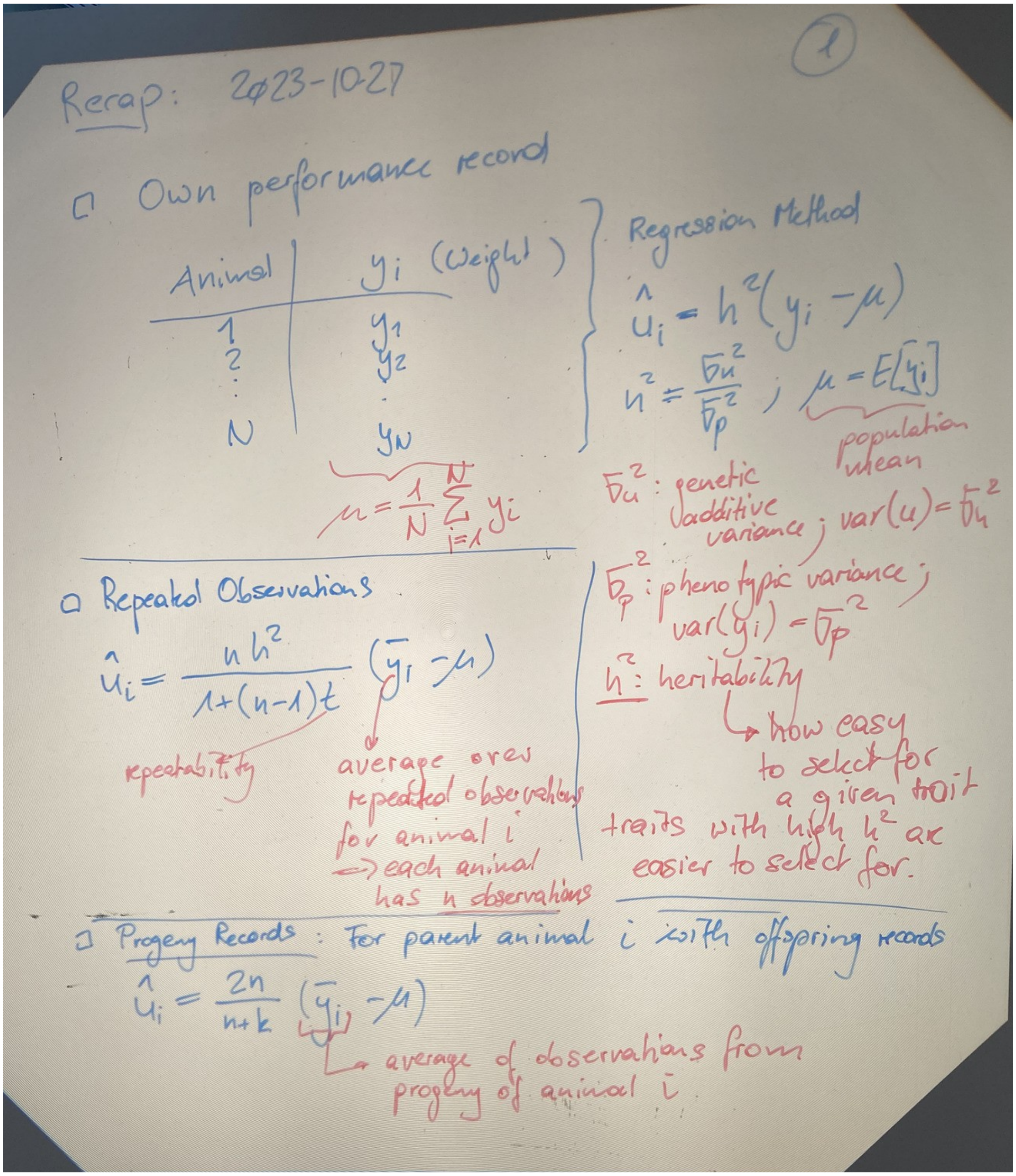
OHP Picture 1



①

Recap: 2023-10-27

☐ Own performance record

Animal	y_i (Weight)
1	y_1
2	y_2
$\vdots$	$\cdot$
N	y_N

$$\underbrace{\mu}_{} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Regression Method

$$\hat{u}_i = h^2(y_i - \mu)$$

$$h^2 \doteq \frac{\sigma_u^2}{\sigma_p^2}\ ;\ \underbrace{\mu = E[y_i]}_{\text{population mean}}$$

σ_u^2: genetic additive variance ; $var(u) = \sigma_u^2$

σ_p^2: phenotypic variance ; $var(y_i) = \sigma_p^2$

h^2: heritability

↳ how easy to select for a given trait

traits with high h^2 are easier to select for.

☐ Repeated Observations

$$\hat{u}_i = \frac{n h^2}{1+(n-1)t}(\bar{y}_i - \mu)$$

repeatability

average over repeated observations for animal i

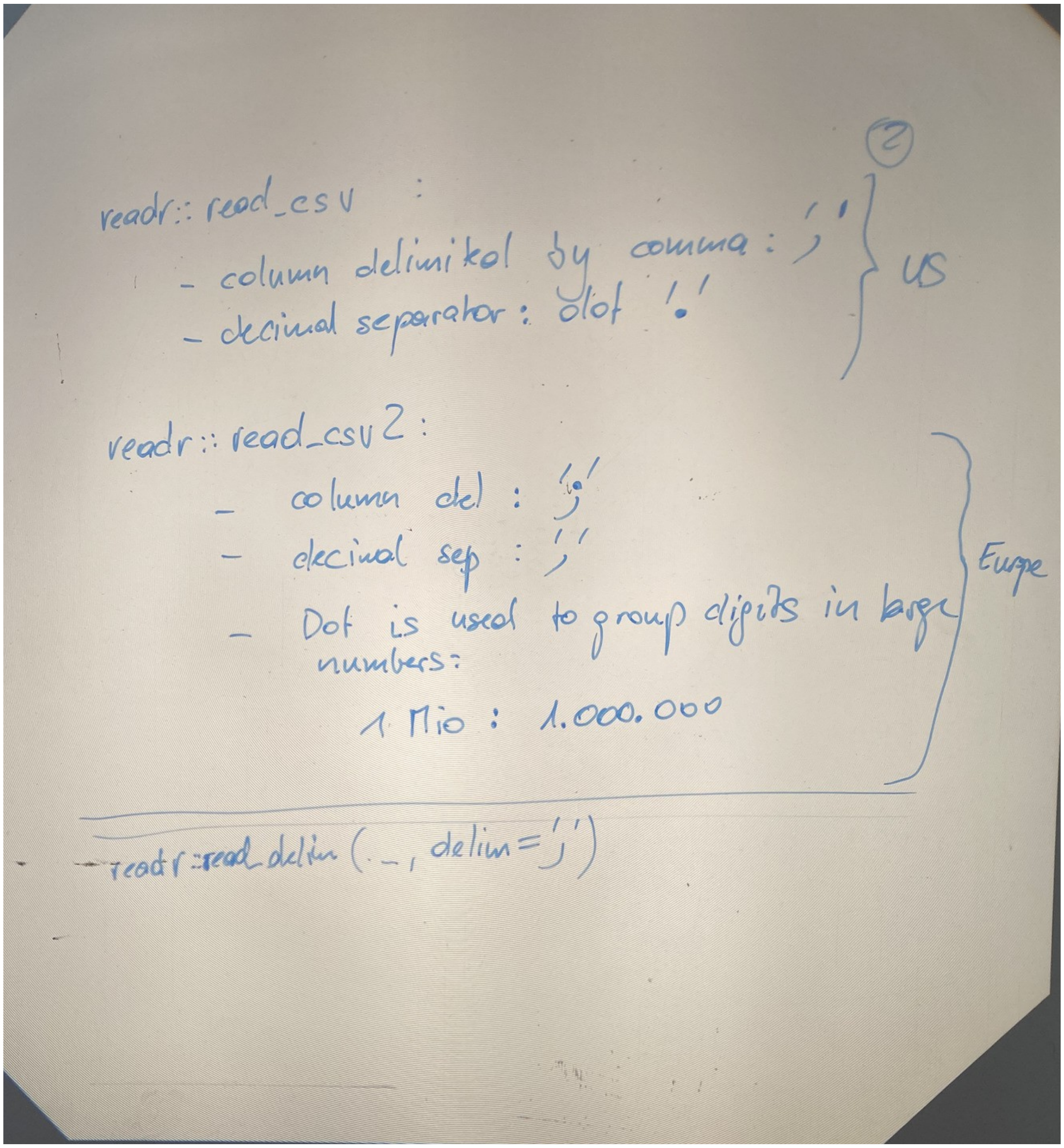
⟹ each animal has n observations

☐ Progeny Records : For parent animal i with offspring records

$$\hat{u}_i = \frac{2n}{n+k}(\bar{y}_i - \mu)$$

↳ average of observations from progeny of animal i

OHP Picture 2



②

readr::read_csv :

- column delimited by comma: ','
- decimal separator: dot '.'

