**Homework on Basic Statistics and Regression Analysis**

Marketing 3597: Marketing Analytics

Fall 2017

**NOTE: This homework is due on Saturday, October 7, before class. Please type the answers and bring a hard copy, no need to print out all the Excel outputs.**

**To Get Ready:**

1. Follow the instructions and get started with R
2. The main R function to use for this homework is lm(). Almost all the functions you would need are included in the handout. If not enough, feel free to seek help using Google search.

**Regression Analysis**

Use data “Coffee\_inClass.csv”, which tracks the weekly sales, promotional spending and price of a particular brand of coffee.

1. Estimate the model , then report the estimation results in the following table.

|  |  |  |
| --- | --- | --- |
| Parameter Name | Coefficient | Standard Error |
| Intercept |  |  |
| Price ($) |  |  |
| Feature (promotion) |  |  |

1. What’s the meaning of the estimated values of each coefficient?
2. Why do we want to report the standard error of each coefficient here?
3. How to calculate the t-stat values for each parameter estimates? Based on the t-stat values, what conclusions do you make regarding the variables to be included in the model?
4. The estimates are obtained using the data provided, where the price level is recorded in dollar. If the price data is recorded in cents instead, when running the same regression model using such data, what would be the value for ? How about its standard error?
5. What other control variables you can think of that can be added to this model, besides those in the data?
6. R square represents the fit of the model to the data.
7. What is the value of R square in this model? What’s the meaning of this number?
8. How is R-square calculated? Can you verify it using the results from your regression estimates? If you named the results from lm() “lm1”, you can access the predicted values of the Y’s using lm1$fitted.values
9. Why do we need the F value in this regression results? How is the F value calculated – using words to explain.
10. Understand the omitted variable bias:
11. What’s the correlation between the two variables: price and promotion (feature)?
12. Drop the variable promotion, and rerun the estimation with Y variable: sales, X variables: price. Compare the **parameter of the price variable** in this estimation vs. that from question 1. Comment on the differences between these two values in these two regression models. What’s the ratio between these two estimation results?