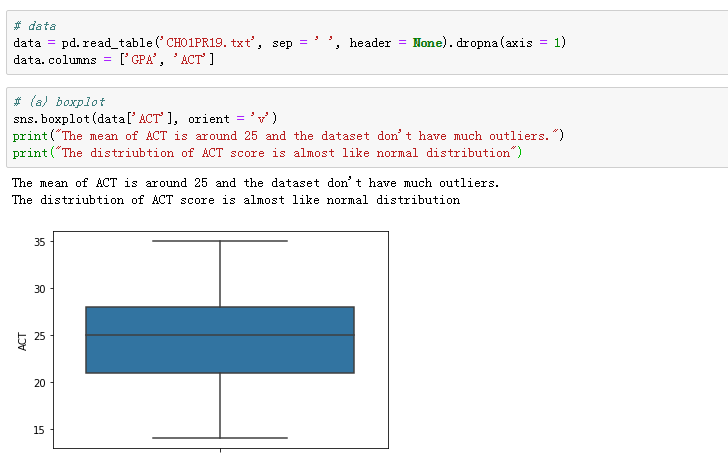
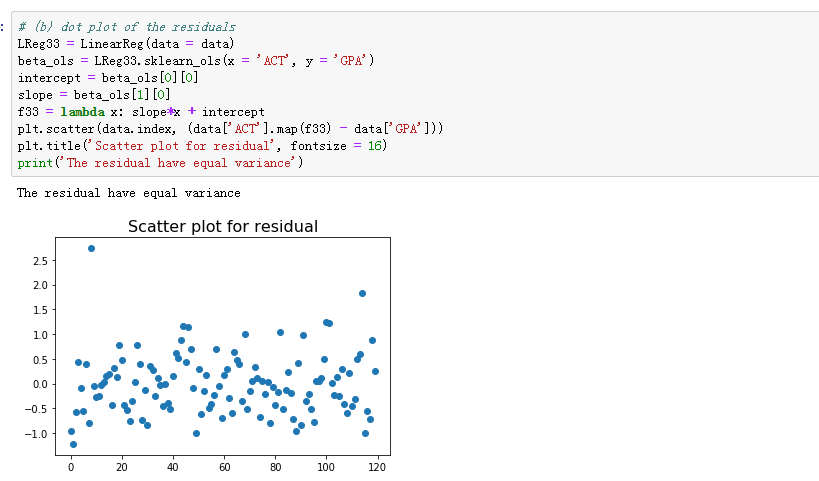
**Linear Regression Assignment 4**

**Name**: Eric Yuan **UNI**: qy2205

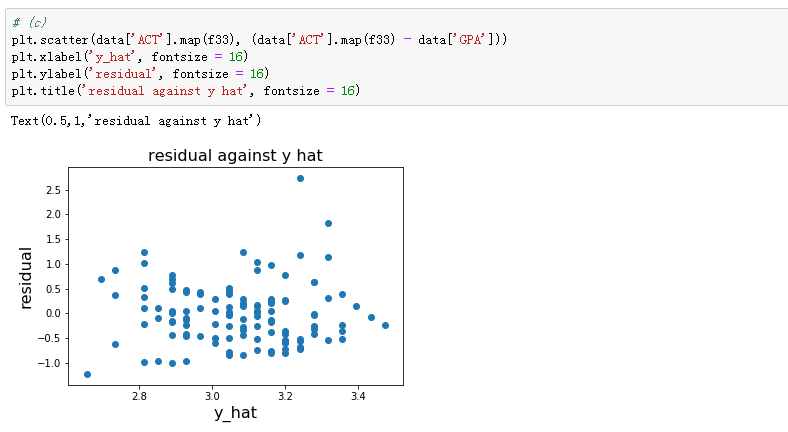
1. **Chapter 3, problem 3 (a, c, d, e, f)**
2. The code is showing below

