heads	0	1	2	3	4	5	100% (freq) For Bin, 1 correct = AI All correct = AI 3st or better																																							
Expected	3.125	15.625	31.25	31.25	15.625	3.125																																								
Actual	6	18	29	34	10	3																																								
	0	E	$\frac{(O-E)^2}{E}$																																											
0 or 1	24	18.75	1.47																																											
2	29	31.25	0.162																																											
3	34	31.25	0.242																																											
4 or 5	13	18.75	1.763																																											

Mark Scheme (Results)

Summer 2007

GCE

GCE Mathematics

Statistics S3 (6691)

June 2007
6691 Statistics S3
Mark Scheme

Question number	Scheme	Marks																																				
<p>1. (a)</p>	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th><i>A</i></th> <th><i>B</i></th> <th><i>C</i></th> <th><i>D</i></th> <th><i>E</i></th> <th><i>F</i></th> <th><i>G</i></th> <th><i>H</i></th> </tr> </thead> <tbody> <tr> <td><i>P</i> Rank</td> <td>2</td> <td>6</td> <td>4</td> <td>3</td> <td>1</td> <td>7</td> <td>8</td> <td>5</td> </tr> <tr> <td><i>Q</i> Rank</td> <td>2</td> <td>8</td> <td>1</td> <td>6</td> <td>3</td> <td>5</td> <td>7</td> <td>4</td> </tr> <tr> <td>d^2</td> <td>0</td> <td>4</td> <td>9</td> <td>9</td> <td>4</td> <td>4</td> <td>1</td> <td>1</td> </tr> </tbody> </table> <p style="text-align: right; margin-right: 20px;">$\sum d^2 = 32$</p> $r_s = 1 - \frac{6 \times 32}{8 \times (8^2 - 1)}$ $= \frac{13}{21} \text{ or AWRT } 0.619$		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>P</i> Rank	2	6	4	3	1	7	8	5	<i>Q</i> Rank	2	8	1	6	3	5	7	4	d^2	0	4	9	9	4	4	1	1	<p>M1A1</p> <p>M1A1</p> <p>M1</p> <p>A1 (6)</p>
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>																														
<i>P</i> Rank	2	6	4	3	1	7	8	5																														
<i>Q</i> Rank	2	8	1	6	3	5	7	4																														
d^2	0	4	9	9	4	4	1	1																														
<p>(b)</p>	<p>$H_0 : \rho = 0$ $H_1 : \rho > 0$ (ρ_s is OK) both</p> <p>r_s 1 tail 5% critical value is 0.6429 (Independent of their H_1)</p> <p>$0.619 < 0.6429$ or not significant</p> <p>So insufficient evidence of a positive correlation between judges competitor <u>is</u> justified</p> <p><u>Or</u></p>	<p>B1</p> <p>B1 (\pm is OK)</p> <p>M1</p> <p>A1f.t. (4)</p> <p style="text-align: center;">10</p>																																				
<p>(a)</p>	<p>1st M1 for attempting to rank both <i>P</i> and <i>Q</i>. 1st A1 for both correct (could be reversed) 2nd M1 for attempting d^2 2nd A1 for $\sum d^2 = 32$. 3rd M1 for correct use of formula for r_s</p>																																					
<p>(b)</p>	<p>M1 for a correct comparison or statement about significance (o.e.) Follow through their r_s provided $0 < r_s < 1$</p> <p>A1f.t. for a conclusion in context. Must mention judges or marks or competitor. If they use correlation they must say it is positive. Follow through their positive r_s with their positive c.v. and ignore hypotheses. So $r_s = 0.667$ they could say competitor's claim is not justified etc.</p>																																					
<p>S.C.</p>	<p><u>No ranking</u> Typical answer (-3.82) can get mark for use of r_s formula and hypotheses in (b) only</p> <p>(a) M0A0M0A0M1A0 (b) B1B1M0A0</p>																																					

Question number	Scheme	Marks						
2. (a)	<p>H_0 : Maths grades are independent of English grades <u>or</u> No association ...</p> <p>H_1 : Maths and English grades are dependent <u>or</u> There is an association ...</p> <p>Expected Frequencies e.g. $\frac{60 \times 40}{120} = 20$</p> <table border="1" data-bbox="823 432 1072 510"> <tr> <td>20</td> <td>27.5</td> <td>12.5</td> </tr> <tr> <td>20</td> <td>27.5</td> <td>12.5</td> </tr> </table> $\sum \frac{(O-E)^2}{E} = 2 \times \left(\frac{5^2}{20} + \frac{2.5^2}{27.5} + \frac{2.5^2}{12.5} \right), = 3.9545\dots \quad \text{AWRT } \underline{3.95} \text{ or } \underline{3.955}$ <p>$\nu = (3-1)(2-1) = 2; \quad \chi_2^2(10\%) \text{ c.v.} = 4.605$</p> <p>$3.95 < 4.605$ or not significant or do not reject H_0 (allow reject H_1)</p> <p>Insufficient evidence of an association between English and maths grades</p> <p><u>or</u> there is support for the Director's belief</p> <p><u>or</u> Student's grades in maths and English are independent</p>	20	27.5	12.5	20	27.5	12.5	<p>B1</p> <p>M1 A1</p> <p>M1, A1</p> <p>B1; B1</p> <p>M1</p> <p>A1 (9)</p> <p>B1 (1)</p> <p>10</p>
20	27.5	12.5						
20	27.5	12.5						
(a)	<p>1st B1 for both hypotheses in terms of independence or association and in context.</p> <p>Must mention Maths and English in at least one of the hypotheses.</p> <p>“relationship” or “correlation” or “connection” or “link” is B0</p> <p>1st M1 for some correct calculation seen</p> <p>1st A1 for all expected frequencies correct. Accept answers without formula seen.</p> <p>2nd M1 for some evidence seen of attempt to calculate test statistic.</p> <p>At least one correct term seen. Follow through their expected frequencies.</p> <p>2nd A1 for AWRT 3.95. Answers only please escalate!</p> <p>3rd M1 for correct comparison or statement – may be implied by correct conclusion.</p> <p>3rd A1 for conclusion in context using “association” or “independence” in connection with grades.</p> <p>Don't insist on seeing English or maths mentioned here.</p> <p>Use ISW for comments if a false statement and correct statement are seen.</p>							
(b)	<p>B1 If they just say expected frequencies are “small” they must go onto mention need to pool.</p>							

Question number	Scheme	Marks
3.	$H_0 : \mu = 18, \quad H_1 : \mu < 18$ $z = \frac{16.5 - 18}{\frac{3}{\sqrt{15}}} = -1.9364\dots$ <p>AWRT – 1.94</p> <p>5% one tail c.v. is $z = (-) 1.6449$ or probability (AWRT 0.026) $(\pm) 1.6449$</p> <p>$-1.94 < -1.6449$ <u>or</u> significant <u>or</u> reject H_0 <u>or</u> in critical region</p> <p>There is evidence that the (mean) time to complete the puzzles has reduced</p> <p><u>Or</u> Robert is getting faster (at doing the puzzles)</p>	<p>B1, B1</p> <p>M1, A1</p> <p>B1</p> <p>M1</p> <p>A1f.t.</p>
	<p>1st & 2nd B1 must see \sim and 18</p> <p>1st M1 for attempting test statistic, allow \pm. Or attempt at critical value for $\bar{X} : \mu - z \times \frac{3}{\sqrt{15}}$</p> <p>1st A1 for AWRT – 1.94. Allow use of $z = +1.94$ to score M1A1. Or critical value = AWRT 16.7.</p> <p>3rd B1 for AWRT 0.026 (i.e. correct probability only) or ± 1.6449. (May be seen in cv formula)</p> <p>2nd M1 for correct comparison or statement relating their test statistic and 1.6449 or their probability and 0.05. Ignore their hypotheses if any or assume they were correct.</p> <p>2nd A1f.t. for conclusion in context which refers to “speed” or “time”. Depends only on previous M</p>	<p>7</p>

Question number	Scheme	Marks																								
<p>4. (a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p>	$\frac{0 \times 17 + 1 \times 31 + \dots}{17 + 31 + \dots} = \left(\frac{200}{100} = 2 \right), \quad \hat{p} = \frac{2}{20} = 0.1 \text{ (Accept } \frac{2}{20} \text{ or 2 per 20)}$ <p>e.g. $r = 100 \times \binom{20}{2} (0.1)^2 (0.9)^{18}$</p> <p style="text-align: right;">$r = 28.5, s = \text{AWRT } 9$</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>x</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>≥ 4</td> </tr> <tr> <td>O_i</td> <td>17</td> <td>31</td> <td>19</td> <td>14</td> <td>19</td> </tr> <tr> <td>E_i</td> <td>12.2</td> <td>27.0</td> <td>28.5</td> <td>19.0</td> <td>13.3</td> </tr> <tr> <td>$\frac{(O-E)^2}{E}$</td> <td>1.89</td> <td>0.59</td> <td>3.17</td> <td>1.32</td> <td>2.44</td> </tr> </table> <p style="text-align: center;">$\sum \frac{(O-E)^2}{E} = \text{AWRT } 9.4$</p> <p>$v = 5 - 2 = 3, \quad \chi_3^2(5\%) = 7.815$</p> <p>$H_0$: Binomial distribution is a good/suitable model/fit [Condone: B(20, 0.1) is...]</p> <p>H_1 : Binomial distribution is not a suitable model both</p> <p>(Significant result) Binomial distribution is not a suitable model</p>	x	0	1	2	3	≥ 4	O_i	17	31	19	14	19	E_i	12.2	27.0	28.5	19.0	13.3	$\frac{(O-E)^2}{E}$	1.89	0.59	3.17	1.32	2.44	<p>M1, A1 (2)</p> <p>M1</p> <p>A1, A1 (3)</p> <p>M1</p> <p>M1A1c.a.o.</p> <p>B1ft, B1ft</p> <p>B1</p> <p>A1cao (7)</p> <p>B1ft (1)</p> <p style="text-align: center;">13</p>
x	0	1	2	3	≥ 4																					
O_i	17	31	19	14	19																					
E_i	12.2	27.0	28.5	19.0	13.3																					
$\frac{(O-E)^2}{E}$	1.89	0.59	3.17	1.32	2.44																					
<p>(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p>	<p>M1 for attempt to find mean or \hat{p} (as printed or better). The 0.1 must be seen in part (a).</p> <p>M1 for correct expression for r or s using the binomial distribution. Follow through their \hat{p}.</p> <p>1st M1 for some pooling (accept $x \geq 5$, obs.freq. ...14, 9, 10 and exp.freq. 19.0, s, 4.3) 2nd M1 for calculation of test statistic (N.B. $x \geq 5$ gives 14.5). One correct term seen. 1st B1ft for number of classes – 2 (N.B. $x \geq 5$ will have $6 - 2 = 4$) 2nd B1ft for the appropriate tables value, ft their degrees of freedom. (NB $\chi_4^2(5\%) = 9.488$) 3rd B1 (for hypotheses) allow just “$X \sim B(20, 0.1)$” for null etc. 2nd A1 for correctly rejecting Binomial model. No ft and depends on 2nd M1.</p> <p>B1ft for independence or constant probability – must mention defective items or defectives Follow through their conclusion in (c). So if they do not reject they may say “defectives occur with probability 0.1”. Stating the value implies constant probability.</p>																									

Question number	Scheme	Marks
5. (a)	$\hat{\mu} = \bar{x} = \frac{361.6}{80}, = \underline{4.52}$ $\hat{\sigma}^2 = s^2 = \frac{1753.95 - 80 \times \bar{x}^2}{79} = (1.51288\dots)$	M1, A1 M1A1ft A1 (5)
(b)	$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A > \mu_B$ $z = \frac{4.52 - 4.06}{\sqrt{\frac{1.51\dots}{80} + \frac{2.50}{60}}} = \left(\frac{0.46}{\sqrt{0.0605\dots}} \right)$ $= (+) 1.8689\dots$	A1 B1 B1 Denominator M1 z dM1 A1 B1
(c)	One tail c.v. is $z = 1.6449$ (AWRT 1.645 or probability AWRT 0.0307 or 0.0308) (significant) there is evidence that diet A is better than diet B <u>or</u> evidence that (mean) weight lost in first week using diet A is more than with B	B1 A1ft (7)
(d)	CLT enables you to assume that \bar{A} and \bar{B} are normally distributed	B1 (1)
(d)	Assumed $\sigma_A^2 = s_A^2$ and $\sigma_B^2 = s_B^2$ (either)	B1 (1)
14		
(a)	2^{nd} M1 for a correct attempt at s or s^2 , A1ft for correct expression for s^2 , ft their mean. N.B. $\sigma_n^2 = 1.49\dots$ so $\frac{80}{79} \times 1.49\dots$ is M1A1ft	
(b)	1^{st} B1 can be given for $\mu_1 = \mu_2$, but 2^{nd} B1 must specify which is A or B . 1^{st} M1 for the denominator, follow through their 1.51. Must have square root can condone 2.50^2 but $\sqrt{\frac{1.51^2}{80} + \frac{2.50^2}{60}}$ is M0. Allow $\sqrt{\frac{1.51}{79} + \frac{2.50}{59}}$ leading to AWRT 1.85 to score M1M1A0 in (b) and can score in (d).	
(c)	2^{nd} dM1 for attempting the correct test statistic, dependent on denominator mark	
(d)	1^{st} A1 for AWRT ± 1.87 , may be implied by a correct probability.	
(d)	2^{nd} A1ft ft their test statistic vs their cv only if H_1 is correct and both Ms are scored	
(c)	B1 for stating <u>either</u> \bar{A} or \bar{B} (but not A or B) are normally distributed	
(d)	B1 for either, can be stated in words in terms of variances or standard deviations.	

Question number	Scheme	Marks
6.	$\bar{x} = \frac{1}{2}(123.5 + 154.7) = \underline{139.1}$ <p style="text-align: right;">2.5758</p> <p>"their 2.5758" $\frac{\sigma}{\sqrt{n}} = 154.7 - 139.1 = 15.6$</p> <p style="text-align: right;">AWRT 1.96</p> <p>"their 1.96" $\frac{\sigma}{\sqrt{n}} = \frac{15.6 \times 1.96}{2.5758} = (11.87\dots)$</p> <p>So 95% C.I. = $139.1 \pm 11.87\dots = (127.22\dots, 150.97\dots)$ AWRT <u>(127, 151)</u></p>	<p>B1</p> <p>B1</p> <p>M1</p> <p>B1</p> <p>M1</p> <p>A1</p> <p style="text-align: right;">6</p>
	<p>1st B1 for mean = 139.1 only</p> <p>1st M1 for UL – mean or mean – LL set equal to z value times standard error or some equivalent expression for standard error. Follow through their 2.5758 provided a z value.</p> <p>May be implied by $\frac{\sigma}{\sqrt{n}} = 6.056\dots$ [N.B. $\frac{15.6}{2.3263} = 6.705\dots$]</p> <p>Condone poor notation for standard error if it is being used correctly to find CI.</p> <p>2nd M1 for full method for semi-width (or width) of 95% interval</p> <p>Follow through their z values for both M marks</p> <p>N.B. Use of 2.60 instead of 2.5758 should just lose 2nd B1 since it leads to AWRT (127, 151)</p>	

Question number	Scheme	Marks
7. (a)	<p>Let $X = L - 4S$ then $E(X) = 19.7 - 4 \times 4.9 = 0.1$ $\text{Var}(X) = \text{Var}(L) + 4^2 \text{Var}(S) = 0.5^2 + 16 \times 0.2^2$ $= 0.89$ $P(X > 0) = [P(Z > -0.10599\dots)]$ $=$ AWRT <u>(0.542 – 0.544)</u></p> <p>(b) $T = S_1 + S_2 + S_3 + S_4$ (May be implied by 0.16) $E(T) = 19.6$ $T \sim N(19.6, 0.16)$ $\text{Var}(T) = 0.16$ or 0.4^2</p> <p>(c) Let $Y = L - T$ $E(Y) = E(L) - E(T) = [0.1]$ $\text{Var}(Y) = \text{Var}(L) + \text{Var}(T) = [0.41]$ Require $P(-0.1 < Y < 0.1)$ $= P(Z < 0) - P(Z < -0.31\dots)$ or $0.5 - P(Z < -0.31\dots)$ or $P(Z < 0.31\dots) - P(Z < 0)$ $= 0.1217$ (tables) or $0.1226\dots$ (calc) AWRT <u>(0.122 – 0.123)</u></p>	<p>M1, A1 M1, M1 A1 M1 A1 (7)</p> <p>M1 B1 A1 (3)</p> <p>M1 M1 M1 M1 A1 (5)</p> <p style="text-align: right;">15</p>
(a)	<p>1st M1 for defining X and attempting $E(X)$ 1st A1 for 0.1. Answer only will score both marks. 2nd M1 for $\text{Var}(L) + \dots$ 3rd M1 for $\dots 4^2 \text{Var}(S)$. For those who don't attempt $L - 4S$ this will be their only mark in (a). 2nd A1 for 0.89 4th M1 for attempting a correct probability, correct expression and attempt to find, which should involve some standardisation: ft their $\sqrt{0.89}$ and their 0.1. If 0.1 is used for $E(X)$ answer should be > 0.5, otherwise M0.</p>	
(c)	<p>1st M1 for a correct method for $E(Y)$, ft their $E(T)$. 2nd M1 for a correct method for $\text{Var}(Y)$, ft their $\text{Var}(T)$. Must have +. 3rd M1 for dealing with the modulus and a correct probability statement. Must be modulus free. May be implied by e.g. $P(Z < \frac{0.2}{\sqrt{\text{their } 0.41}}) - 0.5$, or seeing both 0.378... (or 0.622...) <u>and</u> 0.5 4th M1 for correct expression for the correct probability, as printed or better. E.g. $0.5 + 0.378\dots$ is M0 A1 for AWRT in range.</p>	

Mark Scheme (Results)

June 2008

GCE

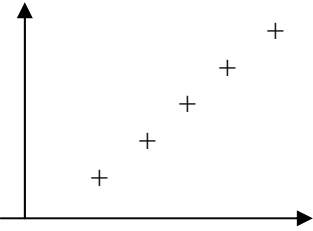
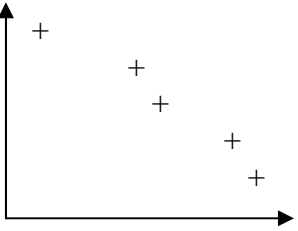
GCE Mathematics (669101)



June 2008
6691 Statistics S3
Mark Scheme

Question number	Scheme	Marks
1. (a)	$\bar{x} = \left(\frac{6046}{36} \right) = 167.94\dots$ $s^2 = \frac{1016338 - 36 \times \bar{x}^2}{35}$ $= 27.0253\dots$	awrt 168 B1 M1 awrt 27.0 A1 (3) (Accept 27)
(b)	99% Confidence Interval is: $\bar{x} \pm 2.5758 \times \frac{5.1}{\sqrt{36}}$ $= (165.755\dots, 170.133\dots)$	M1A1ft B1 awrt (166,170) A1 A1 (5) 8 marks
(a)	M1 for a correct expression for s^2 , follow through their mean, beware it is very “sensitive” $167.94 \rightarrow \frac{999.63..}{35} \rightarrow 28.56\dots$ $167.9 \rightarrow \frac{1483.24..}{35} \rightarrow 42.37\dots$ $168 \rightarrow \frac{274}{35} \rightarrow 7.82$ Use of 36 as the divisor (= 26.3...) is M0A0	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: auto;"> These would all score M1A0 </div>
(b)	M1 for substituting their values in $\bar{x} \pm z \times \frac{5.1 \text{ or } s}{\sqrt{36}}$ where z is a recognizable value from tables 1 st A1 follow through their mean and their z (to 2dp) in $\bar{x} \pm z \times \frac{5.1}{\sqrt{36}}$ Beware: $167.94 \pm 2.5758 \times \frac{5.1^2}{36} \rightarrow (166.07\dots, 169.8\dots)$ but scores B1M0A0A0A0 Correct answer only in (b) scores 0/5 2 nd & 3 rd A marks depend upon 2.5758 and M mark.	

Question number	Scheme	Marks												
2.	$\frac{115 \times 70}{217} = 37.0967\dots \quad \text{or} \quad \frac{1150}{31} \text{ (etc)} \quad \frac{1265}{31}, \frac{1020}{31}, \frac{1122}{31}$ <table border="1" data-bbox="245 376 1197 546"> <thead> <tr> <th>Expected (Obs)</th> <th>A</th> <th>S</th> <th>H</th> </tr> </thead> <tbody> <tr> <td>Boy</td> <td>37.1 (30)</td> <td>37.1 (50)</td> <td>40.8 (35)</td> </tr> <tr> <td>Girl</td> <td>32.9 (40)</td> <td>32.9 (20)</td> <td>36.2 (42)</td> </tr> </tbody> </table> <p>H_0 : There is no association between course and gender</p> <p>H_1 : There is some association between course and gender (both)</p> $\sum \frac{(O - E)^2}{E} = \frac{(37.1 - 30)^2}{37.1} + \frac{(32.9 - 40)^2}{32.9} + \dots + \frac{(36.2 - 42)^2}{36.2}$ <p>= 1.358 + 4.485 + 0.824 + 1.532 + 5.058 + 0.929 = 14.189... awrt 14.2</p> <p>$\nu = (3 - 1)(2 - 1) = 2$, $\chi^2_2(1\%)$ critical value is 9.210 (condone 9.21)</p> <p>Significant result or reject null hypothesis</p> <p>There is evidence of an association between course taken and gender</p> <p>[Correct answers only score full marks]</p>	Expected (Obs)	A	S	H	Boy	37.1 (30)	37.1 (50)	40.8 (35)	Girl	32.9 (40)	32.9 (20)	36.2 (42)	<p>M1</p> <p>A1A1</p> <p>B1</p> <p>M1A1ft</p> <p>A1</p> <p>B1, B1ft</p> <p>M1</p> <p>A1ft (11)</p> <p>11 marks</p>
Expected (Obs)	A	S	H											
Boy	37.1 (30)	37.1 (50)	40.8 (35)											
Girl	32.9 (40)	32.9 (20)	36.2 (42)											
ALT	$\sum \frac{O^2}{E} - N = \frac{30^2}{37.1} + \frac{40^2}{32.9} + \dots + \frac{42^2}{36.2} - 217$	M1A1ft												
	<p>1st M1 for some use of the $\frac{\text{row total} \times \text{col total}}{\text{grand total}}$ formula</p> <p>1st A1 for one correct row or one correct column of expected frequencies to nearest integer</p> <p>2nd A1 for all expected frequencies correct to awrt 1 dp (Allow exact fractions)</p> <p>1st B1 for hypotheses. Independence is OK. Must mention courses and gender at least once. Use of ρ or “correlation” is B0 but allow ISW.</p> <p>2nd M1 for an attempt to calculate test statistic. At least one correct expression, ft expected freq.</p> <p>3rd A1 follow through expected frequencies for at least 3 expressions</p> <p>3rd M1 for a correct statement relating their test statistic and their cv (may be implied by comment)</p> <p>5th A1 for a contextualised comment relating their test statistic and their cv. Ignore their H_0 or H_1 or assume that they were correct. Must mention courses and gender</p>													

Question number	Scheme	Marks																																
<p>3. (a)</p> <p>(i)</p> <p>(b)(i)</p> <p>(ii)</p>	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;">  <p>(i)</p> </div> <div style="text-align: center;">  <p>(ii)</p> </div> </div> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <td>Rank (Judge 1)</td> <td>1</td> <td>4</td> <td>2</td> <td>3</td> <td>5</td> <td>6</td> <td>7</td> </tr> <tr> <td>Rank (Judge 2)</td> <td>1</td> <td>2</td> <td>4</td> <td>3</td> <td>5</td> <td>7</td> <td>6</td> </tr> <tr> <td>d^2</td> <td>0</td> <td>4</td> <td>4</td> <td>0</td> <td>0</td> <td>1</td> <td>1</td> </tr> </tbody> </table> <p style="text-align: right;">$\sum d^2 = 10$</p> $r_s = 1 - \frac{6 \times 10}{7 \times (49 - 1)} = 1 - \frac{5}{28} = \frac{23}{28} \quad \text{or} \quad \text{awrt } \mathbf{0.821}$ <p>(ii) $H_0 : \rho = 0$ $H_1 : \rho > 0$ (Allow ρ_s) ($H_1 : \rho \neq 0$ scores B0)</p> <p>r_s 5% one tail critical value is 0.7143</p> <p>Significant result or reject null hypothesis</p> <p>There is evidence of a (positive) correlation between the judges <u>or</u> the judges agree</p>		A	B	C	D	E	F	G	Rank (Judge 1)	1	4	2	3	5	6	7	Rank (Judge 2)	1	2	4	3	5	7	6	d^2	0	4	4	0	0	1	1	<p>(i) B1</p> <p>(ii) B1B1 (3)</p> <p>M1M1</p> <p>M1A1</p> <p>M1A1 (6)</p> <p>B1,B1</p> <p>B1</p> <p>M1</p> <p>A1ft (5)</p> <p style="text-align: right;">14 marks</p>
	A	B	C	D	E	F	G																											
Rank (Judge 1)	1	4	2	3	5	6	7																											
Rank (Judge 2)	1	2	4	3	5	7	6																											
d^2	0	4	4	0	0	1	1																											
<p>(a) (i)</p> <p>(ii)</p> <p>(b)(i)</p> <p>(ii)</p>	<p>1st B1 for 5 or more points on a straight line of positive gradient</p> <p>2nd B1 for 4 or more points satisfying $-1 < r < 0$</p> <p>3rd B1 for 5 or more points of decreasing ranks not on a straight line</p> <p>1st M1 for attempting to rank one of the judges (at least 2 correct rankings)</p> <p>2nd M1 for ranking both (may be reversed) (at least 2 correct rankings)</p> <p>3rd M1 for attempting d^2.</p> <p>1st A1 for $\sum d^2 = 10$</p> <p>4th M1 for correct use of the r_s formula</p> <p>3rd B1 for the correct critical value - depends upon their $H_1 : \rho > 0$ needs 0.7143, $\rho \neq 0$, 0.7857</p> <p style="padding-left: 40px;">The H_1 may be in words so B0B1 is possible. If no H_1 award for 0.7143 only.</p> <p>5th M1 for a correct statement relating their r_s and their cv (may be implied by correct comment)</p> <p>3rd A1ft follow through their r_s and their cv. Comment in context. Must mention judges.</p> <p style="padding-left: 40px;">Don't insist on "positive" and condone it if they are using $\rho \neq 0$.</p>																																	

Question number	Scheme	Marks
4. (a)	$X = M_1 + M_2 + M_3 + M_4 \sim N(336, 22^2)$ $\mu = 336$ $\sigma^2 = 22^2 \text{ or } 484$ $P(X < 350) = P\left(Z < \frac{350 - 336}{22}\right)$ $= P(Z < 0.64)$ $=$	B1 B1 M1 A1 A1 (5)
(b)	$M \sim N(84, 121) \text{ and } W \sim N(62, 100) \quad \text{Let } Y = M - 1.5W$ $E(Y) = 84 - 1.5 \times 62 = -9$ $\text{Var}(Y) = \text{Var}(M) + 1.5^2 \text{Var}(W)$ $= 11^2 + 1.5^2 \times 10^2 = 346$ $P(Y < 0), \quad = P(Z < 0.48\dots) =$	M1 A1 M1 A1 M1, A1 (6) 11 marks
(a)	2^{nd} B1 for $\sigma = 22$ or $\sigma^2 = 22^2$ or 484 M1 for standardising with their mean and standard deviation (ignore direction of inequality)	
(b)	1^{st} M1 for attempting to find Y . Need to see $\pm(M - 1.5W)$ or equiv. May be implied by $\text{Var}(Y)$. 1^{st} A1 for a correct value for their $E(Y)$ i.e. usually ± 9 . Do not give M1A1 for a “lucky” ± 9 . 2^{nd} M1 for attempting $\text{Var}(Y)$ e.g. $\dots + 1.5^2 \times 10^2$ or $11^2 + 1.5^2 \times \dots$ 3^{rd} M1 for attempt to calculate the correct probability. Must be attempting a probability > 0.5 . Must attempt to standardise with a relevant mean and standard deviation Using $\sigma_M^2 = 11$ or $\sigma_W^2 = 10$ is not a misread.	

Question number	Scheme	Marks
5. (a)	<p>Only cleaners - no managers i.e. not all <u>types</u>. OR Not a random sample 1st 50 may be in same shift/group/share <u>same views</u>. OR Not a random sample (Allow “not a representative sample” in place of “not a random sample”)</p> <p>(b)(i) Label employees (1-550) or obtain an ordered list Select <u>first</u> using <u>random numbers</u> (from 1 - 11) Then select every 11th person from the list</p> <p>(ii) Label managers (1-55) and cleaners (1-495) Use random numbers to select... ...5 managers and 45 cleaners</p> <p>(c) 390, 372 (They must be in this order)</p>	<p>B1g B1h (2)</p> <p>B1 B1 B1</p> <p>M1 M1 A1 (6)</p> <p>B1, B1 (2) 10 marks</p>
(a)	<p>After 1st B1, comments should be in context, i.e. mention cleaners, managers, types of worker etc</p> <p>1st B1g for one row 2nd B1h for both rows. “Not a random sample” only counts once. Score B1B0 or B1B1 or B0B0 on EPEN</p>	
(b)(i)	<p>1st B1 for idea of labelling or getting an ordered list. No need to see 1-550. 2nd B1 selecting first member of sample using random numbers (1-11 need not be mentioned) 3rd B1 selecting every nth where $n = 11$.</p>	
(ii)	<p>1st M1 for idea of <u>two</u> groups and labelling <u>both</u> groups. (Actual numbers used not required) 2nd M1 for use of random numbers within each strata. Don't give for SRS from all 550. “Assign random numbers to managers and cleaners” scores M0M1 A1 for 5 managers <u>and</u> 45 cleaners. (This mark is dependent upon scoring at least one M)</p>	

Question number	Scheme	Marks
6. (a)	$p = \frac{0 \times 11 + 1 \times 21 + \dots}{10 \times (11 + 21 + \dots) \text{ or } 10 \times 100} = \frac{223}{1000} = 0.223 \text{ (*)}$ <p>(Accept $\frac{223}{1000}$)</p>	M1, A1cso (2)
(b)	$r = (0.8)^{10} \times 100 = 10.7374$	awrt 10.74 M1A1
	$s = \binom{10}{2} (0.8)^8 \times (0.2)^2 \times 100 = 30.198\dots$	awrt 30.2 A1
	$t = 100 - [r + s + 26.84 + 20.13 + 8.81] =$	awrt 3.28 A1cao (4)
(c)	H_0 : Binomial ($[n = 10], p = 0.2$) is a suitable model for these data	B1
	H_1 : Binomial ($[n = 10], p = 0.2$) is NOT a suitable model for these data	B1 (2)
(d)	Since $t < 5$, the last two groups are combined	M1
	and $\nu = 4 = 5 - 1$	A1 (2)
(e)	Critical value $\chi_4^2(5\%) = 9.488$	B1
	Not significant or do not reject null hypothesis	M1
	The binomial distribution with $p = 0.2$ is a suitable model for the number of cuttings that do not grow	A1 (3)
13 marks		
(a)	M1 Must show clearly how to get either 223 or 1000. As printed or better.	
	A1cso for showing how to get <u>both</u> 223 and 1000 and reaching $p = 0.223$	
(b)	M1 for any correct method (a correct expression) seen for r or s .	
	1 st A1 for correct value for r awrt 10.74	
	2 nd A1 for $s =$ awrt 30.2	
	3 rd A1 for $t = 3.28$ only	
(c)	B1 for each. The value of p must be mentioned at least once. Accept B(10, 0.2)	
	If hypotheses are correct but with no value of p then score B0B1	
	Minimum is $X \sim B(10, 0.2)$. If just B(10, 0.2) and not B(10, 0.2) award B1B0	
(d)	M1 for combining groups (must be stated or implied by a new table with combined cell seen)	
	A1 for the calculation $4 = 5 - 1$	
(e)	M1 for a correct statement based on 4.17 and their cv(context not required) (may be implied)	
	Use of 4.17 as a critical value scores B0M0A0	
	A1 for a correct interpretation in context and $p = 0.2$ and cuttings mentioned.	

