

Session 2: Association Testing with Quantitative Traits

Analysis of Blood Pressure Data with R

Before you begin:

- Go to the course website and locate the `bpdata.csv` dataset. Load the `bpdata.csv` data file into your R session. This is a comma separated file that can be read directly from the web (if you are connected to the web) using the following command:

```
bpdata=read.csv("http://faculty.washington.edu/  
tathornt/sisg/bpdata.csv",header=TRUE)
```

Alternatively, save the file to your computer and read it into R from the directory where the file is:

```
bpdata=read.table("bpdata.csv",header=TRUE)
```

The `bpdata.csv` dataset contains diastolic and systolic blood pressure measurements for 1000 individuals, and genotype data for 11 SNPs in a candidate gene for blood pressure. Covariates such as gender (`sex`) and body mass index (`bmi`) are included as well.

1. Perform a linear regression in R of systolic blood pressure (`sbp`) on `SNP3` using `lm()`. Compare the estimates, intervals and p-values you get using

- additive (linear) model
- dominant model
- recessive model
- 2 parameter model

(hint; for each case, first add a new column to the data frame, containing the 'predictor' variable you need. Then do the regression, using `lm()`.)

2. Provide plots illustrating the relationship between `sbp` and the three genotypes at `SNP3`. (Hint: try plotting scatterplots and boxplots.)

For question 3 and 4 below, R also has a 'formula' syntax, frequently used when specifying multivariate regression models, with many predictors. To regress outcome on several covariates, the syntax is

```
outcome ~ covariate1 + covariate2 + covariate3
```

3. Now redo the linear regression analysis of `sbp` from question 2 **for the additive model**, but this time adjust for `sex` and `bmi`. Do the results change?
4. What proportion of the heritability of `sbp` is explained by all of the 11 SNPs together?