Project

**Executive Summary:**

For the current study, we are utilizing dataset which has information on several variables for all the school districts in Illinois for 2018-19 academic year. Students at multiple districts take schools statewide, which consequently provide the most consistent data for comparing student enrollment uniformly across all districts. We used the percentage of student enrollment to measure total student enrollment in the school and district as of October 1 of the preceding school year. The unit of analysis for this study was the districts in the state of Illinois for the 2018-19 school year that we were able to secure. Also, we determined the required number of samples or sample size for this current quantitative study by conducting various analysis. The results of those analysis computed for almost all samples of school districts in the state of Illinois. Purposive sampling is conducted to collect samples because sampling for proportionality was not the main concern. This data is collected via secondary sources. This dataset had the data on the district name, city, and student enrollment and various other variables on their databases and records.

**Research Questions:**

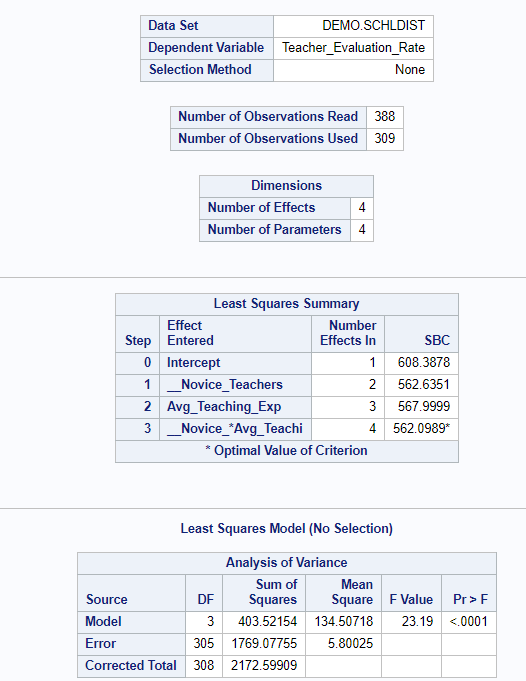
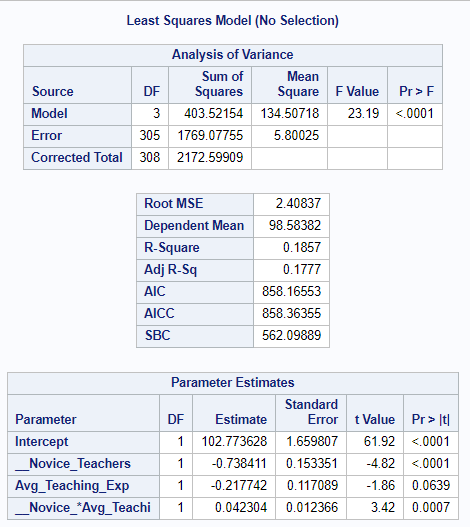
1. Using schldist dataset, conduct the appropriate test to determine

