

# OHP Picture 1

## Recap: Prediction of Breeding Values

□ Simple scenario: Own Performance records  
(Eigenleistung)

- Every animal  $i$  has one observation ( $y_i$ ) of a phenotypic value of a trait of interest  
↳ (birth weight  
· mastitis resistance  
· fertility  
· ... )

Animals	Observations $y_i$ (numbers)
1	$y_1 = 52$
2	$y_2 = 48$
⋮	
$N$	

- Predicted breeding value  $\hat{u}_i$  of animal  $i$   
 $\hat{u}_i = h^2 (y_i - \mu)$  where  $h^2$ : heritability of trait  
 $\mu$ : population mean

- Predictions ( $\hat{u}_i$ ) are associated with errors:  
Quantification of prediction error: accuracy  $r_{\hat{u}, y}$   
Reliability  $r_{\hat{u}}^2 \rightarrow 8\%$  (Bestimmtheitsmass)

# OHP Picture 2

Accuracy of predicted breeding values is also important when quantifying the response to selection (R)

$$R = i \cdot r_{uy} \cdot \sigma_{By} = i \cdot h^2 \cdot \sigma_{By}$$

$\downarrow$  selection intensity  
 $\downarrow$  own performance  
 $\downarrow$  selection response per generation  
 $\rightarrow$  phenotypic standard deviation

## Breeders Equation

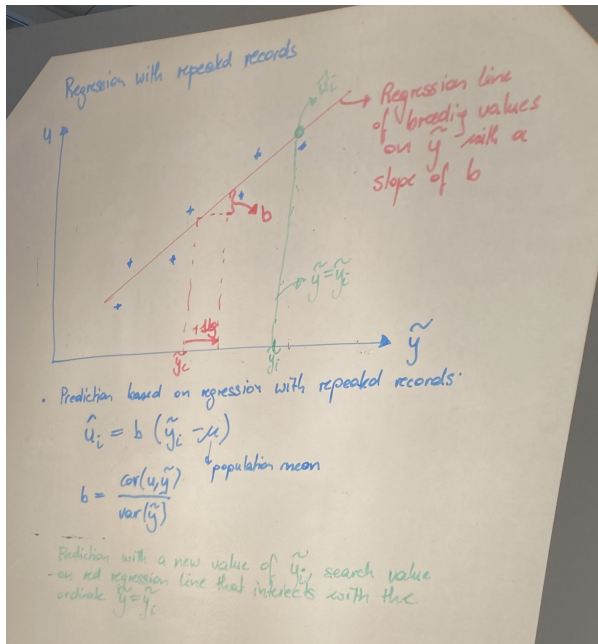
### Repeated Records

□ More than one observation per animal:

Animal $i$	Observation 1	Observation 2 ...	Observation $k$	Mean
1	$y_{11} = 152$	$y_{12} = 181$	$y_{1k} = 515$	$\frac{\sum_{j=1}^k y_{ij}}{k}$
2	$\vdots$			
$\vdots$				
$N$				

$$\bar{y}_i = \frac{1}{k} \sum_{j=1}^k y_{ij} \quad \text{mean}$$

# OHP Picture 3



# OHP Picture 4

Computation of regression coefficient  $b = \frac{\text{cov}(u_i, y)}{\text{var}(y)}$

□ Genetic model:

$$y_{ij} = \mu + u_i + e_{ij}$$

□ Decompose  $e_{ij}$  into a permanent part and a temporary part

$$e_{ij} = pe_i + te_{ij}$$

changes between different observations for animal  $i$

constant over all observations for animal  $i$

$$\Rightarrow y_{ij} = \mu + u_i + pe_i + te_{ij}$$

□ Decomposition at the level of variance:

$$\begin{aligned} \text{var}(y_{ij}) &= \text{var}(\mu + u_i + pe_i + te_{ij}) \\ &= \text{var}(\mu) + \text{var}(u_i) + \text{var}(pe_i) + \text{var}(te_{ij}) \\ &= \sigma_y^2 \quad \left\{ \begin{array}{l} \text{phenotypic} \\ \text{variance} \\ \text{of } n \text{ observations} \end{array} \right. \\ &= \left\{ \begin{array}{l} + 2\text{cov}(\mu, u_i) + 2\text{cov}(\mu, pe_i) + 2\text{cov}(\mu, te_{ij}) \\ + \dots \end{array} \right. \\ &= \text{var}(u_i) + \text{var}(pe_i) + \text{var}(te_{ij}) \\ &= \text{genetic additive variance} = \sigma_A^2 \end{aligned}$$

