OHP Picture 1

Recap 2024-03-18

1

□ Matrix-Vector Notation for linear regression

In contrast to the scalar notation

$$y_i = b_0 + b_1 \cdot x_{1i} + e_i \quad \text{for animal } i$$

Data set on body weight and breast circumference

Intercept

$$
\begin{aligned}
y_1 &= b_0 + b_1 \cdot x_{11} + e_1 & 471 &= b_0 + b_1 \cdot 176 + e_1 \\
y_2 &= b_0 + b_1 \cdot x_{12} + e_2 & 463 &= b_0 + b_1 \cdot 177 + e_2 \\
&\vdots & &\vdots \\
y_{10} &= b_0 + b_1 \cdot x_{110} + e_{10} & 541 &= b_0 + b_1 \cdot 184 + e_2
\end{aligned}
$$

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

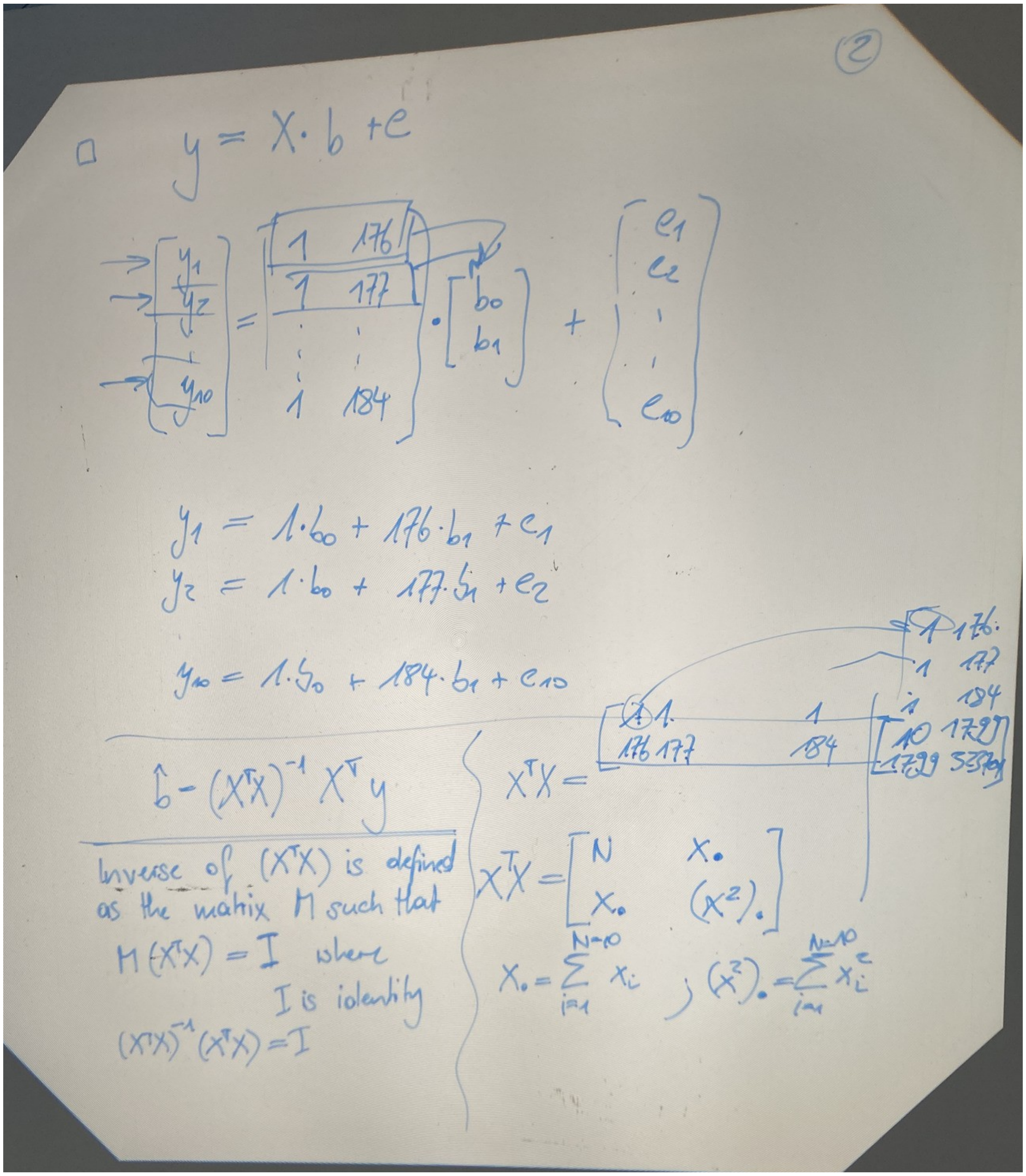
$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix} = \begin{bmatrix} 471 \\ 463 \\ \vdots \\ 541 \end{bmatrix}; \quad b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}; \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_{10} \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 176 \\ 1 & 177 \\ \vdots & \vdots \\ 1 & 184 \end{bmatrix}; \quad \text{Model: } y = X \cdot b + e$$

