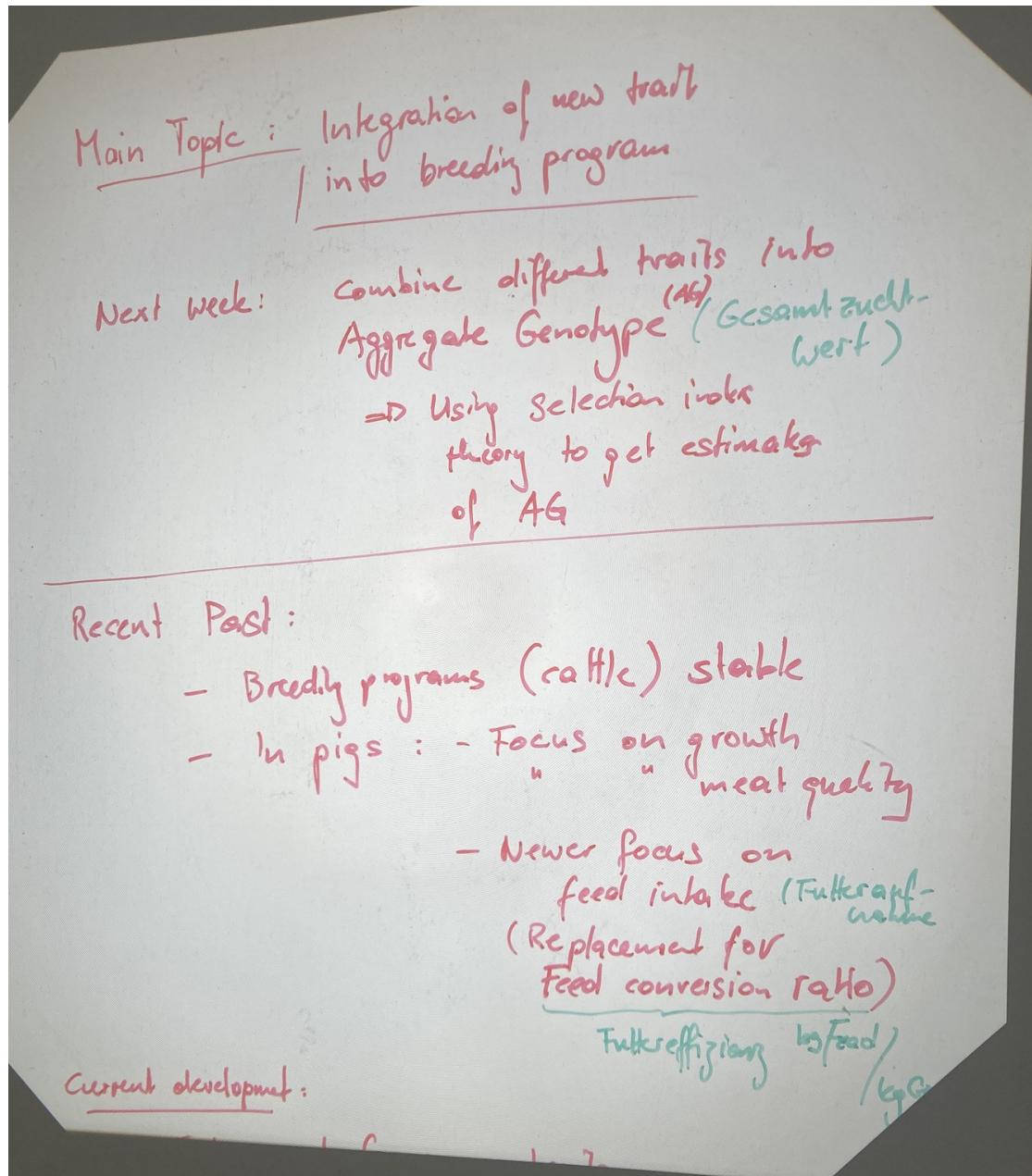
