Question 1

(i)	**EITHER:**			
	$S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 1398.56 - \frac{1}{14} \times 139.8 \times 140.4$ $= -3.434$	M1 for method for S_{xy}		
	$S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 1411.66 - \frac{1}{14} \times 139.8^2 = 15.657$	M1 for method for at least one of S_{xx} or S_{yy}		
	$S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 1417.88 - \frac{1}{14} \times 140.4^2 = 9.869$	A1 for at least one of S_{xy}, S_{xx}, S_{yy} correct		If $\bar{x}$ and $\bar{y}$ used in rounded form, be generous with first A1
	$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-3.434}{\sqrt{15.657 \times 9.869}}$ $= -0.276$	M1 for structure of r A1 (-0.27 to -0.28 to 2dp)		Structure of r needs to be fully correct in all parts – the first two M1 marks must have been earned and $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ applied.
	OR:			
	$\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 1398.56/14 - 9.9857 \times 10.0286$ $= -0.2454$	M1 for method for cov (x,y)		
	$\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(15.657/14)} = \sqrt{1.1184} = 1.0575$	M1 for method for at least one msd		
	$\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(9.869/14)} = \sqrt{0.7049} = 0.8396$	A1 for at least one of cov (x,y), msd(x), msd(y) correct		If $\bar{x}$ and $\bar{y}$ used in rounded form, be generous with first A1
	$r = \frac{\text{cov(x,y)}}{rmsd(x)rmsd(y)} = \frac{-0.2454}{1.0575 \times 0.8396}$ $= -0.276$ NB: using only 3dp in calculating $\bar{x}$ and $\bar{y}$ leads to answer of 0.284 which is still in the acceptable range	M1 for structure of r A1 (-0.27 to -0.28 to 2dp)	**5**	Structure of r needs to be fully correct in all parts – the first two M1 marks must have been earned and $r = \frac{\text{cov(x,y)}}{rmsd(x)rmsd(y)}$ applied.

(ii)	H_0: $\rho = 0$ H_1: $\rho \neq 0$ (two-tailed test) where ρ is the population correlation coefficient For $n = 14$, 5% critical value = –0.5324 Since –0.276 > –0.5324 the result is not significant. Thus we do not have sufficient evidence to reject H_0 There is not sufficient evidence at the 5% level to suggest that there is correlation between birth rate and death rate	B1 for H_0, H_1 in symbols B1 for defining ρ B1 for critical value (+ or -) M1 for a sensible comparison leading to a conclusion (provided that -1 < r < 1) A1 for correct result ft their r B1 ft for conclusion in context	**6**	Condone hypotheses written in words and context. e.g. allow H_0: There is no correlation between x & y, H_1: There is correlation between x & y. (i.e. allow x & y as 'context' since these are defined in the question) NB If hypotheses given only in words and 'association' mentioned then do not award first B1 and last B1 For hypotheses written in words, candidates must make it clear that they are testing for evidence of correlation in the population. One-tailed test cv = (-) 0.4575 Comparison should be between the candidate's value of r from part (i) and an appropriate cv (i.e. the sign of the cv and the sign of r should be the same). NOTE If result not stated but final conclusion is correct award SC1 to replace the final A1 B1
(iii)	The underlying population must have a bivariate Normal distribution. Since the scatter diagram has a roughly elliptical shape.	B1 E1 for elliptical shape	**2**	Not bivariate and Normal
(iv)	Because this data point is a long way from the other data and it is below and to the right of the other data. It does bring the validity of the test into question since this extra data point is so far from the other points and so there is less evidence of ellipticity.	E1 for a long way E1 for below and to the right of. E1 for does cast doubt on validity E1 for less elliptical	**4**	Indication that the point is (possibly) an outlier For identifying the position of this point (allow in terms of x and y) Allow 'no' but only with with suitable explanation e.g. the sample is still too small to provide evidence either for or against the presence of ellipticity.
		TOTAL	**17**	

Question 2

(i)	Mean $= \frac{\Sigma xf}{n} = \frac{0+15+24+27+16+10}{50}$ $= \frac{92}{50} = 1.84$ Variance $= \frac{1}{n-1}\left(\Sigma fx^2 - n\bar{x}^2\right)$ $= \frac{1}{49}\left(258 - 50\times 1.84^2\right)$ $= 1.81$ (to 2 d.p.)	B1 for mean M1 for calculation A1	**3**	Use of MSD gets M1 A0 Standard deviation gets M0 A0 unless "Variance = 1.81" is seen.
(ii)	Because the mean is close to the variance	B1	**1**	Must compare mean and their variance as found in part (i)
(iii)	(*A*) P(No sultanas) $= e^{-1.84}\frac{1.84^0}{0!}$ $= 0.159$ (3 s.f.) (*B*) P(At least two sultanas) = $1 - e^{-1.84}\frac{1.84^0}{0!} - e^{-1.84}\frac{1.84^1}{1!}$ $=1 - 0.159 - 0.292 = 0.549$	M1 for probability calc. A1 M1 for P(1) M1 for 1 – [P(0) + P(1)] used A1 cao	**5**	[1.8 leads to 0.1653] Or attempt to find P(2) + P(3) + P(4) + …+ P(8) Use of $\lambda = 1.8$ loses both accuracy marks [1.8 leads to 1 – 0.4296 = 0.5372]
(iv)	$\lambda = 5\times 1.84 = 9.2$ Using tables: $P(X \geq 10) = 1 - P(X \leq 9)$ $= 1 - 0.5611$ (= 0.4389 **NB ANSWER GIVEN)**	B1 for mean (SOI) M1 for $1 - P(X \leq 9)$ A1	**3**	Any λ

