SAS Project

By - Sashi Pranay Yadav Gobbilla

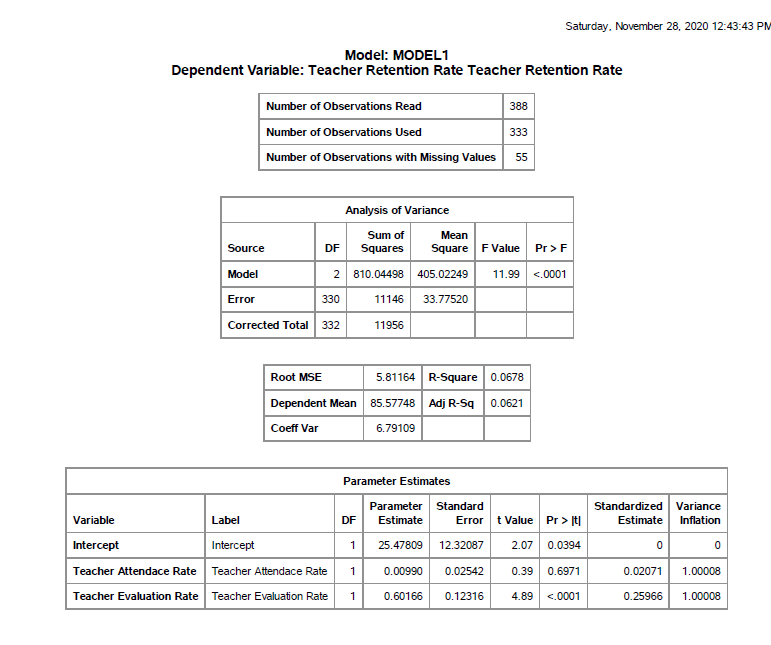
In the given dataset, after thorough analysis I have outlined the following research questions and investigated them through statistical analysis. They are, 1. Analyzing whether the teacher’s retention rate is impacted by teacher’s attendance and evaluation rate. I will use multiple regression analysis as I interpret that there is negligible interaction effect between the continuous variables. 2. To find whether Average Teaching Experience, % Novice Teachers & Degree (Masters) impacts teachers avg. salary. I intend to use moderation analysis as teacher’s salary can be cross product of the independent variables here. A teacher with both masters and bachelor’s degree can have higher salary compared to rest. For these analyses I have considered bachelor’s degree to be the requirement as master’s degree and bachelors are highly correlated to each other. Due to this reason I have chosen master’s degree in analyzing its impact on Teachers average salary rate. 3. To identify relation between student chronic rate and their absenteeism I opted to use simple linear regression. The data set has a total of 388 entities, out of which two of the models I used in analysis used 355 entities as the rest of the entities have missing values.

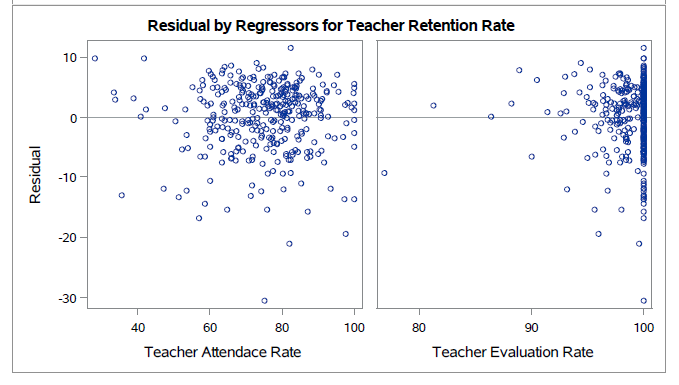
Research Analysis:

1.**Analyzing to conclude that the Teachers Attendance Rate & Teachers Evaluation Rate determines/interprets Teachers retention rate.**

I chose to use Multiple Regression model to analyze the above estimation as there are multiple variables impacting a single variable.

From the below result we can see that,





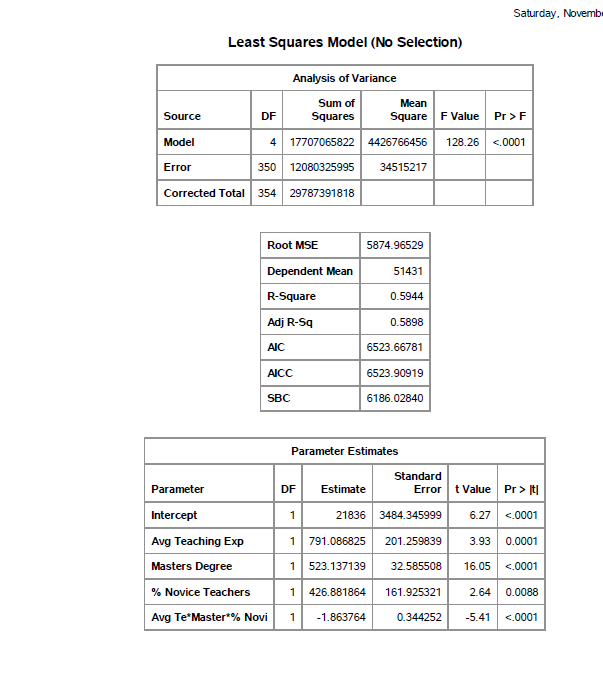
The p value for the resultant F-value of 11.99 for the model is <0.0001. Therefore, the overall model is significant fit for the following analysis as the P-value of F statistic is <0.0001 which is less than nominal alpha value of 0.05.

