GENERAL INSTRUCTIONS

Marks in the mark scheme are explicitly designated as **M**, **A**, **B**, **E** or **G**.

M marks ("method") are for an attempt to use a correct method (not merely for stating the method).

A marks ("accuracy") are for accurate answers and can only be earned if corresponding **M** mark(s) have been earned. Candidates are expected to give answers to a sensible level of accuracy in the context of the problem in hand. The level of accuracy quoted in the mark scheme will sometimes deliberately be greater than is required, when this facilitates marking.

B marks are independent of all others. They are usually awarded for a single correct answer. Typically they are available for correct quotation of points such as 1.96 from tables.

E marks ("explanation") are for explanation and/or interpretation. These will frequently be sub divisible depending on the thoroughness of the candidate's answer.

G marks ("graph") are for completing a graph or diagram correctly.

- Insert part marks in **right-hand** margin in line with the mark scheme. For fully correct parts tick the answer. For partially complete parts indicate clearly in the body of the script where the marks have been gained or lost, in line with the mark scheme.

- Please indicate incorrect working by ringing or underlining as appropriate.

- Insert total in **right-hand** margin, ringed, at end of question, in line with the mark scheme.

- Numerical answers which are not exact should be given to at least the accuracy shown. Approximate answers to a greater accuracy *may* be condoned.

- Probabilities should be given as fractions, decimals or percentages.

- FOLLOW-THROUGH MARKING SHOULD NORMALLY BE USED WHEREVER POSSIBLE. There will, however, be an occasional designation of '**c.a.o.**' for "correct answer only".

- Full credit MUST be given when correct alternative methods of solution are used. If errors occur in such methods, the marks awarded should correspond as nearly as possible to equivalent work using the method in the mark scheme.

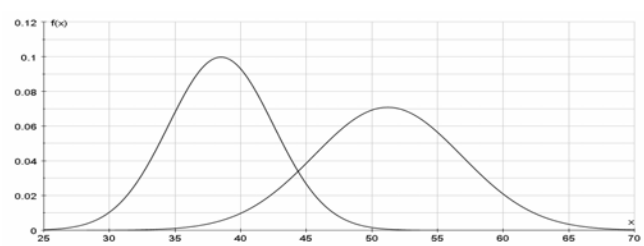
- The following notation should be used where applicable:

Question 1

(i)	Uniform average rate of occurrence; Successive arrivals are independent. Suitable arguments for/against each assumption: Eg Rate of occurrence could vary depending on the weather (any reasonable suggestion)	E1,E1 for suitable assumptions E1, E1 must be in context	**4**
(ii)	Mean = $\frac{\Sigma xf}{n} = \frac{39+40+36+32+15}{100} = \frac{162}{100}$ = 1.62 Variance = $\frac{1}{n-1}\left(\Sigma fx^2 - n\bar{x}^2\right)$ $= \frac{1}{99}\left(430 - 100 \times 1.62^2\right) = 1.69$ (to 2 d.p.)	B1 for mean ***NB answer given*** M1 for calculation A1	**3**
(iii)	Yes, since mean is close to variance	B1FT	**1**
(iv)	P(X = 2) = $e^{-1.62}\frac{1.62^2}{2!}$ = 0.260 (3 s.f.) *Either:* Thus the expected number of 2's is 26 which is reasonably close to the observed value of 20. *Or*: This probability compares reasonably well with the relative frequency 0.2	M1 for probability calc. M0 for tables unless interpolated A1 B1 for expectation of 26 or r.f. of 0.2 E1	**4**
(v)	$\lambda = 5 \times 1.62 = 8.1$ Using tables: P($X \geq$ 10) = 1 – P($X \leq$ 9) = 1 – 0.7041 = 0.2959	B1FT for mean (SOI) M1 for probability from using tables to find 1 – P($X \leq$ 9) A1 FT	**3**
(vi)	Mean no. of items in 1 hour = 360 × 1.62 = 583.2 Using Normal approx. to the Poisson, $X \sim$ N(583.2, 583.2): $P(X \leq 550.5) = P\left(Z \leq \frac{550.5 - 583.2}{\sqrt{583.2}}\right)$ = P($Z \leq$ -1.354) = 1 - Φ(1.354) = 1 – 0.9121	B1 for Normal approx. with correct parameters (SOI) B1 for continuity corr. M1 for probability	**4**

	= 0.0879 (3 s.f.)	using correct tail A1 CAO, (but FT wrong or omitted CC)	
			19

Question 2

(i)	$X \sim N(38.5,16)$ $P(X > 45) = P\left(Z > \frac{45-38.5}{4}\right)$ $= P(Z > 1.625)$ $= 1 - \Phi(1.625) = 1 - 0.9479$ = 0.0521 (3 s.f.) *or* 0.052 (to 2 s.f.)	M1 for standardizing A1 for 1.625 M1 for prob. with tables and correct tail A1 CAO (min 2 s.f.)	**4**
(ii)	From tables $\Phi^{-1}(0.90) = 1.282$ $\frac{x-38.5}{4} = -1.282$ $x = 38.5 - 1.282 \times 4 = 33.37$ So 33.4 should be quoted	B1 for 1.282 seen M1 for equation in *x* and negative z-value A1 CAO	**3**
(iii)	$Y \sim N(51.2, \sigma^2)$ From tables $\Phi^{-1}(0.75) = 0.6745$ $\frac{55-51.2}{\sigma} = 0.6745$ $3.8 = 0.6745\,\sigma$ $\sigma = 5.63$	B1 for 0.6745 seen M1 for equation in σ with z-value A1 ***NB answer given***	**3**
(iv)		G1 for shape G1 for means, shown explicitly or by scale G1 for lower max height in diesel G1 for higher variance in diesel	**4**
(v)	$P(\text{Diesel} > 45) = P\left(Z > \frac{45-51.2}{5.63}\right)$	M1 for prob. calc. for diesel	

	$= P(Z > -1.101) = \Phi(1.101) = 0.8646$ P(At least one over 45) = 1 – P(Both less than 45) = 1 - (1 - 0.0521) x (1 - 0.8646) = 1 - 0.9479 x 0.1354 = 0.8717 NB allow correct alternatives based on: P(D over, P under)+P(D under, P over)+ P(both over) *or* P(D over) + P(P over) – P(both over)	M1 for correct structure M1*dep* for correct probabilities A1 CAO (2 s.f. min)	**4**
			18

Question 3

(i)	$\bar{x} = 4.5$, $\bar{y} = 26.85$ $b = \dfrac{Sxy}{Sxx} = \dfrac{983.6 - 36 \times 214.8/8}{204 - 36^2/8} = \dfrac{17}{42} = 0.405$ OR $b = \dfrac{983.6/8 - 4.5 \times 26.85}{204/8 - 4.5^2} = \dfrac{2.125}{5.25} = 0.405$ hence least squares regression line is: $y - \bar{y} = b(x - \bar{x})$ $\Rightarrow y - 26.85 = 0.405(x - 4.5)$ $\Rightarrow y = 0.405x + 25.03$	B1 for $\bar{x}$ and $\bar{y}$ used (SOI) M1 for attempt at gradient (b) A1 for 0.405 **cao** M1 *indep* for equation of line A1FT for complete equation	**5**
(ii)	$x = 4 \Rightarrow$ predicted $y = 0.405 \times 4 + 25.03 = 26.65$ Residual = 27.5 – 26.65 = 0.85	M1 for prediction A1FT for ± 0.85 B1FT for sign (+)	**3**
(iii)	The new equation would be preferable, since the equation in part (i) is influenced by the unrepresentative point (4,27.5)	B1 E1	**2**
(iv)	H_0: $\rho = 0$; H_1: $\rho > 0$ where ρ represents the population correlation coefficient Critical value at 5% level is 0.3783 Since 0.209 < 0.3783, there is not sufficient evidence to reject H_0, i.e. there is not sufficient evidence to conclude that there is any correlation between cycling and swimming times.	B1 for H_0 and H_1 B1 for defining ρ B1 for 0.3783 M1 for comparison leading to conclusion A1*dep on cv* for conclusion in words	**5**

		in context	
(v)	Underlying distribution must be bivariate normal. The distribution of points on the scatter diagram should be approximately elliptical.	B1 E1	**2**
			17

Question 4

(a) (i)	H_0: $\mu = 166500$; H_1: $\mu > 166500$ Where μ denotes the mean selling price in pounds of the population of houses on the large estate	B1 for both correct B1 for definition of μ	**2**
(ii)	$n = 6$, $\Sigma x = 1018500$, $\bar{x} = £169750$ Test statistic = $\dfrac{169750 - 166500}{14200/\sqrt{6}} = \dfrac{3250}{5797}$ $= 0.5606$ 5% level 1 tailed critical value of z = 1.645 0.5606 < 1.645 so not significant. There is insufficient evidence to reject H_0 It is reasonable to conclude that houses on this estate are not more expensive than in the rest of the suburbs.	B1CAO M1 must include √6 A1FT B1 for 1.645 M1 for comparison leading to a conclusion A1 for conclusion in words in context	**6**

<table>
<tr><td>(b)</td><td>H_0: no association between customer and drink types;
H_1: some association between customer and drink types;</td><td>B1</td><td></td></tr>
<tr><td></td><td>
<table>
<tr><th colspan="2" rowspan="2">Observed</th><th colspan="2">Type of drink</th><th rowspan="2">Row totals</th></tr>
<tr><th>Alcoholic</th><th>Soft drinks</th></tr>
<tr><td rowspan="3">Type of customer</td><td>Business</td><td>54</td><td>63</td><td>117</td></tr>
<tr><td>Tourist</td><td>95</td><td>41</td><td>136</td></tr>
<tr><td>Local</td><td>71</td><td>76</td><td>147</td></tr>
<tr><td colspan="2">Column totals</td><td>220</td><td>180</td><td>400</td></tr>
</table>
</td><td></td><td></td></tr>
<tr><td></td><td>
<table>
<tr><th colspan="2" rowspan="2">Expected</th><th colspan="2">Type of drink</th><th rowspan="2">Row totals</th></tr>
<tr><th>Alcoholic</th><th>Soft drinks</th></tr>
<tr><td rowspan="3">Type of customer</td><td>Business</td><td>64.35</td><td>52.65</td><td>117</td></tr>
<tr><td>Tourist</td><td>74.80</td><td>61.20</td><td>136</td></tr>
<tr><td>Local</td><td>80.85</td><td>66.15</td><td>147</td></tr>
<tr><td colspan="2">Column totals</td><td>220</td><td>180</td><td>400</td></tr>
</table>
</td><td>M1 A1 for expected values (to 2dp)</td><td></td></tr>
<tr><td></td><td>
<table>
<tr><th colspan="2" rowspan="2">Chi squared contribution</th><th colspan="2">Type of drink</th><th rowspan="2">Row totals</th></tr>
<tr><th>Alcoholic</th><th>Soft drinks</th></tr>
<tr><td rowspan="3">Type of customer</td><td>Business</td><td>1.665</td><td>2.035</td><td>3.699</td></tr>
<tr><td>Tourist</td><td>5.455</td><td>6.667</td><td>12.122</td></tr>
<tr><td>Local</td><td>1.200</td><td>1.467</td><td>2.667</td></tr>
</table>
</td><td>M1 for valid attempt at $(O-E)^2/E$

M1dep for summation

A1CAO for X^2</td><td>6</td></tr>
<tr><td></td><td>$X^2 = 18.49$

Refer to $\mathcal{X}_2^2$
Critical value at 5% level = 5.991
Result is significant
There is some association between customer type and type of drink.
NB if H_0 H_1 reversed, or ‘correlation’ mentioned, do not award first B1or final B1 or final E1</td><td>B1 for 2 deg of f
B1 CAO for cv
B1dep on cv
E1</td><td>4</td></tr>
<tr><td></td><td></td><td></td><td>18</td></tr>
</table>

Question 1

(i)	Faults are detected randomly and independently Uniform (mean) rate of occurrence	B1 B1	**2**
(ii)	(*A*) $P(X = 0) = e^{-0.15}\dfrac{0.15^0}{0!} = 0.8607$ (*B*) $P(X \geq 2) = 1 - 0.8607 - e^{-0.15}\dfrac{0.15^1}{1!}$ $= 1 - 0.8607 - 0.1291 = 0.0102$	M1 for probability calc. M0 for tables unless interpolated A1 M1 A1	**4**
(iii)	$\lambda = 30 \times 0.15 = 4.5$ Using tables: $P(X \leq 3) = 0.3423$	B1 for mean (SOI) M1 attempt to find $P(X \leq 3)$ A1	**3**
(iv)	Poisson distribution with $\lambda = 10 \times (0.15 + 0.05) = 2$ $P(X = 5) = e^{-2}\dfrac{2^5}{5!} = 0.0361$ (3 s.f.) or from tables $= 0.9834 - 0.9473 = 0.0361$	B1 for Poisson stated B1 for $\lambda = 2$ M1 for calculation or use of tables A1 FT	**4**
(v)	Mean no. of items in 200 days = $200 \times 0.2 = 40$ Using Normal approx. to the Poisson, $X \sim N(40,40)$: $P(X \geq 50) = P\left(Z > \dfrac{49.5 - 40}{\sqrt{40}}\right)$ $= P(Z > 1.502) = 1 - \Phi(1.502) = 1 - 0.9334$ $= 0.0666$ (3 s.f.)	B1 for Normal approx. (SOI) B1 for both parameters B1 for continuity corr. M1 for probability using correct tail A1 **cao, (**but FT wrong or omitted CC)	**5**
			18

Question 2

(i) (*A*)	$X \sim N(42,3^2)$ $P(X > 50.0) = P\left(Z > \frac{50.0-42.0}{3.0}\right)$ $= P(Z > 2.667)$ $= 1 - \Phi(2.667) = 1 - 0.9962$ $= 0.0038$	M1 for standardizing M1 for prob. calc. with correct tail A1 **NB answer given**	**3**
(i) (*B*)	P(not positive) = 0.9962 P(At least one is out of 7 is positive) $= 1 - 0.9962^7 = 1 - 0.9737$ $= 0.0263$	B1 for use of 0.9962 in binomial expression M1 for correct method A1 CAO	**3**
(i) (*C*)	If an innocent athlete is tested 7 times in a year there is a reasonable possibility (1 in 40 chance) of testing positive. Thus it is likely that a number of innocent athletes may come under suspicion and suffer a suspension so the penalty could be regarded as unfair. *Or* this is a necessary evil in the fight against performance enhancing drugs in sport.	E1 comment on their probability in (i) B E1 for sensible contextual conclusion consistent with first comment	**2**
(ii) (*A*)	B(1000, 0.0038)	B1 for B(,) or Binomial B1 *dep* for both parameters	**2**
(ii) (*B*)	A suitable approximating distribution is Poisson(3.8) P(at least 10 positive tests) $= P(X \geq 10) = 1 - P(X \leq 9)$ $= 1 - 0.9942$ $= 0.0058$ *NB Do not allow use of Normal approximation.*	B1 for Poisson soi B1FT *dep* for $\lambda = 3.8$ M1 for attempt to use $1 - P(X \leq 9)$ A1 FT	**4**
(iii)	P(not testing positive) = 0.995 From tables $z = \Phi^{-1}(0.995) = 2.576$ $\frac{h-48.0}{2.0} = 2.576$ $h = 48.0 + 2.576 \times 2.0 = 53.15$	B1 for 0.995 seen (or implied by 2.576) B1 for 2.576 (FT their 0.995) M1 for equation in *h* and positive z-value A1 CAO	**4**
			18

Question 3

(i)	Rank x: 1, 5, 4, 7, 6, 8, 10, 3, 9, 2 Rank y: 2, 4, 5, 8, 9, 7, 10, 6, 3, 1 d: -1, 1, -1, -1, -3, 1, 0, -3, 6, 1 d^2: 1, 1, 1, 1, 9, 1, 0, 9, 36, 1 $r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 60}{10 \times 99}$ = 0.636 (to 3 s.f.) [*allow* 0.64 to 2 s.f.]	M1 for ranking (allow all ranks reversed) M1 for d^2 A1 CAO for Σd^2 M1 for structure of r_s using their Σd^2 A1 f.t. for $\lvert r_s \rvert < 1$ NB No ranking scores zero	**5**
(ii)	H_0: no association between x and y H_1: positive association between x and y Looking for positive association (one-tail test): Critical value at 5% level is 0.5636 Since 0.636 > 0.5636, there is sufficient evidence to reject H_0, i.e. conclude that there appears to be positive association between temperature and nitrous oxide level.	B1 for H_0 B1 for H_1 NB H_0 H_1 <u>not</u> ito rho B1 for ± 0.5636 (FT their H_1) M1 for comparison with c.v., provided $\lvert r_s \rvert < 1$ A1 for conclusion in words f.t. their r_s and sensible cv	**5**
(iii)	Underlying distribution must be bivariate normal. If the distribution is bivariate normal then the scatter diagram will have an elliptical shape. This scatter diagram is not elliptical and so a PMCC test would not be valid. (Allow comment indicating that the sample is too small to draw a firm conclusion on ellipticity and so on validity)	B1 CAO for bivariate normal B1 indep for elliptical shape E1 dep for conclusion	**3**
(iv)	n=60, PMCC critical value is r = 0.2997 So the critical region is $r \geq 0.2997$	B1 B1 FT their sensible c.v.	**2**
(v)	Any three of the following: • Correlation does not imply causation; • There could be a third factor (causing the correlation between temperature and ozone level); • the claim could be true; • increased ozone could cause higher temperatures.	E1 E1 E1	**3**
			18

Question 4

(i)	H_0: no association between method of travel and type of school; H_1: some association between method of travel and type of school;..	B1 for both	**1**
(ii)	Expected frequency = 120/200 × 70 = 42 Contribution = $(21 - 42)^2 / 42$ = 10.5	M1 A1 M1 for valid attempt at $(O-E)^2/E$ A1 FT their 42 provided O = 21 (at least 1 dp)	**4**
(iii)	$X^2 = 42.64$ Refer to $\mathcal{X}_2^2$ Critical value at 5% level = 5.991 Result is significant There appears to be some association between method of travel and year group. NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1	B1 for 2 deg of f(seen) B1 CAO for cv B1 for significant (FT their c.v. provided consistent with their d.o.f. E1	**4**
(iv)	H_0: $\mu = 18.3$; H_1: $\mu \neq 18.3$ Where μ denotes the mean travel time by car for the whole population. Test statistic $z = \dfrac{22.4 - 18.3}{8.0/\sqrt{20}} = \dfrac{4.1}{1.789}$ = 2.292 10% level 2 tailed critical value of z is 1.645 2.292 > 1.645 so significant. There is evidence to reject H_0 It is reasonable to conclude that the mean travel time by car is different from that by bus.	B1 for both correct B1 for definition of μ M1 (standardizing sample mean) A1 for test statistic B1 for 1.645 M1 for comparison leading to a conclusion A1 for conclusion in words and context	**7**
(v)	The test suggests that students who travel by bus get to school more quickly. This may be due to their journeys being over a shorter distance. It may be due to bus lanes allowing buses to avoid congestion. It is possible that the test result was incorrect (ie implication of a Type I error). More investigation is needed before any firm conclusion can be reached.	E1, E1 for any two valid comments	**2** **18**