(v)	P(2 out of 6 contain at least ten sultanas) $= \binom{6}{2} \times 0.4389^2 \times 0.5611^4 = 0.2864$	M1 for $p^2 \times q^4$ M1 dep for coefficient A1	**3**	$p + q = 1$ Coefficient of 15 as part of a binomial calculation ft if p rounded from part (iv)
(vi)	Use Normal approx with $\mu = np = 60 \times 0.4389 = 26.334$ $\sigma^2 = npq = 60 \times 0.4389 \times 0.5611 = 14.776$ $P(X > 30) = P\left(Z > \frac{30.5 - 26.334}{\sqrt{14.776}}\right)$ $= P(Z > 1.0838) = 1 - \Phi(1.0838)$ $= 1 - 0.8608$ $= 0.1392$	B1 for μ B1 for σ^2 B1 for correct continuity correction M1 for probability using correct tail. FT their μ & σ^2 A1 **cao**	**5**	**SOI** Allow 26.3 Allow 14.8 … …(giving $P(Z > 1.091..) = 0.137$ 3sf) But do not FT wrong or omitted CC
			20	

Question 3

(i)	(A) $P(X < 325)$ $= P\left(Z < \dfrac{325-355}{52}\right)$ $= P(Z < -0.577)$ $= 1 - \Phi(0.577) = 1 - 0.7181$ $= 0.2819$	M1 for standardising M1 for correct structure A1 CAO	**3**	NB When a candidate's answers suggest that (s)he appears to have neglected to use the difference column of the Normal distribution tables penalise the first occurrence only Ignore spurious continuity corrections & allow reversal of numerator i.e. correct tail (including below a negative z) Allow answers which round to 0.282
	(B) $P(300 < X < 400)$ $= P\left(\dfrac{300-355}{52} < Z < \dfrac{400-355}{52}\right)$ $= P(-1.058 < Z < 0.865)$ $= \Phi(0.865) - (1 - \Phi(1.058))$ $= 0.8065 - (1 - 0.8549)$ $= 0.6614$ (0.6615 from GDC)	M1 for standardising both M1 for correct structure A1 CAO	**3**	Penalise spurious continuity corrections Allow 0.663 if penalised inappropriate table use already Use of standard deviation = $\sqrt{52}$ or 52^2 can earn M1 for structure only in each part – max 2/6
(ii)	From tables $\Phi^{-1}(0.2) = -0.8416$ $\dfrac{k-355}{52} = -0.8416$ $k = 355 - 0.8416 \times 52 = 311.2$	B1 for ± 0.8416 seen M1 for equation in k A1 CAO	**3**	NOT 1 – 0.8416 Equation must be equivalent to this. Penalise use of + 0.8416 unless numerator has been reversed. Condone other z values but use of probabilities here, e.g. use of 0.2 or $\Phi(0.2) = 0.5793$, gets M0 A0 Allow 311

(iii)	H_0: $\mu = 355$; H_1: $\mu \neq 355$. Where μ denotes the population mean (reaction time for women) Test statistic = $\dfrac{344-355}{52/\sqrt{25}} = \dfrac{-11}{10.4} = -1.058$ 5% level 2 tailed critical value of z = 1.96 $-1.058 > -1.96$ so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that women have a different reaction time from men in this experiment.	B1 for use of 355 in hypotheses B1 for both correct B1 for definition of μ M1 must include $\sqrt{25}$ A1 B1 for 1.96 M1 for a sensible comparison leading to a conclusion A1 for correct conclusion in words in context	**8**	Use of 355 in hypotheses and hypotheses given in terms of μ not p or x, etc. unless letter used is clearly defined as population mean Allow + 1.058 only if later compared with + 1.96 Or -1.96 Do not accept 'men and women have same reaction time'
		TOTAL	**17**	