**a)** if Novice teachers and Average Teaching Experience has moderation effect on Teachers Evaluation Rate

**Solution:** Moderation Linear Regression (dependent variable is numerical in nature)

**H0:** Novice teachers and Average Teaching Experience has no moderation effect on Teachers Evaluation Rate

**H1:** Novice teachers and Average Teaching Experience has moderation effect on Teachers Evaluation Rate

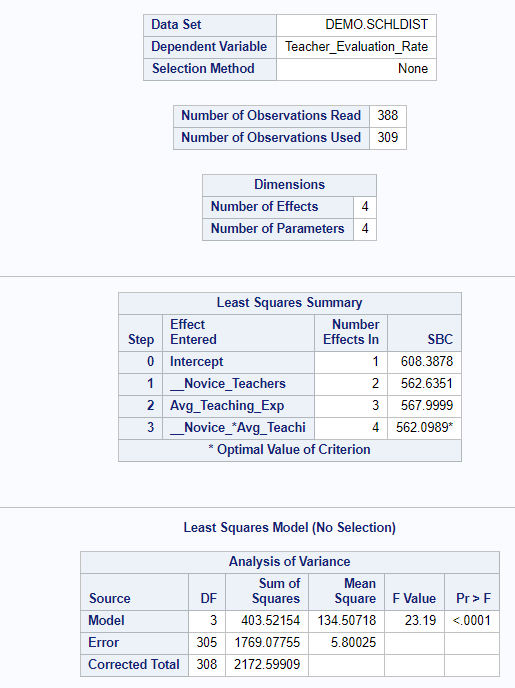
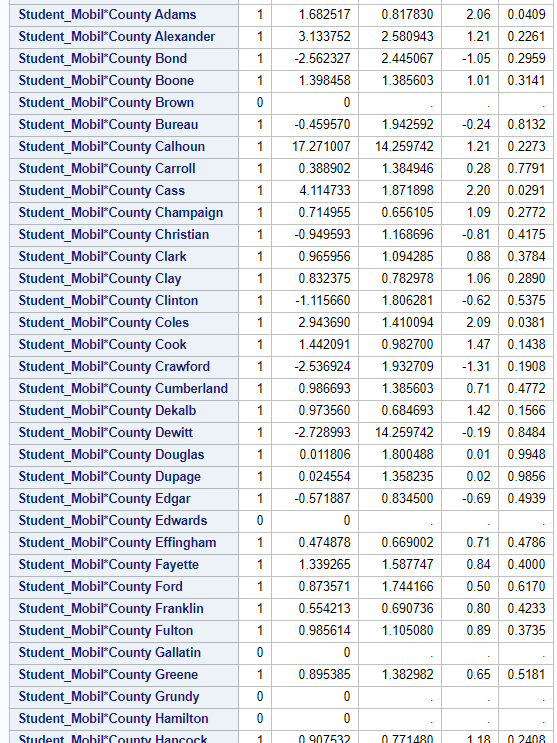
**Conclusion:** From the analysis, we can see that the p-value is less than nominal alpha and hence we can conclude that our **model is significant**. Looking at the parameter estimate, interaction term has p-value 0.0007 which is far less than nominal alpha. So, we can conclude that **there is moderation effect** between Novice teachers and Average teaching experience in impacting Teachers Evaluation rate.

**b)** Examine if County and Student mobility rate interact with each other to impact Average class size (var13)

**Solution:** Moderation Linear Regression(county is categorical in nature)

H0: County and Student mobility rate does not interact with each other to impact Average class size (var13).

H1: County and Student mobility rate interact with each other to impact Average class size (var13).

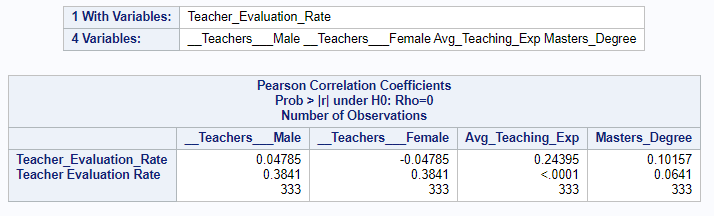
**Conclusion**: From the result, we can see that the **model is significant** as p-value is 0.0003 which is less than nominal alpha. When we look at the parameter estimates, as we have multiple categories in county variable considering the first couple interaction terms Student\_Mobil\*County Adams has a value of 0.0409 which is less than nominal alpha, hence, it has interaction effect on Average class size. Student\_Mobil\*County Alexander has a value of 0.2261 which is larger than nominal alpha concluding that there is no interaction effect on Average class size.

1. Use the schldist Data file for the following. In this question we would like to explore 4 variables (Teachers-Male, Teachers-Female, Master’s Degree and Average Teaching Experience) for their correlation with Teacher Evaluation Rate and Teacher Avg Salary.
2. We would like to explore the correlation between the Teacher Evaluation Rate and Teachers-Male, Teachers-Female, Master’s Degree and Average Teaching Experience. Conduct the three correlation tests and state your findings.

