**Homework 5**

**Statistical Genomics (545)**

**Spring 2021**

Professor: Zhiwu Zhang

TA: Zhou Tang

Due on Monday, April 12, 2021, 3:10PM PST

**Data files**: You can use either the same dataset you used in Homework 2, or switch to a different dataset with the same requirements.

**Hand in:** Email your R RMD and Knit (HTML, PDF, or Word) with subject of “StaGen545 HW5” to [Zhiwu.Zhang@WSU.edu](mailto:Zhiwu.Zhang@WSU.edu). Name your files as

“Homework5\_ firstname\_lastname”.

**Objectives**: 1) Examine the GLM package from previous homework; 2) Statistical power vs. FDR and type I error; 3) mapping resolution; and 4) GWAS methods (GLM, MLM, CMLM, SUPER, FarmCPU, and BLINK).

1. Sample a number of QTNs of your choice from the genetic markers used in Homework 2 and simulate QTN effects from a standard normal distribution. Assign genetic effects for each individual. Simulate normally distributed residual effects with appropriate variance to have a heritability of 0.75. Add residual effects to the genetic effects to create phenotypes. Use the GLM GWAS package you developed in Homework 4 to perform association analyses with three PCs included as covariates. Count the number of false positives for identifying half of your QTNs (20 points).
2. Repeat the simulation in (1) 30 times and examine statistical power vs. FDR at mapping resolution of 100,000 base pairs (20 points).
3. Repeat the simulation in (2) with two additional levels of heritability (0.25, and 0.5) (20 points).
4. Repeat the simulation in (2) with two additional methods of your choice from MLM, CMLM, SUPER, FarmCPU, and BLINK (20 points).
5. Redo one of the above questions (2-4) by excluding the QTNs from the marker dataset prior to GWAS. Describe the impact of inclusion and exclusion of the QTNs (20 points).

**Extra credit**

1. Find another method and demonstrate that it has higher statistical power than BLINK (50 points).