# OHP Picture 5

Decomposition of  $\bar{y}_j^2$  for repeated observations:

$$\text{var}(y_{ij}) = \underbrace{\text{var}(u_i)}_{\text{constant across all observations for animal } i} + \text{var}(p_{e_i}) + \text{var}(t_{e_{ij}})$$

□ Repeatability  $t$ : Tells us the ratio of variance components that are permanent compared to the total variance of all observations

$$t = \frac{\text{var}(u_i) + \text{var}(p_{e_i})}{\text{var}(y_{ij})} \Rightarrow \underline{\text{var}(u_i) + \text{var}(p_{e_i})} = t \cdot \bar{y}_j^2$$

$$1-t = \frac{\text{var}(y_{ij})}{\text{var}(y_{ij})} - \frac{\text{var}(u_i) + \text{var}(p_{e_i})}{\text{var}(y_{ij})}$$

$$= \frac{\text{var}(y_{ij}) - \text{var}(u_i) - \text{var}(p_{e_i})}{\text{var}(y_{ij})} = \frac{\text{var}(t_{e_{ij}})}{\text{var}(y_{ij})}$$

$$\Downarrow$$
$$\underline{\text{var}(t_{e_{ij}})} = (1-t)\bar{y}_j^2$$

# OHP Picture 6

Decomposition of  $\bar{y}_j^2$  for repeated observations:

$$\text{var}(y_{ij}) = \underbrace{\text{var}(u_i) + \text{var}(p_i)}_{\text{constant across all observations for animal } i} + \text{var}(t_{ij})$$

□ Repeatability  $t$ : Tells us the ratio of variance components that are permanent compared to the total variance of all observations

$$t = \frac{\text{var}(u_i) + \text{var}(p_i)}{\text{var}(y_{ij})} \Rightarrow \text{var}(u_i) + \text{var}(p_i) = t \cdot \bar{y}_j^2$$

$$1-t = \frac{\text{var}(y_{ij})}{\text{var}(y_{ij})} - \frac{\text{var}(u_i) + \text{var}(p_i)}{\text{var}(y_{ij})}$$

$$= \frac{\text{var}(y_{ij}) - \text{var}(u_i) - \text{var}(p_i)}{\text{var}(y_{ij})} = \frac{\text{var}(t_{ij})}{\text{var}(y_{ij})}$$

$$\Downarrow$$
$$\text{var}(t_{ij}) = (1-t)\bar{y}_j^2$$

# OHP Picture 7

Computation of  $b_i$ :

$$\text{cov}(u_i, \bar{y}_i) = \text{cov}\left(u_i \left[ \mu + u_i + p_i + \frac{1}{k} \sum_{j=1}^k t_{ij} \right] \right)$$

$$= \frac{1}{k} \sum_{j=1}^k \text{cov}\left(u_i, \left[ \mu + u_i + p_i + t_{ij} \right] \right)$$

↓ genetic model

$$= \underbrace{\frac{1}{k} \sum_{j=1}^k \mu}_{\frac{1}{k} \cdot k \cdot \mu = \mu} + \underbrace{\frac{1}{k} \sum_{j=1}^k u_i}_{u_i} + \underbrace{\frac{1}{k} \sum_{j=1}^k p_i}_{p_i} + \frac{1}{k} \sum_{j=1}^k t_{ij}$$

$$\text{cov}(u_i, \bar{y}_i) = \underbrace{\text{cov}(u_i, \mu)}_{=0} + \underbrace{\text{cov}(u_i, u_i)}_k + \underbrace{\text{cov}(u_i, p_i)}_{=0} + \underbrace{\text{cov}\left(u_i, \frac{1}{k} \sum_{j=1}^k t_{ij}\right)}_{=0}$$

