**Homework 1**

**Statistical Genomics**

**CROPS 545, Spring 2017**

Professor: Zhiwu Zhang

Due on February 1, 2017, Wednesday, 3:10PM, PST

**Objectives**: 1) random variables; 2) derivation of random variables; 3) distribution of random variables; 4) define R functions; 5) t test and 6) Chi-square test.

**Hand in:** Email your report (PDF, limited to five page) and R source code (text file) with subject of “CROPS545 HW1” to [Zhiwu.Zhang@WSU.edu](mailto:Zhiwu.Zhang@WSU.edu). Name your files as following:

Homework1\_ firstname\_lastname.pdf and Homework1\_ firstname\_lastname.R

**Grade components**: 1)Hypothesis or statement; 2) Results; 3) Methods; 4 presentation; 5) R source code (clarity, simplicity and documenting comments)

1. Start from random variables with standard normal distribution (or any other distribution, such as binomial, Chi square, and t), define your own random variable that is function of the random variables. Name your new random variable as your last name and develop a R function to generate the new random variable. The input of your R function should include n, which is number variables to be generated, and parameters for the distribution of the random variable you defined. Note: try not to be the same as the known distributions such as Chi-square, F and t.
2. Sample ten thousand observations from the distribution you defined. Graph their properties and describe the potential application of your distribution in nature.
3. There is an R function, rf(n, df1, df2), to generate n random variables that follow F distribution with degree freedom of df1 and df2. Develop your own R function to do the same thing. Your function should start with normal distribution by using rnorm function in R, followed by derivation of Chi-square distribution variables.
4. Sample 10, 100, 1000 and 100,000 F distributed variables by using the R function you developed. Calculate means of these samples and test them on the null hypothesis that the samples have expected mean of df2/(df2-2)
5. Sample 10, 100, 1000 and 100,000 F distributed variables by using the R function you developed. Calculate variance of these samples and test them on the null hypothesis that the samples have expected variance of