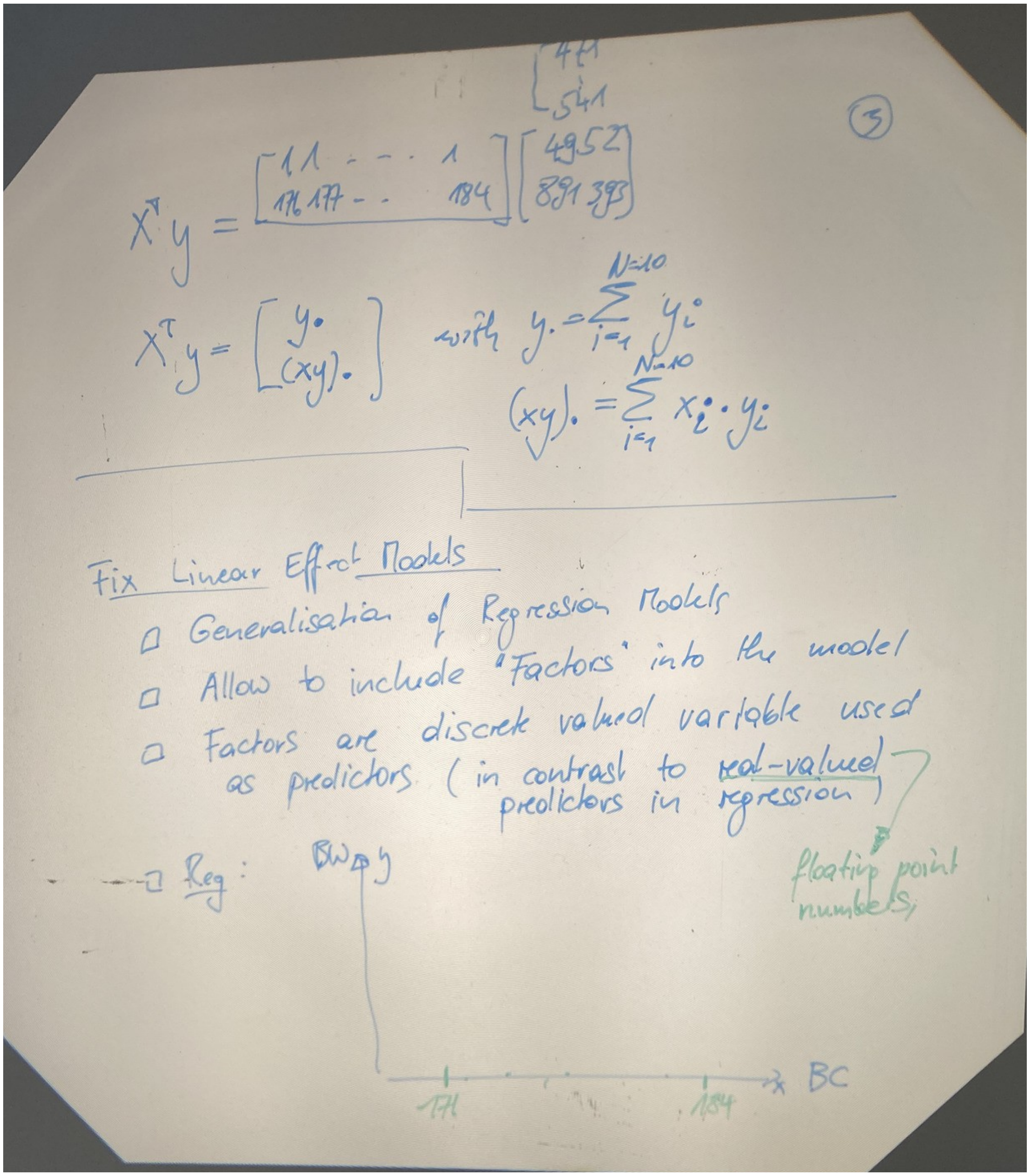
Least Squares Estimate $\hat{b}$ for b

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

OHP Picture 2



OHP Picture 3



3

$$\begin{bmatrix} 4\,\ldots \\ \vdots \\ 541 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 176 & 177 & \cdots & 184 \end{bmatrix} \begin{bmatrix} 4952 \\ 891\,393 \end{bmatrix}$$

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

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

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)