US

readr::read_csv2 :

- column del : ';'
- decimal sep : ','
- Dot is used to group digits in large numbers:

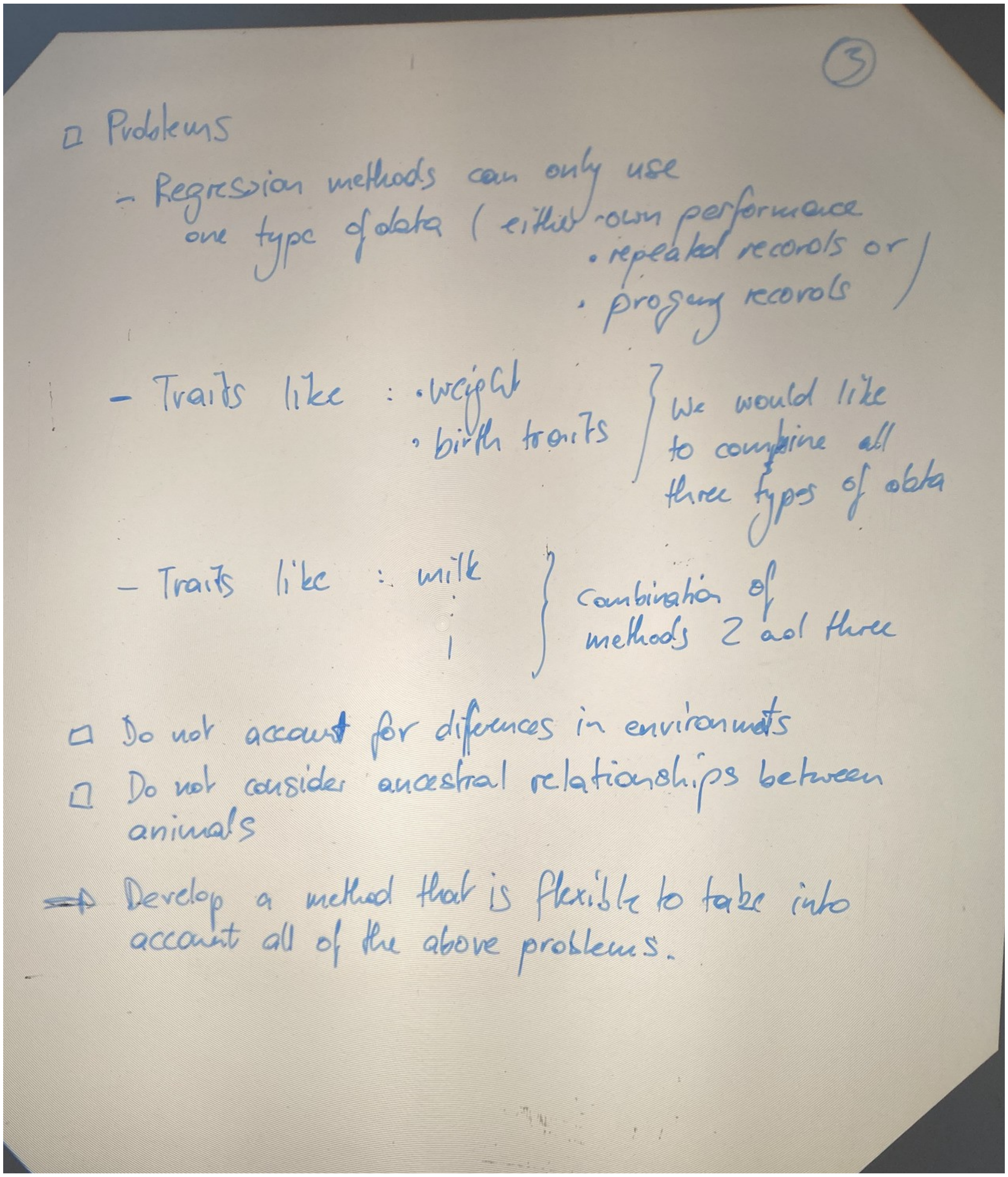
  1 Mio : 1.000.000

Europe

---

readr::read_delim(._, delim=';')

OHP Picture 3



③

☐ Problems

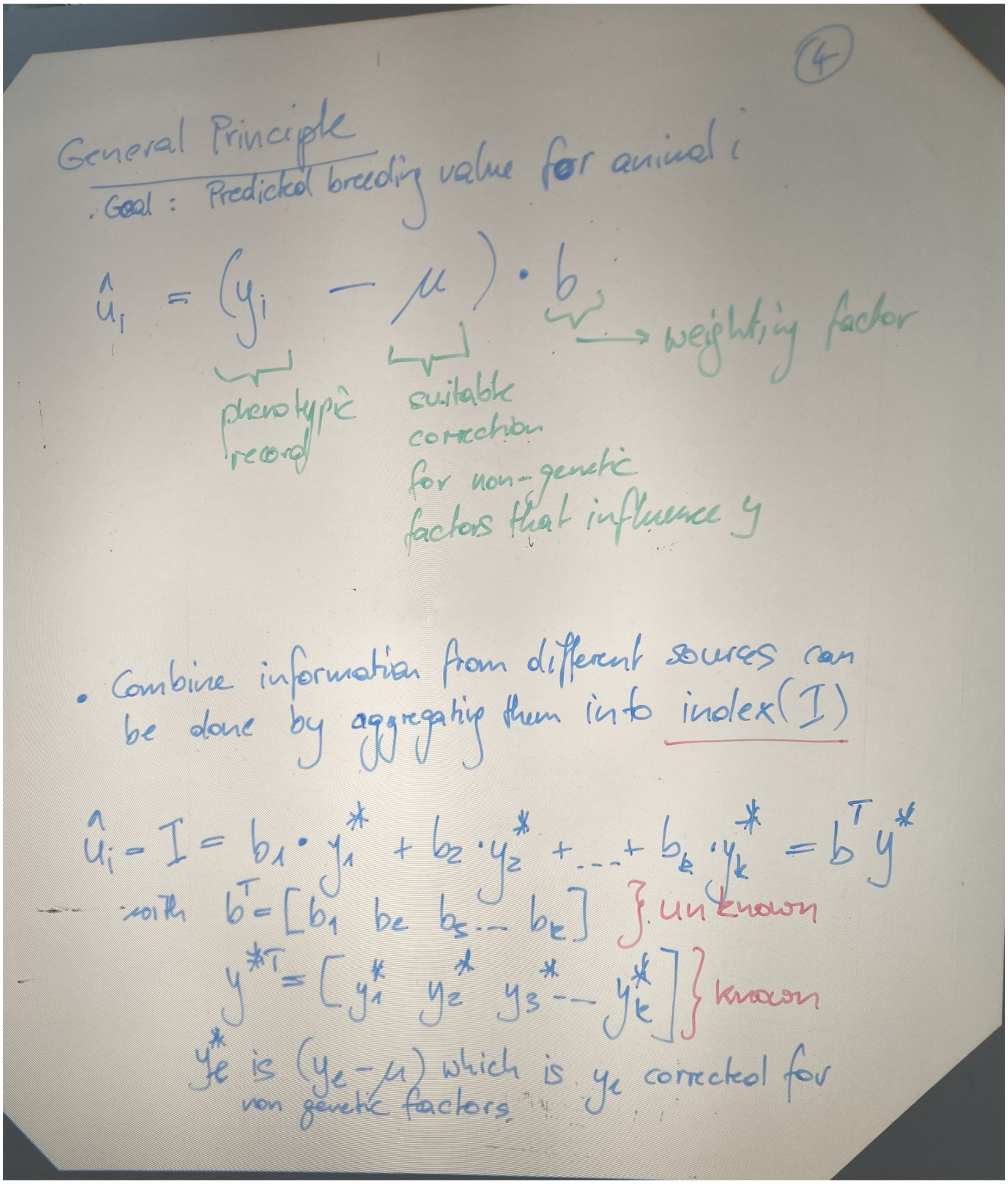
- Regression methods can only use one type of data (either own performance
  - repeated records or
  - progeny records)
- Traits like : • weight
  • birth traits } We would like to combine all three types of data
- Traits like : milk
  ⋮ } Combination of methods 2 and three

☐ Do not account for diferences in environments

☐ Do not consider ancestral relationships between animals

⇒ Develop a method that is flexible to take into account all of the above problems.

