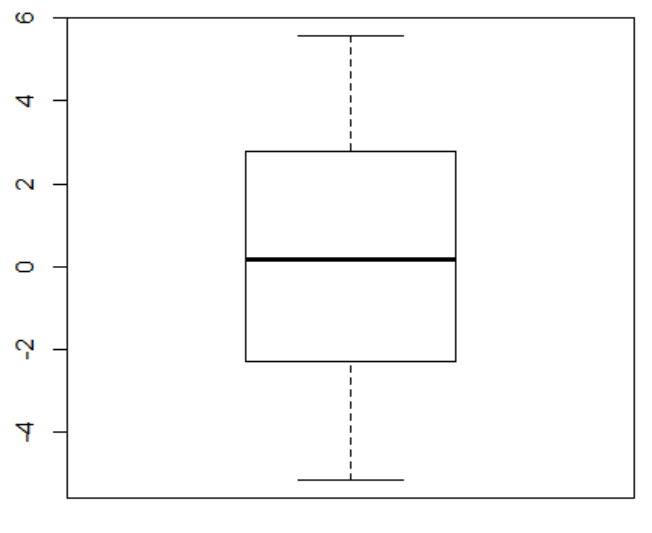
**Homework 3**

**Fangling Zhang**

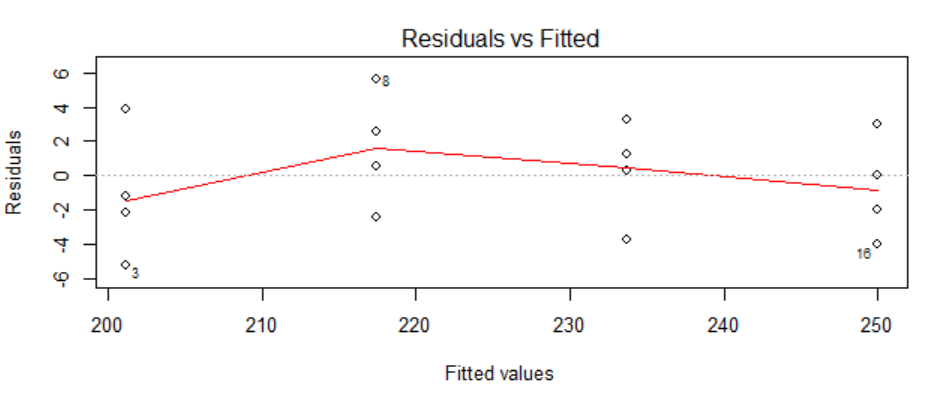
**3.6**

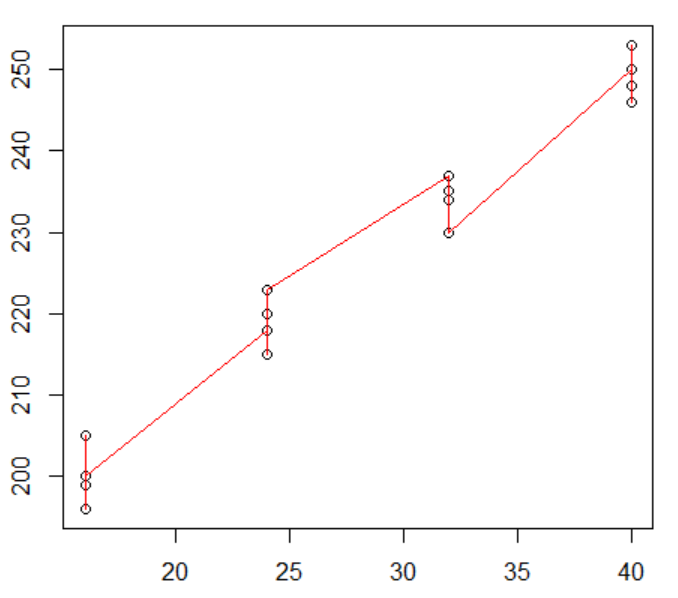
a. the box plot of the residuals is as follows:



We can see that the medium of residuals is about zero and the middle half of the residuals range from -2.5 to 3. The minimun and maximun residuals are about -6 and 6.