Question 4 chi squared calculations

H_0: no association between method of travel and type of school;
H_1: some association between method of travel and type of school;

Observed		Type of school		**Row totals**
		Year 6	Year 11	
Method of travel	Bus	21	49	**70**
	Car	65	15	**80**
	Cycle/Walk	34	16	**50**
Column totals		**120**	**80**	**200**

Expected		Type of school		**Row totals**
		Year 6	Year 11	
Method of travel	Bus	42	28	**70**
	Car	48	32	**80**
	Cycle/Walk	30	20	**50**
Column totals		**120**	**80**	**200**

Chi Squared Contribution		Type of school		**Row totals**
		Year 6	Year 11	
Method of travel	Bus	10.50	15.75	**26.25**
	Car	6.02	9.03	**15.05**
	Cycle/Walk	0.53	0.80	**1.33**
Column totals		**17.05**	**25.58**	**42.64**

(i)	$P(X = 1) = 8 \times 0.1^1 \times 0.9^7$ $= 0.383$	M1 for binomial probability P(*X*=1) A1 (at least 2sf) CAO	**2**
(ii)	$\lambda = 30 \times 0.1 = 3$ (*A*) $P(X = 6) = e^{-3}\frac{3^6}{6!} = 0.0504$(3 s.f.) or from tables = 0.9665 – 0.9161 = 0.0504 (*B*) Using tables: $P(X \geq 8) = 1 - P(X \leq 7)$ = 1 – 0.9881 = 0.0119	B1 for mean SOI M1 for calculation or use of tables to obtain P(*X*=6) A1 (at least 2sf) CAO M1 for correct probability calc' A1 (at least 2sf) CAO	**1** **2** **2**
(iii)	*n* is large and *p* is small	B1, B1 Allow appropriate numerical ranges	**2**
(iv)	$\mu = np = 120 \times 0.1 = 12$ $\sigma^2 = npq = 120 \times 0.1 \times 0.9 = 10.8$	B1 B1	**2**
(v)	$P(X > 15.5) = P\left(Z > \frac{15.5-12}{\sqrt{10.8}}\right)$ $= P(Z > 1.065) = 1 - \Phi(1.065) = 1 - 0.8566$ = 0.1434 NB Allow full marks for use of N(12,12) as an approximation to Poisson(12) leading to $1 - \Phi(1.010) = 1 - 0.8438 = 0.1562$	B1 for correct continuity correction. M1 for probability using correct tail A1 **cao, (**but FT wrong or omitted CC)	**3**
(vi)	From tables $\Phi^{-1}(0.99) = 2.326$ $\frac{x+0.5-12}{\sqrt{10.8}} \geq 2.326$ $x = 11.5 + 2.326 \times \sqrt{10.8} \geq 19.14$ So 20 breakfasts should be carried NB Allow full marks for use of N(12,12) leading to $x \geq 11.5 + 2.326 \times \sqrt{12} = 19.56$	B1 for 2.326 seen M1 for equation in *x* and positive *z*-value A1 CAO (condone 19.64) A1FT for rounding appropriately (i.e. round up if c.c. used o/w rounding should be to nearest integer)	**4**
			18

Question 2

(i)	$X \sim N(49.7, 1.6^2)$ (*A*) $P(X > 51.5) = P\left(Z > \frac{51.5 - 49.7}{1.6}\right)$ $= P(Z > 1.125)$ $= 1 - \Phi(1.125) = 1 - 0.8696 = 0.1304$ (*B*) $P(X < 48.0) = P\left(Z < \frac{48.0 - 49.7}{1.6}\right)$ $= P(Z < -1.0625) = 1 - \Phi(1.0625)$ $= 1 - 0.8560 = 0.1440$ $P(48.0 < X < 51.5) = 1 - 0.1304 - 0.1440 = 0.7256$	M1 for standardizing M1 for prob. calc. A1 (at least 2 s.f.) M1 for appropriate prob' calc. A1 (0.725 – 0.726)	**5**
(ii)	P(one over 51.5, three between 48.0 and 51.5) $= \binom{4}{1} \times 0.7256 \times 0.2744^3 = 0.0600$	M1 for coefficient M1 for 0.7256×0.2744^3 A1 FT (at least 2 sf)	**3**
(iii)	From tables, $\Phi^{-1}(0.60) = 0.2533$, $\Phi^{-1}(0.30) = -0.5244$ $49.0 = \mu + 0.2533\,\sigma$ $47.5 = \mu - 0.5244\,\sigma$ $1.5 = 0.7777\,\sigma$ $\sigma = 1.929$, $\mu = 48.51$	B1 for 0.2533 or 0.5244 seen M1 for at least one correct equation μ & σ M1 for attempt to solve two correct equations A1 CAO for both	**4**
(iv)	Where μ denotes the mean circumference of the entire population of organically fed 3-year-old boys. $n = 10$, Test statistic $Z = \frac{50.45 - 49.7}{1.6/\sqrt{10}} = \frac{0.75}{0.5060} = 1.482$ 10% level 1 tailed critical value of z is 1.282 1.482 > 1.282 so significant. There is sufficient evidence to reject H_0 and conclude that organically fed 3-year-old boys have a higher mean head circumference.	E1 M1 A1(at least 3sf) B1 for 1.282 M1 for comparison leading to a conclusion A1 for conclusion in context	**6**
			18

Question 3

(i)	**EITHER:** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x\Sigma y = 6235575 - \frac{1}{10} \times 4715 \times 13175$ $= 23562.5$ $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 2237725 - \frac{1}{10} \times 4715^2 =$ 14602.5 $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 17455825 - \frac{1}{10} \times 13175^2 =$ 97762.5 $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{23562.5}{\sqrt{14602.5 \times 97762.5}} = 0.624$ **OR:** cov $(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 6235575/10 - 471.5 \times 1317.5$ $= 2356.25$ rmsd$(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(14602.5/10)} = \sqrt{1460.25} = 38.21$ rmsd$(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(97762.5/10)} = \sqrt{9776.25} = 98.87$ $r = \frac{\text{cov(x,y)}}{rmsd(x)rmsd(y)} = \frac{2356.25}{38.21 \times 98.87} = 0.624$	M1 for method for S_{xy} M1 for method for at least one of S_{xx} or S_{yy} A1 for at least one of S_{xy}, S_{xx} or S_{yy} correct M1 for structure of r A1 (0.62 to 0.63) M1 for method for cov (x,y) M1 for method for at least one msd A1 for at least one of S_{xy}, S_{xx} or S_{yy} correct M1 for structure of r A1 (0.62 to 0.63)	**5**
(ii)	H_0: $\rho = 0$ H_1: $\rho \neq 0$ (two-tailed test) where ρ is the population correlation coefficient For n = 10, 5% critical value = 0.6319 Since 0.624 < 0.6319 we cannot reject H_0: There is not sufficient evidence at the 5% level to suggest that there is any correlation between length and circumference.	B1 for H_0, H_1 in symbols B1 for defining ρ B1FT for critical value M1 for sensible comparison leading to a conclusion A1 FT for result B1 FT for conclusion in context	**6**
(iii)	*(A)* This is the probability of rejecting H_0 when it is in fact true. *(B)* Advantage of 1% level – less likely to reject H_0 when it is true. Disadvantage of 1% level – less likely to accept H_1 when H_0 is false.	B1 for 'P(reject H_0)' B1 for 'when true' B1, B1 Accept answers in context	**2** **2**

(iv)	The student's approach is not valid. If a statistical procedure is repeated with a new sample, we should not simply ignore one of the two outcomes. The student could combine the two sets of data into a single set of twenty measurements.	E1 E1 – allow suitable alternatives. E1 for combining samples.	**3**
			18

Question 4

(i)	H_0: no association between musical preference and age; H_1: some association between musical preference and age;	B1	1

Observed		Musical preference			Row totals
		Pop	Classical	Jazz	
Age group	Under 25	57	15	12	**84**
	25 – 50	43	21	21	**85**
	Over 50	22	32	27	**81**
Column totals		**122**	**68**	**60**	**250**

Expected		Musical preference			Row totals
		Pop	Classical	Jazz	
Age group	Under 25	40.992	22.848	20.160	**84**
	25 – 50	41.480	23.120	20.400	**85**
	Over 50	39.528	22.032	19.440	**81**
Column totals		**122**	**68**	**60**	**250**

M1 A2 for expected values (at least 1 dp) (allow A1 for at least one row or column correct)

Contributions		Musical preference			
		Pop	Classical	Jazz	
Age group	Under 25	6.25	2.70	3.30	
	25 – 50	0.06	0.19	0.02	
	Over 50	7.77	4.51	2.94	

M1 for valid attempt at $(O\text{-}E)^2/E$
A1 for all correct

M1dep for summation
A1 for X^2 (27.7 – 27.8) **7**

$X^2 = 27.74$

Refer to χ_4^2 — B1 for 4 deg of f
Critical value at 5% level = 9.488 — B1 CAO for cv
Result is significant — B1FT
There is some association between age group and musical preference. — E1 (conclusion in context) **4**

NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1

(ii)	The values of 6.25 and 7.77 show that under 25's have a strong positive association with pop whereas over 50's have a strong negative association with pop. The values of 4.51 and 2.94 show that over 50's have a reasonably strong positive association with both classical and jazz. The values of 2.70 and 3.30 show that under 25's have a reasonably strong negative associations with both classical and jazz. The 25-50 group's preferences differ very little from the overall preferences.	B1, B1 for specific reference to a value from the table of contributions followed by an appropriate comment B1, B1 (as above for second value) B1, B1 (as above for third value)	**6**
			18

Question 1

(i)	$\bar{t} = 112.8,\ \bar{v} = 0.6$ $b = \frac{Svt}{Svv} = \frac{405.2 - 3\times564/5}{2.20 - 3^2/5} = \frac{66.8}{0.4} = 167$ OR $b = \frac{405.2/5 - 0.6\times112.8}{2.20/5 - 0.6^2} = \frac{13.36}{0.08} = 167$ hence least squares regression line is: $t - \bar{t} = b(v - \bar{v})$ $\Rightarrow t - 112.8 = 167(v - 0.6)$ $\Rightarrow t = 167v + 12.6$	B1 for $\bar{t}$ and $\bar{v}$ used (SOI) M1 for attempt at gradient (*b*) A1 for 167 CAO M1 for equation of line A1 FT	**5**
(ii)	(*A*) For 0.5 litres, predicted time = $= 167 \times 0.5 + 12.6 = 96.1$ seconds (*B*) For 1.5 litres, predicted time = $= 167 \times 1.5 + 12.6 = 263.1$ seconds Any valid relevant comment relating to each prediction such as eg: 'First prediction is fairly reliable as it is interpolation and the data is a good fit' 'Second prediction is less certain as it is an extrapolation'	M1 for at least one prediction attempted A1 for both answers (FT their equation if b>0) NB for reading predictions off the graph only award A1 if accurate to nearest whole number E1 (first prediction) E1 (second prediction)	**4**
(iii)	The *v*-coefficient is the number of additional seconds required for each extra litre of water	E1 for indication of rate wrt *v* E1 *dep* for specifying ito units	**2**
(iv)	$v = 0.8 \Rightarrow$ predicted $t = 167 \times 0.8 + 12.6 = 146.2$ Residual = 156 – 146.2 = 9.8 $v = 1.0 \Rightarrow$ predicted $t = 167 \times 1.0 + 12.6 = 179.6$ Residual = 172 – 179.6 = –7.6	M1 for either prediction M1 for either subtraction A1 *CAO* for absolute value of both residuals B1 for both signs correct.	**4**
(v)	The residuals can be measured by finding the vertical distance between the plotted point and the regression line. The sign will be negative if the point is below the regression line (and positive if above).	E1 for distance E1 for vertical E1 for sign	**3**
			18

Question 2

(a) (i)	$X \sim N(28,16)$ $P(24 < X < 33) = P\left(\frac{24-28}{4} < Z < \frac{33-28}{4}\right)$ $= P(-1 < Z < 1.25)$ $= \Phi(1.25) - (1 - \Phi(1))$ $= 0.8944 - (1 - 0.8413)$ $= 0.8944 - 0.1587$ $= 0.7357$ (4 s.f.) *or* 0.736 (to 3 s.f.)	M1 for standardizing A1 for 1. 25 and -1 M1 for prob. with tables and correct structure A1 CAO (min 3 s.f., to include use of difference column)	**4**
(ii)	$25000 \times 0.7357 \times 0.1 = £1839$ $25000 \times 0.1587 \times 0.05 = £198$ Total = £1839 + £198 = £2037	M1 for either product, (with or without price) M1 for sum of both products with price A1 CAO awrt £2040	**3**
(iii)	$X \sim N(k, 16)$ From tables $\Phi^{-1}(0.95) = 1.645$ $\frac{33-k}{4} = 1.645$ $33 - k = 1.645 \times 4$ $k = 33 - 6.58$ $k = 26.42$ (4 s.f.) *or* 26.4 (to 3 s.f.)	B1 for ±1.645 seen M1 for correct equation in *k* with positive z-value A1 *CAO*	**3**
(b) (i)	H_0: $\mu = 0.155$; H_1: $\mu > 0.155$ Where μ denotes the mean weight in kilograms of the population of onions of the new variety	B1 for both correct & ito μ B1 for definition of μ	**2**
(ii)	Mean weight = 4.77/25 = 0.1908 Test statistic = $\frac{0.1908 - 0.155}{\sqrt{0.005}/\sqrt{25}} = \frac{0.0358}{0.01414}$ $= 2.531$ 1% level 1-tailed critical value of z = 2.326 2.531 > 2.236 so significant. There is sufficient evidence to reject H_0 It is reasonable to conclude that the new variety has a higher mean weight.	B1 M1 must include √25 A1FT B1 for 2.326 M1 For sensible comparison leading to a conclusion A1 for correct, consistent conclusion in words and in context	**6**
			18

Question 3

(i)	Mean = $\frac{\Sigma xf}{n} = \frac{0+20+12+3}{80} = \frac{35}{80}$ (= 0.4375)	B1 for mean ***NB answer given***	**1**
(ii)	Variance = 0.6907^2 = 0.4771 So Poisson distribution may be appropriate, since mean is close to variance	B1 for variance E1*dep on squaring s*	**2**
(iii)	P(*X* = 1) = $e^{-0.4375}\frac{0.4375^1}{1!}$ = 0.282 (3 s.f.) *Either:* Thus the expected number of 1's is 22.6 which is reasonably close to the observed value of 20. *Or*: This probability compares reasonably well with the relative frequency 0.25	M1 for probability calc. M0 for tables unless interpolated (0.2813) A1 B1 for expectation of 22.6 or r.f. of 0.25 E1 for comparison	**4**
(iv)	$\lambda = 8\times0.4375 = 3.5$ Using tables: P(*X* ≥ 12) = 1 – P(*X* ≤ 11) = 1 – 0.9997 = 0.0003	B1 for mean (SOI) M1 for using tables to find 1 – P(*X* ≤ 11) A1 FT	**3**
(v)	The probability of at least 12 free repairs is very low, so the model is not appropriate. This is probably because the mean number of free repairs in the launderette will be much higher since the machines will get much more use than usual.	E1 for 'at least 12' E1 for very low E1	**3**
(vi)	*(A)* $\lambda = 0.4375 + 0.15 = 0.5875$ P(*X* = 3) = $e^{-0.5875}\frac{0.5875^3}{3!}$ = 0.0188 (3 s.f.)	B1 for mean (SOI) M1 A1	**3**
	(B) P(Drier needs 1) = $e^{-0.15}\frac{0.15^1}{1!}$ = 0.129 P(Each needs just 1) = 0.282 ×0.129 = 0.036	B1 for 0.129 (SOI) B1FT for 0.036	**2**
			18

Question 4

<table>
<tr><td>(i)</td><td>H_0: no association between ambition and home location;
H_1: some association between ambition and home location;
<table>
<tr><td colspan="2" rowspan="2">Observed</td><td colspan="2">Home location</td></tr>
<tr><td>City</td><td>Non-city</td></tr>
<tr><td rowspan="2">Ambition</td><td>Good results</td><td>102</td><td>147</td></tr>
<tr><td>Other</td><td>75</td><td>156</td></tr>
</table>
<table>
<tr><td colspan="2" rowspan="2">Expected</td><td colspan="2">Home location</td></tr>
<tr><td>City</td><td>Non-city</td></tr>
<tr><td rowspan="2">Ambition</td><td>Good results</td><td>91.82</td><td>157.18</td></tr>
<tr><td>Other</td><td>85.18</td><td>145.82</td></tr>
</table>
<table>
<tr><td colspan="2" rowspan="2">Contribution to the test statistic</td><td colspan="2">Home location</td></tr>
<tr><td>City</td><td>Non-city</td></tr>
<tr><td rowspan="2">Ambition</td><td>Good results</td><td>1.129</td><td>0.659</td></tr>
<tr><td>Other</td><td>1.217</td><td>0.711</td></tr>
</table>
$X^2 = 3.716$
Refer to χ_1^2
Critical value at 5% level = 3.841
Result is not significant
There is insufficient evidence to conclude that there is any association between home location and ambition.
NB if H_0 H_1 reversed, or ‘correlation’ mentioned, do not award first B1or final B1 or final E1</td><td>B1 in context

M1 A1 for attempt at expected values

M1 for valid attempt at $(O-E)^2/E$

A1CAO for X^2

B1 for 1 dof SOI
B1 CAO for cv
B1 dep on attempt at cv
E1 conclusion in context</td><td>1

