Applied Statistical Methods – Exercise 1

Peter von Rohr

2021-03-01

Problem 1: Linear Regression

Use the example dataset from the course notes which is used to demonstrate how to fit a regression of the response variable `body weight` (BW) on the predictor variable `breast circumference` (BC). The data is shown in the table below.

Table 1: Dataset for Regression of Body Weight on Breast Circumference for ten Animals

Animal	Breast Circumference	Body Weight
1	176	471
2	177	463
3	178	481
4	179	470
5	179	496
6	180	491
7	181	518
8	182	511
9	183	510
10	184	541

Your Tasks

- Compute the regression coefficient using matrix computations. Use the function `solve()` in R to compute the inverse of a matrix.
- Verify your results using the function `lm` in R.

Problem 2: Breeding Values

During the lecture the computation of the breeding values for a given genotype was shown for a completely additive locus which means the genotypic value d of the heterozygous genotypes is 0. In this exercise, we want to compute the general solution for the breeding values of all three genotypes under a monogenic model. The term `monogenic model` is equivalent to a single-locus model.

We are given a single locus G with two alleles G_1 and G_2 which are closely linked to a QTL for a trait of interest. We assume that the population is in Hardy-Weinberg equilibrium at the given locus G. It is important to note here, that the breeding values under this single-locus model are not the same as the direct genomic breeding values. In one of the following exercises, we will come back to this difference.

The allele frequencies are

Allele	Frequency
G_1	p
G_2	q

Allele G_1 is the one with a positive effect on the trait of interest. The genotypic values are given in the following table.

Genotype	Value
G_1G_1	a
G_1G_2	d
G_2G_2	$-a$

Your Task

- Compute the breeding values for all three genotypes G_1G_1, G_1G_2 and G_2G_2.
- Verify the results presented in the lecture by setting $d = 0$ in the breeding values you computed before.

Problem 3: Linkage Between SNP and QTL

In a population of breeding animals, we are given a trait of interest which is determined by a QTL Q on chromosome 1. QTL Q is modelled as a bi-allelic QTL with alleles Q_1 and Q_2. Furthermore we have genotyped our population for two SNPs R and S with two alleles each. One of the SNPs is on chromosome 1 and is closely linked to Q. The other SNP is on chromosome 2 and is unlinked. Figure 1 shows the situation in a diagram.

Based on the following small dataset, determine which of the two SNPs R and/or S is linked to QTL Q.

From the above table it might be difficult to decide which SNP is linked to the QTL. Plotting the data may help. Showing the observations as a function of the genotypes leads to Figure 2.

Your Tasks

- Determine which of the two SNPs R or S is closely linked to the QTL
- Estimate a value for a obtained based on the data
- Try to fit a linear model through the genotypes that SNP which is linked to the QTL using the `lm()` function. The genotype data is available from

`https://charlotte-ngs.github.io/gelasmss2021/data/asm_w02_ex01_p02_genodatafile.csv`

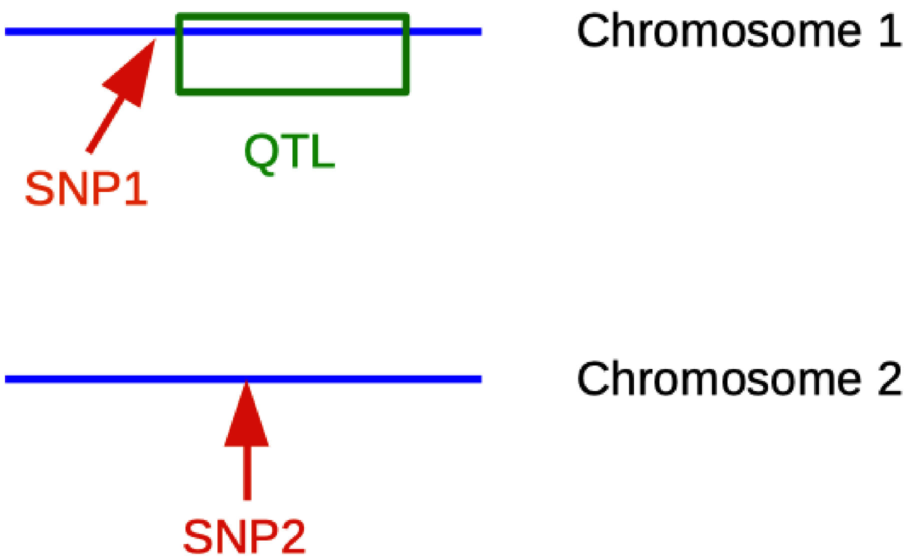

Figure 1: Linkage Between an SNP and a QTL and an independent SNP on a different Chromosome

Table 4: Dataset showing linkage between SNP and QTL

SNP R	SNP S	Observation
R_2R_2	S_1S_1	23.17
R_2R_2	S_2S_2	-27.04
R_1R_2	S_1S_2	-2.79
R_1R_2	S_2S_2	-19.54
R_1R_2	S_2S_2	-24.05
R_1R_2	S_1S_1	25.84
R_1R_2	S_1S_2	-0.36
R_1R_1	S_2S_2	-23.34
R_2R_2	S_1S_2	1.38
R_1R_1	S_1S_2	-1.60
R_1R_2	S_1S_2	-2.97
R_2R_2	S_1S_2	-1.39

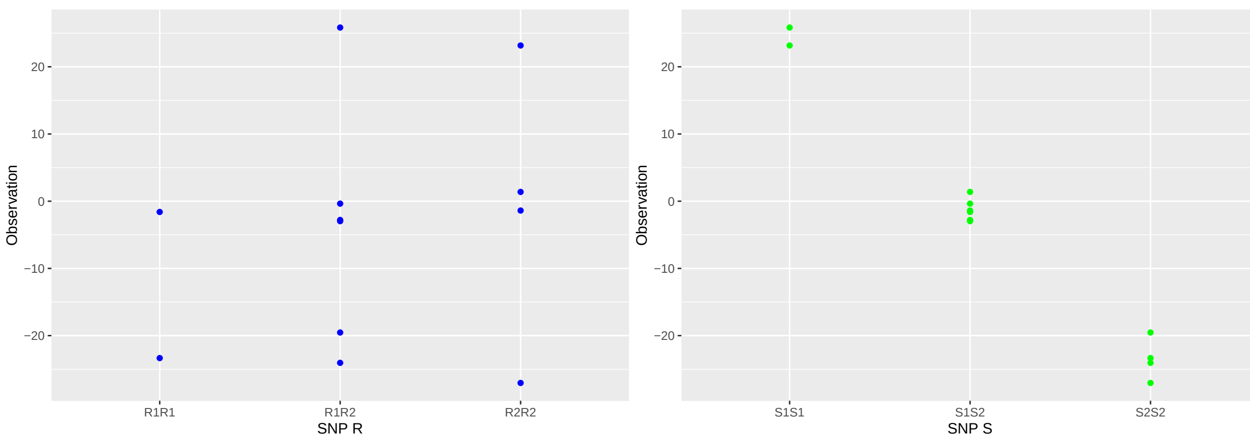

Figure 2: Observations Grouped by SNP Genotypes