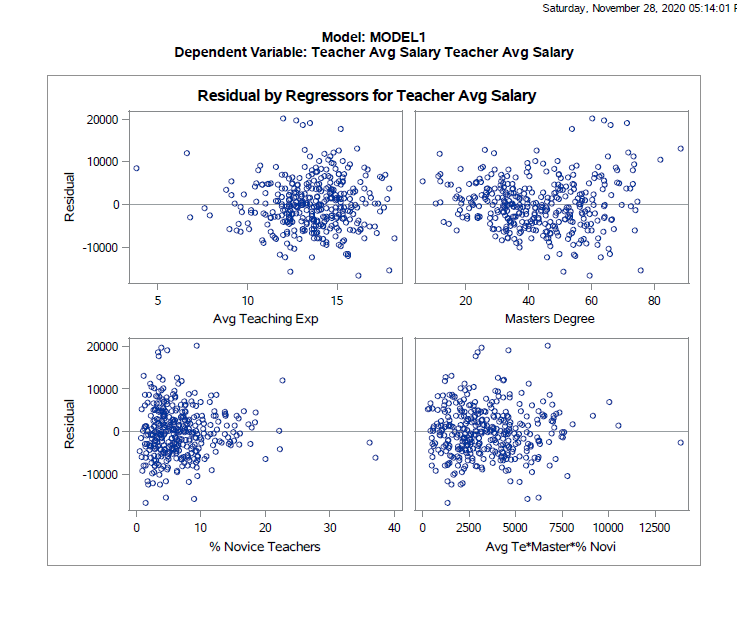
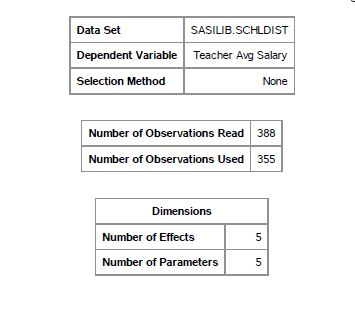
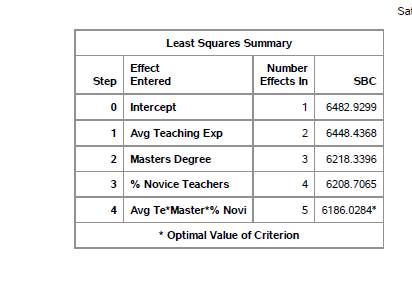
Here, the p value of teacher’s attendance rate is > α (0.05), so this does not impact the teacher’s retention rate. Whereas the p value of teacher’s evaluation rate is < α (0.05), so this variable impacts the teacher’s retention rate. The error rates are minimal as the kernels plot in distribution of residuals graph follows almost the normal distribution.

Hence, I conclude that the model is a good fit and **Teachers evaluation rate impacts the Teachers retention rate compared to the impact of Teachers attendance rate**.

2. **To show that the following variables,** **Average Teaching Experience, % Novice Teachers & Degree (Masters) interact with each other in determining/interpreting the Teachers Avg Salary.**

I intended to use multiple regression of Moderation Analysis to analyze the given statement as there are no categorical variables.





From the above results, p-value for the models statistical F-value of 128.26 is <0.0001. As it is less than α (0.05) we conclude that our model is a good fit.

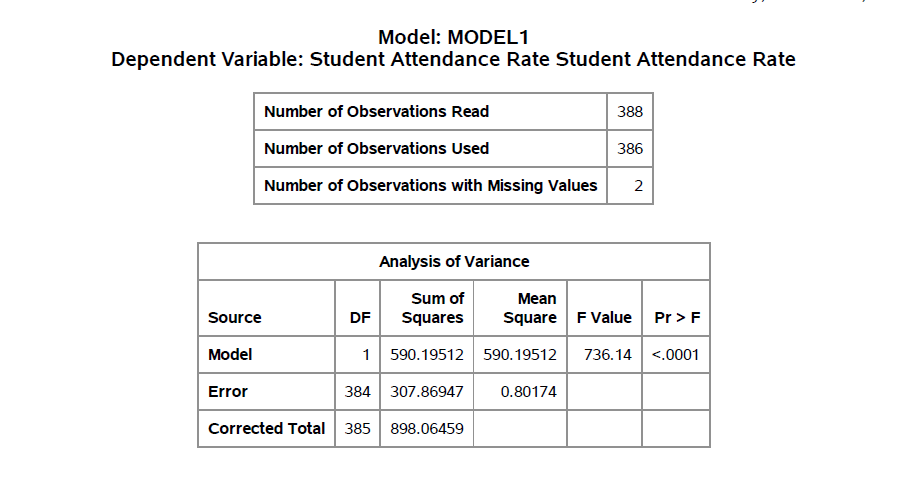
From the parameter estimates table, we see that the p-value of intercept and master’s degree is <0.0001 (< α), Avg. Teaching Exp has a p value of 0.0001 (< α), % novice teacher’ has a p-value of 0.0088(< α), and the cross product of all these variables has a p-value of < 0.0001(< α). So, all the variables impact the teacher’s salary.