4

4</td></tr>
<tr><td>(ii)
(A)</td><td>Expected Country, Results = 249 * 156 / 480 = 80.93
Expected Country, Other = 231 * 156 / 480 = 75.08</td><td>B1
B1</td><td>2</td></tr>
<tr><td>(B)</td><td>Refer to χ_2^2
Critical value at 5% level = 5.991
Result is significant
There is evidence to conclude that there is association between home location and ambition.</td><td>B1 for 2 dof SOI
B1 CAO for cv
E1 for conclusion in context</td><td>3</td></tr>
<tr><td>(C)</td><td>‘Country’ students are much less likely than city or town to have ‘Results’ as their main ambition. Low contributions show that city and town students do not appear to differ markedly in their ambitions.</td><td>E1 for correct obs^n for ‘Country’
E1 for additional correct observation (must refer to contributions)</td><td>2</td></tr>
<tr><td>(iii)</td><td>Conclusion in (i) is valid if only categorizing home location into city and non-city. However if non-city is subdivided into town and country this additional subdivision gives the data more precision and allows the relationship in part (ii) (C) to be revealed.</td><td>E1

E1</td><td>2</td></tr>
<tr><td></td><td></td><td></td><td>18</td></tr>
</table>

Question 1

(i)	$X \sim N(11,3^2)$ $P(X < 10) = P\left(Z < \frac{10-11}{3}\right)$ $= P(Z < -0.333)$ $= \Phi(-0.333) = 1 - \Phi(0.333)$ $= 1 - 0.6304 = 0.3696$	M1 for standardizing M1 for use of tables with their *z*-value M1 *dep* for correct tail A1CAO (must include use of differences)	**4**
(ii)	P(3 of 8 less than ten) $= \binom{8}{3} \times 0.3696^3 \times 0.6304^5 = 0.2815$	M1 for coefficient M1 for $0.3696^3 \times 0.6304^5$ A1 FT (min 2sf)	**3**
(iii)	$\mu = np = 100 \times 0.3696 = 36.96$ $\sigma^2 = npq = 100 \times 0.3696 \times 0.6304 = 23.30$ $Y \sim N(36.96, 23.30)$ $P(Y \geq 50) = P\left(Z > \frac{49.5 - 36.96}{\sqrt{23.30}}\right)$ $= P(Z > 2.598) = 1 - \Phi(2.598) = 1 - 0.9953$ $= 0.0047$	M1 for Normal approximation with correct (FT) parameters B1 for continuity corr. M1 for standardizing and using correct tail A1 **CAO (**FT 50.5 or omitted CC)	**4**
(iv)	H_0: $\mu = 11$; H_1: $\mu > 11$ Where μ denotes the mean time taken by the new hairdresser	B1 for H_0, as seen. B1 for H_1, as seen. B1 for definition of μ	**3**
(v)	Test statistic $= \frac{12.34 - 11}{3/\sqrt{25}} = \frac{1.34}{0.6}$ $= 2.23$ 5% level 1 tailed critical value of z = 1.645 2.23 > 1.645, so significant. There is sufficient evidence to reject H_0 It is reasonable to conclude that the new hairdresser does take longer on average than other staff.	M1 must include $\sqrt{25}$ A1 (FT their μ) B1 for 1.645 M1 for sensible comparison leading to a conclusion A1 for conclusion in words in context (FT their μ)	**5**
			19

Question 2

(i)

x	2.61	2.73	2.87	2.96	3.05	3.14	3.17	3.24	3.76	4.1
y	3.2	2.6	3.5	3.1	2.8	2.7	3.4	3.3	4.4	4.1
Rank x	10	9	8	7	6	5	4	3	2	1
Rank y	6	10	3	7	8	9	4	5	1	2
d	4	-1	5	0	-2	-4	0	-2	1	-1
d^2	16	1	25	0	4	16	0	4	1	1

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\times 68}{10\times 99}$$

= 0.588 (to 3 s.f.) [*allow* 0.59 to 2 s.f.]

M1 for ranking (allow all ranks reversed)

M1 for d^2

A1 for $\Sigma d^2 = 68$

M1 for method for r_s

A1 f.t. for $|r_s| < 1$
NB No ranking scores zero

5

(ii)

H_0: no association between x and y

H_1: positive association between x and y

Looking for positive association (one–tail test): critical value at 5% level is 0.5636

Since 0.588> 0.5636, there is sufficient evidence to reject H_0,
i.e. conclude that there is positive association between true weight x and estimated weight y.

B1 for H_0, in context.

B1 for H_1, in context.

NB H_0 H_1 <u>not</u> ito ρ

B1 for $\pm$ 0.5636

M1 for sensible comparison with c.v., provided $|r_s| < 1$
A1 for conclusion in words & in context, f.t. their r_s and sensible cv

5

(iii)

$\Sigma x = 31.63$, $\Sigma y = 33.1$, $\Sigma x^2 = 101.92$, $\Sigma y^2 = 112.61$,
$\Sigma xy = 106.51$.

$$S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x\Sigma y = 106.51 - \tfrac{1}{10}\times 31.63\times 33.1$$

$= 1.8147$

$$S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 101.92 - \tfrac{1}{10}\times 31.63^2 = 1.8743$$

$$S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 112.61 - \tfrac{1}{10}\times 33.1^2 = 3.049$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{1.8147}{\sqrt{1.8743\times 3.049}} = 0.759$$

M1 for method for S_{xy}

M1 for method for at least one of S_{xx} or S_{yy}

A1 for at least one of S_{xy}, S_{xx}, S_{yy} correct.

M1 for structure of r

A1 (awrt 0.76)

5

(iv)

Use of the PMCC is better since it takes into account not just the ranking but the actual value of the weights. Thus it has more information than Spearman's and will therefore provide a more discriminatory test.

Critical value for rho = 0.5494
PMCC is very highly significant whereas Spearman's is only just significant.

E1 for has values, not just ranks
E1 for contains more information
Allow alternatives.
B1 for a cv
E1 dep

4

19

Question 3

(i)	(*A*) P(*X* = 1) = 0.1712 – 0.0408 = 0.1304 *OR* $= e^{-3.2}\dfrac{3.2^1}{1!}$ = 0.1304 (*B*) P(*X* ≥ 6) = 1 – P(*X* ≤ 5) = 1 – 0.8946 = 0.1054	M1 for tables A1 (2 s.f. WWW) M1 A1	**4**
(ii)	(*A*) λ = 3.2 ÷ 5 = 0.64 P(*X* =1) = $e^{-0.64}\dfrac{0.64^1}{1!}$ = 0.3375 (*B*) P(exactly one in each of 5 mins) = 0.3375^5 = 0.004379	B1 for mean (SOI) M1 for probability A1 B1 (FT to at least 2 s.f.)	**4**
(iii)	Mean no. of calls in 1 hour = 12 × 3.2 = 38.4 Using Normal approx. to the Poisson, *X* ~ N(38.4, 38.4) P(*X* ≤ 45.5) = $P\left(Z \le \dfrac{45.5-38.4}{\sqrt{38.4}}\right)$ = P(*Z* ≤ 1.146) = Φ(1.146) = 0.874 (3 s.f.)	B1 for Normal approx. with correct parameters (SOI) B1 for continuity corr. M1 for probability using correct tail A1 CAO, **(**but FT 44.5 or omitted CC)	**4**
(iv)	(*A*) Suitable arguments for/against each assumption: (*B*) Suitable arguments for/against each assumption:	E1, E1 E1, E1	**4**
			16

Question 4

(i)	H_0: no association between age group and sex; H_1: some association between age group and sex;	B1 (in context)	
	(see tables below)	M1 A1 for expected values (to 2dp) M1 for valid attempt at $(O-E)^2/E$ M1dep for summation	**6**
	$X^2 = 11.84$	A1CAO for X^2	**4**
	Refer to Ξ_2^2 Critical value at 5% level = 5.991 Result is significant There is some association between age group and sex . NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1	B1 for 2 deg of f B1 CAO for cv B1 dep on their cv & X^2 E1 (conclusion in context)	
(ii)	The analysis suggests that there are more females in the under 40 age group and less in the 50 and over age group than would be expected if there were no association. The reverse is true for males. Thus these data do support the suggestion.	E1 E1 E1dep (on at least one of the previous E1s)	**3**
(iii)	Binomial(300, 0.03) soi n = 300, p = 0.03 so *EITHER:* use Poisson approximation to Binomial with $\lambda = np = 9$ Using tables: $P(X \geq 12) = 1 - P(X \leq 11)$ $= 1 - 0.8030 = 0.197$ *OR:* use Normal approximation N(9, 8.73) $P(X > 11.5) = P\left(Z > \frac{11.5-9}{\sqrt{8.73}}\right)$ $= P(Z > 0.846)) = 1 - 0.8012 = 0.199$	B1 CAO *EITHER:* B1 for Poisson B1dep for Poisson(9) M1 for using tables to find $1 - P(X \leq 11)$ A1 *OR:* B1 for Normal B1dep for parameters M1 for using tables with correct tail (cc not required for M1) A1	**5**
			18

Expected		Sex: Male	Sex: Female	**Row totals**
Age group	Under 40	81.84	42.16	**124**
	40 – 49	73.92	38.08	**112**
	50 and over	42.24	21.76	**64**
Column totals		**198**	**102**	**300**

Contribution to test statistic		Sex: Male	Sex: Female
Age group	Under 40	1.713	3.325
	40 – 49	0.059	0.114
	50 and over	2.255	4.378

4767 Statistics 2

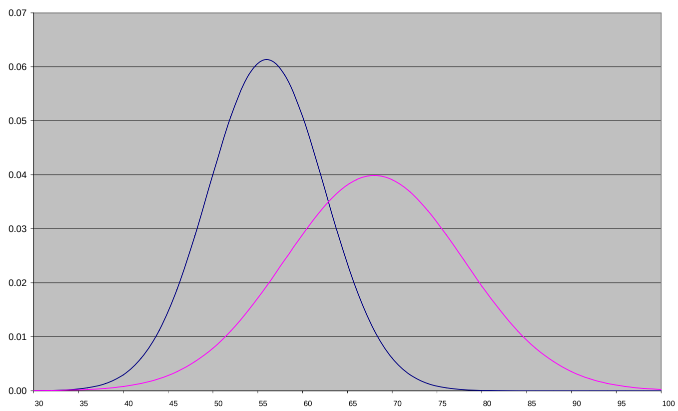
Question 1

(i)	*x* is independent, *y* is dependent since the values of *x* are chosen by the student but the values of *y* are dependent on *x*	B1 E1 dep E1 dep	**3**
(ii)	$\bar{x} = 2.5$, $\bar{y} = 80.63$ $b = \frac{Sxy}{Sxx} = \frac{2530.3 - 30 \times 967.6/12}{90 - 30^2/12} = \frac{111.3}{15} = 7.42$ OR $b = \frac{2530.3/12 - 2.50 \times 80.63}{90/12 - 2.50^2} = \frac{9.275}{1.25} = 7.42$ Hence least squares regression line is: $y - \bar{y} = b(x - \bar{x})$ $\Rightarrow y - 80.63 = 7.42(x - 2.5)$ $\Rightarrow y = 7.42x + 62.08$	B1 for $\bar{x}$ and $\bar{y}$ used (SOI) M1 for attempt at gradient (*b*) A1 for 7.42 **cao** M1 for equation of line A1 FT (*b*>0) for complete equation	**5**
(iii)	*(A)* For *x* = 1.2, predicted growth = 7.42 × 1.2 + 62.08 = 71.0 *(B)* For *x* = 4.3, predicted growth = 7.42 × 4.3 + 62.08 = 94.0 Valid relevant comments relating to the predictions such as : Comment re interpolation/extrapolation Comment relating to the fact that *x* = 4.3 is only just beyond the existing data. Comment relating to size of residuals near each predicted value (need not use word 'residual')	M1 for at least one prediction attempted. A1 for both answers (FT their equation if *b*>0) E1 (first comment) E1 (second comment)	**4**
(iv)	$x = 3 \Rightarrow$ predicted $y = 7.42 \times 3 + 62.08 = 84.3$ Residual = 80 − 84.3 = −4.3	M1 for prediction M1 for subtraction A1 FT (*b*>0)	**3**
(v)	This point is a long way from the regression line. The line may be valid for the range used in the experiment but then the relationship may break down for higher concentrations, or the relationship may be non linear.	E1 E1 for valid in range E1 for *either* 'may break down' *or* 'could be non linear' *or* other relevant comment	**3**
			18

Question 2

(i)	Binomial (94,0.1)	B1 for binomial B1 dep for parameters	**2**
(ii)	*n* is large and *p* is small	B1, B1 Allow appropriate numerical ranges	**2**
(iii)	$\lambda = 94 \times 0.1 = 9.4$ *(A)* $P(X = 4) = e^{-9.4} \dfrac{9.4^4}{4!} = 0.0269$ (3 s.f.) or from tables = 0.0429 − 0.0160 = 0.0269 ***cao*** *(B)* Using tables: $P(X \geq 4) = 1 - P(X \leq 3)$ $= 1 - 0.0160 = 0.9840$ ***cao***	B1 for mean M1 for calculation or use of tables A1 M1 for attempt to find $P(X \geq 4)$ A1 **cao**	**5**
(iv)	P(sufficient rooms throughout August) $= 0.9840^{31} = 0.6065$	M1 A1 FT	**2**
(v)	(*A*) 31 × 94 = 2914 Binomial (2914,0.1)	B1 for binomial B1 dep, for parameters	**2**
	(*B*)Use Normal approx with $\mu = np = 2914 \times 0.1 = 291.4$ $\sigma^2 = npq = 2914 \times 0.1 \times 0.9 = 262.26$ $P(X \leq 300.5) = P\left(Z \leq \dfrac{300.5 - 291.4}{\sqrt{262.26}}\right)$ $= P(Z \leq 0.5619) = \Phi(0.5619) = 0.7130$	B1 B1 B1 for continuity corr. M1 for probability using correct tail A1 **cao,** (but FT wrong or omitted CC)	**5**
			18

Question 3

(i)	$X \sim N(56, 6.5^2)$ P(52.5 < X < 57.5) = $P\left(\frac{52.5-56}{6.5} < Z < \frac{57.5-56}{6.5}\right)$ = P(−0.538 < Z < 0.231) = Φ(0.231) − (1 − Φ(0.538)) = 0.5914 − (1 − 0.7046) = 0.5914 − 0.2954 = 0.2960 (4 s.f.) *or* 0.296 (to 3 s.f.)	M1 for standardizing A1 for -0.538 and 0.231 M1 for prob. with tables and correct structure A1 CAO (min 3 s.f., to include use of difference column)	**4**
(ii)	P(5-year-old < 62) = $P\left(Z < \frac{62-56}{6.5}\right)$ = Φ(0.923) = 0.8220 P(young adult < 62) = $P\left(Z < \frac{62-68}{10}\right)$ = Φ(−0.6) = 1 − 0.7257 = 0.2743 P(One over, one under) = 0.8220 × 0.7257 + 0.1780 × 0.2743 = 0.645	B1 for 0.8220 or 0.1780 B1 for 0.2743 or 0.7257 M1 for either product M1 for sum of both products A1 CAO	**5**
(iii)		G1 for shape G1 for means, shown explicitly or by scale G1 for lower max height in young adults G1 for greater variance in young adults	**4**
(iv)	$Y \sim N(82, \sigma^2)$ From tables $\Phi^{-1}(0.88)$ = 1.175 $\frac{62-82}{\sigma} = -1.175$ −20 = −1.175 σ σ = 17.0	B1 for 1.175 seen M1 for equation in σ with z-value M1 for correct handling of LH tail A1 **cao**	**4**
			17

Question 4

<table>
<tr><td>(i)</td><td>

H_0: no association between sex and subject;
H_1: some association between sex and subject;

OBS	Maths	English	Both	Neither	Row sum
Male	38	19	6	32	**95**
Female	42	55	9	49	**155**
Col sum	**80**	**74**	**15**	**81**	**250**

EXP	Maths	English	Both	Neither	Row sum
Male	30.40	28.12	5.70	30.78	**95**
Female	49.60	45.88	9.30	50.22	**155**
Col sum	**80**	**74**	**15**	**81**	**250**

CONT	Maths	English	Both	Neither
Male	1.900	2.958	0.016	0.048
Female	1.165	1.813	0.010	0.030

$X^2 = 7.94$

Refer to χ_3^2
Critical value at 5% level = 7.815
Result is significant
There is evidence to suggest that there is some association between sex and subject choice.
NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1 or final E1

</td><td>

B1

M1 A2 for expected values
(allow A1 for at least one row or column correct)

M1 for valid attempt at $(O-E)^2/E$
A1
NB These M1 A1 marks cannot be implied by a correct final value of X^2

M1 for summation
A1 **cao** for X^2

B1 for 3 deg of f
B1 CAO for cv

B1

E1

</td><td>

1

7

4

</td></tr>
<tr><td>(ii)</td><td>

H_0: $\mu = 67.4$; H_1: $\mu > 67.4$
Where μ denotes the mean score of the population of students taught with the new method.

Test statistic = $\frac{68.3 - 67.4}{8.9/\sqrt{12}} = \frac{0.9}{2.57}$
= 0.35

10% level 1 tailed critical value of z = 1.282
0.35 < 1.282 so not significant.
There is insufficient evidence to reject H_0
There is insufficient evidence to conclude that the mean score is increased by the new teaching method.

