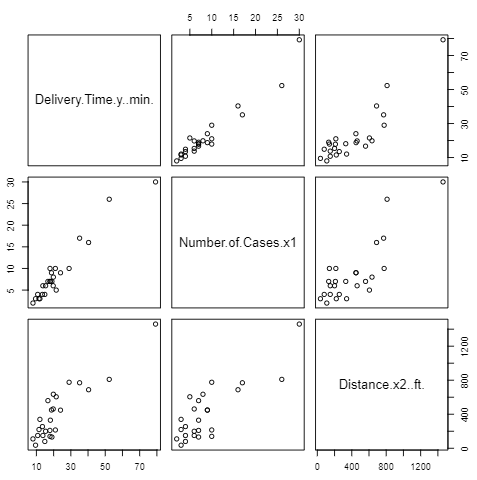
Grace Texana Long Torales

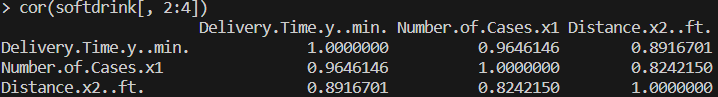
Gtl1500@rit.edu

Homework 7

* 1. While there does seem to be a linear relationship between delivery time (y) and number of cases and distance (regressors), there *also* seems to be a mild linear relationship between the number of cases and distance.



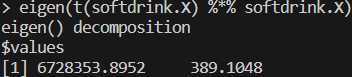
* 1. The correlation matrix corroborates our observations from the matrix plot of the predictors. There is a high correlation between y and the regressors. There is also a high correlation between number of cases and distance.



* 1. The VIFs are <5 and consequently do not suggest that significant multicollinearity is present.



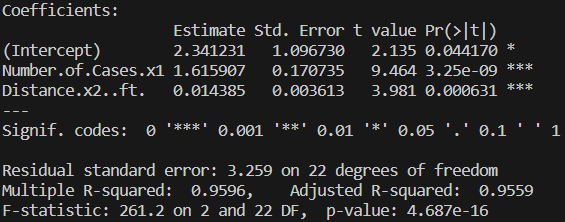
Furthermore, neither eigen value is small—which would have suggested multicollinearity.



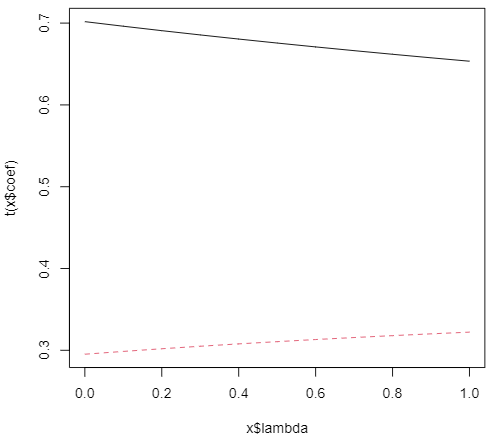
However, 100<κ<1000, suggesting moderate to strong multicollinearity.



On the other hand, R2 is large and the regression coefficients are both statistically significant. This does not indicate multicollinearity. If R2 was large while the regressors were not statistically significant, this would indicate multicollinearity.



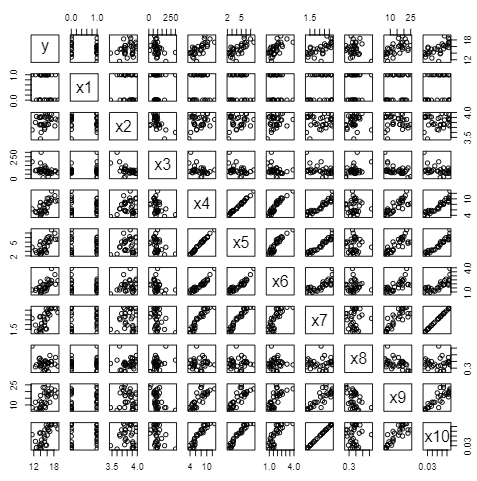
* 1. The multicollinearity diagnostics yielded mixed results, so we will perform ridge regression.



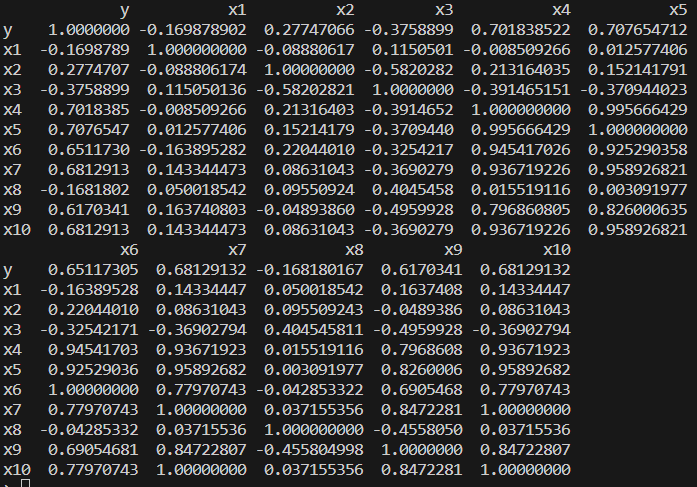
Using generalized cross validation, I found the appropriate value of k to be 0.36.



1. Multicollinearity is definitely hinted at in the matrix plot of the predictors, most notably between x4, x5, x6, x7, and x10, and also between x9 and x10.



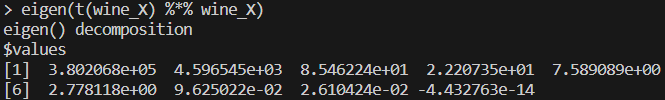
The correlation matrix corroborates my findings and in fact highlights more potential linear relationships than I had originally identified.



In fact, the correlation matrix shows that x7 and x10 are perfectly correlated, making them aliases of each other. I did not notice this until I received the following error while attempting to find the VIFs of the data. It makes sense that these two regressors are linearly related because they both represent a quality of anthocyanins. This error, in itself, is a sign of multicollinearity. After removing x10 from the model, then replacing it and removing x7 from the model, I still received this error both times, suggesting still more multicollinearity.



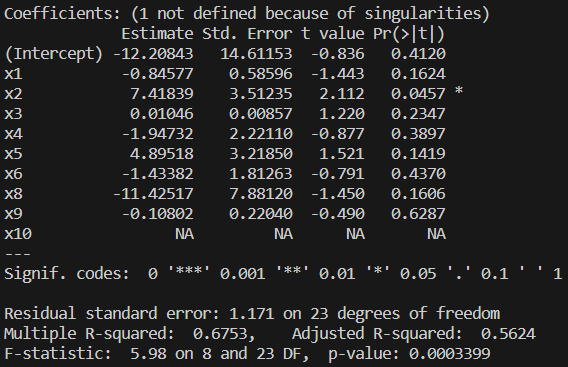
A few of the eigenvalues of X’X are small, which further indicates multicollinearity.



κ>1000 (much larger), suggesting strong multicollinearity.



Although R2<0.8, the F-test fails to reject the null hypothesis that the coefficients are simultaneously equal to 0. At the same time, almost none of the regression coefficients are statistically significant. This is another indicator of multicollinearity.



All in all, I would definitely say there is multicollinearity present in this dataset.