

OHP Picture 1

Recap 2024-03-18 (1)

□ Matrix-Vector Notation for linear regression

In contrast to the scalar notation

$$y_i = b_0 + b_1 \cdot x_{i1} + e_i \quad \text{for animal } i$$

Data set on body weight and broad circumference

Intercept

$$\begin{cases} y_1 = b_0 + b_1 \cdot x_{11} + e_1 & 471 = b_0 + b_1 \cdot 176 + e_1 \\ y_2 = b_0 + b_1 \cdot x_{21} + e_2 & 463 = b_0 + b_1 \cdot 177 + e_2 \\ \vdots \\ y_{10} = b_0 + b_1 \cdot x_{101} + e_{10} & 541 = b_0 + b_1 \cdot 184 + e_{10} \end{cases}$$

Matrix-Vector Notation: Define vectors y, b, e

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix} = \begin{bmatrix} 471 \\ 463 \\ \vdots \\ 541 \end{bmatrix}; \quad b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}; \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_{10} \end{bmatrix}$$
$$X = \begin{bmatrix} 1 & 176 \\ 1 & 177 \\ \vdots & \vdots \\ 1 & 184 \end{bmatrix}; \quad \text{Model: } y = X \cdot b + e$$

Least Squares Estimate \hat{b} for b

$$\hat{b} = (X^T X)^{-1} X^T y$$

OHP Picture 2

(2)

□ $y = X \cdot b + e$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{10} \end{bmatrix} = \begin{bmatrix} 1 & 176 \\ 1 & 177 \\ \vdots & \vdots \\ 1 & 184 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{10} \end{bmatrix}$$

$$y_1 = 1 \cdot b_0 + 176 \cdot b_1 + e_1$$

$$y_2 = 1 \cdot b_0 + 177 \cdot b_1 + e_2$$

$$y_{10} = 1 \cdot b_0 + 184 \cdot b_1 + e_{10}$$

$\hat{b} = (X^T X)^{-1} X^T y$

Inverse of $(X^T X)$ is defined as the matrix M such that $M(X^T X) = I$ where I is identity
 $(X^T X)^{-1} (X^T X) = I$

$X^T X = \begin{bmatrix} N & X_0 \\ X_0 & (X^2)_0 \end{bmatrix}$

$X_0 = \sum_{i=1}^{N=10} x_i$, $(X^2)_0 = \sum_{i=1}^{N=10} x_i^2$

$X^T X = \begin{bmatrix} 11 & 176 & 177 \\ 176 & 177 & 184 \\ 176 & 177 & 184 \end{bmatrix}$

3

$$X^T y = \begin{bmatrix} 11 & \dots & 1 \\ 176 & 177 & \dots & 184 \end{bmatrix} \begin{bmatrix} 4952 \\ 891393 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} y_{\cdot} \\ (xy)_{\cdot} \end{bmatrix} \quad \text{with } y_{\cdot} = \sum_{i=1}^{N=10} y_i$$