</td><td>

B1 for both correct

B1 for definition of μ

M1

A1 **cao**

B1 for 1.282

M1 for comparison

A1 for conclusion in words and in context

</td><td>

7

</td></tr>
<tr><td></td><td></td><td></td><td>19</td></tr>
</table>

4767 Statistics 2

Question 1

(i)	**EITHER:** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x\Sigma y = 880.1 - \frac{1}{48} \times 781.3 \times 57.8$ $= -60.72$ $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 14055 - \frac{1}{48} \times 781.3^2 = 1337.7$ $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 106.3 - \frac{1}{48} \times 57.8^2 = 36.70$ $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-60.72}{\sqrt{1337.7 \times 36.70}} = -0.274$ **OR:** $\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 880.1/48 - 16.28 \times 1.204$ $= -1.265$ $\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(1337.7/48)} = \sqrt{27.87} = 5.279$ $\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(36.70/48)} = \sqrt{0.7646} = 0.8744$ $r = \frac{\text{cov}(x,y)}{rmsd(x)rmsd(y)} = \frac{-1.265}{5.279 \times 0.8744} = -0.274$	M1 for method for S_{xy} M1 for method for at least one of S_{xx} or S_{yy} A1 for at least one of S_{xy}, S_{xx}, S_{yy} correct M1 for structure of *r* A1 CAO (–0.27 to –0.28) M1 for method for cov (x,y) M1 for method for at least one msd A1 for at least one of cov/msd correct M1 for structure of *r* A1 CAO (–0.27 to –0.28)	**5**
(ii)	H_0: $\rho = 0$ H_1: $\rho < 0$ (one-tailed test) where ρ is the population correlation coefficient For $n = 48$, 5% critical value = 0.2403 Since $\lvert -0.274 \rvert > 0.2403$ we can reject H_0: There is sufficient evidence at the 5% level to suggest that there is negative correlation between education spending and population growth.	B1 for H_0, H_1 in symbols B1 for defining ρ B1FT for critical value M1 for sensible comparison leading to a conclusion A1 for result (FT r<0) E1 FT for conclusion in words	**6**
(iii)	Underlying distribution must be bivariate Normal. If the distribution is bivariate Normal then the scatter diagram will have an elliptical shape.	B1 CAO for bivariate Normal B1 indep for elliptical shape	**2**
(iv)	• Correlation does not imply causation • There could be a third factor • increased growth could cause lower spending. Allow any sensible alternatives, including example of a possible third factor.	E1 E1 E1	**3**
(v)	Advantage – less effort or cost Disadvantage – the test is less sensitive (ie is less likely to detect any correlation which may exist)	E1 E1	**2**
			18

Question 2

(i)	(*A*) P(*X* = 2) = $e^{-0.37}\dfrac{0.37^2}{2!}$ = 0.0473 (*B*) P(*X* > 2) = $1-(e^{-0.37}\dfrac{0.37^2}{2!}+e^{-0.37}\dfrac{0.37^1}{1!}+e^{-0.37}\dfrac{0.37^0}{0!})$ = 1 – (0.0473 + 0.2556 + 0.6907) = 0.0064	M1 A1 (2 s.f.) M1 for P(*X* = 1) and P(*X* = 0) M1 for complete method A1 ***NB Answer given***	**5**
(ii)	P(At most one day more than 2) = $\binom{30}{1}\times 0.9936^{29}\times 0.0064+0.9936^{30}$ = = 0.1594 + 0.8248 = 0.9842	M1 for coefficient M1 for $0.9936^{29}\times 0.0064$ M1 for 0.993630 A1 CAO (min 2sf)	**4**
(iii)	λ = 0.37 ×10 = 3.7 P(*X* > 8) = 1 – 0.9863 = 0.0137	B1 for mean (SOI) M1 for probability A1 CAO	**3**
(iv)	Mean no. per 1000ml = 200 × 0.37 = 74 Using Normal approx. to the Poisson, *X* ~ N(74, 74) P(*X* > 90) = $P\left(Z>\dfrac{90.5-74}{\sqrt{74}}\right)$ = P(*Z* >1.918) = 1 – Φ(1.918) = 1 – 0.9724 = 0.0276	B1 for Normal approx. with correct parameters (SOI) B1 for continuity corr. M1 for probability using correct tail A1 CAO (min 2 s.f.)**, (**but FT wrong or omitted CC)	**4**
(v)	P(questionable) = 0.0064 × 0.0137 ×0.0276 = 2.42×10^{-6}	M1 A1 CAO	**2**
			18

Question 3

(i)	$X \sim N(27500, 4000^2)$ $P(X > 25000) = P\left(Z > \frac{25000 - 27500}{4000}\right)$ $= P(Z > -0.625)$ $= \Phi(0.625) = 0.7340$ (3 s.f.)	M1 for standardising A1 for -0.625 M1 *dep* for correct tail A1CAO (must include use of differences)	**4**
(ii)	P(7 of 10 last more than 25000) $= \binom{10}{7} \times 0.7340^7 \times 0.2660^3 = 0.2592$	M1 for coefficient M1 for $0.7340^7 \times 0.2660^3$ A1 FT (min 2sf)	**3**
(iii)	From tables $\Phi^{-1}(0.99) = 2.326$ $\frac{k - 27500}{4000} = -2.326$ $x = 27500 - 2.326 \times 4000 = 18200$	B1 for 2.326 seen M1 for equation in *k* and negative z-value A1 CAO for awrt 18200	**3**
(iv)	H_0: $\mu = 27500$; H_1: $\mu > 27500$ Where μ denotes the mean lifetime of the new tyres.	B1 for use of 27500 B1 for both correct B1 for definition of μ	**3**
(v)	Test statistic $= \frac{28630 - 27500}{4000/\sqrt{15}} = \frac{1130}{1032.8}$ $= 1.094$ 5% level 1 tailed critical value of $z = 1.645$ $1.094 < 1.645$ so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that the new tyres last longer.	M1 must include $\sqrt{15}$ A1 FT B1 for 1.645 M1 *dep* for a sensible comparison leading to a conclusion A1 for conclusion in words in context	**5**
			18

Question 4

(i)	H_0: no association between location and species. H_1: some association between location and species.	B1 for both	**1**
(ii)	Expected frequency = 38/160 × 42 = 9.975 Contribution = $(3 – 9.975)^2$ / 9.975 = 4.8773	M1 A1 M1 for valid attempt at $(O-E)^2/E$ A1 ***NB Answer given***	**4**
(iii)	Refer to χ_4^2 Critical value at 5% level = 9.488 Test statistic X^2 = 32.85 Result is significant There appears to be some association between location and species NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1	B1 for 4 deg of f(seen) B1 CAO for cv M1 Sensible comparison, using 32.85, leading to a conclusion A1 for correct conclusion (FT their c.v.) E1 conclusion in context	**5**
(iv)	• Limpets appear to be distributed as expected throughout all locations. • Mussels are much more frequent in exposed locations and much less in pools than expected. • Other shellfish are less frequent in exposed locations and more frequent in pools than expected.	E1 E1, E1 E1, E1	**5**
(v)	$\frac{24}{53} \times \frac{32}{65} \times \frac{16}{42} = 0.0849$	M1 for one fraction M1 for product of all 3 A1 CAO	**3**
			18

4767 Statistics 2

Question 1

(i)

x	18	43	52	94	98	206	784	1530
y	1.15	0.97	1.26	1.35	1.28	1.42	1.32	1.64
Rank x	1	2	3	4	5	6	7	8
Rank y	2	1	3	6	4	7	5	8
d	-1	1	0	-2	1	-1	2	0
d^2	1	1	0	4	1	1	4	0

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\times 12}{8\times 63}$$

= 0.857 (to 3 s.f.) [*allow* 0.86 to 2 s.f.]

M1 for attempt at ranking (allow all ranks reversed)

M1 for d^2

A1 for $\Sigma d^2 = 12$

M1 for method for r_s

A1 f.t. for $|r_s| < 1$

NB No ranking scores zero

5

(ii)

H_0: no association between X and Y in the population

H_1: some association between X and Y in the population

Two tail test critical value at 5% level is 0.7381

Since 0.857> 0. 7381, there is sufficient evidence to reject H_0,

i.e. conclude that the evidence suggests that there is association between population size X and average walking speed Y.

B1 for H_0

B1 for H_1

B1 for population SOI

NB H_0 H_1 <u>not</u> ito ρ

B1 for ± 0.7381

M1 for sensible comparison with c.v., provided $|r_s| < 1$

A1 for conclusion in words f.t. their r_s and sensible cv

6

(iii)

$\bar{t} = 45$, $\bar{w} = 2.2367$

$$b = \frac{S_{tw}}{S_{tt}} = \frac{584.6 - 270\times 13.42/6}{13900 - 270^2/6} = \frac{-19.3}{1750} = -0.011$$

$$\text{OR } b = \frac{584.6/6 - 45\times 2.2367}{13900/6 - 45^2} = \frac{-3.218}{291.6667} = -0.011$$

hence least squares regression line is:

$$w - \bar{w} = b(t - \bar{t})$$

$$\Rightarrow w - 2.2367 = -0.011(t - 45)$$

$$\Rightarrow w = -0.011t + 2.73$$

B1 for $\bar{t}$ and $\bar{w}$ used (SOI)

M1 for attempt at gradient (b)

A1 CAO for -0.011

M1 for equation of line

A1 FT for complete equation

5

(iv)	(*A*) For $t = 80$, predicted speed $= -0.011 \times 80 + 2.73 = 1.85$ (*B*) The relationship relates to adults, but a ten year old will not be fully grown so may walk more slowly. NB Allow E1 for comment about extrapolation not in context	M1 A1 FT provided b < 0 E1 extrapolation o.e. E1 sensible contextual comment	**4**
		TOTAL	**20**

Question 2

(i)	Binomial(5000,0.0001)	B1 for binomial B1 dep, for parameters	**2**
(ii)	n is large and p is small $\lambda = 5000 \times 0.0001 = 0.5$	B1, B1 (Allow appropriate numerical ranges) B1	**3**
(iii)	$P(X \geq 1) = 1 - e^{-0.5}\frac{0.5^0}{0!} = 1 - 0.6065 = 0.3935$ or from tables $= 1 - 0.6065 = 0.3935$	M1 for correct calculation or correct use of tables A1	**2**
(iv)	P(9 of 20 contain at least one) $= \binom{20}{9} \times 0.3935^9 \times 0.6065^{11}$ $= 0.1552$	M1 for coefficient M1 for $p^9 \times (1-p)^{11}$, p from part (iii) A1	**3**
(v)	Expected number $= 20 \times 0.3935 = 7.87$	M1 A1 FT	**2**
(vi)	Mean $= \frac{\Sigma xf}{n} = \frac{7+4}{20} = \frac{11}{20} = 0.55$ Variance $= \frac{1}{n-1}\left(\Sigma fx^2 - n\bar{x}^2\right)$ $= \frac{1}{19}\left(15 - 20 \times 0.55^2\right) = 0.471$	B1 for mean M1 for calculation A1 CAO	**3**
(vii)	Yes, since the mean is close to the variance, and also as the expected frequency for 'at least one', i.e. 7.87, is close to the observed frequency of 9.	B1 E1 for sensible comparison B1 for observed frequency = 7 + 2 = 9	**3**
		TOTAL	**18**

Question 3

(i)	(*A*) P($X < 120$) = $\text{P}\left(Z < \frac{120-115.3}{21.9}\right)$ = P($Z < 0.2146$) = $\Phi(0.2146) = 0.5849$	M1 for standardizing A1 for $z = 0.2146$ A1 CAO (min 3 sf, to include use of difference column)	**3**
	(*B*) P($100 < X < 110$) = $\text{P}\left(\frac{100-115.3}{21.9} < Z < \frac{110-115.3}{21.9}\right)$ = P($-0.6986 < Z < -0.2420$) = $\Phi(0.6986) - \Phi(0.2420)$ = $0.7577 - 0.5956$ = 0.1621	M1 for standardizing both 100 & 110 M1 for correct structure in calc^n A1 CAO	**3**
	(*C*) From tables $\Phi^{-1}(0.1) = -1.282$ $\frac{k-115.3}{21.9} = -1.282$ $k = 115.3 - 1.282 \times 21.9 = 87.22$	B1 for ±1.282 seen M1 for equation in *k* and negative z-value A1 CAO	**3**
(ii)	From tables, $\Phi^{-1}(0.70) = 0.5244$, $\Phi^{-1}(0.15) = -1.036$ $180 = \mu + 0.5244\,\sigma$ $140 = \mu - 1.036\,\sigma$ $40 = 1.5604\,\sigma$ $\sigma = 25.63$, $\mu = 166.55$	B1 for 0.5244 or ±1.036 seen M1 for at least one equation in μ and σ and Φ^{-1} value M1 dep for attempt to solve two equations A1 CAO for both	**4**
(iii)	$\Phi^{-1}(0.975) = 1.96$ $a = 166.55 - 1.96 \times 25.63 = 116.3$ $b = 166.55 + 1.96 \times 25.63 = 216.8$	B1 for ±1.96 seen M1 for either equation A1 A1 [Allow other correct intervals]	**4**
		TOTAL	**17**

Question 4

(i)	H_0: no association between growth and type of plant; H_1: some association between growth and type of plant;	B1 (in context)	
	<table><tr><td>EXPECTED</td><td>Good</td><td>Average</td><td>Poor</td></tr><tr><td>Coriander</td><td>12.10</td><td>24.93</td><td>17.97</td></tr><tr><td>Aster</td><td>10.56</td><td>21.76</td><td>15.68</td></tr><tr><td>Fennel</td><td>10.34</td><td>21.31</td><td>15.35</td></tr></table>	M1 A2 for expected values (to 2 dp) (allow A1 for at least one row or column correct)	
	<table><tr><td>CONTRIBUTION</td><td>Good</td><td>Average</td><td>Poor</td></tr><tr><td>Coriander</td><td>0.0008</td><td>0.3772</td><td>0.4899</td></tr><tr><td>Aster</td><td>1.2002</td><td>0.6497</td><td>3.4172</td></tr><tr><td>Fennel</td><td>1.2955</td><td>0.0226</td><td>1.2344</td></tr></table>	M1 for valid attempt at $(O-E)^2/E$ A1 for all correct NB These M1A1 marks cannot be implied by a correct final value of X^2	
	$X^2 = 8.69$	M1 for summation A1 for X^2 CAO	
	Refer to χ_4^2	B1 for 4 d.o.f.	
	Critical value at 5% level = 9.488	B1 CAO for cv	
	Result is not significant There is not enough evidence to suggest that there is some association between reported growth and type of plant; NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final A1	M1 A1	**12**
(ii)	Test statistic = $\dfrac{49.2-47}{8.5/\sqrt{50}} = \dfrac{2.2}{1.202} = 1.830$	M1 correct denominator A1	
	1% level 1 tailed critical value of z = 2.326	B1 for 2.326	
	1.830 < 2.326 so not significant. There is not sufficient evidence to reject H_0	M1 (dep on first M1) for sensible comparison leading to a conclusion	
	There is insufficient evidence to conclude that the flowers are larger.	A1 for fully correct conclusion in words in context	**5**
		TOTAL	**17**

4767 Statistics 2

Question 1

(i)	**EITHER:** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x\Sigma y = 316345 - \frac{1}{50} \times 2331.3 \times 6724.3$ $= 2817.8$ $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 111984 - \frac{1}{50} \times 2331.3^2 = 3284.8$ $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 921361 - \frac{1}{50} \times 6724.3^2 = 17036.8$ $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{2817.8}{\sqrt{3284.8 \times 17036.8}} = 0.377$ **OR:** $\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 316345/50 - 46.626 \times 134.486$ $= 56.356$ $\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(3284.8/50)} = \sqrt{65.696} = 8.105$ $\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(17036.8/50)} = \sqrt{340.736} = 18.459$ $r = \frac{\text{cov}(x,y)}{rmsd(x)rmsd(y)} = \frac{56.356}{8.105 \times 18.459} = 0.377$	M1 for method for S_{xy} M1 for method for at least one of S_{xx} or S_{yy} A1 for at least one of S_{xy}, S_{xx} or S_{yy} correct M1 for structure of r A1 (AWRT 0.38) M1 for method for cov (x,y) M1 for method for at least one msd A1 for at least on of cov(x,y), rmsd(x) or rmsd(y) correct M1 for structure of r A1 (AWRT 0.38)	5
(ii)	H_0: $\rho = 0$ H_1: $\rho \neq 0$ (two-tailed test) where ρ is the population correlation coefficient For $n = 50$, 5% critical value = 0.2787 Since 0.377 > 0.2787 we can reject H_0: There is sufficient evidence at the 5% level to suggest that there is correlation between oil price and share cost	B1 for H_0, H_1 in symbols B1 for defining ρ B1FT for critical value M1 for sensible comparison leading to a conclusion A1 for result B1 FT for conclusion in context	6
(iii)	Population The scatter diagram has a roughly elliptical shape, hence the assumption is justified.	B1 B1 elliptical shape E1 conclusion	3
(iv)	Because the alternative hypothesis should be decided without referring to the sample data and there is no suggestion that the correlation should be positive rather than negative.	E1 E1	2
		TOTAL	**16**

Question 2

(i)	Meteors are seen randomly and independently There is a uniform (mean) rate of occurrence of meteor sightings	B1 B1	**2**
(ii)	(*A*) *Either* P(*X* = 1) = 0.6268 – 0.2725 = 0.3543 *Or* P(*X* = 1) = $e^{-1.3}\frac{1.3^1}{1!}$ = 0.3543 (*B*) Using tables: P(*X* ≥ 4) = 1 – P(*X* ≤ 3) = 1 – 0.9569 = 0.0431	M1 for appropriate use of tables or calculation A1 M1 for appropriate probability calculation A1	**4**
(iii)	$\lambda = 10 \times 1.3 = 13$ P(*X* = 10) = $e^{-13}\frac{13^{10}}{10!}$ = 0.0859	B1 for mean M1 for calculation A1 CAO	**3**
(iv)	Mean no. per hour = 60 × 1.3 = 78 Normal approx. to the Poisson, *X* ~ N(78, 78) P(*X* ≥ 100) = $P\left(Z > \frac{99.5-78}{\sqrt{78}}\right)$ = P(*Z* >2.434) = 1 – Φ(2.434) = 1 – 0.9926 = 0.0074	B1 for Normal approx. B1 for correct parameters (SOI) B1 for continuity corr. M1 for correct Normal probability calculation using correct tail A1 CAO**, (**but FT wrong or omitted CC)	**5**
(v)	*Either* P(At least one) = $1 - e^{-\lambda}\frac{\lambda^0}{0!} = 1 - e^{-\lambda} \geq 0.99$ $e^{-\lambda} \leq 0.01$ $-\lambda \leq \ln 0.01$, so $\lambda \geq 4.605$ 1.3 *t* ≥ 4.605, so *t* ≥ 3.54 Answer *t* = 4 *Or* *t* = 1, λ = 1.3, P(At least one) = $1 - e^{-1.3}$ = 0.7275 *t* = 2, λ = 2.6, P(At least one) = $1 - e^{-2.6}$ = 0.9257 *t* = 3, λ = 3.9, P(At least one) = $1 - e^{-3.9}$ = 0.9798 *t* = 4, λ = 5.2, P(At least one) = $1 - e^{-5.2}$ = 0.9944 Answer *t* = 4	M1 formation of equation/inequality using P(*X* ≥ 1) = 1 – P(X = 0) with Poisson distribution. A1 for correct equation/inequality M1 for logs A1 for 3.54 A1 for *t* (correctly justified) M1 at least one trial with any value of *t* A1 correct probability. M1 trial with either *t* = 3 or *t* = 4 A1 correct probability of *t* = 3 <u>and</u> *t* = 4 A1 for *t*	**5**
		TOTAL	**19**

Question 3

(i)	$X \sim N(1720, 90^2)$ $P(X < 1700) = P\left(Z < \frac{1700-1720}{90}\right)$ $= P(Z < -0.2222)$ $= \Phi(-0.2222) = 1 - \Phi(0.2222)$ $= 1 - 0.5879$ $= 0.4121$	M1 for standardising A1 M1 use of tables (correct tail) A1CAO **NB ANSWER GIVEN**	**4**
(ii)	P(2 of 4 below 1700) $= \binom{4}{2} \times 0.4121^2 \times 0.5879^2 = 0.3522$	M1 for coefficient M1 for $0.4121^2 \times 0.5879^2$ A1 FT (min 2sf)	**3**
(iii)	Normal approx with $\mu = np = 40 \times 0.4121 = 16.48$ $\sigma^2 = npq = 40 \times 0.4121 \times 0.5879 = 9.691$ $P(X \geq 20) = P\left(Z \geq \frac{19.5-16.48}{\sqrt{9.691}}\right)$ $= P(Z \geq 0.9701) = 1 - \Phi(0.9701)$ $= 1 - 0.8340 = 0.1660$	B1 B1 B1 for correct continuity corr. M1 for correct Normal probability calculation using correct tail A1 CAO**, (**but FT wrong or omitted CC)	**5**
(iv)	H_0: $\mu = 1720$; H_1 is of this form since the consumer organisation suspects that the mean is below 1720 μ denotes the mean intensity of 25 Watt low energy bulbs made by this manufacturer.	B1 E1 B1 for definition of μ	**3**
(v)	Test statistic = $\frac{1703-1720}{90/\sqrt{20}} = \frac{-17}{20.12}$ $= -0.8447$ Lower 5% level 1 tailed critical value of $z = -1.645$ $-0.8447 > -1.645$ so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that the mean intensity of bulbs made by this manufacturer is less than 1720	M1 must include $\sqrt{20}$ A1FT B1 for –1.645 No FT from here if wrong. Must be –1.645 unless it is clear that absolute values are being used. M1 for sensible comparison leading to a conclusion. FT only candidate's test statistic A1 for conclusion in words in context	**5**
		TOTAL	**20**

Question 4

(i)

H_0: no association between type of car and sex;
H_1: some association between type of car and sex;

B1

EXPECTED	Male	Female
Hatchback	83.16	48.84
Saloon	70.56	41.44
People carrier	51.66	30.34
4WD	17.01	9.99
Sports car	29.61	17.39

M1 A2 for expected values (to 2 dp)

(allow A1 for at least one row or column correct)

CONTRIBUTION	Male	Female
Hatchback	1.98	3.38
Saloon	0.59	1.00
People carrier	3.61	6.15
4WD	0.23	0.40
Sports car	1.96	3.33

M1 for valid attempt at $(O-E)^2/E$
A1 for all correct
NB These M1A1 marks cannot be implied by a correct final value of X^2

$X^2 = 22.62$

M1 for summation
A1 for X^2 CAO

Refer to χ_4^2
Critical value at 5% level = 9.488

B1 for 4 deg of f
B1 CAO for cv

22.62 > 9.488
Result is significant
There is evidence to suggest that there is some association between sex and type of car.