**Solution: Hypothesis:** Two tailed test (Correlation Analysis)

H0: ρ=0

H1: ρ≠0

 **Conclusion:**

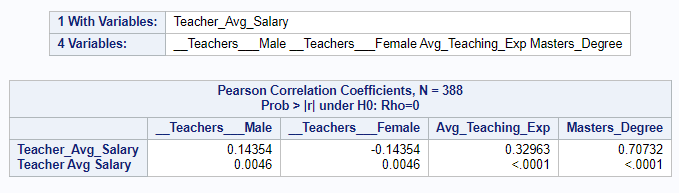
As the P value> α value, **we fail to reject the null hypothesis** and conclude that there is no correlation in case1, case2 and case 3 i.e. in all three correlation tests conducted. Therefore, there is no correlation of Teacher Evaluation rate with Teachers-Male, Teachers-Female, Master’s Degree but Teacher Evaluation rate is correlated with Average Teaching Experience.

1. Similarly explore if Teacher Avg Salary is correlated with Teachers-Male, Teachers-Female, Master’s Degree and Average Teaching Experience.

**Solution: Hypothesis:** Two tailed test (Correlation Analysis)

H0: ρ=0

H1: ρ≠0



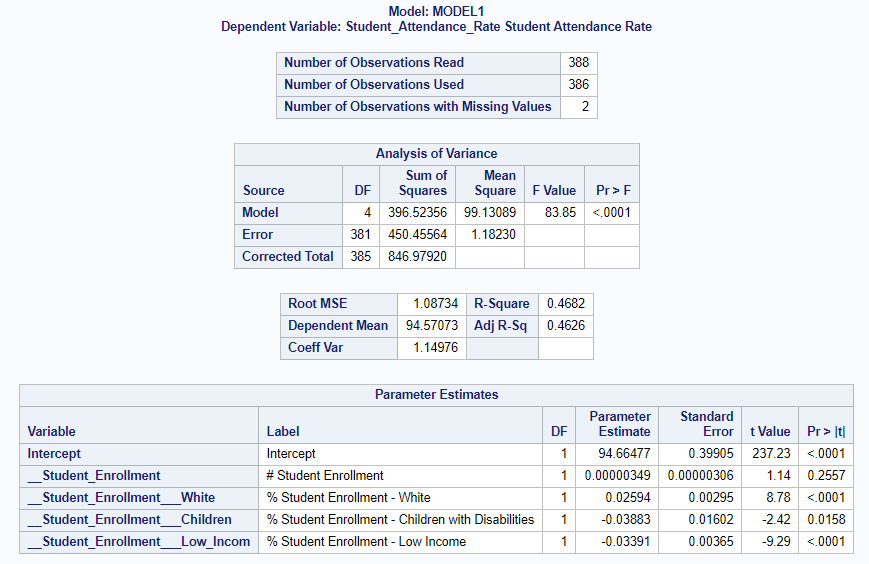
**Conclusion:** As the P value< α value, **we reject the null hypothesis** and conclude that there is correlation in case1, case2 and case 3 i.e. in all three correlation tests conducted. Therefore, there is a correlation of Teacher Average Salary with Teachers-Male, Teachers-Female, Master’s Degree and Average Teaching Experience.