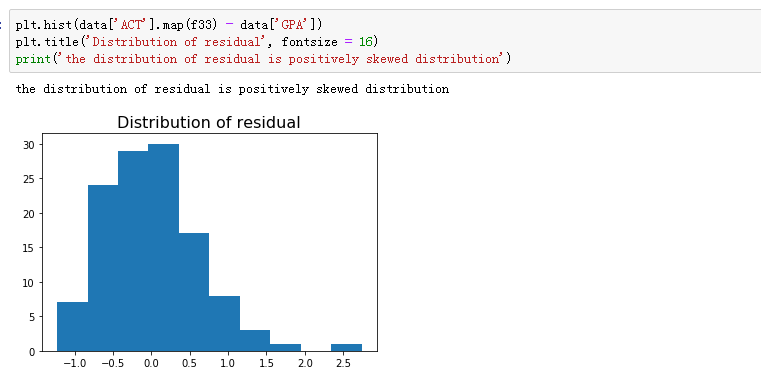






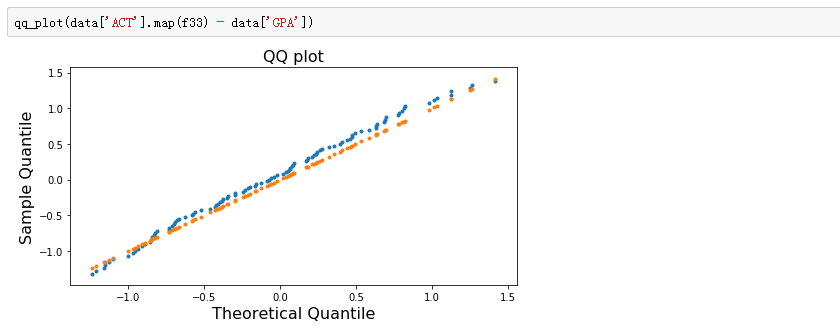
(c)

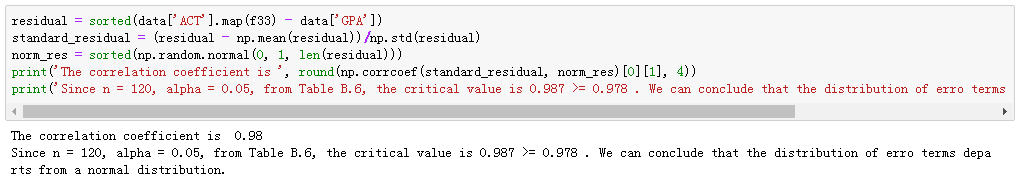




(d)







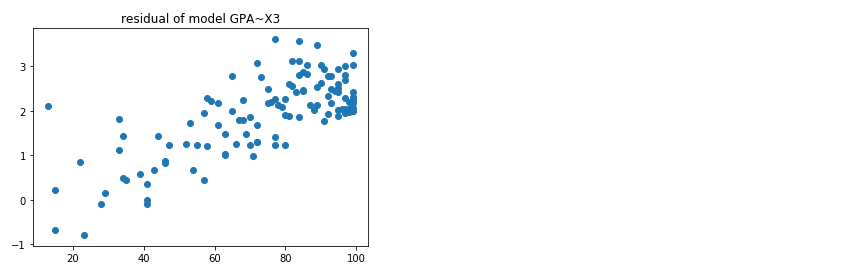
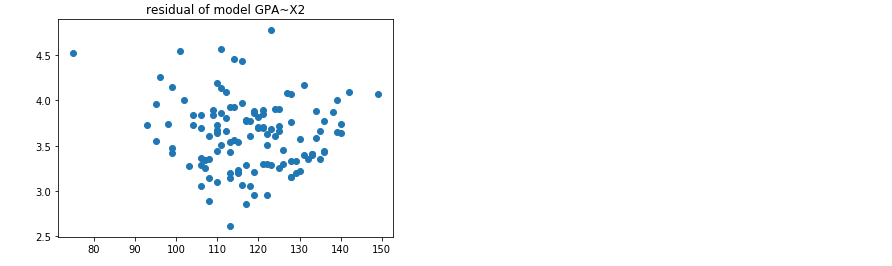
(e)

After calculating, n1 = 65, the mean of d1 = 0.438, n2 = 55, the mean of d2 = 0.5065

, thus the error variance is constant and does not vary with the level of X.