OHP Picture 4



④

General Principle

. Goal : Predicted breeding value for animal i

$$\hat{u}_i = (\underbrace{y_i}_{\text{phenotypic record}} - \underbrace{\mu}_{\text{suitable correction for non-genetic factors that influence } y}) \cdot \underbrace{b}_{} \longrightarrow \text{weighting factor}$$

- Combine information from different sources can be done by aggregating them into index (I)

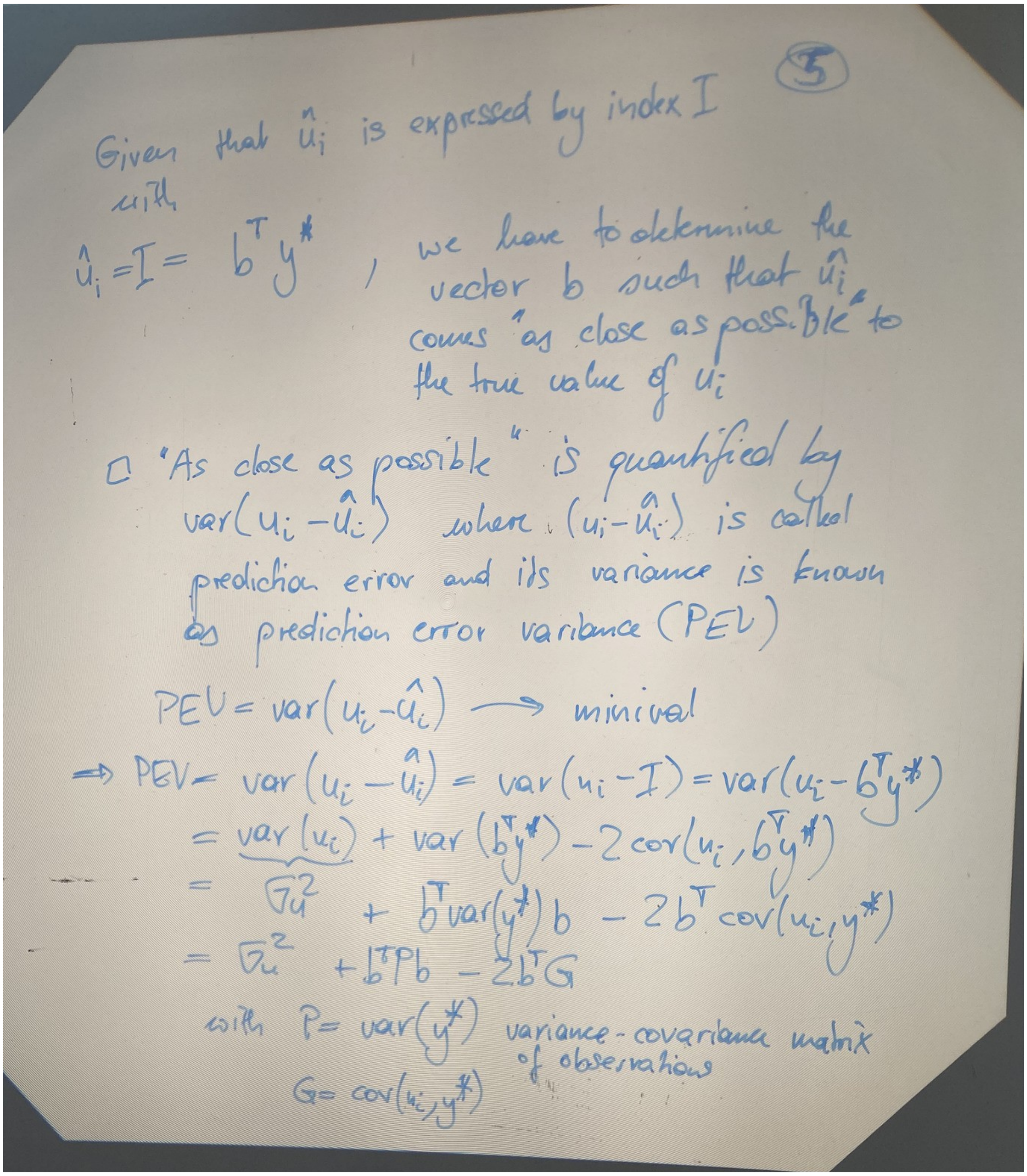
$$\hat{u}_i = I = b_1 \cdot y_1^* + b_2 \cdot y_2^* + \dots + b_k \cdot y_k^* = b^T y^*$$

with $b^T = [b_1 \ b_2 \ b_3 \dots b_k]$ } unknown

$y^{*T} = [y_1^* \ y_2^* \ y_3^* \dots y_k^*]$ } known

y_k^* is $(y_k - \mu)$ which is y_k corrected for non genetic factors

OHP Picture 5



⑤

Given that $\hat{u}_i$ is expressed by index I with

$\hat{u}_i = I = b^T y^*$, we have to determine the vector b such that $\hat{u}_i$ comes "as close as possible" to the true value of u_i

□ "As close as possible" is quantified by $var(u_i - \hat{u}_i)$ where $(u_i - \hat{u}_i)$ is called prediction error and its variance is known as prediction error variance (PEV)