$$(xy)_{\cdot} = \sum_{i=1}^{N=10} x_i \cdot y_i$$

Fix Linear Effect Models

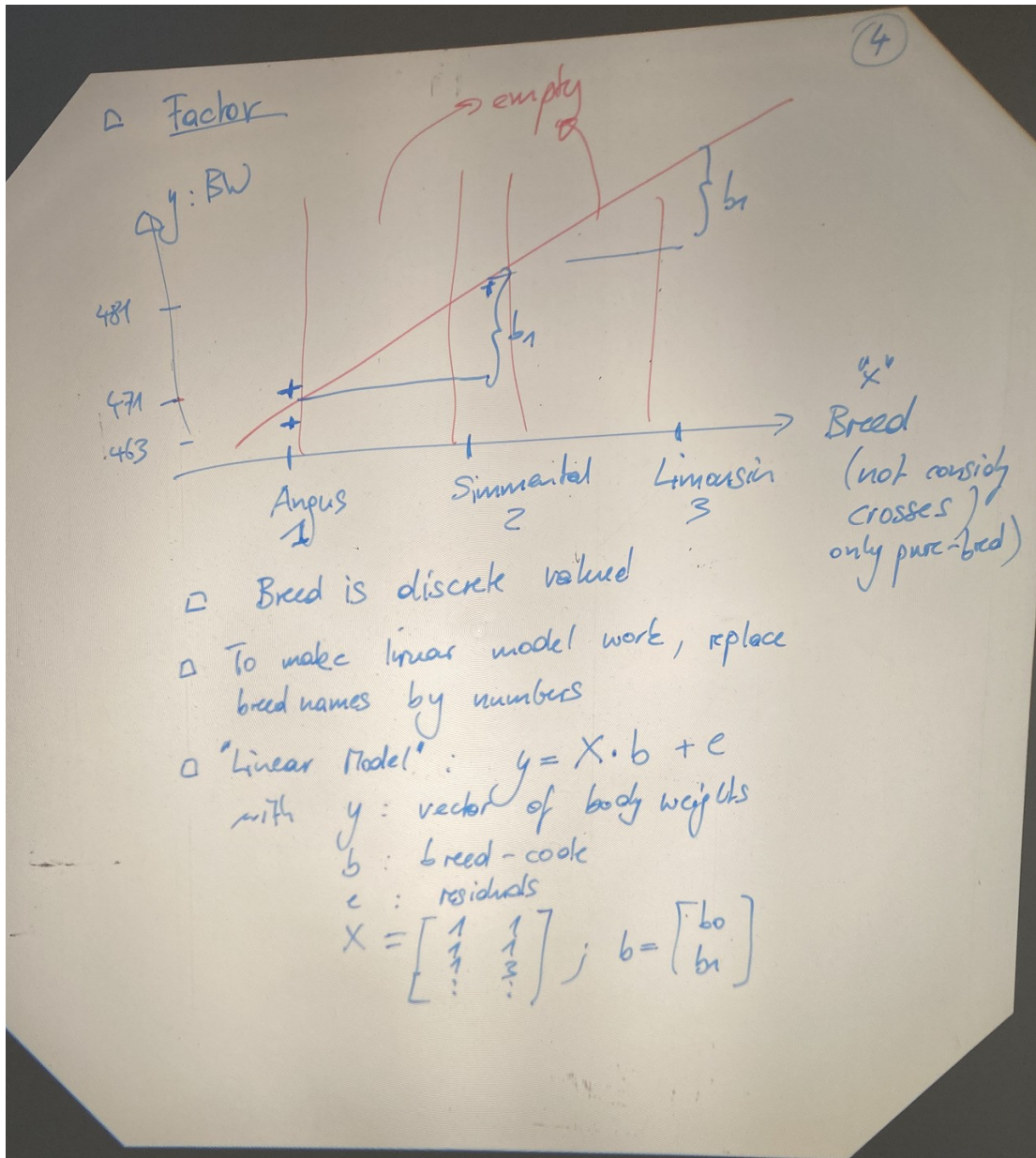
- Generalisation of Regression Models
- Allow to include "Factors" into the model
- Factors are discrete valued variable used as predictors (in contrast to real-valued predictors in regression)

