**Homework 4**

**Statistical Genomics**

**CROPS 545, Spring 2016**

Professor: Zhiwu Zhang

Due on March 23, 2016, Wednesday, 3:10PM, PST

**Data files**: The following files can be download at <http://zzlab.net/GAPIT/data/>.

1. mdp\_numeric.txt from GAPIT demo data. The data file contains 281 individuals (row wise) and 3093 SNPs (column wise) coded as 0/1/2.
2. mdp\_SNP\_information.txt. The file contains SNP ID, chromosome and position.
3. CROP545\_Phenotype.txt. The file contains taxa name and phenotype.
4. CROP545\_Covariates.txt. The file contains taxa name and two covariates.

**Hand in:** Each team (maximum of three people) email your report (PDF, limited to five page), R source code (text file), and user manual/tutorial (PDF, no page limitation) with email subject of “CROPS545 HW4” to [Zhiwu.Zhang@WSU.edu](mailto:Zhiwu.Zhang@WSU.edu). Name your files as following:

Homework4\_ PackageName.pdf and Homework4\_ PackageName.R

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

**Objectives**: Develop your own R package to perform GLM GWAS.

1. The package should contain at least three input: y, X , and C that are R objects of numeric data frame. Their dimensions are n by 1, n by m, and n by t corresponding to phenotype, genotype and covariate data, where n is number of individuals, m is number of markers, and t is number of covariates. The function should return probability values with dimension of 1 by m for the association tests between phenotype and markers. Markers are tested one at a time with covariates in C included as covariates (15 points).
2. The package should also provide additional co-factors to improve performance of GWAS, such PCA. Name your package with an acronym that describes the features of your co-factors. Your package should also automatically eliminate your own co-factors if they are in linear dependent to the covariates provided by users (25 points).
3. Develop a user manual and tutorials (20 points).
4. Perform GWAS on the genotypes, phenotypes and covariates provided (15 points).
5. Demonstrate that your method is superior to the competing method (GWASbyCor) through simulation (25 points).