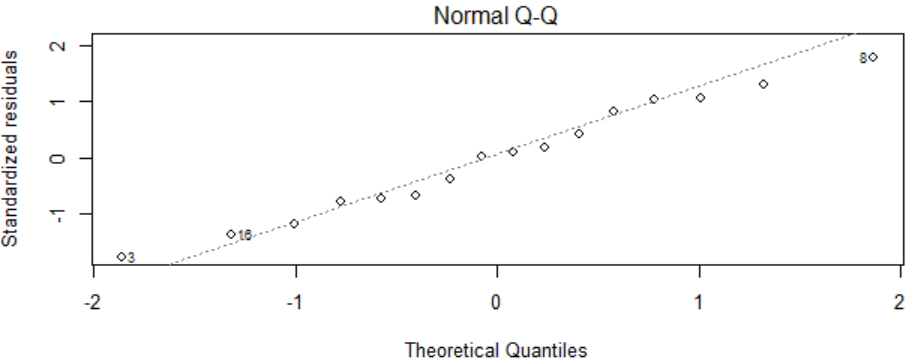
b. The plot of residuals against the fitted values is as follows.

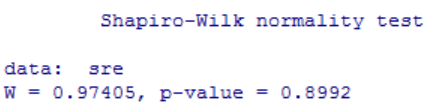




It shows that the resduals displays some systematic tendencies. Hence, some departures form regression model are evident and linear regression is not appropriate here. Another figure below is a loess fit model, we can use it to predit future.

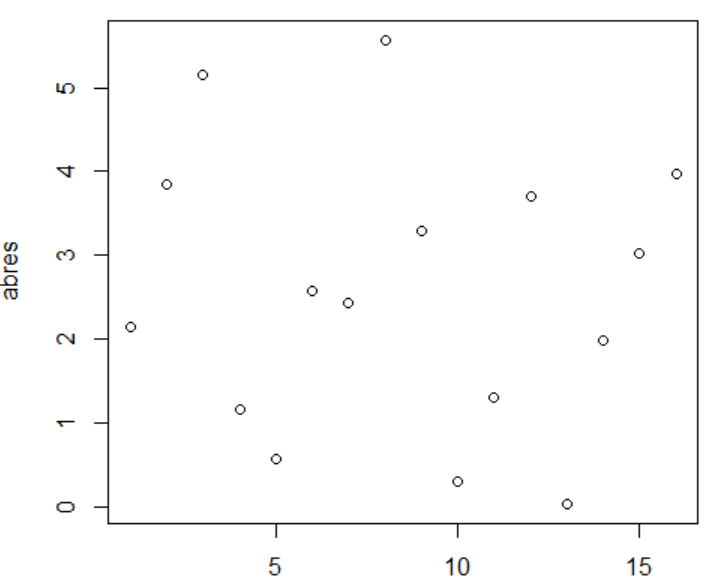
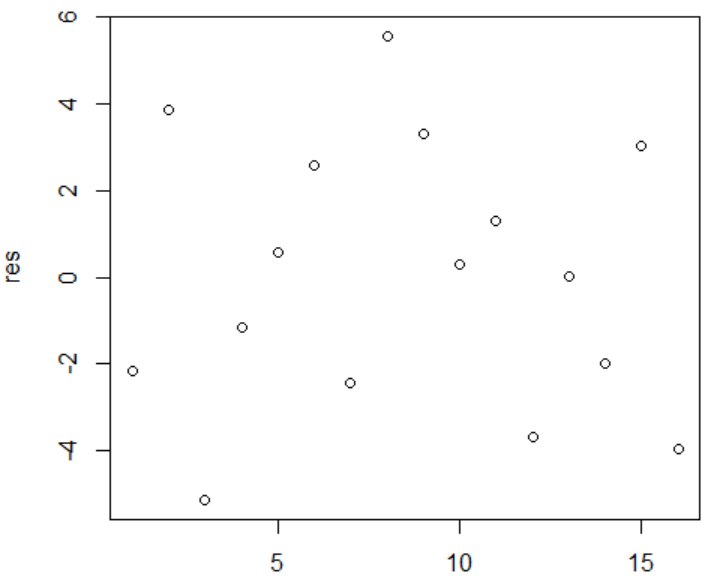
c. The plot of studentized residuals against the appropriate t quantile is as follows:





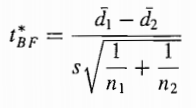
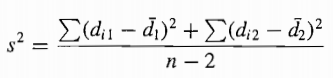
Therefore, we can not refuse the hypothesis that residuals are normally distributed.

d. The left figure is the residuals versus time plot. The right one is the absolute residuals versus time.