Question 4

<table>
<tr><td>(i)</td><td>H_0: no association between pebble size and site
H_1: some association between pebble size and site;</td><td>B1</td><td></td><td>Must be in context
NB if H_0 H_1 reversed, or ‘correlation’ mentioned, do not award first B1 or final E1</td></tr>
<tr><td></td><td><table>
<tr><th>EXPECTED</th><th>Site A</th><th>Site B</th><th>Site C</th></tr>
<tr><td>Large</td><td>13.70</td><td>9.44</td><td>13.86</td></tr>
<tr><td>Medium</td><td>33.33</td><td>22.96</td><td>33.70</td></tr>
<tr><td>Small</td><td>42.96</td><td>29.60</td><td>43.44</td></tr>
</table></td><td>M1 A2 for expected values (to 2 dp)
(allow A1 for at least one row or column correct)</td><td></td><td>1d.p.can get M1A1A0
M1A2 can be implied by correct contributions/final answer</td></tr>
<tr><td></td><td><table>
<tr><th>CONTRIB’N</th><th>Site A</th><th>Site B</th><th>Site C</th></tr>
<tr><td>Large</td><td>0.1226</td><td>0.6940</td><td>1.0731</td></tr>
<tr><td>Medium</td><td>0.8533</td><td>1.5484</td><td>3.7861</td></tr>
<tr><td>Small</td><td>0.3793</td><td>0.3913</td><td>1.2744</td></tr>
</table></td><td>M1 for valid attempt at $(O-E)^2/E$
A1</td><td></td><td>NB These (M1A1) marks cannot be implied by a correct final value of X^2.
A1 for at least 1 row/column correct</td></tr>
<tr><td></td><td>$X^2 = 10.12$</td><td>M1 for summation
A1 for X^2</td><td></td><td>Dependent on previous M1</td></tr>
<tr><td></td><td>Refer to χ_4^2</td><td>B1 for 4 deg of freedom</td><td></td><td></td></tr>
<tr><td></td><td>Critical value at 5% level = 9.488</td><td>B1 CAO for cv</td><td></td><td>Award only if no incorrect working seen</td></tr>
<tr><td></td><td>Result is significant</td><td>B1 ft their ‘sensible’ X^2 and critical value</td><td></td><td>Allow reject H_0. B0 if critical value of 0.711 (lower tail) or 2.776 (t distribution) used.</td></tr>
<tr><td></td><td>There is evidence to suggest that there is some association between pebble size and site</td><td>E1 must be consistent with their X^2</td><td>12</td><td>Dependent on previous B1
SC1 (to replace B1E1 if first B1B1 earned where ‘significant’ not stated but final statement is correct)</td></tr>
</table>

(ii)	Site A Contributes least to X^2 showing that frequencies are as expected if there were no association. OR Contribution of 0.8533 implies that there are (slightly) fewer medium pebbles than expected. Site B Contribution of 1.5484 implies that there are fewer medium pebbles than expected. Site C Contribution of 3.7861 implies that there are **a lot** more medium than expected. NB MAX 3/6 for answers not referring to contributions (explicitly or implicitly).	E2,1,0 E2,1,0 E2,1,0 Need 'a lot more' for E2	**2** **2** **2**	NOTE For each site, some reference to contributions needed (explicitly or implicitly). Award E2 only if no incorrect additional comment made. Allow large/small 'as expected' or 'more than expected' and medium 'as expected' or 'less than expected' for E1 (if contribution not mentioned) Award E2 only if no incorrect additional comment made. Allow large/small 'as expected' or 'more than expected' and medium 'less than expected' for E1 (if contribution not mentioned) Award E2 only if no incorrect additional comment made. Allow large/small 'fewer than expected' and medium 'more than expected' for E1 (if contribution not mentioned)
		TOTAL	**18**	

Additional notes re Q1(ii)

For those carrying out a one-tailed test, B0 B1 B1 M1 A1 B1 is available provided that working is consistent with a one-tailed test being used.
For the final B1 to be earned, the conclusion should refer to alternative hypothesis used. e.g. 'There is not sufficient evidence at the 5% level to suggest that there is a **negative** correlation between birth rate and death rate'.

If the cv is taken from the Spearman's Test table (i.e. -0.5385 and -0.4637) then the third B1 will be lost.
If other 'sensible' cvs are used then only B1 B1 B0 M1 A0 B0 available. Use of t distribution leads to B1 B1 B0 M0 A0 B0 max.

Additional notes re Q3(iii)

Critical Value Method
355 – 1.96 × 52 ÷ √25 gets M1B1
= 334.6… gets A1
334.6 < 344 gets M1for sensible comparison
A1 still available for correct conclusion in words & context

Confidence Interval Method
CI centred on 344
+ or - 1.96 × 52 ÷ √25 gets M1 B1
= (323.62, 364.384) A1
contains 355 gets M1
A1 still available for correct conclusion in words & context

Probability Method
Finding P(sample mean < 344) = 0.1451 gets M1 A1 B1
0.1451 > 0.025 gets M1 for a sensible comparison if a conclusion is . 0.1451 > 0.05 gets M1 A0 unless using one tailed test
A1 still available for correct conclusion in words & context.
Condone P(sample mean > 344) = 0.8549 for M1 but only allow A1 if later compared with 0.975 at which point the final M1and A1 are still available

One-tailed test
Max B1 B0 B1 M1 A1 B1 (for cv = -1.645) M1 A1 (provided that the conclusion relates to H_1: $\mu < 355$, e.g. there is insufficient evidence to suggest that women have a lower reaction time than men in this experiment).

Consistent use of σ = √52
Do not penalise in parts (ii) and (iii).