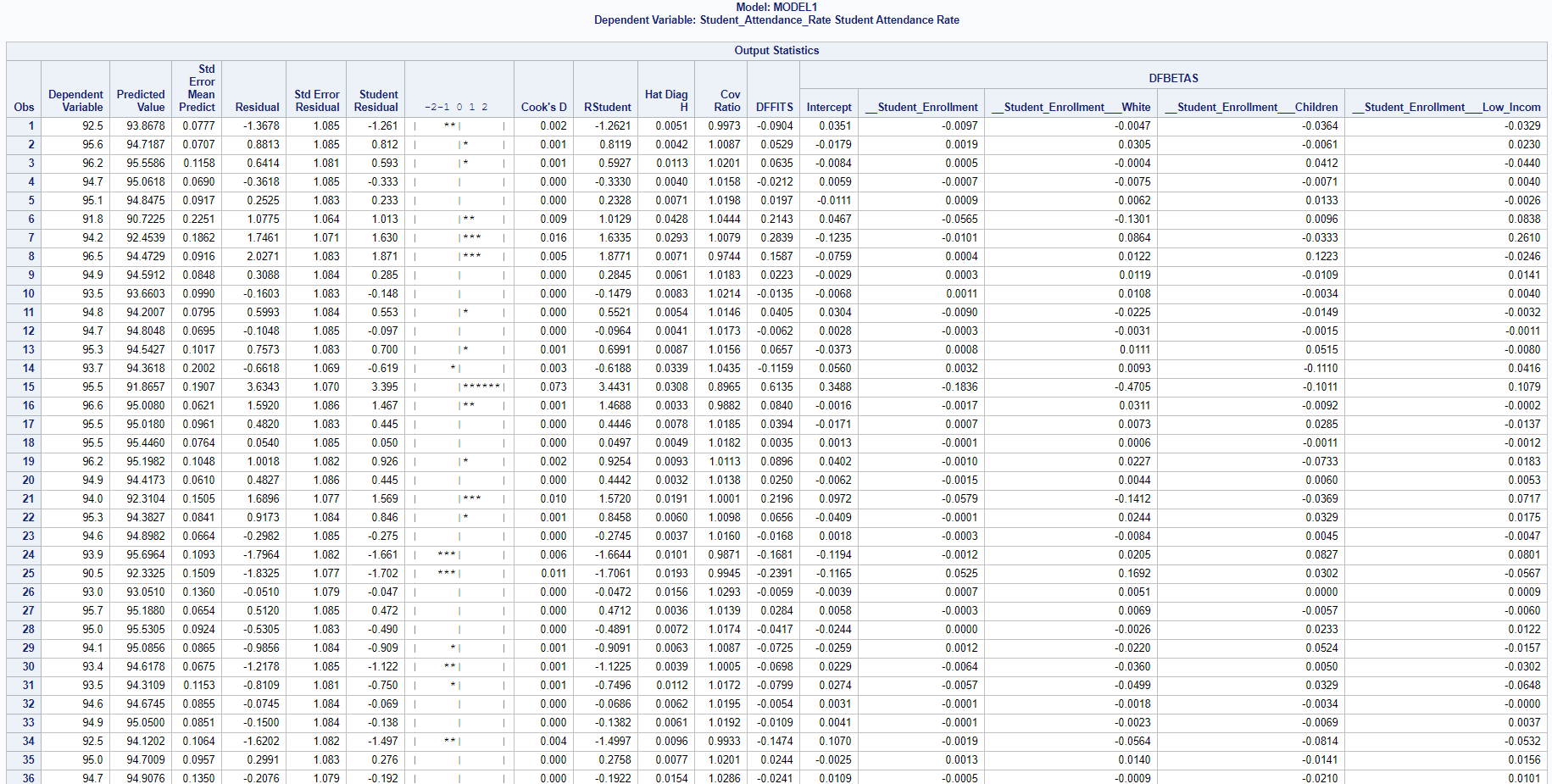
1. Using the “schldist” dataset posted on blackboard, explore if the following factors (Student Enrollment, Student Enrollment – White, Student Enrollment - Children with Disabilities, Student Enrollment - Low Income) impact Student Attendance Rate. What can you conclude about the relative importance of these factors in influencing Student Attendance Rate and Student Mobility Rate? Elaborate the results in detail?

