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 resistence, fertility, ...)

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

- Predicted breeding value $\hat{u}_i$ of animal i

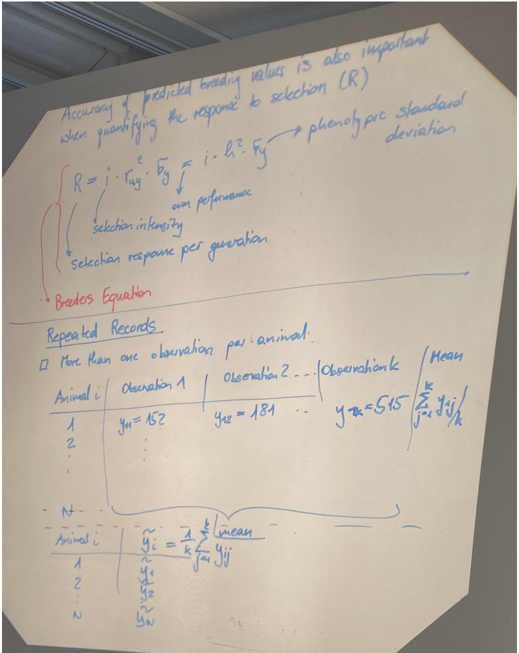
$$\hat{u}_i = h^2 (y_i - \mu)$$

where h^2: heritability of trait

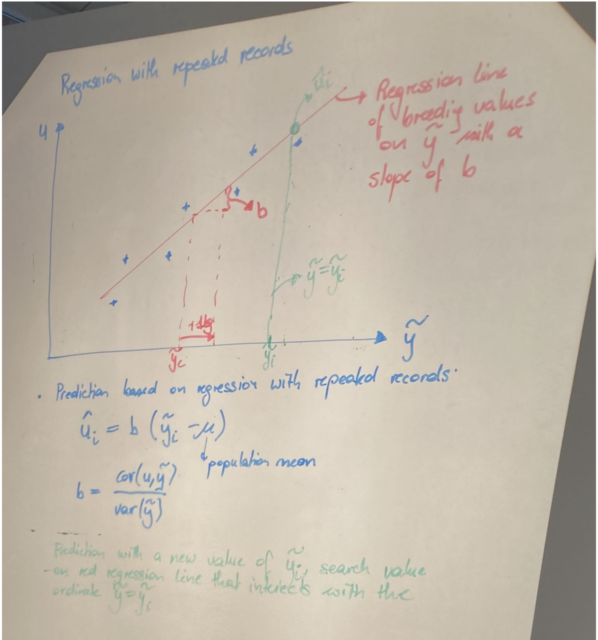
μ: population mean

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

OHP Picture 2



OHP Picture 3



OHP Picture 4

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

- Genetic model:

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

- Decompose e_{ij} into a permanent part and a temporary part

$$e_{ij} = \underbrace{pe_i}_{\text{constant over all observations for animal } i} + \underbrace{te_{ij}}_{\text{changes between different observations for animal } i}$$

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

- Decomposition at the level of variance:

$$\underbrace{var(y_{ij})}_{\text{phenotypic variance of all observations } = \sigma_y^2} = var(\mu + u_i + pe_i + te_{ij})$$

$$= var(\mu) + var(u_i) + var(pe_i) + var(te_{ij})$$

$$\rightarrow 0 \left\{ \begin{array}{l} + 2cov(\mu, u_i) + 2cov(\mu, pe_i) + 2cov(\mu, te_{ij}) \\ + \ldots \end{array} \right.$$

$$= \underbrace{var(u_i)}_{\text{genetic additive variance } = \sigma_u^2} + var(pe_i) + var(te_{ij})$$

OHP Picture 5

Decomposition of σ_y^2 for repeated observations:

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

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

$$t = \frac{var(u_i) + var(pe_i)}{var(y_{ij})} \Rightarrow var(u_i) + var(pe_i) = t \cdot \sigma_y^2$$

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

$$= \frac{var(y_{ij}) - var(u_i) - var(pe_i)}{var(y_{ij})} = \frac{var(te_{ij})}{var(y_{ij})}$$

$$\Downarrow$$

$$var(te_{ij}) = (1-t)\sigma_y^2$$

OHP Picture 6

Decomposition of σ_y^2 for repeated observations:

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

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

$$t = \frac{var(u_i) + var(pe_i)}{var(y_{ij})} \Rightarrow var(u_i) + var(pe_i) = t \cdot \sigma_y^2$$

