Homework 5
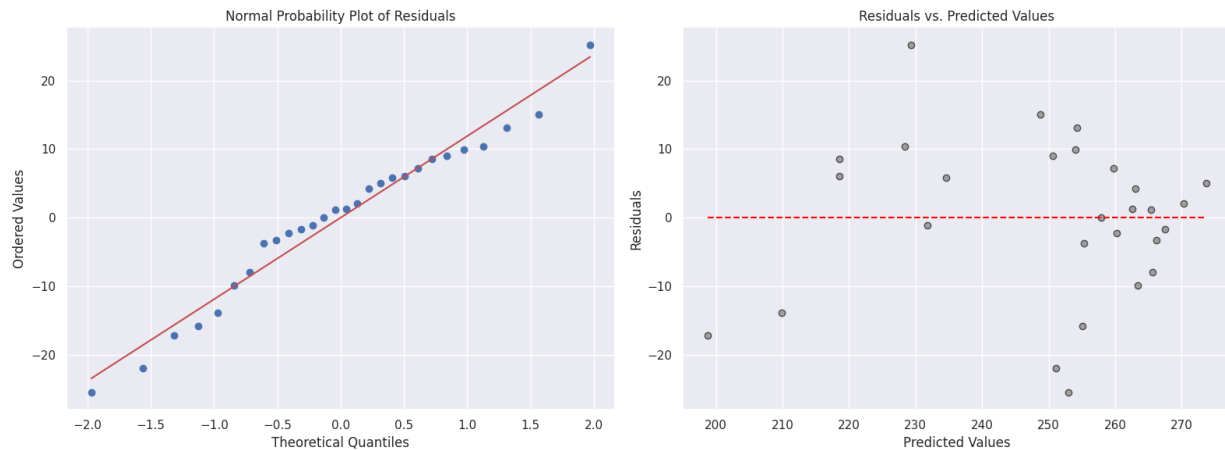Mark McCarthy | S395E958

Question 4.3

1) Consider the simple linear regression model fit to the solar energy data in Problem 2.3.

   1) Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?

   2) Construct and interpret a plot of the residuals versus the predicted response.

From the Normal Probability Plot of the residuals (QQ plot) we can visually asses that the residuals are holding to an acceptable level of normality. There is a concentration of residuals about the point 0.0 on the X axis (Theoretical ideal for a normal distribution on this graph can be visualized by first placing points on a standard normal distribution density curve at regular intervals with respect to spacing on the curve itself. Then projecting those regularly spaced points onto the x axis. This will show an increase in density of projected points as you approach the x values under the peak of the normal curve). This is accompanied by a sample size that is large enough, >20 according to our text (pg. 144), to be a stable plot bolstering this conclusion.

As for the residuals vs. predicted values plot we can see that there is a concentration of points in the higher range of the predicted values. This does not necessarily mean that the plot is out of expectation and the model is defective, however, more predicted values in the lower range would yield a better assessment of model fit.

Question 4.4

1) Consider the multiple regression model fit to the gasoline mileage data in Problem 3.5

1) Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?

2) Construct and interpret a plot of the residuals vs. the predicted response.

3) Construct and interpret the partial regression plots for this model.