Question number	Scheme	Marks
7. (a)	$H_0 : \mu_F = \mu_M \quad H_1 : \mu_F \neq \mu_M \quad (\text{Allow } \mu_1 \text{ and } \mu_2)$ $z = \frac{6.86 - 5.48}{\sqrt{\frac{4.51^2}{200} + \frac{3.62^2}{100}}}$ $= 2.860\dots$ <p>2 tail 5% critical value $(\pm) 1.96$ (or probability awrt 0.0021~0.0022)</p> <p>Significant result or reject the null hypothesis (o.e.)</p> <p>There is evidence of a difference in the (mean) amount spent on junk food by male and female teenagers</p>	B1 M1 A1 awrt $(\pm)2.86$ A1 B1 M1 A1ft (7)
(b)	CLT enables us to assume \bar{F} and \bar{M} are normally distributed	B1 (1)
8 marks		
(a)	$1^{\text{st}} \text{ M1 for an attempt at } \frac{a-b}{\sqrt{\frac{c}{100 \text{ or } 200} + \frac{d}{100 \text{ or } 200}}}$ <p>with 3 of a, b, c or d correct</p> <p>1st A1 for a fully correct expression</p> <p>2nd B1 for ± 1.96 <u>but</u> only if their H_1 is two-tail (it may be in words so B0B1 is OK)</p> <p>If H_1 is one-tail this is automatically B0 too.</p> <p>2nd M1 for a correct statement based on comparison of their z with their cv. May be implied</p> <p>3rd A1 for a correct conclusion in context based on their z and 1.96.</p> <p>Must mention <u>junk food</u> or <u>money</u> and <u>male vs female</u>.</p>	
(b)	B1 for \bar{F} or \bar{M} mentioned. Allow “ <u>mean</u> (amount spent on junk food) is <u>normally distributed</u> ”	
Read the whole statement e.g. “ original distribution is normal so mean is...” scores B0		

Mark Scheme (Results) Summer 2009

GCE

GCE Mathematics (6691/01)

June 2009
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks
Q1	<p>(a) Randomly select a number between 00 and 499 (001 and 500) select every 500th person</p> <p>(bi) <u>Quota</u> Advantage: <u>Representative</u> sample can be achieved (with small sample size) <u>Cheap</u> (costs kept to a minimum) not “quick“ Administration relatively <u>easy</u> Disadvantage Not possible to estimate sampling errors (due to lack of randomness) Not a random process Judgment of interviewer can affect choice of sample – <u>bias</u> Non-response not recorded Difficulties of defining controls e.g. social class</p> <p>(bii) <u>Systematic</u> Advantage: <u>Simple</u> or <u>easy</u> to use not “quick” or “cheap” or “efficient” It is suitable for large <u>samples</u> (not populations) Disadvantage Only random if the ordered list is (truly) random Requires a list of the population <u>or</u> must assign a number to each member of the pop.</p>	<p>B1 B1 (2)</p> <p>B1</p> <p>B1</p> <p>(2)</p> <p>B1</p> <p>B1 (2)</p> <p>[6]</p>
(a)	<p>1st B1 for idea of using random numbers to select the first from 1 - 500 (o.e.) 2nd B1 for selecting every 500th (name on the list)</p> <p style="text-align: center;">If they are clearly trying to carry out <u>stratified</u> sample then score B0B0</p>	
(b)	Score B1 for any one line	
(i)	<p>1st B1 for Quota advantage 2nd B1 for Quota disadvantage</p>	
(ii)	<p>3rd B1 for Systematic Advantage 4th B1 for Systematic Disadvantage</p>	

Question Number	Scheme	Marks
Q2	<p>(a) Limits are $20.1 \pm 1.96 \times 0.5$</p> <p style="text-align: center;"><u>(19.1, 21.1)</u></p> <p>(b) 98 % confidence limits are</p> <p style="text-align: center;">$20.1 \pm 2.3263 \times \frac{0.5}{\sqrt{10}}$</p> <p style="text-align: center;"><u>(19.7, 20.5)</u></p> <p>(c) The growers claim is not correct Since 19.5 does not lie in the interval (19.7, 20.5)</p>	<p>M1 B1 A1cso (3)</p> <p>M1 B1 A1A1 (4)</p> <p>B1 dB1 (2) [9]</p>
	<p>(a) M1 for $20.1 \pm z \times 0.5$. Need 20.1 and 0.5 in correct places with no $\sqrt{10}$ B1 for $z = 1.96$ (or better) A1 for awrt 19.1 <u>and</u> awrt 21.1 but must have scored both M1 and B1 [Correct answer only scores 3/3]</p> <p>(b) M1 for $20.1 \pm z \times \frac{0.5}{\sqrt{10}}$, need to see 20.1, 0.5 and $\sqrt{10}$ in correct places B1 for $z = 2.3263$ (or better) 1st A1 for awrt 19.7 2nd A1 for awrt 20.5 [Correct answer only scores M1B0A1A1]</p> <p>(c) 1st B1 for rejection of the claim. Accept “unlikely” or “not correct” 2nd dB1 Dependent on scoring 1st B1 in this part for rejecting grower’s claim for an argument that supports this. Allow comment on <u>their</u> 98% CI from (b)</p>	

Question Number	Scheme	Marks																																																							
Q3 (a)	<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> </tr> </thead> <tbody> <tr> <td>BMI</td> <td>1</td> <td>6</td> <td>3</td> <td>8</td> <td>4</td> <td>5</td> <td>7</td> <td>2</td> <td>9</td> <td>10</td> </tr> <tr> <td>or</td> <td>10</td> <td>5</td> <td>8</td> <td>3</td> <td>7</td> <td>6</td> <td>4</td> <td>9</td> <td>2</td> <td>1</td> </tr> <tr> <td>Finishing position</td> <td>3</td> <td>5</td> <td>1</td> <td>9</td> <td>6</td> <td>4</td> <td>10</td> <td>2</td> <td>7</td> <td>8</td> </tr> <tr> <td>d^2</td> <td>4</td> <td>1</td> <td>4</td> <td>1</td> <td>4</td> <td>1</td> <td>9</td> <td>0</td> <td>4</td> <td>4</td> </tr> </tbody> </table> <p> $\sum d^2 = 32$ (298) $r_s = 1 - \frac{6 \times 32}{10 \times 99}$ $= 0.80606\dots$ (-0.80606) accept $\pm \frac{133}{165}$ awrt ± 0.806 </p>		A	B	C	D	E	F	G	H	I	J	BMI	1	6	3	8	4	5	7	2	9	10	or	10	5	8	3	7	6	4	9	2	1	Finishing position	3	5	1	9	6	4	10	2	7	8	d^2	4	1	4	1	4	1	9	0	4	4	<p>M1</p> <p>M1</p> <p>M1 A1ft</p> <p>A1 (5)</p>
	A	B	C	D	E	F	G	H	I	J																																															
BMI	1	6	3	8	4	5	7	2	9	10																																															
or	10	5	8	3	7	6	4	9	2	1																																															
Finishing position	3	5	1	9	6	4	10	2	7	8																																															
d^2	4	1	4	1	4	1	9	0	4	4																																															
(b)	<p>$H_0 : \rho = 0, H_1 : \rho > 0,$</p> <p>Critical value is $(\pm)0.5636$</p> <p>$(0.806 > 0.5636$ therefore) in critical region/ reject H_0</p>	<p>B1 B1</p> <p>B1</p> <p>M1 A1ft</p>																																																							
(c)	<p>The lower the BMI the higher the position in the race./ support for doctors belief</p> <p>The position is already ranked OR Position is not Normally distributed</p>	<p>(5)</p> <p>B1 (1)</p> <p>[11]</p>																																																							
(a)	<p>1st M1 for attempt to rank BMI scores</p> <p>2nd M1 for attempt at $\sum d^2$ (<u>must</u> be using ranks)</p> <p>3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.</p> <p>1st A1ft for a correct expression. ft their $\sum d^2$ but only if all 3 Ms are scored</p> <p>2nd A1 awrt ± 0.806 (but sign must be compatible with their $\sum d^2$)</p>	<p>No ranking can score 3rd M1 only</p>																																																							
(b)	<p>2nd B1 for $\rho > 0$ (or < 0 but must be one tail and consistent with their ranking)</p> <p>3rd B1 for critical value that is compatible with their H_1. If one-tail must be ± 0.5636 if two-tail must be ± 0.6485 [Condone wrong sign]</p> <p>M1 for a correct statement relating their r_s with their cv. e.g. “reject H_0”, “in critical region”, “significant result” May be implied by a correct comment</p> <p>A1ft for correct comment in context. Must mention low/high BMI and race/fitness <u>or</u> doctor’s belief. Comment should be <u>one</u>-tailed. Allow positive <u>correlation</u> between... but <u>NOT</u> ...positive <u>relationship</u>...</p>	<p>No H_1 assume one-tail for 3rd B1</p>																																																							
(c)	<p>B1 for a correct and relevant comment either based on the fact that the data was originally partially ordered <u>or</u> on the underlying normal assumption “Quicker” or “easier” score B0</p>																																																								

Question Number	Scheme	Marks
Q4	$X \sim N(55, 3^2) \text{ therefore } \bar{X} \sim N\left(55, \frac{9}{8}\right)$ $P(\bar{X} > 57) = P\left(Z > \frac{57 - 55}{\sqrt{\frac{9}{8}}}\right) = P(Z > 1.8856\dots)$ $= 1 - 0.9706$ $= 0.0294$ <p style="text-align: right;"><u>0.0294~0.0297</u></p>	B1 B1 M1 M1 A1 [5]
ALT	<p>1st B1 for $\bar{X} \sim$ normal and $\mu = 55$, may be implied but must be \bar{X}</p> <p>2nd B1 for $\text{Var}(\bar{X})$ or st. dev of \bar{X} e.g. $\bar{X} \sim N(55, \frac{9}{8})$ or $\bar{X} \sim N\left(55, \left(\frac{3}{\sqrt{8}}\right)^2\right)$ for B1B1</p> <p>Condone use of X if they clearly mean \bar{X} so $X \sim N(55, \frac{9}{8})$ is OK for B1B1</p> <p>1st M1 for an attempt to standardize with 57 and mean of 55 and their st. dev. $\neq 3$</p> <p>2nd M1 for 1 - tables value. Must be trying to find a probability < 0.5</p> <p>A1 for answers in the range 0.0294~0.0297</p> $\sum_{i=1}^8 X_i \sim N(8 \times 55, 8 \times 3^2)$ <p>1st B1 for $\sum X \sim$ normal and mean = 8×55</p> <p>2nd B1 for variance = 8×3^2</p> <p>1st M1 for attempt to standardise with 57×8, mean of 55×8 and their st dev $\neq 3$</p>	

Question Number	Scheme	Marks																		
Q5	<p>(a) $\lambda = \frac{0 \times 40 + 1 \times 33 + 2 \times 14 + 3 \times 8 + 4 \times 5}{100} = 1.05$</p> <p>(b) Using Expected frequency = $100 \times P(X=x) = 100 \times \frac{e^{-1.05} 1.05^x}{x!}$ gives $r = 36.743$ awrt 36.743 or 36.744 $s = 19.290$ 19.29 or awrt 19.290</p> <p>(c) H_0 : Poisson distribution is a suitable model H_1 : Poisson distribution is not a suitable model</p> <table border="1" data-bbox="300 667 1246 1010"> <thead> <tr> <th>Number of goals</th> <th>Frequency</th> <th>Expected frequency</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>40</td> <td>34.994</td> </tr> <tr> <td>1</td> <td>33</td> <td>36.743</td> </tr> <tr> <td>2</td> <td>14</td> <td>19.290</td> </tr> <tr> <td>3</td> <td>8</td> <td>6.752</td> </tr> <tr> <td>≥ 4</td> <td>5</td> <td>2.221</td> </tr> </tbody> </table> <p>$\nu = 4 - 1 - 1 = 2$ CR : $\chi^2_2(0.05) > 5.991$</p> $\sum \frac{(O-E)^2}{E} = \frac{(40-34.9937)^2}{34.9937} + \dots + \frac{(13-8.972443)^2}{8.972443}$ <p style="text-align: center;">[=0.7161...+0.3813...+1.4508...+1.80789..] = 4.356. (ans in range 4.2 – 4.4)</p> <p>Not in critical region Number of goals scored can follow a Poisson distribution / managers claim is justified</p>	Number of goals	Frequency	Expected frequency	0	40	34.994	1	33	36.743	2	14	19.290	3	8	6.752	≥ 4	5	2.221	<p>M1 A1 (2)</p> <p>M1 A1 A1 (3)</p> <p>B1</p> <p>M1</p> <p>B1ft B1</p> <p>M1</p> <p>A1</p> <p>A1 ft (7)</p> <p>[12]</p>
Number of goals	Frequency	Expected frequency																		
0	40	34.994																		
1	33	36.743																		
2	14	19.290																		
3	8	6.752																		
≥ 4	5	2.221																		
	<p>(a) M1 for an attempt to find the mean- at least 2 terms on numerator seen Correct answer only will score both marks</p> <p>(b) M1 for use of correct formula (ft their mean). 1st A1 for r, 2nd A1 for s (19.29 OK)</p> <p>(c) 1st B1 Must have both hypotheses and mention Poisson at least once inclusion of their value for mean in hypotheses is B0 but condone in conclusion 1st M1 for an attempt to pool ≥ 4 2nd B1ft for $n - 1 - 1 = 2$ i.e realising that they must subtract 2 from their n 3rd B1 for 5.991 only 2nd M1 for an attempt at the test statistic, at least 2 correct expressions/values (to 3sf) 1st A1 for answers in the range 4.2~4.4 2nd A1 for correct comment in context based on their test statistic and their cv that mentions goals or manager. Dependent on 2nd M1 Condone mention of Po(1.05) in conclusion Score A0 for inconsistencies e.g. “significant” followed by “manager’s claim is justified”</p>																			

Question Number	Scheme	Marks
Q6 (a)	<p>$\mu_U \sim$ mean length of upper shore limpets, $\mu_L \sim$ mean length of lower shore limpets</p> <p>$H_0 : \mu_u = \mu_L$</p> <p>$H_1 : \mu_u < \mu_L$ both</p> $\text{s.e.} = \sqrt{\frac{0.42^2}{120} + \frac{0.67^2}{150}}$ $= 0.0668$ $z = \frac{5.05 - 4.97}{0.0668} = (\pm)1.1975 \quad \text{awrt } \pm \underline{1.20}$ <p>Critical region is $z \geq 1.6449$, or probability = awrt (0.115 or 0.116) $z = \pm 1.6449$</p> <p>(1.1975 < 1.6449) therefore not in critical region / accept H_0/not significant (or $P(Z \geq 1.1975) = 0.1151$, $0.1151 > 0.05$ or z not in critical region)</p> <p>There is no evidence that the limpets on the upper shore are shorter than the limpets on the lower shore.</p>	<p>B1</p> <p>M1</p> <p>A1</p> <p>dM1 A1</p> <p>B1</p> <p>M1</p> <p>A1 (8)</p> <p>B1</p> <p>B1</p> <p>(2)</p> <p>[10]</p>
(a)	<p>1st B1 If μ_1, μ_2 used then it must be clear which refers to upper shore. Accept sensible choice of letters such as u and l.</p> <p>1st M1 Condone minor slips e.g. $\frac{0.67^2}{120}$ or $\frac{0.67}{150} + \frac{0.42^2}{120}$ etc i.e. swapped n or one sd and one variance but M0 for $\sqrt{\frac{0.67}{150} + \frac{0.42}{120}}$</p> <p>1st A1 can be scored for a fully correct expression. May be implied by awrt 1.20</p> <p>2nd dM1 is dependent upon the 1st M1 but can fit their se value if this mark is scored.</p> <p>2nd A1 for awrt (+) 1.20</p> <p>3rd M1 for a correct statement based on their z value and their cv. No cv is M0A0 If using probability they must compare their p (<0.5) with 0.05 (o.e) so can allow $0.884 < 0.95$ to score this 3rd M1 mark. May be implied by their contextual statement and M1A0 is possible.</p>	
(b)	<p>3rd A1 for a correct comment to accept null hypothesis that mentions <u>length</u> of <u>limpets</u> on the two <u>shores</u>.</p> <p>1st B1 for one correct statement. Accept "samples are independent"</p> <p>2nd B1 for both statements</p>	

Question Number	Scheme	Marks
Q7 (a)	<p>Estimate of Mean = $\frac{600.9}{5} = 120.18$</p> <p>Estimate of Variance = $\frac{1}{4} \left\{ 72216.31 - \frac{600.9^2}{5} \right\}$ or $\frac{0.148}{4} = 0.037$</p> <p>(b) $P(-0.05 < \mu - \hat{\mu} < 0.05) = 0.90$ or $P(-0.05 < \bar{X} - \mu < 0.05) = 0.90$ [\leq is OK]</p> $\frac{0.05}{\frac{0.2}{\sqrt{n}}} = 1.6449$ $n = \frac{1.6449^2 \times 0.2^2}{0.05^2}$ $n = 43.29\dots$ $n = 44$	<p>M1A1</p> <p>M1 A1ft A1 (5)</p> <p>B1</p> <p>M1 A1</p> <p>dM1</p> <p>A1</p> <p>A1 (6) [11]</p>
(a)	<p>1st M1 for an attempt at $\sum x$ (accept 600 to 1sf)</p> <p>1st A1 for $\frac{600.9}{5} = \text{awrt } 120$ or awrt 120.2. No working give M1A1 for awrt 120.2</p> <p>2nd M1 for the use of a correct formula including a reasonable attempt at $\sum x^2$ (Accept 70 000 to 1sf) or $\sum (x - \bar{x})^2 = 0.15$ (to 2 dp)</p> <p>2nd A1ft for a correct expression with correct $\sum x^2$ but can fit their <u>mean</u> (for expression - no need to check values if it is incorrect)</p> <p>3rd A1 for 0.037 Correct answer with no working scores 3/3 for variance</p> <p>(b) B1 for a correct probability statement <u>or</u> “width of 90% CI = $0.05 \times 2 = 0.1$”</p> <p>1st M1 for $\frac{0.05}{\frac{0.2}{\sqrt{n}}} = z$ value <u>or</u> $2 \times \frac{0.2}{\sqrt{n}} \times z = 0.1$</p> <p>Condone 0.5 instead of 0.05 <u>or</u> missing 2 <u>or</u> 0.05 for 0.1 for M1</p> <p>1st A1 for a correct equation including 1.6449</p> <p>2nd dM1 Dependent upon 1st M1 for rearranging to get $n = \dots$ Must see “squaring”</p> <p>2nd A1 for $n = \text{awrt } 43.3$</p> <p>3rd A1 for rounding up to get $n = 44$</p> <p>Using e.g. 1.645 instead of 1.6449 can score all the marks except the 1st A1</p>	<p>1st B1 may be implied by 1st A1 scored or correct equation.</p>

Question Number	Scheme	Marks
Q8 (a)	$E(4X-3Y) = 4E(X) - 3E(Y)$ $= 4 \times 30 - 3 \times 20$ $= 60$	M1 A1 (2)
(b)	$\text{Var}(4X-3Y) = 16 \text{Var}(X) + 9 \text{Var}(Y)$ $= 16 \times 9 + 9 \times 4$ $= 180$	16 or 9; adding M1; M1 A1 (3)
(c)	$E(B) = 80$ $\text{Var}(B) = 16$ $E(B - A) = 20$ $\text{Var}(B - A) = 196$ $P(B - A > 0) = P\left(Z > \frac{-20}{\sqrt{196}}\right) = [P(Z > -1.428\dots)]$ $= 0.923 \dots$	$E(B) - E(A)$ ft on 180 and 16 B1 B1 M1 A1ft stand. using their mean and var dM1 awrt 0.923 – 0.924 A1 (6) [11]
(a)	M1 for correct use of $E(aX + bY)$ formula	
(b)	1 st M1 for $16\text{Var}(X)$ or $9\text{Var}(Y)$ 2 nd M1 for <u>adding</u> variances Key points are the 16, 9 and +. Allow slip e.g using $\text{Var}(X)=4$ etc to score Ms	
(c)	1 st M1 for attempting $B - A$ and $E(B - A)$ or $A - B$ and $E(A - B)$ This mark may be implied by an attempt at a correct probability e.g. $P\left(Z > \frac{0 - (80 - 60)}{\sqrt{180 + 16}}\right)$. To be implied we must see the “0” 1 st A1ft for $\text{Var}(B - A)$ can ft their $\text{Var}(A) = 180$ and their $\text{Var}(B) = 16$ 2 nd dM1 Dependent upon the 1 st M1 in part (c). for attempting a correct probability i.e. $P(B-A > 0)$ or $P(A-B < 0)$ and standardising with their mean and variance. They must standardise properly with the 0 to score this mark 2 nd A1 for awrt 0.923 ~ 0.924	

Mark Scheme (Results) Summer 2010

GCE

GCE Statistics S3 (6691/01)

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information, please call our GCE line on 0844 576 0025, our GCSE team on 0844 576 0027, or visit our website at www.edexcel.com.

If you have any subject specific questions about the content of this Mark Scheme that require the help of a subject specialist, you may find our **Ask The Expert** email service helpful.

Ask The Expert can be accessed online at the following link:

<http://www.edexcel.com/Aboutus/contact-us/>

Summer 2010

Publications Code UA024774

All the material in this publication is copyright

© Edexcel Ltd 2010

Hypothesis Tests (Final M1A1)

For an incorrect comparison (e.g. probability with z value) even with a correct statement and/or comment award M0A0

For a correct or no comparison with more than one statement one of which is false
Award M0A0 (This is compatible with the principle above of contradictory statements being penalised)

Apply these rules to all questions

June 2010
Statistics S3 6691
Mark Scheme

Question Number	Scheme	Marks
Q1	<p>$H_0: \mu = 80, H_1: \mu > 80$</p> $z = \frac{83 - 80}{\frac{15}{\sqrt{100}}} = 2$ <p>$2 > 1.6449$ (accept 1.645 or better)</p> <p>Reject H_0 <u>or</u> significant result <u>or</u> in the critical region Managing director's claim is supported.</p>	<p>B1,B1</p> <p>M1A1</p> <p>B1</p> <p>M1</p> <p>A1</p> <p style="text-align: right;">7</p>
<p>2nd M1A1</p> <p>Critical Region</p>	<p>1st B1 for H_0. They must use μ not x, p, λ or \bar{x} etc</p> <p>2nd B1 for H_1 (must be > 80). Same rules about μ.</p> <p>1st M1 for attempt at standardising using 83, 80 and $\frac{15}{\sqrt{100}}$. Can accept \pm.</p> <p>May be implied by $z = \pm 2$</p> <p>1st A1 for + 2 only</p> <p>3rd B1 for ± 1.6449 seen (or probability of 0.0228 or better)</p> <p>2nd M1 for a correct statement about "significance" or rejecting H_0 (or H_1) based on their z value and their 1.6449 (provided it is a recognizable critical value from normal tables) <u>or</u> their probability (< 0.5) and significance level of 0.05. Condone their probability > 0.5 compared with 0.95 for the 2nd M1</p> <p>2nd A1 for a correct contextualised comment. Must mention "director" and "claim" <u>or</u> "time" and "use of Internet". No follow through.</p> <p>If no comparison or statement is made but a correct contextualised comment is given the M1 can be implied. If a comparison is made it must be <u>compatible</u> with statement otherwise M0 e.g. comparing 0.0228 with 1.6449 is M0 or comparing probability 0.9772 with 0.05 is M0 comparing -2 with - 1.6449 is OK provided a correct statement accompanies it condone -2 $>$ -1.6449 provided their statement correctly rejects H_0.</p> <p>They may find a critical region for \bar{X}: $\bar{X} > 80 + \frac{15}{\sqrt{100}} \times 1.6449 = \text{awrt } 82.5$</p> <p>1st M1 for $80 + \frac{15}{\sqrt{100}} \times (z \text{ value})$</p> <p>3rd B1 for 1.645 or better</p> <p>1st A1 for awrt 82.5</p> <p>The rest of the marks are as per the scheme.</p>	

Question Number	Scheme	Marks
Q2	<p style="text-align: center;">[$P \sim N(90,9)$ and $J \sim N(91,12)$]</p> <p>(a) $(J - P) \sim N(1, 21)$ $P(J < P) = P(J - P < 0)$ $= P\left(Z < \frac{0-1}{\sqrt{21}}\right)$ $= P(Z < -0.2182\dots)$ $= 1 - 0.5871 = 0.4129$ calculator (0.4136....)</p> <p>(b) $X = (J_1 + J_2 + \dots + J_{60}) - (P_1 + P_2 + \dots + P_{60})$ $E(X) = 60 \times 91 - 60 \times 90 = 60$ [stated as $E(X) = 60$ or $X \sim N(60, \dots)$] $\text{Var}(X) = 60 \times 9 + 60 \times 12 = 1260$ $P(X > 120) = P\left(Z > \frac{120-60}{\sqrt{1260}}\right)$ $= P(Z > 1.69030\dots)$ $= 1 - 0.9545 = 0.0455$</p>	<p>M1, A1</p> <p>dM1</p> <p>A1 (4)</p> <p>M1</p> <p>B1</p> <p>A1</p> <p>M1</p> <p>A1 (5)</p> <p style="text-align: right;">9</p>
Use of means	<p>(a) 1st M1 for attempting $J - P$ and $E(J - P)$ or $P - J$ and $E(P - J)$ 1st A1 for variance of 21 (Accept $9 + 12$). Ignore any slip in μ here. 2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on previous M so if $J - P$ (or $P - J$) is not being used score M0 If their method is not crystal clear then they must be attempting $P(Z < -ve \text{ value})$ or $P(Z > +ve \text{ value})$ i.e. their probability <u>after</u> standardisation should lead to a prob. < 0.5 so e.g. $P(J - P < 0)$ leading to 0.5871 is M0A0 unless the M1 is clearly earned. 2nd A1 for awrt 0.413 or 0.414</p> <p style="text-align: center;">The first 3 marks may be implied by a correct answer</p> <p>(b) 1st M1 for a clear attempt to identify a correct form for X. This may be implied by correct variance of 1260 B1 for $E(X) = 60$. Can be awarded even if they are using $X = 60J - 60P$. Allow $P - J$ and -60 1st A1 for a correct variance. If 1260 is given the M1 is scored by implication. 2nd M1 for attempting a correct probability and standardising with 120 and their 60 and 1260 If the answer is incorrect a full <u>expression</u> must be seen following through their values for M1 e.g. $P\left(Z > \frac{120 - \text{their } 60}{\sqrt{\text{their variance}}}\right)$. If using -60, should get $P\left(Z < \frac{-120 - -60}{\sqrt{\text{their variance}}}\right)$</p> <p>Attempt to use $\bar{J} - \bar{P}$ for 1st M1, $E(\bar{J} - \bar{P}) = 1$ for B1 and $\text{Var}(\bar{J} - \bar{P}) = 0.35$ for A1 Then 2nd M1 for standardisation with 2, and their 1 and 0.35</p>	

Question Number	Scheme	Marks
Q3 (a)	$E \sim N(0, 0.5^2)$ or $X \sim N(w, 0.5^2)$ $P(E < 0.6) = P\left(Z < \frac{0.6}{0.5}\right)$ or $P(X - w < 0.6) = P\left(Z < \frac{0.6}{0.5}\right)$ $= P(Z < 1.2)$ $= 2 \times 0.8849 - 1 = 0.7698$ awrt 0.770	M1 A1 (2)
(b)	$\bar{E} \sim N\left(0, \frac{1}{64}\right)$ or $\bar{X} \sim N\left(w, \frac{0.5^2}{16}\right)$ $P(\bar{E} < 0.3) = P\left(Z < \frac{0.3}{\frac{1}{8}}\right)$ or $P(\bar{X} - w < 0.3) = P\left(Z < \frac{0.3}{\frac{1}{8}}\right)$ $= P(Z < 2.4)$ $= 2 \times 0.9918 - 1 = 0.9836$ awrt 0.984	M1 M1, A1 A1 (4)
(c)	$35.6 \pm 2.3263 \times \frac{1}{8}$ (35.3, 35.9)	M1 B1 A1, A1 (4) 10
(a)	1 st M1 for identifying a correct probability (they must have the 0.6) and attempting to standardise. Need . This mark can be given for 0.8849 - 0.1151 seen as final answer. 1 st A1 for awrt 0.770. NB an answer of 0.3849 or 0.8849 scores M0A0 (since it implies no) M1 may be implied by a correct answer	
(b)	1 st M1 for a correct attempt to define \bar{E} or \bar{X} but must attempt $\frac{\sigma^2}{n}$. Condone labelling as E or X This mark may be implied by standardisation in the next line. 2 nd M1 for identifying a correct probability statement using \bar{E} or \bar{X} . Must have 0.3 and 1 st A1 for correct standardisation as printed or better 2 nd A1 for awrt 0.984 The M marks may be implied by a correct answer.	
Sum of 16, not means	1 st M1 for correct attempt at suitable sum distribution with correct variance ($= 16 \times \frac{1}{4}$) 2 nd M1 for identifying a correct probability. Must have 4.8 and 1 st A1 for correct standardisation i.e. need to see $\frac{4.8}{\sqrt{4}}$ or better	
(c)	M1 for $35.6 \pm z \times \frac{0.5}{\sqrt{16}}$ B1 for 2.3263 or better. Use of 2.33 will lose this mark but can still score $\frac{3}{4}$ 1 st A1 for awrt 35.3 2 nd A1 for awrt 35.9	

Question Number	Scheme	Marks																																
Q4 (a)	<table border="1" data-bbox="311 302 1236 548"> <tr> <td>Distance rank</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> </tr> <tr> <td>Depth rank</td> <td>1</td> <td>2</td> <td>4</td> <td>3</td> <td>6</td> <td>7</td> <td>5</td> </tr> <tr> <td>d</td> <td>0</td> <td>0</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>2</td> </tr> <tr> <td>d^2</td> <td>0</td> <td>0</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>4</td> </tr> </table> <p data-bbox="223 627 351 672">$\sum d^2 = 8$</p> <p data-bbox="223 683 526 862"> $r_s = 1 - \frac{6 \times 8}{7 \times 48}$ $= \frac{6}{7} = 0.857142$ </p> <p data-bbox="1141 795 1292 840">awrt 0.857</p>	Distance rank	1	2	3	4	5	6	7	Depth rank	1	2	4	3	6	7	5	$ d $	0	0	1	1	1	1	2	d^2	0	0	1	1	1	1	4	<p data-bbox="1353 414 1404 448">M1</p> <p data-bbox="1353 481 1404 515">M1</p> <p data-bbox="1353 627 1444 660">M1A1</p> <p data-bbox="1353 716 1404 750">M1</p> <p data-bbox="1353 817 1404 851">A1</p> <p data-bbox="1468 851 1516 884">(6)</p> <p data-bbox="1353 884 1404 918">B1</p> <p data-bbox="1353 929 1404 963">B1</p> <p data-bbox="1353 974 1404 1008">M1</p> <p data-bbox="1353 1019 1428 1052">A1ft</p> <p data-bbox="1468 1064 1516 1097">(4)</p> <p data-bbox="1468 1176 1516 1209">10</p>
Distance rank	1	2	3	4	5	6	7																											
Depth rank	1	2	4	3	6	7	5																											
$ d $	0	0	1	1	1	1	2																											
d^2	0	0	1	1	1	1	4																											
(a)	<p data-bbox="223 1243 1053 1276">1st M1 for an attempt to rank the depths against the distances</p> <p data-bbox="223 1276 1053 1310">2nd M1 for attempting d for their ranks. Must be using ranks.</p> <p data-bbox="223 1310 925 1355">3rd M1 for attempting $\sum d^2$ (must be using ranks)</p> <p data-bbox="223 1355 893 1400">1st A1 for sum of 8 (or 104 for reverse ranking)</p> <p data-bbox="223 1400 1500 1489">4th M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.</p> <p data-bbox="223 1489 1500 1534">2nd A1 for awrt (\pm) 0.857. Sign should correspond to ranking (so use of 104 should get -0.857)</p>																																	
(b)	<p data-bbox="223 1568 1500 1601">1st B1 for both hypotheses in terms of ρ, H_1 must be one tail and compatible with their ranking</p> <p data-bbox="223 1601 718 1646">2nd B1 for cv of 0.8929 (accept \pm)</p> <p data-bbox="223 1646 1436 1691">M1 for a correct statement relating their r_s with their cv but cv must be such that $cv < 1$</p> <p data-bbox="223 1691 1436 1780">A1ft for a correct contextualised comment. Must mention “researcher” and “claim” <u>or</u> “distance (from bank)” and “depth (of water)”</p> <p data-bbox="223 1780 1117 1825">Follow through their r_s and their cv (provided it is $cv < 1$)</p> <p data-bbox="223 1825 718 1859">Use of “association” is A0</p>																																	

Question Number	Scheme	Marks																																																					
Q5	<table border="1" data-bbox="223 291 1212 465"> <thead> <tr> <th>Finances</th> <th>Worse</th> <th>Same</th> <th>Better</th> <th></th> </tr> </thead> <tbody> <tr> <td>Income</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Under £15 000</td> <td>10.54</td> <td>10.54</td> <td>12.92</td> <td>34</td> </tr> <tr> <td>£15 000 and above</td> <td>20.46</td> <td>20.46</td> <td>25.08</td> <td>66</td> </tr> <tr> <td></td> <td>31</td> <td>31</td> <td>38</td> <td>100</td> </tr> </tbody> </table> <p data-bbox="223 510 1085 544">H_0 : State of finances and income are independent (not associated)</p> <p data-bbox="223 555 1085 589">H_1 : State of finances and income are not independent (associated)</p> <table border="1" data-bbox="223 638 821 963"> <thead> <tr> <th>O_i</th> <th>E_i</th> <th>$\frac{(O_i - E_i)^2}{E_i}$</th> <th>$\frac{O_i^2}{E_i}$</th> </tr> </thead> <tbody> <tr> <td>14</td> <td>10.54</td> <td>1.1358....</td> <td>18.59..</td> </tr> <tr> <td>11</td> <td>10.54</td> <td>0.0200....</td> <td>11.48..</td> </tr> <tr> <td>9</td> <td>12.92</td> <td>1.1893...</td> <td>6.269..</td> </tr> <tr> <td>17</td> <td>20.46</td> <td>0.5851...</td> <td>14.12..</td> </tr> <tr> <td>20</td> <td>20.46</td> <td>0.0103...</td> <td>19.55..</td> </tr> <tr> <td>29</td> <td>25.08</td> <td>0.6126...</td> <td>33.53..</td> </tr> </tbody> </table> <p data-bbox="223 996 1324 1070">$\sum \frac{(O_i - E_i)^2}{E_i} = 3.553... \quad \text{or} \quad \sum \frac{O_i^2}{E_i} - 100 = 103.553... - 100 = 3.553... \quad (\text{awrt } \mathbf{3.55})$</p> <p data-bbox="223 1081 486 1115">$\nu = (3 - 1)(2 - 1) = 2$</p> <p data-bbox="223 1126 359 1160">cv is 5.991</p> <p data-bbox="223 1171 1117 1205">3.553 < 5.991 so insufficient evidence to reject H_0 <u>or</u> not significant</p> <p data-bbox="223 1216 1173 1249">There is no evidence of association between state of finances and income.</p>	Finances	Worse	Same	Better		Income					Under £15 000	10.54	10.54	12.92	34	£15 000 and above	20.46	20.46	25.08	66		31	31	38	100	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$	14	10.54	1.1358....	18.59..	11	10.54	0.0200....	11.48..	9	12.92	1.1893...	6.269..	17	20.46	0.5851...	14.12..	20	20.46	0.0103...	19.55..	29	25.08	0.6126...	33.53..	<p data-bbox="1353 387 1401 454">M1 A1</p> <p data-bbox="1353 544 1401 577">B1</p> <p data-bbox="1353 790 1401 824">M1</p> <p data-bbox="1353 857 1401 891">A1</p> <p data-bbox="1353 1014 1401 1048">A1</p> <p data-bbox="1353 1081 1401 1115">B1</p> <p data-bbox="1353 1126 1401 1160">B1</p> <p data-bbox="1353 1171 1401 1205">M1</p> <p data-bbox="1353 1216 1401 1249">A1</p> <p data-bbox="1481 1294 1520 1328">10</p>
Finances	Worse	Same	Better																																																				
Income																																																							
Under £15 000	10.54	10.54	12.92	34																																																			
£15 000 and above	20.46	20.46	25.08	66																																																			
	31	31	38	100																																																			
O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$																																																				
14	10.54	1.1358....	18.59..																																																				
11	10.54	0.0200....	11.48..																																																				
9	12.92	1.1893...	6.269..																																																				
17	20.46	0.5851...	14.12..																																																				
20	20.46	0.0103...	19.55..																																																				
29	25.08	0.6126...	33.53..																																																				
	<p data-bbox="223 1350 1252 1417">1st M1 for some use of $\frac{\text{Row Total} \times \text{Col.Total}}{\text{Grand Total}}$. May be implied by correct E_i</p> <p data-bbox="223 1429 805 1462">1st A1 for all expected frequencies correct</p> <p data-bbox="223 1473 1436 1541">B1 for both hypotheses. Must mention “state” or “finances” and “income” at least once Use of “relationship” or “correlation” or “connection” is B0</p> <p data-bbox="223 1552 1492 1585">2nd M1 for at least two correct terms (as in 3rd or 4th column) or correct expressions with their E_i</p> <p data-bbox="223 1597 1500 1630">2nd A1 for all correct terms. May be implied by a correct answer.(2 dp or better-allow eg 1.13...)</p> <p data-bbox="223 1641 1468 1675">3rd M1 for a correct statement linking their test statistic and their cv . Must be χ^2 not normal.</p> <p data-bbox="223 1686 1468 1787">4th A1 for a correct comment in context - must mention “state” or “finances” and “income” condone “relationship” or “connection” here but not “correlation”. No follow through. e.g. “There is no evidence of a relationship between finances and income”</p>																																																						