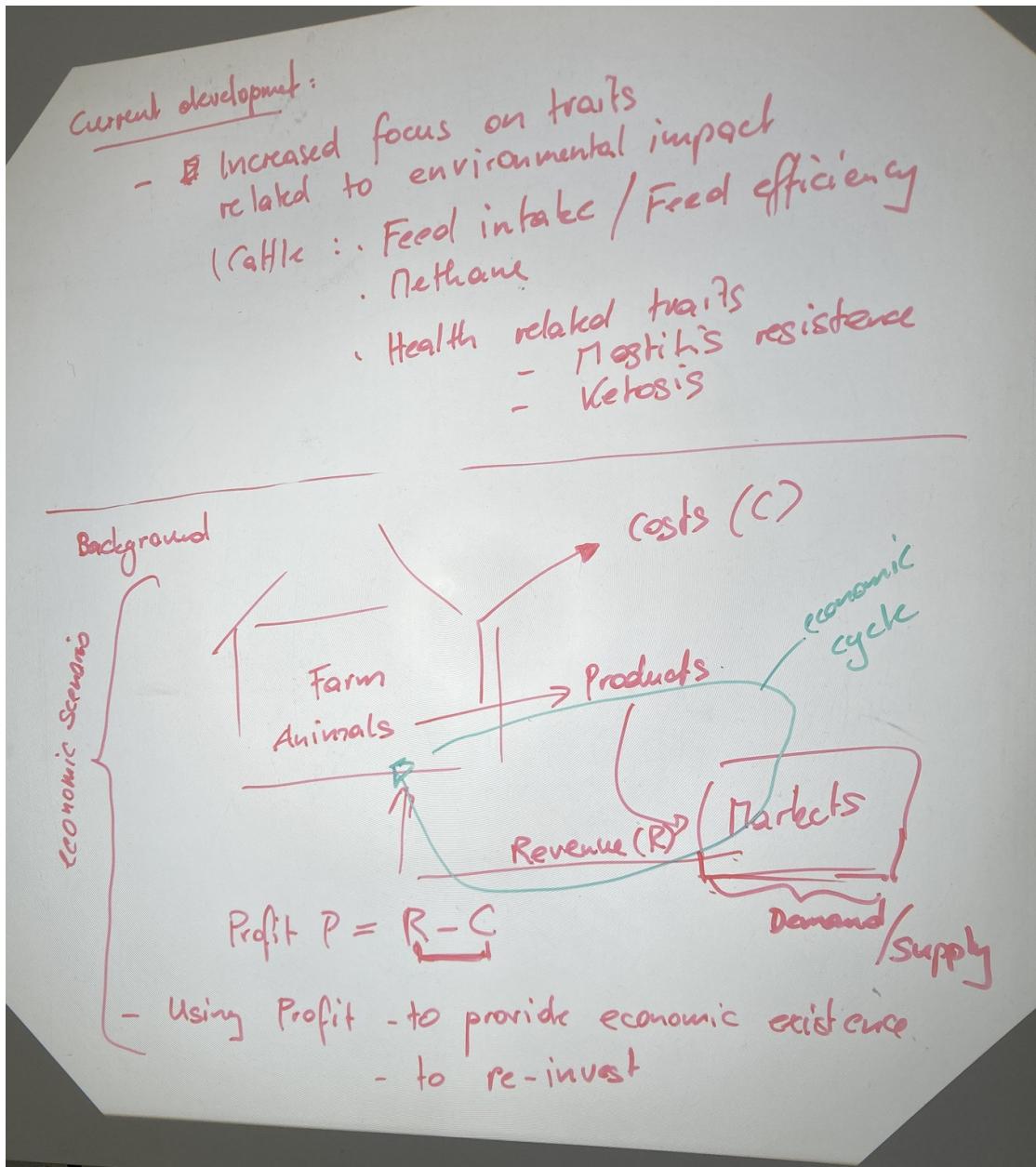