4) Compute the studentized residuals and the R-student residuals for this model. What information is conveyed by these scaled residuals?
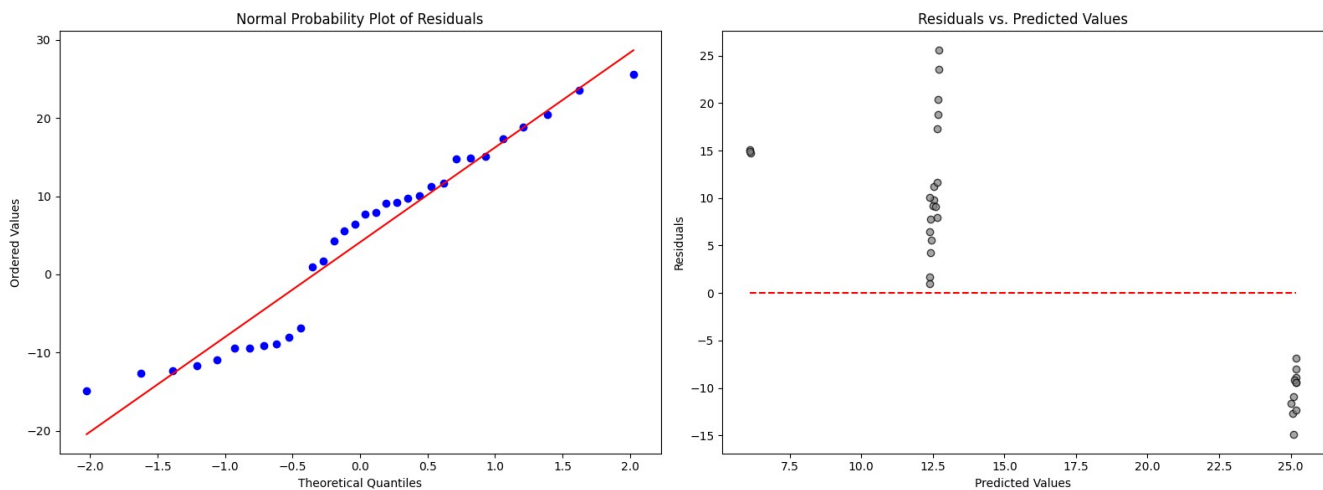
1 & 2 : The normal probability plot shows a somewhat normal plot of studentized residuals although the deviation in the lower quantiles seems odd. The plot of the residuals vs. the predicted values though points toward the model having deficiencies. This particular behavior does not seem to fit with any of the examples in the text but I would suspect either non-linearity or a mistake in the generation of the plot.