Question Number	Scheme						Marks																																			
Q6	<table border="1"> <thead> <tr> <th>Distance from centre of site (m)</th> <th>0-1</th> <th>1-2</th> <th>2-4</th> <th>4-6</th> <th>6-9</th> <th>9-12</th> </tr> </thead> <tbody> <tr> <td>$b - a$</td> <td>1</td> <td>1</td> <td>2</td> <td>2</td> <td>3</td> <td>3</td> </tr> <tr> <td>No of artefacts</td> <td>22</td> <td>15</td> <td>44</td> <td>37</td> <td>52</td> <td>58</td> </tr> <tr> <td>$P(a \leq X < b)$</td> <td>$\frac{1}{12}$</td> <td>$\frac{1}{12}$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{6}$</td> <td>$\frac{1}{4}$</td> <td>$\frac{1}{4}$</td> </tr> <tr> <td>$228 \times P(a \leq X < b)$</td> <td>19</td> <td>19</td> <td>38</td> <td>38</td> <td>57</td> <td>57</td> </tr> </tbody> </table>						Distance from centre of site (m)	0-1	1-2	2-4	4-6	6-9	9-12	$b - a$	1	1	2	2	3	3	No of artefacts	22	15	44	37	52	58	$P(a \leq X < b)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$	$228 \times P(a \leq X < b)$	19	19	38	38	57	57	M1 A1 A1
	Distance from centre of site (m)	0-1	1-2	2-4	4-6	6-9	9-12																																			
	$b - a$	1	1	2	2	3	3																																			
	No of artefacts	22	15	44	37	52	58																																			
	$P(a \leq X < b)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$																																			
	$228 \times P(a \leq X < b)$	19	19	38	38	57	57																																			
	<table border="1"> <thead> <tr> <th>Class</th> <th>O_i</th> <th>E_i</th> <th>$\frac{(O_i - E_i)^2}{E_i}$</th> <th>$\frac{O_i^2}{E_i}$</th> </tr> </thead> <tbody> <tr> <td>0-1</td> <td>22</td> <td>19</td> <td>$\frac{9}{19} = 0.4736\dots$</td> <td>25.57...</td> </tr> <tr> <td>1-2</td> <td>15</td> <td>19</td> <td>$\frac{16}{19} = 0.8421\dots$</td> <td>11.84...</td> </tr> <tr> <td>2-4</td> <td>44</td> <td>38</td> <td>$\frac{36}{38} = 0.9473\dots$</td> <td>50.94...</td> </tr> <tr> <td>4-6</td> <td>37</td> <td>38</td> <td>$\frac{1}{38} = 0.0263\dots$</td> <td>36.02...</td> </tr> <tr> <td>6-9</td> <td>52</td> <td>57</td> <td>$\frac{25}{57} = 0.4385\dots$</td> <td>47.43...</td> </tr> <tr> <td>9-12</td> <td>58</td> <td>57</td> <td>$\frac{1}{57} = 0.0175\dots$</td> <td>59.01...</td> </tr> </tbody> </table>						Class	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$	0-1	22	19	$\frac{9}{19} = 0.4736\dots$	25.57...	1-2	15	19	$\frac{16}{19} = 0.8421\dots$	11.84...	2-4	44	38	$\frac{36}{38} = 0.9473\dots$	50.94...	4-6	37	38	$\frac{1}{38} = 0.0263\dots$	36.02...	6-9	52	57	$\frac{25}{57} = 0.4385\dots$	47.43...	9-12	58	57	$\frac{1}{57} = 0.0175\dots$	59.01...	M1 A1
	Class	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$																																					
	0-1	22	19	$\frac{9}{19} = 0.4736\dots$	25.57...																																					
	1-2	15	19	$\frac{16}{19} = 0.8421\dots$	11.84...																																					
2-4	44	38	$\frac{36}{38} = 0.9473\dots$	50.94...																																						
4-6	37	38	$\frac{1}{38} = 0.0263\dots$	36.02...																																						
6-9	52	57	$\frac{25}{57} = 0.4385\dots$	47.43...																																						
9-12	58	57	$\frac{1}{57} = 0.0175\dots$	59.01...																																						
<p>H_0: <u>continuous uniform</u> distribution <u>is</u> a good fit</p>						B1																																				
<p>H_1: <u>continuous uniform</u> distribution <u>is not</u> a good fit</p>																																										
$\sum \frac{(O_i - E_i)^2}{E_i} = \frac{313}{114} = 2.75 \quad \text{or} \quad \sum \frac{O_i^2}{E_i} - 228 = 230.745\dots - 228 = \dots \quad (\text{awrt } \mathbf{2.75})$						dM1A1																																				
$\nu = 6 - 1 = 5$						B1																																				
$\chi_5^2(0.05) = 11.070$ (ft their ν i.e. $\chi_\nu^2(0.05)$)						B1ft																																				
<p>2.75 < 11.070, insufficient evidence to reject H_0</p>						M1																																				
<p>Continuous uniform distribution is a suitable model</p>						A1																																				
12																																										
<p>1st M1 for calculation of at least 3 widths and attempting proportions/probs. <u>or</u> for 1:2:3 ratio seen 1st A1 for correct probabilities 2nd A1 for all correct expected frequencies 2nd M1 for attempting $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$, at least 3 correct expressions or values. Follow through their E_i provided they are not all = 38 3rd A1 for a correct set of calcs - 3rd or 4th column. (2 dp or better and allow e.g. 0.94...) 3rd dM1 dependent on 2nd M1 for attempting a correct sum or calculation (must see at least 3 terms and +) The first three Ms and As can be implied by a test statistic of awrt 2.75 4th M1 for a correct statement based on their test statistic (> 1) and their cv (> 3.8) Contradictory statements score M0 e.g. "significant" do not reject H_0. 5th A1 for a correct comment suggesting that continuous uniform model is suitable. No ft</p>																																										

Question Number	Scheme	Marks
Q7	<p>(a) Label full time staff 1-6000, part time staff 1-4000 Use random numbers to select Simple random sample of 120 full time staff and 80 part time staff</p> <p>(b) Enables estimation of statistics / errors for each strata <u>or</u> “reduce variability” <u>or</u> “more representative” <u>or</u> “reflects population structure” NOT “more accurate”</p> <p>(c) $H_0: \mu_f = \mu_p, \quad H_1: \mu_f \neq \mu_p$ (accept μ_1, μ_2) $\text{s.e.} = \sqrt{\frac{21}{80} + \frac{19}{80}}, \quad z = \frac{52 - 50}{\sqrt{\frac{21}{80} + \frac{19}{80}}} = (2\sqrt{2})$ $= 2.828\dots$ (awrt 2.83) Two tailed critical value $z = 2.5758$ (or prob of awrt 0.002 (<0.005) or 0.004 (<0.01)) [2.828 > 2.5758 so] significant evidence to reject H_0 There is evidence of a difference in policy awareness between full time and part time staff</p> <p>(d) Can use mean full time and mean part time ~ Normal</p> <p>(e) Have assumed $s^2 = \sigma^2$ or variance of sample = variance of population</p> <p>(f) $2.53 < 2.5758$, not significant <u>or</u> do not reject H_0 So there is insufficient evidence of a difference in mean awareness</p> <p>(g) Training course has closed the gap between full time staff and part time staff’s mean awareness of company policy.</p>	<p>M1 M1 A1 (3)</p> <p>B1 (1)</p> <p>B1 M1,M1 A1 B1 dM1 A1ft (7)</p> <p>B1 B1 (2)</p> <p>B1 (1)</p> <p>M1 A1ft (2)</p> <p>B1 (1)</p> <p>17</p>
	<p>(a) 1st M1 for attempt at labelling full-time and part-time staff. One set of correct numbers. 2nd M1 for mentioning use of random numbers 1st A1 for s.r.s. of 120 full-time and 80 part-time</p> <p>(c) 1st M1 for attempt at s.e. - condone one number wrong . NB correct s.e. = $\sqrt{\frac{1}{2}}$ 2nd M1 for using their s.e. in correct formula for test statistic. Must be $\frac{\pm(52 - 50)}{\sqrt{\frac{p}{q} + \frac{r}{s}}}$ 3rd dM1 dep. on 2nd M1 for a correct statement based on their normal cv and their test statistic 2nd A1 for correct comment in context. Must mention “scores” or “policy awareness” and types of “staff”. Award A0 for a one-tailed comment. Allow ft</p> <p>(d) 1st B1 for mention of mean(s) <u>or</u> use of \bar{X}, provided \bar{X} clearly refers to full-time or part-time 2nd B1 for stating that distribution can be assumed normal e.g. “mean score of the test is normally distributed” gets B1B1</p> <p>(f) M1 for correct statement (may be implied by correct contextualised comment) A1 for correct contextualised comment. Accept “no difference in mean scores”. Allow ft</p> <p>(g) B1 for correct comment in context that implies training was effective. This must be supported by their (c) and (f). Condone one-tailed comment here.</p>	

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481

Email publications@linneydirect.com

Order Code UA024774 Summer 2010

For more information on Edexcel qualifications, please visit www.edexcel.com/quals

Edexcel Limited. Registered in England and Wales no.4496750
Registered Office: One90 High Holborn, London, WC1V 7BH

Mark Scheme (Results)

June 2011

GCE Statistics S3 (6691) Paper 1

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information, please call our GCE line on 0844 576 0025 or visit our website at www.edexcel.com.

If you have any subject specific questions about the content of this Mark Scheme that require the help of a subject specialist, you may find our **Ask The Expert** email service helpful.

Ask The Expert can be accessed online at the following link:
<http://www.edexcel.com/Aboutus/contact-us/>

June 2011

Publications Code UA028846

All the material in this publication is copyright

© Edexcel Ltd 2011

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - M marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - A marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - B marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes and can be used if you are using the annotation facility on ePEN.

- bod – benefit of doubt
- ft – follow through
- the symbol \checkmark will be used for correct ft
- cao – correct answer only
- cso - correct solution only. There must be no errors in this part of the question to obtain this mark
- isw – ignore subsequent working
- awrt – answers which round to
- SC: special case
- oe – or equivalent (and appropriate)
- dep – dependent
- indep – independent
- dp decimal places
- sf significant figures
- * The answer is printed on the paper
- \square The second mark is dependent on gaining the first mark

June 2011
Statistics S3 6691
Mark Scheme

Question Number	Scheme	Marks
1.	X_1, X_2, \dots, X_n is a random sample of size n , for large n , drawn from a population of any distribution with mean μ and variance σ^2 then \bar{X} is (approximately) $N\left(\mu, \frac{\sigma^2}{n}\right)$	B1 B1 B1 (3) 3
	1 st B for large sample or equivalent 2 nd B for ‘population of any distribution’ or ‘any population’ 3 rd B require mean or symbol and normal (parameters not required)	

Question Number	Scheme	Marks																																								
2. (a)	<table border="1" data-bbox="296 376 1161 591"> <thead> <tr> <th>Town</th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <td><i>h</i> rank</td> <td>1</td> <td>5</td> <td>2</td> <td>3</td> <td>7</td> <td>4</td> <td>6</td> </tr> <tr> <td><i>c</i> rank</td> <td>4</td> <td>3</td> <td>2</td> <td>1</td> <td>6</td> <td>7</td> <td>5</td> </tr> <tr> <td>d</td> <td>3</td> <td>2</td> <td>0</td> <td>2</td> <td>1</td> <td>3</td> <td>1</td> </tr> <tr> <td>d^2</td> <td>9</td> <td>4</td> <td>0</td> <td>4</td> <td>1</td> <td>9</td> <td>1</td> </tr> </tbody> </table> $\sum d^2 = 28$ $r_s = 1 - \frac{6 \times 28}{7 \times 48}$ $= 0.5$	Town	A	B	C	D	E	F	G	<i>h</i> rank	1	5	2	3	7	4	6	<i>c</i> rank	4	3	2	1	6	7	5	$ d $	3	2	0	2	1	3	1	d^2	9	4	0	4	1	9	1	M1 M1 M1A1 M1 A1 (6)
Town	A	B	C	D	E	F	G																																			
<i>h</i> rank	1	5	2	3	7	4	6																																			
<i>c</i> rank	4	3	2	1	6	7	5																																			
$ d $	3	2	0	2	1	3	1																																			
d^2	9	4	0	4	1	9	1																																			
(b)	$H_0 : \rho = 0, H_1 : \rho \neq 0$ Critical values are $r_s = \pm 0.7857$ $0.5 < 0.7857$ insufficient evidence to reject H_0 Councillor's claim is supported.	B1 B1ft M1 A1ft (4) 10																																								

Question Number	Scheme	Marks
<p>Notes (a)</p> <p>(b)</p>	<p>1st M1 for an attempt to rank the hardship against calls</p> <p>2nd M1 for attempting d for their ranks. Must be using ranks.</p> <p>3rd M1 for attempting $\sum d^2$ (must be using ranks)</p> <p>1st A1 for sum of 28 (or 84)</p> <p>4th M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.</p> <p>2nd A1 for awrt 0.5 (or -0.5)</p> <p>1st B1 for both hypotheses in terms of ρ, H_1 must be two tail.</p> <p>2nd B1 for cv of ± 0.7857 (or 0.7143 to ft from 1-tailed H_1)</p> <p>M1 for a correct statement relating their r_s with their cv but cv must be such that $cv < 1$</p> <p>A1ft for a correct contextualised comment. Must mention “Councillor” and “claim” <u>or</u> “hardship” and “number of calls (to the emergency services)”</p> <p>Follow through their r_s and their cv (provided it is $cv < 1$)</p> <p>Condone use of “association” in conclusion for A1</p> <p>Condone ‘positive’ in conclusion.</p>	

Question Number	Scheme	Marks																																																
3.	<table border="1" data-bbox="292 333 1131 562"> <thead> <tr> <th>Defect Type Shift</th> <th>D₁</th> <th>D₂</th> <th></th> </tr> </thead> <tbody> <tr> <td>First Shift</td> <td>47.25</td> <td>15.75</td> <td>63</td> </tr> <tr> <td>Second Shift</td> <td>56.25</td> <td>18.75</td> <td>75</td> </tr> <tr> <td>Third Shift</td> <td>46.5</td> <td>15.5</td> <td>62</td> </tr> <tr> <td></td> <td>150</td> <td>50</td> <td>200</td> </tr> </tbody> </table> <p data-bbox="276 640 1042 674">H₀ : Type of defect is independent of Shift (no association)</p> <p data-bbox="276 689 1046 723">H₁ : Type of defect is not independent of Shift (association)</p> <table border="1" data-bbox="292 804 1121 1133"> <thead> <tr> <th><i>O</i></th> <th><i>E</i></th> <th>$\frac{(O-E)^2}{E}$</th> <th>$\frac{O_i^2}{E_i}$</th> </tr> </thead> <tbody> <tr> <td>45</td> <td>47.25</td> <td>0.1071...</td> <td>42.857...</td> </tr> <tr> <td>18</td> <td>15.75</td> <td>0.3214...</td> <td>20.571..</td> </tr> <tr> <td>55</td> <td>56.25</td> <td>0.02777...</td> <td>53.777...</td> </tr> <tr> <td>20</td> <td>18.75</td> <td>0.0833...</td> <td>21.333...</td> </tr> <tr> <td>50</td> <td>46.5</td> <td>0.2634...</td> <td>53.763...</td> </tr> <tr> <td>12</td> <td>15.5</td> <td>0.7903...</td> <td>9.290...</td> </tr> </tbody> </table> <p data-bbox="276 1178 973 1261">$\frac{(O-E)^2}{E} = 1.5934..$ or $\frac{O_i^2}{E_i} - 200 = 201.5934 - 200 = 1.5934..$</p> <p data-bbox="276 1272 528 1305">$\nu = (3-1)(2-1) = 2$</p> <p data-bbox="276 1317 496 1350">$\chi^2_2(0.10) = 4.605$</p> <p data-bbox="276 1361 895 1395">1.59 < 4.605 so insufficient evidence to reject H₀</p> <p data-bbox="276 1406 1002 1440">Insufficient evidence to support manager's belief /claim.</p>	Defect Type Shift	D ₁	D ₂		First Shift	47.25	15.75	63	Second Shift	56.25	18.75	75	Third Shift	46.5	15.5	62		150	50	200	<i>O</i>	<i>E</i>	$\frac{(O-E)^2}{E}$	$\frac{O_i^2}{E_i}$	45	47.25	0.1071...	42.857...	18	15.75	0.3214...	20.571..	55	56.25	0.02777...	53.777...	20	18.75	0.0833...	21.333...	50	46.5	0.2634...	53.763...	12	15.5	0.7903...	9.290...	<p data-bbox="1318 528 1406 562">M1A1</p> <p data-bbox="1318 674 1358 707">B1</p> <p data-bbox="1318 1099 1406 1133">M1A1</p> <p data-bbox="1318 1178 1358 1211">A1</p> <p data-bbox="1318 1272 1358 1305">B1</p> <p data-bbox="1318 1317 1382 1350">B1ft</p> <p data-bbox="1318 1361 1358 1395">M1</p> <p data-bbox="1318 1406 1358 1440">A1</p> <p data-bbox="1433 1451 1473 1485">10</p>
Defect Type Shift	D ₁	D ₂																																																
First Shift	47.25	15.75	63																																															
Second Shift	56.25	18.75	75																																															
Third Shift	46.5	15.5	62																																															
	150	50	200																																															
<i>O</i>	<i>E</i>	$\frac{(O-E)^2}{E}$	$\frac{O_i^2}{E_i}$																																															
45	47.25	0.1071...	42.857...																																															
18	15.75	0.3214...	20.571..																																															
55	56.25	0.02777...	53.777...																																															
20	18.75	0.0833...	21.333...																																															
50	46.5	0.2634...	53.763...																																															
12	15.5	0.7903...	9.290...																																															

Question Number	Scheme	Marks
Notes	<p>1st M1 for some use of $\frac{\text{Row Total} \times \text{Col.Total}}{\text{Grand Total}}$ May be implied by correct E_i</p> <p>1st A1 for all expected frequencies correct</p> <p>B1 for both hypotheses. Must mention “defect” and “shift” at least once</p> <p>Use of “relationship” or “correlation” or “connection” is B0</p> <p>2nd M1 for at least two correct terms (as in 3rd or 4th column) or correct expressions with their E_i</p> <p>2nd A1 for all correct terms. May be implied by a correct answer.(2 dp or better-allow eg 0.10...)</p> <p>3rd M1 for a correct statement linking their test statistic and their cv . Must be χ^2 not normal.</p> <p>4th A1 for a correct comment in context - must mention “manager’s belief” or “shift” and “defect type” - condone “relationship” or “connection” here but not “correlation”. No follow through e.g. “There is evidence of a relationship between shift and type of defect”</p>	

Question Number	Scheme	Marks
4. (a)	$\bar{x} = \frac{5320}{80} = 66.5$ $s^2 = \frac{392000 - 80 \times (66.5)^2}{79}$ $= 483.797\dots$ <p style="text-align: right;">awrt 484</p>	M1,A1 M1A1ft A1 (5)
(b)	$H_0: \mu_m = \mu_{nm}, \quad H_1: \mu_m > \mu_{nm}$ <p>(accept μ_1, μ_2 with definition)</p> $z = \frac{69.0 - 66.5}{\sqrt{\frac{483.797}{80} + \frac{446.44}{60}}}$ $= 0.6807$ <p>One tailed cv 1.6449 0.752)</p> <p>0.6807 < 1.6449 (or 0.248 > 0.05) insufficient evidence to reject H_0 Mean money spent is not greater with music playing.</p> <p style="text-align: right;">awrt 0.681 (Probability is awrt</p>	B1B1 M1dM1 A1 B1 dM1 A1ft (8) 13

Question Number	Scheme	Marks
(b)	<p style="text-align: center;">Notes</p> <p>No definition award B1B0. 1st M1 for attempt at s.e. - condone one number wrong or switched 60 & 80 . 2nd dM1 for using their s.e. in correct formula for test statistic. 3rd dM1 dep. on 2nd M1 for a correct statement based on their normal cv and their test statistic 2nd A1 for correct comment in context. Must mention “money spent” and “music playing”. Allow ft.</p> <p>Critical Region for (b) Standard error x z value for 2nd M1 Standard error x 1.6449= awrt 6.04 for 1st A1 2.5<6.04</p>	

Question Number	Scheme									Marks
5.										
(a)	Hurricanes: occur singly / are independent or occur at random / are a rare event / at a constant rate									B1B1 (2)
(b)	From data $\frac{1 \times 2 + 2 \times 5 + 3 \times 17 + \dots + 7 \times 12}{80} = 4.4875$									M1A1 (2)
(c)	No of hurricanes, h	0	1	2	3	4	5	6	7+	M1A1A1 (3)
	$80P(X = h)$	0.9	4.038	$r=9.06\dots$	13.55	$s=15.205$	13.647...	10.206...	13.388...	
	Combine to give expected frequencies >5	13.9991		13.55	15.205...	13.647...	10.206...	13.388...		
(d)	Observed	7		17	20	12	12	12		M1
	$\frac{(O - E)^2}{E}$	3.499...		0.876...	1.511...	0.198...	0.315...	0.143...		
	$\frac{O_i^2}{E_i}$	3.500...		21.322...	26.306...	10.551...	14.108...	10.755..		
	H_0 : Poisson distribution is a good fit o.e. H_1 : Poisson distribution is not a good fit o.e. $\sum \frac{(O_i - E_i)^2}{E_i} = 6.545\dots$ or $\frac{O_i^2}{E_i} = 86.545 - 80 = 6.545\dots$ (awrt 6.55 or 6.54) $\nu = 6 - 2 = 4$ cv is 9.488 (ft their ν i.e. $\chi_{\nu}^2(0.05)$ $6.545 < 9.488$ so insufficient evidence to reject H_0 (Hurricanes) can be modelled by a Poisson distribution									B1 A1 B1 B1ft A1
										(6) 13

Question Number	Scheme	Marks
<p>(b)</p> <p>(c)</p> <p>(d)</p>	<p style="text-align: center;">Notes</p> <p>M for at least 2 terms on numerator. 359/80 only award M0A0</p> <p>M for 80xPoisson probability with 4.4875 and either 2 or 4.</p> <p>1st A1 for awrt 9.06 and 2nd A1 for awrt 15.20 or 15.21</p> <p>1st M1 for some pooling and attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$, at least 3 correct expressions or values.</p> <p>1st B1 no value for parameter permitted</p> <p>2nd A1 for a correct comment suggesting that Poisson model is suitable.</p> <p>No ft</p>	

Question Number	Scheme	Marks
6. (a)	$L = A_1 + A_2 + \dots + A_6$ Mean is $E(L) = 6 \times 20 = 120$ Standard deviation is $\sqrt{\text{Var}(W)} = \sqrt{6 \times 5^2} = 5\sqrt{6} = 12.247\dots$ 12.2	B1 B1 (2)
(b)	$P(L > 110) = P\left(Z > \left(\frac{110 - 120}{12.247\dots}\right)\right)$ $= P(Z < 0.8164\dots)$ $= 0.7939 \text{ (or } 0.7929 \text{ using interpolation or } 0.79289 \text{ by calc)}$	M1 A1 (2)
(c)	Let $X = 4B - \sum_1^6 A_i$ $E(X) = 140 - 120 = 20$ $\text{Var}(X) = 16 \times 8^2 + 6 \times 5^2 = 1174$ $P(X < 0) = P\left(Z < \frac{-20}{\sqrt{1174}}\right) = P(Z < -0.583\dots)$ $= 0.2797 \text{ (or } 0.2810 \text{ if no interpolation) or } 0.27971 \text{ by calc.}$	B1 M1M1A1 M1 A1 (6) 10

Question Number	Scheme	Marks
	Notes	
(b)	M1 for identifying a correct probability (they must have the 110) and attempting to standardise with their mean and sd. This can be implied by the correct answer. A1 for awrt 0.794 or 0.793	
(c)	Accept ± 20 for B mark. Only award for probability statement if 2 terms in var 1 st M1 for 1024, 2 nd M1 for 150 3 rd M for standardising with their mean and 2 term sd and finding probability < 0.5 2 nd A1 for awrt 0.280 or 0.281	

Question Number	Scheme	Marks
7. (a)	$H_0: \mu = 250, H_1: \mu < 250,$ $z = \frac{248 - 250}{\frac{5.4}{\sqrt{90}}}$ $= -3.513\dots$ <p>awrt - 3.51 Critical value -1.6449 -3.513.. < -1.6449 so sufficient evidence to reject H_0 Manager's claim is justified.</p>	B1 M1 A1 B1 A1 (5)
(b)	98% CI for μ is $248 \pm 2.3263 \times \frac{5.4}{\sqrt{90}}$ <p>= awrt (247,249) 2.33</p> <p>dependent upon z value awrt</p>	M1B1 A1A1 (4)
(c)	Hypothesis test is significant or CI does not contain stated weight. (Manager should ask the company to investigate if their) stated weight is too high o.e.	B1 B1 (2)
(d)	$P(\bar{x} - \mu < 1) = 0.98$ $\frac{1}{\frac{3}{\sqrt{n}}} = 2.3263$ $n = (3 \times 2.3263)^2 = 48.7\dots$ <p>Sample size 49 required.</p>	M1 A1 dM1A1 A1 (5) 16

Question Number	Scheme	Marks
	Notes	
(a)	1 st B1 for H_0 and for H_1 (must be <250) They must use μ not x , p , λ or \bar{x} etc	
(b)	1 st M1 for attempt at standardising using 248, 250 and sd. Can accept \pm . Critical region: $250 - 0.936 = 249.064$ for M1A1 (and compare with 248.)	
(d)	3 rd B1 for ± 1.6449 seen (or probability of 0.0002 or better)	
	2 nd A1 for a correct contextualised comment. Must mention "Manager" and "claim" <u>or</u> "weight" and "stated weight". No follow through.	
	2.3263 or better for B mark. Any z value replacing 2.3263 award M.	
	1 st M for LHS = z value >1	
	1 st A for RHS awrt 2.33	
	2 nd A1 for answers in the range 48.7-48.9	
	3 rd A1 don't condone \geq	

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467

Fax 01623 450481

Email publication.orders@edexcel.com

Order Code UA028846 June 2011

For more information on Edexcel qualifications, please visit
www.edexcel.com/quals

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual




Llywodraeth Cynulliad Cymru
Welsh Assembly Government





Mark Scheme (Results)

Summer 2012

GCE Statistics S3 (6691) Paper 1

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2012

Publications Code UA033146

All the material in this publication is copyright

© Pearson Education Ltd 2012

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

Hypothesis Tests (Final M1A1)

For an incorrect comparison (e.g. probability with z value) even with a correct statement and/or comment award MOA0

For a correct or no comparison with more than one statement one of which is false
Award MOA0 (This is compatible with the principle above of contradictory statements being penalised)

Apply these rules to all questions

June 2012
6691 Statistics S3
Mark Scheme

Question Number	Scheme	Marks																																																						
1 (a)	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;">X</th> <th style="text-align: center;">Y</th> <th style="text-align: center;">Rank X</th> <th style="text-align: center;">Rank Y</th> <th style="text-align: center;">d</th> <th style="text-align: center;">d²</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">62</td><td style="text-align: center;">54</td><td style="text-align: center;">3</td><td style="text-align: center;">2</td><td style="text-align: center;">1</td><td style="text-align: center;">1</td></tr> <tr><td style="text-align: center;">56</td><td style="text-align: center;">47</td><td style="text-align: center;">4</td><td style="text-align: center;">5</td><td style="text-align: center;">-1</td><td style="text-align: center;">1</td></tr> <tr><td style="text-align: center;">87</td><td style="text-align: center;">71</td><td style="text-align: center;">1</td><td style="text-align: center;">1</td><td style="text-align: center;">0</td><td style="text-align: center;">0</td></tr> <tr><td style="text-align: center;">54</td><td style="text-align: center;">50</td><td style="text-align: center;">5</td><td style="text-align: center;">3</td><td style="text-align: center;">2</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">65</td><td style="text-align: center;">49</td><td style="text-align: center;">2</td><td style="text-align: center;">4</td><td style="text-align: center;">-2</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">15</td><td style="text-align: center;">25</td><td style="text-align: center;">6</td><td style="text-align: center;">8</td><td style="text-align: center;">-2</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">12</td><td style="text-align: center;">30</td><td style="text-align: center;">7</td><td style="text-align: center;">7</td><td style="text-align: center;">0</td><td style="text-align: center;">0</td></tr> <tr><td style="text-align: center;">10</td><td style="text-align: center;">44</td><td style="text-align: center;">8</td><td style="text-align: center;">6</td><td style="text-align: center;">2</td><td style="text-align: center;">4</td></tr> </tbody> </table>	X	Y	Rank X	Rank Y	d	d ²	62	54	3	2	1	1	56	47	4	5	-1	1	87	71	1	1	0	0	54	50	5	3	2	4	65	49	2	4	-2	4	15	25	6	8	-2	4	12	30	7	7	0	0	10	44	8	6	2	4	M1 M1
X	Y	Rank X	Rank Y	d	d ²																																																			
62	54	3	2	1	1																																																			
56	47	4	5	-1	1																																																			
87	71	1	1	0	0																																																			
54	50	5	3	2	4																																																			
65	49	2	4	-2	4																																																			
15	25	6	8	-2	4																																																			
12	30	7	7	0	0																																																			
10	44	8	6	2	4																																																			
	$\sum d^2 = 18$ $r_s = 1 - \frac{6 \times 18}{8(64-1)} = 0.7857...$	A1 M1A1 (5)																																																						
1(b)	$H_0 : \rho = 0$ $H_0 : \rho > 0$ Critical region $r_s > 0.6429$ (0.7857 > 0.6429 sufficient evidence to) reject H_0 There is evidence of agreement between the scores awarded by each manager	B1 B1 B1 M1 A1ft (5)																																																						
1(c)	(A and D are now) tied ranks (for Manager Y) Average rank (awarded to A and D) and use $r_s = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$	B1 B1 (2)																																																						
1(a)	Notes 1st M1 for an attempt to rank score X and score Y 2nd M1 for attempting d ² for their ranks. Must be using ranks. 1st A1 for sum of 18 3rd M1 for use of the correct formula with their $\sum d^2$. If answer is not correct an expression is required.	B1 B1 (2)																																																						
1(b)	2nd A1 for awrt 0.786 1st B1 for null hypotheses in terms of ρ or ρ_s 2nd B1 for alt hyp as given 3rd B1 for cv of +0.6429 (or 0.7381 if two tailed from hyp) M1 for a correct statement relating their r_s with their cv but cv must be such that $ cv < 1$ A1ft for a correct contextualised comment. Must mention “scores / rankings” and “manager” Follow through their r_s and their cv (provided it is $ cv < 1$ Use of “association” is A0	B1 B1 B1 M1 A1ft A0																																																						
1(c)	1st B1 Tied ranks can be implied by 2.5, 6.5 or both 2 or 6 or description. 2nd B1 Average rank implied by 2.5 or 6.5 or description and ‘use of pmcc’.	B1 B1																																																						
Total 12																																																								

Question Number	Scheme	Marks								
2(a)	Sampling frame within each species of fish in the lake impossible to obtain.	B1 (1)								
2(b)	Quota sampling	B1 (1)								
2(c)	Advantages: Sample can be obtained quickly Costs are kept to a minimum Administration of survey is easy Disadvantages: Not possible to estimate sampling errors Process not random Surveyor may not be able to identify species of fish easily	B1 B1 (2)								
2(d)	<table border="1" data-bbox="408 745 1150 1037"> <thead> <tr> <th data-bbox="408 745 778 786">Species</th> <th data-bbox="778 745 1150 786">Quota</th> </tr> </thead> <tbody> <tr> <td data-bbox="408 786 778 869">Trout</td> <td data-bbox="778 786 1150 869">$\frac{1400}{2450} \times 30 = 17.14$</td> </tr> <tr> <td data-bbox="408 869 778 952">Bass</td> <td data-bbox="778 869 1150 952">$\frac{600}{2450} \times 30 = 7.35$</td> </tr> <tr> <td data-bbox="408 952 778 1037">Pike</td> <td data-bbox="778 952 1150 1037">$\frac{450}{2450} \times 30 = 5.51$</td> </tr> </tbody> </table> <p data-bbox="209 1077 1350 1144">Fish are caught from the lake until the quota of 17 trout, 7 bass and 6 pike are reached.</p> <p data-bbox="209 1144 1350 1211">If a fish is caught and the species quota is full, then this is ignored.</p> <p data-bbox="209 1301 1350 1335">Notes</p>	Species	Quota	Trout	$\frac{1400}{2450} \times 30 = 17.14$	Bass	$\frac{600}{2450} \times 30 = 7.35$	Pike	$\frac{450}{2450} \times 30 = 5.51$	B1B1B1 B1 (4) Total 8
Species	Quota									
Trout	$\frac{1400}{2450} \times 30 = 17.14$									
Bass	$\frac{600}{2450} \times 30 = 7.35$									
Pike	$\frac{450}{2450} \times 30 = 5.51$									
2(a)	‘You can’t / it’s very difficult to number all the fish’ or equivalent									
2(c)	Correct answer to (b) required. Some detail required.									
2(d)	1 st B1 any one correct calculation seen or implied 2 nd B1 all correct to at least 1 dp 3 rd B1 for 17,7,6 4 th B1 accept equivalent statement. Require comment on what to do with ‘extra fish’.									