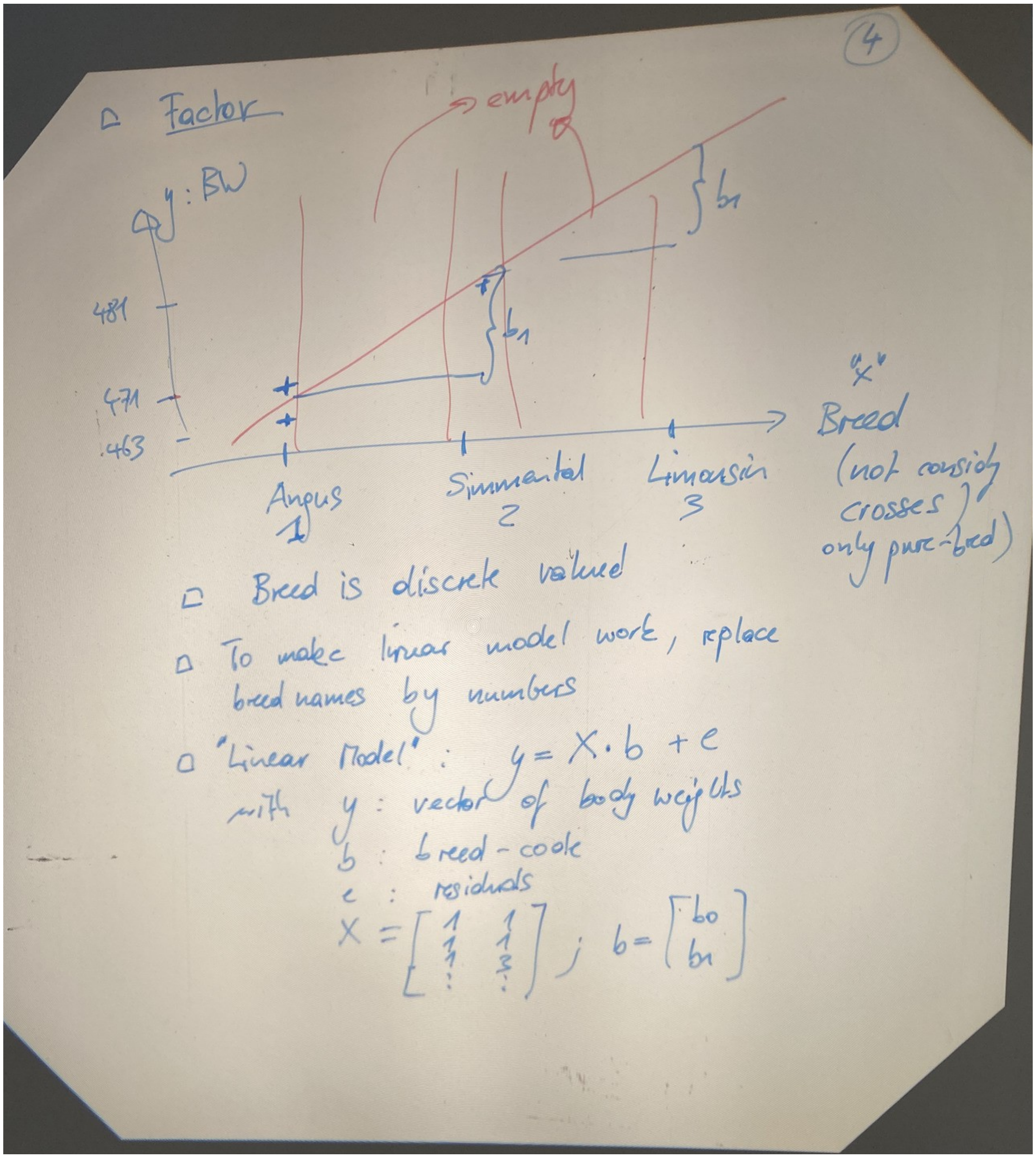
floating point numbers

- Reg: BW y