Partial Regression Plot
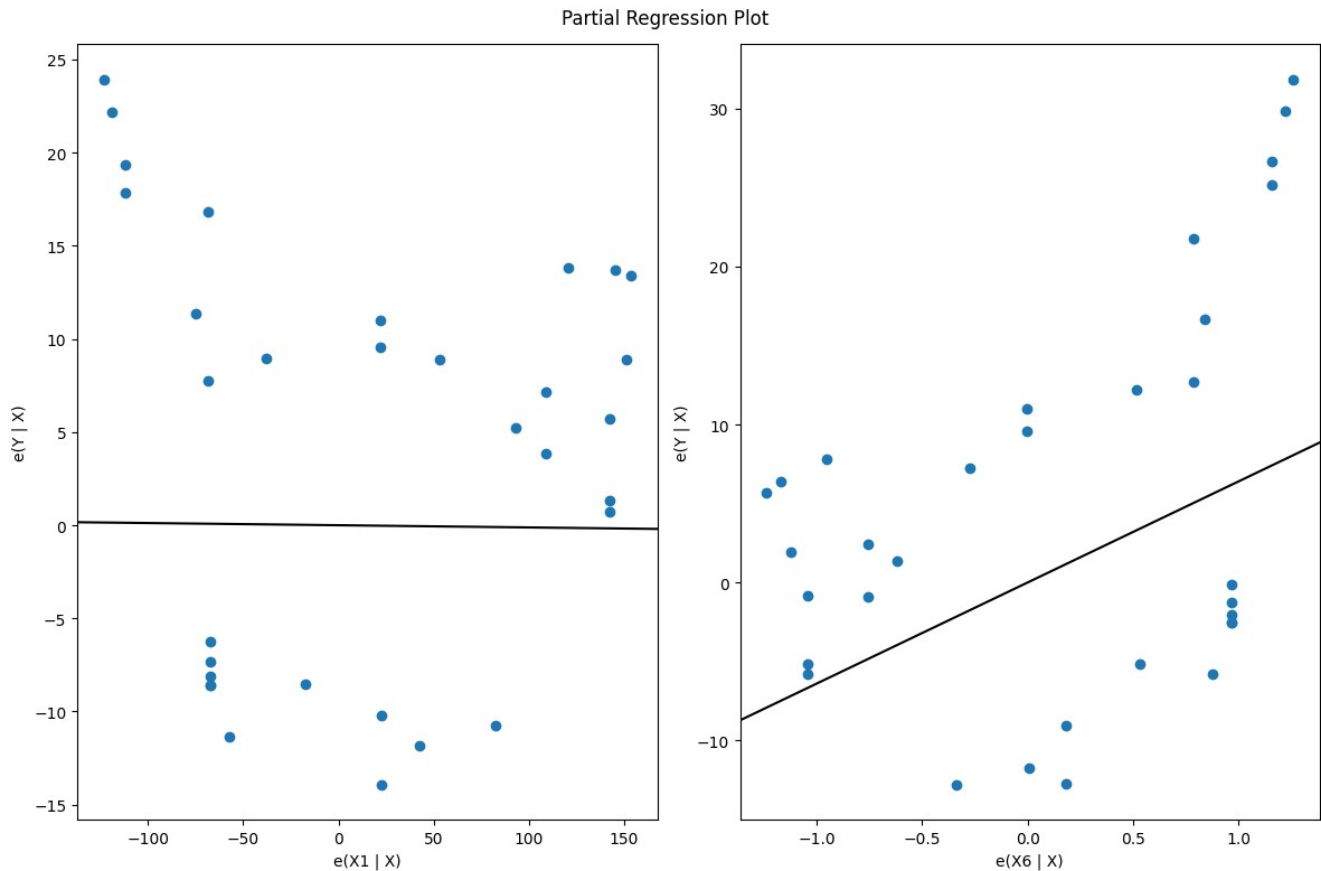
3 : The partial regression plots of x1 and x6 show that there is a stronger linear relationship between the target (mpg) and X6 (carburetor barrels) than with X1 (displacement). This makes sense since the mpg of a naturally aspirated engine should be influenced more by the efficiency of the carburetor than by the size of the engine cylinder volume for engine blocks that are in from our sample taken from the year 1975. (I have a little experience working on cars)

 4 : Studentized residuals: -6.841753166118225 -8.906064596114254 15.069346312470595 6.415348528556257 14.741412065026994 -14.865046516467956 9.760068456779134 9.21244353022032 23.539770302210957 18.831126601437226 -9.44930444611321 25.562004221937084 9.08656197704369 14.889257638736535 7.9166779058480605 5.6115056881414205 -11.650529363047877 -10.916266689763122 -8.036880836115923 4.223918211509071 11.205619579360997 10.054186147702453 -9.14040636361424 20.426158391092876 17.32151678761853 -12.686494192800847 11.66821561174285

7.747116929884731 1.652517595410586 0.9627282878515568 -12.344132240565262 -9.44930444611321

R-student residuals : -0.5403230165882502 -0.7033506722126488 1.1910673653435326 0.504695201180888 1.1797509975429303 -1.182891388762239 0.7963008560087801 0.7484819583469705 1.8733369792661712 1.5032601169437327 -0.7462526868506485 2.0302123688544587 0.7401348468018655 1.171483150105881 0.6410396111152387 0.4510977251650692 -0.9160868056804978 -0.8686658229052776 -0.6347074487918494 0.33753036319568835 0.9142399464376864 0.7869554530426701 -0.7276373330940092 1.630589072103735 1.4025805416762864 -1.007066643929348 0.9432437028573782 0.6190667196013825 0.1300034903105092 0.0757377942534198 -0.977174331955986 -0.7462526868506485

4 : Studentized residuals are useful for identifying outliers by standardizing residuals to have a more uniform scale.

R-student residuals take it a step further by recalculating the standard deviation of residuals excluding the observation in question, making them more sensitive to influential points.