M1 sensible comparison leading to a conclusion
A1

NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final A1

12

(ii)

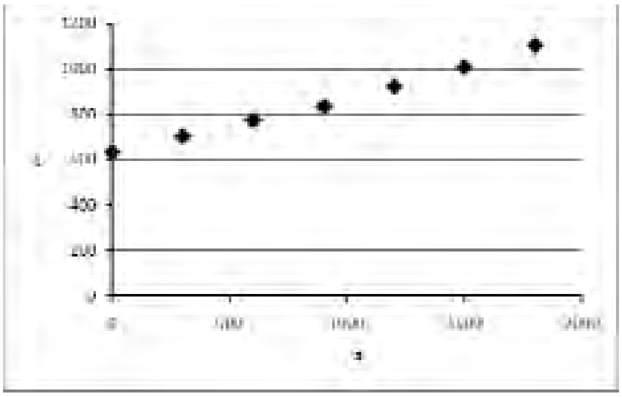
- In hatchbacks, male drivers are more frequent than expected. E1
- In saloons, male drivers are slightly more frequent than expected. E1
- In people carriers, female drivers are much more frequent than expected. E1
- In 4WDs the numbers are roughly as expected E1
- In sports cars, female drivers are more frequent than expected. E1

5

TOTAL **17**

Deleted: ¶

4767 Statistics 2

1	(i)	 	G1 For values of *a* G1 for values of *t* G1 for axes	[3]
	(ii)	*a* is independent, *t* is dependent since the values of *a* are not subject to random variation, but are determined by the runways which the pilot chooses, whereas the values of *t* are subject to random variation.	B1 E1dep E1dep	[3]
	(iii)	$\bar{a} = 900$, $\bar{t} = 855.2$ $b = \frac{S_{at}}{S_{aa}} = \frac{6037800 - 5987 \times 6300/7}{8190000 - 6300^2/7} = \frac{649500}{2520000} = 0.258$ OR $b = \frac{6037800/7 - 855.29 \times 900}{8190000/7 - 900^2} = \frac{92785}{360000} = 0.258$ hence least squares regression line is: $t - \bar{t} = b(a - \bar{a})$ $\Rightarrow t - 855.29 = 0.258\,(a - 900)$ $\Rightarrow t = 0.258a + 623$	B1 for $\bar{a}$ and $\bar{t}$ used (SOI) M1 for attempt at gradient (*b*) A1 for 0.258 **cao** M1 for equation of line A1 FT for complete equation	[5]
	(iv)	(*A*) For *a* = 800, predicted take–off distance $= 0.258 \times 800 + 623 = 829$ (*B*) For *a* = 2500, predicted take–off distance $= 0.258 \times 2500 + 623 = 1268$ Valid relevant comments relating to the predictions such as: First prediction is interpolation so should be reasonable Second prediction is extrapolation and may not be reliable	M1 for at least one prediction attempted A1 for both answers (FT their equation if *b*>0) E1 (first comment) E1 (second comment)	[4]
	(v)	$a = 1200 \Rightarrow$ predicted $t = 0.258 \times 1200 + 623 = 933$ Residual = 923 – 933 = –10 The residual is negative because the observed value is less than the predicted value.	M1 for prediction M1 for subtraction A1 FT E1	[4]
			Total	**[19]**

2	**(i)**	P(1 of 10 is faulty) $= \binom{10}{1} \times 0.02^1 \times 0.98^9 = 0.1667$	M1 for coefficient M1 for probabilities A1	**[3]**
	(ii)	n is large and p is small	B1, B1 Allow appropriate numerical ranges	**[2]**
	(iii)	$\lambda = 150 \times 0.02 = 3$ (*A*) $\mathrm{P}(X = 0) = \mathrm{e}^{-3}\dfrac{3^0}{0!} = 0.0498$ (3 s.f.) or from tables $= 0.0498$	B1 for mean (soi) M1 for calculation or use of tables A1	**[3]**
		(*B*) Expected number = 3 Using tables: $\mathrm{P}(X > 3) = 1 - \mathrm{P}(X \le 3)$ $= 1 - 0.6472 = 0.3528$	B1 expected no = 3 (soi) M1 A1	**[3]**
	(iv)	(*A*) Binomial(2000,0.02)	B1 for binomial B1 for parameters	**[2]**
		(*B*) Use Normal approx with $\mu = np = 2000 \times 0.02 = 40$ $\sigma^2 = npq = 2000 \times 0.02 \times 0.98 = 39.2$ $\mathrm{P}(X \le 50) = \mathrm{P}\left(Z \le \dfrac{50.5 - 40}{\sqrt{39.2}}\right)$ $= \mathrm{P}(Z \le 1.677) = \Phi(1.677) = 0.9532$ NB Poisson approximation also acceptable for full marks	B1 B1 B1 for continuity corr. M1 for probability using correct tail A1 CAO	**[5]**
			Total	**[18]**

3	**(i)**	(*A*) $P(X < 50)$ $= P\left(Z < \frac{50-45.3}{11.5}\right)$ $= P(Z < 0.4087)$ $= \Phi(0.4087)$ $= 0.6585$	M1 for standardising M1 for correct structure of probability calc' A1 CAO inc use of diff tables 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	**[3]**
		(*B*) $P(45.3 < X < 50)$ $= 0.6585 – 0.5$ $=0.1585$	M1 A1	**[2]**
	(ii)	From tables $\Phi^{-1}(0.9) = 1.282$ $\frac{k-45.3}{11.5} = 1.282$ $k = 45.3 + 1.282 \times 11.5 = 60.0$	B1 for 1.282 seen M1 for equation in *k* A1 CAO	**[3]**
	(iii)	P(score = 111) $=P(110.5 < Y < 111.5)$ $= P\left(\frac{110.5-100}{15} < Z < \frac{111.5-100}{15}\right)$ $= P(0.7 < Z < 0.7667)$ $= \Phi(0.7667) – \Phi(0.7)$ $= 0.7784 – 0.7580$ $= 0.0204$	B1 for both continuity corrections M1 for standardising M1 for correct structure of probability calc' A1 CAO	**[4]**
	(iv)	From tables, $\Phi^{-1}(0.3) = –0.5244$, $\Phi^{-1}(0.8) = 0.8416$ $22 = \mu + 0.8416\,\sigma$ $15 = \mu – 0.5244\,\sigma$ $7 = 1.3660\,\sigma$ $\sigma = 5.124$, $\mu = 17.69$	B1 for 0.5244 or 0.8416 seen M1 for at least one equation in z, μ & σ A1 for both correct M1 for attempt to solve two appropriate equations A1 CAO for both	**[5]**
			TOTAL	**[17]**

4	(i)	H_0: no association between size of business and recycling service used. H_1: some association between size of business and recycling service used.	B1 for both	[1]
	(ii)	Expected frequency = 78/285 × 180 = 49.2632 Contribution = $(52 – 49.2632)^2$ / 49.2632 = 0.1520	M1 A1 M1 for valid attempt at $(O\text{-}E)^2/E$ A1 ***NB Answer given*** Allow 0.152	[4]
	(iii)	Test statistic X^2 = 0.6041 Refer to $\mathcal{X}_2^2$ Critical value at 5% level = 5.991 Result is not significant There is no evidence to suggest any association between size of business and recycling service used. NB if H_0 H_1 reversed, or 'correlation' mentioned in part (i), do not award B1in part (i) or E1 in part (iii).	B1 B1 for 2 deg of f(seen) B1 CAO for cv B1 for not significant E1	[5]
	(iv)	H_0: $\mu = 32.8$; H_1: $\mu < 32.8$ Where μ denotes the population mean weight of rubbish in the bins. Test statistic = $\dfrac{30.9-32.8}{3.4/\sqrt{50}} = -\dfrac{1.9}{0.4808} = -3.951$ 5% level 1 tailed critical value of z = –1.645 –3.951 < –1.645 so significant. There is sufficient evidence to reject H_0 There is evidence to suggest that the weight of rubbish in dustbins has been reduced.	B1 for use of 32.8 B1 for both correct B1 for definition of μ M1 must include √50 A1 B1 for ±1.645 M1 for sensible comparison leading to a conclusion A1 for conclusion in words in context	[8]
			TOTAL	**[18]**

Question 1

<table>
<tr><td>(i)</td><td>
<table>
<tr><td>x</td><td>6</td><td>17</td><td>9</td><td>20</td><td>13</td><td>15</td><td>11</td><td>14</td></tr>
<tr><td>y</td><td>6</td><td>13</td><td>10</td><td>11</td><td>9</td><td>7</td><td>12</td><td>15</td></tr>
<tr><td>Rank x</td><td>8</td><td>2</td><td>7</td><td>1</td><td>5</td><td>3</td><td>6</td><td>4</td></tr>
<tr><td>Rank y</td><td>8</td><td>2</td><td>5</td><td>4</td><td>6</td><td>7</td><td>3</td><td>1</td></tr>
<tr><td>d</td><td>0</td><td>0</td><td>2</td><td>-3</td><td>-1</td><td>-4</td><td>3</td><td>3</td></tr>
<tr><td>d^2</td><td>0</td><td>0</td><td>4</td><td>9</td><td>1</td><td>16</td><td>9</td><td>9</td></tr>
</table>

$\Sigma d^2 = 48$

$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6\times 48}{8\times 63}$
$=$ 0.429 (to 3 s.f.) [*allow* 0.43 to 2 s.f.]</td><td>M1 for attempt at ranking (allow all ranks reversed)

M1 for d^2

A1 CAO for Σd^2

M1 for method for r_s

A1 f.t. for $|r_s| < 1$
NB No ranking scores zero</td><td>5</td></tr>
<tr><td>(ii)</td><td>H_0: no association between X and Y in the population
H_1: some positive association between X and Y in the population

One tail test critical value at 5% level is 0.6429
Since 0.429 < 0.6429, there is insufficient evidence to reject H_0,

i.e. conclude that there is not enough evidence to show positive association between the two judges' scores.</td><td>B1 for H_0
B1 for H_1
B1 for population SOI
NB H_0 H_1 <u>not</u> ito ρ
B1 for ± 0.6429
M1 for sensible comparison with c.v., provided that $|r_s| < 1$
A1 for conclusion in context f.t. their r_s and sensible cv</td><td>3

3</td></tr>
<tr><td>(iii)</td><td>A bivariate Normal distribution is required.

Scatter diagram.

Suitable discussion</td><td>B1

G1 labelled axes
G1 correct points
E1
E1</td><td>5</td></tr>
<tr><td></td><td></td><td>**TOTAL**</td><td>**16**</td></tr>
</table>

Question 2

(i)	Counts have a uniform average rate of occurrence All counts are independent	E1 E1	**2**
(ii)	Variance = 3.4	B1	**1**
(iii)	(A) *Either* $P(X = 3) = 0.5584 - 0.3397 = 0.2187$ *Or* $P(X = 3) = e^{-3.4}\frac{3.4^3}{3!} = 0.2186$ (B) Using tables: $P(X \geq 3) = 1 - P(X \leq 2)$ $= 1 - 0.3397$ $= 0.6603$	M1 for use of tables or calculation A1 M1 for 1 - P($X \leq$ 2) M1 correct use of Poisson tables A1	**2** **3**
(iv)	$\lambda = 12 \times 3.4 = 40.8$ $P(X = 40) = e^{-40.8}\frac{40.8^{40}}{40!} = 0.0625$	B1 for mean M1 for calculation A1	**3**
(v)	Mean no. per hour = 12 × 3.4 = 40.8 Using Normal approx. to the Poisson, $X \sim N(40.8, 40.8)$ $P(X \geq 40) = P\left(Z > \frac{39.5 - 40.8}{\sqrt{40.8}}\right)$ $= P(Z > -0.2035) = \Phi(0.2035)$ $= 0.5806$	B1 for Normal approx. B1 for correct parameters (SOI) B1 for correct continuity corr. M1 for probability using correct tail A1 CAO (3 s.f.)	**5**
(vi)	Overall mean = 4.8 $P(X \geq 8) = 1 - P(X \leq 7)$ $= 1 - 0.8867 = 0.1133$	B1 for 4.8 M1 A1	**3**
		TOTAL	**19**

Question 3

(i)	(A) $P(X < 65) = P\left(Z < \frac{65-63}{5.2}\right)$ $= P(Z < 0.3846)$ $= \Phi(0.3846) = 0.6497$ (B) $P(60 < X < 65) = P\left(\frac{60-63}{5.2} < Z < \frac{65-63}{5.2}\right)$ $= P(-0.5769 < Z < 0.3846)$ $= \Phi(0.3846) - (1 - \Phi(0.5769))$ $= 0.6497 - (1 - 0.7181)$ $= 0.3678$	M1 for standardizing M1 for structure A1 CAO (min 3 s.f.), 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 M1 for standardizing both M1 for correct structure A1 CAO 3s.f.	**3** **3**
(ii)	P(All 5 between 60 and 65) $= 0.3678^5 = 0.00673$	M1 A1 FT (min 2sf)	**2**
(iii)	From tables $\Phi^{-1}(0.95) = 1.645$ $\frac{k-63}{5.2} = -1.645$ $x = 63 - 5.2 \times 1.645 = 54.45$ mins	B1 for ±1.645 seen M1 for correct equation in k A1 CAO	**3**
(iv)	H_0: $\mu = 63$ minutes; H_1: $\mu < 63$ minutes. Where μ denotes the population mean time on the new course. Test statistic $= \frac{61.7-63}{5.2/\sqrt{15}} = \frac{-1.3}{1.3426}$ $= -0.968$ 5% level 1 tailed critical value of $z = 1.645$ $-0.968 > -1.645$ so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that the new course results in lower times.	B1 for use of 63 B1 for both correct B1 for definition of μ M1 must include √15 A1 B1 for ±1.645 M1 for sensible comparison leading to a conclusion A1 FT for correct conclusion in words in context	**3** **5**
			19

Question 4

(i)	H_0: no association between category of runner and type of running; H_1: some association between category of runner and type of running;	B1	1
	<table><tr><th>EXPECTED</th><th>Junior</th><th>Senior</th><th>Veteran</th></tr><tr><td>Track</td><td>5.13</td><td>7.84</td><td>6.03</td></tr><tr><td>Road</td><td>6.48</td><td>9.90</td><td>7.62</td></tr><tr><td>Both</td><td>5.40</td><td>8.25</td><td>6.35</td></tr></table>	M1 A2 for expected values (to 2 dp) (allow A1 for at least one row or column correct)	
	<table><tr><th>CONTRIBUTN</th><th>Junior</th><th>Senior</th><th>Veteran</th></tr><tr><td>Track</td><td>2.9257</td><td>0.0032</td><td>2.6949</td></tr><tr><td>Road</td><td>0.9468</td><td>0.3663</td><td>2.5190</td></tr><tr><td>Both</td><td>0.3615</td><td>0.3694</td><td>0.0192</td></tr></table>	M1 for valid attempt at $(O-E)^2/E$ A1 for all correct NB These M1A1 marks cannot be implied by a correct final value of X^2	
	$X^2 = 10.21$	M1 for summation A1 for X^2	7
	Refer to X_4^2	B1 for 4 deg of f	
	Critical value at 5% level = 9.488	B1 CAO for cv	
	Result is significant	B1 FT their 'sensible' X^2	
	There is evidence to suggest that there is some association between category of runner and type of running. NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1or final E1	E1 must be consistent with their X^2	4
(ii)	• Juniors appear be track runners more often than expected and road less often than expected.	E1 E1	
	• Seniors tend to be as expected ín all three categories of running.	E1 E1	
	• Veterans tend to be road runners more than expected and track runners less than expected.	E1 E1	6
		TOTAL	**18**

Question 1

(i)	**EITHER:** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 1398.56 - \frac{1}{14} \times 139.8 \times 140.4$ $= -3.434$ $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 1411.66 - \frac{1}{14} \times 139.8^2 = 15.657$ $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 1417.88 - \frac{1}{14} \times 140.4^2 = 9.869$ $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-3.434}{\sqrt{15.657 \times 9.869}}$ $= -0.276$	M1 for method for S_{xy} M1 for method for at least one of S_{xx} or S_{yy} A1 for at least one of S_{xy}, S_{xx}, S_{yy} correct M1 for structure of r A1 (-0.27 to -0.28 to 2dp)		If $\bar{x}$ and $\bar{y}$ used in rounded form, be generous with first A1 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: $\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 1398.56/14 - 9.9857 \times 10.0286$ $= -0.2454$ $\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(15.657/14)} = \sqrt{1.1184} = 1.0575$ $\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(9.869/14)} = \sqrt{0.7049} = 0.8396$ $r = \frac{\text{cov(x,y)}}{rmsd(x)rmsd(y)} = \frac{-0.2454}{1.0575 \times 0.8396}$ $= -0.276$ 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 method for cov (x,y) M1 for method for at least one msd A1 for at least one of cov (x,y), msd(x), msd(y) correct M1 for structure of r A1 (-0.27 to -0.28 to 2dp)	5	If $\bar{x}$ and $\bar{y}$ used in rounded form, be generous with first A1 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 $= \frac{1}{n-1}\left(\Sigma fx^2 - n\bar{x}^2\right)$ $= \frac{1}{49}\left(258 - 50\times1.84^2\right)$ $= 1.81$ (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 λ = 1.8 loses both accuracy marks [1.8 leads to 1 – 0.4296 = 0.5372]
(iv)	$\lambda = 5\times1.84 = 9.2$ 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) $P(X < 325)$ $= P\left(Z < \frac{325-355}{52}\right)$ $= P(Z < -0.577)$ $= 1 - \Phi(0.577) = 1 - 0.7181$ $= 0.2819$	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) $P(300 < X < 400)$ $= P\left(\frac{300-355}{52} < Z < \frac{400-355}{52}\right)$ $= P(-1.058 < Z < 0.865)$ $= \Phi(0.865) - (1 - \Phi(1.058))$ $= 0.8065 - (1 - 0.8549)$ $= 0.6614$ (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 $\Phi^{-1}(0.2) = -0.8416$ $\frac{k-355}{52} = -0.8416$ $k = 355 - 0.8416 \times 52 = 311.2$	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 $\sigma = \sqrt{52}$
Do not penalise in parts (ii) and (iii).

