Cai Ying Zhi U2140245C

a) $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 62$

$\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 276.5$

$S_{xx} = \sum_{i=1}^{10}(x-\bar{x})^2 = 2760$

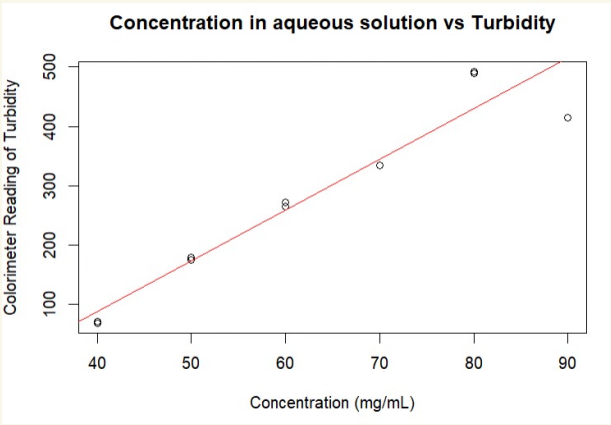
$S_{xy} = 23540 = \sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y})$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = 8.528986$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = -252.2971$

Values were obtained from R

∴ Model is $\hat{y}_i = -252.2971 + 8.528986x_i$



b) $e_i = y_i - \hat{y}_i$, $\hat{\beta}_0 = -252.247$, $\hat{\beta}_1 = 8.528986$, used for calculating e_i

$e_1 = 69 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 40)$

$= -19.862$

$e_2 = 175 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 50)$

$= 0.848$

$e_3 = 272 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 60)$

$= 12.558$

$e_4 = 335 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 70)$

$= -9.732$

$e_5 = 490 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 80)$

$= 59.978$

$e_6 = 415 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 90)$

$= -100.312$

$e_7 = 72 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 40)$

$= -16.862$

$e_8 = 265 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 60)$

$= 5.558$

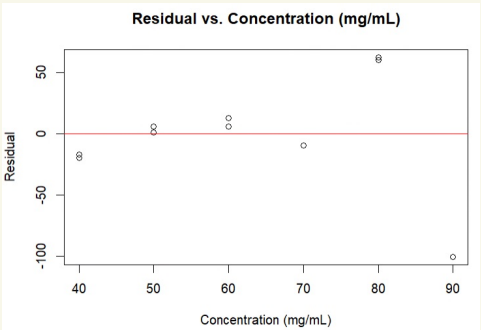
$e_9 = 492 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 80)$

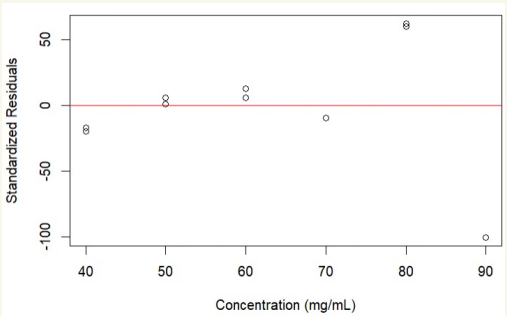
$= 61.978$

$e_{10} = 180 - (\hat{\beta}_0 + \hat{\beta}_1 \cdot 50)$

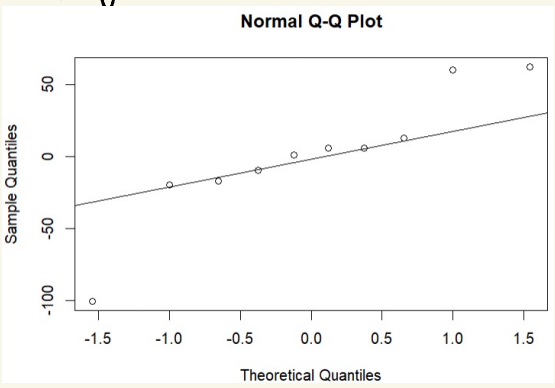
$= 5.848$

Continue on next page





Residual plot has a quadratic pattern,
So regression function for data may not be linear.
However, data size is not big enough as there are only 10 points.
Residuals are more spread out for X=80 and X=90, and have fan shape, implying the variance of error terms is not constant.
More data is needed.



Linearity of QQ Plot suggest error term follow normal distribution.

c)

	DF	SS	MS	F	P
SSR	1	200772.3	200772.3	86.829	1.434×10^{-5}
SSE	8	18498.2	2312		
SSTO	9	219270.5			

$SSR = \beta_1^2 S_{xx} = 200772.3$, $MS_{Reg} = SSR$

$SSTO = \sum_{i=1}^{10} (y_i - \bar{y})^2 = 219270.5$

$SSE = SSTO - SSR = 18498.2$, $S^2 = \frac{SSE}{n-2} = 2312$, $F = \frac{MS_{reg}}{s^2}$

$R^2 = \frac{SSR}{SSTO} = \frac{200772.3}{219270.5}$, $p = 1.434 \times 10^{-5}$

R^2 is very high, which implies SLR model is a good fit to the data.

X and Y have strong linear relationship

Small p-value for F-test, imply X significantly contribute to the explanation of response variable Y.