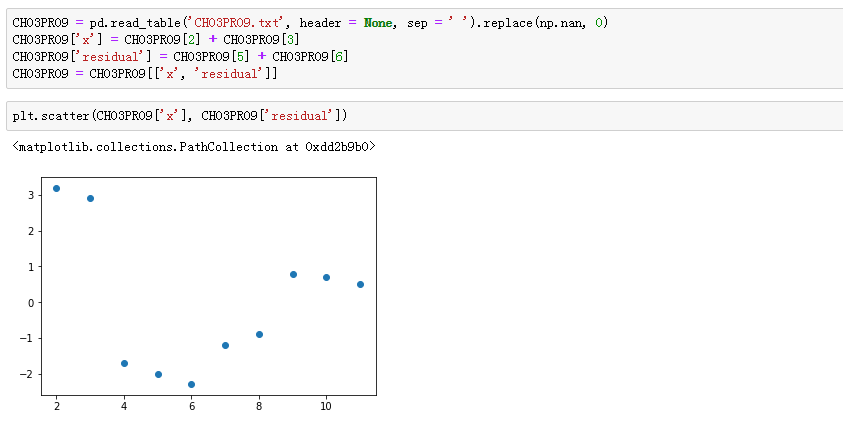
(f)





The residual of linear model GPA~X2 has equal variance. The residual of linear model GPA~X3 are increasing with X3. Thus, we can add intelligence test score (X2) to the original model.

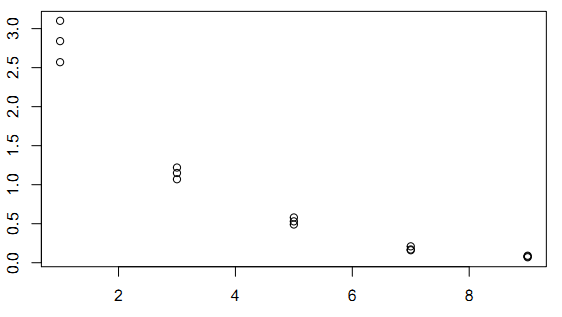
1. **Chapter 3 problem 9**



The variance of residual is big when x is small and small when x is big. We might use sqrt transformation to x to alleviate this problem.

1. **Chapter 3 problem 16**

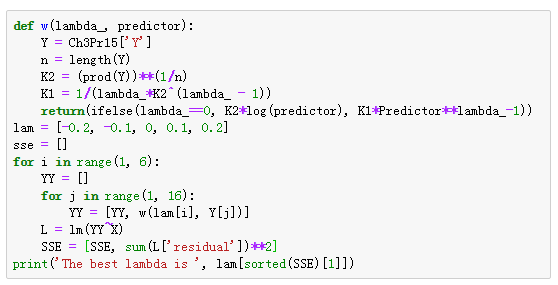
(a)



Based on the plot, *Y* andseems has correlated relationship, so we might apply log transformation to y.

(b)

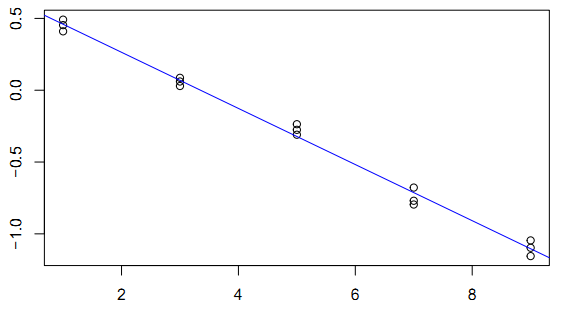
Where:

 C:\Users\ADMINI~1\AppData\Local\Temp\WeChat Files\177760f3c25593acc90905ba7f9477a.png

(c)

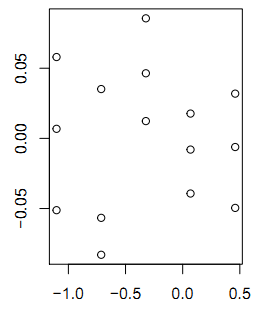
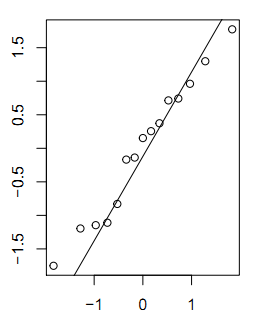
Run the regression, we could get:

(d)



Based on the plot, the regression result seems good.

(e)

**Figure 1: Residuals against fitted values Figure 2: Q-Q plot**

Based on the plot, we could see the residuals are normally distributed.

(f)

1. **Chapter 3 problem 23**

Set as different levels of . On each level, there are observations. Therefore, . The observed value of the response variable for the level of by

Full model:

Where:

are parameters

are independent

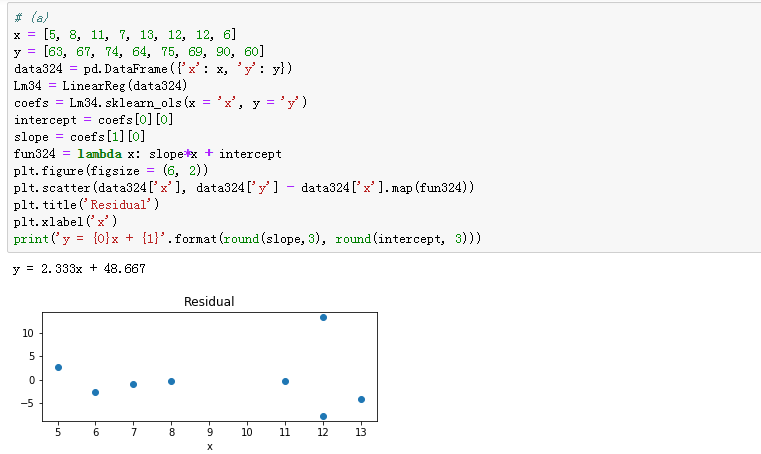
Reduced model:

Test Statistics:

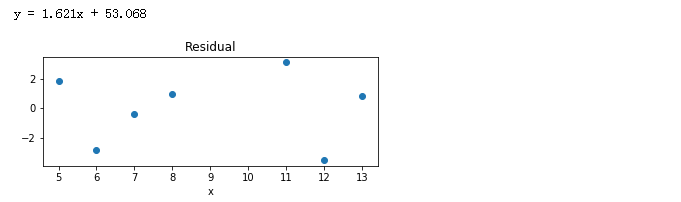
When n = 20, c = 10, = 10, .

1. **Chapter 3 problem 24**

(a)



(b)



The slope and intercept has changed a lot which means (12, 90) might be a outlier.

(c)

Where:

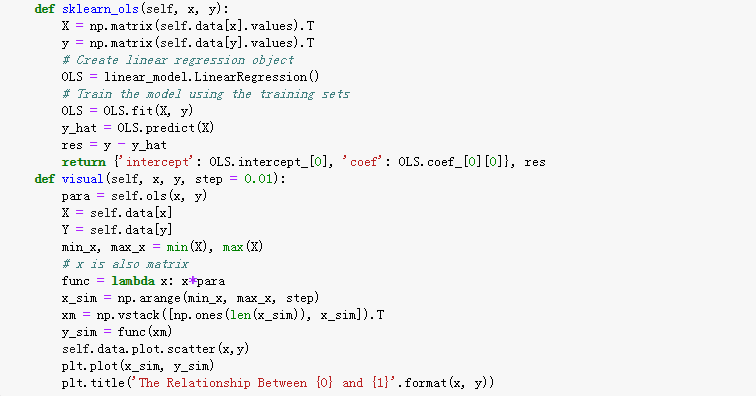
Thus,

is not in this interval which means it is an outlier.