$$= \text{cov}(u_i, u_i) = \text{var}(u_i) = \sigma_u^2$$

genetic additive  
variance

# OHP Picture 8

$$\begin{aligned} \text{var}(\tilde{y}_i) &= \text{var}\left(\frac{1}{k} \sum_{j=1}^k [u + u_i + p_i + k_{ij}]\right) \\ &= \text{var}\left(u + u_i + p_i + \frac{1}{k} \sum_{j=1}^k k_{ij}\right) \\ &= \underbrace{\text{var}(u)}_{=0} + \text{var}(u_i) + \text{var}(p_i) \\ &\quad + \text{var}\left(\frac{1}{k} \sum_{j=1}^k k_{ij}\right) + \underbrace{2 \text{cov}(u, u_i)}_{=0} t \dots \end{aligned}$$

$$t \cdot \tilde{\sigma}_y^2 \leftarrow \text{var}(u_i) + \text{var}(p_i) + \frac{1}{k} \text{var}(k_{ij}) = (1-t) \tilde{\sigma}_y^2$$

using repeatability relations

$$\begin{aligned} \Rightarrow \text{var}(\tilde{y}_i) &= t \cdot \tilde{\sigma}_y^2 + \frac{1}{k} (1-t) \tilde{\sigma}_y^2 \\ &= \frac{1}{k} [k \cdot t + (1-t)] \tilde{\sigma}_y^2 \\ &= \frac{1 + (k-1)t}{k} \tilde{\sigma}_y^2 \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{Regression coefficient } b &= \frac{\text{cov}(u, \tilde{y}_i)}{\text{var}(u, \tilde{y}_i)} = \frac{\tilde{\sigma}_u^2}{\frac{1 + (k-1)t}{k} \tilde{\sigma}_y^2} \\ &= \frac{k \tilde{\sigma}_u^2}{1 + (k-1)t \tilde{\sigma}_y^2} = \frac{k h^2}{1 + (k-1)t} \end{aligned}$$



# OHP Picture 9

$$\begin{aligned}
 \text{var}(\tilde{y}_i) &= \text{var}\left(\frac{1}{k} \sum_{j=1}^k [u + u_i + p_{e_i} + t e_{ij}]\right) \\
 &= \text{var}\left(u + u_i + p_{e_i} + \frac{1}{k} \sum_{j=1}^k t e_{ij}\right) \\
 &= \underbrace{\text{var}(u)} + \text{var}(u_i) + \text{var}(p_{e_i}) \\
 &\quad + \text{var}\left(\frac{1}{k} \sum_{j=1}^k t e_{ij}\right) + \underbrace{2 \text{cov}(u, u_i) + \dots}_{=0}
 \end{aligned}$$

$$t \cdot \bar{v}_y^2 = \underbrace{\text{var}(u_i) + \text{var}(p_{e_i})}_{t \cdot \bar{v}_y^2} + \frac{1}{k} \underbrace{\text{var}(t e_{ij})}_{(1-t) \bar{v}_y^2}$$

using repeatability relations

$$\Rightarrow \text{var}(\tilde{y}_i) = t \cdot \bar{v}_y^2 + \frac{1}{k} (1-t) \bar{v}_y^2$$

$$= \frac{1}{k} [k \cdot t + (1-t)] \bar{v}_y^2$$

$$= \frac{1 + (k-1)t}{k} \bar{v}_y^2$$

$$\begin{aligned}
 \Rightarrow \text{Regression coefficient } b &= \frac{\text{cov}(u_i, \tilde{y}_i)}{\text{var}(u_i, \tilde{y}_i)} = \frac{\bar{v}_u^2}{\frac{1 + (k-1)t}{k} \bar{v}_y^2} \\
 &= \frac{k \bar{v}_u^2}{1 + (k-1)t \bar{v}_y^2} = \frac{k h^2}{1 + (k-1)t}
 \end{aligned}$$

# OHP Picture 10

1. Own Performance
  2. Repeated Records
  3. Progeny Records
- } same animal, limited use in dairy cattle  
comes from breeding programs in dairy cattle

□ Dataset

Animal $i$ (Bulls)	daughters - half-sibs			Average across offspring
	offspring 1	offspring 2	... offspring $k$	
1	$y_{11}$			$\bar{y}_1$
2				$\bar{y}_2$
⋮				
$N$				$\bar{y}_N$