RECOGNISING ACHIEVEMENT

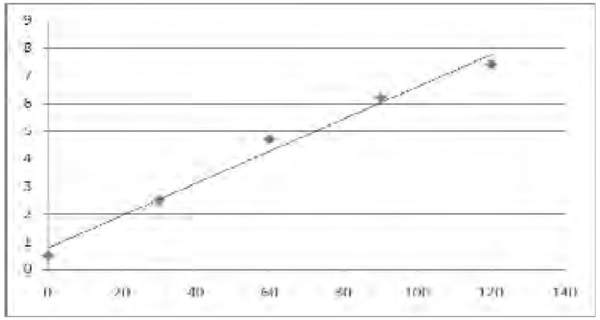
GCE

Mathematics (MEI)

Advanced GCE

Unit **4767:** Statistics 2

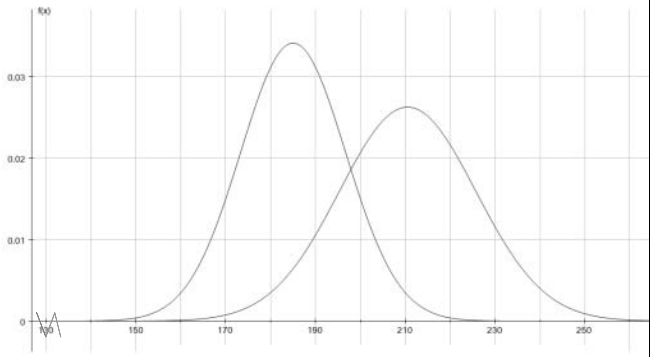
Mark Scheme for June 2011

1 (i)		G1 for axes G1 For values of x G1 for values of y	**3**	Condone axes drawn either way. Axes should show some indication of scale. If not then Max G1 if points 'visibly correct'. If axes are scaled and only one point is incorrectly plotted, allow max G2/3.
1 (ii)	$\bar{x} = 60,\ \bar{y} = 4.26$ $b = \frac{S_{xy}}{S_{xx}} = \frac{1803 - 300\times 21.3/5}{27000 - 300^2/5} = \frac{525}{9000} = 0.0583$ OR $b = \frac{1803/5 - 60\times 4.26}{27000/5 - 60^2} = \frac{105}{1800} = 0.0583$ hence least squares regression line is: $y - \bar{y} = b(x - \bar{x})$ $\Rightarrow\ y - 4.26 = 0.0583\,(x - 60)$ $\Rightarrow\ y = 0.0583x + 0.76$	B1 for $\bar{x}$ and $\bar{y}$ used appropriately (SOI) M1 for attempt at gradient (b) A1 for 0.0583 **cao** M1 for equation of line A1 FT for complete equation	**5**	B1 for means can be implied by a correct value of b using either method. Allow $\bar{y} = 4.3$ Attempt should be correct – e.g. evidence of either of the two suggested methods should be seen. Allow 0.058 Condone $0.058\dot{3}$ and $\frac{7}{120}$ Dependent on first M1. Values must be substituted to earn M1. Condone use of their b for FT provided $b>0$. Final equation must be simplified. $b = 0.058$ leads to $y = 0.058x + 0.78$
1 (iii)	Regression line plotted on graph The fit is good	G1 G1 E1 for good fit	**3**	Line must pass through their $(\bar{x}, \bar{y})$ and y-intercept. E0 for notably inaccurate graphs/lines

1 (iv)	$x = 30 \Rightarrow$ predicted $y = 0.0583 \times 30 + 0.76 = 2.509$ Residual = 2.5 – 2.509 = –0.009	B1 for prediction M1 for subtraction A1 FT	**3**	Using their **equation** Subtraction can be 'either way' but for the final mark the sign of the residual must be correct. FT sensible equations only – e.g. no FT for $y = 0.071x$ leading to +0.37. [$c = 0.78$ leads to a residual of -0.02]
1 (v)	(*A*) For $x = 45$, $y = 0.0583 \times 45 + 0.76 = 3.4$ (*B*) For $x = 150$, $y = 0.0583 \times 150 + 0.76 = 9.5$	M1 for at least one prediction attempted A1 for **both answers** (FT their equation provided their $b>0$)	**2**	Prediction obtained from their equation.
1 (vi)	This is **well** below the predicted value … …suggesting that the model breaks down for larger values of *x*.	E1 for **well** below E1 extrapolation	**2**	Some indication that the value (8.7) is **significantly** below what is expected (9.5) is required for the first E1. Simply pointing out that it is 'below' is not sufficient. The second E1 is available for a suitable comment relating to the model being suitable only for values within the domain of the given points. Allow other sensible comments for either E1. E.g. The data might be better modelled by a curve', 'there may be other factors affecting yield',
			18	

2 (i)	Independently means that the arrival time of each car is unrelated to the arrival time of any other car. Randomly means that the arrival times of cars are not predictable. At a uniform average rate means that the average rate of car arrivals does not vary over time.	E1 E1 E1	**3**	NOTE **Each answer must be 'in context' and 'clear'** Allow sensible alternative wording. SC1 For ALL answers not in context but otherwise correct.
2 (ii)	P(At most 1 car) $= e^{-0.62}\dfrac{0.62^0}{0!} + e^{-0.62}\dfrac{0.62^1}{1!}$ $= 0.5379\ldots + 0.3335\ldots = 0.871$	M1 for either M1 for sum of both A1 CAO	**3**	$1.62e^{-0.62}$ Allow 0.8715 not 0.872 or 0.8714 Allow 0.87 without wrong working seen
2 (iii)	New $\lambda = 10 \times 0.62 = 6.2$ P(more than 5 in 10 mins) = 1 – 0.4141 = 0.5859	B1 for mean (SOI) M1 for probability A1 CAO	**3**	Use of $1 - P(X \le 5)$ with any λ Allow 0.586
2 (iv)	Poisson with mean 37.2	B1 for Poisson B1 for mean 37.2	**2**	Dependent on first B1 Condone P(37.2, 37.2)
2 (v)	Use Normal approx with $\mu = \sigma^2 = \lambda = 37.2$ $P(X \ge 40) = P\left(Z > \dfrac{39.5 - 37.2}{\sqrt{37.2}}\right)$ $= P(Z > 0.377) = 1 - \Phi(0.377) = 1 - 0.6469$ $= 0.3531$	B1 for Normal (SOI) B1 for parameters B1 for 39.5 M1 for correct use of Normal approximation using correct tail A1 **cao**	**5**	Allow 0.353
			16	

3 (i)	P(Apple weighs at least 220g) $= P\left(Z > \frac{220-210.5}{15.2}\right)$ $= P(Z > 0.625)$ $= 1 - \Phi(0.625) = 1 - 0.7340$ $= 0.2660$	M1 for standardising M1 for correct structure A1 CAO inc use of diff tables	**3**	Condone numerator reversed but penalise continuity corrections i.e. 1 – Φ(positive z value) Allow 0.266 but not 0.27
3 (ii)	P(All 3 weigh at least 220g) = $0.2660^3 = 0.0188$	M1 A1 FT	**2**	M1 for their answer to part (i) cubed Allow 0.019 and 0.01882
3 (iii)	(*A*) Binomial (100, 0.0188) (*B*) Use a Poisson distribution with $\lambda = 100\times0.0188 = 1.88$ P(At most one) = $e^{-1.88}\frac{1.88^0}{0!} + e^{-1.88}\frac{1.88^1}{1!}$ $= 0.1525 + 0.2869 = 0.4394$ (*C*) n is large and p is small	B1 for binomial B1 for parameters B1 for Poisson SOI B1 for Poisson mean M1 for either probability M1 for sum of both A1 CAO For 0.44 or better B1	**2** **5** **1**	Second B1 dependent on first B1 FT their answer to part (ii) for second B1 Consistent with $p < 0.1$ from part (iii) (*A*) FT answer to part (ii) with $p < 0.1$ Dependent on both previous B1 marks Allow 0.4395 but not 0.4337 Dependent on use of Poisson in part (iii) *B* Allow n is large and np <10 & n is large and $np \approx npq$
3(iv)(*A*)	$\Phi^{-1}(0.1) = -1.282$ $\frac{170-185}{\sigma} = -1.282$ $1.282\,\sigma = 15$ $\sigma = 11.70$	B1 for ±1.282 M1 for correct equation as written o.e. A1 CAO	**3**	Do not allow 1 – 1.282 Allow M1 if different ***z*-value** used Without incorrect working seen. Allow 11.7

3(iv)(*B*)	 Cox's Braeburns	G1 for shape G1 for means, shown explicitly or by scale G1 for lower max height for Braeburns G1 for greater width (variance) for Braeburns	**4**	Ignore labelling of vertical axis. Two intersecting, adjacent Normal curves Means at 185 and 210.5
		TOTAL	**20**	
4(a)(i)	H_0: no association between amount spent and sex H_1: some association between amount spent and sex	B1 for both	**1**	Hypotheses must be the right way round, in context and must not mention 'correlation'.
4(a)(ii)	Expected frequency = 62×102÷200 = 31.62 Contribution = $(34 – 31.62)^2$ / 31.62 = 0.1791	B1 M1 A1 for valid attempt at $(O-E)^2/E$ ***NB Answer given***	**3**	Do not allow 31.6

4(a)(iii)	Refer to X_4^2 Critical value at 5% level = 9.488 3.205 < 9.488 Result is not significant There is insufficient evidence to suggest any association between amount spent and sex. .	B1 for 4 deg of freedom B1 CAO for cv M1 A1 for not significant E1	**5**	Allow p = 0.524 0.524 > 0.05 Conclusion must be stated to earn A1 here. Allow 'do not reject H_0' and condone 'accept H_0' or 'reject H_1'. FT if cv consistent with their d.o.f. Dependent on previous A1 and final comment must be in context and not mention correlation. SC1 for correct final conclusion where previous A1 omitted but M1 awarded.
4 (b)	H_0: $\mu = 400$; H_1: $\mu < 400$ Where μ denotes the population mean (weight of the loaves). $\bar{x} = 396.5$ Test statistic $= \frac{396.5-400}{5.7/\sqrt{6}} = \frac{-3.5}{2.327} = -1.504$ 5% level 1 tailed critical value of z = –1.645 –1.504 > –1.645 so not significant. There is insufficient evidence to reject H_0 There is insufficient evidence to suggest that the true mean weight of the loaves is lower than the minimum specified value of 400 grams.	B1 for H_0 B1 for H_1 B1 for definition of μ B1 for sample mean M1 must include √6 A1FT their sample mean B1 for ±1.645 M1 for sensible comparison leading to a conclusion A1 for conclusion in context	**9**	Hypotheses in words must refer to population mean. Condone numerator reversed for M1 but award A1 only if test statistic of 1.504 is compared with a positive z-value. Dependent on previous M1 FT their sample mean only if hypotheses are correct.
		TOTAL	**18**	

Additional notes re Q1 parts (ii), (iv) and (v)

Part (ii) 'x on y' max B1
Part (iv) $x = 16.9y - 12.02$ leads to a prediction of $x = 30.23$ and a residual of -0.23 B1 M1 A1 available.
Part (v) 'x on y' not appropriate here so award 0 if 'x on y' used.

Additional notes re Q2 parts (i) & (v)

Part (i)
Independent – Allow 'not linked to' or 'no association' or 'unrelated' 'not affected by', 'not connected to', 'not influenced by'
DO NOT ACCEPT 'not together' or 'not dependent'
Random – Allow 'not predictable' or 'no pattern' or 'could happen at any time' or 'not specific time'
Uniform average rate – Allow 'average (rate) is constant over time' DO NOT ACCEPT 'average constant' or 'average rate and uniform' – be generous over defining 'average' and/or 'rate'.
Part (v) If Binomial distribution stated in part (iv), allow B1 B0 B1 M0 A0 max

Additional notes re Q3 part (iii) where $p > 0.1$

(iii) B – as scheme unless a Normal approximation is more suitable ($p > 0.1$). If so, award B1 B1 for Normal and correct parameters. The remaining marks are dependent on both these B1 marks being awarded. M1 for the correct continuity correction ($P(X < 1.5)$) and depM1 for the correct tail but award A0.
(iii) C – 'n is large and p is not too small' or '$np > 10$'

Additional notes re Q4(b)

<u>σ estimated</u>
sample mean, 7.079… used in place of 5.7, the given value of the population mean, leads to a test statistic of -1.212… This gets M1A0 & the remaining marks are still available.

<u>Critical Value Method</u>
400 – 1.645 × 5.7 ÷ √6… gets M1B1 …= 396.17…gets A1
400 + 1.645 × 5.7 ÷ √6 gets M1B1A0.
396.5 > 396.2 gets M1for sensible comparison (and B1 for 396.5)
A1 still available for correct conclusion in words & context

<u>90% Confidence Interval Method</u>
CI centred on 396.5 (gets B1 for 396.5)
+ or – 1.645 × 5.7 ÷ √6 gets M1 B1
= (392.67, 400.33) A1
contains 400 gets M1
A1 still available for correct conclusion in words & context

<u>Probability Method</u>
Finding P(sample mean < 396.5) = 0.0663 gets M1 A1 (and B1 for 396.5)
0.0663 > 0.05 gets M1 for a sensible comparison if a conclusion is made and also gets the B1 for 0.0663 (to replace the B1 for cv = 1.645).
A1 still available for correct conclusion in words & context.
Condone P(sample mean >396.5) = 0.9337 for M1 and B1 for 0.9337 but only allow A1 if later compared with 0.95 at which point the final M1and A1 are still available

<u>Two-tailed test</u>
Max B1 B0 B1 B1 M1 A1 B1 (for cv = -1.96) M1 A0

Question			Answer	Marks	Guidance
1	(i)		**EITHER**		
			$S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x\Sigma y = 600.41 - \frac{1}{10} \times 113.69 \times 52.81 = 0.01311$	M1	For method for S_{xy}
			$S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 1292.56 - \frac{1}{10} \times 113.69^2 = 0.01839$	M1	For method for at least one of S_{xx} or S_{yy}
			$S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 278.91 - \frac{1}{10} \times 52.81^2 = 0.02039$	A1	For at least one of S_{xy}, S_{xx} or S_{yy} correct
			$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{0.01311}{\sqrt{0.01839 \times 0.02039}} = 0.677$	M1 A1	For fully correct structure of r For answer rounding to 0.68
			OR		
			$\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 600.41/10 - 11.369 \times 5.281 = 0.001311$	M1	For method for cov (x,y)
			$\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(0.01839/10)} = \sqrt{0.001839} = 0.04288$	M1	For method for at least one msd or rmsd
			$\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(0.02039/10)} = \sqrt{0.002039} = 0.04516$	A1	For at least one of cov (x,y), msd or rmsd correct
			$r = \frac{\text{cov}(x,y)}{rmsd(x)rmsd(y)} = \frac{0.001311}{0.04288 \times 0.04516} = 0.677$	M1 A1	For fully correct structure of r For answer rounding to 0.68
				[5]	Methods mixed – max M0M1A1M0A0
1	(ii)		H_0: $\rho = 0$ H_1: $\rho \neq 0$ (two-tailed test)	B1	For H_0, H_1 in symbols. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population correlation coefficient.
			where ρ is the population correlation coefficient	B1	For defining ρ. Condone omission of "population" if correct notation ρ is used, but if ρ is defined as the **sample** correlation coefficient then award **B0**.
			For n = 10, 10% critical value = 0.5494	B1	CAO Note that critical values for a one-tailed test at the 10% level are not available in tables.