$$PEV = var(u_i - \hat{u}_i) \longrightarrow \text{minimal}$$

$$\Rightarrow PEV = var(u_i - \hat{u}_i) = var(u_i - I) = var(u_i - b^T y^*)$$

$$= \underbrace{var(u_i)} + var(b^T y^*) - 2\, cor(u_i, b^T y^*)$$

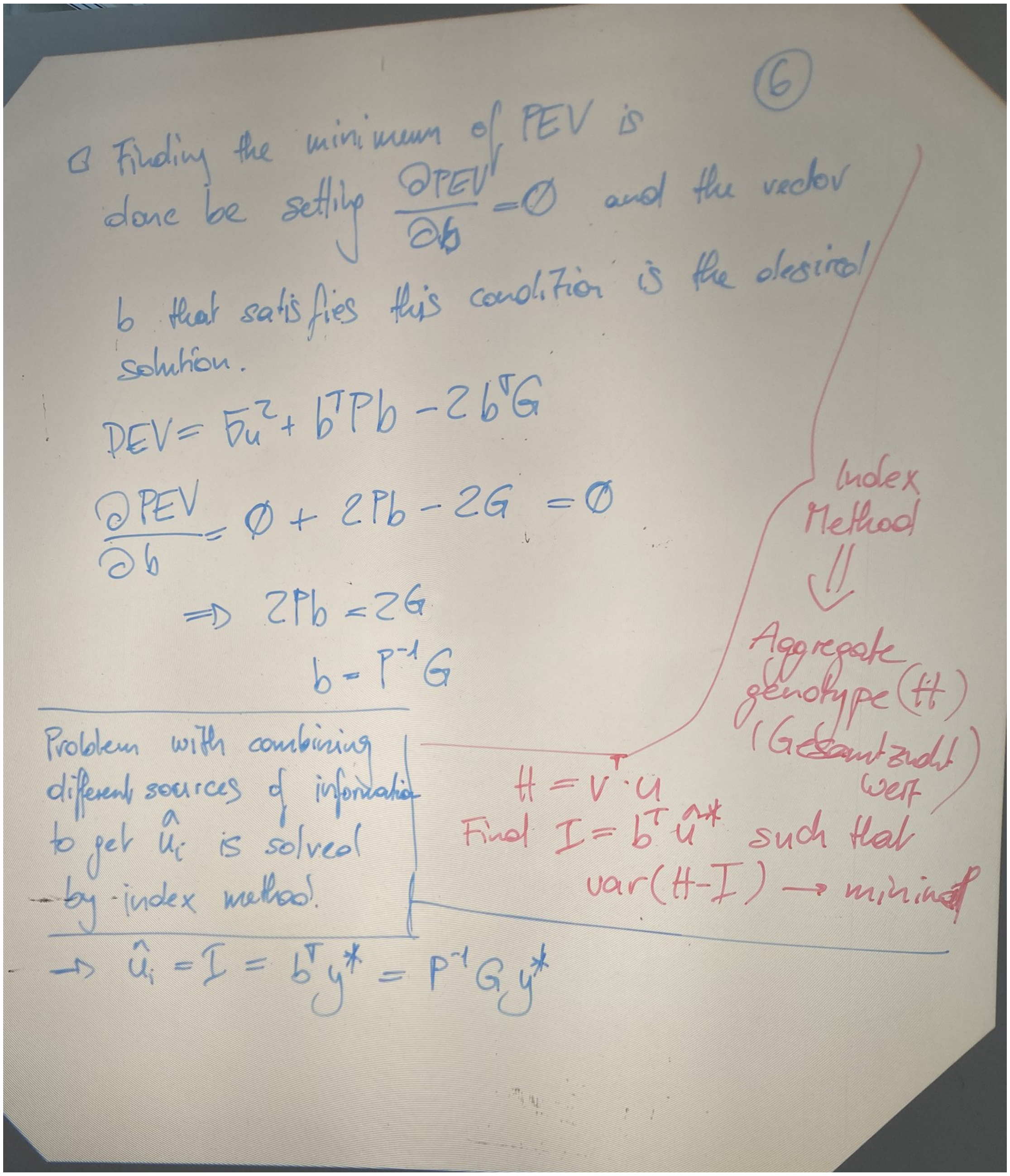
$$= \sigma_u^2 + b^T var(y^*)\, b - 2 b^T cov(u_i, y^*)$$

$$= \sigma_u^2 + b^T P b - 2 b^T G$$

with $P = var(y^*)$ variance-covariance matrix of observations

$G = cov(u_i, y^*)$

OHP Picture 6



⑥

- Finding the minimum of PEV is done be setting $\frac{\partial PEV}{\partial b} = 0$ and the vector b that satisfies this condition is the desired solution.

$$PEV = \sigma_u^2 + b^T P b - 2 b^T G$$

$$\frac{\partial PEV}{\partial b} = 0 + 2Pb - 2G = 0$$

$$\Rightarrow 2Pb = 2G$$

$$b = P^{-1} G$$

Index Method

⇓

Aggregate genotype (H)
(Gesamtzuchtwert)

$$H = v^T \cdot u$$

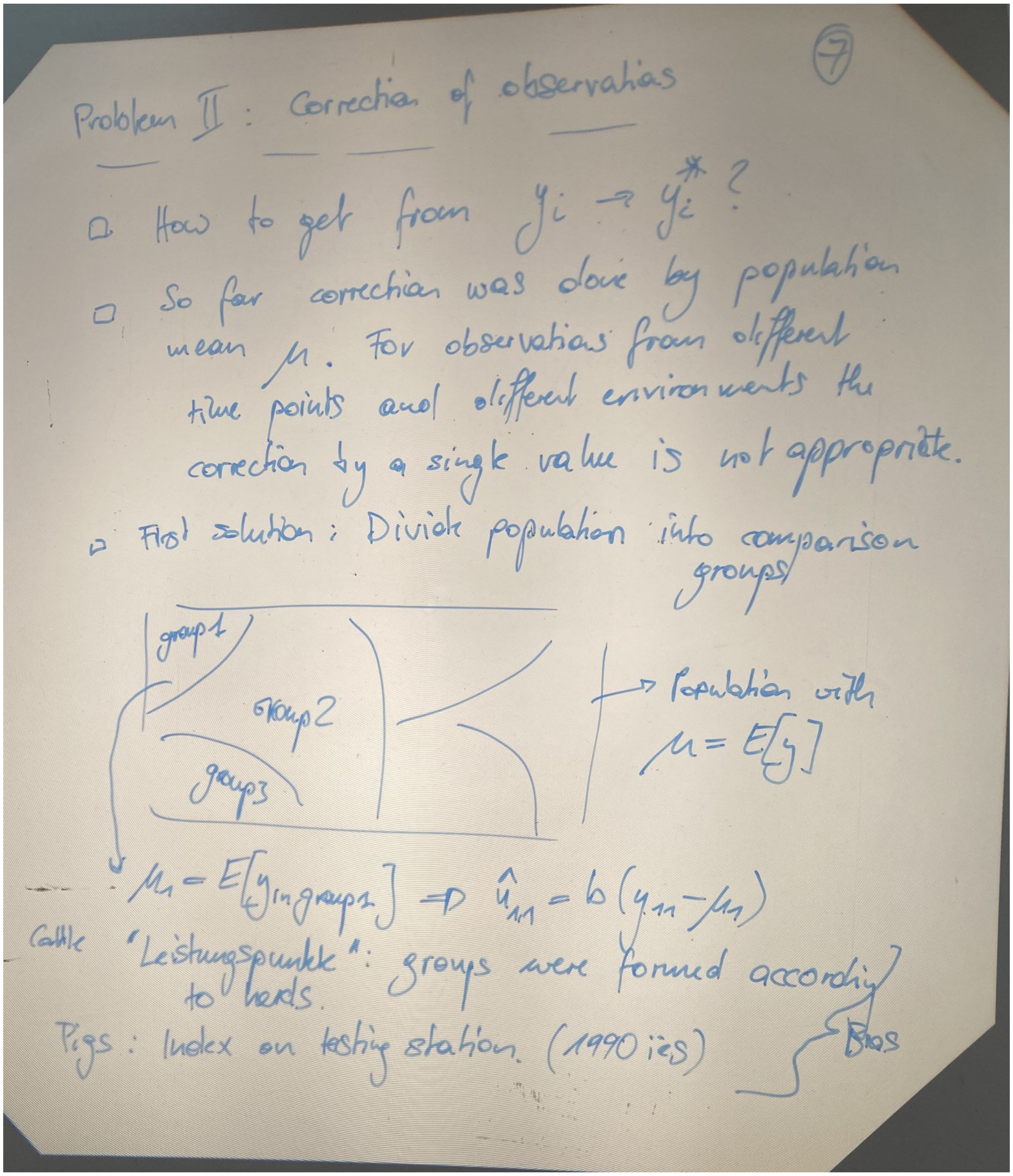
Find $I = b^T \hat{u}^*$ such that

var(H-I) → minimal

Problem with combining different sources of information to get $\hat{u}_i$ is solved by index method.

$$\Rightarrow \hat{u}_i = I = b^T y^* = P^{-1} G y^*$$

OHP Picture 7



⑦

Problem II: Correction of observations

- ☐ How to get from $y_i \rightarrow y_i^*$?
- ☐ So far correction was done by population mean μ. For observations from different time points and different environments the correction by a single value is not appropriate.
- ☐ First solution: Divide population into comparison groups

