**Homework 2 - Group assignment**

Using the store level scanner data provided to you in class (filename will be similar to spagsauc\_groc\_1114\_1165). You need to link UPC from this file with information in the prod\_sauce.xls file to get brand names etc.

Prepare a report that summarizes the data. At a minimum, it should answer the following questions.

Summarize data: No need for statistical testing for these questions.

1. What are the top 6 brands in the category in terms of dollar sales? What are the market shares of the 6 brands (assuming there are only 6 brands in the market).
2. Which companies are the major players in the category? Which company owns which brands?
3. Create a 7th brand called “Other” that has all other brands that are not in the top 6.
4. Find average prices, display, features of each of the 7 brands.
5. What are the top 5 regions in terms of dollar sales?
6. What are the top 10 store chains that sell a lot of your category in terms of dollar sales?
7. What is the average price per unit of 7 brands by week? Plot the average price by week (I wish to see a line plot of price by week). Comment on your findings.
8. Assume you are manager of a brand (out of the top 6). Write a short paragraph stating what you learned from this descriptive analysis (steps 1-7).

Statistical Analysis

1. Do large stores (top 3 stores) have higher average price per unit than small stores (stores ranked 8-10) for brand 1 (the top brand in Q1). Test and report your results and comments.
2. Develop three additional hypotheses linking useful variables to dollar sales, test them and report your findings.
3. For the top brand: run a regression model with weekly dollar sales as dependent variable. Use average weekly price per unit, average display, average feature, and other useful variables in your regression and answer the following questions:
4. What is the R-sq and adj R-sq of the model?
5. Which coefficients are significant?
6. Which variables are most important in explaining sales?
7. Interpret the meaning of the price coefficient? What is the price elasticity?
8. Interpret the meaning of the display coefficient?
9. Test whether there is an interaction between display, feature and price. Comment on your findings.
10. Test whether the effect of price is non-linear. Comment on your findings.
11. Test using VIF and COLLIN whether there is multicollinearity in the model? Comment on your findings.
12. Test for presence of heteroscedasticity using White test. Do A WLS if needed. Comment on your findings.