Statistical Theory and Methodology as it Applies to the Study of Genetics
Saleh Wehelie
Mentor: Seth Temple

This quarter, I delved into statistical genetics and learned how to to compute measures of relatedness through kinship and inbreeding as well as the risk of an individual inheriting a conditional trait (or disease) based on known family trees. After studying this material, I also learned about the applications of biostatistics and statistical genetics in the real world and how it has a unique role in safeguarding public health and improving lives through quantitative research.

The main subjects I reviewed for my project were kinship, inbreeding, identity by descent (IBD), and genotype probabilities. We can calculate the relatedness between individuals (kinship and inbreeding) through something called the Path Counting Formula given a pedigree.

$$\psi(B,C) = \sum_{A} \sum_{P(A)} (1 + f_A)(½)^{n(P(A))+1}$$

This equation above represents the Path Counting Formula to calculate the kinship between two individuals B and C which then returns us the value for the kinship coefficient. Identity by descent (IBD) is often used to define if two genes between two individuals both come from the same ancestral gene and is a commonly used term to describe the degree of relatedness between two individuals. Two individuals are IBD if both individuals were to inherit the same haplotype in the same locus (chromosomal location) from a shared common ancestor.

The main focus of my project was conducted using genotype probabilities. Given two pedigree charts, one with inbred offspring and one without, and given that individual 5 has a recessive genotype aa that is present in the population at 20% frequency, the goal was to calculate the probability individual 6 (6 being the full sibling to individual 5 in both pedigree charts) also having the homozygous recessive genotype aa making the alleles in both siblings IBD. We can calculate the genotype probability of individual 6 by finding the conditional probability through the Law of Total Probability. The conditional probability is the likelihood of an event occurring based on the occurrence of another event and the Law of Total Probability is a fundamental rule relating marginal probabilities to conditional probabilities:

$$P(Gk = aa | Gj = aa, T) = \frac{P(Gk = aa, Gj = aa | T)}{P(Gj = aa | T)}$$

Using the probability formula above, we were able to infer that the probability of individual 6 inheriting the same genotype as individual 5 was 44.5% in the case that they were both products of inbreeding and 36.0% in the case they were not inbred. From these results we can conclude that inbreeding significantly increases the chances of homozygous recessive genotypes also increasing the risk of disorders caused by these recessive genes.

Overall, through the SPA DRP program, I was able to gain valuable experience of research in biostatistics and statistical genetics which has helped immensely in learning more about the everyday life of statisticians as I work to pursue this career path.