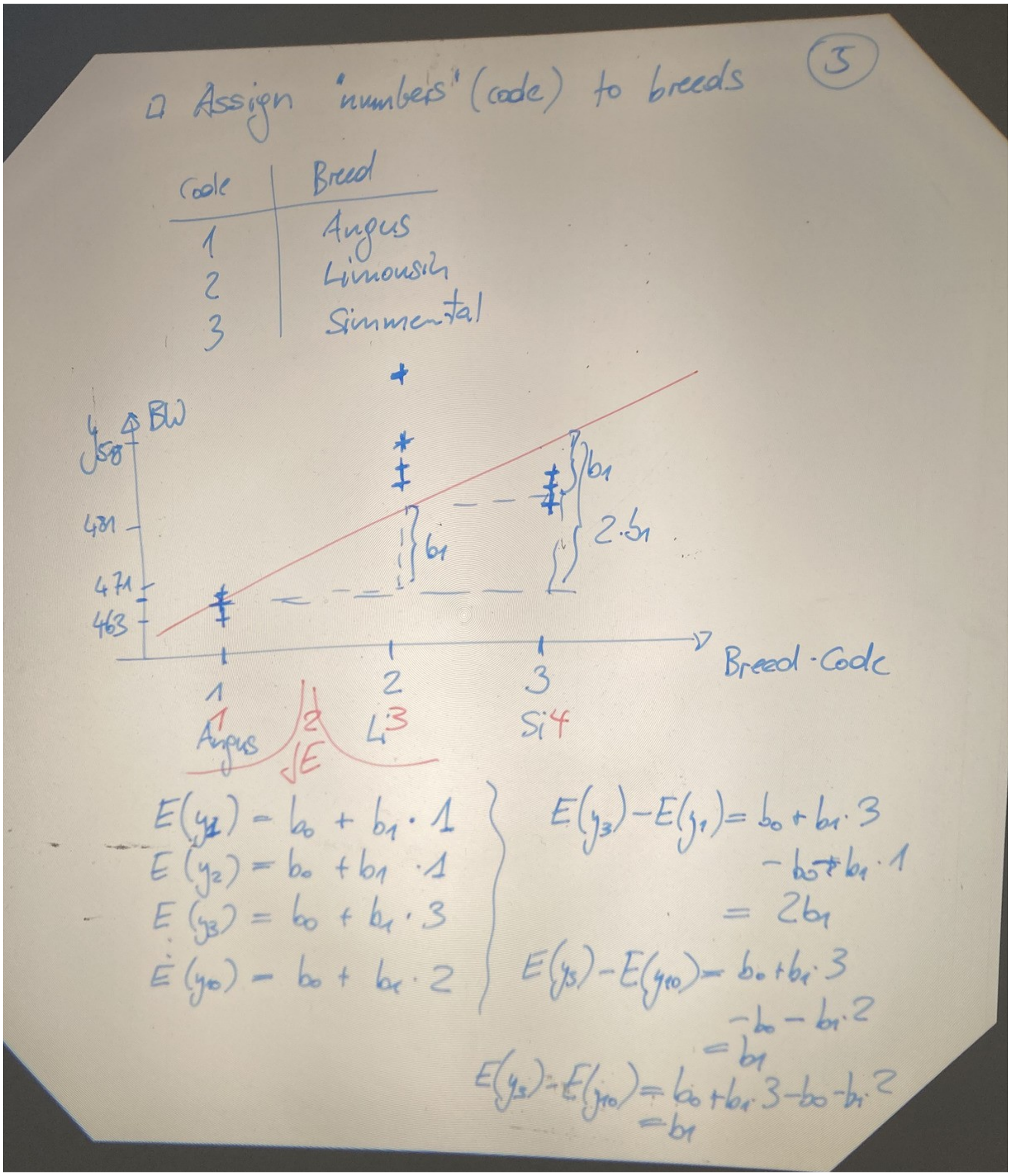
BC

171 184

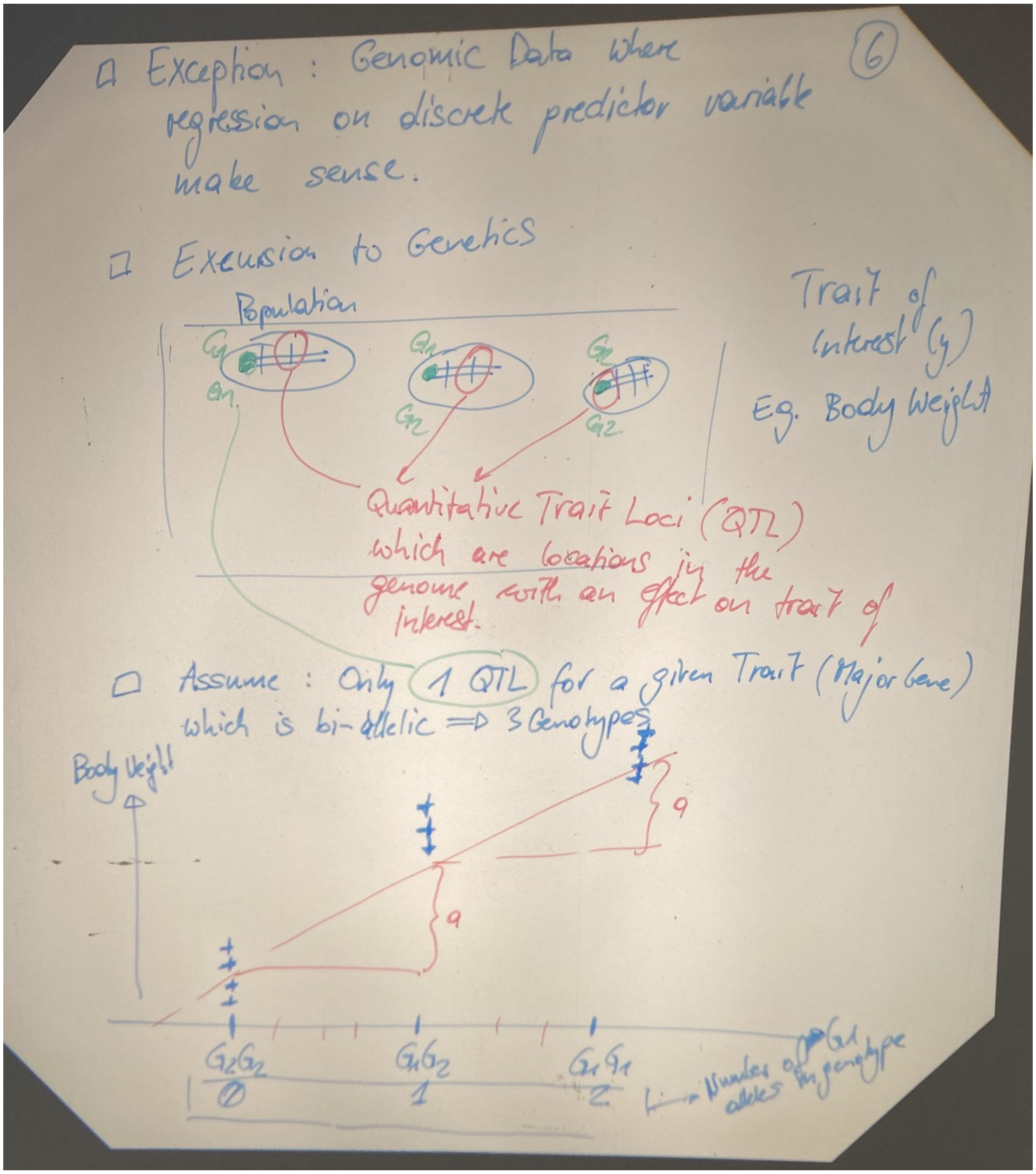
OHP Picture 4



