Least Squares Regression Line

Q1, (OCR 4767, Jan 2007, Q1i-iii)

(i)	$\bar{t} = 112.8,\ \bar{v} = 0.6$ $b = \frac{Svt}{Svv} = \frac{405.2 - 3 \times 564/5}{2.20 - 3^2/5} = \frac{66.8}{0.4} = 167$ OR $b = \frac{405.2/5 - 0.6 \times 112.8}{2.20/5 - 0.6^2} = \frac{13.36}{0.08} = 167$ 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 × 0.5 + 12.6 = 96.1 seconds (*B*) For 1.5 litres, predicted time = = 167 ×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**

Q2, (Jan 2008, Q1i-iii)

(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**

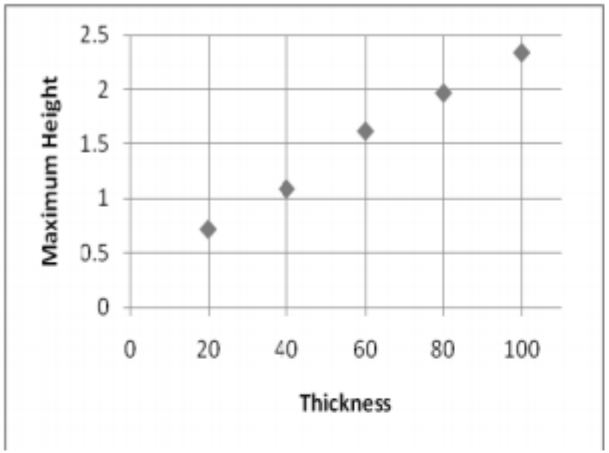
Q3, (OCR 4732, Jan 2005, Q9)

(i)	$\frac{264 - \frac{90 \times 15}{5}}{1720 - \frac{90^2}{5}}$ or $\frac{264 - 5 \times 18 \times 3}{1720 - 5 \times 18^2}$	M1		Formula correctly used
	= –0.06 AG	A1		–0.06 correctly obtained
	$y - {}^{15}/_{5} = -0.06(x - {}^{90}/_{5})$	M1		or $a = {}^{15}/_{5} - (-0.06) \times {}^{90}/_{5}$
	$y = 4.08 - 0.06x$	A1	4	Complete equation correct
(ii)	Substitute $x = 20.5$ ($y = 2.85$) Substitute $x = 19.5$ ($y = 2.91$) $2.91 - 2.85 = 0.06$	M1 M1 A1	3	Allow 20 ($y = 2.88$) or 20.49 Answer 0.06 or –0.06, c.w.d
(iii)	–0.6, 0.5	B1 B1	2	–0.6 correct 0.5 correct
(iv)	1.5 Calculated equation minimises this quantity	B1 B1	2	Not "Low value for Σe^2 means points near line"
(v)	$\bar{e} = \Sigma e_i/5$ = 0 $\Sigma e_i^2/5$ (– her $\bar{e}$)^2 = 0.3	M1 A1 M1 A1	4	$\Sigma e_i/5$ used Answer 0, cwd, cao $\Sigma e_i^2/5$ 0.3 only, must see -0^2 or -0 in variance. ie: No working: $\bar{e} = 0$: M1A1; Var = 0.3: M1A0

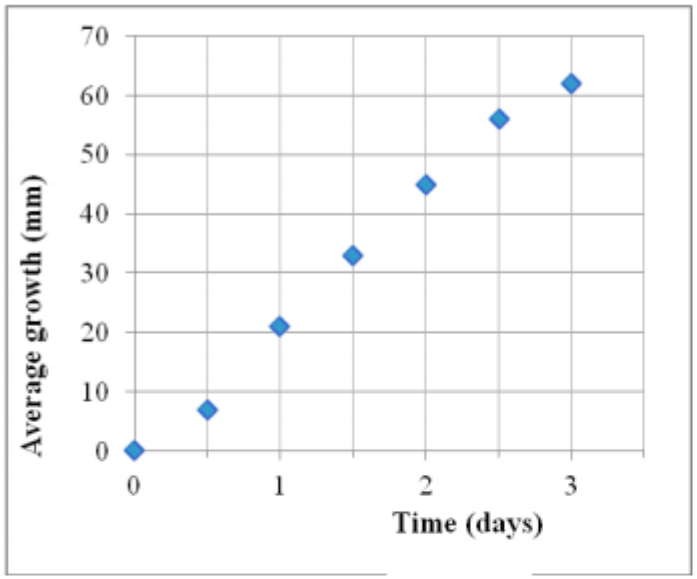
Q4, (OCR 4732, Jan 2009, Q2)