The plots suggest that the error variance is constant with time.

e. Brown-Forsythe test function:

where 

We get s^2=18.28 and tBF=7.48.

To control the α risk at 0.05, we requre t(0.975,14)=3.33.

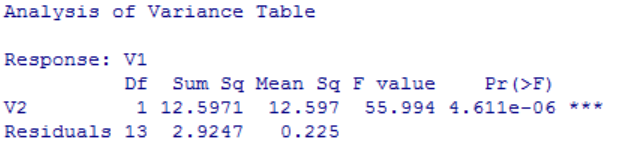
Here tBF>3.33, so we conclude the error variance is not constant.

This conclution is consistant with part (b).

**15.**

a. Y=2.575-0.324X

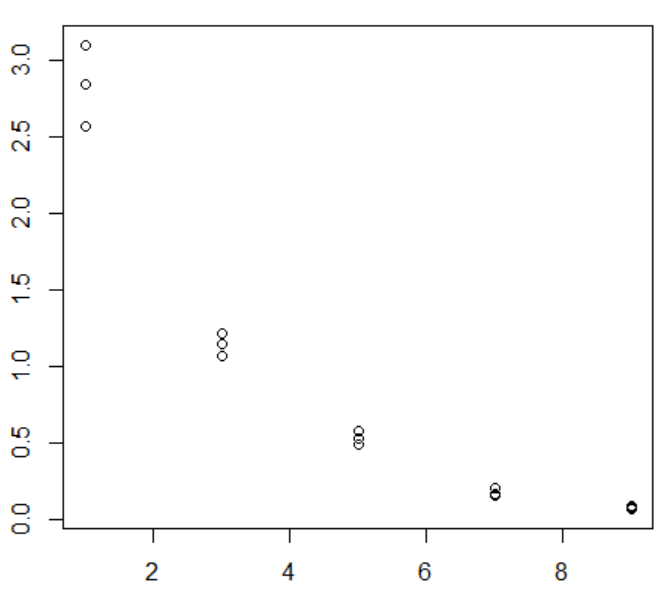
b. By using F test, p-value here is 4.611e-6, much smaller than α=0.025. Hence, we can refuse the hypothesis that a linear regression function is appropriate here. The conclusion is there is lack of fit of a linear regression function.



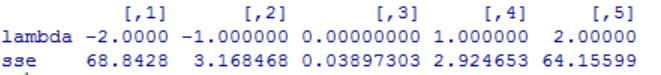
c. No, it does not indicate what regression function is appropriate.

**16.**

a. From the scatter plot below, we might try to transform Y by log or SQRT.



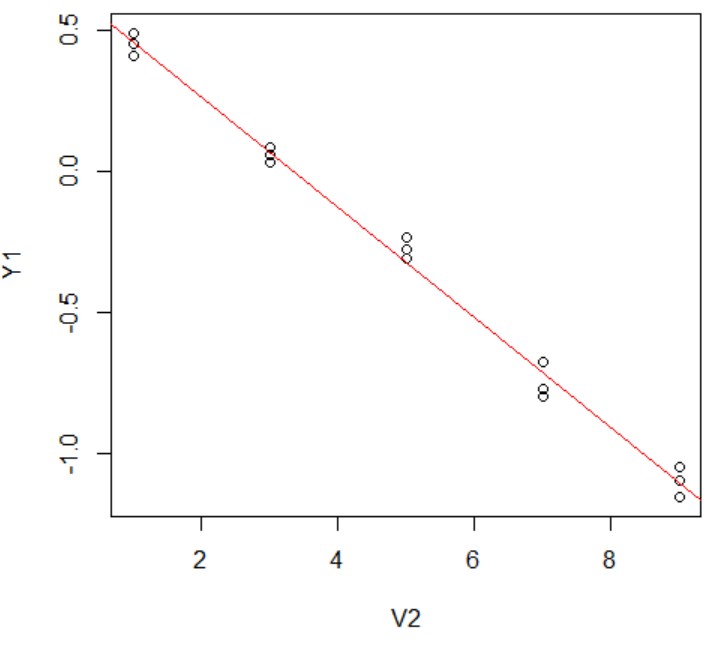
b. For each λ (lambda), the responding SSE are as follows:



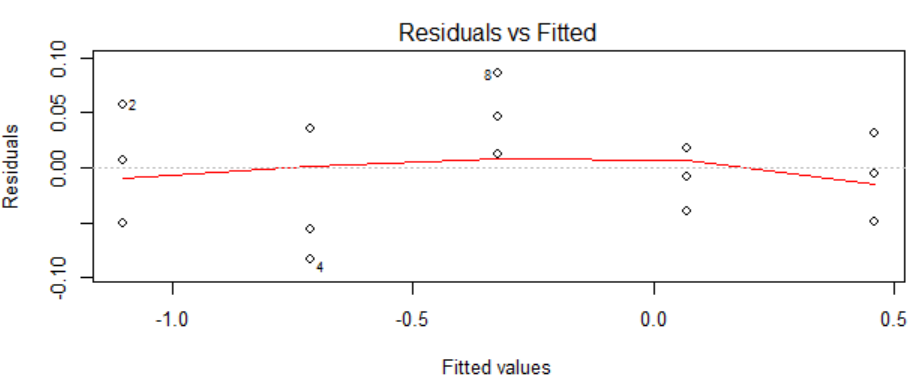
When λ=0, we get minimum SSE. Hence, Y’= log(Y) is suggested.

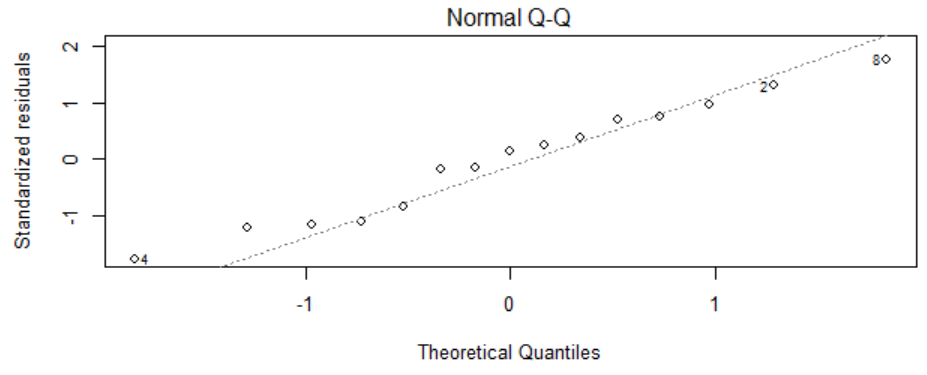
c. Y’=0.65-0.20X

d. From the following figure, the regression line appear to be a good fit to the transformed data.



e. The plot of the residuals against the fitted values suggests that the linear regression model is appropriate here. The normal probability plot suggests that the errors are normally distributed.

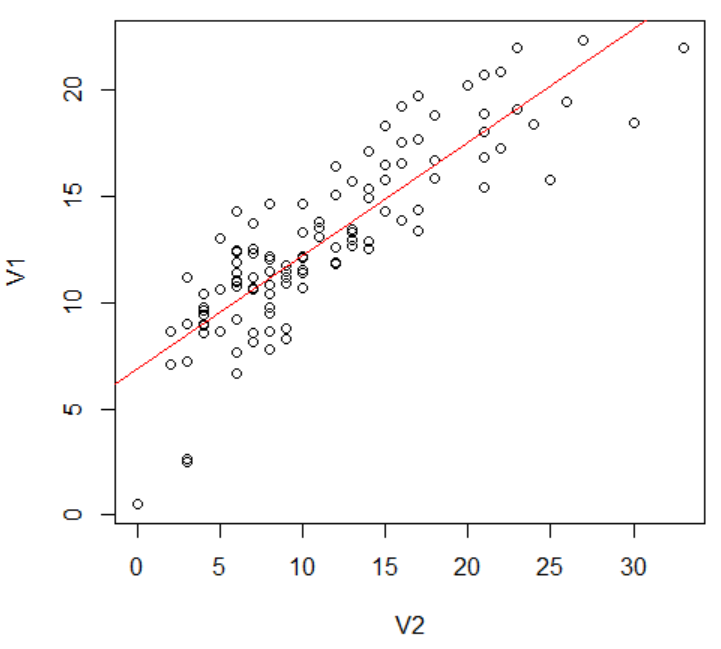




f. The estimated regression function in the original units:

