Fixed Linear Effects Models

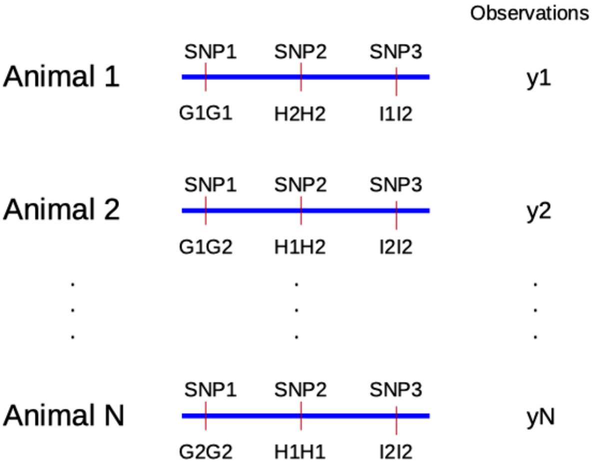
Peter von Rohr

25.02.2019

Background

- Given a population of N animals
- Each animal has information on genotypes at loci G, H and I
- Each animal has an observation for one quantitative trait of interest y
- **Goal**: Predict genomic breeding values

Data


Observations
SNP1
SNP2
SNP3
Animal 1
y1
G1G1
H2H2
I1I2
SNP1
SNP2
SNP3
Animal 2
y2
G1G2
H1H2
I2I2
.
.
.
SNP1
SNP2
SNP3
Animal N
yN
G2G2
H1H1
I2I2

Two Types Of Models

1. **Genetic** Model: How can we decompose the phenotype into genetic part and non-genetic environmental part
2. **Statistical** Model: How to estimate unknown parameters from a dataset

Goals:

1. Use genetic model to show how observations and genetic information can be used to predict breeding values.
2. Use statistical techniques to do the prediction

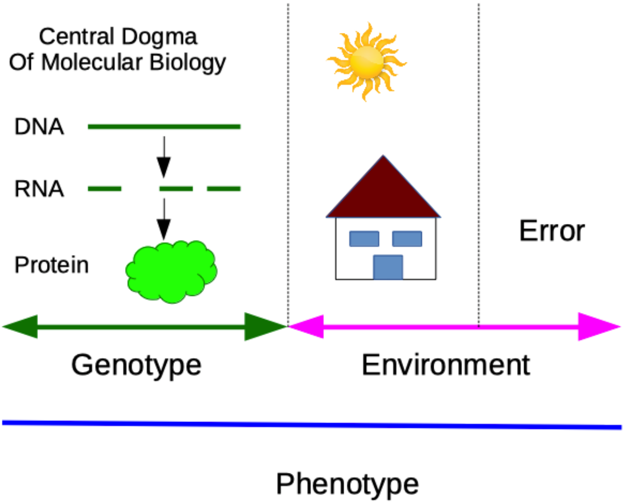
Genetic Model

- simple model from quantitative genetics to split phenotypic observation into
  - genetic part g and
  - environmental part e

$$y = g + e$$

- environment: split into
  - known environmental factors: `herd`, `year`, `...` (β)
  - unknown random error (ϵ)
- polygenic model: use a finite number of loci to model genetic part of phenotypic observation

Genetic Model (II)



Polygenic Model

- Component g can be decomposed into contributions g_j of single loci

$$g = \sum_{j=1}^{k} g_j$$

- Assume that loci are additive, hence genotypic values g_j depends on a_j with $d_j = 0$
- Genotypic values at locus j can either be $-a_j$, 0 or a_j
- Breeding values based on locus j depends on a_j.

Genotypic Value

- Genotypic value g_i for animal i over all loci

$$g_i = M_i \cdot a$$

where M_i is a row vector with elements -1, 0 and 1 and a is the vector of all genotypic values of the positive homozygous genotypes of all loci.

Phenotypic Value

- Collecting all components for an observation y_i for animal i

$$y_i = W_i \cdot \beta + M_i \cdot a + \epsilon_i$$

- all animals in the population

$$y = W \cdot \beta + M \cdot a + \epsilon$$

- combining $b^T = \left[\begin{array}{cc} \beta & a \end{array} \right]$ and $X = \left[\begin{array}{cc} W & M \end{array} \right]$

$$y = X \cdot b + \epsilon$$

Statistical Model

- genetic model from statistics point of view
- phenotypic observation as response y
- vector b (known environment and genotypic values) as unknown parameter
- fixed predictor variales in matrix X
- vector ϵ as random error terms

$\rightarrow$ Fixed Linear Effects Model

Parameter Estimation

- use regression model
- regression means both response and predictors are continuous
- example dataset: body weight on breast circumference

Regression Dataset

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

Regression Model

- response y: body weight
- predictor x: breast circumference
- model for observation y_i

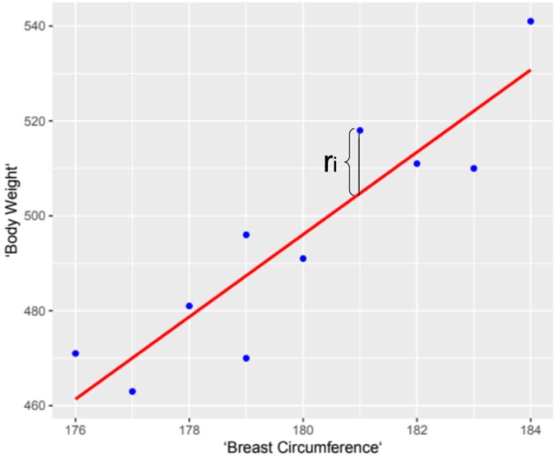
$$y_i = x_i * b + \epsilon_i$$

- meaning of b: change x_i by one unit $\rightarrow$ y_i changes on average by b units.
- use case: measure x_{N+1} for animal $N+1$ with unknown weight and use b to predict y_{N+1}

Least Squares

- How to find b such that y is best approximated by x
- Residuals $r_i = y_i - x_i * b$
- Minimization of sum of squared residuals (SSQ_R)
- Use $\hat{b}$ at minimal SSQ_R as estimate

LSQ Diagram



540
520
500
480
460
'Body Weight'
r_i
176
178
180
182
184
'Breast Circumference'