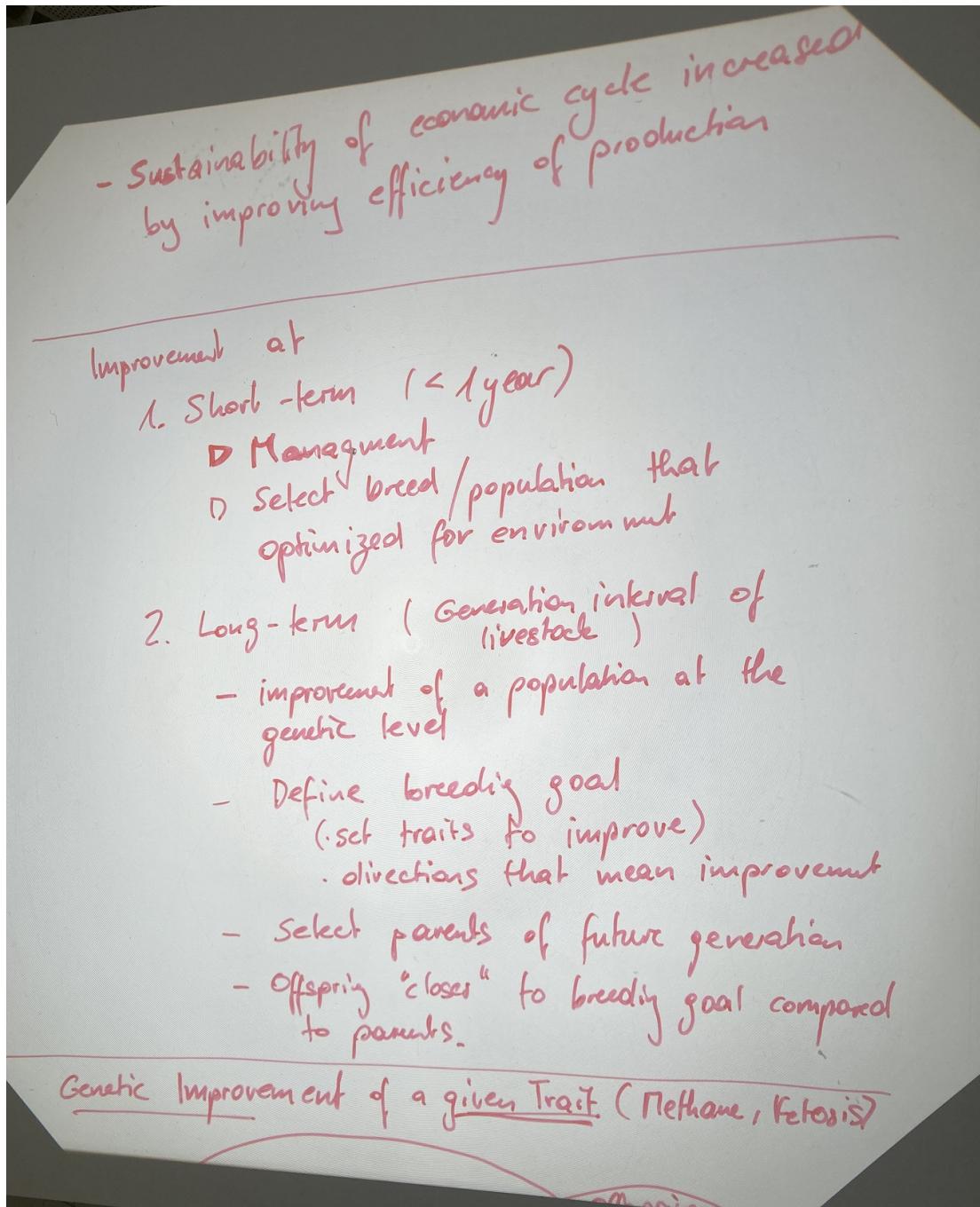
OHP Picture 1



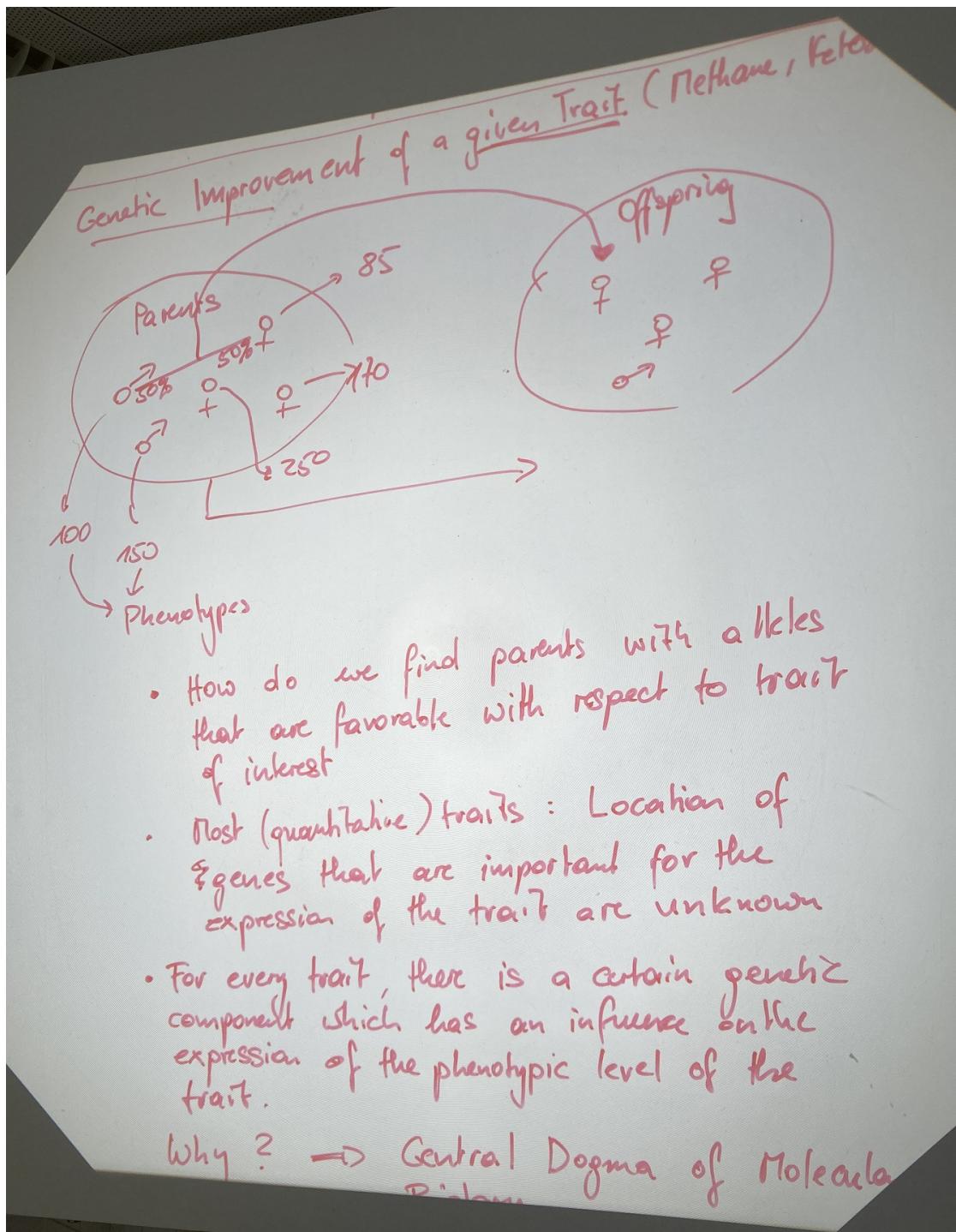
OHP Picture 2



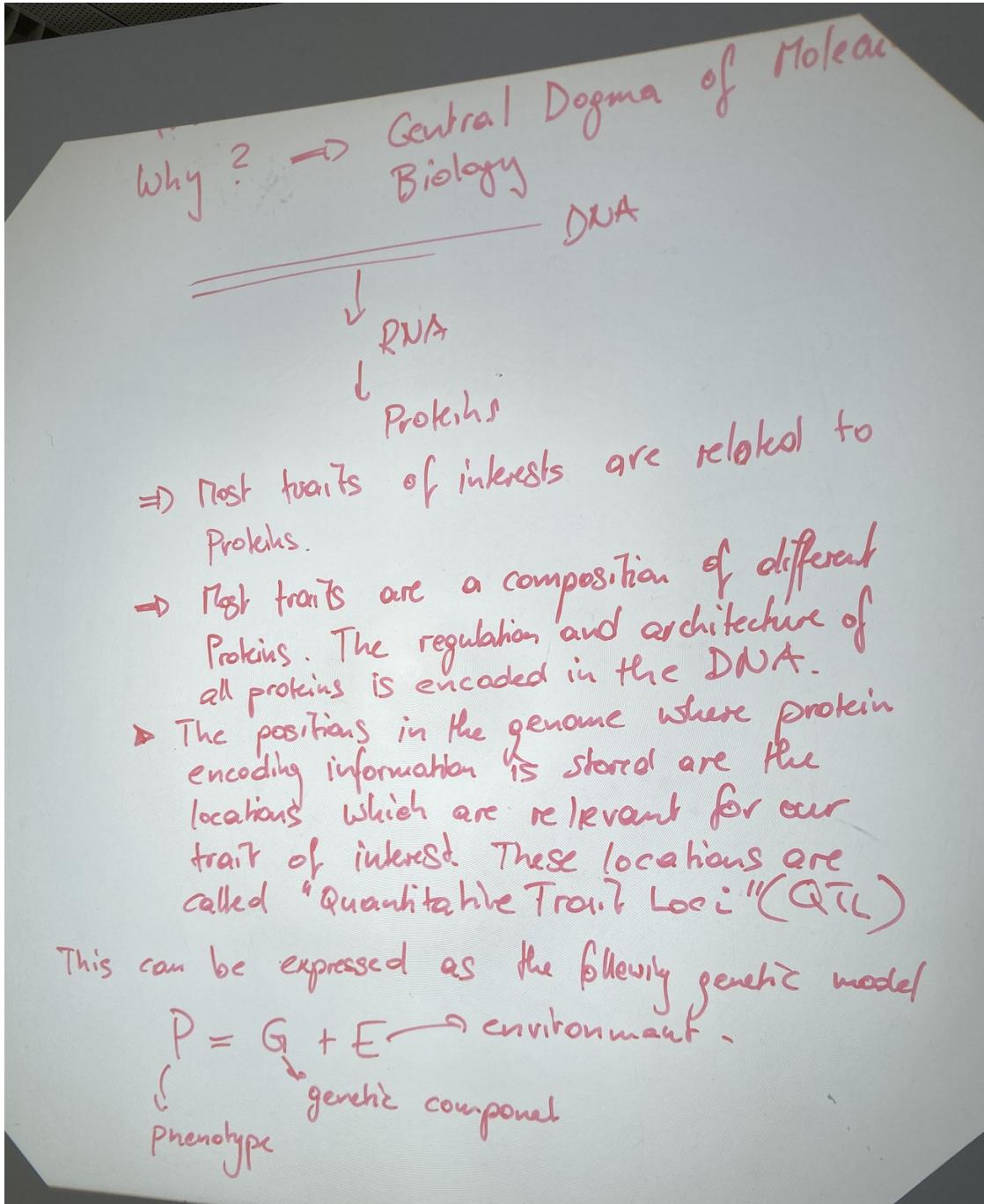
OHP Picture 3



OHP Picture 4



OHP Picture 5



OHP Picture 6

- Animals with best genetic component are identified using statistical models.
- In Statistical models :
Unknown genetic components are estimated given known phenotypic observations, known pedigree relationships and known genotypes for marker positions.
- In stochastic relationships, there is always a certain level of uncertainty, furthermore not all influence factors are known.
Hence we need statistical modelling to account for uncertainty and to find the most important influence factors.
- Statistical Model:
 - Response variable y : Phenotypic observations of trait of interest (Methane, Ketosis)
 - Predictor variables : x_1, x_2, \dots, x_k

