

Update On GWAS Analysis

Some EDA and Report On Multivariate Shrinkage

Omid Shams Solari¹

¹ UC Berkeley

26 May 2016

Table of Contents

1 EDA

2 Ridge Regression and Multivariate Shrinkage

3 Immediate Goals

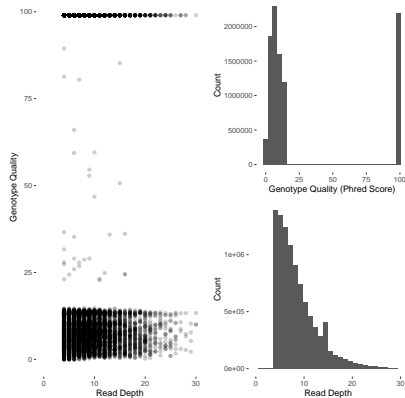


FIGURE – Genotype Distribution

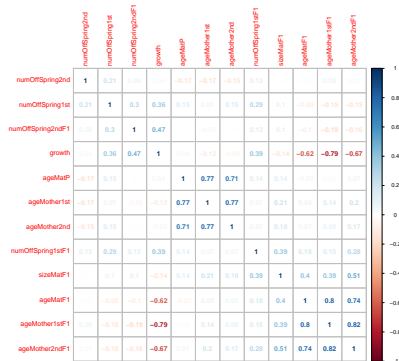


FIGURE – Phenotype Correlation Matrix

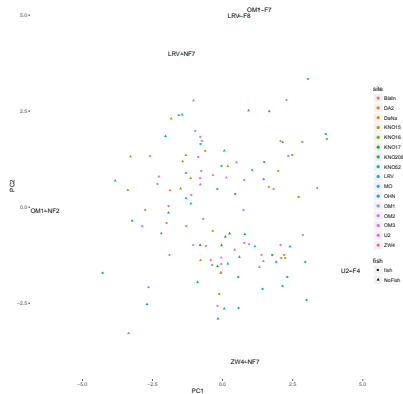


FIGURE – First two principal components

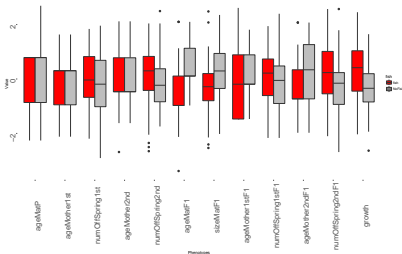


FIGURE – The MSE of the two class imbalance estimators vs. number of samples on a log-log scale.

TABLE – Two-tailed two-sample test of significance of the difference in means

Phenotype	t-Statistic	P-value
ageMatP	-0.2384	0.8119
ageMother1st	-1.6165	0.1089
numOffSpring1st	0.8418	0.4016
ageMother2nd	-1.9701	0.0512
numOffSpring2nd	1.4582	0.1475
ageMatF1	-2.9446	0.0039
sizeMatF1	-3.4709	0.0007
ageMother1stF1	-2.9482	0.0038
numOffSpring1stF1	0.9070	0.3662
ageMother2ndF1	-3.6670	0.0003
numOffSpring2ndF1	2.1287	0.0354
growth	4.0152	0.0001

<https://github.com/osolari/GWAS>

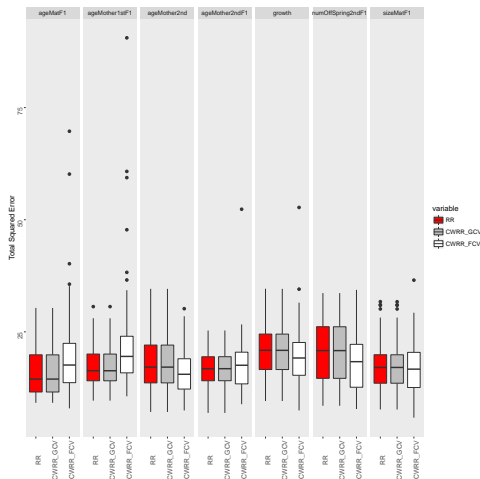


FIGURE – Total Squared Error Comparison of RR , C&W-RR-GCV and C&W-RR-FCV

- Implement Sparse Canonical Correlation Analysis Method.
- Use Random Forest to find reciprocal links.