Name:

Graded by: TA

**TOTAL POINTS: /100**

Problem 1: Sample 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).

* Order all the QTNs by their P values (4 pts)

Full Points

* Find the median of the P values as threshold (4 pts)

Full Points

* Count the number of SNPs with P values smaller than the threshold (4 pts)

Full Points

* Commenting (4 pts)
* Full Points
* Bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 2: Repeat the simulation in (1) 30 times and examine statistical power vs. FDR at mapping resolution of 100,000 base pairs (20 points).

* Set resolution correctly (3 pts)

Full Points

* Calculate power average correctly (2 pts)

Full Points

* Calculate FDR average correctly (2 pts)

Full Points

* Conduct the replicates correctly (3 pts)

Full Points

* Plot ROC curve (2 pts)

Full Points

* Commenting (4 pts)
* Full Points
* Bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 3: Repeat the simulation in (2) with two additional levels of heritability (0.25, and 0.5) (20 points).

* Conduct the simulation with different heritability (3 pts)

Full Points

* Calculate power and FDR average correctly (3 pts)

Full Points

* Conduct the replicates correctly (3 pts)

Full Points

* Plot ROC curve (3 pts)

Full Points

* Commenting (4 pts)
* Full Points
* Bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 4: Repeat the simulation in (2) with two additional methods of your choice from MLM, CMLM, SUPER, FarmCPU, and BLINK (20 points).

* Conduct the simulation with different methods (3 pts)

Full Points

* Calculate power and FDR average correctly (3 pts)

Full Points

* Conduct the replicates correctly (3 pts)

Full Points

* Plot ROC curve (3 pts)

Full Points

* Commenting (4 pts)
* Full Points
* Bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 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).

* Exclude QTNs from marker dataset before GWAS (4 pts)

Full Points

* Plot ROC curve (4 pts)

Full Points

* Explain the difference between inclusion and exclusion of QTNs (4 pts)

Full Points

* Commenting (4 pts)
* Full Points
* Bug-free (4 pts)

Full Points

Comments:

POINTS: 20

**Extra Credit**

Problem 6: Find another method and demonstrate that it has higher statistical power than BLINK (50 points).

You get either no credit or full credit if all of the following are satisfied:

1. The method you choose has higher statistical power than BLINK.
2. The conclusion is based on at least 30 replicates
3. At least three datasets are used for comparison.

Comments:

POINTS: 50