Lecture 9 Exercises:

ALL of the code and scripts are inside of the file: Lecture_9_code.txt. Because the code for doing some of these tasks is quite involved, please copy and paste from that file, but do make efforts to understand what the code is generally doing.

Exercise 1. Read in the variants in Gene2 and read in trait2. Analyze the rare variants in Gene2 using both SKAT and also weighted count based collapsing where the weights are the same as in SKAT (i.e. beta density function with parameters 1 and 25).

Exercise 2. Now run the omnibus SKAT, but consider setting rho (r.corr) = 0 and rho = 1 and rho = 0.5. What do you notice?

Exercise 3. Run the omnibus version of SKAT, but use the "optimal" approach which searches across a range of rho values.

Exercise 4. Power calculations. Suppose we're writing a grant to do a new project in which we would like to sequence a bunch of people with some continuous trait and then apply SKAT to run the analysis exome wide (i.e. we will be testing 20,000 genes such that the alpha level is 2.5e-6)

We posit the following:

- Average region length will be 5kb
- 20% of the variants with MAF < 5% will be causal
- 20% of the causal variants will decrease the trait value while 80% will increase the trait value
- We assume the magnitude of the effect sizes for the causal variants is equal to $-c \text{Log}_{10} \text{MAF}$ where we set c such that a variant with allele frequency of 1/10,000 is 2.

Let's further assume that we do not have prior sequencing data so we will need to use simulated chromosomes from COSI (let's further assume European ancestry for simplicity).

Calculate the anticipated power if our sample size is 1000.

How many subjects would we need to have 80% power?