4767 Statistics 2

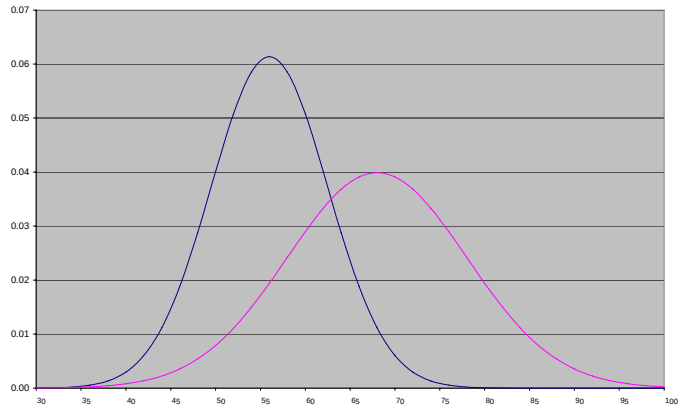
Question 1

(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 ⇒ predicted *y* = 7.42 × 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)* $P(X = 4) = e^{-9.4}\ \frac{9.4^4}{4!} = 0.0269$ (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*) $31 \times 94 = 2914$ Binomial (2914,0.1) (*B*)Use Normal approx with $\mu = np = 2914 \times 0.1 = 291.4$ $\sigma^2 = npq = 2914 \times 0.1 \times 0.9 = 262.26$ $P(X \leq 300.5) = P\left(Z \leq \frac{300.5 - 291.4}{\sqrt{262.26}}\right)$ $= P(Z \leq 0.5619) = \Phi(0.5619) = 0.7130$	B1 for binomial B1 dep, for parameters B1 B1 B1 for continuity corr. M1 for probability using correct tail A1 **cao,** (but FT wrong or omitted CC)	**2** **5**
			18

Question 3

(i)	$X \sim N(56, 6.5^2)$ $P(52.5 < X < 57.5) = P\left(\frac{52.5-56}{6.5} < Z < \frac{57.5-56}{6.5}\right)$ $= P(-0.538 < Z < 0.231)$ $= \Phi(0.231) - (1 - \Phi(0.538))$ $= 0.5914 - (1 - 0.7046)$ $= 0.5914 - 0.2954$ $= 0.2960$ (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)	$P(\text{5-year-old} < 62) = P\left(Z < \frac{62-56}{6.5}\right)$ $= \Phi(0.923) = 0.8220$ $P(\text{young adult} < 62) = P\left(Z < \frac{62-68}{10}\right)$ $= \Phi(-0.6) = 1 - 0.7257 = 0.2743$ P(One over, one under) $= 0.8220 \times 0.7257 + 0.1780 \times 0.2743$ $= 0.645$	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 $\Phi^{-1}(0.88) = 1.175$ $\frac{62-82}{\sigma} = -1.175$ $-20 = -1.175\,\sigma$ $\sigma = 17.0$	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

(i)

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^2_3
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

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

1

7

4

(ii)

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 = $\dfrac{68.3 - 67.4}{8.9/\sqrt{12}} = \dfrac{0.9}{2.57}$
= 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.

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

7

19