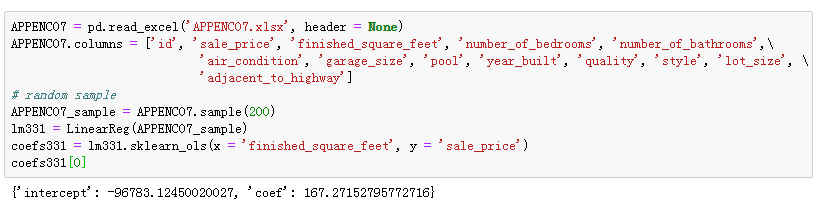
1. **Chapter 3 problem 31**

Notice that in this question, we use the class coded in homework1 and qq\_plot function coded in this homework.

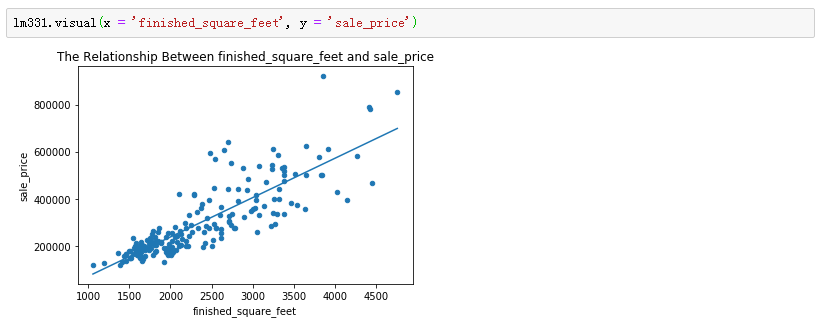




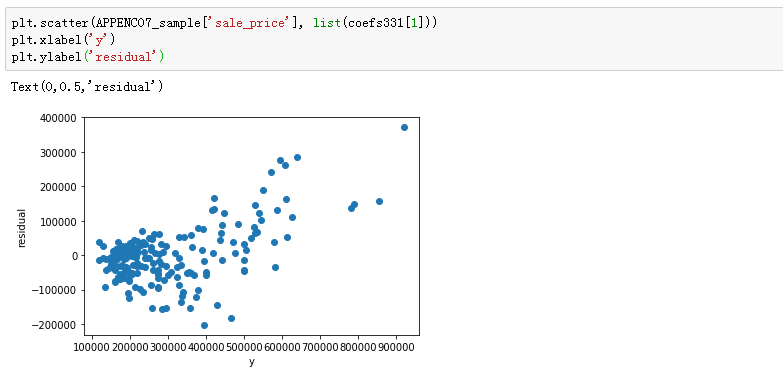
Clean the data and run the regression



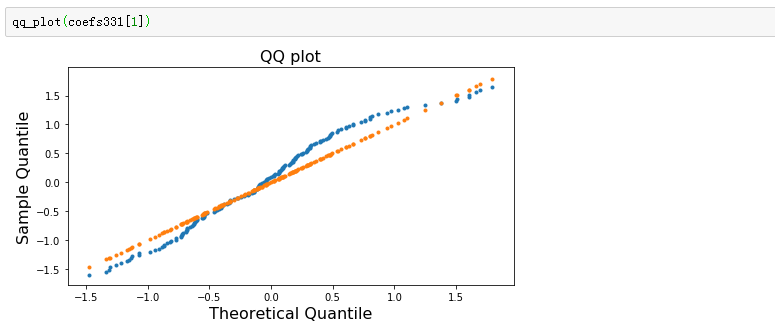