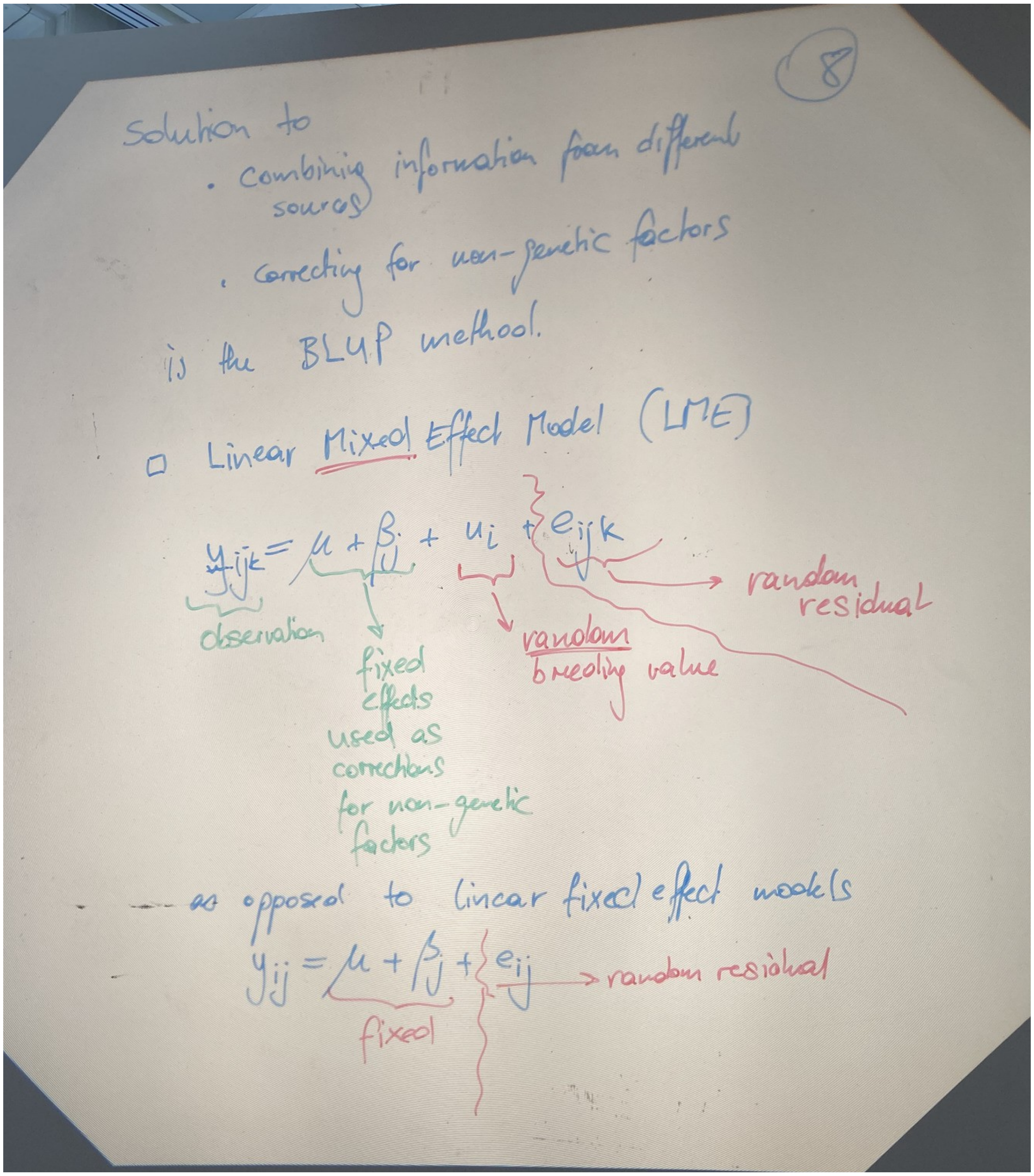
$\mu_1 = E[y \text{ in group } 1] \Rightarrow \hat{u}_{11} = b(y_{11} - \mu_1)$

Cattle "Leistungspunkte": groups were formed according to herds.

Pigs: Index on testing station (1990 ies)

Bias

OHP Picture 8



8

Solution to
- combining information from different sources
- correcting for non-genetic factors

is the BLUP method.

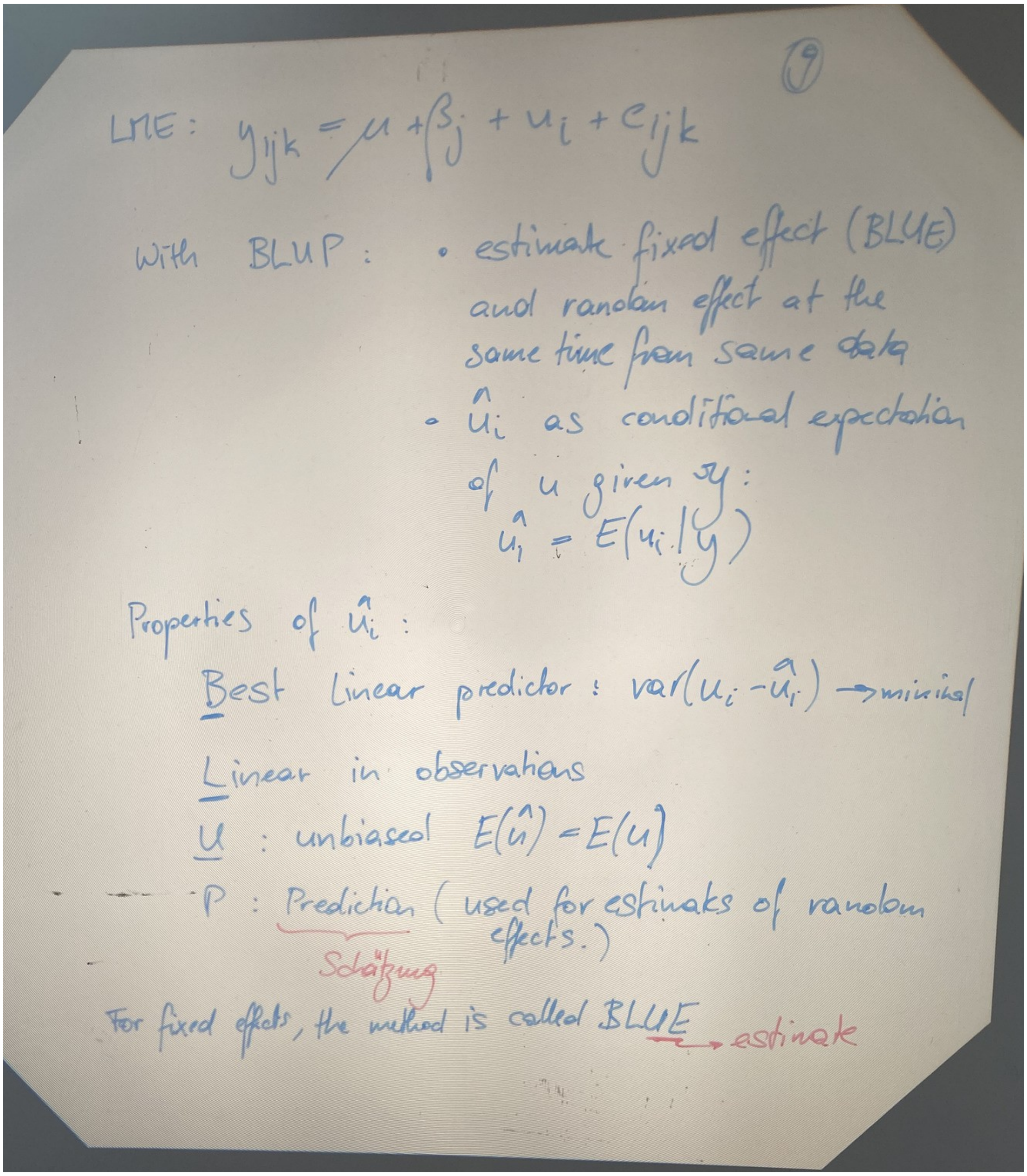
□ Linear Mixed Effect Model (LME)

$$\underbrace{y_{ijk}}_{\text{observation}} = \underbrace{\mu + \beta_j}_{\text{fixed effects used as corrections for non-genetic factors}} + \underbrace{u_i}_{\text{random breeding value}} + \underbrace{e_{ijk}}_{\text{random residual}}$$

-- as opposed to linear fixed effect models

$$y_{ij} = \underbrace{\mu + \beta_j}_{\text{fixed}} + e_{ij} \longrightarrow \text{random residual}$$