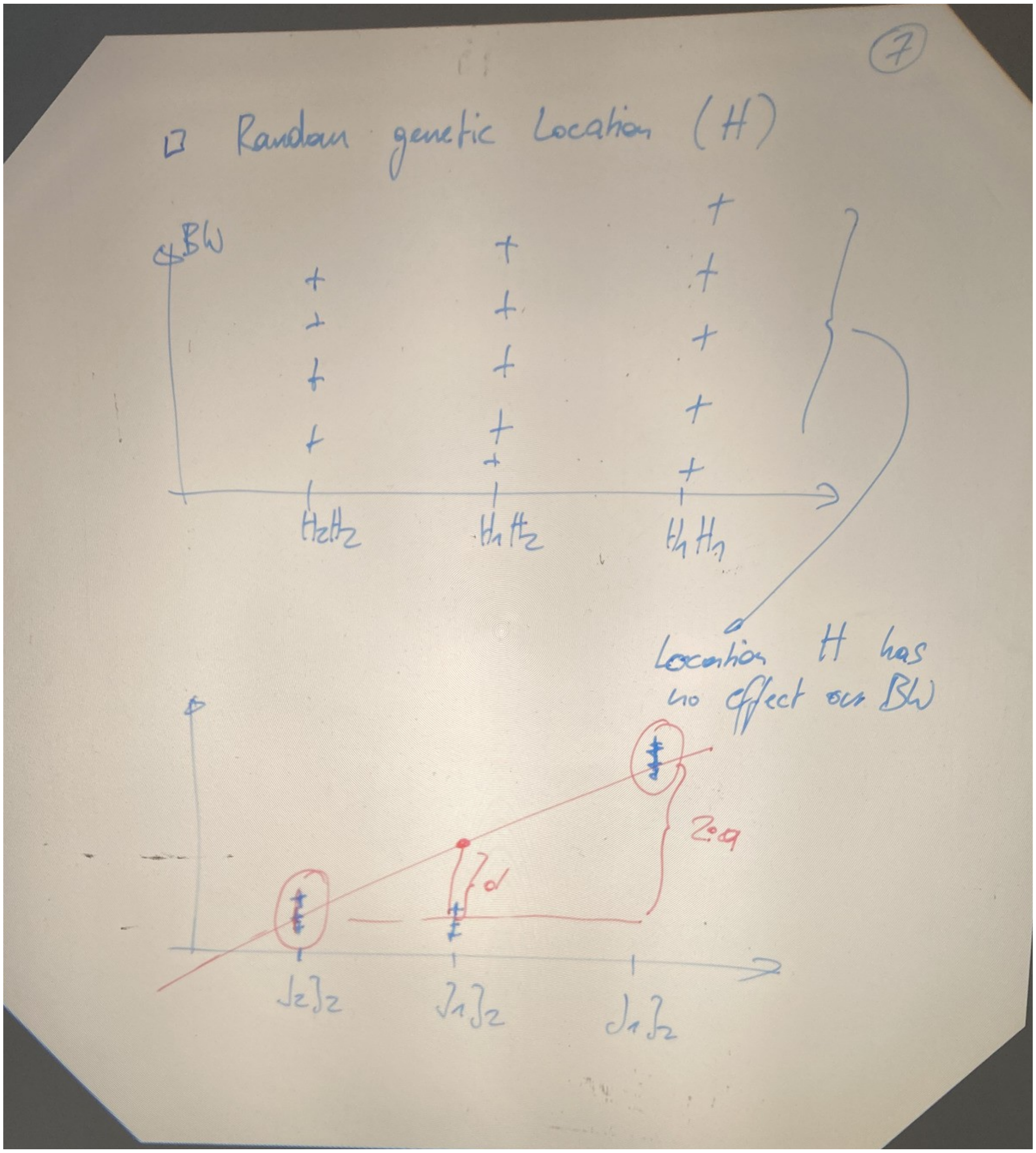
OHP Picture 5



OHP Picture 6



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

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