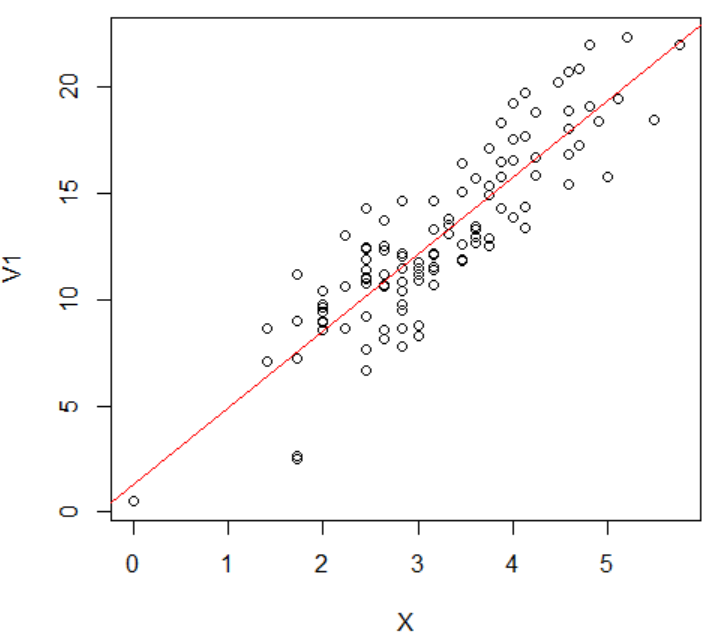
**18.**

a. A linear relation appear adequate here. Do not need a transformation on X or Y, because the relationship between X and Y does not seem to be in other shape.

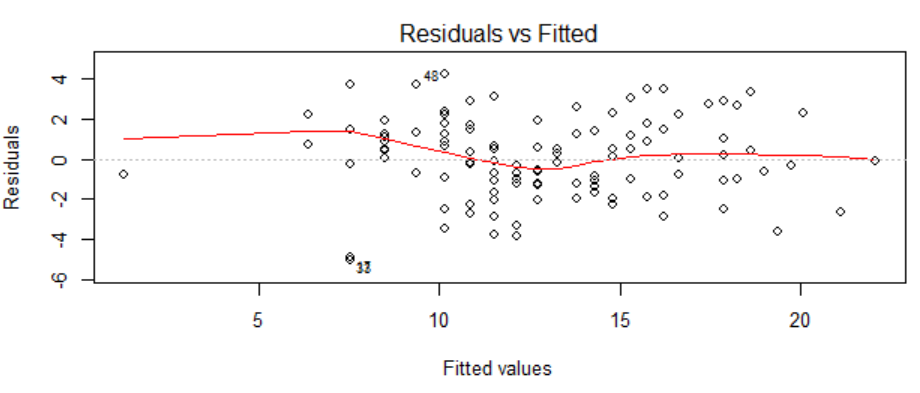


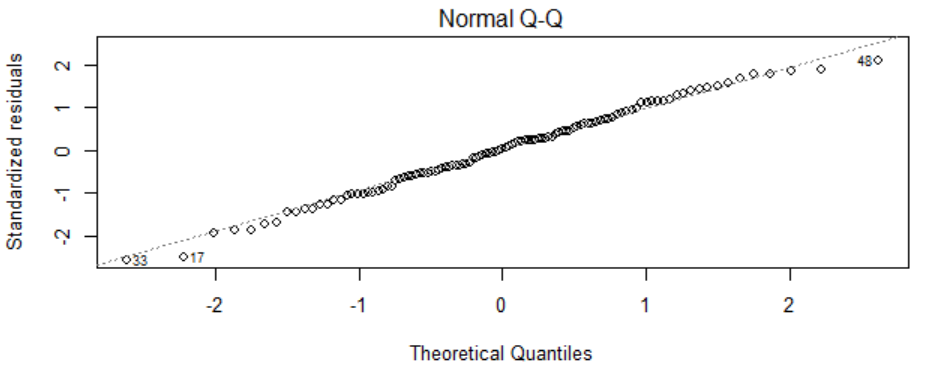
b. Y=1.25+3.62X’

c. The regression line also appears to be a good fit to the transformed data.



d. The plot of residuals and fitted values shows the linear model is appropriate here. The normal probability plot shows that the errors are normally distributed.





f. The estimated regression function in the original units: Y=1.25+3.62

**21.**