

# Genetic Evaluation

Peter von Rohr

02 Oktober 2020

# Consequences of Definition of Breeding Value

- ▶ Based on the average of a large number of offspring, because
  - ▶ offspring inherit a random sample of parental alleles
  - ▶ average over a large number of offspring reduces sampling effect
- ▶ The breeding value is defined as a deviation from the population mean
  - ▶ population mean depends on allele frequencies which are specific for each population
  - ▶ hence breeding values can only be compared within one population.
- ▶ Because the breeding value is defined as a deviation its expected value of the breeding value is 0

# The Basic Model

$$\begin{aligned}y_{ij} &= v_i + e_{ij} \\ &= \mu + u_i + d_i + i_i + e_{ij}\end{aligned}$$

where

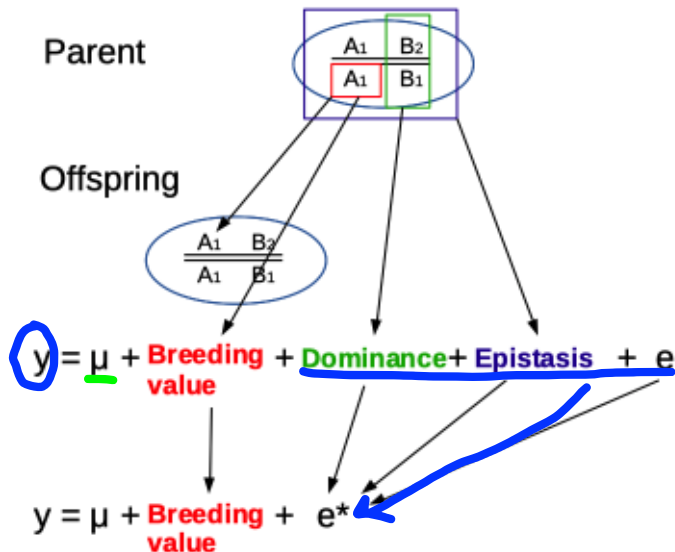
$y_{ij}$   $j^{th}$  record of animal  $i$

$\mu$  population mean

$v_i$  genotypic value, corresponding to the sum of all additive ( $u$ ), dominance ( $d$ ) and epistatic ( $i_i$ ) effects of the genotype of animal  $i$

$e_{ij}$  random environmental effects of animal  $i$

## Re-arranging Terms



## New Model

$$y_{ij} = \mu_i + u_i + e_{ij}^*$$

where

- $y_{ij}$   $j^{th}$  record of animal  $i$
- $\mu_i$  identifiable fixed environmental effect
- $u_i$  sum of all additive ( $u$ ) genetic effects of the genotype of animal  $i$
- $e_{ij}^*$  dominance, epistatic and random environmental effects of animal  $i$

# Infinitesimal Model

- ▶ Central Limit Theorem for  $u_i$  and  $e_{ij}$  lead to **multivariate normal distributions** with
  - ▶  $E(u) = 0$  and  $E(e) = 0$  and
  - ▶ Known variances and co-variances
  - ▶ No co-variances between  $u_i$  and  $e_{ij}$
- ▶  $\mu$  is assumed to be constant for a given evaluation
- ▶ Phenotypic observation  $y_{ij}$  is the sum of two normally distributed random variables, therefore
  - ▶  $y_{ij}$  also follows a multivariate normal distribution
  - ▶  $E(y) = \mu$