OHP Picture 9



⑨

LME: $y_{ijk} = \mu + \beta_j + u_i + e_{ijk}$

with BLUP:
- estimate fixed effect (BLUE) and random effect at the same time from same data
- $\hat{u}_i$ as conditional expectation of u given y: $\hat{u}_i = E(u_i | y)$

Properties of $\hat{u}_i$:

<u>B</u>est Linear predictor: $var(u_i - \hat{u}_i) \rightarrow$ minimal

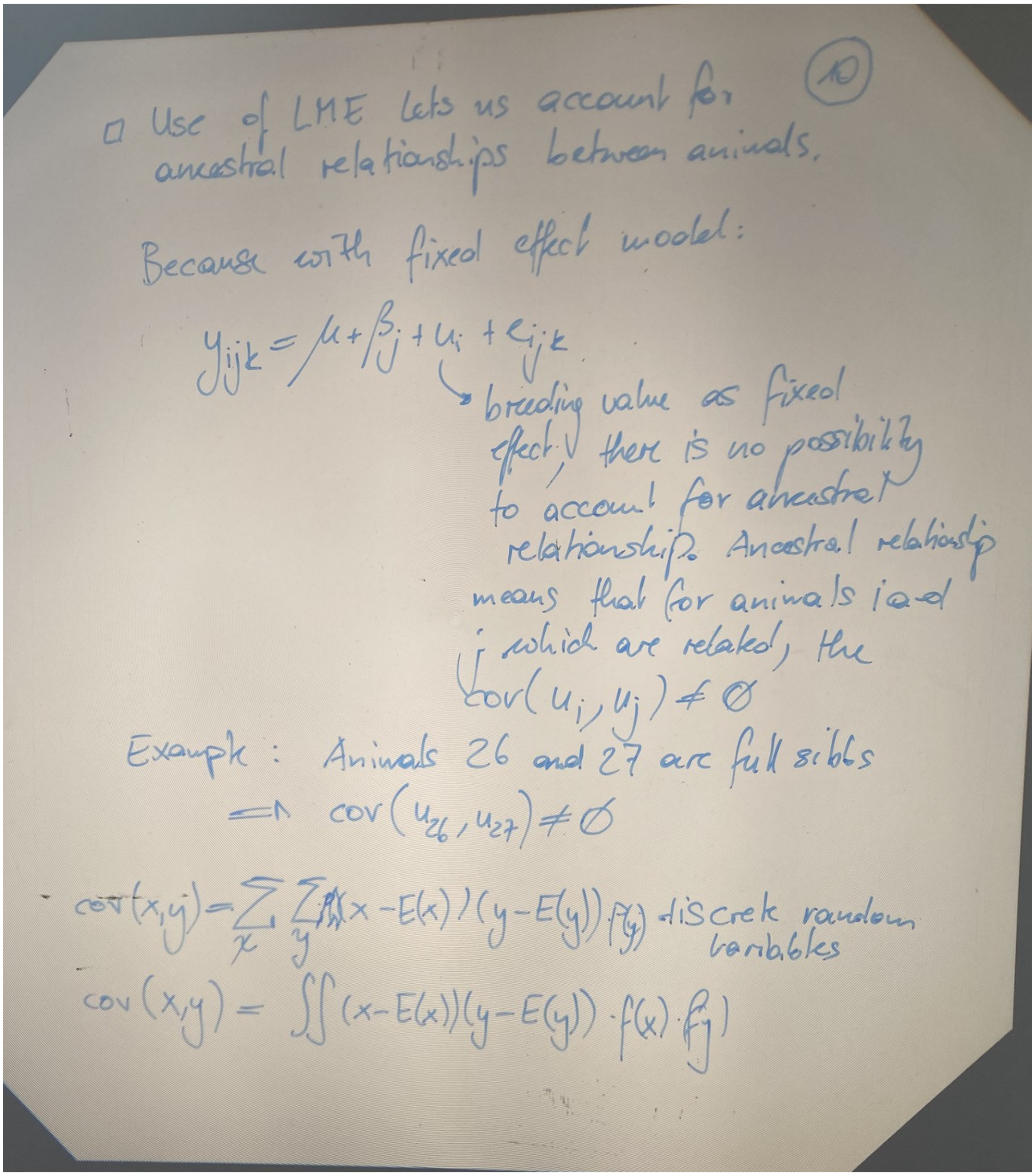
<u>L</u>inear in observations

<u>U</u>: unbiased $E(\hat{u}) = E(u)$

P: $\underbrace{\text{Prediction}}_{\text{Schätzung}}$ (used for estimates of random effects.)

For fixed effects, the method is called BLU$\underset{\rightarrow \text{ estimate}}{E}$

OHP Picture 10



⑩

☐ Use of LME lets us account for ancestral relationships between animals.

Because with fixed effect model:

$$y_{ijk} = \mu + \beta_j + u_i + e_{ijk}$$

→ breeding value as fixed effect, there is no possibility to account for ancestral relationship. Ancestral relationship means that for animals i and j which are related, the $cov(u_i, u_j) \neq 0$

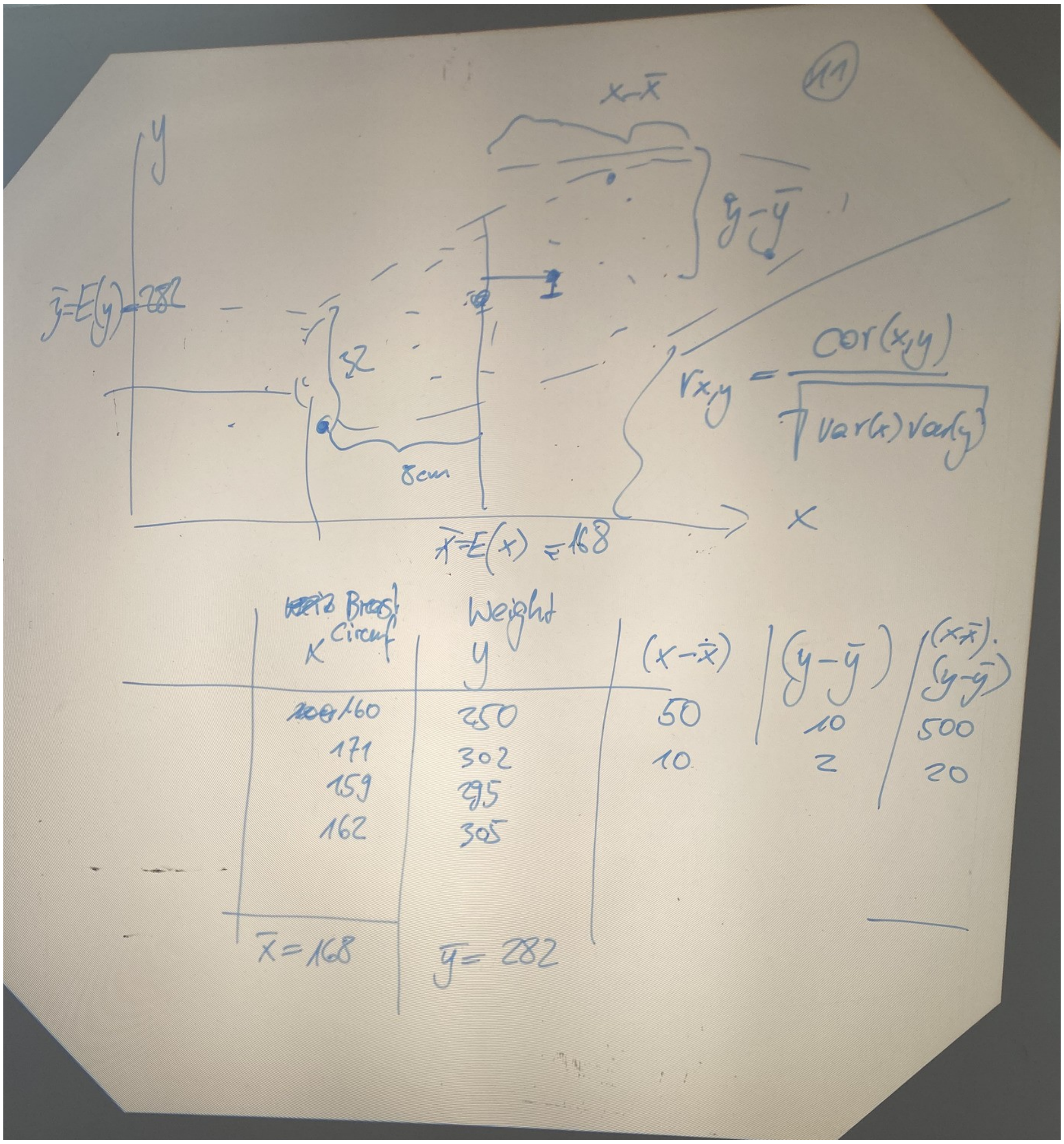
Example: Animals 26 and 27 are full sibbs