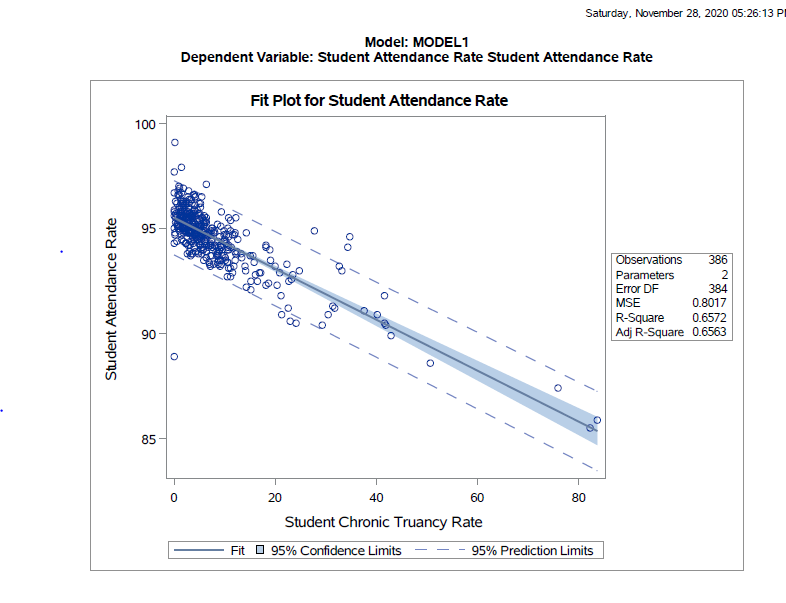
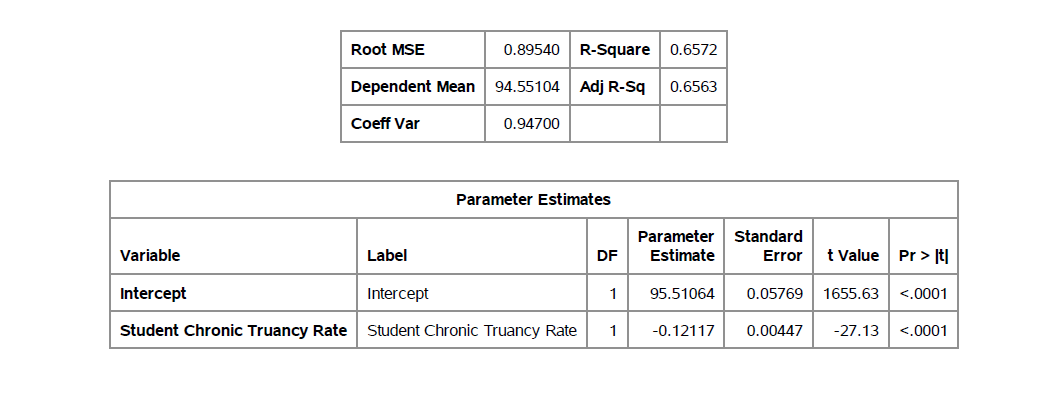
Hence, the following model is significant and conclude that **Average Teaching Experience, % Novice Teachers & Degree (Masters) interact with each other in determining/interpreting the Teachers Avg Salary.**

3.

**Claiming that in the School District dataset the Student Chronic Truancy rate determines/interprets the Student Attendance rate.**

Using Simple linear regression analysis, I am analyzing the level of relation between both the variables.





Here, P-value is < α for F-value of 736.14, intercept estimate is 95.51, parameter estimate for student chronic Truancy rate is -0.12. R-square value is 0.6572.

Analysis of variance shows that the P value (< 0.0001) is less than nominal alpha (0.05), hence we can conclude that the model is a good fit and it is analyzed that 65.72% variance in Student Attendance Rate is being determined by Student Chronic Truancy rate.

Here from the parameter estimates table, we can see that for every 1 unit increase in Student Chronic Truancy Rate the Student Attendance Rate is expected to decrease by 0.12117.

Hence, I conclude that the model is a good fit for the analysis and conclude that **Student Chronic Truancy rate impacts the Student Attendance rate.**