			Since 0.677 > 0.5494 the result is significant.	M1	For sensible comparison leading to a conclusion provided that $\|r\| < 1$. The comparison can be in the form of a diagram as long as it is clear and unambiguous. Sensible comparison: e.g. 0.677 > 0.5494 is 'sensible' whereas 0.677 > - 0.5494 is 'not sensible'. Reversed inequality sign e.g. 0.677 < 0.5494 etc. gets max M1 A0.	
			(Thus we have sufficient evidence to) reject H_0	A1*	For reject H_0 o.e. FT their r and critical value from 10% 2-tail column.	
			There is sufficient evidence at the 10% level to suggest that there is correlation between times for the first and last sections.	E1dep* **[6]**	For correct, non-assertive conclusion in context. Allow 'x and y' for context. E0 if H_0 and H_1 not stated, reversed or mention a value other than zero for ρ in H_0. Do not allow 'positive correlation' or 'association'	
1	**(iii)**		The underlying population must have a bivariate Normal distribution.	B1	Condone "bivariate Normal distribution", "underlying bivariate Normal distribution", but **do not allow** "the **data** have a bivariate Normal distribution"	
			The points in the scatter diagram should have a roughly elliptical shape.	E1 **[2]**	Condone 'oval' or suitable diagram	
1	**(iv)**		The hypothesis test has shown that there appears to be correlation.	E1	For relevant comment relating to the test result or positive value of r in supporting (unless FT leads to not supporting) the commentator's suggestion. Or correlation does not imply causation. There may be a third factor. For questioning the use of the word 'must'	
			However it could be that there is a third causal factor	E1 **[2]**	Allow any two suitable, statistically based comments.	
1	**(v)**	(*A*)	Yes because the critical value at the 1% level is 0.7646 which is larger than the test statistic	B1* E1dep* **[2]**	B1 for 0.7646 seen E1 for comment consistent with their (ii) provided $r\| < 1$	

1	**(v)**	(*B*)	One advantage of a 1% level is that one is less likely to reject the null hypothesis when it is true. One disadvantage of a 1% level is that one is more likely to accept the null hypothesis when it is false.	E1 E1 **[2]**	o.e. Wording must be clear. o.e.	
2	**(i)**		Binomial(1200,1/300)	B1 B1dep **[2]**	For binomial. For parameters Allow B(1200, 1/300) and B(1200, 0.00333)	
2	**(ii)**		Because n is large and p is small	E1, E1 **[2]**	Allow n is large and $np < 10$. Allow "sample is large" for n is large and "mean ≈ variance" for "p is small"	
2	**(iii)**		$\lambda = 1200 \times 1/300 = 4$ (*A*) $\mathrm{P}(X = 1) = e^{-4}\dfrac{4^1}{1!} = 0.0733$ (3 s.f.) or from tables $= 0.0916 - 0.0183 = 0.0733$ (*B*) Using tables: $\mathrm{P}(X > 4) = 1 - \mathrm{P}(X \le 4)$ $= 1 - 0.6288 = 0.3712$	B1 M1 A1 M1 A1 **[5]**	For λ FT their p For attempt to find $\mathrm{P}(X = 1)$ using Poisson p.d.f. or tables Allow answers which round to 0.073 www. FT their λ (= np). No FT for $\lambda = 1/300$. For finding $1 - \mathrm{P}(X \le 4)$ CAO For answers rounding to 0.371 www	
2	**(iv)**		$\mu = 80$ $\sigma^2 = 80$	B1 B1 **[2]**	If symbols/words used then they must be correct. Allow σ^2 rounding to 79.7 from original binomial. FT their λ (= np)	
2	**(v)**	(*A*)	$\mathrm{P}(Y \ge 90) = \mathrm{P}\left(Z \ge \dfrac{89.5 - 80}{\sqrt{80}}\right)$ $= \mathrm{P}(Z > 1.062) = 1 - \Phi(1.062)$ $= 1 - 0.8559 = 0.1441$	B1 M1 A1cao **[3]**	For correct continuity correction. For probability using correct tail and structure (condone omission of c.c.) $\sigma^2 = 79.73$ leads to $\mathrm{P}(Z > 1.064)$ $\sigma^2 = 79.73$ leads to $1 - 0.8563 = 0.1437$. Allow 0.144 www. NOTE 0.1441 from B(24000, 1/300) gets 0/3	

2	**(v)**	(*B*)	$P(Y \le k) > 0.05$ From tables $\Phi^{-1}(0.05) = -1.645$ $\frac{(k+0.5)-80}{\sqrt{80}} = -1.645$ $k + 0.5 = 80 - (1.645 \times \sqrt{80}) = 65.29$ $k > 64.79$ So least value of $k = 65$	 B1 M1 A1 A1 **[4]**	For ±1.645 seen For correct equation in *k* seen or equivalent – e.g. allow +1.645 used if numerator reversed. FT their μ, σ^2 and z-value. Condone omission of, or incorrect, continuity correction. A1 for 65.29 or 64.79 or 65.79 (σ^2 = 79.73 leads to 65.31 or 64.81 or 65.81) Allow 3s.f. For rounding **64.79 or 64.81 up to give *k* = 65**. See additional notes for alternative method	
3	**(i)**		$P(X \ge 750) = P\left(Z \ge \frac{750 - 751.4}{2.5}\right)$ $= P(Z > -0.56) = \Phi(0.56) = 0.7123$	M1 M1 A1 **[3]**	For standardizing For correct structure (M0 if continuity correction used) CAO Allow 0.712 www	
3	**(ii)**		P(all 6 at least 750ml) = 0.7123^6 = 0.1306	M1 A1 **[2]**	For (their answer to part (i))^6 FT 3s.f.	
3	**(iii)**		$P(Y=0) = \binom{25}{0} \times 0.8694^{25}$ (= 0.0302) $P(Y=1) = \binom{25}{1} \times 0.8694^{24} \times 0.1306$ (= 0.1135) P(*Y*=0) + P(*Y*=1) = 0.144 $P(Y \ge 2) = 1 - 0.144$ =0.856	M1 M1 M1dep A1 **[4]**	For using Binomial(25, *p*) with their *p* from part (ii) For correct structure of either P(*Y* = 0) or P(*Y* = 1) with their *p* from part (ii) M0 if *p* and *q* reversed For 1 – sum of both probabilities CAO	

3	(iv)		$P(Z < \frac{750-\mu}{2.5}) = 0.02$ $\Phi^{-1}(0.02) = -2.054$ $\frac{750-\mu}{2.5} = -2.054$ $\mu = 750 + 2.054 \times 2.5$ = 755.1	B1 M1 M1 A1 **[4]**	For ±2.054 seen. Allow ±2.055 For correct equation as seen or equivalent. FT σ = √2.5. M0 if c.c. used. For correctly rearranging their equation (if 750 used in numerator) for μ, FT their *z* cao Condone 755 or 5 s.f. rounding to 755.1 www	
3	(v)		$P(Z < \frac{750-751.4}{\sigma}) = 0.02$ $\frac{750-751.4}{\sigma} = -2.054$ $\sigma = \frac{-1.4}{-2.054}$ = 0.682	M1 M1 A1 **[3]**	For correct equation as seen or equivalent For correctly rearranging their equation (if 750 used in numerator) for σ unless this leads to σ < 0 cao Allow answers rounding to 0.68 www	
3	(vi)		Probably easier to change the mean (as reducing the standard. deviation would require a much more accurate filling process). However increasing the mean would result in fewer bottles being filled overall and so less profit for the owners, so **reducing** the **standard deviation** would be **preferable** to the vineyard owners.	E1 E1 **[2]**	 For "preferable to reduce the standard deviation" with valid reason.	
4	(a)	(i)	Expected frequency = 67/150 × 57 = 25.46 Contribution = $(34 - 25.46)^2$ / 25.46 = 2.8646	B1 M1 A1 **[3]**	For 25.46 For valid attempt at $(O-E)^2/E$ Correct values used to give answer which rounds to 2.8646 ***NB Answer given***	

4	**(a)**	**(ii)**	H_0: no association between type of cake and classification of person. H_1: some association between type of cake and classification of person. Test statistic $X^2 = 12.86$ Refer to χ_3^2 Critical value at 1% level = 11.34 Result is significant There is evidence to suggest association between type of cake and classification of person. NB if H_0 H_1 reversed, omitted or 'correlation' mentioned, do not award first B1 or final E1	B1 B1 B1 B1 E1 **[5]**	For both hypotheses in context For 3 degrees of freedom CAO For cv. No FT from here if wrong/omitted For significant For correct, non-assertive conclusion, in context.	
4	**(b)**		$\bar{x} = 4.995$ H_0: $\mu = 5$ H_1: $\mu < 5$ Where μ denotes the **mean** content of the bags of flour (in the population) Test statistic $= \dfrac{4.995 - 5.0}{0.0072/\sqrt{8}} = \dfrac{-0.005}{0.002546} = -1.964$ Lower 5% level 1 tailed critical value of $z = -1.645$ $-1.964 < -1.645$ so significant. There is sufficient evidence to reject H_0 There is sufficient evidence to suggest that the average contents of bags is less than 5kg.	B1 B1 B1 B1 M1* A1 B1* M1 dep* A1 **[9]**	For 4.995 seen For use of 5 in hypotheses. For both correct. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population mean. For definition of μ in context. Condone omission of "population" if correct notation μ is used, but if μ is defined as the **sample** mean then award **B0**. must include $\sqrt{8}$ FT their $\bar{x}$. Allow +1.964 only if later compared with +1.645 For –1.645 No FT from here if wrong. Must be –1.645 unless it is clear that absolute values are being used. For sensible comparison with correct c.v. leading to a conclusion. For non-assertive conclusion in words and in context. No FT here. See additional notes.	

ADDITIONAL NOTES REGARDING QUESTION 2 (v) *B*

M1 for using a trial and improvement method with N(80,80) or N(80, 79.73) to find $P(Y \leq k)$ for any k. The distribution being used needs to be made clear.

A1 for $P(Y \leq 66) = 0.0587\ldots$ (0.0584… from $\sigma^2 = 79.73$) or $P(Y \leq 65) = 0.0467\ldots$ (0.0464… from $\sigma^2 = 79.73$)
A1 for both
Final A1 not available if 66 and 65 used

Or
A1 for $P(Y \leq 65.5) = 0.0524\ldots$ (0.0521… from $\sigma^2 = 79.73$) or $P(Y \leq 64.5) = 0.0415\ldots$ (0.0412… from $\sigma^2 = 79.73$)
A1 for both
A1 for least value of $k = 65$, dependent on previous two A marks earned.

ADDITIONAL NOTES REGARDING QUESTION 4 (b)

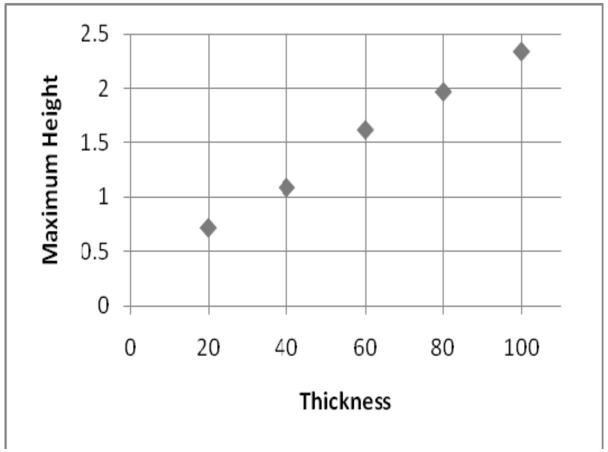
Critical Value Method
$5 - 1.645 \times 0.0072 \div \sqrt{8}$ gets M1*B1*
= 4.9958… gets A1
4.995 < 4.99581.. gets M1dep* for sensible comparison
A1 still available for correct conclusion in words & context

"Confidence Interval" Method
$4.995 + 1.645 \times 0.0072 \div \sqrt{8}$ gets M1* B1*
= 4.9991.. gets A1
NOTE that the final M1dep* A1 available only if 1.645 used.
5 > 4.9991… gets M1
A1 still available for correct conclusion in words & context

Probability Method
Finding P(sample mean < 4.995) = 0.0248 gets M1* A1 B1
0.0248 < **0.05*** gets M1dep* for a sensible comparison if a conclusion is made.
A1 available for a correct conclusion in words & context.
Condone P(sample mean > 4.995) = 0.9752 for M1 but only allow A1 B1 if later compared with 0.95, at which point the final M1and A1 are still available

ADDITIONAL NOTE REGARDING OVER-SPECIFICATION OF ANSWERS

Over-specification by providing final answers correct to 5 or more significant figures will be penalised. When this applies, candidates may lose no more than 2 marks per question and no more than 4 marks in total. The only exception to this rule is in Question 3 part (iv) – see guidance note.

Question			Answer	Marks	Guidance	
1	(i)			G1 G2,1,0 **[3]**	G1 For axes suitably labelled with some indication of **linear** scale provided. G2 for points plotted correctly. G1 if 4 points plotted correctly. G0 if two or more incorrectly plotted/omitted points. Special Case SC1 for points visibly correct on axes where no indication of scale has been provided.	Allow x & y Allow axes reversed.
1	(ii)		Thickness is the independent variable since the values of ‘Thickness’ are not subject to random variation, but are determined by the manufacturer.	E1 **[1]**	Allow explanations referring to thickness being “controlled” by the manufacturer. Allow equivalent interpretations.	
1	(iii)		$\bar{t} = 60,\ \bar{h} = 1.548$ $b = \frac{S_{th}}{S_{tt}} = \frac{546.8 - (300 \times 7.74/5)}{22000 - 300^2/5} = \frac{82.4}{4000} =$ 0.0206	B1 M1* A1	For $\bar{t}$ and $\bar{h}$ used. SOI (e.g. can be implied by $b = 0.0206$) For attempt at **calculating** gradient (b) for h on t. For 0.0206 cao	
			OR $b = \frac{546.8/5 - (60 \times 1.548)}{22000/5 - 60^2} = \frac{16.48}{800} =$ 0.0206 hence least squares regression line is: $h - \bar{h} = b(t - \bar{t})$			

Question			Answer	Marks	Guidance	
			$\Rightarrow$ $h - 1.548 = 0.0206\ (t - 60)$	M1 dep*	For equation of line, using their b, $b > 0$, and passing through their $(\bar{t}, \bar{h})$	
			$\Rightarrow$ $h = 0.0206\ t + 0.312$	A1 [5]	Final equation must have h as the subject. **CAO** Allow $h = 0.021\ t + 0.31$, Allow $h = 0.021\ t + 0.288$ NOTE If equation given in terms of y and x then A0 unless x & y defined appropriately	
1	**(iv)**	**(*A*)**	$(0.0206 \times 70) + 0.312 = 1.754$ Likely to be reliable as interpolation	B1 E1 **[2]**	Allow 1.75 FT their equation provided $b > 0$	
1	**(iv)**	**(*B*)**	$(0.0206 \times 120) + 0.312 = 2.784$ Could be unreliable as extrapolation	B1 E1 **[2]**	Allow 2.78 FT their equation provided $b > 0$ Condone “reliable as 120 is not too far away from the data used to produce the equation”	
1	**(v)**		Thickness = 40 $\Rightarrow$ predicted max height $= (0.0206 \times 40) + 0.312 = 1.136$ Residual = 1.09 – 1.136 = –0.046	 M1 M1 A1 **[3]**	 For prediction. FT their equation provided $b > 0$ For difference between 1.09 and prediction. Allow –0.05	
1	**(vi)**		Regression line gives a prediction of $(0.0206 \times 200) + 0.312 = 4.432$ This is **well above** the observed value. It could be that the relationship breaks down for larger thickness, or that the relationship is not linear	B1* E1 dep* E1 **[3]**	B1 for obtaining a prediction from regression equation or from graph E1 for noting the large difference between prediction and actual value E1 for suitable interpretation regarding the relationship between maximum height and thickness	

Question			Answer	Marks	Guidance	
2	(i)	(A)	$P(X = 0) = \dfrac{e^{-2.1}2.1^0}{0!}$	M1	For calculation	
			$= 0.1225$	A1	**CAO** Allow 0.122	
			Or from tables $P(X = 0) = 0.1225$			
				[2]		
2	(i)	(B)	$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.3796$	M1	M1 for **use of** correct structure. i.e. M0 for use of $1 - P(X \leq 2)$ or $1 - 0.6796$ Using $\lambda = 2.0$ leading to $1 - 0.4060$ gets M1	
			$= 0.6204$	A1 **[2]**	CAO Allow 0.6203, 0.620	
2	(i)	(C)	New $\lambda = 5 \times 2.1 = 10.5$	B1	For mean (SOI)	
			P(Between 5 and 10 in 5 mins) $= 0.5207 - 0.0211$	M1	For $P(X \leq 10) - P(X \leq 4)$ used.	e.g. $1 - 0.9379$ leads to B0M1A0
			$= 0.4996$	A1 **[3]**	CAO Allow 0.500, 0.50. Condone 0.5 www.	
2	(ii)		Mean number in 60 minutes $= 60 \times 2.1 = 126$			
			Using Normal approx. to the Poisson,	B1	For Normal approx.	
			$X \sim N(126, 126)$	B1	For correct parameters (SOI)	
			$P(X \geq 130) = P\left(Z \geq \dfrac{129.5 - 126}{\sqrt{126}}\right)$	B1	For correct continuity correction	
			$= P(Z > 0.3118) = 1 - \Phi(0.3118)$	M1	For correct probability structure	
			$= 1 - 0.6224$ $= 0.3776$	A1	CAO**, (**Do not FT wrong or omitted CC). Allow 0.378www & 0.3775	
				[5]		

Question			Answer	Marks	Guidance	
2	(iii)		(Because if butterflies are blown in pairs,) the events will no longer be occurring singly.	E1 **[1]**		Accept 'not independent'
2	(iv)		P(3 or fewer) = P(3 or fewer individuals and no pairs) + P(0 or 1 individual and 1 pair)			
			= (0.9068 × 0.8187) + (0.4932 × (0.9825 – 0.8187))	M1 M1	For P(0 pairs) (= 0.8187) For P(1 pair) (= 0.1638 or 0.1637)	First two M1s can be awarded for 0.9825
			= (0.9068 × 0.8187) + (0.4932 × 0. 1638)	M2	For structure M2 for correct 6 combinations identified and their probabilities added, M1 for 5 combinations identified and their probabilities added.	
			= 0.7424 + 0.0808			
			= 0.8232	A1	CAO Allow awrt 0.823	
			Or			
			using *D* for the number of pairs and *S* for the number of singles $P(D=0)\times P(S=0)=e^{-0.2}\times e^{-1.7}$ = 0.1495… $P(D=0)\times P(S=1)=e^{-0.2}\times 1.7e^{-1.7}$ = 0.2542… $P(D=0)\times P(S=2)=e^{-0.2}\times 1.7^2e^{-1.7}\div 2$ = 0.2161… $P(D=0)\times P(S=3)=e^{-0.2}\times 1.7^3e^{-1.7}\div 3!$ = 0.1224… $P(D=1)\times P(S=0)=0.2e^{-0.2}\times e^{-1.7}$ = 0.0299… $P(D=1)\times P(S=1)=0.2e^{-0.2}\times 1.7e^{-1.7}$ = 0.0508…			
			Or			
			$P(D=0)\times P(S=3)+P(D=1)\times P(S=1)$ $P(D=0)\times P(S=2)+P(D=1)\times P(S=0)$ $P(D=0)\times P(S=1)$ $P(D=0)\times P(S=0)$		P(3 butterflies) (= 0.1733…) P(2 butterflies) (= 0.2460…) P(1 butterfly) (= 0.2542…) P(0 butterflies) (= 0.1495…)	