Plot the regression line



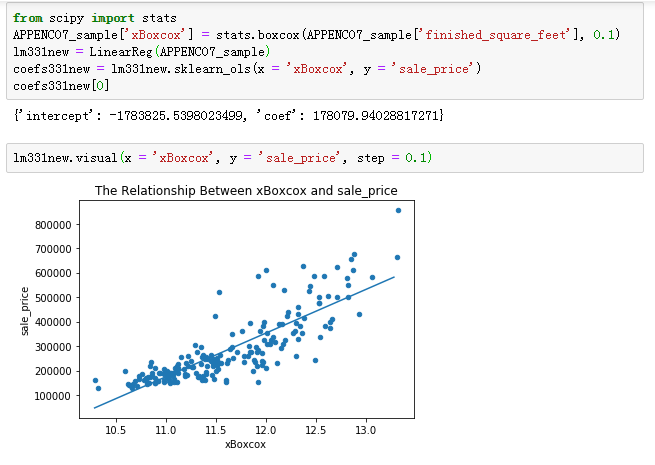
Plot the residual



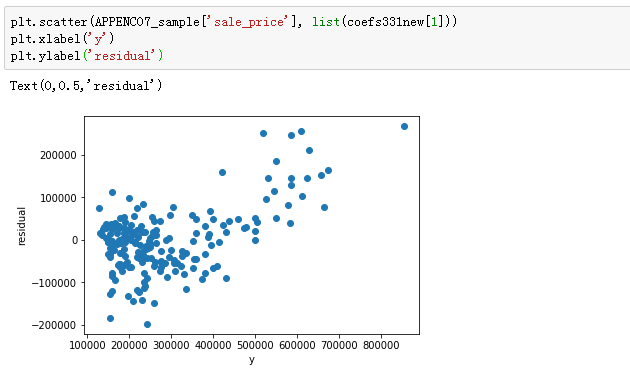
QQ plot

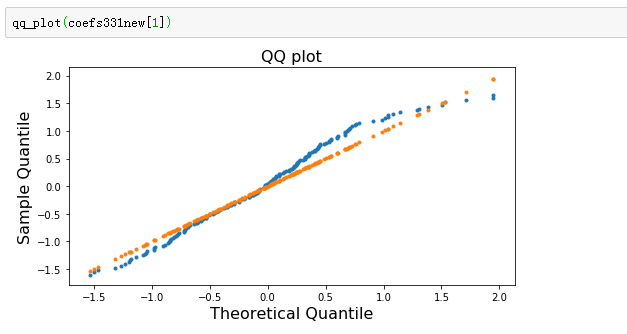


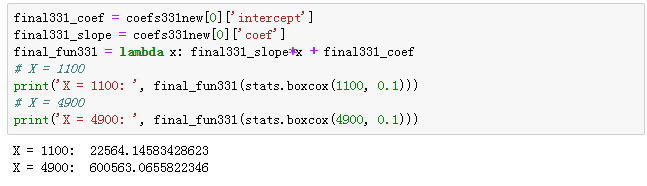
Box Cox transformation



New residual plot



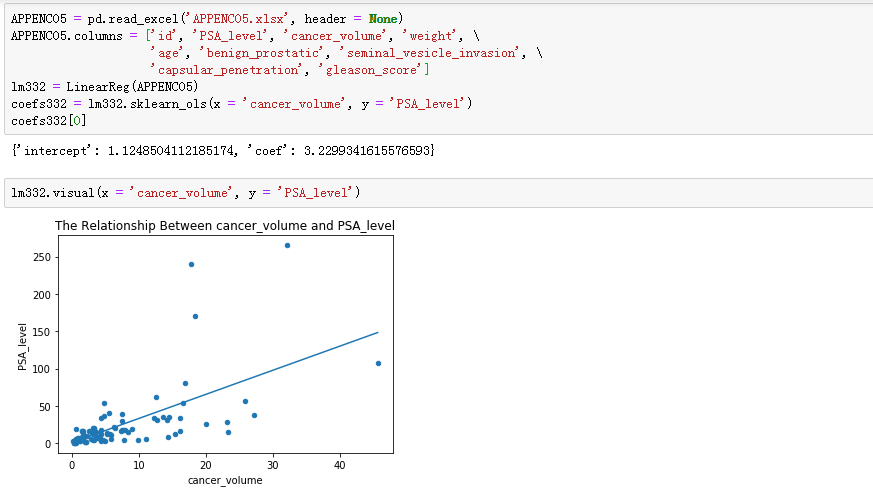


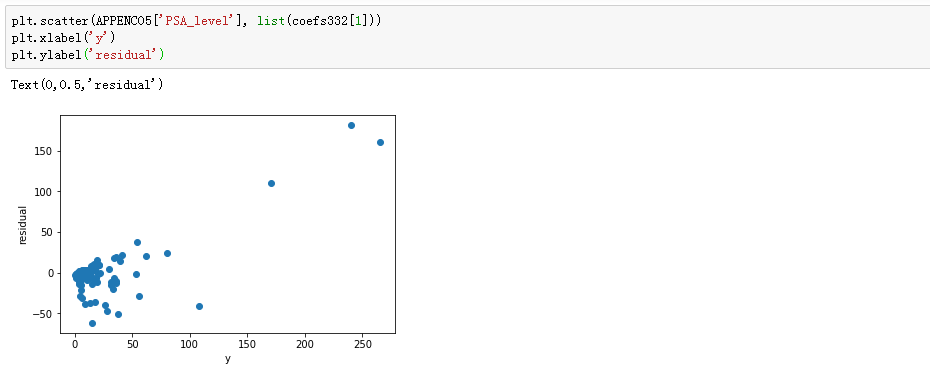


