

Alexandre Lamy  
10/16/17

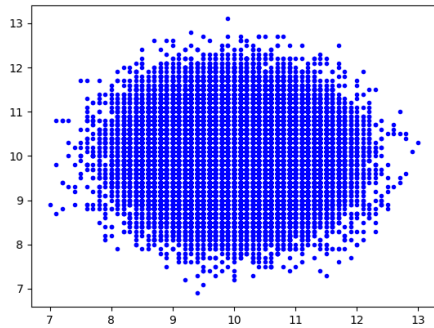
## Clean Data Generation and Naive Buhmbox Cutoff Results

Overview of work done:

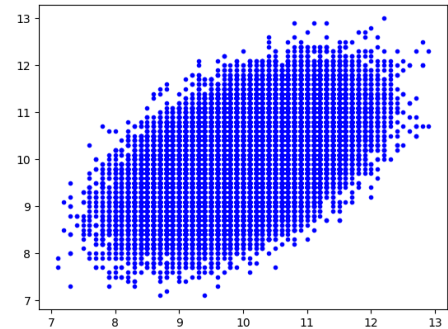
- cleaned up data generation process
- got pipeline from data generation into standard Buhmbox working

Details:

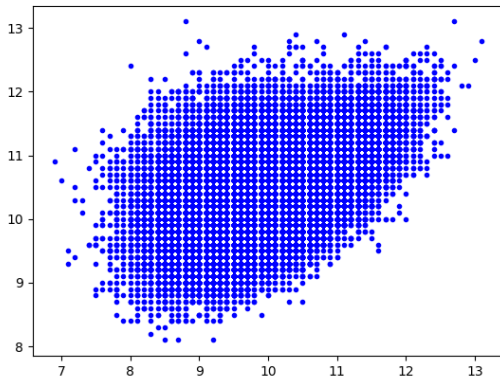
- 100 000 Individuals generated



Independent snps  
100 each



Shared SNPS  
50 independent each, 50 shared each



Heterogeneous individuals

3 phenotypes generated:

1 has 25 independent SNPs and 75 shared with 3

2 has 100 independent SNPs

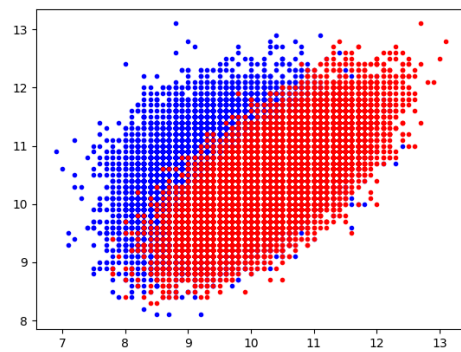
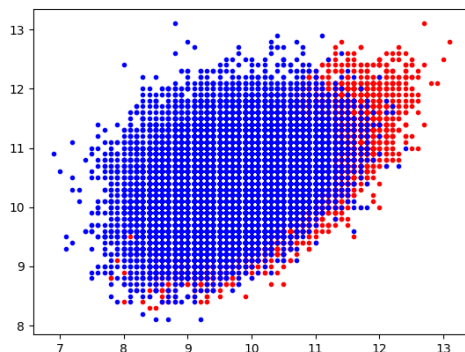
3 has 25 independent SNPs and 75 shared with 1

Then only two phenotypes are actually kept:

phenotype1 = phen1

phenotype2 = max(phen2, phen3)

Below blue is individuals who have phenotype2=phen2, red is phenotype2 = phen3.



- Looked at “cases” for phenotype2, i.e. individuals with phenotype2 values that have z-score of 1.5 or more. Then looked at snps for phenotype1.
- Got the following results:
  - Buhmbox score for independent phenotypes: -0.5972, 0.7928, -0.0560
  - Buhmbox score for pleiotropy: -12.3058, -12.7265, -11.9245
  - Buhmbox score for heterogeneity: 28.6874, 29.0802, 28.7971

Next goals:

- Modify buhmbox to use weights rather than cases/controls (which induce an arbitrary cutoff).