BIOST 551/STAT 551: Autumn Quarter 2014
Homework 1
Due Thursday, October 9, 2014

Show All Work

1. [20 points] In class, we discussed the FecB gene that affects fecundity (litter size) in Merino Sheep. Assume that the three genotypes at the FecB gene yield the following mean litter sizes:

Genotype	Mean Litter Size
bb	1.48
Bb	2.17
BB	2.66

   (a) If the frequency of allele B for the FecB gene is .18 in the sheep population, provide the mean litter size in the population. Also provide a mean litter size in the population that is scaled relative to the average of the mean litter sizes for the two homozygotes.

   (b) What is mean litter size if the frequency of the B allele is doubled (frequency of the B allele is .36) in the population?

   (c) What are the average allele effects for the B and b alleles for the two population allele frequencies given in questions 1(a) and 1(b) above?

   (d) What is the average effect of allele substitution for the two population allele frequencies given in questions 1(a) and 1(b) above?

2. [40 points] A single autosomal locus with two alleles, D and d, contributes to a phenotype of interest. One thousand individuals sampled from a population were both screened for the phenotype and genotyped at the locus. A file named "hw1data.txt" containing the phenotype values and genotypes for the sample individuals can be downloaded from the course website in the "Data Sets" section. Feel free to use a statistical software of your choice for this question. To read the text file into R, the following command can be used:

```
mydata=read.table("hw1data.txt",header=TRUE).
```

   (a) Using linear regression, answer the following questions:

      i. What is the average effect of allele substitution from the linear regression model? What are the estimated average allele effects for the D and d allele? Give your linear regression model.

ii. What are the estimated breeding values for each genotype. (The breeding value for a genotype is defined to be the sum of the allelic effects for the two alleles of the genotype).

iii. Give an estimate of the narrow sense heritability (h^2) for this locus from the linear regression.

(b) Now, using only the mean phenotype values for the three possible genotypes and allele frequency estimates for the D and d alleles from the sample, estimate the average allelic effects for the D and d alleles. How do these estimates compare to your estimates using the linear regression model for question 2(a) above?

3. [40 points] A phenotype Y is influenced by a single autosomal locus. The locus is bi-allelic with alleles A and a and the allele frequencies are p and q, respectively, in a population. The population is in Hardy-Weinberg equilibrium (HWE) at this locus, and the phenotype has the following model:

$$Y = \mu_Y + \alpha_1 X_1 + \alpha_2 X_2 + \beta I_{X_1=1} + \epsilon$$

where μ_Y is the population mean of the phenotype, X_1 and X_2 are the number of copies, respectively, of the A and a allele an individual has, $I_{X_1=1}$ is an indicator for the heterozygotes genotype (i.e., takes a value of 1 if the genotype is heterozygous, and 0 if the genotype is homozygous), and ϵ is a random effect for the environmental deviation that is independent of genotype at the locus where $\epsilon \sim N(0, \sigma^2)$ (that is, ϵ follows a normal distribution with mean 0 and variance σ^2). Note that $X_2 = 2 - X_1$.

(a) Calculate the expected value and variance of the phenotype Y in the population in terms of the parameters p, q, α_1, α_2, β, and σ^2.

(b) What is the broad sense heritability (H^2) of this phenotype in the population?