**Homework 1. Review of Basic Statistics**

Marketing 3597: Marketing Analytics

Fall 2017

**NOTE: This homework is due on Saturday, September 30, 2017, before class. Please *type the answers* and bring a hard copy, no need to print out all the code, unless you want to.**

**To Get Ready:**

1. If you have never used R programming, please following the instructions in the file **R Starter and Statistics Review.html** to get yourself familiar with the R environment, after installing R and R studio following the instructions.
2. Before you start on the homework, you are suggested to review the basic statistics concept mentioned in the html file posted on Google drive.

**Basic Statistics**

1. If you have a series of numbers, such as X = (2.3, 4.2, 3, 5.6, 7), calculate the mean, sample variance and population variance among these numbers.
2. In the above calculations, why the population variance and the sample variance have different values. When would these two values be very close to each other?
3. If you have another series of numbers Y = (3.4, 5.3, 4, 4, 7.4), calculate the covariance and correlation between these two series (X, Y). Please use two ways to do these calculations (1) use the functions provided by R; (2) by hand.
4. In the above calculations of the correlation, what’s the sign of the correlation? If you run a regression using Y values as the dependent variable, and X as the independent variable, the coefficient of the slope will be positive or negative? Think about this question before you run a regression. Then use Excel to run a regression and see whether your answer was right. Comment on what you find.

*R code help:*

*To estimate a regression of Y on X, using R, you can use the function* ***lm()***

X = c(2.3, 4.2, 3, 5.6, 7)

Y = c(3.4, 5.3, 4, 4, 7.4)

lmresults=lm(Y~X)

summary(lmresults)

*If for some reason you could not get this running on your computer using R, it is ok, input these numbers into Excel, and run the regression using Excel. You may find* [*this link*](http://www.wikihow.com/Run-Regression-Analysis-in-Microsoft-Excel) *useful for instructions on how to run regression using Excel.*

1. What are the two parameters used to describe Normal distribution? What’s the meaning of each parameter?
2. What’s the relationship between the PDF and CDF for the same distribution, evaluated at the same value of the random variable?
3. What are the features of a Normal distribution?