**Homework 3**

**Statistical Genomics**

**CROPS 545, Spring 2017**

Professor: Zhiwu Zhang

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

**Data files**: mdp\_numeric.txt from GAPIT demo data. The data file can be download from <http://www.zzlab.net/GAPIT/GAPIT_Tutorial_Data.zip>. The data contains 281 individuals (row wise) and 3093 SNPs (column wise) coded as 0/1/2. The SNP ID, chromosome and position is indicated by a file named mdp\_SNP\_information.txt

**Hand in:** Email your report (PDF, limited to five page) and R source code (text file) with subject of “CROPS545 HW3” to [Zhiwu.Zhang@WSU.edu](mailto:Zhiwu.Zhang@WSU.edu). Name your files as following:

Homework3\_ firstname\_lastname.pdf and Homework3\_ firstname\_lastname.R

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

**Objectives**: 1) Simulate phenotype from genotype; 2) GWAS by correlation; 3) Evaluate true and false positives.

1. Sample 10 SNPs as QTNs out of the 3093 SNPs. Simulate QTN effects from a standard normal distribution. Assign genetic effects for each of the 281 individuals. Simulate normal distributed residual effects with appropriate variance to have a heritability of 0.75. Add residual effects to genetic effect to create phenotypes. You can either use the G2P R function or code everything by yourself. Describe the distribution of genetic effect, residual effects and phenotypes and explore the relationship among them (20 points).
2. Perform GWAS by using the correlation method. You can either use the GWASbyCor R function or code everything by yourself. Create Manhattan plot and label the positions of the QTNs (20 points).
3. Find number of QTNs among top ten associated SNPs (20 points).
4. Count number of SNPs with P values smaller than the P value of the seventh significant QTN (20 points).
5. Redo (3-4) for 100 replicates. Report the averages and standard deviations (20 points).

**Extra credit**

1. Simulated phenotypes from genotypes so that the phenotypes skewed normal distribution due to genetic effect with a long tail on the right (25 points, report is limited to one extra page).