Question Number	Scheme	Marks
3(a)	($X_1, X_2, X_3, \dots, X_n$ is a random) sample of size n , for n is large , (from a population with mean μ and variance σ^2) then \bar{X} is (approximately) Normal.	B1 B1
		(2)
3 (b)	$\bar{x} = \frac{1740000}{100} = 17400$	B1
	$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}, = 17400 \pm 1.96 \times \frac{5000}{\sqrt{100}}$ [16420,18380]	M1, B1 A1A1
3(c)	\bar{X} : Normal (approx) by CLT, and normal needed to find CI.	B1,B1 (5) (2)
3 (d)	20000 above upper confidence limit (not just outside) Complaint justified.	B1ft dB1ft (2)
3(b)	Notes Recognisable z value required for method. 2 nd B1 1.96 or better seen award Final A1s accept 3sf if correct expression seen. 5/5 for [16420,18380]	Total 11

Question Number	Scheme	Marks																																																
4	<p>H_0 : Egg yield and breed of chicken are independent (not associated) H_1 : Egg yield and breed of chicken are dependent (associated)</p> <table border="1" data-bbox="316 405 1235 672"> <thead> <tr> <th>Egg Yield Breed</th> <th>Low</th> <th>Medium</th> <th>High</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Leghorn</td> <td>$\frac{100 \times 36}{150} = 24$</td> <td>$\frac{100 \times 84}{150} = 56$</td> <td>$\frac{100 \times 30}{150} = 20$</td> <td>100</td> </tr> <tr> <td>Cornish</td> <td>$\frac{50 \times 36}{150} = 12$</td> <td>$\frac{50 \times 84}{150} = 28$</td> <td>$\frac{50 \times 30}{150} = 10$</td> <td>50</td> </tr> <tr> <td>Total</td> <td>36</td> <td>84</td> <td>30</td> <td>150</td> </tr> </tbody> </table> <table border="1" data-bbox="225 779 1348 1099"> <thead> <tr> <th>O</th> <th>E</th> <th>$\sum \frac{(O-E)^2}{E}$</th> <th>$\sum \frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>22</td> <td>24</td> <td>0.166667</td> <td>20.2</td> </tr> <tr> <td>52</td> <td>56</td> <td>0.285714</td> <td>48.3</td> </tr> <tr> <td>26</td> <td>20</td> <td>1.8</td> <td>33.8</td> </tr> <tr> <td>14</td> <td>12</td> <td>0.333333</td> <td>16.3</td> </tr> <tr> <td>32</td> <td>28</td> <td>0.571429</td> <td>36.6</td> </tr> <tr> <td>4</td> <td>10</td> <td>3.6</td> <td>1.6</td> </tr> </tbody> </table> <p>$\sum \frac{(O-E)^2}{E} = 6.757... \text{ or } \sum \frac{O^2}{E} - 100 = 6.757...$ $\nu = 2, \chi_2^2(5\%) = 5.991$ (6.757 > 5.991 so sufficient evidence to) reject H_0 Egg yield and breed of chicken are dependent (associated)</p> <p>Notes B1 for both hypotheses. Must mention “yield” and “breed” in both but condone ditto marks. Use of “relationship” or “correlation” or “connection” is B0 1st M1 for some use of $\frac{\text{Row Total} \times \text{Col. Total}}{\text{Grand Total}}$. May be implied by a correct E_i 1st A1 for all expected frequencies correct 2nd M1 for at least two correct terms or correct expressions with their E_i 2nd A1 for all correct terms. May be implied by a correct answer (2 sf or better) 3rd M1 for a correct statement linking their test statistic and their cv. Must be χ^2 not normal. 4th A1 for a correct comment in context - must mention “egg yield” and “breed of chicken” - condone “relationship” or “connection” here but not “correlation”. No follow through e.g. “There is no evidence of a relationship between egg yield and breed of chicken” is A0 whatever their test stat or cv.</p>	Egg Yield Breed	Low	Medium	High	Total	Leghorn	$\frac{100 \times 36}{150} = 24$	$\frac{100 \times 84}{150} = 56$	$\frac{100 \times 30}{150} = 20$	100	Cornish	$\frac{50 \times 36}{150} = 12$	$\frac{50 \times 84}{150} = 28$	$\frac{50 \times 30}{150} = 10$	50	Total	36	84	30	150	O	E	$\sum \frac{(O-E)^2}{E}$	$\sum \frac{O^2}{E}$	22	24	0.166667	20.2	52	56	0.285714	48.3	26	20	1.8	33.8	14	12	0.333333	16.3	32	28	0.571429	36.6	4	10	3.6	1.6	<p>B1 M1A1 M1A1 A1 B1B1ft M1 A1 (10) Total 10</p>
Egg Yield Breed	Low	Medium	High	Total																																														
Leghorn	$\frac{100 \times 36}{150} = 24$	$\frac{100 \times 84}{150} = 56$	$\frac{100 \times 30}{150} = 20$	100																																														
Cornish	$\frac{50 \times 36}{150} = 12$	$\frac{50 \times 84}{150} = 28$	$\frac{50 \times 30}{150} = 10$	50																																														
Total	36	84	30	150																																														
O	E	$\sum \frac{(O-E)^2}{E}$	$\sum \frac{O^2}{E}$																																															
22	24	0.166667	20.2																																															
52	56	0.285714	48.3																																															
26	20	1.8	33.8																																															
14	12	0.333333	16.3																																															
32	28	0.571429	36.6																																															
4	10	3.6	1.6																																															

Question Number	Scheme	Marks
5(a)	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A \neq \mu_B$ $z = \frac{\pm(80 - 74)}{\sqrt{\frac{100}{29} + \frac{225}{26}}}$ $z = \pm 1.7247\dots$ <p style="text-align: right;">awrt ± 1.72</p> <p>$1.7247 > 1.6449$ o.e. so reject H_0</p> <p>There is evidence of a difference in the (mean) scores of their students.</p>	<p>B1</p> <p>M1A1</p> <p>A1</p> <p>dM1</p> <p>A1</p> <p style="text-align: right;">(6)</p>
5(b)	<p>(For $z=1.6$, test above not significant so no evidence of a difference.)</p> <p>For Mr A's claim, $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A > \mu_B$, and critical value is $z=1.2816$</p> <p>(Both z values significant,) Mr Alan's claim supported.</p>	<p>B1, B1</p> <p>B1</p> <p style="text-align: right;">(3)</p>
5(a)	<p>Notes</p> <p>1st M1 for attempt at s.e. (condone one number wrong) and for using their s.e. in correct formula for test statistic.</p> <p>1st A1 for correct expression for se</p> <p>2nd dM1 dep. on 1st M1 for a correct statement based on their normal cv and their test statistic</p> <p>3rd A1 for correct comment in context. Must mention "scores" and "students / groups/classes" Award A0 for a one-tailed comment.</p>	<p style="text-align: right;">Total 9</p>
5(b)	<p>1st B1 Alternative hyp should be clearly defined</p>	

Question Number	Scheme	Marks																		
6(a)	$\text{Mean} = \frac{1 \times 16 + 2 \times 20 + \dots + 6 \times 8}{100} = 2.91 \text{ **ag**}$	M1A1 (2)																		
6(b)	$p = \frac{2.91}{6} = 0.485$ $a = 100 \times C_3^6 \times 0.485^3 \times 0.515^3 = 31.17$ $b = 100 \times 0.485^6 = 1.3(0)$	B1 M1A1 A1 (4)																		
6(c)	<p>H_0 : Binomial is a good fit H_1 : Binomial is a not a good fit</p>	B1																		
	<table border="1" data-bbox="225 786 1334 958"> <thead> <tr> <th data-bbox="225 786 408 882">Number of defective items</th> <th data-bbox="408 786 592 882">0 or 1</th> <th data-bbox="592 786 775 882">2</th> <th data-bbox="775 786 959 882">3</th> <th data-bbox="959 786 1142 882">4</th> <th data-bbox="1142 786 1334 882">5 or 6</th> </tr> </thead> <tbody> <tr> <td data-bbox="225 882 408 920">O</td> <td data-bbox="408 882 592 920">22</td> <td data-bbox="592 882 775 920">20</td> <td data-bbox="775 882 959 920">23</td> <td data-bbox="959 882 1142 920">17</td> <td data-bbox="1142 882 1334 920">18</td> </tr> <tr> <td data-bbox="225 920 408 958">E</td> <td data-bbox="408 920 592 958">12.41</td> <td data-bbox="592 920 775 958">24.82</td> <td data-bbox="775 920 959 958">31.17</td> <td data-bbox="959 920 1142 958">22.01</td> <td data-bbox="1142 920 1334 958">9.59</td> </tr> </tbody> </table>	Number of defective items	0 or 1	2	3	4	5 or 6	O	22	20	23	17	18	E	12.41	24.82	31.17	22.01	9.59	M1
Number of defective items	0 or 1	2	3	4	5 or 6															
O	22	20	23	17	18															
E	12.41	24.82	31.17	22.01	9.59															
	$\sum \frac{(O - E)^2}{E} = \frac{(22 - 12.41)^2}{12.41} + \frac{(20 - 24.82)^2}{24.82} + \dots + \frac{(18 - 9.59)^2}{9.59} = 18.998\dots \text{ awrt } 19.0$ <p>$\nu = 5 - 2 = 3$ degrees of freedom $\chi_3^2(5\%) = 7.815$ $18.998\dots > 7.815$ so reject H_0 Binomial is a not a good fit (and is not a good model for the number of defective items in samples of size 6)</p>	M1A1 B1 B1ft M1 A1 (8)																		
6(a) 6(b) 6(c)	<p>Notes</p> <p>1st M At least 2 correct terms on numerator and 100 for denominator.</p> <p>0.485 can be implied by at least 1 correct answer.</p> <p>Accept awrt 2dp for final answers</p> <p>Clear use of Binomial and x100 required for method.</p> <p>Parameters in hyps award B0</p> <p>1st M1 for combining either 0 and 1 or 5 and 6 or both. Require at least 1 value in a combined correct.</p> <p>2nd M1 for attempting $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$, at least 2 correct expressions or values.</p> <p>2nd A1 for a correct comment suggesting that Binomial model is not suitable. No ft Condone parameters here.</p>	Total 14																		

Question Number	Scheme	Marks
7(a)	$M : N(177, 25), F : N(163, 16)$ $E(M - F) = 177 - 163 = 14$ $\text{Var}(M - F) = 25 + 16 = 41$ $M - F : N(14, 41)$ $P(M - F > 0) = P\left(Z > \frac{-14}{\sqrt{41}}\right) \text{ or } P\left(Z < \frac{14}{\sqrt{41}}\right)$ $= P(Z < 2.186\dots)$ $= 0.9854 \quad \text{or } 0.9856 \text{ by calculator} \quad \text{awrt } 0.985 \text{ or } 0.986$	B1 M1A1 M1 A1 (5)
7(b)	$W = M_1 + M_2 + \dots M_6 + F_1 + F_2 + \dots F_4$ $E(W) = 6 \times 177 + 4 \times 163$ $= 1714$ $\text{Var}(W) = 6 \times 25 + 4 \times 16$ $= 214$ $P(W < 1700) = P\left(Z < \frac{1700 - 1714}{\sqrt{214}}\right) \text{ or } P\left(Z > \frac{1714 - 1700}{\sqrt{214}}\right)$ $= P(Z < -0.957\dots)$ $= 1 - 0.8315$ $= 0.1685$ $\text{awrt } Z < -0.96 \text{ or } Z > 0.96$ $\text{awrt } 0.169$ <p>(0.1693 by calculator)</p>	B1 M1 A1 M1 A1 A1 (6) Total 11
7(a) and (b)	Notes Condone reversed sds for method in (b) Accept metres: 2.14 award M1A0 in metres. 2nd M1s for identifying a correct probability and attempting to standardise with their mean and sd. Require explicit sd or accept 1156 for M1A0. This can be implied by the correct answer.	

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com
Order Code UA033146 Summer 2012

For more information on Edexcel qualifications, please visit our website
www.edexcel.com

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual




Llywodraeth Cynulliad Cymru
Welsh Assembly Government





Mark Scheme (Results)

Summer 2013

GCE Statistics 3 (6691/01R)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2013

Publications Code UA037014

All the material in this publication is copyright

© Pearson Education Ltd 2013

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.
3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes:

- bod – benefit of doubt
 - ft – follow through
 - the symbol \checkmark will be used for correct ft
 - cao – correct answer only
 - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper
 - \square The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.
 5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
 6. If a candidate makes more than one attempt at any question:
 - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
 - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.
 7. Ignore wrong working or incorrect statements following a correct answer.
 8. In some instances, the mark distributions (e.g. M1, B1 and A1) printed on the candidate's response may differ from the final mark scheme.

Question Number	Scheme	Marks
1.	Label females 1 – 100 (or 0 – 99) and males 1 – 300 (or 0 – 299) Using <i>random numbers</i> for <u>each group</u> in range 1 – 100 (0 – 99) <u>select 15</u> females and using 1 – 300 (or 0 – 299) select <u>45</u> <u>males</u>	B1 B1 B1 [Total 3]
	Notes	
	1 st B1 for labelling\numbering\listing females <u>and</u> males 2 nd B1 for use of random numbers or “randomly select” in <u>each group</u> (may be implied) 3 rd B1 for selecting the correct number of females <u>and</u> males e.g. randomly select 45 males and 15 females scores 2 nd and 3 rd B marks since randomly selecting and the “each group” is implied, If using systematic sampling within each strata allow 1 st B1 and 3 rd B1 (if earned) but 2 nd B0	

Question Number	Scheme	Marks
2.	$X \sim N(40, 3^2) \quad \bar{X} \sim N\left(40, \frac{9}{n}\right)$ <p style="text-align: right;">(Condone $Y \sim N\left(40, \frac{9}{n}\right)$)</p> $P(\bar{X} > 42) = P\left(Z > \frac{42 - 40}{\sqrt{\frac{9}{n}}}\right)$ $\frac{42 - 40}{\sqrt{\frac{9}{n}}} \geq 1.6449$ $n \geq 6.087$ $n = 7$	B1 M1 B1 dM1 A1 [Total 5]
	<p>1st B1 for stating the correct distribution for \bar{X}. May be implied if correctly used in line 2 and no incorrect version seen elsewhere.</p> <p>1st M1 for an attempt to standardise with 42, 40 and their $\sqrt{\frac{9}{n}}$, must have n. Allow \pm</p> <p>2nd B1 for using $z = \pm 1.6449$ (or better)</p> <p>2nd dM1 Dep on 1st M1 for forming an equation in n or \sqrt{n}. Allow “=” or “<” i.e. setting their standardised expression = their z value ($z > 1.5$)</p> <p>A1 for $n = 7$ only The A1 must follow from correct working so e.g. $n < 6.087$ leading to $n = 7$ is A0</p>	

Question Number	Scheme											Marks
3 (a)	Town	A	B	C	D	E	F	G	H	I	J	M1
	Pop	1	2	3	4	5	6	7	8	9	10	
	Empl	2	1	3	5	4	6	10	8	9	7	
	d	1	1	0	1	1	0	3	0	0	3	
	d ²	1	1	0	1	1	0	9	0	0	9	
	$\sum d^2 = 22$											M1A1
	$r_s = 1 - \frac{6 \times 22}{10 \times 99}$											dM1
	$= \frac{143}{165} = 0.866$											awrt 0.867
(b)	$H_0: \rho = 0 ; H_1: \rho > 0$ CV = 0.6485											B1
	in critical region / significant/ reject H ₀											B1
	evidence of <u>positive</u> correlation between population and no. of employees											M1 A1 (4)
(c)	CV = <u>0.6319</u>											B1
	[not in critical region / not significant/ do not reject H ₀]											
	No evidence of <u>positive</u> correlation											B1 (2)
(d)	No evidence to suggest that as pop' increased the no. of employees increased <u>linearly</u> . Villages <u>ranked</u> highly for pop' were also <u>ranked</u> highly for the no. of employees.											B1 B1
												(2)
ALT	Alternate for part (d) if different conclusions in part (b) and part (c) Data probably not (bivariate) normal therefore Spearman's coefficient is more suitable than the product moment correlation coefficient.											[Total 13]
Notes												
(a)	1 st M1 for an attempt to rank no of employees against the populations 2 nd M1 for attempting $\sum d^2$ (must be using ranks) ft their ranks 1 st A1 for 22 3 rd dM1 dep on 1 st M1 for use of the correct formula with their $\sum d^2$. If ans. is not correct an expr' is required.											
(b)	1 st B1 for both hypotheses in terms of ρ , H ₁ must be one tail and compatible with their ranking M1 for a correct statement relating their r_s ($ r_s < 1$) with their cv but cv must be such that $ cv < 1$ A1 for a correct contextualised comment that is rejecting H ₀ Must mention "population" and "no. of employees" and "positive correlation". Follow through their r_s and their cv (provided it is $ cv < 1$ Use of "association" is A0											
(c)	1 st B1 for 0.6319 2 nd B1 does not require context just no <u>positive</u> correlation mentioned											
(d)	1 st B1 for a comment relating to pmcc (i) no <u>linear</u> relationship <u>or</u> (ii) pmcc requires (joint) normal distribution 2 nd B1 for a second comment relating to Spearman's (i) there is a (non-linear) relationship between <u>rank</u> s.. <u>or</u> (ii) data not (joint) normal so Spearman's is better											

Question Number	Scheme	Marks																									
4 (a)	$\frac{282 \times 100}{600}$ (Do not accept $282 - 114.2 - 90.2 - 30.6$ (o.e.))	B1 (1)																									
4 (b)	9	B1 (1)																									
4 (c)	2.5 or better (Do not accept 0.025)	B1 (1)																									
4 (d)	<p>H_0: hair colour occurs in the ratio 2:6:1:3 H_1: hair colour does not occur in the ratio 2:6:1:3</p> <table border="1" data-bbox="220 589 1326 902"> <thead> <tr> <th></th> <th>black</th> <th>brown</th> <th>red</th> <th>blonde</th> </tr> </thead> <tbody> <tr> <td>observed</td> <td>105</td> <td>282</td> <td>48</td> <td>165</td> </tr> <tr> <td>expected</td> <td>100</td> <td>300</td> <td>50</td> <td>150</td> </tr> <tr> <td>$\frac{(O_i - E_i)^2}{E_i}$</td> <td>0.25</td> <td>1.08</td> <td>0.08</td> <td>1.5</td> </tr> <tr> <td>$\frac{O_i^2}{E_i}$</td> <td>110.25</td> <td>265.08</td> <td>46.08</td> <td>181.5</td> </tr> </tbody> </table> <p> $\sum \frac{(O_i - E_i)^2}{E_i} = 2.91 \quad \text{or} \quad \sum \frac{O_i^2}{E_i} - 600 = 602.91 - 600 = 2.91 \quad (\text{awrt } \mathbf{2.91})$ $\nu = 3$ cv is 7.815 $[2.91 < 7.815]$ so insufficient evidence to reject H_0 <u>or</u> not significant There is evidence to suggest that hair colour does occur in the given ratio. </p>		black	brown	red	blonde	observed	105	282	48	165	expected	100	300	50	150	$\frac{(O_i - E_i)^2}{E_i}$	0.25	1.08	0.08	1.5	$\frac{O_i^2}{E_i}$	110.25	265.08	46.08	181.5	B1 B1 expected M1 A1 A1 B1 B1 dM1 A1 (9) [Total 12]
	black	brown	red	blonde																							
observed	105	282	48	165																							
expected	100	300	50	150																							
$\frac{(O_i - E_i)^2}{E_i}$	0.25	1.08	0.08	1.5																							
$\frac{O_i^2}{E_i}$	110.25	265.08	46.08	181.5																							
Notes																											
4 (d)	<p>1st B1 for both hypotheses. Must mention hair colour and ratio e.g. "hair colour in the given ratio" Allow use of ditto 2nd B1 for all 4 correct expected frequencies 1st M1 for at least 2 correct calculations from 3rd or 4th row 1st A1 for all correct calculations to at least 3sf if row 4 If awrt 2.91 is seen with no incorrect working award B1M1A1A1 2nd dM1 Dep on 1st M1 for a correct statement linking their test statistic and their cv (cv > 3.5) 3rd A1 for a correct comment in context - must mention "hair colour" and "ratios" <u>or</u> "model" e.g. "There is evidence of to support the given model" No follow through If hypotheses are the wrong way round score A0.</p>																										

Question Number	Scheme	Marks
<p>5 (a)</p> <p>$\bar{x} = \frac{1}{2}(118.8 + 121.2) = 120$</p> <p>“their 1.6449” $\frac{\sigma}{\sqrt{n}} = 121.2 - 120$</p> <p>“their 2.3263” $\frac{\sigma}{\sqrt{n}} = 2.3263 \times \left(\frac{121.2 - 120}{1.6449} \right)$</p> <p>So 98% C.I. = $120 \pm 1.424\dots = (118.3028\dots, 121.699\dots)$</p> <p>(b) awrt $(118\pi, 122\pi)$ or $(371/372, 382/383)$</p> <p>(c) $P(\text{All}) = (0.98)^3 = 0.941$</p>	<p>1.6449 (or better)</p> <p>2.3263 (or better)</p> <p>awrt (118, 122)</p>	<p>B1</p> <p>B1 M1</p> <p>B1 dM1</p> <p>A1</p> <p>B1ft</p> <p>M1 A1</p> <p>(6)</p> <p>(1)</p> <p>(2)</p> <p>[Total 9]</p>
Notes		
<p>(a)</p> <p>(c)</p>	<p>NB in part (a) only lose one of the B1 marks for not using the percentage points table</p> <p>1st B1 for $\bar{x} = 120$</p> <p>2nd B1 for 1.6449 or better in an attempt (could be $1.6449\sigma = k$ or even $1.6449\sigma^2 = k$) Condone strange notation for standard error (E) here if it is <u>used</u> correctly</p> <p>1st M1 for an attempt to find “width” or “half-width” of a 90% CI ft their z value ($z > 1.5$) e.g. for $z E = 121.2 - 120$ (o.e.) N.B. $E = 0.7295\dots$ Condone missing 2 here.</p> <p>3rd B1 for 2.3263 or better in an attempt at CI. If score 2nd B0 for using 1.64 or 1.645 allow 3rd B1 for 2.32 or 2.33 here</p> <p>2nd dM1 for a correct attempt at “width” or “half-width” of a 98% CI ft their z value ($z > 2$) Dependent on 1st M1 and ft their value or expression for s.e.</p> <p>A1 for lower limit in range $[118, 118.35)$ <u>and</u> upper limit in range $(121.65, 122]$</p> <p>Answer only of awrt (118, 122) with no incorrect working seen scores 6/6/ if 1.6449 and 2.3263 are seen and 5/6 (B1B1M1B0M1A1) otherwise.</p> <p>M1 for a correct expression i.e. $(0.98)^3$</p> <p>A1 for awrt 0.941</p>	

Question Number	Scheme	Marks
6 (a)	$\text{Var}(X) = \frac{(a+5-a+1)^2}{12} [=3]$ $\bar{X} \sim N\left(a+2, \frac{3}{50}\right)$	M1 A1, A1ft (3)
(b)	$17.2 - 1.96 \times \sqrt{\frac{3}{50}} < \mu < 17.2 + 1.96 \times \sqrt{\frac{3}{50}}$ $17.2 - 1.96 \times \sqrt{\frac{3}{50}} < a + 2 < 17.2 + 1.96 \times \sqrt{\frac{3}{50}}$ $14.7 < a < 15.7$	B1 M1 B1 A1 (4) [Total 7]
Notes		
(a)	M1 for a correct expression for $\text{Var}(X)$ in terms of a <u>or</u> $\text{Var}(X) = 3$ 1 st A1 for normal and correct mean must be $a + 2$ NB $N(17.2, \dots)$ is A0 and $N(17.2, \frac{3}{50})$ is M1A0A1 2 nd A1ft for correct $\text{Var}(\bar{X})$, i.e. (their "3")/50	
(b)	1 st B1 for correct use of $z = 1.96$ in an attempt e.g. $\bar{x} \pm z\sigma$ or $\bar{x} \pm z\sigma^2$ M1 for $17.2 \pm z \times \sqrt{\frac{3}{50}}$ where $ z > 1.5$ accept just + or just - Answer of (16.7, 17.7) scores B1M1B0A0 2 nd B1 for either of the inequalities with $a + 2$ and any z ($ z > 1.5$) or $a = 15.2 \pm z \times \sqrt{\frac{3}{50}}$ A1 for awrt 14.7 and 15.7	

Question Number	Scheme	Marks
<p>7 (a)</p>	<p>$H_0: \mu_a = \mu_b, \quad H_1: \mu_a < \mu_b$</p> $\text{s.e.} = \sqrt{\frac{25^2}{100} + \frac{10^2}{150}}, \quad z = \frac{67 - 60}{\sqrt{\frac{25^2}{100} + \frac{10^2}{150}}} \quad \text{CR} = 1.6449 \times \sqrt{\frac{25^2}{100} + \frac{10^2}{150}}$ <p>$z = \pm 2.6616\dots \quad \quad \quad = \pm 4.326\dots \quad (\text{awrt } \mathbf{2.66/4.33})$</p> <p>One tailed critical value $z = 1.6449$ (or prob of awrt 0.004 (<0.05)) [Condone 0.996 if compared correctly with 0.95 for the B1] 2.6616 > 1.6449 so] significant evidence to reject H_0 There is evidence that the amount of lead present in the soil has decreased.</p>	<p>B1</p> <p>M1,dM1</p> <p>A1</p> <p>B1</p> <p>dM1</p> <p>A1ft</p> <p>(7)</p>
	Notes	
<p>(a)</p>	<p>1st B1 for both hypotheses in terms of μ not words. Accept μ_1, μ_2 etc if there is some indication of which is which e.g $X \sim N(67, 25^2)$ implies X is “before”.</p> <p>1st M1 for attempt at s.e. - condone one number wrong or mis-matched variances i.e. $\sqrt{\frac{p}{q} + \frac{r}{s}}$ (3 of p, q, r & s correct) or $\sqrt{\frac{10^2}{100} + \frac{25^2}{150}}$</p> <p>2nd dM1 Dep on 1st M1 for using their s.e. in correct formula for test statistic. Num of $\pm (67 - 60)$ <u>or</u> for correct expression for CR</p> <p>3rd dM1 dep. on 2nd M1 for a correct statement based on their normal cv ($cv > 1.5$) and their test statistic</p> <p>2nd A1ft for correct comment in context. Must mention “lead” or “soil” and “factory”. Allow ft If hypotheses are the wrong way round score A0 If hypotheses are not for a difference between 2 means award A0</p>	<p>B1</p> <p>(1)</p>
<p>(b)</p>	<p>B1 must mention <u>mean</u> and <u>normal</u>. In words or symbols e.g. $\bar{X} \sim N(\dots$</p>	<p>B1</p> <p>(1)</p>
		[Total 9]

Question Number	Scheme	Marks
<p>8 (a)</p> <p>Let $W = D_1 - D_2$ $W \sim N(0, 2.88)$ $P(W > 3) = 2 \times P(W > 3)$ $= 2 \times P\left(Z > \frac{3-0}{\sqrt{2.88}}\right)$ $= 2 \times P(Z > 1.76776\dots)$ $= 2 \times (1 - 0.9616)$ $= 0.0768$</p> <p>(b)</p> <p>Let $T = 5C - 4D$ or $4D - 5C$ or $C - \frac{4}{5}D$ or $\frac{4}{5}D - C$ $T \sim N(\pm 4, 39.04)$ or $N(\pm 0.8, 1.5616)$ $P(T < 0) = P\left(Z < \frac{0-4}{\sqrt{39.04}}\right)$ or $P\left(Z < \frac{0-0.8}{\sqrt{1.5616}}\right)$ $= P(Z < -0.64018\dots)$ $= (1 - 0.7389)$ $= 0.2611$</p> <p>(c)</p> <p>Let $P = D + D + D + D + D + D + B$ Let $Q = C + C + C + C + C + C + B$ $P \sim N(352, 13.64)$ and $Q \sim N(292, 8.84)$ [Let $R = P - Q$ $R \sim N(\pm 60, 22.48)$</p> <p>$P(R > 50) = P\left(Z > \frac{50-60}{\sqrt{22.48}}\right)$ $= P(Z > -2.10\dots)$ $= 0.9821$</p>	<p>awrt 0.077</p> <p>awrt 0.261</p> <p>awrt 0.982 ~ 0.983</p>	<p>M1 A1, A1 M1 dM1 A1</p> <p>M1 A1 A1 M1 A1</p> <p>M1 A1, A1 M1 dM1 A1</p> <p>(6)</p> <p>(5)</p> <p>(6)</p> <p>[Total 17]</p>
Notes		
<p style="text-align: center;">Award full marks in each part for a correct answer with no incorrect working seen.</p> <p>(a) 1st M1 for explicitly defining a suitable W and attempt to find the distribution of W. May be implied by sight of $N(0, 2.88)$ 1st A1 for normal and mean of 0, 2nd A1 for variance of 2.88. Award M1A1A1 for $N(0, 2.88)$ seen. 2nd M1 for realising need $2 \times P(W > 3)$ 3rd dM1 Dep on 1st M1 for standardising with 3, 0 and their s.d. Must lead to $P(Z > +ve)$ (o.e.)</p> <p>(b) 1st M1 for explicitly defining a suitable T but may be implied by sight of one of these normals 1st A1 for normal and correct mean, 2nd A1 for correct variance. Accept awrt 3sf i.e. 39.0, 1.56 2nd M1 for standardising with 0 and their mean and their s.d. Must lead to $P(Z < -ve)$ (o.e.)</p> <p>(c) 1st M1 for explicitly defining a correct P or Q. May be implied by a correct distribution for P or Q 1st A1 for a correct distribution for P 2nd A1 for a correct distribution for Q 2nd M1 for attempting R and obtaining its distribution- ft their P and Q means and variances 3rd dM1 for attempting $P(R > 50)$ <u>and</u> standardising with 50 and their $E(R)$ and their $\sqrt{\text{Var}(R)}$ Dependent on 2nd M1. Must lead to a $P(Z > -ve)$ (o.e.)</p>		

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467

Fax 01623 450481

Email publication.orders@edexcel.com

Order Code UA037014 Summer 2013

For more information on Edexcel qualifications, please visit our website
www.edexcel.com

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual



Llywodraeth Cynulliad Cymru
Welsh Assembly Government





Mark Scheme (Results)

Summer 2013

GCE Statistics 3 (6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2013

Publications Code UA037011

All the material in this publication is copyright

© Pearson Education Ltd 2013

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.
3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes:

- bod – benefit of doubt
 - ft – follow through
 - the symbol \surd will be used for correct ft
 - cao – correct answer only
 - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper
 - \square The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.
 5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
 6. If a candidate makes more than one attempt at any question:
 - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
 - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.
 7. Ignore wrong working or incorrect statements following a correct answer.
 8. In some instances, the mark distributions (e.g. M1, B1 and A1) printed on the candidate's response may differ from the final mark scheme.