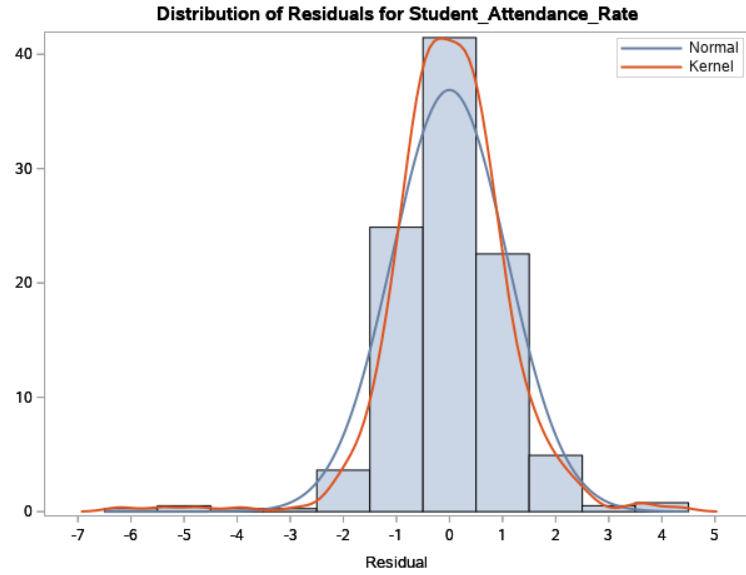
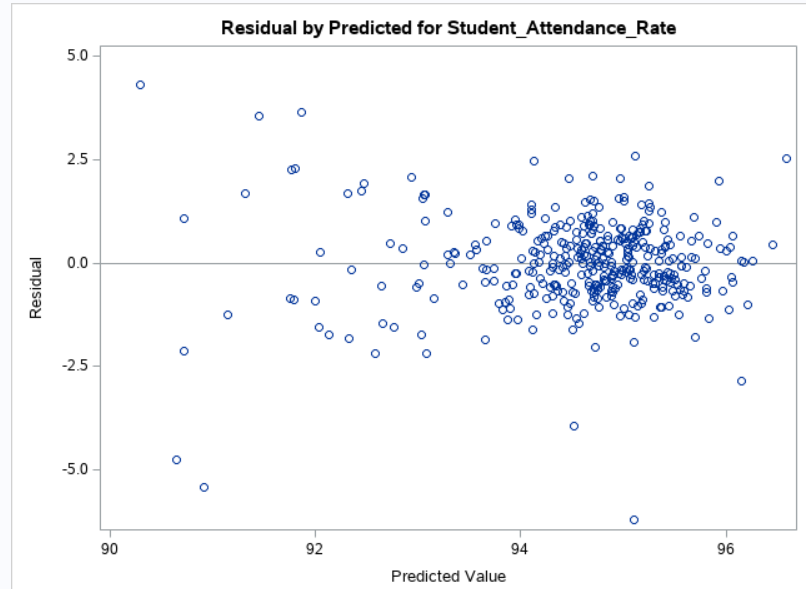
**Solution: Hypothesis:** Two tailed test (Multiple Regression test)

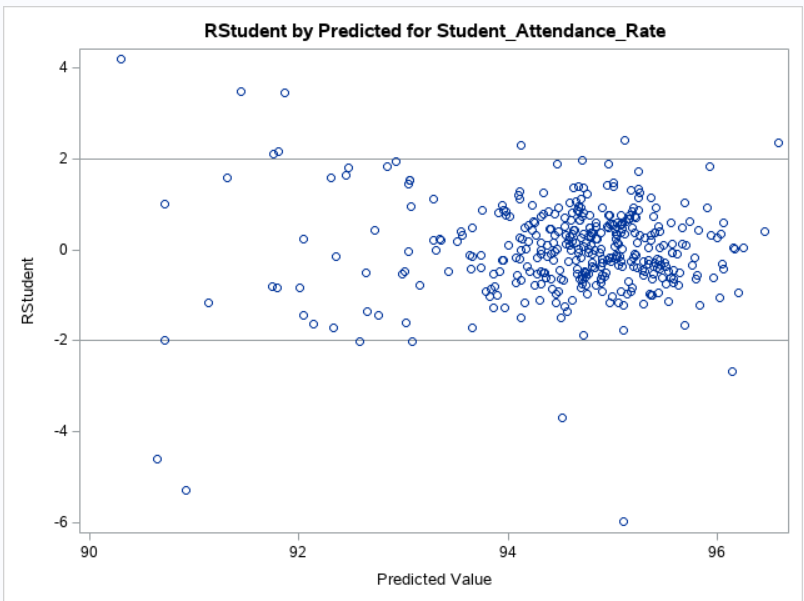
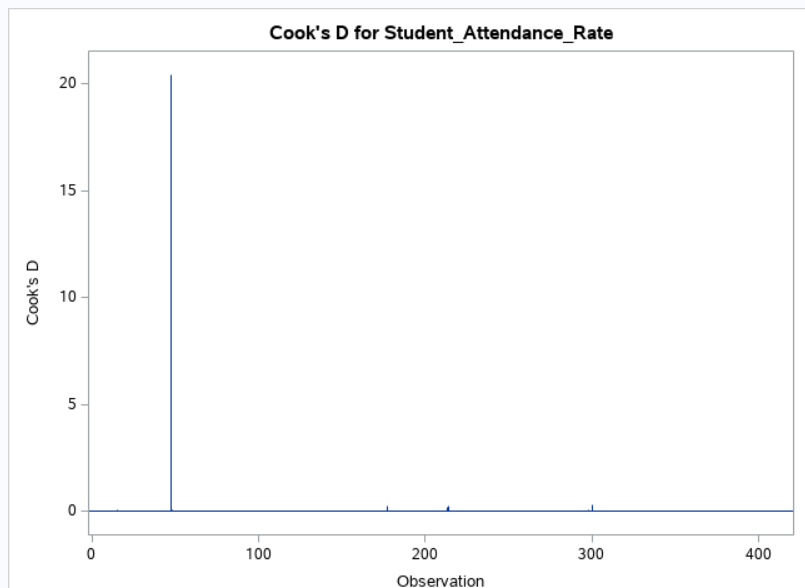
H0: ρ=0