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

$$= \frac{var(y_{ij}) - var(u_i) - var(pe_i)}{var(y_{ij})} = \frac{var(te_{ij})}{var(y_{ij})}$$

$$\Downarrow$$

$$var(te_{ij}) = (1-t)\sigma_y^2$$

OHP Picture 7

Computation of b:

$$cov(u_i, \tilde{y}_i) = cov\left(u_i, \left[\mu + u_i + pe_i + \frac{1}{k}\sum_{j=1}^{k} te_{ij}\right]\right)$$

$$\tilde{y}_i = \frac{1}{k}\sum_{j=1}^{k} \underbrace{y_{ij}}_{\text{genetic model}} = \frac{1}{k}\sum_{j=1}^{k}\left[\mu + u_i + pe_i + te_{ij}\right]$$

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

$$cov(u_i, \tilde{y}_i) = \underbrace{cov(u_i, \mu)}_{=0} + cov(u_i, u_i) + \underbrace{cov(u_i, pe_i)}_{=0} + \underbrace{cov\left(u_i, \frac{1}{k}\sum_{j=1}^{k} te_{ij}\right)}_{=0}$$

$$= cov(u_i, u_i) = \underbrace{var(u_i)}_{\text{genetic additive variance}} = \sigma_u^2$$

OHP Picture 8

$$\operatorname{var}(\tilde{y}_i) = \operatorname{var}\left(\frac{1}{k}\sum_{j=1}^{k}\left[\mu + u_i + pe_i + te_{ij}\right]\right)$$

$$= \operatorname{var}\left(\mu + u_i + pe_i + \frac{1}{k}\sum_{j=1}^{k} te_{ij}\right)$$

$$= \overbrace{\operatorname{var}(\mu)}^{=0} + \operatorname{var}(u_i) + \operatorname{var}(pe_i) + \operatorname{var}\left(\frac{1}{k}\sum_{j=1}^{k} te_{ij}\right) + \underbrace{2\operatorname{cov}(\mu, u_i) + \ldots}_{=0}$$

$$\underbrace{t \cdot \sigma_y^2}_{} = \overbrace{\operatorname{var}(u_i) + \operatorname{var}(pe_i)} + \frac{1}{k}\underbrace{\operatorname{var}(te_{ij})}_{(1-t)\sigma_y^2}$$

using repeatability relations

$$\Rightarrow \operatorname{var}(\tilde{y}_i) = t \cdot \sigma_y^2 + \frac{1}{k}(1-t)\sigma_y^2$$

$$= \frac{1}{k}\left[k \cdot t + (1-t)\right]\sigma_y^2$$

$$= \frac{1 + (k-1)t}{k}\sigma_y^2$$

$$\Rightarrow \text{Regression coefficient } b = \frac{\operatorname{cov}(u, \tilde{y})}{\operatorname{var}(u, \tilde{y})} = \frac{\sigma_u^2}{\frac{1+(k-1)t}{k}\sigma_y^2}$$

$$= \frac{k\sigma_u^2}{(1+(k-1)t)\,\sigma_y^2} = \frac{k h^2}{1+(k-1)t}$$

OHP Picture 9

$$\mathrm{var}(\tilde{y}_i) = \mathrm{var}\left(\frac{1}{k}\sum_{j=1}^{k}\left[\mu + u_i + pe_i + te_{ij}\right]\right)$$

$$= \mathrm{var}\left(\mu + u_i + pe_i + \frac{1}{k}\sum_{j=1}^{k} te_{ij}\right)$$

$$= \overbrace{\mathrm{var}(\mu)}^{=\emptyset} + \mathrm{var}(u_i) + \mathrm{var}(pe_i)$$

$$+ \mathrm{var}\left(\frac{1}{k}\sum_{j=1}^{k} te_{ij}\right) + \underbrace{2\,\mathrm{cov}(\mu, u_i) + \ldots}_{=\emptyset}$$

$$= \overbrace{\mathrm{var}(u_i) + \mathrm{var}(pe_i)}^{t \cdot \sigma_y^2} + \frac{1}{k}\underbrace{\mathrm{var}(te_{ij})}_{(1-t)\sigma_y^2}$$

using repeatability relations

$$\Rightarrow \mathrm{var}(\tilde{y}_i) = t \cdot \sigma_y^2 + \frac{1}{k}(1-t)\sigma_y^2$$

$$= \frac{1}{k}\left[k \cdot t + (1-t)\right]\sigma_y^2$$

$$= \frac{1 + (k-1)t}{k}\sigma_y^2$$

$$\Rightarrow \text{Regression coefficient } b = \frac{\mathrm{cov}(u, \tilde{y})}{\mathrm{var}(u, \tilde{y})} = \frac{\sigma_u^2}{\frac{1+(k-1)t}{k}\sigma_y^2}$$

$$= \frac{k\sigma_u^2}{1+(k-1)t\,\sigma_y^2} = \frac{k h^2}{1+(k-1)t}$$

OHP Picture 10

1. Own Performance
2. Repeated Records

} same animal, limited use in dairy cattle

3. Progeny Records — comes from breeding programs in dairy cattle

---

- Dataset

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

$\underbrace{\text{offspring 1} \quad \text{offspring 2} \quad \cdots \quad \text{offspring } k}_{\text{daughters – halfsibs}}$