$\Rightarrow cov(u_{26}, u_{27}) \neq 0$

$cov(x,y) = \sum_x \sum_y (x - E(x))(y - E(y)) \cdot f(y)$ discrete random variables

$cov(x,y) = \iint (x - E(x))(y - E(y)) \cdot f(x) \cdot f(y)$

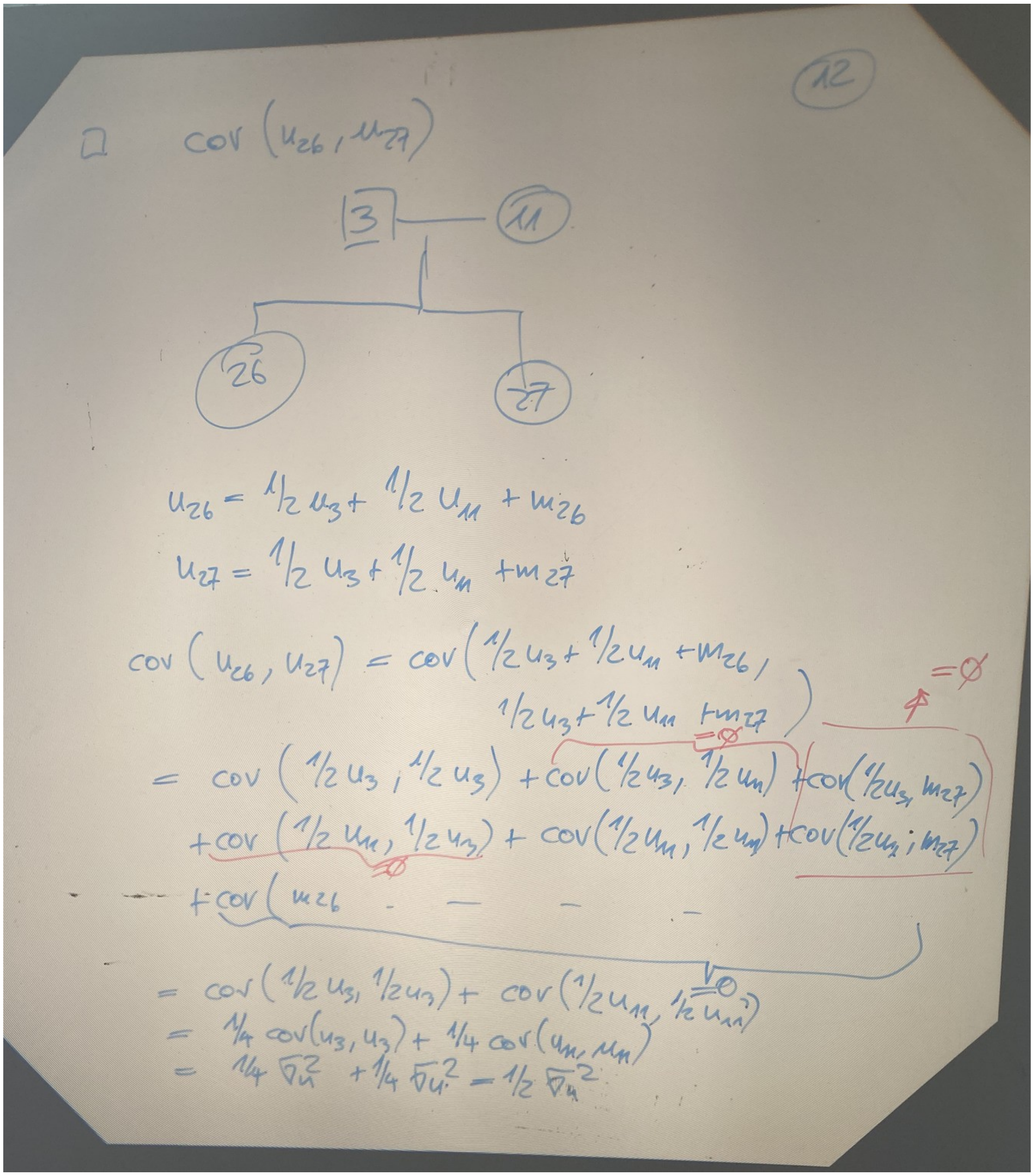
OHP Picture 11



$$r_{x,y} = \frac{cor(x,y)}{\sqrt{var(x)\,var(y)}}$$

~~[illegible]~~ Breast Circuf x	Weight y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x}) \cdot (y-\bar{y})$
~~200~~ 160	250	50	10	500
171	302	10	2	20
159	295			
162	305			
$\bar{x}=168$	$\bar{y}=282$			

OHP Picture 12



(12)

☐ $\mathrm{cov}(u_{26}, u_{27})$

$$u_{26} = \tfrac{1}{2} u_3 + \tfrac{1}{2} u_{11} + m_{26}$$

$$u_{27} = \tfrac{1}{2} u_3 + \tfrac{1}{2} u_{11} + m_{27}$$

$$\mathrm{cov}(u_{26}, u_{27}) = \mathrm{cov}\left(\tfrac{1}{2} u_3 + \tfrac{1}{2} u_{11} + m_{26}, \tfrac{1}{2} u_3 + \tfrac{1}{2} u_{11} + m_{27}\right)$$

$$= \mathrm{cov}\left(\tfrac{1}{2} u_3, \tfrac{1}{2} u_3\right) + \overbrace{\mathrm{cov}\left(\tfrac{1}{2} u_3, \tfrac{1}{2} u_{11}\right)}^{=0} + \boxed{\mathrm{cov}\left(\tfrac{1}{2} u_3, m_{27}\right)} \xrightarrow{} = 0$$

$$+ \underbrace{\mathrm{cov}\left(\tfrac{1}{2} u_{11}, \tfrac{1}{2} u_3\right)}_{0} + \mathrm{cov}\left(\tfrac{1}{2} u_{11}, \tfrac{1}{2} u_{11}\right) + \boxed{\mathrm{cov}\left(\tfrac{1}{2} u_{11}; m_{27}\right)}$$