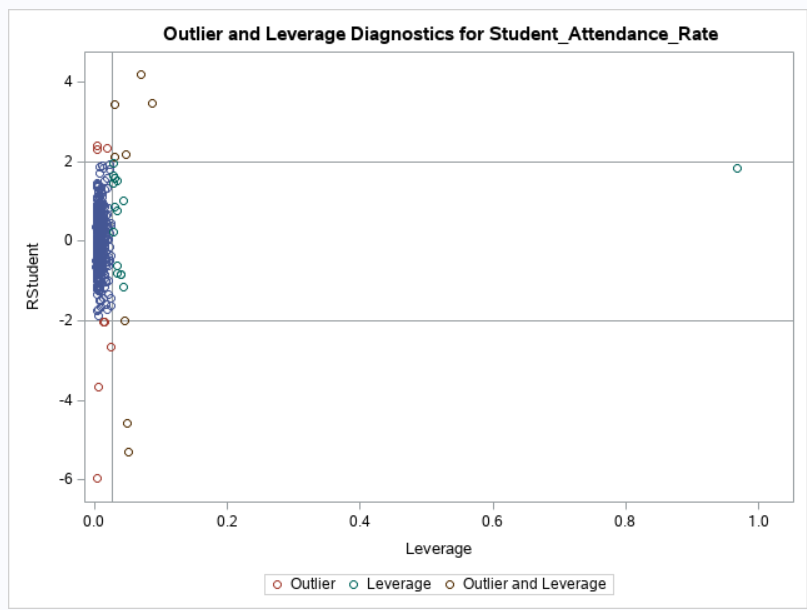
H1: ρ≠0







**Conclusion:** From the results of running a multiple regression model, the F-value is 83.85 and p-value for the F statistic is less than 0.0001. Also, the adjusted r-square value 0.4626 indicating that our model is significant. Hence, we **reject the null** and conclude that our model overall is a good. Also, we can make the following conclusions.

**Parameter Estimates:**

* For variables Student Enrollment-White, Student Enrollment - Children with Disabilities, Student Enrollment -Low Income has p-value less than .0001. Therefore, we can conclude that Student Enrollment-White, Student Enrollment - Children with Disabilities, Student Enrollment -Low Income have positive significant effect on Student Attendance Rate.
* For other variable Student Enrollment, look at the p-value associated with its T value and it is more than a nominal alpha of .05, we would conclude that it has no influence on our dependent variable, Student Attendance Rate.

**Distribution of Residual for Happiness:**

* From the distribution of residuals diagram, the errors do not follow normal distribution and has higher deviations.
* When we look at the residual by predicted for Student Attendance Rate, there does seem to be very strong megaphone effect and we can conclude that both assumptions of normality and equal variance are met by the model.

**Cook’s D Graph:**

* The Cook’s D graph indicates that we do have one observation that is highly different from other with Cook’s D value 0.0782 and observation number 85.

**Outliers & Leverages Graph:**

* Similarly, from outliers and leverage graph we can conclude that there are no observations that may qualify under influential observations label.

Overall, the result is surprising. I never thought that the Student Enrollment will have no influence on Student Attendance Rate. Also, to some extent Student Enrollment should have affected Student Attendance Rate, as it is also important factor.