Question			Answer	Marks	Guidance	
			Or $P(D \le 1) \times P(S \le 1) = 0.98247 \times 0.49324$ $D(D = 0) \times P(S = 2) = 0.21613$ $D(D = 0) \times P(S = 3) = 0.12247$	**[5]**		
3	(i)	(*A*)	$P(X < 450) = P\left(Z < \frac{450-435}{30}\right)$ $= P(Z < 0.5) = \Phi(0.5)$ $= 0.6915$	M1 M1 A1 **[3]**	For standardising. M0 if 'continuity correction' applied For correct structure CAO Allow 0.692	
3	(i)	(*B*)	$P(400 < X < 450)$ $= P\left(\frac{400-435}{30} < Z < \frac{450-435}{30}\right)$ $= P(-1.1667 < X < 0.5)$ $= \Phi(0.5) - \Phi(-1.1667)$ $= 0.6915 - 0.1216$ $= 0.5699$	 M1 B1 A1 **[3]**	 For correct structure For use of difference column to obtain 0.8784, 0.8783, 0.1216 or 0.1217. Condone 0.8782 or 0.1218 FT "their 0.6915" – 0.1216 (or 0.1217)	
3	(ii)		P(all 5 between 400 and 450) $= 0.5699^5$ $= 0.0601$	 M1 A1 **[2]**	 FT Allow 0.060	

Question			Answer	Marks	Guidance	
3	(iii)		$P(Y < 350) = 0.2$, $P(Y > 390) = 0.1$			
			$P(Z < \frac{350-\mu}{\sigma}) = 0.2$			
			$\Phi^{-1}(0.2) = -0.8416$			
			$\frac{350-\mu}{\sigma} = -0.8416$	M1	For equation as seen or equivalent with their –ive z value	If 'continuity corrections' applied allow M marks but do not award final A marks
			$P(Z > \frac{390-\mu}{\sigma}) = 0.1$			
			$\Phi^{-1}(0.9) = 1.282$	B1	For 1.282 or -0.8416	
			$\frac{390-\mu}{\sigma} = 1.282$	M1	For equation as seen or equivalent with their +ive z value	
			$350 = \mu - 0.8416\,\sigma$ $390 = \mu + 1.282\,\sigma$ $2.1236\,\sigma = 40$			
			$\sigma = 18.84$	A1	Allow 18.8	Answers to max 2 d.p.
			$\mu = 350 + (0.8416 \times 18.84) = 365.85$	A1 **[5]**	Allow 365.86, 366, 365.9	
3	(iv)		$\Phi^{-1}(0.975) = 1.96$	B1	For using a suitable pair of z values e.g. $\pm$ 1.96	
			$a = 365.85 - (1.96 \times 18.84)$	M1	For either equation provided that a suitable pair of z-values is used. e.g. +2.326 and –1.751	
			$= 328.9$	A1	FT their μ and σ to 2 d.p. (A0 if 'continuity correction' used)	Accept any correct values of a and b.
			$b = 365.85 + (1.96 \times 18.84)$ $= 402.8$	A1 **[4]**	FT their μ and σ to 2 d.p. (A0 if 'continuity correction' used)	

Question			Answer	Marks	Guidance	
4	(a)		H_0: no association between grade and hours worked H_1: some association between grade and hours worked;	B1	Hypotheses in context	NB These M1A1 marks cannot be implied by a correct final value of X^2
			Expected values table (see below)	M1 A1	Any row/column correct For expected values (to 2 dp)	
			Contributions table (see below)	M1 A1	For valid attempt at $(O-E)^2/E$. Any row column correct. For all correct	
			$X^2 = 2.347$	B1		
			Refer to χ_1^2	M1	For 1 deg of freedom. No FT from here if wrong.	
			Critical value at 5% level = 3.841	A1	CAO for cv or p-value = 0.1255. SC1 for cv or p-value if 1 dof not seen.	
			Result is not significant. There is insufficient evidence to suggest that there is any association between hours worked and grade.	E1	For conclusion in context. NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1 or final E1	
				[9]		

Expected values:

	Less than 5hrs	At least 5hrs
A or B	17.05	13.95
C or lower	15.95	13.05

Contributions:

	Less than 5hrs	At least 5hrs
A or B	0.5104	0.6238
C or lower	0.5456	0.6669

Question			Answer	Marks	Guidance	
4	(b)		$\bar{x} = 417.79$	B1	For $\bar{x}$	
			H_0: $\mu = 420$;	B1	For use of 420 in hypotheses. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population mean.	
			H_1 : $\mu \neq 420$	B1	For both correct	
			Where μ denotes the mean volume of the cans of tomato purée (in the population)	B1	For definition of μ. Condone omission of "population" if correct notation μ is used, but if μ is defined as the **sample** mean then award **B0**.	
			Test statistic $= \frac{417.79 - 420}{3.5/\sqrt{10}} = \frac{-2.21}{1.107} = -1.997$	M1* A1	must include $\sqrt{10}$ FT their $\bar{x}$	
			Lower 1% level 2 tailed critical value of z = – 2.576	B1*	For –2.576 Must be –2.576 unless it is clear that absolute values are being used.	
			–1.997 > – 2.576	M1 dep*	For sensible comparison leading to a conclusion.	
			So not significant. There is insufficient evidence to reject H0			
			There is insufficient evidence to suggest that the average volumes of the cans of tomato purée is not 420ml	A1 **[9]**	For conclusion in words in context provided that correct cv used. FT only candidate's test statistic.	

ADDITIONAL NOTES REGARDING QUESTION 4 (b)

Critical Value Method
420 – 2.576 × 3.5 ÷ √10 gets M1*B1*
= 417.148… gets A1
417.79 > 417.148.. gets M1dep* for sensible comparison
A1 still available for correct conclusion in words & context

Confidence Interval Method
CI centred on 417.79 + or – 2.5756 × 3.5 ÷ √10 gets M1* B1*
= (414.93…, 420.64..) gets A1
NOTE that the final M1dep* A1 available only if 2.576 used.
"Contains 420" gets M1dep*
A1 still available for correct conclusion in words & context

Probability Method
Finding P(sample mean < 417.79) = 0.0229 gets M1* A1 B1*
0.0229 > **0.005*** gets M1dep* for a sensible comparison if a conclusion is made.
A1 available for a correct conclusion in words & context.
Condone P(sample mean > 417.79) = 0.9771 for M1* but only allow A1 B1* if sensible comparison made, at which point the final M1dep* and A1 are still available

ADDITIONAL NOTE REGARDING OVER-SPECIFICATION OF ANSWERS
Over-specification by providing final answers correct to 5 or more significant figures will be penalised. When this applies, candidates may lose no more than 2 marks per question and no more than 4 marks in total. The only exception to this rule is in Question 3 parts (iii) & (iv) – see guidance notes.

Question		Answer	Marks	Guidance
1	(i)	**EITHER:**		
		$S_{xy} = \sum xy - \frac{1}{n}\sum x \sum y = 40.66 - \frac{1}{60} \times 43.62 \times 55.15$	M1*	For method for S_{xy}
		$= 0.56595$		
		$S_{xx} = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2 = 32.68 - \frac{1}{60} \times 43.62^2$	M1*	For method for at least one of S_{xx} or S_{yy}
		$= 0.96826$		
		$S_{yy} = \sum y^2 - \frac{1}{n}\left(\sum y\right)^2 = 51.44 - \frac{1}{60} \times 55.15^2$	A1	For at least one of S_{xy}, S_{xx} or S_{yy} (to 2 sf) Note Allow 0.57322 for S_{xy} and 0.76634 for S_{yy} from rounding mean of y to 0.919.
		$= 0.74796$		
		$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{0.56595}{\sqrt{0.96826 \times 0.74796}} = 0.665$	M1 dep* A1 **[5]**	For structure of r For answer rounding to 0.66 or 0.67
		OR:		
		$\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 40.66/60 - (43.62/60 \times 55.15/60)$ $= 0.0094325$	M1*	For method for cov (x,y)
		$\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(0.96826/60)} = \sqrt{0.016137\ldots} = 0.1270$	M1*	For method for at least one msd or rmsd
		$\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(0.74796/60)} = \sqrt{0.012466} = 0.1117$	A1	For at least one of cov (x,y), msd or rmsd correct (to 2 sf)
		$r = \frac{\text{cov}(x,y)}{\text{rmsd}(x)\text{rmsd}(y)} = \frac{0.0094325}{0.1270 \times 0.1117} = 0.665$	M1 dep* A1	For structure of r For answer rounding to0.66 or 0.67
			[5]	Methods mixed – max M0M1A1M0A0

Question			Answer	Marks	Guidance	
1	(ii)		H_0: $\rho = 0$ H_1: $\rho > 0$ (one-tailed test)	B1	For H_0, H_1 in symbols. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population correlation coefficient.	
			where ρ is the population correlation coefficient	B1	For defining ρ. Condone omission of "population" if correct notation ρ is used, but if ρ is defined as the **sample** correlation coefficient then award **B0**. Allow "ρ is the pmcc".	
			For n = 60, 5% critical value = 0.2144	B1	For critical value	
			Since 0.665 > 0.2144, the result is significant.	M1	For sensible comparison leading to a conclusion provided that $\|r\| < 1$. The comparison can be in the form of a diagram as long as it is clear and unambiguous. Sensible comparison: e.g. 0.665 > 0.2144 is 'sensible' whereas 0.665 > - 0.2144 is 'not sensible'. Reversed inequality sign e.g. 0.665 < 0.2144 etc. gets max M1 A0.	
			Thus we have sufficient evidence to reject H_0	A1	For reject H_0 o.e. FT their r and critical value from 5% 1-tail column.	
			There is sufficient evidence at the 5% level to **suggest** that there is **positive** correlation between FEV1 before and after the two-week course.	E1	For correct, **non-assertive** conclusion in context (allow 'x and y' for context). E0 if H_0 and H_1 not stated, reversed or mention a value other than zero for ρ in H_0.	
				[6]		

Question			Answer	Marks	Guidance	
1	**(iii)**		The underlying population must have a bivariate Normal distribution. Yes, since the scatter diagram appears to have a roughly elliptical shape.	B1 E1 **[2]**	Condone "bivariate Normal distribution", "underlying bivariate Normal distribution", but **do not allow** "the **data** have a bivariate Normal distribution" Condone 'oval' or suitable diagram	
1	**(iv)**		The significance level is the probability of rejecting the null hypothesis when in fact it is true.	E1* E1dep* **[2]**	For "probability of rejecting H_0" or "probability of a significant result". For "when H_0 is true"	
1	**(v)**		$\sum x = 43.62 + 0.45 = 44.07$ $\sum y = 55.15 - 0.45 = 54.70$ $\sum xy = 40.66$ $\sum x^2 = 32.68 + 1 - 0.55^2 = 33.3775$ $\sum y^2 = 51.44 - 1 + 0.55^2 = 50.7425$	B1 B1 B1 **[3]**	For $\sum x$ or $\sum y$ or $\sum xy$ For $\sum x^2$ or $\sum y^2$ (to 2 dp) For all correct (ignore n)	
2	**(i)**		P(At least one has red hair) = $1 - 0.97^{10}$ = 0.263	M1 A1 **[2]**	M1 for $1 - 0.97^{10}$ Allow 0.26	
2	**(ii)**		(Because X is binomially distributed), n is large and p is small. Mean = 1.8	E1 E1 B1 **[3]**	Allow "sample is large" for n is large Allow "$np < 10$" or "mean ≈ variance" for "p is small" Do not allow "the probability is small"	

Question			Answer	Marks	Guidance	
2	(iii)	(A)	$P(X = 2) = e^{-1.8}\frac{1.8^2}{2!} = 0.2678$ OR = 0.7306 – 0.4628 = 0.2678	M1 A1 **[2]**	For calculation for P(X = 2) FT their mean. Allow answer to 3sf.	
2	(iii)	(B)	$P(X > 2) = 1 - P(X \le 2) = 1 - 0.7306$ = 0.2694	M1 A1 **[2]**	$1 - P(X \le 2)$ used. e.g. $1 - P(X \le 2) = 1 - 0.4628$ gets M0 CAO	
2	(iv)		The mean (np = 1.8) is too small It is not appropriate to use a Normal approximation	E1* E1dep* **[2]**	For "mean is too small" or "mean < 10" For "not appropriate". Do not allow "p is too small".	
2	(v)		Binomial(5000, 0.03)	B1* B1dep* **[2]**	For binomial, or B(,) For parameters	
2	(vi)		Mean 5000 × 0.03 = 150 Variance = 5000 × 0.03 × 0.97 = 145.5 Using Normal approx. to the binomial, $X \sim N(150, 145.5)$ $P(X \ge 160) = P\left(Z \ge \frac{159.5 - 150}{\sqrt{145.5}}\right)$ $= P(Z > 0.7876) = 1 - \Phi(0.7876) = 1 - 0.7846$ = 0.215 (to 3 sig.fig.)	B1 B1 B1 M1 A1 **[5]**	For mean (soi) For variance (soi) For continuity corr. For probability using correct tail and structure (condone omission of/incorrect c.c.) CAO, (Do not FT wrong or omitted CC) Allow 0.2155. Do not allow 0.216	

Question			Answer	Marks	Guidance	
3	(i)		$\text{P}(Y = 76) = \text{P}\left(\frac{75.5-76}{12} \le Z \le \frac{76.5-76}{12}\right)$	B1	For one correct continuity correction used	
			$= \text{P}(-0.04166\ldots < Z < 0.04166\ldots)$ $= \Phi(0.04166\ldots) - (1 - \Phi(0.04166\ldots))$	M1	For standardizing	
			$= 2 \times \Phi(0.04166\ldots) - 1$ $= 2 \times 0.5167 - 1$	M1	For correctly structured probability calculation.	
			$= 0.0334$	A1	CAO inc use of diff tables. Allow 0.0330 – 0.0340 www.	
				[4]		
3	(ii)		$\text{P}(Y \ge 80) = \text{P}\left(Z \ge \frac{79.5-76}{12}\right)$	B1	For correct cc used	
			$= \text{P}(Z > 0.2917) = 1 - \Phi(0.2917)$	M1	For correct structure	
			$= 1 - 0.6148 = 0.3852 = 0.385$ to 3 sig fig	A1	CAO do not allow 0.386	
				[3]		
3	(iii)		$3 \times 0.3852 \times 0.6148^2 = 0.4368$	M1	$3 \times$ their $p \times (1 -$ their $p)^2$	
				A1	FT their p. Allow 2sf if working seen.	
				[2]		

Question			Answer	Marks	Guidance	
3	(iv)		**EITHER:** P(Score $\geq k$) = 0.1			
			$\Phi^{-1}(0.9) = 1.282$	B1	For 1.282	
			$\frac{k-76}{12} = 1.282$	M1	Allow k – 0.5 used for k. Positive z used.	
			$k = 76 + (1.282 \times 12) = 91.38$ or $k = 76 + 0.5 + (1.282 \times 12) = 91.88$	A1	For 91.38 or 91.88	
			91.38 > 90.5 or 91.88 > 91	M1	Relevant comparison (e.g. diagram)	www
			so lowest reported mark = 92	A1		
			OR Trial and improvement method	M1	M1 for attempt to find P(Mark ≥ integer)	
			P(Mark ≥ 91) = P(Score ≥ 90.5) = 0.1135	A1	A1 for 0.1135	
			P(Mark ≥ 92) = P(Score ≥ 91.5) = 0.0982	A1	A1 for 0.0982	
			P(Mark ≥ 91) > 10% and P(Mark ≥ 92) < 10%	M1	M1 for comparisons	www
			so lowest reported mark = 92	A1		
				[5]		
3	(v)		$P(Y \leq 50) = 0.2$			
			$P(Z \leq \frac{50.5-\mu}{12}) = 0.2$	B1	For 50.5 used	
			$\frac{50.5-\mu}{12} = \Phi^{-1}(0.2) = -0.8416$	B1	For –0.8416. Condone – 0.842 Condone 0.8416 if numerator reversed.	
				M1	For structure.	
			$\mu = 50.5 + (12 \times 0.8416) = 60.6$	A1	CAO	
				[4]		

Question			Answer	Marks	Guidance	
4	(i)		H_0: no association between sex and artist preferred H_1: some association between sex and artist preferred	B1	For both hypotheses in context	
			EXPECTED: Monet / Renoir / Degas / Cézanne Male: 12.13 / 28 / 13.07 / 16.8 Female: 13.87 / 32 / 14.93 / 19.2	M1 A2	For expected values (to 2 dp where appropriate) (allow A1 for at least one row or column correct)	
			CONTRIB'N: Monet / Renoir / Degas / Cézanne Male: 1.4081 / 0.3214 / 1.8626 / 0.2881 Female: 1.2321 / 0.2813 / 1.6298 / 0.2521	M1 A2	For valid attempt at $(O–E)^2/E$ For all correct (to 2 dp) and presented in a table or clear list. (Allow A1 for at least one row or column correct)	NB:These three marks cannot be implied by a correct final value of X^2
			$X^2 = 7.28$	B1	Allow 7.27	www
			Refer to χ_3^2	B1	for 3 deg of f	
			Critical value at 10% level = 6.251	B1	CAO for cv No FT from here if wrong or omitted, unless *p*-value used instead	B1 for *p*-value = 0.0636
			Result is significant	B1	FT their X^2	
			There is evidence to **suggest** that there is some association between sex and artist preferred	E1	For correct (FT their X^2), non-assertive conclusion, in context.	
			NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1 or final E1	**[12]**		

Question			Answer	Marks	Guidance	
4	(ii)		Monet: **More females and fewer males than expected** prefer Monet, as indicated by l**arge contribution(s)** (of 1.4081 and 1.2321).	E1* E1dep*	FT their table of contributions	NB MAX 3/6 for answers not referring to contributions (explicitly or implicitly).
			Renoir: Preferences are much **as expected**, as indicated by s**mall contributions.**	E1		
			Degas: **Fewer females and more males than expected** prefer Degas, as indicated by **large contribution(s)** (of 1.8626 and 1.6298).	E1* depE1*		
			Cézanne: Preferences are much **as expected**, as indicated by s**mall contributions**.	E1		SC1 Renoir and Cézanne have correct comments for both but without referring to contributions
				[6]		