OHP Picture 7

• Statistical Model:

- Response variable y : Phenotypic observations of trait of interest (Methane, ketosis)
- Predictor variables: x_1, x_2, \dots, x_k
- Error / Residual: e
- function $m(x_1, x_2, \dots, x_k)$ with $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$
 $\Leftrightarrow m(\underline{x})$
- Function $m(\underline{x})$ is used to relate the predictor (\underline{x}) to the response y
- The simplest class of relationships between \underline{x} and y is a linear relationship, for animali
 $y_i = m(x_i) + e_i$
- Example Data set

Animal	Body Weight	Breast Circumference
1	471	176

OHP Picture 8

□ Example Dataset

Animal	Body Weight	Breast Circumference
1	471	176
2	:	:
:	:	:
10	541	184

Response Predictor

□ Specify $m(x)$ for the example dataset:

For animal i :

$$y_i = b \cdot x_i + c_i$$

↓ Known ↓ Unknown ↓ Known
 Known Unknown Estimate from data

defines the slope of regression line

□ Data used as input:

$$y_1 = b \cdot x_1 + e_1 \Leftrightarrow 471 = b \cdot 176 + e_1$$

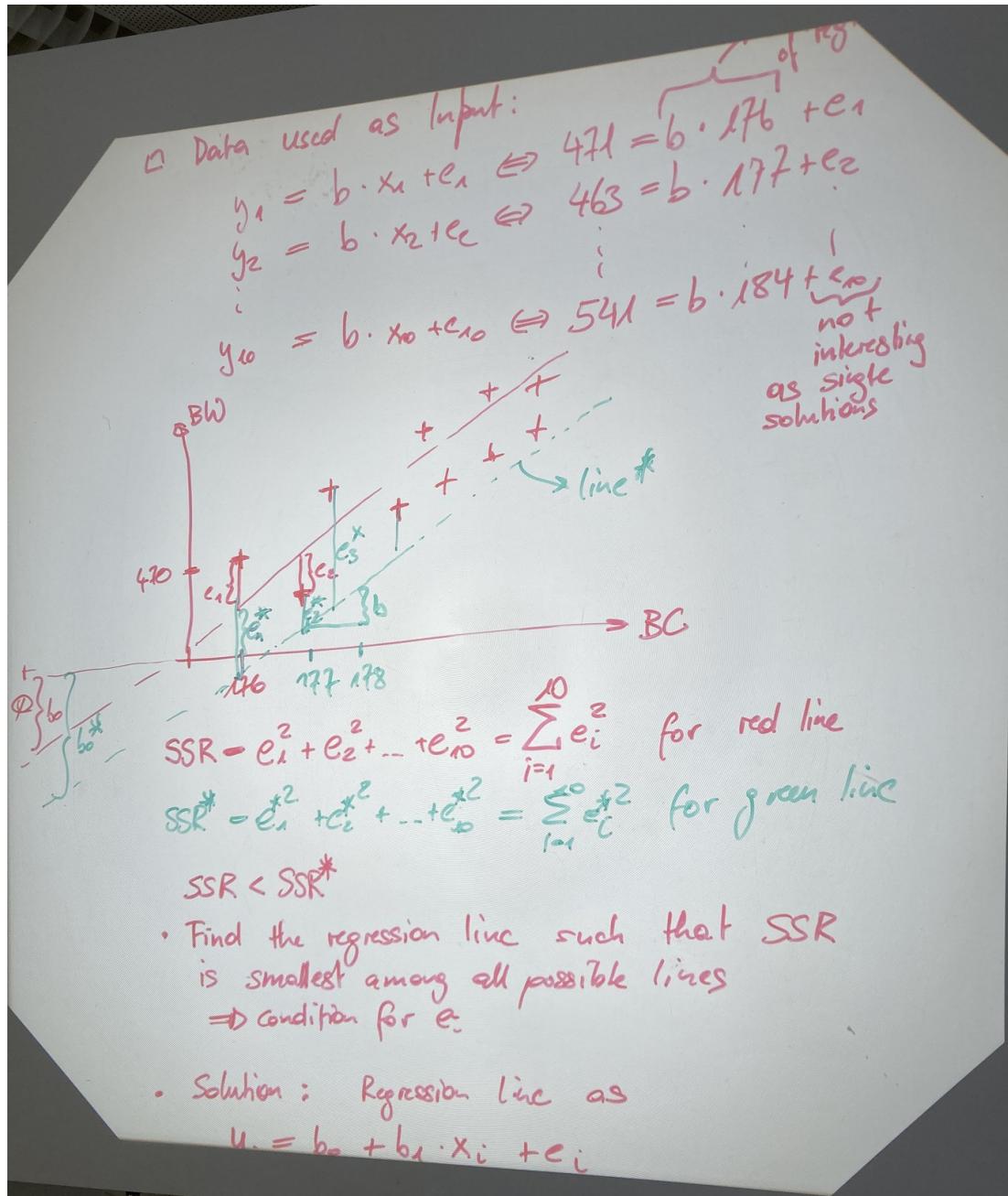
$$y_2 = b \cdot x_2 + e_2 \Leftrightarrow 463 = b \cdot 177 + e_2$$