Question Number	Scheme				Marks																																				
<p>1.</p> <p>H_0 : Cholesterol level is independent of intake of saturated fats(no association) H_1 : Cholesterol level is not independent of intake of saturated fats (association)</p> <table border="1" data-bbox="280 506 1134 801"> <thead> <tr> <th>O</th> <th>E</th> <th>$\frac{(O-E)^2}{E}$</th> <th>$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>12</td> <td>7.6</td> <td>2.547... or $\frac{242}{95}$</td> <td>18.947... or $\frac{360}{19}$</td> </tr> <tr> <td>8</td> <td>12.4</td> <td>1.56129... or $\frac{242}{155}$</td> <td>5.161... or $\frac{160}{31}$</td> </tr> <tr> <td>26</td> <td>30.4</td> <td>0.6368... or $\frac{121}{190}$</td> <td>22.236... or $\frac{845}{38}$</td> </tr> <tr> <td>54</td> <td>49.6</td> <td>0.3903... or $\frac{121}{310}$</td> <td>58.790... or $\frac{3645}{62}$</td> </tr> </tbody> </table> <p>$\sum \frac{(O-E)^2}{E} = 5.1358234..$ or $\frac{1.2^2}{7.6} + \frac{8^2}{12.4} + \frac{26^2}{30.4} + \frac{54^2}{49.6} - 100 = 5.14$ (awrt 5.14)</p> <p>$\nu = (2-1)(2-1) = 1$ $\chi_1^2(0.05) = 3.841$ $5.14 > 3.841$ so sufficient evidence to reject H_0 [Condone “accept H_1”] Association between cholesterol level and saturated fat intake</p>	O	E	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$	12	7.6	2.547... or $\frac{242}{95}$	18.947... or $\frac{360}{19}$	8	12.4	1.56129... or $\frac{242}{155}$	5.161... or $\frac{160}{31}$	26	30.4	0.6368... or $\frac{121}{190}$	22.236... or $\frac{845}{38}$	54	49.6	0.3903... or $\frac{121}{310}$	58.790... or $\frac{3645}{62}$	<table border="1"> <thead> <tr> <th>Cholesterol Level</th> <th>High</th> <th>Low</th> <th></th> </tr> </thead> <tbody> <tr> <td>High</td> <td>7.6</td> <td>12.4</td> <td>20</td> </tr> <tr> <td>Low</td> <td>30.4</td> <td>49.6</td> <td>80</td> </tr> <tr> <td></td> <td>38</td> <td>62</td> <td>100</td> </tr> </tbody> </table>	Cholesterol Level	High	Low		High	7.6	12.4	20	Low	30.4	49.6	80		38	62	100				<p>M1A1</p> <p>(2)</p> <p>B1</p> <p>(1)</p> <p>dM1</p> <p>A1</p> <p>A1</p> <p>(3)</p> <p>B1</p> <p>B1</p> <p>(2)</p> <p>M1</p> <p>A1</p> <p>(2)</p> <p>Total 10</p>
	O	E	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$																																					
	12	7.6	2.547... or $\frac{242}{95}$	18.947... or $\frac{360}{19}$																																					
	8	12.4	1.56129... or $\frac{242}{155}$	5.161... or $\frac{160}{31}$																																					
	26	30.4	0.6368... or $\frac{121}{190}$	22.236... or $\frac{845}{38}$																																					
	54	49.6	0.3903... or $\frac{121}{310}$	58.790... or $\frac{3645}{62}$																																					
	Cholesterol Level	High	Low																																						
	High	7.6	12.4	20																																					
	Low	30.4	49.6	80																																					
		38	62	100																																					
Notes																																									
<p>Minimum working use part marks: E_i (2), Hyp (1), 5.14 (3), 3.841 (2), Conclusion (2)</p> <p>1st M1 for some use of $\frac{\text{Row Total} \times \text{Col.Total}}{\text{Grand Total}}$. May be implied by correct E_i</p> <p>1st A1 for all expected frequencies correct. Allow M1A0 for E_i rounded to integers</p> <p>1st B1 for both hypotheses. Must mention “cholesterol” and “fats” at least once Use of “relationship” or “correlation” or “connection” is B0</p> <p>2nd dM1 for at least 2 correct terms (as in 3rd or 4th column) or correct expressions with their E_i Dependent on 1st M1 Accept 2sf accuracy for the M mark</p> <p>2nd A1 for all correct terms. May be implied by a correct ans.(2 dp or better) Allow truncation eg 2.54... 3rd A1 for awrt 5.14</p> <p>2nd B1 for correct degrees of freedom (may be implied by a cv of 3.841)</p> <p>3rd M1 for a correct statement linking their test statistic and their cv(cv could be 2.705 or > 3.5) Contradictory statements score M0 e.g. “significant, do not reject H_0”</p> <p>4th A1 for a correct comment in context - must mention “cholesterol” and “fats” condone “relationship” or “connection” here but not “correlation”. e.g. “There is evidence of a relationship between cholesterol level and fat intake” No follow through. If e.g hypotheses are the wrong way round A0 here.</p>																																									

Question Number	Scheme									Marks	
2(a)	Uni	A	B	C	D	E	F	G			M1A1A1
	Staff-Stu	2	4	3	5	7	1	6			
	Satisfaction	3	2	6	4	5	1	7			
	[d]	-1	2	-3	1	2	0	-1			
	d^2	1	4	9	1	4	0	1	20		
	$r_s = 1 - \frac{6 \times 20}{7(49-1)} = 0.642857\dots \quad \left(\text{accept } \frac{9}{14} \right) \quad \text{(awrt } \mathbf{0.643} \text{)}$									dM1A1 (5)	
(b)	<p>$H_0: \rho = 0$ $H_1: \rho \neq 0$ ($\rho > 0$) Critical value is ± 0.7857 (± 0.7143 for a one tailed test) $0.643 < cv$ so insufficient evidence to reject H_0 There is insufficient evidence to suggest a (positive) correlation between staff-student ratio and satisfaction.</p>									B1 B1 B1ft (3) Total 8	
Notes											
(a)	<p>1st M1 for an attempt to rank the staff-students ratio <u>or</u> satisfaction (at least 4 correct) 1st A1 for correct rankings for both (one or both may be reversed) 2nd A1 for $\sum d^2 = 20$ or correct d^2 row (NB $\sum d^2 = 92$ for one set of reversed ranks) 2nd dM1 for use of the correct formula, follow through their $\sum d^2$ (Dependent on 1st M1) If answer is not correct, a correct expression is required. 3rd A1 If $\sum d^2 = 20$ for awrt 0.643 <u>or</u> if $\sum d^2 = 92$ for awrt -0.643 (accept $\pm \frac{9}{14}$)</p>										
(b)	<p>1st B1 for both hypotheses in terms of ρ, one tail H_1 must be compatible with their ranking Hypotheses just in words e.g. “no correlation” score B0 2nd B1 for cv of 0.7857 <u>or</u> 0.7143 for one-tailed test (accept \pm) Their cv must be compatible with their H_1 which may be in words If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible. 3rd B1ft for a correct contextualised comment. Must mention “ratio” or “no. of students per member of staff” <u>and</u> “satisfaction” Follow through their r_s and their cv (provided it is $cv < 1$) Don’t insist on the word “positive” for a one-tailed test Use of “association” is B0 Independent of 1st B1 so if $r_s > cv$ must say there is sufficient evidence of(o.e.) and if $r_s < cv$ must say insufficient evidence of ... (o.e.) regardless of their hypotheses Contradictory statements score B0 (This mark is just testing interpretation of comparison of their r_s and their cv)</p>										

Question Number	Scheme	Marks
<p>3(a)i</p> <p>e.g.</p> <p>3(a)ii</p> <p>e.g.</p> <p>e.g.</p> <p>3(b)</p> <p>3(c)</p>	<p>Quota Sampling: Advantages: Fieldwork can be done quickly, <u>or</u> administering the test is easy, <u>or</u> costs are kept to a minimum (cheap), <u>or</u> gives estimates for each course. <u>or</u> OK for large populations <u>or</u> sampling frame not required (o.e.) Disadvantages: Non-random process <u>or</u> not possible to estimate the sampling errors, <u>or</u> non response not recorded, <u>or</u> interviewer can introduce bias in sample choice. (o.e.)</p> <p>Stratified Sampling: Advantages: Can give accurate estimates as it is a random process, <u>or</u> gives estimates for each course <u>or</u> representative of [BUT not “proportional” to] the whole population. (o.e.) Disadvantages: Sampling frame required, <u>or</u> strata may not be clear as some students overlap courses <u>or</u> not suitable for large populations. (o.e.)</p> <p>Total enrolments=1000 (may be implied by calculations)</p> <p>Leisure and Sport=$\frac{420}{1000} \times 100 = 42$</p> <p>Information Technology=$\frac{337}{1000} \times 100 = 33.7 = 34$</p> <p>Health and Social Care=$\frac{200}{1000} \times 100 = 20$</p> <p>Media Studies=$\frac{43}{1000} \times 100 = 4.3 = 4$</p> <p>The college’s information system would be used to identify each student and which course they are enrolled on. i.e. idea of sampling frame or list for each course. Use of random numbers to select required number of students from each course</p>	<p>B1</p> <p>B1</p> <p>B1</p> <p>M1</p> <p>A1</p> <p>B1</p> <p>B1</p> <p>(2)</p> <p>(3)</p> <p>(2)</p> <p>Total 7</p>
Notes		
<p>(a)</p> <p>(b)</p> <p>(c)</p>	<p style="text-align: center;">Do not penalise for lack of context in part (a)</p> <p>1st B1 for an advantage and a disadvantage for quota sampling (must be 1st or labelled (i)) 2nd B1 for an advantage and a disadvantage for stratified sampling (2nd or labelled (ii)) Do not allow opposite pairs e.g. “quicker/easier” for quota sampling and “takes a long time/more difficult” for stratified <u>or</u> quota “easy to use” but strat. “hard for large populations” Do not allow same reason for both e.g. “gives estimates for each course”</p> <p>M1 for one correct calculation, ft their “1000” A1 for 42, 34, 20 and 4 only</p> <p>1st B1 for some mention of a suitable <u>sampling frame</u>. Need not give the specific term but a suitable source of <u>list</u> is required for all students <u>in each course</u>. 2nd B1 for mentioning use of <u>random numbers</u> or some random selection process <u>for each course</u>. If they are describing systematic sampling score B0 here</p>	

Question Number	Scheme	Marks																														
4	(a) $\bar{x} = \frac{8 \times 1.5 + 12 \times 4 + 13 \times 5.5 + 9 \times 7 + 8 \times 10}{50} = \frac{274.5}{50} = 5.49$ (*)	B1cso																														
	$s^2 = \frac{8 \times 1.5^2 + 12 \times 4^2 + 13 \times 5.5^2 + 9 \times 7^2 + 8 \times 10^2}{49} - \frac{50}{49} 5.49^2 = 6.88$ (*)	M1, A1cso																														
	(b) $a = 50 \times P(6 < X < 8) = 50 \times P(0.194.. < Z < 0.956..)$ $a = 12.81$ (tables) <u>or</u> 12.68 (calc) $b = 50 - (28.85 + a)$ $= 8.34$ (tables) <u>or</u> 8.47 (calc)	M1 A1 A1ft																														
	(c) H_0 : Normal distribution is a good fit H_1 : Normal distribution is not a good fit	B1																														
	<table border="1"> <thead> <tr> <th>Class</th> <th>O</th> <th>E</th> <th>$\frac{O^2}{E}$</th> <th>$\frac{(O-E)^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>0-3</td> <td>8</td> <td>8.56</td> <td>7.4766...</td> <td>0.0366...</td> </tr> <tr> <td>3-5</td> <td>12</td> <td>12.73</td> <td>11.31186....</td> <td>0.0418...</td> </tr> <tr> <td>5-6</td> <td>13</td> <td>7.56</td> <td>22.354497...</td> <td>3.9144...</td> </tr> <tr> <td>6-8</td> <td>9</td> <td>12.68 or (12.81)</td> <td>(6.32) ~ 6.38801..</td> <td>1.0680...~ (1.13)</td> </tr> <tr> <td>8-12</td> <td>8</td> <td>(8.34) or 8.47</td> <td>7.556080...~ (7.67)</td> <td>(0.013) ~ 0.0260..</td> </tr> </tbody> </table>	Class	O	E	$\frac{O^2}{E}$	$\frac{(O-E)^2}{E}$	0-3	8	8.56	7.4766...	0.0366...	3-5	12	12.73	11.31186....	0.0418...	5-6	13	7.56	22.354497...	3.9144...	6-8	9	12.68 or (12.81)	(6.32) ~ 6.38801..	1.0680...~ (1.13)	8-12	8	(8.34) or 8.47	7.556080...~ (7.67)	(0.013) ~ 0.0260..	M1
	Class	O	E	$\frac{O^2}{E}$	$\frac{(O-E)^2}{E}$																											
	0-3	8	8.56	7.4766...	0.0366...																											
	3-5	12	12.73	11.31186....	0.0418...																											
	5-6	13	7.56	22.354497...	3.9144...																											
	6-8	9	12.68 or (12.81)	(6.32) ~ 6.38801..	1.0680...~ (1.13)																											
8-12	8	(8.34) or 8.47	7.556080...~ (7.67)	(0.013) ~ 0.0260..																												
$\sum \frac{O^2}{E} - N = 5.087 \dots \sim 5.1400 \dots$ awrt (5.09 ~ 5.14)	A1																															
$\nu = 5 - 3 = 2$ (for 5 - 3 or 2 can be implied by 5.991 seen)	B1																															
$\chi^2(0.05) = 5.991$	B1																															
5.09 < 5.991 so insufficient evidence to reject H_0 Normal distribution is a good fit.	M1 A1																															
	Total 14																															
Notes																																
(a)	B1cso for denominator of 50 and at least 3 products on num <u>or</u> 274.5 on num M1 for a correct expression with at least 3 correct products on num <u>or</u> $\frac{1844.25}{49} - \frac{1507.005}{49}$ <u>or</u> $\frac{337.245}{49}$ <u>or</u> $\left(\frac{7377}{200} - 5.49^2\right) \times \frac{50}{49}$ etc Allow 3sf accuracy A1cso for 6.88 with M1 scored and no incorrect working seen																															
(b)	M1 a full method for a or b using the normal dist. Correct use of (6), 8, 5.49 and $\sqrt{6.88}$ seen 1 st A1 for a in range 12.68 ~ 12.81 or b in range 8.34~ 8.47 or awrt these values 2 nd A1ft for $50 - 28.85 -$ their a (or b) (but requires M1). Allow awrt 3sf. Must add up to 50																															
(c)	1 st B1 for both hypotheses. B0 if they include 5.49 or 6.88. Condone $X \sim N(\mu, \sigma^2)$ etc 1 st M1 for attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$, at least 3 correct expressions or values. 1 st A1 for at least 4 correct calcs - 3 rd or 4 th column. (2 dp or better and allow e.g. 7.47) Allow any value in the ranges for the last two rows. 2 nd A1 for a test statistic that is awrt 5.09 ~ 5.14. Award M1A1A1 if this is obtained. 2 nd M1 for a correct statement based on their test statistic (> 1) and their cv (> 3.8) Contradictory statements score M0 e.g. "significant" do not reject H_0 . 3 rd A1 for a correct comment suggesting that normal model is suitable <u>or</u> manager's belief is correct. No f t . Condone mention of 5.49 or 6.88 here. Hypotheses wrong way round scores A0																															

Question Number	Scheme	Marks
<p>5 (a)</p> <p>(b)</p>	<p>Let $L \sim N(50, 25)$ and $S \sim N(15, 9)$ Let $X = L - (S_1 + S_2 + S_3)$ $E(X) = 50 - 3 \times 15 = 5$ $\text{Var}(X) = 25 + 3 \times 9 = 52$ $P(X < 0) = P\left(Z < \frac{-5}{\sqrt{52}}\right)$ $= P(Z < -0.693..)$ $= 0.244$ or 0.2451 (tables) (awrt 0.244 ~ 0.245)</p> <p>Let $Y = L - 3S$ $E(Y) = 50 - 3 \times 15 = 5$ $\text{Var}(Y) = 25 + 3^2 \times 9 = 106$ $P(Y > 0) = P\left(Z > \frac{-5}{\sqrt{106}}\right)$ $= P(Z > -0.4856..)$ $= 0.686$ or 0.6879 (tables) (awrt 0.686 ~ 0.688)</p>	<p>B1 B1 M1A1 dM1 A1 (6)</p> <p>B1 B1 M1A1 dM1 A1 (6)</p> <p>Total 12</p>
Notes		
<p>(a)</p> <p>(b)</p>	<p>1st B1 for forming a suitable variable X <u>explicitly</u> seen. Do not give for $L - 3S$ but allow $L - (S + S + S)$ 2nd B1 for $E(X) = 5$ (or -5 if their X is defined the other way around) 1st M1 for an attempt at $\text{Var}(X) = \text{Var}(L) + 3\text{Var}(S)$. Do not condone 5 for “25” or 3 for “9” 1st A1 for 52 2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on 1st M1 so if X is not being used or wrong variance score M0 If their method is not crystal clear then they must be attempting $P(Z < \text{-ve value})$ or $P(Z > \text{+ve value})$ i.e. their probability <u>after</u> standardisation should lead to a prob. < 0.5 2nd A1 for awrt 0.244 ~ 0.245 Correct ans. only scores 5/6 (or 6/6 if 1st B1) but must be clearly labelled as (a) or the first answer.</p> <p>1st B1 for defining a new variable $[Y =]_{\pm} (L - 3S)$. May be implied by a correct variance. 2nd B1 for $E(Y) = 5$ (or -5 if their Y is defined as $Y = 3S - L$) 1st M1 for an attempt at $\text{Var}(Y) = \text{Var}(L) + 3^2 \text{Var}(S)$. Do not condone 5 for “25” or 3 for “9” 1st A1 for 106 only 2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on 1st M1 so if Y is not being used or wrong variance score M0 If their method is not crystal clear then they must be attempting $P(Z > \text{-ve value})$ or $P(Z < \text{+ve value})$ i.e. their probability <u>after</u> standardisation should lead to a prob. > 0.5 2nd A1 for an awrt 0.686 ~ 0.688 Correct answer only scores 6/6 but must be clearly labelled as (b) or the second answer.</p>	

Question Number	Scheme	Marks
<p>6 (a)</p> <p>$H_0 : \mu_{new} - \mu_{old} = 1$ $H_1 : \mu_{new} - \mu_{old} > 1$</p> $z = \frac{7 - 5.5 - 1}{\sqrt{\frac{0.5}{60} + \frac{0.75}{70}}} = 3.62254\dots$ <p>(awrt 3.62)</p> <p>Critical value $z = 1.6449$ (allow \pm) $[3.62 > 1.6449]$ so sufficient evidence to reject H_0 Evidence that the mean yield of new variety is more than 1 kg greater than the old variety.</p> <p>(b) Mean yield is normally distributed Sample size is large. Must state or imply that in this case sample size is large</p>		<p>B1 B1</p> <p>M1 A1A1 A1</p> <p>B1 dM1</p> <p>A1</p> <p>(9)</p> <p>B1 B1</p> <p>(2)</p> <p>Total 11</p>
Notes		
<p>(a)</p> <p>ALT</p> <p>(b)</p>	<p>1st & 2nd B1 for hypotheses. Accept μ_1, μ_2 or μ_A, μ_B etc if there is some indication of which is which e.g. $A \sim N(\mu_A, 0.5)$</p> <p>1st M1 for an attempt at se. Condone switching 0.5 and 0.75 $\sqrt{\frac{0.5 \text{ or } 0.75}{60} + \frac{0.75 \text{ or } 0.5}{70}}$</p> <p>1st A1 for a correct expression for denominator of test statistic or 0.138... or $\sqrt{0.0190\dots}$</p> <p>2nd A1 for a correct numerator of test statistic (must have the - 1)</p> <p>3rd A1 for awrt 3.62 [Allow - 3.62 from numerator of $5.5 - 7 - - 1$ and compatible H_1]</p> <p>3rd B1 for ± 1.6449 seen <u>or</u> probability of 0.0002 (tables) or 0.000145...(calc) [allow 0.0001]</p> <p>2nd dM1 dep. on 1st M1 for a correct statement based on their normal cv and their test statistic</p> <p>2nd A1 for correct comment in context. Must mention “yield” <u>and</u> “varieties” or “old” and “new” <u>and</u> “1” If second B mark is B0 award A0 here</p> <p>Pooled estimate: If they calculate $s_p = \sqrt{0.41845\dots} = 0.64688\dots$ allow 1st M1, 1st A1 for expression (or awrt 0.114) and 2nd A1 if numerator correct but A0 for test statistic (4.39)</p> <p>1st B1 for mention of <u>mean</u> (yield) and <u>normal</u> (distribution) 2nd B1 for mention of <u>sample</u> (size) being <u>large</u> in <u>this case</u></p>	

Question Number	Scheme	Marks
7 (a)	$\hat{\mu} = \bar{x} = \frac{33.29}{8} = 4.16125 \quad (\text{awrt } \mathbf{4.16})$ $\hat{\sigma}^2 = s^2 = \frac{4.12^2 + 5.12^2 + \dots - 8 \times \bar{x}^2}{7}$ $\hat{\sigma}^2 = s^2 = \frac{141.4035 - 138.528013}{7} = 0.41078\dots \quad (\text{awrt } \mathbf{0.411})$	B1 M1 A1 (3)
(b)	$\sum x = 33.29 + 32 \times 4.55 = 178.89, \quad (\text{awrt } \mathbf{179})$ $\sum x^2 = "141.4035" + 31 \times 0.25 + 32 \times 4.55^2 (= 811.6335) \quad (\text{awrt } \mathbf{812})$	B1 M1A1
	$\text{Combined sample: } s^2 = \frac{811.6335 - \frac{178.89^2}{40}}{39} = 0.29724865\dots \quad (\text{awrt } \mathbf{0.297})$	M1A1
	$\frac{s}{\sqrt{n}} = \frac{\sqrt{0.297\dots}}{\sqrt{40}} = 0.0862 \quad (\text{awrt } \mathbf{0.0862})$	M1A1 (7)
(c)	$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = \frac{178.89}{40} \pm 1.96 \frac{0.67}{\sqrt{40}}$ $= (4.2646\dots, 4.67988\dots) \quad \text{awrt } (\mathbf{4.26}[\text{or } 4.265], \mathbf{4.68})$	M1B1 A1 (3)
	Notes	Total 13
(a)	<p>M1 for an attempt at s^2: correct denom, clear attempt at $\sum x^2$ and ft their \bar{x} Ans only 2/2</p>	
(b)	<p>B1 for correct sum or mean or fully correct expression (accept mean = awrt 4.47) May be in (c) 1st M1 for their $141.4035 + 31 \times 0.25 + 32 \times 4.55^2$ or "141.4035" + 7.75 + 662.48 (accept 3sf) Beware: $32(0.25 + 4.55^2) + "141.4035"$ = awrt 812 but scores M0A0. 1st A1 for a fully correct expression (all to 3sf or better) or answer only = awrt 812 2nd M1 for a correct expression using their values 3rd M1 dependent on using a changed s^2 (not their 0.411 or 0.25) for $\frac{\sqrt{"0.297"}}{\sqrt{40}}$ This s^2 must be based on a <u>combination</u> of their 0.411 and 0.25 e.g. 0.661</p>	
(c)	<p>M1 for $\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$ for any $z (> 1.5)$ and ft their \bar{x} based on combining their 4.16 and 4.55, do not award for simply using 4.55 or their 4.16. Condone $\sigma = \sqrt{\text{their } 0.297}$ or their (b) B1 for $z = 1.96$ used in an attempt at a CI, may for example miss \sqrt{n} A1 for both limits awrt 3sf. Allow lower limit of 4.265</p>	

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467

Fax 01623 450481

Email publication.orders@edexcel.com

Order Code UA037011 Summer 2013

For more information on Edexcel qualifications, please visit our website
www.edexcel.com

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Ofqual




Llywodraeth Cynulliad Cymru
Welsh Assembly Government





Mark Scheme (Results)

Summer 2014

Pearson Edexcel International A Level in
Statistics 3
(WST03/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2014

Publications Code IA040147

All the material in this publication is copyright

© Pearson Education Ltd 2014

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL IAL MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.
3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
 - ft – follow through
 - the symbol \surd will be used for correct ft
 - cao – correct answer only
 - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper
 - \square The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.
 5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
 6. Ignore wrong working or incorrect statements following a correct answer.

Question Number	Scheme	Marks
1. (a)	165, 8	B1 [1]
(b)	Select <u>every 6th person</u> {having chosen the first person by} Selecting a random number between 1 and 6 or selecting a random number and then loop back to start when you reach the end.	B1 dB1
(b)(ii)	The <u>list</u> is alphabetical and <u>has not been sorted by gender</u> .	B1 [3]
(c)	Label male members 1- 180, female members 1 – 120 <u>Use random numbers</u> to select a ... Simple random sample of <u>30 male</u> members and <u>20 female</u> members	M1 M1 A1 [3]
(d)	Any one of <ul style="list-style-type: none"> • It (a stratified sample) is <u>not biased</u> as the members are chosen randomly. • You <u>can estimate</u> the <u>sampling errors</u> (for a stratified sample) • It (a stratified sample) gives <u>more accurate estimates</u> as it is a random process. • A quota sample may <u>be biased</u> (whereas a stratified sample is not). • It's <u>not possible</u> to <u>estimate/find</u> the <u>sampling errors</u> for a <u>quota sample</u> (whereas you can for a stratified sample) 	B1 [1]
Notes		
(a)	B1 165 followed by 8 or 008.	
(b)(i)	1st B1 For selecting every 6 th (name on the list) 2nd dB1 <i>is dependent on the first B1 mark being awarded.</i> For idea of using random numbers to select first from 1 to 6 or 0 to 5 (o.e.) or selecting a random number between 1 and 300 and then looping back when the end of the list has been reached.	
(b)(ii)	B1 A comment that implies <u>the list</u> (or sampling frame) has not been sorted by gender. Note B0 for “the ordered list is not truly random” Note B0 for “sample does not divide the members into gender.”	
(c)	1st M1 For suitable labelling of all 180 males <u>and</u> all 120 females. E.g. Allow labelling female members 181 – 300. Also allow labelling male members 0 – 179 and female members either 0 to 119 or 180 to 299. 2nd M1 For use of random numbers to select males and females. A1 For 30 males <u>and</u> 20 females (dependent on 2nd M1 only) Note A simple random sample of 30 males and 20 females scores 2 nd M1 and A1.	
(d)	Note B0 for “a stratified sample can reflect the population structure.” B0 for “estimates obtained from each of the strata.”	

Question Number	Scheme	Marks
<p>2.</p> <p>(a)</p> <p>(b)</p> <p>(c)</p>	<p>X follows a continuous uniform distribution over $[\alpha - 3, 2\alpha + 3]$</p> <p>$\{E(\bar{X}) = \mu = \frac{2\alpha + 3 + \alpha - 3}{2}$ $= \frac{3\alpha}{2}$. So \bar{X} is a biased estimator.</p> <p>bias $\left\{ = \frac{3\alpha}{2} - \alpha \right\} = \pm \frac{\alpha}{2}$</p> <p>$k = \frac{2}{3}$</p> <p>$\alpha = \frac{2}{3}\bar{X} = \frac{2}{3}(8)$</p> <p>Max value $= 2\left(\frac{16}{3}\right) + 3$ $= \frac{41}{3}$</p>	<p>M1</p> <p>A1</p> <p>B1</p> <p>[3]</p> <p>B1</p> <p>[1]</p> <p>M1</p> <p>M1</p> <p>A1</p> <p>[3]</p> <p>7</p>
Notes		
<p>(a)</p> <p>(c)</p>	<p>M1 Using the formula $\left(\frac{a+b}{2}\right)$ or getting $\frac{3\alpha}{2}$</p> <p>A1 $\frac{3\alpha}{2}$ and concluding. Allow A1 for $\frac{3\alpha}{2} \neq \alpha$.</p> <p>Note Also allow A1 for bias $= \pm \frac{\alpha}{2} \neq 0$</p> <p>1st M1 An attempt to use the sample data given to find \bar{x} and multiply by their k. Allow full expression for \bar{x} or $\frac{\sum x}{n}$.</p> <p>Note 1st M1 can be implied by a correct recovery leading to $\alpha = \frac{16}{3}$</p> <p>2nd M1 $2 \times$ "their α" + 3 where their α is a function of the sample mean - which found by applying $\frac{\sum x}{n}$ from the data values given in the question.</p> <p>Note $2(13) + 3 = 39$ is M0M0A0</p>	

Question Number	Scheme	Marks						
3. (a)	$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A > \mu_B$ $\text{s.e.} = \sqrt{\frac{35^2}{80} + \frac{28^2}{100}} \quad \{ = 4.81170448... \}$ $z = \frac{532 - 520}{4.8117...} ; = 2.4939...$ <p>One tailed c.v. $Z = 2.3263$ or CR: $Z \geq 2.3263$ or p-value = awrt $0.006 < 0.01$ or "0.994" > 0.99</p> <p>[in the CR/significant/Reject H_0 / "0.006" < 0.01 / "0.994" > 0.99]</p> <p>Conclude either</p> <ul style="list-style-type: none"> that the <u>mean weight</u> of <u>grapefruit</u> from <u>farm A</u> is <u>greater</u> than that of <u>farm B</u>. that the <u>average weight</u> of <u>grapefruit</u> from <u>farm A</u> is <u>greater</u> than that of <u>farm B</u>. that the <u>grocer's</u> belief is <u>correct</u>. 	<p>B1</p> <p>M1 A1</p> <p>dM1; awrt 2.49 A1</p> <p>Critical value of 2.3263 Or a correct probability comparison. B1</p> <p>A correct conclusion in context which is based on <i>their</i> z-value and <i>their</i> critical value, where $c.v. > 1$. A1</p> <p>[7] 7</p>						
Notes								
	<p>B1 If μ_1, μ_2 used then it must be clear which refers to farm A and to farm B.</p> <p>1st M1 Condone minor slips e.g. $\sqrt{\frac{35^2}{100} + \frac{28^2}{80}}$ or $\sqrt{\frac{35}{80} + \frac{28^2}{100}}$ etc. i.e. swapped n or one s.d. and one variance.</p> <p>1st A1 $\text{s.e.} = \sqrt{\frac{35^2}{80} + \frac{28^2}{100}}$. Or can be implied by s.e. = awrt 4.81</p> <p>2nd dM1 <i>is dependent upon the 1st M1.</i> You can follow through their s.e. if 1st M1 mark has been awarded.</p> <p>Note M1A1dM1 is scored for writing $z = \pm \frac{(532 - 520)}{\sqrt{\frac{35^2}{80} + \frac{28^2}{100}}}$</p> <p>Special Case SC: M1A0M0A0 for $\text{s.e.} = \sqrt{\frac{35}{80} + \frac{28}{100}} \{ = 0.847... \}$</p> <p>Final A1 Dependent on the first two method marks being scored. For a contextualised comment which is rejecting H_0. Contradictory statements score final A0. E.g. "significant, do not reject H_0".</p> <p>Alternative method for 2nd "M1A1B1" marks: Let $D = \bar{x}_A - \bar{x}_B$</p> <table border="1" data-bbox="324 1638 1429 1806"> <tr> <td>$2.3263 = \frac{D - 0}{4.8117...}$</td> <td>dM1: dependent upon the 1st M1 for $\frac{D}{\text{their "4.8117..."}} = 2.3263 / 2.32 / 2.33$</td> </tr> <tr> <td></td> <td>A1: $D = \text{awrt } 11.2$</td> </tr> <tr> <td>So, $D = 11.193$</td> <td>B1: 2.3263</td> </tr> </table>	$2.3263 = \frac{D - 0}{4.8117...}$	dM1: dependent upon the 1 st M1 for $\frac{D}{\text{their "4.8117..."}} = 2.3263 / 2.32 / 2.33$		A1: $D = \text{awrt } 11.2$	So, $D = 11.193$	B1: 2.3263	
$2.3263 = \frac{D - 0}{4.8117...}$	dM1: dependent upon the 1 st M1 for $\frac{D}{\text{their "4.8117..."}} = 2.3263 / 2.32 / 2.33$							
	A1: $D = \text{awrt } 11.2$							
So, $D = 11.193$	B1: 2.3263							

Question Number	Scheme										Marks			
4. (a)	Man	A	B	C	D	E	F	G	H	I	J	Attempt to rank both for x and for w . (at least four correct).	M1	
	Rank x	1	2	3	4	5	6	7	8	9	10			
	Rank w	2	7	4	3	1	9	6	5	8	10			
	or	Man	A	B	C	D	E	F	G	H	I	J	For finding the difference between each of the ranks and evaluating $\sum d^2$. $\sum d^2 = 64$	M1
	Rank x	10	9	8	7	6	5	4	3	2	1			
	Rank w	9	4	7	8	10	2	5	6	3	1			
		$\sum d^2 = 1 + 25 + 1 + 1 + 16 + 9 + 1 + 9 + 1 + 0; = 64$											A1	
		$r_s = 1 - \frac{6(64)}{10(99)}; = 0.6121212\dots$										Using $1 - \frac{6 \sum d^2}{10(99)}$ with their $\sum d^2$	dM1;	
												$\frac{101}{165}$ or awrt 0.612	A1	
													[5]	
(b)	$H_0 : \rho = 0, H_1 : \rho > 0$					Both hypotheses stated correctly					B1			
	Critical Value $r_s = 0.5636$ or CR: $r_s \geq 0.5636$					Critical value of 0.5636					B1			
	Either													
	<ul style="list-style-type: none"> Since $r_s = 0.6121\dots$ lies <u>in</u> the <u>CR</u> Result is <u>significant</u> Reject H_0 (condone H_1) 										see notes	M1		
	conclude that there is a <u>positive correlation</u> between systolic <u>blood pressure</u> and <u>weight</u> .										Conclusion in context	A1		
												[4]		
(c)	Both either													
	<ul style="list-style-type: none"> Critical Value $r = 0.5494$ CR: $r \geq 0.5494$ 													
	and either													
	<ul style="list-style-type: none"> Since $r = 0.5114$ <u>does not lie in the CR</u> Result is <u>not significant</u> Do not reject H_0 (or accept H_0) 											M1		
	Conclude that there is <u>no positive correlation</u>										Context not required here.	A1		
												[2]		
(d)	Either													
	<ul style="list-style-type: none"> A comment that conveys both the ideas “as x increases, w increases” and “the relationship is non-linear” “There is a positive correlation” and “the relationship is non-linear” Data is not (bi-variate) normal 										Any one of these or equivalent.	B1		
												[1]		
												12		

Notes	
4. (a)	<p>3rd dM1 <i>is dependent on 1st M1</i> for use of $1 - \frac{6 \sum d^2}{10(99)}$ with their $\sum d^2$</p> <p>Note If a candidate finds $\sum d^2 = 266$, leading to $r_s = \text{awrt } -0.612$ then award M1M1A1M1A1.</p>
(b)	<p>1st B1 Both hypotheses stated in terms of ρ.</p> <p>M1 For a correct statement relating their r_s ($r_s < 1$) with their c.v. where $\text{their c.v.} < 1$</p> <p>A1 For a contextualised comment which is rejecting H_0, which must mention “<u>positive correlation</u>”, “<u>blood pressure</u>” and “<u>weight</u>”. (Use of “association” is A0.) Follow through their r_s with their c.v. (provided $\text{their c.v.} < 1$)</p> <p>Two-tailed test Applying a two-tailed test scores a maximum of B0B1M1A0 So Award SC B0B1 for $H_0: \rho = 0$, $H_1: \rho \neq 0$ followed by critical value $r_s = (\pm) 0.6485$ and allow access to the M1 mark only.</p>

Question Number	Scheme	Marks																																																				
<p>5. (a)</p>	<p>H_0 : There is no association between type of drink and gender (independent) H_1 : There is an association between type of drink and gender (dependent)</p> <table border="1" data-bbox="272 331 935 520"> <thead> <tr> <th>Expected</th> <th>Tea</th> <th>Coffee</th> <th>Hot Chocolate</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Male</td> <td>46.53</td> <td>34.31</td> <td>13.16</td> <td>94</td> </tr> <tr> <td>Female</td> <td>52.47</td> <td>38.69</td> <td>14.84</td> <td>106</td> </tr> <tr> <td>Total</td> <td>99</td> <td>73</td> <td>28</td> <td>200</td> </tr> </tbody> </table> <table border="1" data-bbox="272 611 881 968"> <thead> <tr> <th>Observed</th> <th>Expected</th> <th>$\frac{(O - E)^2}{E}$</th> <th>$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>57</td> <td>46.53</td> <td>2.3559...</td> <td>69.8259...</td> </tr> <tr> <td>26</td> <td>34.31</td> <td>2.0127...</td> <td>19.7027...</td> </tr> <tr> <td>11</td> <td>13.16</td> <td>0.3545...</td> <td>9.1945...</td> </tr> <tr> <td>42</td> <td>52.47</td> <td>2.0892...</td> <td>33.6192...</td> </tr> <tr> <td>47</td> <td>38.69</td> <td>1.7849...</td> <td>57.0949...</td> </tr> <tr> <td>17</td> <td>14.84</td> <td>0.3144...</td> <td>19.4744...</td> </tr> <tr> <td colspan="2">Totals</td> <td>8.9116...</td> <td>208.9116...</td> </tr> </tbody> </table> <p>$X^2 = \sum \frac{(O - E)^2}{E} \text{ or } \sum \frac{O^2}{E} - 200 ; = 8.9116...$</p> <p>$\nu = (2 - 1)(3 - 1) = 2$</p> <p>$\chi^2_2(0.05) = 5.991 \Rightarrow \text{CR: } X^2 \geq 5.991$</p> <p>[in the CR/significant/Reject H_0]</p> <p>conclude that there is an association between type of <u>drink</u> preferred and <u>gender</u>. (or they are not independent.)</p>	Expected	Tea	Coffee	Hot Chocolate	Total	Male	46.53	34.31	13.16	94	Female	52.47	38.69	14.84	106	Total	99	73	28	200	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	57	46.53	2.3559...	69.8259...	26	34.31	2.0127...	19.7027...	11	13.16	0.3545...	9.1945...	42	52.47	2.0892...	33.6192...	47	38.69	1.7849...	57.0949...	17	14.84	0.3144...	19.4744...	Totals		8.9116...	208.9116...	<p>Correct hypotheses B1</p> <p>Some attempt at (Row Total)(Column Total) (Grand Total) M1</p> <p>Can be implied by at least one correct E_i to 1d.p.</p> <p>All expected frequencies are correct. Condone exact fractions. A1</p> <p>At least 2 correct terms for $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. dM1</p> <p>Accept 2 sf accuracy for the dM1 mark.</p> <p>At least 5 correct $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ terms to either 2 dp or better. A1</p> <p>Allow truncation.</p> <p>For applying either $\sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 200$ dM1</p> <p>8.9 or awrt (8.88 - 8.91) A1</p> <p>$\nu = 2$ B1</p> <p>5.991 or ft $\chi^2_{\text{their } \nu}(0.05)$ B1ft</p> <p>A correct conclusion in context which is based on <i>their</i> X^2-value and <i>their</i> χ^2-critical value. A1</p>
	Expected	Tea	Coffee	Hot Chocolate	Total																																																	
Male	46.53	34.31	13.16	94																																																		
Female	52.47	38.69	14.84	106																																																		
Total	99	73	28	200																																																		
Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																																																			
57	46.53	2.3559...	69.8259...																																																			
26	34.31	2.0127...	19.7027...																																																			
11	13.16	0.3545...	9.1945...																																																			
42	52.47	2.0892...	33.6192...																																																			
47	38.69	1.7849...	57.0949...																																																			
17	14.84	0.3144...	19.4744...																																																			
Totals		8.9116...	208.9116...																																																			
<p>(b)</p>	<p>$\chi^2_2(0.005) = 10.597 \Rightarrow \text{CR: } X^2 \geq 10.597$</p> <p>[not in the CR/not significant/do not reject H_0]</p> <p>Either</p> <ul style="list-style-type: none"> Conclude there is no association between type of drink preferred and gender (or they are independent). The conclusion would change (if a correct H_0 has been rejected in part (a)). 	<p>Critical value of 10.597 B1</p> <p>Any one of these. B1</p> <p>[10]</p> <p>[2]</p> <p>12</p>																																																				