# Central Limit Theorem

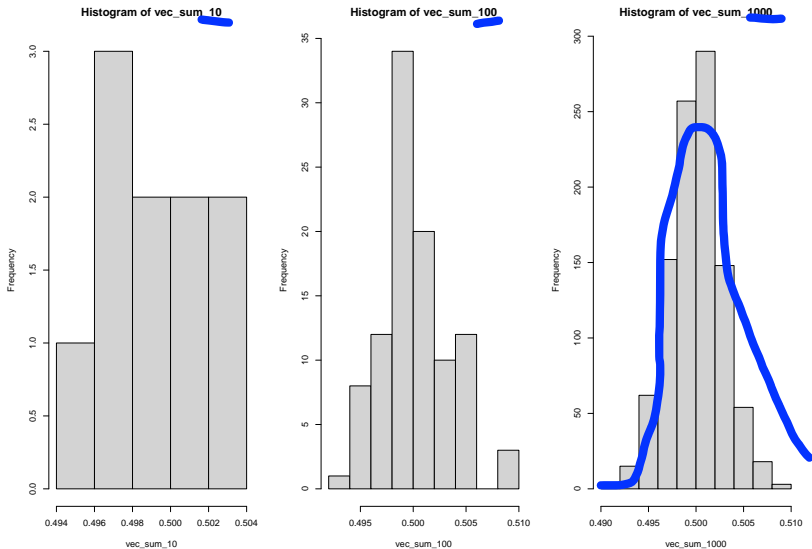
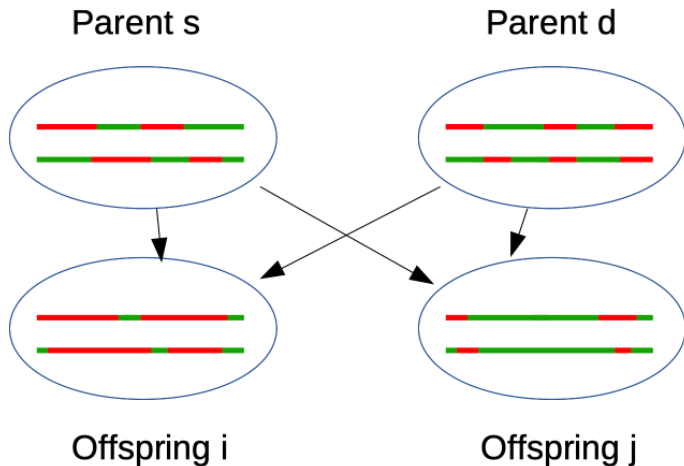


Figure 1: Distribution of Sums of Different Numbers of Components

# Decomposition of Breeding Value



$$u_i = 1/2u_s + 1/2u_d + m_i$$

$$u_j = 1/2u_s + 1/2u_d + m_j$$

**Mendelian sampling:**  
caused by the fact  
that not every full sib  
receives the same sample  
of alleles from the s and d



# Basic Principle of Predicting Breeding Values

Breeding values are predicted according to the following two steps.

1. Observations corrected for the appropriate mean performance values of animals under the same conditions
  - ▶ conditions are described by the effects captured in  $\mu_j$ .
2. The corrected observations are weighted by a certain factor
  - ▶ factor reflects the amount of information available for prediction

**1: correct  $y$  for an appropriate mean:  $(y - \mu)$**

**2: weighting factor for the corrected observation**

# Animal's Own Performance - Single Record

Data from the complete population:

Animal ID | Observation

1 |  $y_1$

2 |  $y_2$

...

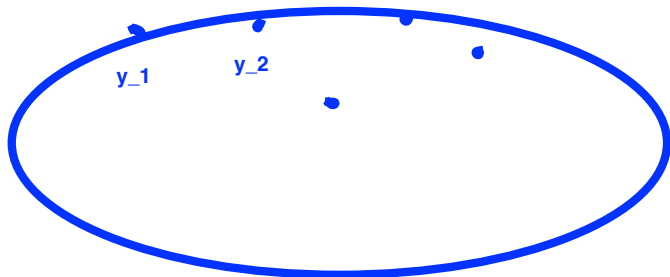
N |  $y_N$

Predict breeding values  $u_i$   
based on observations and  
based on new genetic model

- ▶ one phenotypic observation per animal
- ▶ search for prediction  $\hat{u}_i$  of the breeding value  $u_i$  of animal  $i$
- ▶ assume  $u_i$  and  $y_i$  known for a certain population

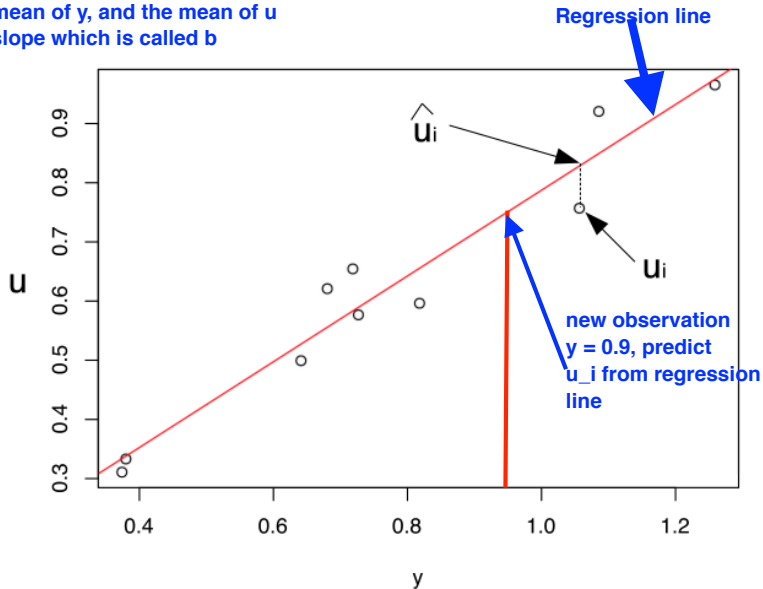
→ plot

population



## Plot $u$ against $y$

Regression line is determined by  
> mean of  $y$ , and the mean of  $u$   
> slope which is called  $b$



