Name:

Graded by: TA

**TOTAL POINTS: /100**

Problem 1: Use the GAPIT.Phenotype.Simulation function to simulate phenotypes with heritability of 75% controlled by 10 QTNs having effects with standard normal distribution. Display the QTN effects. Create 3D plots on scatter plots of total genetic effects (breeding values), residual effects, and phenotypes of individuals (5 points).

* Display the QTN effects (1 pt)

Full Points

* 3D plot (2 pts)

Full Points

* Commenting (1 pt)
* Full Points
* Bug-free (1 pt)

Full Points

Comments:

POINTS: 5

Problem 2: Conduct GWAS with BLINK on the entire population using phenotypes dependent variable. Identify SNPs that pass 1% threshold (after Bonferroni multiple test correction). Display their estimated effects against their true effects. (5 points).

* GWAS with BLINK (1 pts)

Full Points

* Correct threshold (1 pt)

Full Points

* Display effects (1 pt)

Full Points

* Commenting (1 pt)
* Full Points
* Bug-free (1 pt)

Full Points

Comments:

POINTS: 5

Problem 3: Randomly sample 80% of individuals from the entire population as the training population and the rest as the testing population. Conduct GWAS with BLINK on the training population using phenotype as the dependent variable. Identify SNPs that pass the 1% threshold (after Bonferroni multiple test correction). Plot the estimated effects of the associated SNPs against their true effects (15 points).

* Population division (6 pts)

Full Points

* GWAS on training (2 pts)

Full Points

* Display effects (1 pt)

Full Points

* Commenting (3 pts)
* Full Points
* Bug-free (3 pts)

Full Points

Comments:

POINTS: 15

Problem 4: Use the effects estimated in (3) for SNPs that passed the 1% threshold (after Bonferroni multiple test correction) to predict the phenotypes of the individuals in the training population and testing population for Marker-Assisted Selection (MAS). Calculate the correlation between the observed and predicted phenotypes for training population and testing population separately. Explain the difference between the training and testing populations (15 points).

* Prediction for training population (2 pts)

Full Points

* Prediction for testing population (2 pts)

Full Points

* Correlation in training population (2 pts)

Full Points

* Correlation in testing population (2 pts)

Full Points

* Explanation (5 pts)

Full Points

* Commenting (1 pt)
* Full Points
* Bug-free (1 pt)

Full Points

Comments:

POINTS: 15

Problem 5: Conduct ridge regression on the training population to estimate all marker effects. Use the estimated marker effects to predict the phenotypes of the individuals in the training population and testing population. Calculate the correlation between the observed and predicted phenotypes for training population and testing population separately. Compare ridge regression with MAS (15 points).

* Prediction for training population (2 pts)

Full Points

* Prediction for testing population (2 pts)

Full Points

* Correlation in training population (2 pts)

Full Points

* Correlation in testing population (2 pts)

Full Points

* Explanation (5 pts)

Full Points

* Commenting (1 pts)
* Full Points
* Bug-free (1 pts)

Full Points

Comments:

POINTS: 15

Problem 6: Conduct gBLUP on both training population and testing population with the phenotypes masked in the testing population. Calculate the correlation between the observed and predicted phenotypes for training population and testing population separately. Compare gBLUP with ridge regression. (15 points).

* Masking phenotypes of testing population (5 pts)

Full Points

* gBLUP (3 pts)

Full Points

* Correlation in training population (1 pt)

Full Points

* Correlation in testing population (1 pts)

Full Points

* Commenting (2 pts)
* Full Points
* Bug-free (3 pts)

Full Points

Comments:

POINTS: 15

Problem 7: Redo (6) with the SNPs identified in (3) as fixed effects for GWAS-assisted GS (15 points).

* Associated SNP as fixed effects (5 pts)

Full Points

* Masking phenotypes of testing population (3 pts)

Full Points

* gBLUP (1 pt)

Full Points

* Prediction accuracy (1 pt)

Full Points

* Commenting (2 pts)
* Full Points
* Bug-free (3 pts)

Full Points

Comments:

POINTS: 15

Problem 8: Redo (6) with the SNPs identified in (2) as fixed effects for GWAS-assisted GS. Explain the difference from (7) (15 points).

* Associated SNPs from (2) (2 pts)

Full Points

* Masking phenotypes of testing population (1 pt)

Full Points

* gBLUP (1 pt)

Full Points

* Prediction accuracy (1 pt)

Full Points

* Explain the difference (5 pts)

Full Points

* Commenting (2 pts)
* Full Points
* Bug-free (3 pts)

Full Points

Comments:

POINTS: 15

**Extra Credit**

Problem 9: Repeat (3-8) at least 30 times and compare prediction accuracy based on their averages and standard errors (30 points).

* Compare among MAS, ridge regression and gBLUP (5 pts)

Full Points

* Compare gBLUP with and without incorporating associated SNPs from (3) (5 pts)

Full Points

* Explain the differences between incorporating associated SNPs from (2) and (3) (10 pts)

Full Points

* Commenting (5 pts)
* Full Points
* Bug-free (5 pts)

Full Points

Problem 10: Redo (2) using residual instead of phenotype as the dependent variable. Take the top ten SNPs with smallest P values as the associated SNPs and redo (9) (30 points).

* Compare gBLUP with and without incorporating associated SNPs from (3) (10 pts)

Full Points

* Explain the differences between incorporating associated SNPs from (2) and (3) (10 pts)

Full Points

* Commenting (5 pts)
* Full Points
* Bug-free (5 pts)

Full Points