		Notes
5. (a)	1st B1	For both hypotheses. Must mention “drink” and “gender” or “sex” at least once. Use of “relationship” or “correlation” or “connection” is B0.
	2nd dM1	Dependent on the first method mark. At least 2 correct terms (as in 3 rd or 4 th column) or <i>correct expressions</i> with their E_i
	2nd A1	All correct terms to either 2 d.p. or better. Allow truncated answers.
	3rd dM1	Dependent on the second method mark. For applying either $\sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 200$
	3rd A1	8.9 or awrt (8.88 – 8.91)
	2nd B1	$\nu = 2$ This mark can be implied by a correct critical value of 5.991
	Note	If 8.9 or awrt (8.88 – 8.91) is seen (from a calculator) without the expected frequencies stated then award special case M0A0M1A1M1A1.
	Final A1	Dependent on the third method mark. A correct contextualised conclusion which is rejecting H_0 . Must mention “drink” and “gender” or “sex”. No follow through. If e.g. hypotheses are the wrong way round A0 here.
	Note	Contradictory statements score A0. E.g. “significant, do not reject H_0 ”.
	Note	Condone “relationship” or “connection” here but not “correlation”. e.g. “There is evidence of a relationship between grades and gender”
Note	Full accuracy gives $X^2 = 8.911619\dots$ and p-value 0.0116 to 0.0117	

Question Number	Scheme	Marks																																																														
6. (a)	$\hat{p} = \frac{0(2) + 1(21) + 2(45) + 3(42) + 4(12) + 5(3)}{8(2+21+45+42+12+3) \text{ or } 8(125)} \left\{ = \frac{300}{1000} \right\} = 0.3 (*)$	Answer is given. See notes. M1 A1cso [2]																																																														
(b)	$r = 125 \times {}^8C_3(0.3)^3(0.7)^5 \{= 31.76523...\} \text{ (formula)}$ or $r = 125 \times (0.8059 - 0.5518) \{= 31.7625\} \text{ (tables)}$ $s = 125 - (7.21 + 24.71 + 37.06 + \text{their } r + 17.02 + 5.83) \{= 1.40477... \text{ or } 1.4075\}$ or $s = 125 \times (1 - 0.9887) \{= 1.4125\}$	M1																																																														
	$r = 31.76523 \text{ or } 31.7625 \text{ or } 31.7575$ $s = 1.40477 \text{ or } 1.4075 \text{ or } 1.4125$	$r = \text{awrt } 31.77 \text{ or } r = \text{awrt } 31.76$ $s = 1.4 \text{ or awrt } 1.40 \text{ or } s = \text{awrt } 1.41$ A1 A1 [3]																																																														
(c)	<table border="1" data-bbox="272 636 1328 1066"> <thead> <tr> <th># failed tasks</th> <th>O_i</th> <th>E_i</th> <th>Comb O_i</th> <th>Comb E_i</th> <th>$\frac{(O - E)^2}{E}$</th> <th>$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>2</td> <td>7.21</td> <td>2</td> <td>7.21</td> <td>3.7648...</td> <td>0.5548...</td> </tr> <tr> <td>1</td> <td>21</td> <td>24.71</td> <td>21</td> <td>24.71</td> <td>0.5570...</td> <td>17.8470...</td> </tr> <tr> <td>2</td> <td>45</td> <td>37.06</td> <td>45</td> <td>37.06</td> <td>1.7011...</td> <td>54.6411...</td> </tr> <tr> <td>3</td> <td>42</td> <td>31.77 (31.76)</td> <td>42</td> <td>31.77 (31.76)</td> <td>3.2941... (3.3016...)</td> <td>55.5241... (55.5416...)</td> </tr> <tr> <td>4</td> <td>12</td> <td>17.02</td> <td>12</td> <td>17.02</td> <td>1.4806...</td> <td>8.4606...</td> </tr> <tr> <td>5</td> <td>3</td> <td>5.83</td> <td rowspan="2">3</td> <td>7.23</td> <td>2.4748...</td> <td>1.2448...</td> </tr> <tr> <td>≥ 6</td> <td>0</td> <td>1.40 (1.41)</td> <td>(7.24) {7.25}</td> <td>(2.4831...)</td> <td>(1.2431...)</td> </tr> <tr> <td colspan="5" style="text-align: center;">Totals</td> <td>13.2724... (13.2882...)</td> <td>138.2724... (138.2882...)</td> </tr> </tbody> </table> <p> $X^2 = \sum \frac{(O - E)^2}{E} \text{ or } \sum \frac{O^2}{E} - 125 ;= \text{awrt } 13.3$ </p> <p> $v = 6 - 1 - 1 = 4$ </p> <p> $\chi_4^2(0.05) = 9.488 \Rightarrow \text{CR: } X^2 \geq 9.488$ </p> <p> H_0 : Binomial distribution is a good(or suitable) model (or fit). H_1 : Binomial distribution is not a suitable model. </p> <p>[in the CR/significant/Reject H_0]</p> <p>Binomial distribution is not a suitable model.</p>	# failed tasks	O_i	E_i	Comb O_i	Comb E_i	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	0	2	7.21	2	7.21	3.7648...	0.5548...	1	21	24.71	21	24.71	0.5570...	17.8470...	2	45	37.06	45	37.06	1.7011...	54.6411...	3	42	31.77 (31.76)	42	31.77 (31.76)	3.2941... (3.3016...)	55.5241... (55.5416...)	4	12	17.02	12	17.02	1.4806...	8.4606...	5	3	5.83	3	7.23	2.4748...	1.2448...	≥ 6	0	1.40 (1.41)	(7.24) {7.25}	(2.4831...)	(1.2431...)	Totals					13.2724... (13.2882...)	138.2724... (138.2882...)	For applying either $\sum \frac{(O - E)^2}{E} \text{ or } \sum \frac{O^2}{E} - 125$ awrt 13.3 see notes a correct ft for their $\chi_k^2(0.05)$, where $k = n - 1 - 1$ from their n . Correct hypotheses A correct conclusion (context not required here) which is based on their X^2 -value and their χ^2 -critical value. dM1 A1 B1 ft B1 B1 A1 [8]
# failed tasks	O_i	E_i	Comb O_i	Comb E_i	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																																																										
0	2	7.21	2	7.21	3.7648...	0.5548...																																																										
1	21	24.71	21	24.71	0.5570...	17.8470...																																																										
2	45	37.06	45	37.06	1.7011...	54.6411...																																																										
3	42	31.77 (31.76)	42	31.77 (31.76)	3.2941... (3.3016...)	55.5241... (55.5416...)																																																										
4	12	17.02	12	17.02	1.4806...	8.4606...																																																										
5	3	5.83	3	7.23	2.4748...	1.2448...																																																										
≥ 6	0	1.40 (1.41)		(7.24) {7.25}	(2.4831...)	(1.2431...)																																																										
Totals					13.2724... (13.2882...)	138.2724... (138.2882...)																																																										
(d)	Following from a correct conclusion in part (c), a comment conveying either <ul style="list-style-type: none"> p is not constant employer's belief is not justified. 	B1 [1] 14																																																														

Notes	
6. (a)	<p>M1 Must show clearly how to get either 300 or 1000. A1 cso Showing how to get <u>both</u> 300 and 1000 and reaching $p = 0.3$</p>
(b)	<p>M1 For any correct method (or a correct expression) for finding either r or s. A1 $r = \text{awrt } 31.77$ or $r = \text{awrt } 31.76$ A1 $s = 1.4$ or $\text{awrt } 1.40$ or $s = \text{awrt } 1.41$</p>
(c)	<p>1st M1 For an attempt to pool 5 failed tasks and ≥ 6 failed tasks ONLY. Note Give 1st M0 for pooling 0 failed tasks and 1 failed task. 2nd M1 For an attempt at the test statistic, at least 2 correct expressions/values (to awrt 2 d.p. or truncated 2 d.p.) 3rd dM1 <i>Dependent on the second method mark.</i> For applying either $\sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 125$ 1st A1 awrt 13.3 1st B1ft For their evaluated $n - 1 - 1$. i.e. realising that they must subtract 2 from their n. 2nd B1 For a correct ft for their $\chi_k^2(0.05)$, where $k = n - 1 - 1$ from their n. 3rd B1 Must have both hypotheses and mention Binomial at least once. Inclusion of 0.3 for p in hypotheses is B0 but condone in conclusion. Final A1 <i>Dependent on the 2nd and 3rd Method marks only.</i> A correct conclusion (context not required) which is rejecting H_0. Note No follow through on their hypotheses if they are stated the wrong way round. Note Contradictory statements score A0. E.g. “significant, do not reject H_0”. Note Condone mentioning of Bin(8, 0.3) in conclusion Note Full accuracy gives a combined expected frequency of 7.245956..., $\frac{(O - E)^2}{E} = 2.4880$, $\frac{O^2}{E} = 1.2421$, $X^2 = 13.28333...$ Note p-value for the test is 0.0099 to 0.0100 Note No combining gives $X^2 = 13.58...$ Note Combining 0/1 and 4/5/≥ 6 gives $X^2 = 11.02$</p>

Question Number	Scheme		Marks
<p>7. (a)</p> <p>(b)</p>	<p>$X = 4Y - 3W$, $Y \sim N(40, 3^2)$, $W \sim N(50, 2^2)$; Y, W are independent.</p> <p>$\{E(X) = 4E(Y) - 3E(W) = 4(40) - 3(50)\} \Rightarrow E(X) = 10$ $E(X) = 10$ (seen or implied)</p> <p>$\text{Var}(X) = 16\text{Var}(Y) + 9\text{Var}(W)$ Either $(4^2)\text{Var}(Y)$ or $+(3^2)\text{Var}(W)$</p> <p>$\{\text{Var}(X) = 16(9) + 9(4)\} \Rightarrow \text{Var}(X) = 180$ For adding the variances</p> <p>$\{\text{So } X \sim N(10, 180)\}$ $\text{Var}(X) = 180$</p> <p>$\{P(X > 25) = \} P\left(Z > \frac{25 - 10}{\sqrt{180}}\right)$ Standardising (\pm) with their mean</p> <p>$= P(Z > 1.11803\dots)$ and their standard deviation</p> <p>$= 1 - 0.8686$ awrt ± 1.12</p> <p>$= 0.1314$ (or $0.131777\dots$) awrt 0.131 or awrt 0.132</p> <hr/> <p>$A = \sum_{i=1}^3 Y_i$, $C \sim N(115, \sigma^2)$; $P(A - C < 0) = 0.2$; A, C are independent.</p> <p>$\{E(A - C) = 3E(Y) - E(C) = 3(40) - (115)\} \Rightarrow E(A - C) = 5$ $E(A - C) = 5$</p> <p>$\text{Var}(A - C) = 3\text{Var}(Y) + \text{Var}(C)$ $3\text{Var}(Y)$ and a + ...</p> <p>$\{\text{Var}(A - C) = 3(9) + \sigma^2\} \Rightarrow \text{Var}(A - C) = 27 + \sigma^2$ $\text{Var}(A - C) = 27 + \sigma^2$</p> <p>$\{\text{So } A - C \sim N(5, 27 + \sigma^2)\}$</p> <p>$\{P(A - C < 0) = 0.2\} \Rightarrow P\left(Z < \frac{-5}{\sqrt{27 + \sigma^2}}\right) = 0.2$</p> <hr/> <p>$\frac{-5}{\sqrt{27 + \sigma^2}} = k$ ($= -0.8416$) Standardising (\pm) with their mean and their standard deviation</p> <p>which is in terms of σ^2 and setting the result equal to k, where k is in the interval $[0.84, 0.85]$.</p> <hr/> <p>± 0.8416 or awrt ± 0.8416</p> <hr/> <p>Correct equation. See notes</p> <hr/> <p>$\sigma^2 = \left(\frac{-5}{-0.8416}\right)^2 - 27 \Rightarrow \sigma^2 = \dots$ Squaring and rearranging leading to a positive value for σ^2.</p> <hr/> <p>$\sigma^2 = 8.2962\dots$ ($= 8.4308\dots$ from using -0.84) awrt 8.3 or awrt 8.4</p> <p>($= 8.2945\dots$ from calculator, so need awrt 8.29 for full marks if no prior working is shown.)</p>	<p>B1</p> <p>M1</p> <p>M1</p> <p>A1</p> <p>M1</p> <p>A1</p> <p>A1</p> <p>[7]</p> <p>B1</p> <p>M1</p> <p>A1</p> <p>M1</p> <p>B1</p> <p>A1</p> <p>dM1</p> <p>A1 cso</p> <p>[8]</p> <p>15</p>	
<p>(a)</p> <p>(b)</p>	<p>Note Condone applying reversed variances, e.g. $16(4) + 9(9)$ for the first 2 method marks.</p> <p>Note $\text{Var}(X) = 180$ with no working gets M1M1A1</p> <p>Note $\text{Var}(X) = 48$ with no working gets M0M1A0</p> <p>Note $\text{Var}(X) = 108$ with no working gets M1M0A0</p> <p>Note $\text{Var}(X) = 24$ with no working gets M0M0A0</p> <p>2nd M1 Allow $\frac{\pm \text{their } E(A - C)}{\sqrt{\text{their } \text{Var}(A - C)}} = k$, where k is in the interval $(0.84, 0.85)$.</p> <p>2nd B1 For either -0.8416 or 0.8416</p> <p>2nd A1 E.g. Allow $\frac{-5}{\sqrt{27 + \sigma^2}} = [-0.85, -0.84]$ or $\frac{5}{\sqrt{27 + \sigma^2}} = [0.84, 0.85]$</p> <p>3rd M1 <i>Dependent on the 2nd M1 mark being awarded.</i></p>		



Mark Scheme (Results)

Summer 2014

Pearson Edexcel GCE in Statistics 3R
(6691/01R)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2014

Publications Code UA040138

All the material in this publication is copyright
© Pearson Education Ltd 2014

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
 - ft – follow through
 - the symbol \surd will be used for correct ft
 - cao – correct answer only
 - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - d... or dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper or ag- answer given
 - \square or d... The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
6. If a candidate makes more than one attempt at any question:
 - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
 - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.
7. Ignore wrong working or incorrect statements following a correct answer.

Question Number	Scheme									Marks	
<p>1.</p> <p>a)</p> <p>b)</p> <p>c)</p> <p>d)</p>	Car model	A	B	C	D	E	F	G	H	<p>M1</p> <p>M1A1</p> <p>M1</p> <p>A1</p> <p>(5)</p> <p>B1</p> <p>B1</p> <p>M1A1ft</p> <p>(4)</p> <p>B1</p> <p>(1)</p> <p>B1</p> <p>(1)</p> <p>(11 marks)</p>	
	Sales rank	8	6	1	5	4	7	2	3		
	Fuel efficiency rank	8	1	5	6	2	7	4	3		
	d^2	0	25	16	1	4	0	4	0		
	$\sum d^2 = 50$										
	$r_s = 1 - \frac{6 \sum d^2}{8(64 - 1)} = 1 - \frac{6 \times 50}{8 \times 63}$										
	$r_s = \frac{204}{504} = 0.40476 \dots$										
	awrt 0.405										
	$H_0: \rho_s = 0 \quad H_1: \rho_s > 0 \quad (\text{accept } \rho_s \text{ or } \rho)$										
	1 tail critical value $\rho = 0.6429$										
Test value is not in critical region so insufficient evidence to reject H_0											
No significant evidence at 5% level to support journalist's belief											
Underlying (bivariate) Normal distribution											
Evidence does not support Normal distribution since mean < median or (negative) skew, oe											

Notes	
a)	M1 for attempting to rank at least one set of data A1 for at least one set of data ranked correctly (NB this mark comes after 2 nd M1) M1 for attempting Σd^2 M1 for correct use of formula for r_s
b)	B1 for H_0 and H_1 correct (condone \leq for H_0) 2 nd B1 allow 0.7381 if their $H_1: \rho_s \neq 0$ M1 for correct statement relating their test statistic and critical value A1ft their test statistic, H_1 and critical value but must be in context.
c)	B1 require Normal distribution, ignore additional assumptions
d)	B1 require not Normal and valid reason

Question Number	Scheme	Marks
<p>2)</p> <p>(a)</p> <p>(i)</p> <p>(ii)</p> <p>(b)</p>	<p>Expected value = $\frac{50 \times 74}{200} = 18.5$</p> <p>$\chi^2$ contribution = $\frac{(27-18.5)^2}{18.5} = 3.905405405 = 3.91$ to 3sfs</p> <p>H_0: users age and main mobile phone use are independent/ no association between users age and main mobile phone use H_1: users age and main mobile phone use are not independent/ some association between users age and main mobile phone use</p> <p>$\nu = 4$</p> <p>Critical value $\chi^2 = 9.488$</p> <p>Test statistic is in critical region therefore significant evidence to reject H_0 and accept H_1. Evidence at 5% level that age and main phone use are not independent.</p>	<p>B1 cso</p> <p>B1 cso</p> <p>(2)</p> <p>B1</p> <p>B1</p> <p>B1ft</p> <p>M1</p> <p>A1ft</p> <p>(5)</p> <p>(7 marks)</p>
Notes		
<p>(b)</p>	<p>3rd B1 ft on their value of ν</p> <p>M1 for attempt to compare test statistic and their critical value</p> <p>A1 ft on test statistic and critical value but must be comment in context. (A0 if hypotheses are the wrong way around)</p>	

Question Number	Scheme	Marks
<p>3) (a)</p> <p>(b)</p>	$P(S > 2C) = P(S - 2C > 0)$ $E[S - 2C] = 4.9 - 2 \times 2.5 = -0.1$ $\text{Var}(S - 2C) = 0.64 + 4 \times 0.16 = 1.28$ $P(S - 2C > 0), = P(Z > \frac{0 - -0.1}{\sqrt{1.28}})$ $= P(Z > 0.08838\dots)$ $= 0.4641 \quad (\text{tables}), \text{ or } 0.4648\dots (\text{calculator})$ <p style="text-align: right;">accept awrt 0.464 or 0.465</p> $\text{Let } T = S_1 + S_2 + \dots + S_{100}$ $E[T] = 100 \times 4.9 = 490$ $\text{Var}(T) = 100 \times 0.64 = 64$ $P(T < 500) = P(Z < \frac{500 - 490}{\sqrt{64}})$ $= P(Z < 1.25)$ $= 0.8944$	<p>B1</p> <p>M1A1</p> <p>M1, M1</p> <p>A1</p> <p>(6)</p> <p>M1A1</p> <p>A1</p> <p>M1</p> <p>A1</p> <p>(5)</p> <p>(11 marks)</p>
Notes		
<p>(a)</p> <p>(b)</p>	<p>1st M1 for $\dots \pm 4\text{Var}(C)$ 2nd M1 for $P(S - 2C > 0)$ 3rd M1 ft their expectation and variance but not if $\text{Var}(S - 2C)$ is negative. (Should lead to $P(Z > +ve)$)</p> <p>1st M1 for attempt to find mean or variance of total</p> <p>1st A1 either correct</p> <p>2nd A1 both correct</p> <p>2nd M1 for standardising using 500, their mean and their sd leading to $P(Z < +ve)$ o.e.</p> <p>Sample mean, $\bar{x} = \frac{660 + \alpha}{5} = 132 + \frac{\alpha}{5}$</p>	

Question Number	Scheme	Marks
4)	<p>Test statistic, $z = \frac{132 + \frac{\alpha}{5} - 160}{\frac{6}{\sqrt{5}}}$</p> <p>Critical z values is 1.6449</p> <p>Therefore the test statistic is significant if</p> $\frac{132 + \frac{\alpha}{5} - 160}{\frac{6}{\sqrt{5}}} > 1.6449$ <p>Therefore</p> $132 + \frac{\alpha}{5} - 160 > 1.6449 \times \frac{6}{\sqrt{5}}$ $\alpha > 5 \left(1.6449 \times \frac{6}{\sqrt{5}} + 28 \right)$ $\alpha > 162.0686493 \dots$ <p>Accept awrt 162.1</p>	<p>M1A1ft</p> <p>B1</p> <p>M1</p> <p>A1</p> <p>(6)</p> <p>(6 marks)</p>
Notes		
	<p>1st A1 ft on their \bar{x}</p> <p>1st B1 given for 1.6449 seen (condone sign)</p> <p>3rd M1 <u>inequality</u> using their test statistic, accept incorrect signs for M1</p>	

Question Number	Scheme	Marks
<p>5)</p> <p>(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p>	$S_E^2 = \frac{1}{n-1} \left(\sum x^2 - \frac{(\sum x)^2}{n} \right) = \frac{1}{119} \left(956909 - \frac{10650^2}{120} \right)$ $= \frac{11721.5}{119} = 98.5$ <p> $H_0: \mu_F = \mu_E,$ $H_1: \mu_F \neq \mu_E,$ </p> $\bar{x}_E = \frac{10650}{120} = 88.75 \quad \text{and} \quad \bar{x}_F = \frac{6510}{70} = 93$ <p>Test statistic, $z = \frac{93-88.75-0}{\sqrt{\frac{151}{70} + \frac{98.5}{120}}} = 2.4627 \dots$</p> <p>Critical values, $z = (\pm)2.5758$</p> <p>Test stat is not in critical region</p> <p>Insufficient evidence to reject H_0 at 1% level</p> <p>No significant evidence of a difference in mean lengths of English and French films</p> <p>By CLT we can assume that the mean of a large sample has a Normal distribution</p> <p>On a list, label English films 1 – 724 and French films 1-473 (oe)</p> <p>Use random number table/generator to select</p> $\frac{724}{724+473} \times 190 = 115 \text{ English films and}$ $\frac{473}{1197} \times 190 = 75 \text{ French films}$	<p>M1</p> <p>A1 (2)</p> <p>B1</p> <p>M1</p> <p>M1A1</p> <p>B1ft</p> <p>M1</p> <p>A1ft (7)</p> <p>B1 (1)</p> <p>B1</p> <p>M1A1</p> <p>(3)</p> <p>(13 marks)</p>

Notes	
(a)	<p>Alternative</p> $S_E^2 = \frac{n}{n-1} \left(\frac{\sum x^2}{n} - \bar{x}^2 \right) = \frac{120}{119} \left(\frac{956909}{120} - 88.75^2 \right) = 98.5$
(b)	<p>1st B1 needs both H₀ and H₁, can be in words</p> <p>2nd B1ft on their H₁</p> <p>1st M1 for attempt @ both means (\bar{x}_E may be in (a))</p> <p>2nd M1 for attempt at correct test statistic, ft their values</p> <p>3rd M1 for attempt to compare their test stat and critical values</p> <p>A1 ft on their test and critical values but must include comment in context</p>
(c)	<p>Require mention of mean of <i>E</i> or <i>F</i> and normal distribution</p>
(d)	<p>M1 requires use of <u>random numbers</u> and attempt to find correct sample sizes</p> <p>A1 both 115 and 75 found.</p>

Question Number	Scheme	Marks						
6)	Independence of each occurrence (of a fake coin)	B1						
(a)	Constant probability for each occurrence (of a fake)	B1 (2)						
(b)	$r = 150 \times P(X = 2) = 150 \times \binom{20}{2} \times 0.05^2 \times 0.95^{18}$ $r = 28.3015\dots$ <p style="text-align: right;">awrt 28.3</p> $s = 150 - (53.8 + 56.6 + 28.3 + 8.9) = 2.4$	M1 A1 A1ft (3)						
(c)	<p>H_0: Bin(20, 0.05) is a suitable model H_1: Bin(20, 0.05) is not a suitable model</p> <p>Combining last two groups</p> <table border="1" data-bbox="370 999 1240 1108"> <tr> <td></td> <td style="text-align: center;">≥ 3</td> </tr> <tr> <td style="text-align: center;">Observed frequency</td> <td style="text-align: center;">19</td> </tr> <tr> <td style="text-align: center;">Expected frequency</td> <td style="text-align: center;">11.3</td> </tr> </table> <p>$\nu = 4 - 1 = 3$</p> <p>Critical value, $\chi^2(0.05) = 7.815$ (accept 9.488 if their $\nu = 4$)</p> <p>Test statistic, $\sum \frac{(O-E)^2}{E} = \frac{(43-53.8)^2}{53.8} + \frac{(62-56.6)^2}{56.6} + \dots$</p> $= 2.168\dots + 0.515\dots + 0.186\dots + 5.246\dots$ $= 8.117 \quad (\text{accept } 10.16 \text{ if groups not combined})$ <p>In critical region, sufficient evidence to reject H_0, accept H_1</p> <p>Significant evidence at 5% level to reject the manager's model</p>		≥ 3	Observed frequency	19	Expected frequency	11.3	B1 M1 B1 B1ft M1 A1ft A1ft (7)
	≥ 3							
Observed frequency	19							
Expected frequency	11.3							

Question Number	Scheme	Marks
(d)	$\nu = 4 - 2 = 2$ 4 classes due to pooling 2 restrictions (equal total and mean/proportion)	B1 B1 (2)
(e)	H_0 : Binomial distribution is a good model H_1 : Binomial distribution is not a good model Critical value, $\chi^2(0.05) = 5.991$ Test statistic is not in critical region, insufficient evidence to reject H_0 Accept the assistant manager's model for the number of fake coins per bag.	B1 B1 B1 (3) (17 marks)
Notes		
(b)	M1A1 for one of r or s correct A1ft for other value if using $150 - \dots$ and answer must be >0	
(c)	1 st B1 can be in words but must include $p = 0.05$ 3 rd B1 ft on their ν Test statistic alternative method $\text{Test stat} = \sum \frac{O^2}{E} - 150 = \frac{43^2}{53.8} + \frac{62^2}{56.6} + \dots - 150 = 8.117 \dots$ 1 st A1 ft if their groups not combined 2 nd A1 ft their test and critical values but must be comment in context e.g. mention of "manager's model" <u>or</u> "fake coins"	
(d)	1 st B1 evidence that pooling is required 2 nd B1 must have correct reasons for restrictions.	

Question Number	Scheme	Marks
<p>7) (a) (i)</p>	$\bar{x} = \frac{10.01+9.97+9.93+\dots}{8} = 9.9775$ <p>95% CI $\bar{x} \pm 1.96 \times \frac{0.08}{\sqrt{8}}$</p> <p>95% CI for μ (9.92, 10.03)</p>	<p>M1</p> <p>B1M1</p> <p>A1</p> <p>(4)</p>
(ii)	10.00 is within confidence interval so accept that pump may be performing correctly (although sample mean is low).	<p>B1</p> <p>(1)</p>
(b)	<p>Upper limit of CI is</p> $9.96 + 1.6449 \times \frac{0.08}{\sqrt{n}} < 10.00$ $\frac{1.6449 \times 0.08}{\sqrt{n}} < 0.04$ $\sqrt{n} > \frac{1.6449 \times 0.08}{0.04}$ <p>$n > 10.82\dots$ therefore minimum $n = 11$</p>	<p>B1, M1A1ft</p> <p>M1</p> <p>A1 cao</p> <p>(5)</p> <p>(10 marks)</p>

Notes		
(a)		
(i)	1 st M1 attempt to find sample mean B1 for correct z value A1 limits correct to 2 decimal places (or more)	
(b)	B1 for correct z value 1 st M1A1, ft their z value	



Mark Scheme (Results)

Summer 2014

Pearson Edexcel GCE in Statistics 3
(6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2014

Publications Code UA040135

All the material in this publication is copyright

© Pearson Education Ltd 2014

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
 - ft – follow through
 - the symbol \surd will be used for correct ft
 - cao – correct answer only
 - cso – correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - d... or dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper or ag- answer given
 - \square or d... The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
6. If a candidate makes more than one attempt at any question:
 - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
 - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.
7. Ignore wrong working or incorrect statements following a correct answer.

Question Number	Scheme	Marks
1(a)	(This is a sample where) every (possible) sample (of size n) has an equal chance of being chosen.	B1 (1)
(b)	‘When it is impossible to provide a sampling frame ’ or a correct example with an indication of sampling frame being impossible.	B1 (1)
(c)(i)	A list/register of all the students.	B1
(ii)	Number the students (from 0 to 74, 1 to 75 etc.)	B1
	Using the random no. table read off the nos. and identify or select the students allocated those nos.	B1 (3) Total 5
Notes		
(a)	Require all / each / every etc sample and same/equal etc chance / probability etc for B1	
(b)	Require impossible / no / doesn't exist etc and sampling frame for B1	
(c)(i)	Require list/register etc and all/every/75 etc students for B1 List of 8 students is B0	
(ii)	First B1 accept ‘ in the corresponding position ’ o.e. if numbering omitted Second B1 require both for mark.	

Question Number	Scheme	Marks
2a(i) (ii)(iii) (b)	Only contains known data / function of data only / no population parameters therefore it is a statistic (ii) and (iii) contain unknown parameters / population parameters / μ and / or σ therefore it is not a statistic . $(E(\frac{3X_1 - X_{20}}{2}) = \frac{3\mu - \mu}{2} =) \mu$ $\text{Var}(\frac{3X_1 - X_{20}}{2}) = \frac{9\sigma^2 + \sigma^2}{2^2}$ $= \frac{5\sigma^2}{2}$	B1 B1d B1 B1d (4) B1 M1 A1 (3) Total 7
Notes		
(a)(i) (b)	First B1 for known / no unknowns o.e. in (i) Second B1 dependent on first B1 for 'Yes' / is a statistic o.e. in (i) Third B1 for unknowns o.e. in both (ii) and (iii) Fourth B1 dependent on third B1 for 'No' / not a statistic o.e. in both (ii) and (ii) B1 for μ M1 for some squaring on numerator or denominator and must add on numerator A1 for $\frac{5\sigma^2}{2}$ o.e.	

Question Number	Scheme	Marks																																													
3	<table border="1" style="width: 100%; border-collapse: collapse; margin-bottom: 10px;"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="3" style="text-align: center;">Happiness</th> </tr> <tr> <th style="text-align: center;">Not happy</th> <th style="text-align: center;">Fairly happy</th> <th style="text-align: center;">Very happy</th> </tr> </thead> <tbody> <tr> <th rowspan="2" style="text-align: center;">Gender</th> <th style="text-align: center;">Female</th> <td style="text-align: center;">13.51</td> <td style="text-align: center;">41.77</td> <td style="text-align: center;">30.71</td> </tr> <tr> <th style="text-align: center;">Male</th> <td style="text-align: center;">8.49</td> <td style="text-align: center;">26.23</td> <td style="text-align: center;">19.29</td> </tr> </tbody> </table> <p> H_0 : Happiness and gender are independent/ not associated H_1 : Happiness and gender are not independent/ associated </p> <table border="1" style="width: 100%; border-collapse: collapse; margin-bottom: 10px;"> <thead> <tr> <th style="text-align: center;">O</th> <th style="text-align: center;">E</th> <th style="text-align: center;">$\frac{(O - E)^2}{E}$</th> <th style="text-align: center;">$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">9</td> <td style="text-align: center;">13.51</td> <td style="text-align: center;">1.508</td> <td style="text-align: center;">5.996</td> </tr> <tr> <td style="text-align: center;">43</td> <td style="text-align: center;">41.77</td> <td style="text-align: center;">0.0361</td> <td style="text-align: center;">44.264</td> </tr> <tr> <td style="text-align: center;">34</td> <td style="text-align: center;">30.71</td> <td style="text-align: center;">0.351</td> <td style="text-align: center;">37.637</td> </tr> <tr> <td style="text-align: center;">13</td> <td style="text-align: center;">8.49</td> <td style="text-align: center;">2.402</td> <td style="text-align: center;">19.915</td> </tr> <tr> <td style="text-align: center;">25</td> <td style="text-align: center;">26.23</td> <td style="text-align: center;">0.0575</td> <td style="text-align: center;">23.829</td> </tr> <tr> <td style="text-align: center;">16</td> <td style="text-align: center;">19.29</td> <td style="text-align: center;">0.560</td> <td style="text-align: center;">13.274</td> </tr> </tbody> </table> <p> $\sum \frac{(O - E)^2}{E} = 4.91$ or $\sum \frac{O^2}{E} - N = 144.91 - 140 = 4.91$ $\nu = (3 - 2)(2 - 1) = 2$ $\sum \frac{(O - E)^2}{E} < 5.991$ 4.91 < 5.991 so 'insufficient evidence to reject H_0' or 'Accept H_0' No association between gender and happiness. </p>			Happiness			Not happy	Fairly happy	Very happy	Gender	Female	13.51	41.77	30.71	Male	8.49	26.23	19.29	O	E	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	9	13.51	1.508	5.996	43	41.77	0.0361	44.264	34	30.71	0.351	37.637	13	8.49	2.402	19.915	25	26.23	0.0575	23.829	16	19.29	0.560	13.274	<p>M1 A1</p> <p>B1 dM1</p> <p>A1 A1</p> <p>B1 B1ft</p> <p>M1 A1</p> <p style="text-align: right;">(10) Total 10</p>
				Happiness																																											
		Not happy	Fairly happy	Very happy																																											
Gender	Female	13.51	41.77	30.71																																											
	Male	8.49	26.23	19.29																																											
O	E	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																																												
9	13.51	1.508	5.996																																												
43	41.77	0.0361	44.264																																												
34	30.71	0.351	37.637																																												
13	8.49	2.402	19.915																																												
25	26.23	0.0575	23.829																																												
16	19.29	0.560	13.274																																												
Notes																																															
	<p>1st M1 for some use of $\frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$. May be implied by at least 1 correct E_i</p> <p>1st A1 awrt 13.5, 41.8, 30.7, 8.5, 26.2 and 19.3 Allow M1A0 for E_i rounded to integers</p> <p>1st B1 for both hypotheses. Must mention "happiness" and "gender" at least once.</p> <p>Use of "relationship" or "correlation" or "connection" is B0</p> <p>2nd dM1 for at least 2 correct terms (in 3rd or 4th columns) or correct expressions with their E_i</p> <p>Dependent on 1st M1. Accept 2sf accuracy for the M mark.</p> <p>2nd A1 for all correct terms (2sf or better). May be implied by a correct ans</p> <p>Allow truncation e.g. 44.2...</p> <p>3rd A1 awrt 4.91 . Condone 4.915</p> <p>2nd B1 for correct degrees of freedom (may be implied by a cv of 5.991)</p> <p>3rd B1ft for cv that follows from their degrees of freedom</p> <p>3rd M1 for a correct statement linking their test statistic and their cv</p> <p>Contradictory statements score M0 e.g. "significant, do not reject H_0"</p> <p>Condone "reject H_1 "</p> <p>4th A1 for a correct comment in context - must mention "gender" and "happiness"</p> <p>Condone "relationship" or "connection" here but not "correlation".</p> <p>e.g. "There is no evidence of a relationship between gender and happiness"</p> <p>No follow through. If e.g hypotheses are the wrong way around A0 here.</p> <p>SC Use of calculator with no working may get M0A0B1M1A0A1B1B1M1A1</p>																																														

Question Number	Scheme	Marks
4	$E(A) = E(B) + 4E(C) - 3E(D)$ $= 22$ $\text{Var}(A) = \text{Var}(B) + 16\text{Var}(C) + 9\text{Var}(D)$ $= 168.25$ $P(A < 45) = P\left(Z < \frac{45 - 22}{\sqrt{168.25}}\right)$ $= P(Z < 1.773)$ $= 0.9616$	M1 A1 M1 A1 M1 A1 awrt 0.962 (6) Total 6
Notes		
	1 st M1 for $E(4C) = 4E(C)$ and $-E(3D) = -3E(D)$ 1 st A1 for 22 cao 2 nd M1 for use of $\text{Var}(aX) = a^2\text{Var} X$ and + their '9Var(D)' 2 nd A1 for 168.25 cao 3 rd M1 for standardising using their mean and their sd 3 rd A1 for awrt 0.962. NB Calculator gives 0.961899....	