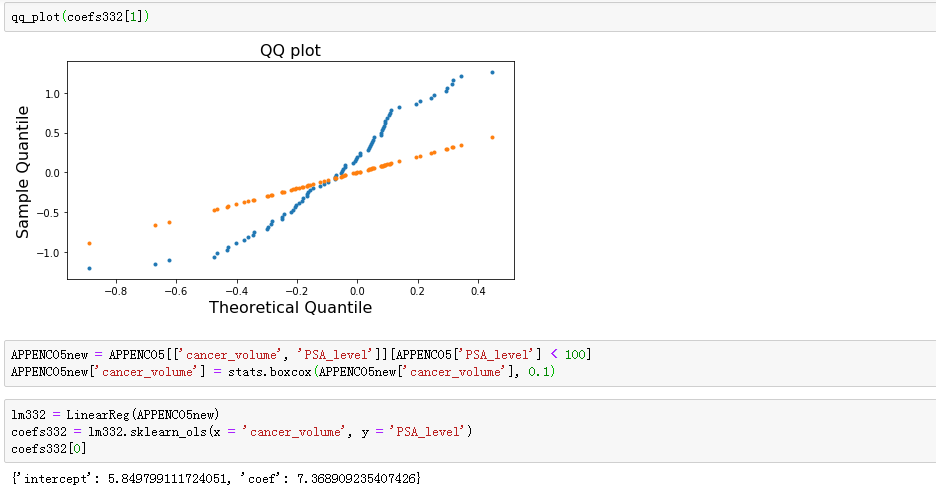
Based on the analysis and plot, we could see that after box-cox transformation, the residual almost have equal variance and follow the normal distribution. The strength of the model is that it satisfies all the assumptions, but the weakness is that it not perfectly satisfies all and the parameters are not robust enough.