(i)(a)	$\dfrac{8736.9 - \dfrac{202 \times 245.3}{7}}{7300 - \dfrac{202^2}{7}}$ or $\dfrac{1658.24}{1470.86}$ $= 1.127\ldots$ (= 1.13 **AG**)	M1 A1 2	correct sub in any correct formula for b eg $\dfrac{236.8921}{210.1249}$ must see 1.127… ; 1.127.. alone: M1A1
(b)	$y - {}^{245.3}/_{7} = 1.13(x - {}^{202}/_{7})$ $y = 1.1x + 2.5$ (or 2.4) or $y = 1.13x + 2.43$	M1 A1 2	or $a = {}^{245.3}/_{7} - 1.13 \times {}^{202}/_{7}$ 2 sfs suff. (exact: $y = 1.127399..x + 2.50934\ldots$)
(ii)(a)	$(1.1(..) \times 30 + 2.5(..)) = 35.5$ to 36.5	B1f 1	
(b)	$(1.1(..) \times 100 + 2.5(..)) = 112.4$ to 115.6	B1f 1	
(iii)	(a) Reliable (b) Unreliable because extrapolated	B1 B1 2	Both reliable: B1 Ignore extras (a) more reliable than (b) B1 because (a) within data or (b) outside data B1

Q5, (OCR 4767, Jan 2013, Q1)

(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.
(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.	
(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: $h - \bar{h} = b(t - \bar{t})$			

Q6, (OCR 4767, Jun 2015, Q1i,ii,iv,v)

(i)			G1* G1dep* G1dep* **[3]**	Both axes labeled (allow *t* and *y*) with indication of scale for values of time BOD if (0,0) not clearly visible for values of average growth BOD if (0,0) not clearly visible. BOD if confusion arises from points plotted for part (v).	Allow axes interchanged Condone *x* for *t* (evenly spaced) visually correct SC1 for points having the correct distribution and G0* awarded. Line through origin should appear but this is rewarded in part **(v)**
(ii)		$\bar{t} = 1.5,\ \bar{y} = 32$ $b = \frac{S_{yt}}{S_{tt}} = \frac{490 - (224 \times 10.5/7)}{22.75 - 10.5^2/7} = \frac{154}{7} = 22$ OR $b = \frac{490/7 - (32 \times 1.5)}{22.75/7 - 1.5^2} = \frac{22}{1} = 22$ hence least squares regression line is: $y - \bar{y} = b(t - \bar{t})$ $\Rightarrow\ y - 32 = 22\,(t - 1.5)$ $\Rightarrow\ y = 22\,t - 1$	B1 M1 A1 M1 A1 **[5]**	For $\bar{t}$ and $\bar{y}$ seen or implied by final answer. For attempt at gradient (*b*) For 22 cao For equation of line CAO	Seen either in calculating *b* or in forming the equation of the line. Correct structure needed. See additional notes. FT their $\bar{t}$ and $\bar{y}$ for M1 With their $b > 0$, $\bar{t}$ and $\bar{y}$ A0 for $y = 22x - 1$

(iv)		$(22 \times 5) - 1 = 109$ Likely to be **unreliable as extrapolation** (oe)	B1 B1 **[2]**	Estimate calculated using equation	FT their equation
(v)		$a = \frac{490}{22.75} = 21.5\,38\ldots = 21.5$ (3 s.f.) Equation is $y = 21.5t$ Line plotted on diagram	M1 A1 A1 A1 **[4]**	 Allow $y = 21.54t$ CAO For line correctly plotted CAO A0 if axes not scaled or $a \neq 21.5$ to 3 sf	 Allow $y = (280/13)t$ Through (0,0) and between (3, 64) and (3,65)

Q7, (OCR 4732, Jun 2016, Q2ii)

ii	a	'Increased' and Positive gradient or positive coeff of n or 'Output goes up by 0.6 each month' Both needed	B1 **[1]**	'Increased' and values of z shown as follows: at least 6 values or 1st and last values or 1st, or 2nd or 3rd or 4th **and** 9th or10th or 11th or 12th ie 17.6 or 18.2 or 18.8 or 19.4 **and** 22.4 or 23 or 23.6 or 24.2	'Increased' and 'Value of $0.6n$ increases as n increases'	NOT: 'Increased' and 'Value of z incr *as* n incr' 'z incr as no. of mths incr
ii	b	$\bar{n} = 6.5$ or $\frac{78}{12}$ oe seen $\bar{z} = 0.6\times$'6.5' + 17 alone, eg not ÷12 or $17 = \bar{z} - 0.6\times$'6.5' oe $\bar{z} = 20.9$	B1 M1 A1 **[3]**	 or (0.6×1+17+0.6×2+17+ . . .0.6×12+17)÷12 oe or '250.8'÷12 M1 ft their '6.5' only if comes from ....÷12 cao	 Long method, all correct terms seen and ÷12 M1 NB ans 20.9 may not score the B1	
ii	c	Total output = "20.9" × 12 251 (3 sf)	M1 A1f **[2]**	or 0.6×1+17+0.6×2+17+ . . .0.6×12+17 oe or eg $\frac{88}{5} + \frac{91}{5} + \ldots + \frac{121}{5}$ oe ft their (ii)(b)	Long method, all correct terms seen Not ISW, eg 25100 scores A0, even if 251 seen	
			10			