Question Number	Scheme	Marks																								
5(a)	The seeds are independent / There are a fixed number of seeds in a row / There are only two outcomes to the seed germinating – either it germinates or it does not / The probability of a seed germinating is constant	B1 B1 (2)																								
(b)	$\frac{(0 \times 2) + (1 \times 6) + (2 \times 11) + (3 \times 19) + (4 \times 25) + (5 \times 32) + (6 \times 16) + (7 \times 9)}{120 \times 7} = \frac{504}{840} = 0.6^{**}$	M1 A1cso (2)																								
(c)	$p = 0.6 \quad q = 0.4$ $s = 120 \times 21q^5p^2 = 120 \times 21 \times 0.4^5 \times 0.6^2 = 9.29$ $t = 120 \times 35q^3p^4 = 120 \times 35 \times 0.4^3 \times 0.6^4 = 34.84$	B1 B1 (2)																								
(d)	H_0 : A binomial distribution is a suitable model. H_1 : A binomial distribution is not a suitable model.	B1 M1																								
	<table border="1"> <thead> <tr> <th>Observed number of rows</th> <th>19</th> <th>19</th> <th>25</th> <th>32</th> <th>25</th> </tr> </thead> <tbody> <tr> <td>Expected number of rows</td> <td>11.55</td> <td>23.22</td> <td>34.84</td> <td>31.35</td> <td>19.04</td> </tr> <tr> <td>$\frac{(O-E)^2}{E}$</td> <td>4.81</td> <td>0.77</td> <td>2.78</td> <td>0.013</td> <td>1.87</td> </tr> <tr> <td>$\frac{O^2}{E}$</td> <td>31.26</td> <td>15.55</td> <td>17.94</td> <td>32.66</td> <td>32.83</td> </tr> </tbody> </table>	Observed number of rows	19	19	25	32	25	Expected number of rows	11.55	23.22	34.84	31.35	19.04	$\frac{(O-E)^2}{E}$	4.81	0.77	2.78	0.013	1.87	$\frac{O^2}{E}$	31.26	15.55	17.94	32.66	32.83	
Observed number of rows	19	19	25	32	25																					
Expected number of rows	11.55	23.22	34.84	31.35	19.04																					
$\frac{(O-E)^2}{E}$	4.81	0.77	2.78	0.013	1.87																					
$\frac{O^2}{E}$	31.26	15.55	17.94	32.66	32.83																					
	$v = 5 - 2 = 3$ Critical value for $\chi^2 = 11.345$ $\sum \frac{(O-E)^2}{E} = 10.23$ or $\sum \frac{O^2}{E} - N = 130.23 - 120 = 10.23$ $10.23 < 11.345$ therefore do not reject H_0 A binomial is a suitable model.	B1ft B1ft M1A1 A1 (7)																								
Total 13																										

Notes

(a)	Any two and at least one must have context. 2 correct, no context B1B0. Do not award B0B1.	
(b)	M1 require at least two correct terms in numerator and /(120x7) or /120 then /7 A1 cso as given answer	
(c)	Cao for each B1	
(d)	1 st B1 for both hypotheses. B0 if they include 0.6 Condone $X \sim B(n,p)$ etc 1 st M1 for using some combined columns (<8) 2 nd B1ft follows from 'their no of columns' -2 3 rd B1ft follows from the degrees of freedom 2 nd M1 for attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ with at least 2 nd (3 seeds) and 4 th (5 seeds) accurate to 2sf Contradictory statements score M0 e.g. "significant" do not reject H_0 1 st A1 for awrt 10.2 2 nd A1 dependent on 2 nd M for a correct comment suggesting that binomial model is suitable. No follow through . Condone mention of 0.6 here. Hypotheses wrong way round scores A0	

Question Number	Scheme	Marks
6(a)	$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$ $E(\bar{X}) = \frac{1}{n} E(X_1 + \dots + X_n)$ $= \frac{1}{n} (E(X_1) + \dots + E(X_n))$ $= \frac{1}{n} (\mu + \dots + \mu)$ $= \frac{n\mu}{n} = \mu$	B1cso (1)
(b)	$\bar{x} = \frac{1}{5} (197 + 203 + 205 + 201 + 195)$ $= 200.2(\text{g})$ $s^2 = \frac{1}{n-1} (\sum x^2 - n\bar{x}^2) \quad \text{or} \quad \frac{n}{n-1} \text{Var } x$ $= \frac{1}{5-1} (200469 - 5(200.2^2))$ $= 17.2$	B1 M1 A1 (3)
(c)	<p>We require $2 \times 1.25 \geq \text{Width of confidence interval}$</p> $2.5 \geq \frac{2 \times 1.96 \times 4.8}{\sqrt{n}} \quad \text{or} \quad 1.25 \geq \frac{1.96 \times 4.8}{\sqrt{n}} \quad \text{or} \quad \frac{1.25}{4.8} \geq \frac{1.96}{\sqrt{n}}$ $\sqrt{n} \geq \frac{2 \times 1.96 \times 4.8}{2.5} = 7.5264$ $n \geq 56.6(5)$ <p>Minimum sample size is 57</p>	M1B1 A1 A1 (4)
Notes		
(a)	B1 cso: require $E(\bar{X}) = \mu$ with at least 1 correct intermediate step and no incorrect working.	
(b)	B1 for 200.2 or $\frac{1001}{5}$ M1 for use of correct formula. Accept $\frac{1}{4} S_{xx} = \frac{1}{4} \times 68.8$ A1 for awrt 17.2	
(c)	M1 for use of any equivalent expression. Accept equality. Accept their s instead of 4.8 B1 for 1.96 seen with s.e. 1 st A1 for 56.6(5) 2 nd A1 for 57. Must follow from correct working e.g. $\sqrt{n} \leq 7.5264$ resulting in $n = 57$ award A0	
Total 8		

Question Number	Scheme	Marks
7(a)	$z = \pm 3.2905$ $\sigma = \frac{30}{3.2905}$ $\sigma = 9.117 **$	B1 M1 A1cso (3)
(b)	$H_0 : \mu = 1000 \quad H_1 : \mu < 1000$ <p>mean weight = 999.54</p> $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(999.54 - 1000)}{\frac{9.117}{\sqrt{10}}} = -0.160 \quad \text{or} \quad \frac{c - 1000}{\sqrt{\frac{83.12}{10}}} = -2.3263 \therefore \text{CR } c < 993.29$ <p>1% critical value = - 2.3263</p> <p>- 2.3263 < -0.160</p> <p>Accept H_0/ not in critical region</p> <p>There is no evidence that that the machine is delivering packets of mean weight less than 1 kg</p>	B1 B1 M1A1 B1 dM1 A1ft (7) Total 10
Notes		
(a)	M1 for 30/‘their $ z $ ’, >1 A1 cso as given answer	
(b)	1 st B1 both hypotheses correct. Accept 1kg in hypotheses if consistent units used in working usually either kg or g. 2 nd B1 999.54 (g) or 0.99954 (kg) 1 st M1 for standardising using their mean allow \pm , 1000 and $\frac{9.117}{\sqrt{10}}$ o.e. in kg 1 st A1 awrt -0.160 unless clearly using $ z $ (stated) then accept 0.160 or CR awrt 993 Condone -0.16 if fully correct expression seen. 3 rd B1 ± 2.3263 sign consistent with test statistic or $p = 0.4364 > 0.01$ NB $p = 0.5636 < 0.99$ 2 nd dM1 dependent upon 1 st M for a correct statement linking their test statistic and their cv Contradictory statements score M0 e.g. “significant, do not reject H_0 ” 2 nd A1 for correct conclusion in context. Must mention ‘machine’ and ‘packets’.	

Question Number	Scheme	Marks																																								
8(a)	$r = \frac{9.3433}{\sqrt{0.0632 \times 1957.5556}}$ $= 0.840$	M1 A1 (2)																																								
(b)	$H_0 : \rho = 0 \quad H_1 : \rho > 0$ Critical value = 0.5822 $0.840 > 0.5822$ There is evidence to reject H_0 . There is evidence of a positive correlation between a man's height and his weight.	B1 B1 M1 A1ft (4)																																								
(c)	<table border="1" data-bbox="354 436 1263 567"> <thead> <tr> <th>Man</th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> </tr> </thead> <tbody> <tr> <td>Actual weight</td> <td>1</td> <td>2</td> <td>7</td> <td>3</td> <td>4</td> <td>5</td> <td>8</td> <td>6</td> <td>9</td> </tr> <tr> <td>Peter's order</td> <td>1</td> <td>4</td> <td>2</td> <td>6</td> <td>3</td> <td>8</td> <td>5</td> <td>9</td> <td>7</td> </tr> <tr> <td>d^2</td> <td>0</td> <td>4</td> <td>25</td> <td>9</td> <td>1</td> <td>9</td> <td>9</td> <td>9</td> <td>4</td> </tr> </tbody> </table> $\sum d^2 = 70$ $r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$ $= 1 - \frac{6 \times 70}{9(81 - 1)}$ $= 0.417$	Man	A	B	C	D	E	F	G	H	I	Actual weight	1	2	7	3	4	5	8	6	9	Peter's order	1	4	2	6	3	8	5	9	7	d^2	0	4	25	9	1	9	9	9	4	B1 B1 M1A1 dM1 A1 (6)
Man	A	B	C	D	E	F	G	H	I																																	
Actual weight	1	2	7	3	4	5	8	6	9																																	
Peter's order	1	4	2	6	3	8	5	9	7																																	
d^2	0	4	25	9	1	9	9	9	4																																	
(d)	$H_0 : \rho = 0 \quad H_1 : \rho > 0$ Critical value 0.600 $(0.417 < 0.600)$ There is insufficient evidence to reject H_0 . Peter does not have the ability to correctly order men, by weight, from their photograph.	B1 B1 M1 A1 (4) Total 16																																								

Notes	
(a)	<p>M1 Clear use of $r = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$</p> <p>A1 0.840 cao</p>
(b)	<p>1st B1 for both hypotheses in terms of ρ, one tail H_1 must be compatible with their r</p> <p>Hypotheses just in words e.g. “no correlation” score B0</p> <p>2nd B1 for 0.5822 cao</p> <p>M1 for a statement comparing ‘their r’ with ‘their cv’</p> <p>A1 for a correct contextualised comment. Must mention positive correlation, be carrying out a 1-tailed test and mention height and weight.</p> <p>Follow through their r and their cv (provided their $cv < 1$ and their $r < 1$)</p>
(c)	<p>1st B1 for attempt to rank actual weight / Peter’s order with at least 4 correct</p> <p>2nd B1 for correct rankings for both (one or both may be reversed)</p> <p>1st M1 for use of $\sum d^2$ with at least 4 values correct and attempt to add</p> <p>1st A1 for 70 or 170 with reversed rankings</p> <p>2nd dM1 for use of the correct formula, follow through their $\sum d^2$. Dependent on 1st M1</p> <p>If answer is not correct, a correct expression is required.</p> <p>2nd A1 for awrt 0.417 or $\frac{5}{12}$</p>
(d)	<p>1st B1 for both hypotheses in terms of ρ or ρ_s. One tail H_1 must be compatible with their ranking</p> <p>Hypotheses just in words e.g. “no correlation” score B0</p> <p>2nd B1 for cv of 0.6(00) cao</p> <p>Their cv must be compatible with their H_1 which may be in words</p> <p>M1 for statement comparing ‘their r’ with ‘their cv’</p> <p>A1 for a correct contextualised comment. Must mention Peter and Men.</p> <p>Follow through their r and their cv (provided their $cv < 1$ and their $r_s < 1$)</p>



Mark Scheme (Final)

Summer 2015

Pearson Edexcel International A Level in
Statistics 3 (WST03/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2015

Publications Code IA042726

All the material in this publication is copyright

© Pearson Education Ltd 2015

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL IAL MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
 - ft – follow through
 - the symbol \surd will be used for correct ft
 - cao – correct answer only
 - cso - correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - d... or dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper or ag- answer given
 - \square or d... The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
6. If a candidate makes more than one attempt at any question:
 - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
 - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.
7. Ignore wrong working or incorrect statements following a correct answer.

**June 2015
WST03 Statistics 3 Mark Scheme**

Question Number	Scheme	Marks
1. (a)	$\{w\} = 018$ or 18	018 or 18
(b)	$\{x\} = 18$	18
(c)	$\{\text{prob} =\} 0$	0
(d)	<p>Advantage: Any one of:</p> <ul style="list-style-type: none"> • <u>Simple</u> or <u>easy</u> to use also allow “quick” or “efficient” (o.e.) • It is suitable for large samples (or populations) • Gives a good spread of the data <p>Disadvantage: Any one of:</p> <ul style="list-style-type: none"> • The alphabetical list is (probably) <u>not random</u> • <u>Biased</u> since the list is not (truly) random • <u>Some combinations</u> of names are <u>not possible</u> 	<p>B1</p> <p>[1]</p> <p>B1</p> <p>[1]</p> <p>B1</p> <p>[1]</p> <p>B1</p> <p>B1</p> <p>[2]</p> <p>(Total 5)</p>
Notes		
(d)	<p>If no labels are given treat the 1st reason as an advantage and the 2nd as a disadvantage</p> <p>1st B1: for advantage</p> <p>2nd B1: for disadvantage</p> <p>“it requires a sampling frame” is 2nd B0 since the alphabetical list is given.</p> <p>Note: Do not score both B1 marks for opposing advantages and disadvantages.</p>	

Question Number	Scheme										Marks																																																									
<p>2. (a)</p> <p>Judge 1</p> <p>Judge 2</p> <p>or</p> <p>Judge 1</p> <p>Judge 2</p> <p>$\sum d^2 = 4 + 1 + 1 + 4 + 1 + 1 + 0 + 9 + 1$</p> <p>or $0 + 1 + 1 + 1 + 9 + 4 + 1 + 1 + 4 = 22$</p> <p>$r_s = 1 - \frac{6(22)}{9(80)}; = 0.8166666...$</p> <p>(b)</p> <p>$H_0 : \rho = 0, H_1 : \rho > 0$</p> <p>Critical Value = 0.7833 <u>or</u> CR: $r_s \geq 0.7833$</p> <p>Since $r_s = 0.8166...$ it lies in the CR, <u>or</u> reject H_0 (o.e.)</p> <p>The two <u>judges</u> (or “they”) are in <u>agreement</u> <u>or</u> there is a <u>positive correlation</u> between the ranks of the two <u>judges</u>.</p>	<table style="width:100%; border-collapse: collapse;"> <tr> <td style="text-align: center;"><i>A</i></td> <td style="text-align: center;"><i>B</i></td> <td style="text-align: center;"><i>C</i></td> <td style="text-align: center;"><i>L</i></td> <td style="text-align: center;"><i>N</i></td> <td style="text-align: center;"><i>R</i></td> <td style="text-align: center;"><i>S</i></td> <td style="text-align: center;"><i>T</i></td> <td style="text-align: center;"><i>Y</i></td> <td></td> <td></td> </tr> <tr> <td style="text-align: center;">6</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">9</td> <td style="text-align: center;">2</td> <td style="text-align: center;">8</td> <td style="text-align: center;">1</td> <td style="text-align: center;">5</td> <td style="text-align: center;">7</td> <td></td> <td></td> </tr> <tr> <td style="text-align: center;">8</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> <td style="text-align: center;">7</td> <td style="text-align: center;">3</td> <td style="text-align: center;">9</td> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">6</td> <td></td> <td></td> </tr> <tr> <td></td> <td style="text-align: center;"><i>S</i></td> <td style="text-align: center;"><i>N</i></td> <td style="text-align: center;"><i>B</i></td> <td style="text-align: center;"><i>C</i></td> <td style="text-align: center;"><i>T</i></td> <td style="text-align: center;"><i>A</i></td> <td style="text-align: center;"><i>Y</i></td> <td style="text-align: center;"><i>R</i></td> <td style="text-align: center;"><i>L</i></td> <td></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> <td style="text-align: center;">6</td> <td style="text-align: center;">7</td> <td style="text-align: center;">8</td> <td style="text-align: center;">9</td> <td></td> <td></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> <td style="text-align: center;">2</td> <td style="text-align: center;">8</td> <td style="text-align: center;">6</td> <td style="text-align: center;">9</td> <td style="text-align: center;">7</td> <td></td> <td></td> </tr> </table> <p style="text-align: right;">$\sum d^2 = 22$</p> <p style="text-align: right;">$\frac{49}{60}$ or awrt 0.817</p> <p style="text-align: right;">0.7833</p>	<i>A</i>	<i>B</i>	<i>C</i>	<i>L</i>	<i>N</i>	<i>R</i>	<i>S</i>	<i>T</i>	<i>Y</i>			6	3	4	9	2	8	1	5	7			8	4	5	7	3	9	1	2	6				<i>S</i>	<i>N</i>	<i>B</i>	<i>C</i>	<i>T</i>	<i>A</i>	<i>Y</i>	<i>R</i>	<i>L</i>		1	2	3	4	5	6	7	8	9			1	3	4	5	2	8	6	9	7			<p>M1</p> <p>M1</p> <p>A1</p> <p>M1;</p> <p>A1</p> <p>[5]</p> <p>B1</p> <p>B1</p> <p>M1</p> <p>A1ft</p> <p>[4]</p> <p>(Total 9)</p>
<i>A</i>	<i>B</i>	<i>C</i>	<i>L</i>	<i>N</i>	<i>R</i>	<i>S</i>	<i>T</i>	<i>Y</i>																																																												
6	3	4	9	2	8	1	5	7																																																												
8	4	5	7	3	9	1	2	6																																																												
	<i>S</i>	<i>N</i>	<i>B</i>	<i>C</i>	<i>T</i>	<i>A</i>	<i>Y</i>	<i>R</i>	<i>L</i>																																																											
1	2	3	4	5	6	7	8	9																																																												
1	3	4	5	2	8	6	9	7																																																												
Notes																																																																				
<p>(a)</p> <p>False Ranking</p> <p>(b)</p> <p> cv >1</p>	<p>1st M1 for an attempt to rank at least one row (at least 4 correct)</p> <p>2nd M1 for an attempt at d^2 row (may be implied by sight of $\sum d^2 = 22$ or 221 for reverse ranks)</p> <p>1st A1 for $\sum d^2 = 22$ (or 221 if reverse ranking is used) Can be implied by correct answer.</p> <p>3rd M1 for use of the correct formula with their $\sum d^2$ (if it is clearly stated)</p> <p style="text-align: center;">If the answer is not correct then a correct expression is required</p> <p>e.g Alphabetic ranking: Gives Judge 1: 7 5 2 3 8 1 9 6 4</p> <p style="text-align: right;">Judge 2: 7 8 5 2 3 9 4 1 6 $\sum d^2 = 162$ and $r_s = -0.35$</p> <p>Scores: M0(for ranking), M1(for attempt at d^2 row), A0, M1 (for use of their $\sum d^2$), A0 i.e. 2 out of 5</p> <p style="text-align: center;">Can follow through their r_s in (b)</p> <p>1st B1 for both hypotheses stated correctly in terms of ρ (allow ρ_s) H_1 must be compatible with ranking</p> <p>2nd B1 for cv = 0.7833 (independent of their H_1 (no 2-tail value in tables) <u>but</u> compatible sign with their r_s)</p> <p>M1: for a correct statement (in words) relating their r_s with their critical value.</p> <p style="text-align: center;">e.g. “reject H_0”, “in critical region”, “significant”, “positive correlation”</p> <p>May be implied by a correct contextual comment.</p> <p>If their cv is $cv > 1$ (often from using normal tables) award M0A0</p> <p>If $their\ r_s > their\ cv$ then “significant” (o.e.) for M1 and “<u>judges are in agreement</u>” (o.e.) for A1ft</p> <p>If $their\ r_s < their\ cv$ then “not significant” (o.e.) for M1 and “<u>judges don’t agree</u>” (o.e.) for A1ft</p> <p>A1ft: for a correct follow through conclusion in context.</p> <p style="text-align: center;">“positive correlation” alone scores M1 A0</p> <p style="text-align: center;">For reverse ranking should still say “<u>judges are in agreement</u>”</p>																																																																			

Question Number	Scheme	Marks																																																				
<p>3. (a)</p> <p>(b)</p> <p>(c)</p>	$\hat{\lambda} = \frac{0(47) + 1(57) + 2(46) + 3(35) + 4(9) + 5(6)}{200} = \frac{320}{200} = 1.6$ <p>Full exp' or at least 2 products and 320/200 seen</p> <p>Using $r = 200 \times \frac{e^{-1.6}(1.6)^2}{2!}$</p> <p>$s = 200 - (40.38 + 64.61 + \text{their } r + 27.57 + 11.03) \{= 4.72449139...\}$ <u>or</u> their $r + s = 56.41$</p> <p>$r = 51.68550861...$ and $s = 4.72449139...$ $r = \text{awrt } \mathbf{51.69}$ and $s = \text{awrt } \mathbf{4.72}$</p> <p>$H_0$: Poisson (distribution) is a suitable/ sensible (model) H_1 : Poisson (distribution) is not a suitable/ sensible (model).</p> <table border="1" data-bbox="236 689 1359 1025"> <thead> <tr> <th>Number of accidents</th> <th>Observed</th> <th>Expected</th> <th>Combined Observed</th> <th>Combined Expected</th> <th>$\frac{(O - E)^2}{E}$</th> <th>$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>47</td> <td>40.38</td> <td>47</td> <td>40.38</td> <td>1.0853</td> <td>54.7053</td> </tr> <tr> <td>1</td> <td>57</td> <td>64.61</td> <td>57</td> <td>64.61</td> <td>0.8963</td> <td>50.2863</td> </tr> <tr> <td>2</td> <td>46</td> <td>51.69</td> <td>46</td> <td>51.69</td> <td>0.6264</td> <td>40.9364</td> </tr> <tr> <td>3</td> <td>35</td> <td>27.57</td> <td>35</td> <td>27.57</td> <td>2.0024</td> <td>44.4324</td> </tr> <tr> <td>4</td> <td>9</td> <td>11.03</td> <td rowspan="2">15</td> <td rowspan="2">15.75</td> <td rowspan="2">0.0357</td> <td rowspan="2">14.2857</td> </tr> <tr> <td>≥ 5</td> <td>6</td> <td>4.72</td> </tr> <tr> <td colspan="5" style="text-align: right;">Totals</td> <td>4.6461</td> <td>204.6461</td> </tr> </tbody> </table> $X^2 = \sum \frac{(O - E)^2}{E} \text{ or } \sum \frac{O^2}{E} - 200 := 4.6461$ <p>$v = 5 - 1 - 1 = 3$ awrt 4.65</p> <p>$\chi^2_3(0.10) = 6.251 \Rightarrow \text{CR: } X^2 \geq 6.251$ 3</p> <p>[Since $X^2 = 4.6461$ does not lie in the CR, then there is insufficient evidence to reject H_0]</p> <p>The number of accidents per day can be modelled by a Poisson distribution <u>or</u> the supervisor's belief is correct. 6.251</p>	Number of accidents	Observed	Expected	Combined Observed	Combined Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	0	47	40.38	47	40.38	1.0853	54.7053	1	57	64.61	57	64.61	0.8963	50.2863	2	46	51.69	46	51.69	0.6264	40.9364	3	35	27.57	35	27.57	2.0024	44.4324	4	9	11.03	15	15.75	0.0357	14.2857	≥ 5	6	4.72	Totals					4.6461	204.6461	<p>B1 * [1]</p> <p>M1</p> <p>M1</p> <p>A1 [3]</p> <p>B1</p> <p>M1</p> <p>M1; A1</p> <p>B1 ft</p> <p>B1 ft</p> <p>A1 ft</p> <p>[7] (Total 11)</p>
Number of accidents	Observed	Expected	Combined Observed	Combined Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																																																
0	47	40.38	47	40.38	1.0853	54.7053																																																
1	57	64.61	57	64.61	0.8963	50.2863																																																
2	46	51.69	46	51.69	0.6264	40.9364																																																
3	35	27.57	35	27.57	2.0024	44.4324																																																
4	9	11.03	15	15.75	0.0357	14.2857																																																
≥ 5	6	4.72																																																				
Totals					4.6461	204.6461																																																
Notes																																																						
<p>(b)</p> <p>(c)</p> <p>No pooling</p>	<p>Note: Allow A1 for $s = \text{awrt } 4.74$ (found as a result of using expected values to full accuracy.)</p> <p>1st B1: for <u>both</u> hypotheses and mentioning Poisson at least once. Allow Poisson is a "good fit/model" but <u>not</u> "good method" Inclusion of 1.6 for mean in hypotheses is B0 but condone in conclusion.</p> <p>1st M1: For an attempt to pool 4 accidents and ≥ 5 accidents <u>or</u> pool when $E_i < 5$ No pooling is M0</p> <p>2nd M1: For an attempt at the test statistic, at least 2 correct expressions/values (to awrt 2 d.p.)</p> <p>1st A1: For awrt 4.65 (score M1M1A1 if awrt 4.65 seen) If no pooling can allow 2nd M1 if $X^2 = 5.33$ is seen</p> <p>2nd B1ft: For $n - 1 - 1$ i.e. subtracting 2 from their n. B1B1 may be implied by 6.251 (if pooling) or 7.779 for no pooling</p> <p>3rd B1ft: For a correct ft for their $\chi^2_k(0.10)$, where $k = n - 1 - 1$ from their n.</p> <p>2nd A1ft: (Dep. on the 2nd M1) For correct comment in context based on their test statistic and their critical value that mentions accidents or supervisor. Condone mention of Po(1.6) in conclusion Score A0 for inconsistencies e.g. "significant" followed by "supervisor's belief is justified"</p> <p>Note: Full accuracy gives a combined expected frequency of 15.76, $\frac{(O - E)^2}{E} = 0.0366$, $\frac{O^2}{E} = 14.2766$, $X^2 = 4.64855...$ and p-value 0.199</p>																																																					

Question Number	Scheme	Marks
<p>4. (a)</p> <p>(b)</p>	<p>Let $X =$ weight of a sack of potatoes, $X \sim N(25.6, 0.24^2)$</p> <p>So $D = X_1 - X_2 \sim N(0, 2(0.24)^2)$ or $D \sim N(0, 0.1152)$</p> <p>$\{P(D > 0.5) = \}$ $2P(D > 0.5)$</p> $= 2 \times P\left(Z > \frac{0.5}{\sqrt{0.1152}}\right)$ $= 2 \times P(Z > 1.4731\dots)$ <u>or</u> $= 2(1 - 0.9292)$ $= 0.1416$ <p>Let $Y =$ weight of an empty pallet, $Y \sim N(20.0, 0.32^2)$</p> <p>So $T = X_1 + X_2 + \dots + X_{30} + Y$</p> <p>$T \sim N(30(25.6) + 20, 30(0.24)^2 + 0.32^2)$</p> <p>$T \sim N(788, 1.8304)$</p> <p>$\{P(T > 785) = \}$ $P\left(Z > \frac{785 - 788}{\sqrt{1.8304}}\right)$</p> $= P(Z > -2.2174\dots)$ $= 0.9868$	<p>Attempt at D and $D \sim N(0, \dots)$ $(0.24)^2 + (0.24)^2$; 0.1152 $2 \times P(D > 0.5)$ can be implied</p> <p>M1 A1; A1 dM1 dM1 awrt 0.141 or awrt 0.142</p> <p>B1 M1 A1 M1 awrt 0.987</p> <p>[6] [5] (Total 11)</p>
Notes		
(a)	<p>1st M1: For clear definition of D and normal distribution with mean of 0 (Can be implied by 3rd M1)</p> <p>1st A1: for correct use of $\text{Var}(X_1 - X_2)$ formula</p> <p>2nd A1: for 0.1152</p> <p>2nd dM1: For realising need $2 \times P(D > 0.5)$ (Dependent on 1st M1 i.e. must be using suitable D)</p> <p>3rd dM1: Dep on 1st M1 for standardising with 0.5, 0 and their s.d. ($\neq 0.24$) Must lead to $P(Z > +ve)$ (o.e.) $P(Z > 1.47)$ implies 1st M1 1st A1 2nd A1 and 3rd M1 Correct answer only will score 6 out of 6</p>	
(b)	<p>B1: For a mean of $30(25.6) + 20$. Can be implied by 788.</p> <p>1st M1: For $30(0.24)^2 + 0.32^2$. Can be implied by 1.8304 or awrt 1.83</p> <p>Allow M1 for swapping error i.e. $30 \times 0.32^2 + 0.24^2$ if the expression is seen</p> <p>1st A1: For normal and correct variance of 1.8304 or awrt 1.83. Normality may be implied by standardisation</p> <p>2nd M1: For standardising with 785 with their mean and st. dev. ($\neq 0.24$) Must lead to $P(Z > -ve)$ oe.</p> <p>2nd A1: awrt 0.987 Correct answer only will score 5 out of 5</p> <p>Note: Calculator answers are (a) 0.14071..., (b) 0.98670...</p>	

Question Number	Scheme		Marks																															
<p>5.</p>	H_0 : Grades and gender are independent (or not associated) H_1 : Grades and gender are dependent (or associated)		B1 (1)																															
	<table border="1"> <thead> <tr> <th>Observed</th> <th>Male</th> <th>Female</th> </tr> </thead> <tbody> <tr> <td>Distinction</td> <td>37</td> <td>44</td> </tr> <tr> <td>Merit</td> <td>127</td> <td>96</td> </tr> <tr> <td>Unsatisfactory</td> <td>36</td> <td>20</td> </tr> </tbody> </table>		Observed	Male	Female	Distinction	37	44	Merit	127	96	Unsatisfactory	36	20	An attempt to convert percentages to observed frequencies. M1 All observed frequencies are correct. A1																			
	Observed	Male	Female																															
	Distinction	37	44																															
	Merit	127	96																															
	Unsatisfactory	36	20																															
	<table border="1"> <thead> <tr> <th>Expected</th> <th>Male</th> <th>Female</th> <th>Totals</th> </tr> </thead> <tbody> <tr> <td>Distinction</td> <td>45</td> <td>36</td> <td>81</td> </tr> <tr> <td>Merit</td> <td>123.889</td> <td>99.111</td> <td>223</td> </tr> <tr> <td>Unsatisfactory</td> <td>31.111</td> <td>24.889</td> <td>56</td> </tr> <tr> <td>Totals</td> <td>200</td> <td>160</td> <td>360</td> </tr> </tbody> </table>		Expected	Male	Female	Totals	Distinction	45	36	81	Merit	123.889	99.111	223	Unsatisfactory	31.111	24.889	56	Totals	200	160	360	Some attempt at (Row Total)(Column Total) (Grand Total) M1 Can be implied by a correct E_i A1 All expected frequencies are correct to nearest integer.											
	Expected	Male	Female	Totals																														
	Distinction	45	36	81																														
	Merit	123.889	99.111	223																														
	Unsatisfactory	31.111	24.889	56																														
	Totals	200	160	360																														
<table border="1"> <thead> <tr> <th>Observed</th> <th>Expected</th> <th>$\frac{(O - E)^2}{E}$</th> <th>$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>37</td> <td>45</td> <td>1.422</td> <td>30.422</td> </tr> <tr> <td>44</td> <td>36</td> <td>1.778</td> <td>53.778</td> </tr> <tr> <td>127</td> <td>123.889</td> <td>0.078</td> <td>130.189</td> </tr> <tr> <td>96</td> <td>99.111</td> <td>0.098</td> <td>92.987</td> </tr> <tr> <td>36</td> <td>31.111</td> <td>0.768</td> <td>41.657</td> </tr> <tr> <td>20</td> <td>24.889</td> <td>0.960</td> <td>16.071</td> </tr> <tr> <td colspan="2">Totals</td> <td>5.104</td> <td>365.104</td> </tr> </tbody> </table>		Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	37	45	1.422	30.422	44	36	1.778	53.778	127	123.889	0.078	130.189	96	99.111	0.098	92.987	36	31.111	0.768	41.657	20	24.889	0.960	16.071	Totals		5.104	365.104	At least 2 correct terms for $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i . M1 Accept 2 sf accuracy for the M1 mark. A1 All correct $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ terms to either 2 dp or better. Allow truncation. (\Rightarrow by awrt 5.1 if 3 rd M1 seen)
Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																															
37	45	1.422	30.422																															
44	36	1.778	53.778																															
127	123.889	0.078	130.189																															
96	99.111	0.098	92.987																															
36	31.111	0.768	41.657																															
20	24.889	0.960	16.071																															
Totals		5.104	365.104																															
$X^2 = \sum \frac{(O - E)^2}{E} \text{ or } \sum \frac{O^2}{E} - 360 ; = \text{awrt } 5.1$		awrt 5.1 A1 (7)																																
$\nu = (3 - 1)(2 - 1) = 2$		($\nu =$) 2 (Can be implied by 5.991) B1																																
$\chi^2_2(0.05) = 5.991 \Rightarrow \text{CR: } X^2 \geq 5.991$		For 5.991 only B1																																
Since $X^2 = 5.1$ does not lie in the CR, then there is insufficient evidence to reject H_0		M1																																
Business Studies <u>grades</u> and <u>gender</u> are independent <u>or</u> There is no association between Business Studies <u>grades</u> and <u>gender</u> . <u>Or</u> <u>Head of department's</u> (belief) is correct		A1ft (4)																																
<p style="text-align: right;">[12] (Total 12)</p>																																		
Notes																																		
<p>5.10 only</p>	Final M1: For a correct statement linking their test statistic and their critical value (> 3.8) Note: Contradictory statements score M0. E.g. "significant, do not reject H_0 ". Final A1ft: For a correct ft statement in context – must mention "grades" and "gender" or "sex" <u>or</u> "head of department" Condone "relationship" or "connection" here but not "correlation". e.g. "There is no evidence of a relationship between grades and gender" Just seeing 5.10... only can imply 1 st 3 Ms but loses 1 st 3 As so can score 4 out of 7 (Qu says "show..") Note: Full accuracy gives $X^2 = 5.104356...$ and p-value 0.0779																																	

