

Elements of R for Genetics & Bioinformatics

Session 5. Simulation, and permutation tests

1. The **power** of a test (at Type I error rate of, say, 0.05) is the probability that its p-value will be less than 0.05. One way to estimate the power is by simulating the study many times and calculating the proportion of simulations where the p-value is less than 0.05.

Suppose we are to test whether there is a significant change in blood pressure (from e.g. before treatment to after treatment) in 100 individuals. An appropriate statistical test is a one-sample t-test, and it is reasonable to assume that the 100 subjects' changes are sampled from a Normal distribution, with mean 2mmHg, and standard deviation 7mmHg.

Using `rnorm()`, `t.test()` and `replicate()`, compute the power of this test, by simulation. Keen people: compare your answer to that computed by the `power.t.test()` function.

2. In this question we investigate differences in mean systolic blood pressure by genotype, for the data on 11 SNPs you saw before. To obtain p-values for linear regression analysis of the association between blood pressure and genotype for a single variant, we can use e.g.

```
#read in the data, and merge
justsnps    <- read.table("example-snp.txt",header=TRUE)
sampleinfo  <- read.csv("example-pheno.csv", header=TRUE)
merged      <- merge(justsnps, sampleinfo, by="id")

#computing p-value for linear regression of SBP on snp3
lm.temp    <- lm(merged$sbp~as.numeric(merged[,4]))
summary(lm.temp)
coef(summary(lm.temp))[2,4]
```

To do this in a loop, we could use `apply()`, e.g.

```
all.p <- apply(merged[,2:12], 2, function(snp_i){
  n.minor <- as.numeric(factor(snp_i)) # turn CC/CT/TT into 0/1/2
  lm1 <- lm(merged$sbp~n.minor)        # do regression
  coef(summary(lm1))[2,4]              # extract p-value
})
```

The 'best' (i.e. minimum) p-value is obtained by e.g.

```
min(all.p)
```

Use a permutation test of this minimum p-value across all 11 SNPs to assess whether there is *really* a statistically significant effect of any of these SNPs on systolic blood pressure.