10-15-18

* He will email pages that talk about this from grant
* Grant proposes how to deal with correlation that exists in our tests
* He will send papers he found recently that deal with model goodness of fit
* Burk-jones stat (GOF stat) we will want to look at (also assumes iid p-values)
* Pre-print he has for group using burk-jones stat in genetics and worry about correlations between stats
* At least get through grant and some of papers by next week
* Once we have a good grasp of goodness of fit concept we can start some basic simulations and writing some code to look at behavior of stats or tests
* Doesn’t expect me to have read all by next week
* We are getting closer to proposal stage, don’t need to read the other papers to start doing some work
* Once we setup problem we could setup some basic simulations to get feet wet
* i.e. simple iid cases with diff distributions of p-values to get feeling for how these behave even before we introduce correlation
* Looking at only candidate snps only gets what we aready knew to look at as candidates (huge jump)
* HC are hoped to be more sensitive when number of tests gets big and sparsity of true signals gets increased
* We might use burk-jones
* Local levels
* Estimate correlation structure of p-values from data
* Get idea of sampling dist under null and use that to perform test trough simulation
* One approach is no change to test stats above but just look at simulation of sampling dist under given correlation structure of p-values
* May leave some power out because what was good for iid may not be good for correlation
* Also may want to adjust test stat to take correlation into account in first place
* Pick some appropriate models in first place
* Might be mathematical statistics portion of this project mary sara will know better and this will make stats dept happy
* Include in email reference to lambda mean calcs
* Individual p-values for us come from regress transcript level on genotype of eqtl locus with grm lmm,
* In gwas phenotype is the same across tests you just change the candidate snp
* In eqtl you change the phenotype
* People typically don’t do entire genome, and do eqtl of those snps against 10,000 genes or subset of that
* That leaves out a lot of tests
* Most of them will be null
* They mostly do not do asymptotically
* Once not iid asymptotics get really challenging
* If we can figure out the asymptotics that’s great, cause then you can compute stuff exactly
* Might be mathematical models of why statistics work in iid case then you might be able to take that model or idea to the non-iid case and perform similar actions, this introduces new way yo do the test
* Work: has higher power
* Idea we can get higher power

10-22

* Know two versions of burk-jones stats
* Try to reproduce Mark’s Table 1 from proposal
  + Fold in new stuff like BJ stats
  + i.e one paper deals with BJ and adding in correlation
* try simulations just uinder independence assumption for example
* could be doing that at same time as doing some of these other papers
* and approach to correlation
* send brief update to her

10-29

* m=80, mu=2.146
* m=40, mu=2.54
* m=20, mu=3.035
* empirical estimation of correlation of the z scores
* use multiple SNPS to estimate the correlation structure between the z-scores
* Z~MVN(0,Sig)
* To start make sig diag=1 non-diag=rho for some rho so symmetry means exchangeable and thus doesn’t matter which 80 or 40 you chose to set mu=1
* H0: Calculate the critical values from 500 draws with sigma estimated
* H1: start off with all off diagonal for Z mvn equal to row

11-5-18

* Figure out the Cholesky
  + Band diagonal (vary number of bands i.e. just first or second, etc diagonal off main)
  + All correlated at rho
  + Blocks
    - Quantify extent of concentration of eps false null in single blocks versus many blocks
    - i.e. 80 blocks, all 80 false null in one block -> one non null per block
  + follow example of “Set-Based Tests for Genetic Association Using the Generalized Berk-Jones Statistic”
* Look at donoho jin paper pg 19 and try to remake fig 10 empirically with simple grid search
  + Map out gof for parameter space (mu,eps) grid search and then find heat map for each stat for its rel order and abs power value
* Look more into score test (possible in trans eqtl)
  + Maybe look more at heritability and score test of heritability