Name:

Graded by: TA

**TOTAL POINTS: /100**

Problem 1: Choose a dataset. Display various statistics (position, missing rates, MAF) (20 points)

* Visualize marker locations (4 points)

Full Points

* Calculate and visualize **sample** missing rate (4 points)

Full Points

* Calculate and visualize **marker** missing rate (4 points)

Full Points

* Calculate and visualize minor allele frequency (4 points)

Full Points

* Commenting and bug-free (4 pts)

Full Points

* EXTRA CREDIT: Novel data source, must be sharable and well-described for others to use (5 pts)

Full Points

Comments:

POINTS: 20

Problem 2: Randomly select 5%, 25%, and 50% of data and set to missing. Impute with stochastic imputation. Calculate imputation accuracy as correlation coefficient or match proportion. Iterate 30 times, report average, sd, and number of reps. Describe the relationship between missing rate and imputation accuracy.

* Randomly set values to NA at 5%, 25%, and 50% (4 pts)

Full Points

* Stochastic imputation algorithm implemented correctly (4 pts)

Full Points

* Report average accuracy, sd, and number of iterations (at least 30x) (4 pts)

Full Points

* Effectively describe the relationship between missing rate and imputation accuracy (4 pts)

Full Points

* Commenting and bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 3: Redo the process of problem 2 using KNN method.

* KNN implemented correctly (4 pts)

Full Points

* Report average accuracy, sd, and number of iterations (at least 30x) (4 pts)

Full Points

* Effectively describe the relationship between missing rate and imputation accuracy (4 pts)

Full Points

* Describe the difference between stochastic and KNN imputation (4 pts)

Full Points

* Commenting and bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 4: Switch KNN to calculate distances between markers instead of individuals.

* Correctly modify KNN to use marker distances rather than individuals (4 pts)

Full Points

* Report average accuracy, sd, and number of iterations (4 pts)

Full Points

* Effectively describe the relationship between missing rate and imputation accuracy (4 pts)

Full Points

* Describe the difference between using individuals and markers for KNN (4 pts)

Full Points

* Commenting and bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 5: Fix the missing rate to 25% and perform imputation with BEAGLE.

* Were they able to get BEAGLE to run? (if not done using the R ‘system’ command, did they document their process?) (4 pts)

Full Points

* Calculate imputation accuracy (4 pts)

Full Points

* Report average accuracy, sd, and number of iterations? (4 pts)

Full Points

* Describe the advantage of BEAGLE vs KNN (4 pts)

Full Points

* Commenting and bug-free (4 pts)

Full Points

Comments:

POINTS: 20

Problem 6 (Extra Credit): Redo problem 3, but calculate accuracies separately for each allele.

* Calculate accuracy correctly using only homozygotes (4 pts)

Full Points

* Report average accuracy, sd, and number of iterations (4 pts)

Full Points

* Effectively describe the difference between major and minor allele imputation (4 pts)

Full Points

* Describe patterns observed for this alternate way of calculating accuracy (key is MAF!) (4 pts)

Full Points

* Commenting and bug-free (4 pts)

Full Points

Problem 7 (Extra Credit): Perform imputation with another method; demonstrate an accuracy superior to KNN and BEAGLE.

* The algorithm is explained and implemented correctly (4 pts)

Full Points

* Report average accuracy, sd, and number of iterations (4 pts)

Full Points

* Effectively compare the accuracy to both KNN and BEAGLE (4 pts)

Full Points

* Describe why this method performs better (4 pts)

Full Points

* Commenting and bug-free (4 pts)

Full Points