□ Reg: $BW \Delta y$

floating point numbers

176 184 x BC

OHP Picture 4

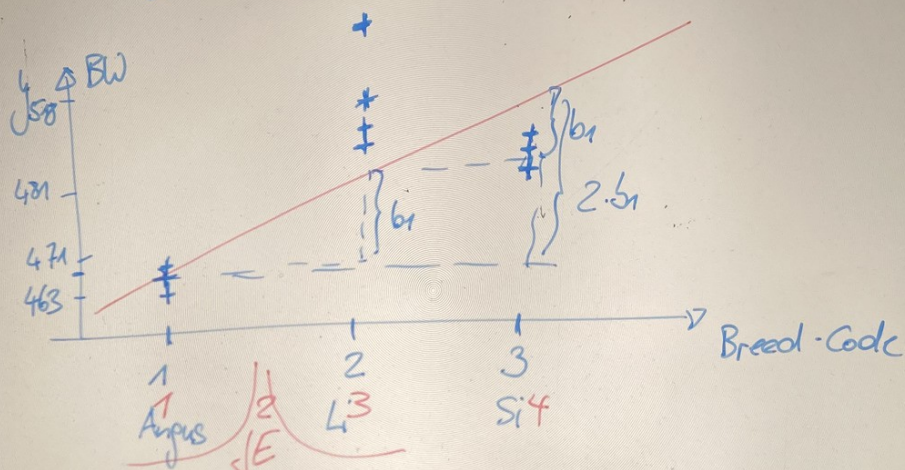


OHP Picture 5

□ Assign "numbers" (code) to breeds

⑤

Code	Breed
1	Angus
2	Limousin
3	Simmental



$$\left. \begin{aligned} E(y_1) &= b_0 + b_1 \cdot 1 \\ E(y_2) &= b_0 + b_1 \cdot 2 \\ E(y_3) &= b_0 + b_1 \cdot 3 \\ E(y_0) &= b_0 + b_1 \cdot 0 \end{aligned} \right\}$$

$$\begin{aligned} E(y_3) - E(y_1) &= b_0 + b_1 \cdot 3 \\ &\quad - b_0 + b_1 \cdot 1 \\ &= 2b_1 \end{aligned}$$

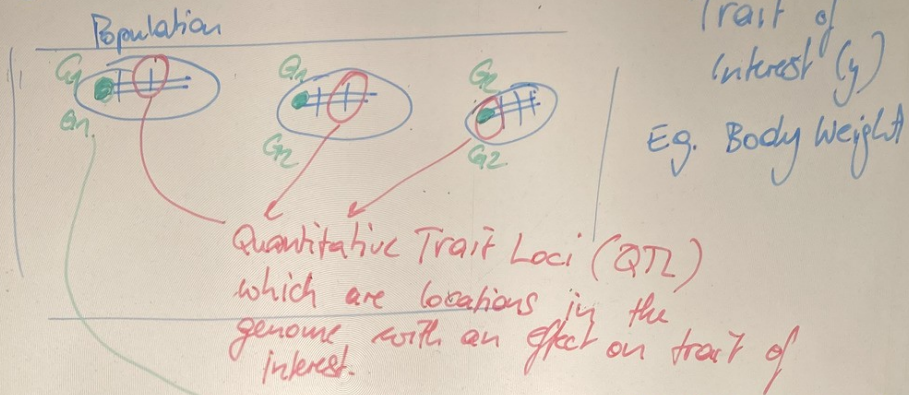
$$\begin{aligned} E(y_3) - E(y_0) &= b_0 + b_1 \cdot 3 \\ &\quad - b_0 - b_1 \cdot 0 \\ &= 3b_1 \end{aligned}$$

$$\begin{aligned} E(y_2) - E(y_0) &= b_0 + b_1 \cdot 2 - b_0 - b_1 \cdot 0 \\ &= 2b_1 \end{aligned}$$

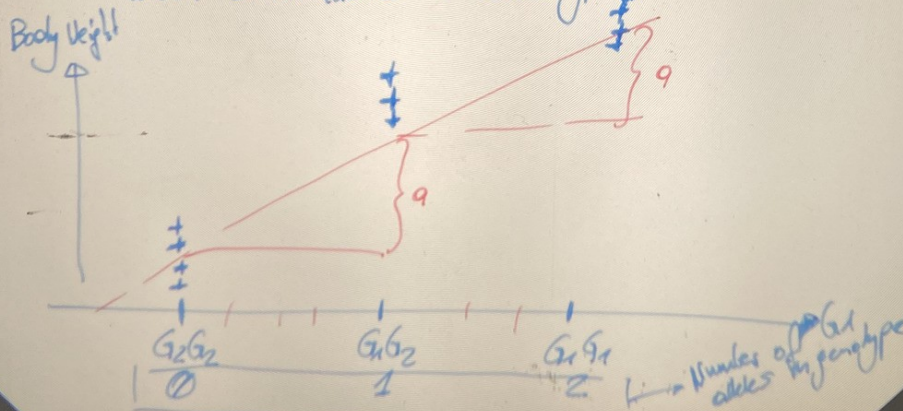
OHP Picture 6

□ Exception: Genomic Data where regression on discrete predictor variable make sense. (6)