$$\vdots$$

$$y_{10} = b \cdot x_{10} + e_{10} \Leftrightarrow 541 = b \cdot 184 + e_{10}$$

↑ BW + + F + + - -
 not interesting as size - height

OHP Picture 9



OHP Picture 10

. Solution: Regression line as
 $y_i = b_0 + b_1 \cdot x_i + e_i$
 For vectors $\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix}$ and matrix \underline{X}
 where $\underline{X} = \begin{bmatrix} 1 & 176 \\ 1 & 177 \\ 1 & 184 \end{bmatrix}$ and the vector \underline{b}
 with $\underline{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$, then applying the
 least-squares condition gives
 $\hat{\underline{b}} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}$

• In R: Function `lm()`
 $\text{lm}(\text{Body Weight} \sim \text{Breast Circumference, data=df})$

In Real dataset:

- The number of predictors can be large
- In ChatGPT: 10^{10}
- In such datasets, there might be

OHP Picture 11

In Real dataset:

- The number of predictors can be large
 - In Chat GPT: 10^{10}
- In Livestock datasets, there might be 20 - 100 predictors, for environmental factors
- Genomic datasets: Genotypes at 150K positions.

Animal	Body Weight	SNP ₁	SNP ₂	...	SNP ₁₅₀₀₀₀	---
1						
:						
N						

Linear Model:

$$y_i = b_0 + b_{SNP_1} \cdot x_{i,1} + b_{SNP_2} \cdot x_{i,2} + \dots + b_{SNP_{150000}} \cdot x_{i,150000} + e_i$$

$$\underline{b} = \begin{bmatrix} b_0 \\ b_{SNP_1} \\ b_{SNP_2} \\ \vdots \\ b_{SNP_{100}} \\ b_{SNP_{150000}} \\ b_{BC} \end{bmatrix}$$

Least Squares Solution
can no longer be used
because $(X^T X)^{-1}$ cannot
be computed.

OHP Picture 12

Linear Model:

$$\bar{y}_i = \bar{b}_0 + \bar{b}_{SNP_1} \cdot \bar{x}_{i,1} + \bar{b}_{SNP_2} \cdot \bar{x}_{i,2} + \dots + \bar{b}_{SNP_{15000}} \cdot \bar{x}_{i,15000} + \epsilon_i$$

$\underline{b} = \begin{bmatrix} b_0 \\ b_{SNP_1} \\ b_{SNP_2} \\ b_{SNP_{15000}} \\ \vdots \\ b_{SNP_{15000}} \end{bmatrix}$

Important

Least Squares Solution can no longer be used because $(X^T X)^{-1}$ cannot be computed.

In R: `lm(--)` \Rightarrow Error

Solution:

1. Separate meaningful predictors from predictors which are not important
 \Rightarrow Model selection.