**Homework on Basic Statistics and Regression Analysis**

Marketing 3597: Marketing Analytics

Fall 2017

**NOTE: This homework is due on Saturday, October 14, before class. Please type the answers and bring a hard copy, no need to print out all the R outputs.**

In the lecture, we practiced conducting marketing mix modeling using a regression model. In the majority of the exercise in class, we have the Y variable to be the sales in that time period. Towards the end, we realized that if we use the changes in sales as a Y variable, the model is quite improved based on the adjusted R square value. Please try estimating a few regression models with the sales change as Y variable, and with different specifications of X variables.

In other words, the models you are going to estimate are based on the basic model

Here the variable is generated using R in class. The other variables are from data, except the dummy variable is generated in class as well. Try estimating the following alternative models (1-3), and show the results.

**Please do not just copy the R output, but put the results in tables or a nice format, ready to present to your managers. Think about what are the important statistics you want to show for a regression model.**

1. Lagged X variable

Similar to what we did in class, instead of using all X variables in the current time period, use all X variables in the past time period. Compare the results from this model with those from the basic model. Which one do you like better? Why?

1. Lagged Y variable

Similar to what we did in class, add the lagged Y variable in the model. Compare the results from this model with those from the basic model. Which one do you like better? Why?

1. Dummy variables
2. Create one dummy variable for the three months in the summer, from June to August.
3. Create three dummy variables for the three months from June to August, one for each month.

Compare the results of the above two models in their dummy variable estimates.

1. Interaction variables

In class, we demonstrated three different model specifications with different interactions. Please estimate two models, each with a different set of interactions. Feel free to use those we tried in class, but for a different Y variable.

1. Nonlinear transformation of the X variables
2. Add the square terms of the X variables, together with the linear form. How do you like the estimated results?
3. Use ln-transformation for all the X variables, and include only the ln(X+1) into the model, without the linear term, and how do you like the results?
4. Do these nonlinear transformation make sense in this context? Do they provide better fit comparing to the basic model?
5. No need to estimate, comment on why we do not include lagY variable in this model, with as the Y variable in this regression.