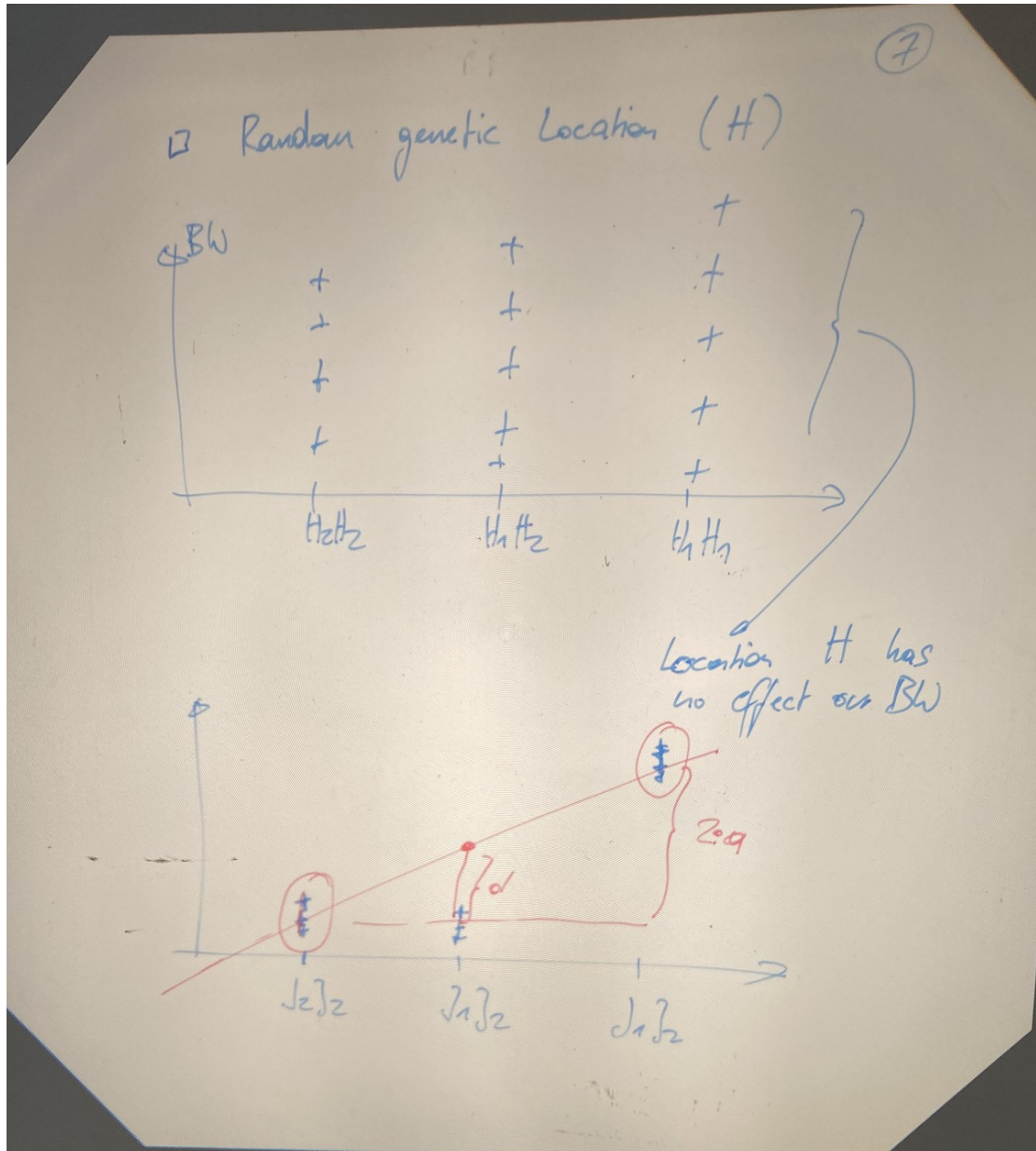
□ Excursion to Genetics



□ Assume: Only 1 QTL for a given Trait (Major Gene) which is bi-allelic \Rightarrow 3 Genotypes



OHP Picture 7



OHP Picture 8

⑧

- Fixed Linear Effect Model also known as "Regression on dummy variables"
- Solution to problem with Breeds as a Factor
 - Not just use one predictor per factor, but use as many predictors as there are levels in the factor.
- Eg. : Factor : Breed
Levels : Angus, Limousin, Simmental

$$E(y_i) = b_0 + \cancel{b_{Angus}} \cdot x_{Angus} + b_{Limousin} \cdot x_{Limousin} + b_{Simmental} \cdot x_{Simmental}.$$