Project#1

Regression

02/19/2020

This is a group project, and students should work in a group of size 3. Include all the R code, hypothesis testing, one or two lines of explanation for any output. The report should be organized, printed,and stapled. The due data of this project is **Wednesday** 02/19/2020.

```
Movies = read.csv("C3 2008Movies.csv")
```

The 2008Movies file contains data on movies released in 2008.

1. Calculate a regression modelto predict box office from run time. Interpret the R^2 value and test statistic for the slope in the context of this problem.
2. Create indicator variables for the genre and MPAA ratting.Use the best subsets regression to determine a appropriate regression model
    a. Validate the model assumptions.
    b. Look at residual plots and check for heteroskedasticity (unequal variance), multicollinearity, correlation of errors, and outliers. Transform the data if it is appropriate.
    c. submit your suggested least squares regression formula along with a limited number of appropriate graphs that provide justification for your model. Describe why you believe this model is the best.
    d. Test the overall model adequcy.
3. Conduct an extra sum of squares test to determine if one or more interaction terms (or quadratic terms) should be included in the model. You can choose any other terms to test.
4. Test whether average run time is the same for different Genre. Clearly show your hypothesis test.
5. Check equality of variance of run time for Genre type.