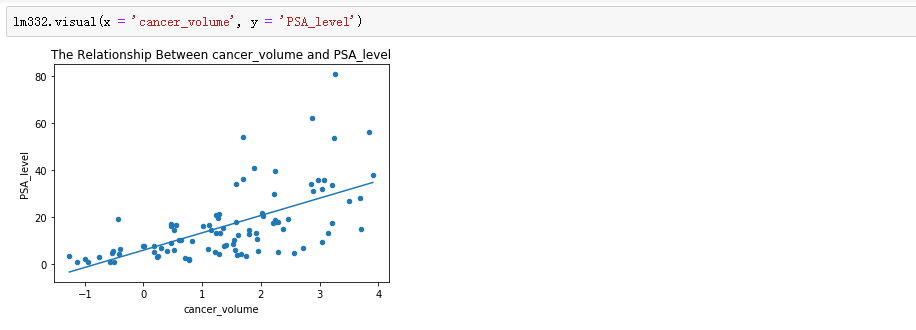
1. **Chapter 3 problem 32**

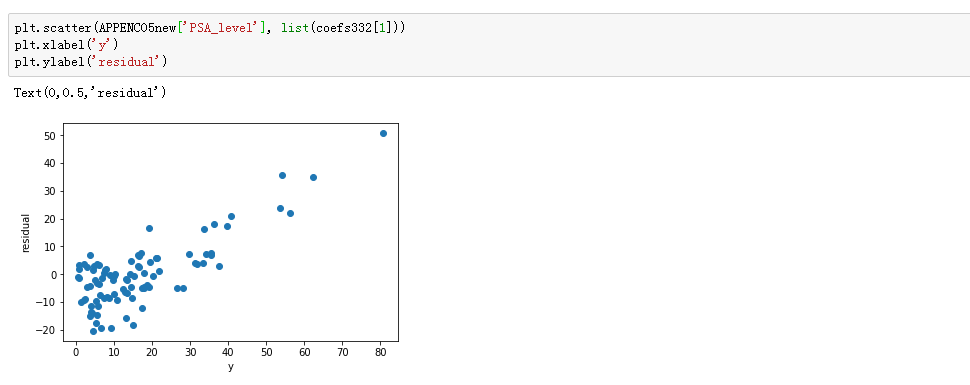
The techniques we use are almost the same as techniques we used in problem 31. The difference between those two problems is we remove the outliers in the dataset in this question.

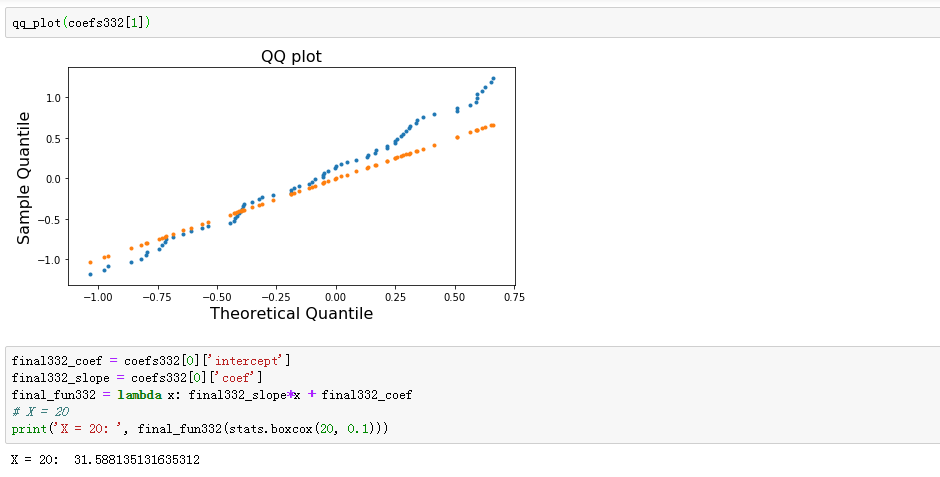












Based on the analysis and plot, we could see that after removing outliers and box-cox transformation, the residuals almost have equal variance and follow the normal distribution. The strength of the model is that it satisfies all the assumptions and the parameters are robust, but the weakness is that the residuals seem not follow normal distribution when x above 2.