□ Regression to predict breeding values:

→ Predicted breeding value of animal  $i$ :

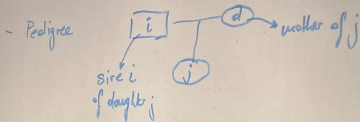
$$\hat{u}_i = b(\bar{y}_i - \mu); \quad b = \frac{\text{cov}(u_i, \bar{y}_i)}{\text{var}(\bar{y}_i)}$$

where  $\bar{y}_i = \frac{1}{k} \sum_{j=1}^k y_{ij}$

# OHP Picture 11

Decomposition of  $\bar{y}_i = \frac{1}{k} \sum_{j=1}^k y_{ij}$

- Genetic model:  $y_{ij} = \mu + u_j + e_j$



- Breeding value  $u_j$  of daughters  $j$ :

$$u_j = \frac{1}{2} u_i + \frac{1}{2} u_d + m_j$$

→ Insert into genetic model:

$$y_{ij} = \mu + \frac{1}{2} u_i + \frac{1}{2} u_d + m_j + e_j \quad ; \text{ do that for all daughters of sire } i$$

$$\bar{y}_i = \frac{1}{k} \sum_{j=1}^k [\mu + \frac{1}{2} u_i + \frac{1}{2} u_d + m_j + e_j]$$

$$= \mu + \frac{1}{2} u_i + \frac{1}{k} \sum_{j=1}^k \frac{1}{2} u_d + \frac{1}{k} \sum_{j=1}^k m_j + \frac{1}{k} \sum_{j=1}^k e_j$$

# OHP Picture 12

$$\begin{aligned}
 * \text{cov}(u_i, \bar{y}_i) &= \text{cov}\left(u_i, \mu + \frac{1}{2}u_i + \frac{1}{k} \sum_{j=1}^k \frac{1}{2}u_{dj} + \frac{1}{k} \sum_{j=1}^k m_j\right) \\
 &\quad + \frac{1}{k} \sum_{j=1}^k e_j \\
 &= \underbrace{\text{cov}(u_i, \mu)}_{=0} + \text{cov}\left(u_i, \frac{1}{2}u_i\right) + \text{cov}\left(u_i, \frac{1}{k} \sum_{j=1}^k \frac{1}{2}u_{dj}\right) \\
 &\quad + \underbrace{\text{cov}\left(u_i, \frac{1}{k} \sum_{j=1}^k m_j\right)}_{=0} + \underbrace{\text{cov}\left(u_i, \frac{1}{k} \sum_{j=1}^k e_j\right)}_{=0}
 \end{aligned}$$

values

Assume that sire  $i$  and dams  $d_j$  are unrelated

$$- \frac{1}{2} \text{cov}(u_i, u_i) = -\frac{1}{2} \text{var}(u_i) = -\frac{1}{2} \sigma_u^2$$

$$\Rightarrow \text{cov}(u_i, u_{dj}) = 0$$

$$\text{var}(\bar{y}_i) : \text{using } \bar{y}_i = \mu + \frac{1}{2}u_i + \frac{1}{k} \sum_{j=1}^k \frac{1}{2}u_{dj} + \frac{1}{k} \sum_{j=1}^k m_j + \frac{1}{k} \sum_{j=1}^k e_j$$

$$\text{var}(\bar{y}_i) = \underbrace{\text{var}\left(\frac{1}{2}u_i\right)}_{\text{permanent part for sire } i} + \text{var}\left(\frac{1}{k} \sum_{j=1}^k \frac{1}{2}u_{dj}\right) + \text{var}\left(\frac{1}{k} \sum_{j=1}^k m_j\right) + \text{var}\left(\frac{1}{k} \sum_{j=1}^k e_j\right)$$

For a given sire  $i$ :  
 permanent part:  $\text{var}\left(\frac{1}{2}u_i\right)$   
 across all daughters

## OHP Picture 13

Computation with Variances:

Random variable  $X$  (continuous)

$$E[X] = \int x f(x) dx \quad \text{with } f(x) \text{ density of } X$$

$$\text{Var}[X] = \int (x - E(x))^2 f(x) dx$$

$$\text{var}[a \cdot x] = \int a^2 (x - E(x))^2 f(x) dx$$

$$= a^2 \int (x - E(x))^2 f(x) dx$$

$$= a^2 \cdot \text{var}(x)$$

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)$$

$$\text{cov}(X, Y+Z) = \text{cov}(X, Y) + \text{cov}(X, Z)$$

# OHP Picture 14

$$\begin{aligned} \text{var}(\bar{y}_i) &= \text{var}\left(\frac{1}{2}u_i\right) + \text{var}\left(\frac{1}{k} \sum_{j=1}^k \frac{1}{2}u_{ij}\right) \\ &+ \text{var}\left(\frac{1}{k} \sum_{j=1}^k w_j\right) + \text{var}\left(\frac{1}{k} \sum_{j=1}^k c_j\right) \\ &= \frac{1}{4}\text{var}(u_i) + \frac{1}{k} \text{var}\left(\sum_{j=1}^k \frac{1}{2}u_{ij}\right) \\ &= t \cdot \bar{v}_y^2 + \frac{1}{k}(1-t)\bar{v}_y^2 \quad \text{with } t = \frac{h^2}{4} \end{aligned}$$

Regression Coefficient:

$$b = \frac{\text{cov}(u_i, \bar{y}_i)}{\text{var}(\bar{y}_i)} = \frac{1/2 \bar{v}_u^2}{\left(t + \frac{1-t}{k}\right) \bar{v}_y^2}$$

$$= \frac{1/2 h^2 \bar{v}_y^2}{\left(\frac{h^2}{4} + \left(1 - \frac{h^2}{4}\right) \frac{1}{k}\right) \bar{v}_y^2}$$

$$= \frac{1/2 h^2 k}{\frac{kh^2}{4} + \left(1 - \frac{h^2}{4}\right)} = \frac{2kh^2}{kh^2 + (4-h^2)}$$

$$= \frac{2k}{k + (4-h^2)/h^2} = \frac{2k}{k + \beta} \quad \text{with } \beta = \frac{4-h^2}{h^2}$$

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

# OHP Picture 15

$$\begin{aligned} \text{var}(\bar{y}_i) &= \text{var}\left(\frac{1}{2}u_i\right) + \text{var}\left(\frac{1}{k} \sum_{j=1}^k \frac{1}{2}u_{ij}\right) \\ &= \text{var}\left(\frac{1}{k} \sum_{j=1}^k u_{ij}\right) + \text{var}\left(\frac{1}{k} \sum_{j=1}^k e_j\right) \\ &= \frac{1}{4} \text{var}(u_i) + \frac{1}{k^2} \text{var}\left(\sum_{j=1}^k \frac{1}{2}u_{ij}\right) \\ &= t \cdot \bar{v}_y^2 + \frac{1}{k} (1-t) \bar{v}_y^2 \quad \text{with } t = \frac{4}{k} \end{aligned}$$

Regression coefficient:

$$\begin{aligned} b &= \frac{\text{cov}(u_i, \bar{y}_i)}{\text{var}(\bar{y}_i)} = \frac{\frac{1}{2} \bar{v}_u^2}{\left(t + \frac{(1-t)}{k}\right) \bar{v}_y^2} \\ &= \frac{\frac{1}{2} t^2 \bar{v}_y^2}{\left(\frac{t^2}{4} + (1 - \frac{t^2}{4}) / k\right) \bar{v}_y^2} \\ &= \frac{\frac{1}{2} t^2 k}{\frac{t^2}{4} + (1 - \frac{t^2}{4})} = \frac{2kt^2}{kt^2 + (4-t^2)} \\ &= \frac{2k}{k + (4-t^2)/t^2} = \frac{2k}{k + \beta} \quad \text{with } \beta = \frac{4-t^2}{t^2} \end{aligned}$$

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

# OHP Picture 16

Reading Data into R:

- traditional : `read.table(...)`  $\Rightarrow$  `data.frame`
- new : readr package : `read_csv()`  
function

$\Downarrow$   
tibble  
(modern data frame)

- writing data to a file

`readr::write_csv(tbl_weight, file = "weight_data.csv")`

$\rightarrow$  column separated by ";" } US-style  
decimal delimiter : "." }

column sep by ";"  
decimal : "."