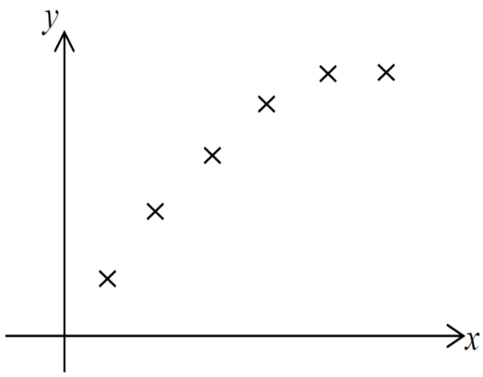
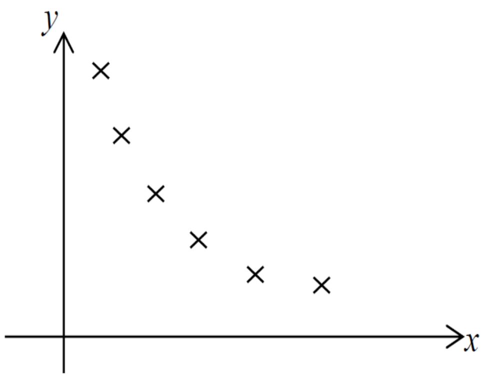
Correlation and Regression

1 MJC/2/9

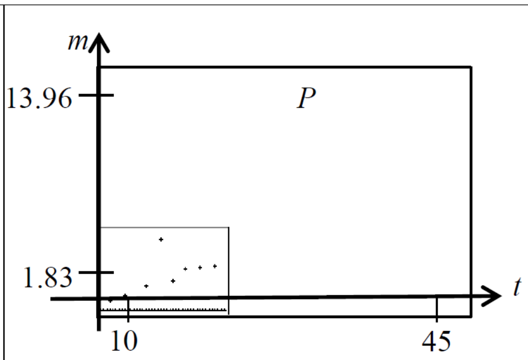
(a)



(aii)



bi)



(bii) With the removal of the point P, **<u>as t increases, m increases at an approximately constant rate</u>** which is consistent with a linear model.

With the removal of the point P, **<u>the remaining points lie close to a straight line</u>** which is consistent with a linear model.

(biii) $m = 0.204t - 0.0323$ (3 s.f)

(biv) $m = 0.16934t + 0.72035$

$$\frac{40.21+k}{8} = 0.16934(31) + 0.72035$$

$\therefore k = 7.55$ (2 d.p)

2 HCI/2/9

9(i)

$\overline{x}=\dfrac{17.21+k}{6}$, $\overline{n}=3.5$

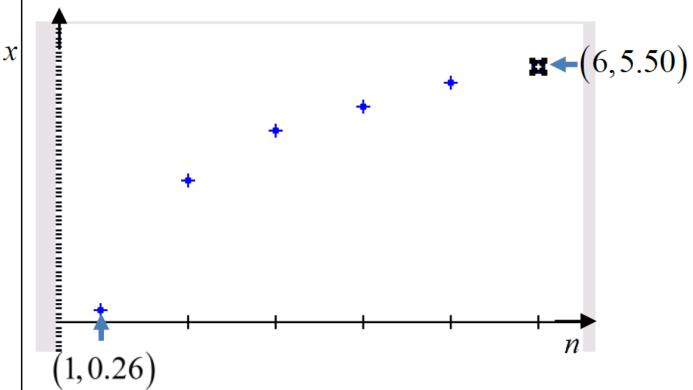
$(\overline{n},\overline{x})$ lies on the regression line, so

$x=0.943n+0.484$

$\dfrac{17.21+k}{6}=0.943(3.5)+0.484$

$k=5.50(2\text{ dp})$

9(ii)

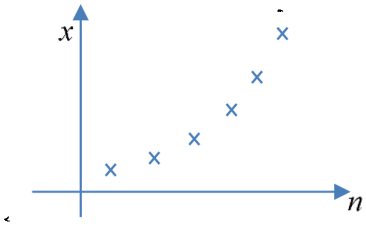


9 (iii)

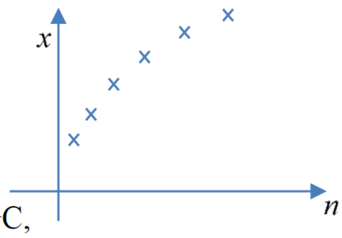
From the scatter plot, x and n have a curvilinear relationship. Therefore **<u>a linear model</u>** <u>is inappropriate</u> even though the product moment correlation coefficient is relatively high (i.e. 0.915).

9 (iv)

The graph of $x=a+bn^2$ is concave upwards (or increase at increasing rate) similar to the scatter plot



The graph of $x=a+b\ln n$ is concave downwards (or increase at decreasing rate) .



From GC,

$r_B=0.8212061688\approx0.821(3\text{ s.f.})$

$r_A=0.9859197289\approx0.986(3\text{ s.f.})$ which is closer to 1 and hence suggested a relatively stronger linear relationship between x and $\ln n$ as compared to x and n^2. Therefore <u>**Model (A)** is more appropriate</u>.

9 (iv)	Using GC, the regression line of x on $\ln n$ is $x = 0.638570346 + 2.869411323\ln n$ $b = 2.869411323 \approx 2.87$ $a = 0.638570346 \approx 0.639$ For $x \geq 10$, $2.869411323\ln n + 0.638570346 \geq 10$ $\ln n \geq 3.262491362$ $n \geq 26.1$ The number of reported cases is at least 1,000 in the 27th month
9 (v)	The value of a represents the **estimated** number of cases **(in hundreds)** of the virus reported in the **first month**.