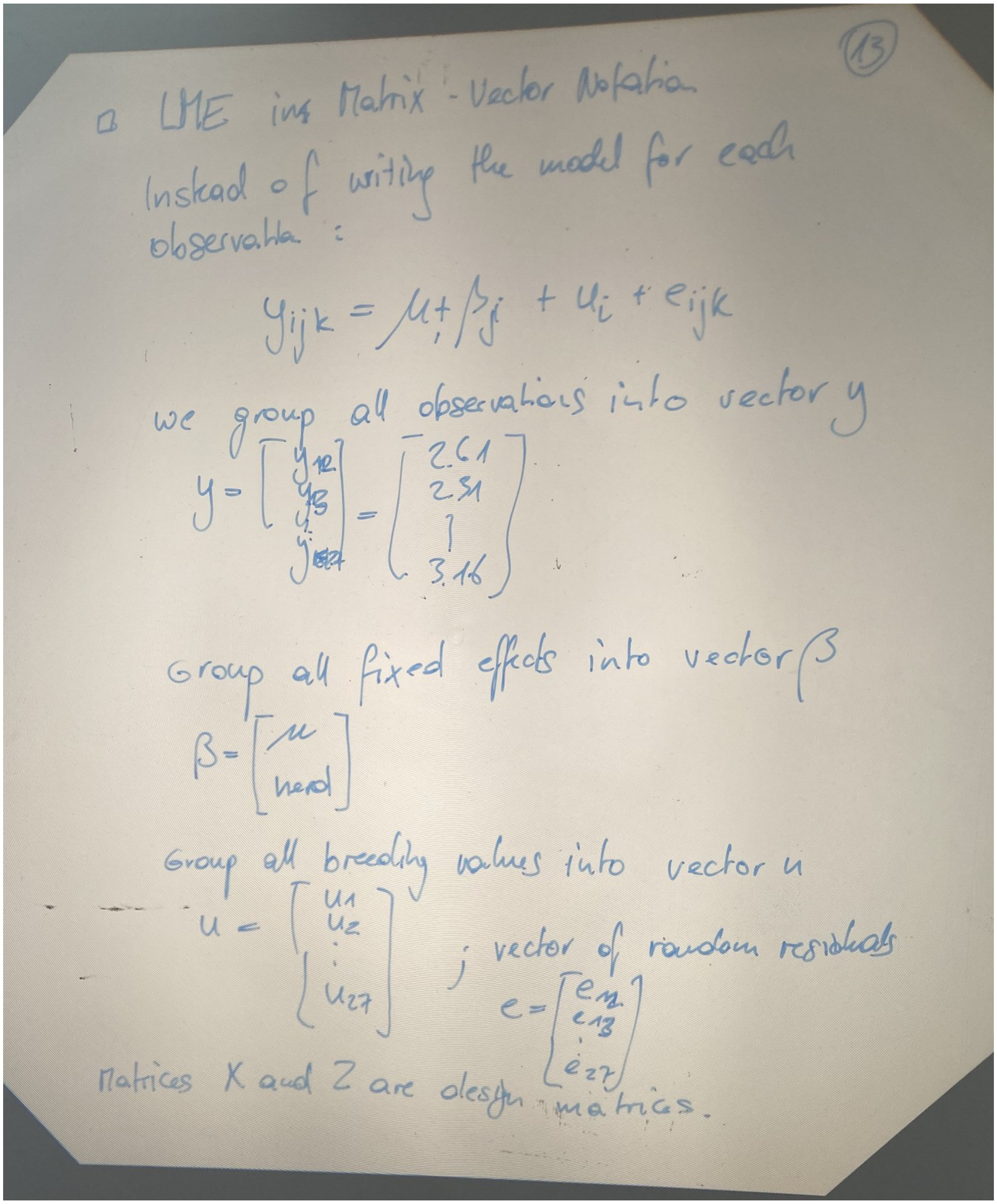
$$+ \underbrace{\mathrm{cov}(m_{26} \quad - \quad - \quad - \quad -}_{=0}$$

$$= \mathrm{cov}\left(\tfrac{1}{2} u_3, \tfrac{1}{2} u_3\right) + \mathrm{cov}\left(\tfrac{1}{2} u_{11}, \tfrac{1}{2} u_{11}\right)$$

$$= \tfrac{1}{4} \mathrm{cov}(u_3, u_3) + \tfrac{1}{4} \mathrm{cov}(u_{11}, u_{11})$$

$$= \tfrac{1}{4} \sigma_u^2 + \tfrac{1}{4} \sigma_u^2 = \tfrac{1}{2} \sigma_u^2$$

OHP Picture 13



⑬

☐ LME in Matrix - Vector Notation

Instead of writing the model for each observation :

$$y_{ijk} = \mu + \beta_j + u_i + e_{ijk}$$

we group all observations into vector y

$$y = \begin{bmatrix} y_{12} \\ y_{13} \\ \vdots \\ y_{27} \end{bmatrix} = \begin{bmatrix} 2.61 \\ 2.31 \\ \vdots \\ 3.16 \end{bmatrix}$$

Group all fixed effects into vector β

$$\beta = \begin{bmatrix} \mu \\ herd \end{bmatrix}$$

Group all breeding values into vector u

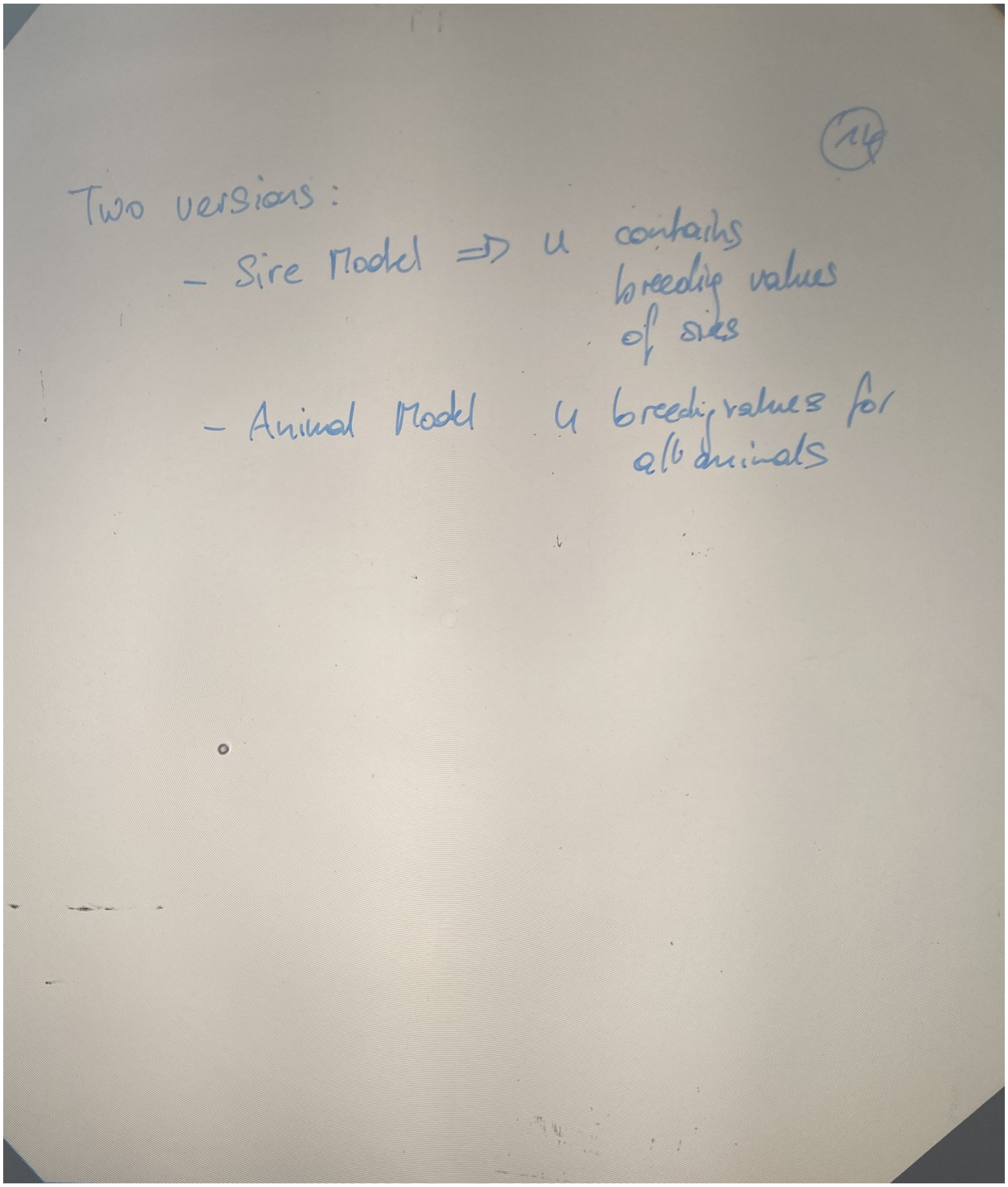
$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{27} \end{bmatrix}$$

; vector of random residuals

$$e = \begin{bmatrix} e_{12} \\ e_{13} \\ \vdots \\ e_{27} \end{bmatrix}$$

Matrices X and Z are design matrices.

OHP Picture 14



⑭

Two versions:

- Sire Model ⇒ u contains breeding values of sires
- Animal Model u breeding values for all animals