- Regression to predict breeding values:
  - → Predicted breeding value of animal i:

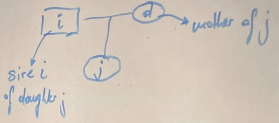
$$\hat{u}_i = b(\bar{y}_i - \mu) \; ; \; b = \frac{cov(u_i, \bar{y}_i)}{var(\bar{y}_i)}$$

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



- 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:

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

$$\rightarrow \bar{y}_i = \frac{1}{k}\sum_{j=1}^{k}\left[\mu + \frac{1}{2}u_i + \frac{1}{2}u_{d_j} + m_j + e_j\right]$$

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

OHP Picture 12

* $\operatorname{cov}(u_i, \bar{y}_i) = \operatorname{cov}\left(u_i, \mu + \frac{1}{2}u_i + \frac{1}{k}\sum_{j=1}^{k}\frac{1}{2}u_{dj} + \frac{1}{k}\sum_{j=1}^{k} m_j + \frac{1}{k}\sum_{j=1}^{k} e_j\right)$

$$= \underbrace{\operatorname{cov}(u_i, \mu)}_{=0} + \operatorname{cov}\left(u_i, \frac{1}{2}u_i\right) + \operatorname{cov}\Big(u_i, \overbrace{\frac{1}{k}\sum_{j=1}^{k}\frac{1}{2}u_{dj}}^{\text{mothers}}\Big) + \underbrace{\operatorname{cov}\left(u_i, \frac{1}{k}\sum_{j=1}^{k} m_j\right)}_{=0} + \underbrace{\operatorname{cov}\left(u_i, \frac{1}{k}\sum_{j=1}^{k} e_j\right)}_{=0}$$

Assume that sire i and dams d_j are unrelated

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

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

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

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

For a given sire i:

permanent part across all daughters : $\operatorname{var}\left(\frac{1}{2}u_i\right)$

OHP Picture 13

Computation with Variances:

Random variable X (continuous)

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

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

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

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

$$= a^2 \cdot var(x)$$

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

$$cov(X, Y+Z) = cov(X,Y) + cov(X,Z)$$

OHP Picture 14

$$\mathrm{var}(\bar{y}_i) = \mathrm{var}\left(\tfrac{1}{2} u_i\right) + \mathrm{var}\left(\tfrac{1}{k} \sum_{j=1}^{k} \tfrac{1}{2} u_{dj}\right) + \mathrm{var}\left(\tfrac{1}{k} \sum_{j=1}^{k} m_j\right) + \mathrm{var}\left(\tfrac{1}{k} \sum_{j=1}^{k} e_j\right)$$

$$= \tfrac{1}{4} \mathrm{var}(u_i) + \tfrac{1}{k^2} \mathrm{var}\left(\sum_{j=1}^{k} \tfrac{1}{2} u_{dj}\right)$$

$$= t \cdot \sigma_y^2 + \tfrac{1}{k}(1-t)\sigma_y^2 \quad \text{with } t = \tfrac{h^2}{4}$$

Regression Coefficient:

$$b = \frac{\mathrm{cov}(u, \bar{y}_1)}{\mathrm{var}(\bar{y}_i)} = \frac{\tfrac{1}{2}\sigma_u^2}{\left(t + \frac{(1-t)}{k}\right)\sigma_y^2}$$

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

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

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

$$\hat{u}_c = b(\bar{y}_1 - \mu) = \frac{2k}{k+\beta}(\bar{y}_1 - \mu)$$

OHP Picture 15

$$var(\bar{y}_i) = var\left(\tfrac{1}{2}u_i\right) + var\left(\frac{1}{k}\sum_{j=1}^{k} \tfrac{1}{2} u_{dj}\right)$$

$$+ var\left(\frac{1}{k}\sum_{j=1}^{k} m_j\right) + var\left(\frac{1}{k}\sum_{j=1}^{k} e_j\right)$$

$$= \tfrac{1}{4} var(u_i) + \frac{1}{k^2} var\left(\sum_{j=1}^{k} \tfrac{1}{2} u_{dj}\right)$$

$$= t \cdot \sigma_y^2 + \frac{1}{k}(1-t)\sigma_y^2 \quad \text{with } t = \frac{h^2}{4}$$

Regression Coefficient:

$$b = \frac{cov(u, \bar{y}_i)}{var(\bar{y}_i)} = \frac{\tfrac{1}{2}\sigma_u^2}{\left(t + \frac{(1-t)}{k}\right)\sigma_y^2}$$

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

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

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

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

OHP Picture 16

Reading Data into R:

- traditional : read.table(...) ⇒ data.frame
- new : readr package : read_csv() function

⇓

tibble
(modern data frame)

- writing data to a file

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

→ column separated by ","
decimal delimiter : "."
} US-settings

column sep by ";"
decimal : ","