Question Number	Scheme	Marks																																																																
5.	<p>Mark Scheme for candidates who use percentages instead of observed values.</p> <p>H_0 : Grades and gender are independent (or not associated) “grades” and “gender” mentioned at least once.</p> <p>H_1 : Grades and gender are dependent (or associated)</p> <table border="1" style="width: 100%; margin: 10px 0;"> <thead> <tr> <th>Observed</th> <th>Male</th> <th>Female</th> </tr> </thead> <tbody> <tr> <td>Distinction</td> <td>18.5</td> <td>27.5</td> </tr> <tr> <td>Merit</td> <td>63.5</td> <td>60.0</td> </tr> <tr> <td>Unsatisfactory</td> <td>18.0</td> <td>12.5</td> </tr> </tbody> </table> <table border="1" style="width: 100%; margin: 10px 0;"> <thead> <tr> <th>Expected</th> <th>Male</th> <th>Female</th> <th>Totals</th> </tr> </thead> <tbody> <tr> <td>Distinction</td> <td>23</td> <td>23</td> <td>46</td> </tr> <tr> <td>Merit</td> <td>61.75</td> <td>61.75</td> <td>123.5</td> </tr> <tr> <td>Unsatisfactory</td> <td>15.25</td> <td>15.25</td> <td>30.5</td> </tr> <tr> <td>Totals</td> <td>100</td> <td>100</td> <td>200</td> </tr> </tbody> </table> <table border="1" style="width: 100%; margin: 10px 0;"> <thead> <tr> <th>Observed</th> <th>Expected</th> <th>$\frac{(O - E)^2}{E}$</th> <th>$\frac{O^2}{E}$</th> </tr> </thead> <tbody> <tr> <td>18.5</td> <td>23</td> <td>0.8804</td> <td>14.8804</td> </tr> <tr> <td>27.5</td> <td>23</td> <td>0.8804</td> <td>32.8804</td> </tr> <tr> <td>63.5</td> <td>61.75</td> <td>0.0496</td> <td>65.2996</td> </tr> <tr> <td>60.0</td> <td>61.75</td> <td>0.0496</td> <td>58.2996</td> </tr> <tr> <td>18.0</td> <td>15.25</td> <td>0.4959</td> <td>21.2459</td> </tr> <tr> <td>12.5</td> <td>15.25</td> <td>0.4959</td> <td>10.2459</td> </tr> <tr> <td colspan="2" style="text-align: center;">Totals</td> <td>2.8518</td> <td>202.8518</td> </tr> </tbody> </table> <p>$X^2 = \sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - \frac{360}{200} = 2.8518$</p> <p>$\nu = (3 - 1)(2 - 1) = 2$</p> <p>$\chi^2_2(0.05) = 5.991 \Rightarrow \text{CR: } X^2 \geq 5.991$</p> <p>Since $X^2 = 2.86$ does not lie in the CR, then there is insufficient evidence to reject H_0</p>	Observed	Male	Female	Distinction	18.5	27.5	Merit	63.5	60.0	Unsatisfactory	18.0	12.5	Expected	Male	Female	Totals	Distinction	23	23	46	Merit	61.75	61.75	123.5	Unsatisfactory	15.25	15.25	30.5	Totals	100	100	200	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	18.5	23	0.8804	14.8804	27.5	23	0.8804	32.8804	63.5	61.75	0.0496	65.2996	60.0	61.75	0.0496	58.2996	18.0	15.25	0.4959	21.2459	12.5	15.25	0.4959	10.2459	Totals		2.8518	202.8518	<p>B1</p> <p>These marks cannot be obtained. M0 A0</p> <p>Some attempt at $\frac{(\text{Row Total})(\text{Column Total})}{(\text{Grand Total})}$ M1</p> <p>Can be implied by one of these E_i's</p> <p>Expected frequencies are not correct. A0</p> <p>At least 2 “correct” terms for $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i. M1</p> <p>Accept 2 sf accuracy for the M1 mark.</p> <p>This mark cannot be obtained. A0</p> <p>This mark cannot be obtained. A0</p> <p>$(\nu =) 2$ (Can be implied by 5.991) B1</p> <p>For 5.991 only B1</p> <p>M1</p> <p>Not available since comes from incorrect working. A0</p> <p style="text-align: right;">[12] (Total 12)</p>
Observed	Male	Female																																																																
Distinction	18.5	27.5																																																																
Merit	63.5	60.0																																																																
Unsatisfactory	18.0	12.5																																																																
Expected	Male	Female	Totals																																																															
Distinction	23	23	46																																																															
Merit	61.75	61.75	123.5																																																															
Unsatisfactory	15.25	15.25	30.5																																																															
Totals	100	100	200																																																															
Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																																																															
18.5	23	0.8804	14.8804																																																															
27.5	23	0.8804	32.8804																																																															
63.5	61.75	0.0496	65.2996																																																															
60.0	61.75	0.0496	58.2996																																																															
18.0	15.25	0.4959	21.2459																																																															
12.5	15.25	0.4959	10.2459																																																															
Totals		2.8518	202.8518																																																															
Notes																																																																		
If a candidate uses percentages rather than observed values then they can obtain a maximum of 6 marks . They can get B1 M0A0 M1A0 M1A0A0 B1B1M1A0.																																																																		

Question Number	Scheme	Marks
6. (a)	$\left\{ \hat{\mu} = \frac{\sum x}{n} = \frac{1570}{50} = \right\} \bar{x} = 31.4$ $\left\{ \hat{\sigma}^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \right\} s_x^2 = \frac{49467.58 - 50(31.4)^2}{50 - 1}$ $= 3.460816...$	$\bar{x} = \mathbf{31.4}$ B1 cao M1 A1ft awrt 3.46 A1 [4]
(b)	[Let $Y =$ time taken to complete obstacle course in the afternoon.] $H_0: \mu_x = \mu_y, H_1: \mu_x > \mu_y$ $(z =) \frac{31.4 - 30.9}{\sqrt{\frac{3.46}{50} + \frac{3.03}{50}}}$ $= 1.38781...$ CR: $Z \geq 1.6449$ or probability = awrt 0.082 or awrt 0.083 Since $z = 1.38781...$ does not lie in the CR, then there is insufficient evidence to reject H_0 Conclude that the <u>mean time</u> to complete the obstacle course is the same for the early <u>morning</u> and late <u>afternoon</u> .	B1 M1 A1ft awrt 1.39 A1 1.6449 or better B1 M1 A1 [7]
(c)	\bar{X} and \bar{Y} are both approx. normally distributed or $\bar{X} - \bar{Y}$ normal (Condone \bar{x} and \bar{y})	B1 [1]
(d)	Have assumed $s^2 \approx \sigma^2$ or variance of sample \approx variance of population	B1 [1]
(Total 13)		

Notes

(a)	B1: 31.4 cao Allow 31 minutes, 24 seconds. 1 st M1: A correct expression for either s or s^2 (ignore label) 1 st A1ft: A correct expression for s^2 with their ft \bar{x} . 3 rd A1: awrt 3.46 (Correct answer scores 3 out of 3)
(b)	1 st B1: Both hypotheses stated correctly, with some indication of which μ is which. Eg: μ_M, μ_A 1 st M1: For an attempt at $\frac{a-b}{\sqrt{\frac{c}{50} + \frac{d}{50}}}$ with at least 3 of a, b, c or d correct. Allow \pm 1 st A1ft: for $\pm \frac{\text{their } 31.4 - 30.9}{\sqrt{\frac{\text{their } 3.46}{50} + \frac{3.03}{50}}}$ Allow $D = \bar{x} - \bar{y}$ $1.64 \sim 1.65 = \frac{D - 0}{\sqrt{\frac{3.46}{50} + \frac{3.03}{50}}}$ [SE = 0.360277..] 2 nd A1: for awrt 1.39 (possibly \pm) (Allow for CV $D =$ awrt 0.593) (NB $d = 0.5$) Correct answer scores M1A1ftA1 <u>but</u> $0 - (31.4 - 30.9) \rightarrow -1.39$ loses this 2 nd A mark 2 nd B1: Critical value of 1.6449 or better (seen). Allow for probability = awrt 0.082 or awrt 0.083 Note: p-values are 0.0823 (tables) and 0.0826 (calculator). 2 nd M1: For a correct statement linking their test statistic and their critical value. Note: Contradictory statements score M0. E.g. "significant, do not reject H_0 ". 3 rd A1: For a correct statement in context that accepts H_0 (no ft) Condone "no difference in mean times" Must mention " <u>mean time</u> ", " <u>morning</u> " and " <u>afternoon</u> " or " <u>both times of day</u> "
(c)	B1 E.g. $\bar{X} \sim N(...)$ need both. Allow in words e.g "sample means are normally distributed"
(d)	B1 condone only mentioning "x" or "y" <u>but</u> watch out for $s_x = s_y$ or $\sigma_x = \sigma_y$ which scores B0

Question Number	Scheme	Marks
<p>7.</p> <p>(a)</p> <p>(b)</p>	<p>Let $X =$ score on a die</p> <p>$E(S) = 3.5$, $\text{Var}(S) = \frac{35}{12}$</p> <p>So, $\bar{S} \sim N\left(3.5, \frac{\left(\frac{35}{12}\right)}{40}\right)$ or $\bar{S} \sim N\left(3.5, \frac{7}{96}\right)$</p> <p>$P(\bar{S} < 3) = P\left(Z < \frac{3 - 3.5}{\sqrt{\frac{7}{96}}}\right) \{= P(Z < -1.85164...)\}$</p> <p>$\{= 1 - 0.9678\} = 0.0322$</p>	<p>$E(S) = 3.5$ B1</p> <p>$\text{Var}(S) = \frac{35}{12}$ or awrt 2.92 B1</p> <p>[2]</p> <p>B1ft</p> <p>M1</p> <p>0.032 to 0.0322 A1</p> <p>[3]</p> <p>(Total 5)</p>
Notes		
(a)	<p>2nd B1 allow awrt 2.92</p>	
ALT ES	<p>(b) B1ft for $\bar{S} \sim N\left(3.5, \frac{\left(\frac{35}{12}\right)}{40}\right)$ seen or implied. Follow through their $E(S)$ and their $\text{Var}(S)$</p> <p>NB $\frac{7}{96} = 0.07291\dot{6}$ accept awrt 0.0729</p> <p>M1 for an attempt to standardise with 3, their mean (>3) and $\sqrt{\frac{\text{their Var}(S)}{40}}$. Must lead to $P(Z < -ve)$</p> <p>A1 for 0.032 ~ 0.0322</p> <p>B1ft for $\sum S \sim N\left(140, \frac{350}{3}\right)$ where 140 is $40 \times$ their $E(S)$ and variance is $40 \times$ their $\text{Var}(S)$</p> <p>M1 for $P\left(Z < \frac{120 - 140}{\sqrt{\frac{350}{3}}}\right)$ or $P\left(Z < \frac{119.5 - 140}{\sqrt{\frac{350}{3}}}\right) \{= P(Z < -1.8979...)\}$</p> <p>A1 for 0.032~0.0322 or (with continuity correction) 0.0287 (tables) or 0.0289 (calculator).</p>	

Question Number	Scheme	Marks
<p>8. (a)</p> <p>(b)</p> <p>(c)</p>	$\left\{ \bar{x} = \frac{29.74 + 31.86}{2} \right\} \Rightarrow \bar{x} = 30.8$ <p style="text-align: right;">$\bar{x} = 30.8$</p> <p style="text-align: right;">This can be implied. See note.</p> $"1.96" \left(\frac{\sigma}{\sqrt{n}} \right) = 31.86 - 30.8 \quad \text{or} \quad 2("1.96") \left(\frac{\sigma}{\sqrt{n}} \right) = 31.86 - 29.74$ $SE_{\bar{x}} = \frac{31.86 - 30.8}{1.96} = 0.540816... = 0.54 \text{ (2dp)}$ <p style="text-align: right;">awrt 0.54</p> <p>A 90% CI for μ is $\bar{x} \pm 1.6449 \left(\frac{\sigma}{\sqrt{n}} \right)$</p> $= 30.8 \pm 1.6449(0.54)$ <p style="text-align: right;">(their \bar{x}) \pm (their z)(their $SE_{\bar{x}}$ from (a))</p> $= (29.91, 31.69)$ <p style="text-align: right;">(awrt 29.9, awrt 31.7)</p> <p>Let X = number of confidence intervals containing μ or Y = number of confidence intervals not containing μ So $X \sim \text{Bin}(4, 0.9)$ or $Y \sim \text{Bin}(4, 0.1)$</p> $P(X \geq 3) \text{ or } P(Y \leq 1) = {}^4C_3(0.9)^3(0.1) + (0.9)^4$ $= 0.2916 + 0.6561 = 0.9477$ <p style="text-align: right;">${}^4C_3(0.9)^3(0.1) + (0.9)^4$ 0.9477 or 0.948</p>	<p>B1</p> <p>M1</p> <p>A1</p> <p style="text-align: right;">[3]</p> <p>B1</p> <p>M1</p> <p>A1</p> <p style="text-align: right;">[3]</p> <p>M1</p> <p>A1 oe</p> <p>A1</p> <p style="text-align: right;">[3]</p> <p style="text-align: right;">(Total 9)</p>
Notes		
<p>(a)</p> <p>(b)</p> <p>(c)</p>	<p>B1: $\bar{x} = 30.8$ may be implied by $1.96 \left(\frac{\sigma}{\sqrt{n}} \right) = [31.86 - 30.8] = 1.06$ <u>or</u> $2(1.96) \left(\frac{\sigma}{\sqrt{n}} \right) = 31.86 - 29.74$</p> <p>M1: A correct equation for either a width or a half-width involving a z-value $1.5 \leq z \leq 2$</p> <p>Eg: "their z" $\left(\frac{\sigma}{\sqrt{n}} \right) = 31.86 - "30.8"$ ft their \bar{x} <u>or</u> $2("their z") \left(\frac{\sigma}{\sqrt{n}} \right) = 31.86 - 29.74$</p> <p>or "their z" $(SE_{\bar{x}}) = 31.86 - "30.8"$ <u>or</u> $2("their z")(SE_{\bar{x}}) = 31.86 - 29.74$ are fine for M1.</p> <p>A1: 0.54 or awrt 0.54 Must be seen as final answer to (a) NB $\frac{53}{98}$ as final answer is A0</p> <p>Condone $\bar{x} \pm 1.96\sigma = \dots$ for B1 and M1 but A0 even if they say "σ = standard error = 0.54"</p> <p>Otherwise answer only of 0.54 scores 3 out of 3</p> <p>B1 for use of 1.6449 or better in an attempt at a CI formula. Need at least 1.6449 (their SE)</p> <p>M1 for attempt at CI ft their values and provided $1 \leq z \leq 1.7$</p> <p>M1: States or applies either $X \sim \text{Bin}(4, 0.9)$ <u>or</u> $Y \sim \text{Bin}(4, 0.1)$</p> <p>Condone M1 for $0.9^4 + 0.9^3 \times 0.1$ (o.e.)</p> <p>1st A1: ${}^4C_3(0.9)^3(0.1) + (0.9)^4$ or $(0.9)^4 + {}^4C_1(0.1)(0.9)^3$ oe</p> <p>2nd A1: 0.9477 or 0.948</p>	



Mark Scheme (Results)

Summer 2015

Pearson Edexcel GCE
in Statistics 3 (6691/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2015

Publications Code UA042717

All the material in this publication is copyright

© Pearson Education Ltd 2015

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

PEARSON EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M** marks: Method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
 - **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B** marks are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.

3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
 - ft – follow through
 - the symbol \checkmark will be used for correct ft
 - cao – correct answer only
 - cso – correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - d... or dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper or ag- answer given
 - \square or d... The second mark is dependent on gaining the first mark
4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

5. For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.
6. If a candidate makes more than one attempt at any question:
 - If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
 - If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.
7. Ignore wrong working or incorrect statements following a correct answer.

June 2015
6691 S3
Mark Scheme

Question Number	Scheme	Marks																																												
1. (a)	Label all the books from 1 – 160 (o.e.) Use random numbers to select the 10 books	B1 B1 (2)																																												
(b)	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="padding: 2px;">Book</th> <th style="padding: 2px;"><i>A</i></th> <th style="padding: 2px;"><i>B</i></th> <th style="padding: 2px;"><i>C</i></th> <th style="padding: 2px;"><i>D</i></th> <th style="padding: 2px;"><i>E</i></th> <th style="padding: 2px;"><i>F</i></th> <th style="padding: 2px;"><i>G</i></th> <th style="padding: 2px;"><i>H</i></th> <th style="padding: 2px;"><i>I</i></th> <th style="padding: 2px;"><i>J</i></th> </tr> </thead> <tbody> <tr> <td style="padding: 2px;">Borrow rank</td> <td style="padding: 2px;">1</td> <td style="padding: 2px;">2</td> <td style="padding: 2px;">3</td> <td style="padding: 2px;">4</td> <td style="padding: 2px;">5</td> <td style="padding: 2px;">6</td> <td style="padding: 2px;">7</td> <td style="padding: 2px;">8</td> <td style="padding: 2px;">9</td> <td style="padding: 2px;">10</td> </tr> <tr> <td style="padding: 2px;">Page rank</td> <td style="padding: 2px;">1</td> <td style="padding: 2px;">6</td> <td style="padding: 2px;">4</td> <td style="padding: 2px;">2</td> <td style="padding: 2px;">8</td> <td style="padding: 2px;">3</td> <td style="padding: 2px;">10</td> <td style="padding: 2px;">7</td> <td style="padding: 2px;">5</td> <td style="padding: 2px;">9</td> </tr> <tr> <td style="padding: 2px;">d^2</td> <td style="padding: 2px;">0</td> <td style="padding: 2px;">16</td> <td style="padding: 2px;">1</td> <td style="padding: 2px;">4</td> <td style="padding: 2px;">9</td> <td style="padding: 2px;">9</td> <td style="padding: 2px;">9</td> <td style="padding: 2px;">1</td> <td style="padding: 2px;">16</td> <td style="padding: 2px;">1</td> </tr> </tbody> </table> <p style="margin-top: 10px;">$r_s = 1 - \frac{6 \times 66}{10(100-1)}, [= 1 - 0.4] = 0.6$ <u>0.6</u></p>	Book	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	Borrow rank	1	2	3	4	5	6	7	8	9	10	Page rank	1	6	4	2	8	3	10	7	5	9	d^2	0	16	1	4	9	9	9	1	16	1	M1 M1 M1,A1 (4)
Book	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>																																				
Borrow rank	1	2	3	4	5	6	7	8	9	10																																				
Page rank	1	6	4	2	8	3	10	7	5	9																																				
d^2	0	16	1	4	9	9	9	1	16	1																																				
(c)	<p>$H_0: \rho = 0 \quad H_1: \rho > 0$ Critical value is 0.5636 $0.6 > cv$ so significant result and sufficient evidence to reject H_0 There is support for the librarian's belief <u>or</u> there is evidence of a correlation between the number of pages in a book and the number of times it is borrowed.</p>	B1 B1 B1ft (3)																																												
Notes		Total 9																																												
(a)	<p>1st B1 for labelling\numbering\listing\using sampling frame of all 160 books 2nd B1 for use of random numbers\selection and mentioning the number 10</p>																																													
(b)	<p>1st M1 for an attempt to rank the number of pages (at least 4 correct) Allow reverse ranks 2nd M1 for attempt at d^2 row (may be implied by sight of $\sum d^2 = 66$ or 264 for reverse ranks) 3rd M1 for use of the correct formula, follow through their $\sum d^2$ if clearly stated If answer is not correct, a correct expression is required. A1 for 0.6 (or -0.6 for reverse ranks)</p>																																													
(c)	<p>1st B1 for both hypotheses in terms of ρ, one tail H_1 (compatible with ranks) Allow use of ρ_s Hypotheses just in words e.g. “no correlation” score B0. 2nd B1 for cv of 0.5636 [If they have a two tail H_1 then allow 0.6485] Allow \pm for reverse ranking but must be same sign as r_s If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible. 3rd B1ft for a correct contextualised comment. Must mention “librarian” (or he) <u>or</u> “number of pages” and “borrowing” Follow through their r_s and their cv (provided it is $cv < 1$) Don't insist on the word “positive” or “negative” for a one-tailed test Use of “association” is B0 Independent of 1st B1 so if $r_s > cv$ must say there is sufficient evidence of(o.e.) and if $r_s < cv$ must say insufficient evidence of ... (o.e.) regardless of their hypotheses</p>																																													

Question Number	Scheme	Marks
<p>2. (a)</p>	<p>$H_0 : \mu_g - \mu_s = 1.5$ [$g =$ in a group, $s =$ on their own]</p> <p>$H_1 : \mu_g - \mu_s > 1.5$</p> $\text{s.e.} = \sqrt{\frac{2.1^2}{80} + \frac{1.4^2}{65}} = [\sqrt{0.08527\dots}] = [0.292]$ $z = \frac{8.7 - 6.6 - 1.5}{\sqrt{\frac{2.1^2}{80} + \frac{1.4^2}{65}}}$ <p style="text-align: center;">= 2.0546.... awrt 2.05(5)</p> <p>cv 1% one tailed = 2.3263</p> <p style="text-align: center;">Not significant, accept H_0</p> <p>Insufficient evidence that using plan as part of a group leads to weight loss of more than 1.5 kg than using plan on one's own or researcher's belief not supported</p>	<p>B1</p> <p>B1</p> <p>M1</p> <p>dM1</p> <p>A1</p> <p>B1</p> <p>dM1</p> <p>A1ft</p> <p style="text-align: right;">(8)</p>
<p>(b)</p>	<p>Since sample is large Central Limit Theorem (CLT) applies No need to <u>assume</u> normal distribution</p>	<p>B1</p> <p>dB1</p> <p style="text-align: right;">(2)</p>
Notes		
<p>(a)</p>	<p>1st & 2nd B1 for hypotheses. Accept μ_1, μ_2 or μ_A, μ_B etc if there is some indication of which is which e.g. $G \sim N(\mu_g, 8.7)$</p> <p>1st M1 for an attempt at se with 3 out of 4 values correct. Condone switching 2.1 and 1.4</p> $\sqrt{\frac{2.1^2 \text{ or } 1.4^2}{80} + \frac{1.4^2 \text{ or } 2.1^2}{65}}$ <p>2nd dM1 dependent on 1st M1 for a correct numerator (must have -1.5) and ft their se.</p> <p>1st A1 for awrt 2.05</p> <p>3rd B1 for ± 2.3263 or better seen or probability of awrt 0.02</p> <p>3rd dM1 dep. on 1st M1 for a correct statement based on their normal cv and their test statistic</p> <p>2nd A1ft for correct comment in context. Must mention "plan" and "group or individual" and "1.5" <u>or</u> "researcher" and "belief or claim"</p> <p>NB Use of cv for difference in means D will have $D = 1.5 + 2.3263 \times \text{s.e.} = \text{awrt } 2.18$ and requires sight of $d = 2.1$ with a comment for the 3rd M1</p>	
<p>(b)</p>	<p>1st B1 for mentioning "large samples" and "CLT"</p> <p>2nd dB1 dependent on 1st B1 for stating no need to assume normality (since CLT assures it)</p>	

Question Number	Scheme	Marks
3. (a)	Label staff (from 1 – 16) and children (from 1 – 40) Use random numbers to select 4 staff and 10 children	B1 B1 B1 (3)
(b)	$\bar{x} = \hat{\mu} = 31.2142\dots$ awrt <u>31.2</u> $s^2 = \frac{26983 - 14 \times "31.2\dots"}{13}$ $= 1026.33\dots$ awrt <u>1030</u>	B1 M1 A1ft A1 (4)
(c)	" $\frac{\sqrt{1026.33\dots}}{\sqrt{14}}$ " , = 8.562.. awrt <u>8.56</u>	M1, A1 (2)
(d)	The variation within each stratum is quite small (o.e.) The difference in the means will be quite large , (so variations from the overall mean will be large giving a larger overall s.e.)	B1 B1 (2)
Notes		Total 11
(a)	1 st B1 for labelling\numbering\listing staff <u>and</u> children 2 nd B1 for use of random numbers or “randomly select” in <u>each group</u> (may be implied) 3 rd B1 for selecting the correct number of staff <u>and</u> children e.g. randomly select 4 staff and 10 children scores 2 nd and 3 rd B marks since randomly selecting and the “each group” is implied,	
(b)	B1 for awrt 31.2 M1 for a correct expression ft their \bar{x} and allow transcription error in $\sum x^2$ e.g. 29683 1 st A1ft for a fully correct expression ft their \bar{x} only 2 nd A1 for awrt 1030	
(c)	M1 for attempting $\frac{\text{"their } s\text{"}}{\sqrt{14}}$ (must have 14) A1 for awrt 8.56	
(d)	1 st B1 for a suitable comment about variation (se) suggesting that variation (se) within strata is less than that overall 2 nd B1 for a suitable reason about means , pointing out that the individuals’ weights will vary a lot from the overall mean and so overall s.e. will be higher.	

Question Number	Scheme	Marks
<p>4. (a)</p> <p>(b)</p>	$H_0 : \mu = 0.5 \quad H_1 : \mu \neq 0.5$ <p>(Significance level =)10%</p> <p>(0.5 is in the interval so not significant, accept H_0, can accept) $\mu = 0.5$</p> $1.6449 \times \frac{\sigma}{\sqrt{100}} = 0.0247$ $\sigma = 0.15016 \quad \text{or} \quad \frac{10 \times 0.0247}{1.6449} \quad (\text{awrt } 0.15)$ $0.479 \pm 1.96 \times \frac{\sigma}{\sqrt{150}}$ <p style="text-align: right;">awrt <u>(0.455, 0.503)</u></p>	<p>B1 dB1 B1 (3)</p> <p>M1 B1 A1 M1 B1 A1 (6)</p> <p>Total 9</p>
Notes		
	<p>(a) 1st B1 for both hypotheses in terms of μ. 2nd dB1 for 10% but accept 5% if they have a one-tail test as H_1 3rd B1 for a correct comment leading to accepting H_0 Ignore any 'further calculations'.</p> <p>(b) 1st M1 for $z \frac{\sigma}{\sqrt{100}} = k$, using $n = 100$ and where $z > 1.5$ and $0.02 < k < 0.03$ 1st B1 for 1.6449 or better in an attempt (could be $1.6449\sigma = k$ or even $1.6449 \sigma^2 = k$) 1st A1 for a correct expression for σ e.g. awrt 0.15 2nd M1 for $\bar{x} \pm z \times \frac{\sigma}{\sqrt{150}}$ for any $z (> 1)$ and ft their σ and allow $\bar{x} \in (0.4633, 0.5127)$ Allow use of letter σ without a value. 2nd B1 for 1.96 or better in an attempt (could be 1.96σ or even $1.96 \sigma^2$) 2nd A1 for awrt 0.455 <u>and</u> awrt 0.503</p>	

Question Number	Scheme	Marks
<p>5 (i)</p> <p>(ii)(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p>	<p>Let $R = B_1 + B_2 + B_3 + B_4 + B_5 - 5H$ so $E(R) = -25$ (o.e.)</p> $\text{Var}(R) = 5 \times 6^2 + 5^2 \times 4^2$ $R \sim N(-25, \sqrt{580^2})$ <p>$P(R > 0) = P(Z > \frac{0-25}{\sqrt{580}}) = P(Z > 1.04)$, = 0.149619...(calc) <u>or</u> 0.1492 (tables)</p> <p>$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{5}\right)$</p> $\text{Var}(D) = \sigma^2 + \frac{\sigma^2}{5} \left[= \frac{6\sigma^2}{5} \right], \text{ so } D \sim N\left(0, \frac{6\sigma^2}{5}\right)$ <p>$P(Y_1 > \bar{X} + \sigma) = P(D > \sigma) = P\left(Z > \frac{\sigma}{\sqrt{\frac{6}{5}\sigma}}\right)$</p> <p>= $P(Z > 0.912\dots) = 0.181(3 \text{ dp})$ (*)</p> <p>Since U_1 and \bar{U} are not independent (so variance formula cannot be used) Can be implied e.g. U_1 used to calculate \bar{U}, U_1 and \bar{U} from same sample o.e.</p> <p>Let $F = U_1 - \bar{U} = U_1 - \frac{(U_1 + U_2 + U_3 + U_4 + U_5)}{5} = \frac{4U_1 - (U_2 + U_3 + U_4 + U_5)}{5}$</p> $\text{Var}(F) = \frac{4^2\sigma^2 + 4\sigma^2}{5^2} = 0.8\sigma^2, \text{ so } F \sim N(0, 0.8\sigma^2)$ <p>$P(F > \sigma) = P\left(Z > \frac{\sigma}{\sigma\sqrt{0.8}}\right) = P(Z > 1.118\dots)$</p> <p>= 0.1314 (tables) or 0.131776...(calc) awrt 0.131-0.132</p>	<p>B1</p> <p>M1A1</p> <p>dM1 A1</p> <p>(5)</p> <p>B1</p> <p>M1, A1</p> <p>(3)</p> <p>M1</p> <p>A1cso</p> <p>(2)</p> <p>B1</p> <p>(1)</p> <p>M1, A1</p> <p>dM1, A1</p> <p>M1</p> <p>A1cso</p> <p>(6)</p> <p>Total 17</p>
Notes		
<p>(i)</p> <p>(ii)(a)</p> <p>(ii)(b)</p> <p>(c)</p> <p>(d)</p>	<p>1st B1 for $E(R) = -25$ (or 25 if their R is defined the other way around)</p> <p>1st M1 for an attempt at $\text{Var}(R) = 5\text{Var}(B) + 25\text{Var}(H)$. Condone swapping of 6^2 and 4^2</p> <p>1st A1 for normal and correct variance (ft their mean)</p> <p>2nd dM1 for attempting the correct probability and standardising with their mean and sd. This mark is dependent on 1st M1 so if R is not being used or M0 for variance score M0 If their method is not crystal clear then they must be attempting $P(Z > +ve \text{ value})$ o.e</p> <p>2nd A1 for answer in the range [0.149, 0.150]</p> <p>B1 for correct distribution of \bar{X} (may be implied for a correct answer for D)</p> <p>M1 for correct attempt at $\text{Var}(D)$ (ft their $\text{Var}(\bar{X})$) [A1 needs must be fully correct]</p> <p>M1 for expressing probability in terms of D and standardising</p> <p>A1cso for seeing $P(Z > 0.912\dots)$ or prob of $1 - 0.8186$ (tables) or $0.180655\dots$(calc)</p> <p>B1 correct statement that should mention U_1 and \bar{U}</p> <p>1st M1 for forming an expression in terms of $U_1\dots U_5$ only</p> <p>1st A1 for collecting U_1 terms and getting in a form where $\text{Var}(aX \pm bY)$ can be used.</p> <p>2nd dM1 for a correct expression for $\text{Var}(\text{their } F)$. Dependent on 1st M1.</p> <p>2nd A1 for a correct distribution for F</p> <p>3rd M1 attempting a correct prob and standardising using their $\text{Var}(F)$, σ must cancel</p> <p>3rd A1cso for awrt 0.131 or 0.132</p>	

Question Number	Scheme	Marks																									
<p>6. (a)</p>	<p>$H_0 : U[0, 10]$ is a suitable model $H_1 : U[0, 10]$ is not a suitable model</p> <table border="1" data-bbox="280 248 1034 465"> <thead> <tr> <th>D</th> <th>O_i</th> <th>E_i</th> <th>$\frac{(O_i - E_i)^2}{E_i}$</th> <th>$\frac{O_i^2}{E_i}$</th> </tr> </thead> <tbody> <tr> <td>0 – 4</td> <td>22</td> <td>40</td> <td>8.1</td> <td>12.1</td> </tr> <tr> <td>4 – 7</td> <td>39</td> <td>30</td> <td>2.7</td> <td>50.7</td> </tr> <tr> <td>7 – 9</td> <td>25</td> <td>20</td> <td>1.25</td> <td>31.25</td> </tr> <tr> <td>9 – 10</td> <td>14</td> <td>10</td> <td>1.6</td> <td>19.6</td> </tr> </tbody> </table> <p>Values of D Expected Freq 4th or 5th col $\chi^2 = 13.65$</p> <p>$\nu = 3, \chi_3^2(1\%) = 11.345$ [Reject H_0,] the uniform distribution over $[0, 10]$ is not a suitable model</p>	D	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$	0 – 4	22	40	8.1	12.1	4 – 7	39	30	2.7	50.7	7 – 9	25	20	1.25	31.25	9 – 10	14	10	1.6	19.6	<p>B1 B1 M1A1 M1 A1 B1, B1 A1 (9) M1, A1 B1ft (3) M1, A1 (2) B1 (1) B1 (1) B1 (1) B1 B1 (2) Total 19</p>
D	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$																							
0 – 4	22	40	8.1	12.1																							
4 – 7	39	30	2.7	50.7																							
7 – 9	25	20	1.25	31.25																							
9 – 10	14	10	1.6	19.6																							
Notes																											
<p>(a)</p>	<p>2nd B1 for the correct values for D (can be implied by 40, 30, 20, and 10.) 1st M1 for at least 2 expected frequencies or clear use of a correct formula e.g. $0.4N$ 1st A1 for all the correct E_i 2nd M1 for at least 2 correct calculations from 4th or 5th column 2nd A1 for a test statistic of 13.65 (accept 13.7 to 3 sf) Awr 13.7 only scores 2nd B1M1A1M1A1 3rd A1 for a correct conclusion rejecting the uniform model. Award provided their test statistic > 11.345</p>																										
	<p>(b) M1 for some attempt to use πR^2 to find r</p>																										
	<p>(c) M1 for a correct statement that it is not significant A1 for correctly stating that Henry’s model is suitable o.e.</p>																										
	<p>(d) B1 Independence or association mentioned at least once if ditto marks used. Allow connection but not correlation.</p>																										
	<p>(f) B1 for recognising there is an $E_i < 5$ and need for pooling/combining oe</p>																										
	<p>(g) 2ndB1 for correctly stating that Phoebe’s belief is not supported by the data oe (depends on their cv being > 1.411)</p>																										

