Fixed Linear Effects Models

Peter von Rohr

2022-03-07

Extension of Dataset on Body Weight

Animal	BC	Body Weight	BCS	HEI	Breed
1	176	471	5.0	161	Angus
2	177	463	4.2	121	Angus
3	178	481	4.9	157	Simmental
4	179	470	3.0	165	Angus
5	179	496	6.8	136	Simmental
6	180	491	4.9	123	Simmental
7	181	518	4.4	163	Limousin
8	182	511	4.4	149	Limousin
9	183	510	3.5	143	Limousin
10	184	541	4.7	130	Limousin

Include Breed into Model

- Breed has an influence on body weight
- Predictor variables must be numeric
- Breed must be converted to numeric code
- Assignment of codes to breeds is rather arbitrary

Breed Codes

Code	Breed
1	Angus
2	Limousin
3	Simmental

Modelling Effect of Breed

- Simplification: "breed" is the only predictor
- Expected body weight (y_i) for animal i

$$E(y_i) = b_0 + b_1 x_i$$

Problems

- Nothing wrong with previous model
- But the following relations might give a hint to some problems

$$E(\text{BW Angus}) = b_0 + b_1$$
$$E(\text{BW Limousin}) = b_0 + 2b_1$$
$$E(\text{BW Simmental}) = b_0 + 3b_1$$

This means, for example, that

$$E(\text{BW Limousin}) - E(\text{BW Angus}) =$$
$$E(\text{BW Simmental}) - E(\text{BW Limousin})$$
$$E(\text{BW Simmental}) - E(\text{BW Angus}) =$$
$$2\left[E(\text{BW Limousin}) - E(\text{BW Angus})\right]$$

Consequences

- Allocation of numerical codes imposes relations between expected values
- Relations might be unreasonable
- Regression analysis only yields estimates for b_0 and b_1, effects of other breeds are determined
- Conclusion: regression on numerical codes of discrete variables are in most cases unreasonable
- Exception: Estimation of marker effects

Linear Regression Analysis for Genomic Data