# Regression

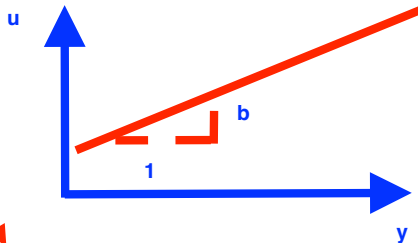
- ▶ red line denotes **regression line** from  $u_i$  onto  $y_i$
- ▶ because phenotypes have genetic basis  $\rightarrow$  connection between  $u_i$  and  $y_i$
- ▶ measure for slope: regression coefficient  $b$
- ▶ new genetic model can be interpreted as regression

$$u_i = y_{ij} - \mu_i - e_{ij}^*$$

- ▶ Allowing for different slopes in a statistical model, introduces  $b$

$$u_i = b * (y_{ij} - \mu_i) + e_{ij}^{**}$$

# Regression Coefficient



$$\begin{aligned} b &= \frac{\text{cov}(u, y)}{\text{var}(y)} \\ &= \frac{\text{cov}(u, \mu + u + e)}{\text{var}(y)} \\ &= \frac{\text{cov}(u, u)}{\text{var}(y)} \\ &= \frac{\text{var}(u)}{\text{var}(y)} = h^2 \end{aligned}$$

where  $h^2$  is called **heritability**

# Prediction

- ▶ Given a new  $y_i$ , what would be the predicted  $u_i$ ?
- ▶ Use regression line and compute  $\hat{u}_i$

$$\begin{aligned}\hat{u}_i &= b * (y_i - \mu) \\ &= h^2 * (y_i - \mu)\end{aligned}$$

# Accuracy

- Measured as correlation between true breeding value  $u$  and selection criterion  $y$

$$\begin{aligned}r_{u,y} &= \frac{\text{cov}(u, y)}{\sigma_u \sigma_y} \\&= \frac{\sigma_u^2}{\sigma_u \sigma_y} \\&= \frac{\sigma_u}{\sigma_y} \\&= h \quad \text{Square root of the heritability}\end{aligned}$$

In practice: heritabilities are considered to be high when they are between 0.4 - 0.6.  
They considered medium between 0.2 to 0.4  
Below 0.2, they are considered low. Most modern traits, have very low heritabilities.  
This simple method does not yield very accurate predictions.

# Response To Selection

- ▶ Why is the accuracy important?
- ▶ Response to selection depends on it
- ▶ Example of single record

$$R = i * \underbrace{r_{u,y}^2}_{\text{accuracy}} * \sigma_y = i * h^2 * \sigma_y$$

**i: selection intensity ==> how many animals are selected as parents**  
**\sigma\_y: phenotypic standard deviation**



# Repeated Records

Example: Dairy cattle: cows with repeated lactations

- ▶ Additional component of variation in the phenotypic observations  $y$   
 $\text{var}(te)$ : temporary environment

$$\text{var}(y) = \text{var}(u) + \text{var}(pe) + \text{var}(te)$$

$\text{var}(u)$ : genetic additive

$\text{var}(pe)$ : permanent environment

$$t = \frac{\text{var}(u) + \text{var}(pe)}{\text{var}(y)}$$

$t$  is repeatability

Before:

$$\text{var}(e) = \text{var}(pe) + \text{var}(te)$$

- ▶ Predicted breeding value

$$\hat{u}_i = b(\tilde{y}_i - \mu)$$

mean of repeated observations

# Regression Coefficient

$$b = \frac{\text{cov}(u, \tilde{y})}{\text{var}(\tilde{y})}$$

$$\text{cov}(u, \tilde{y}) = \text{cov}(u, u + pe + \frac{1}{n} \sum_{k=1}^n te_k) = \sigma_u^2$$

$$\text{var}(\tilde{y}) = \text{var}(u) + \text{var}(pe) + \frac{1}{n} \text{var}(te)$$

$$\begin{aligned}\text{var}(\tilde{y}) &= t * \sigma_y^2 + \frac{1}{n} (1 - t) * \sigma_y^2 \\ &= \frac{1}{n} (n * t + (1 - t)) \sigma_y^2 \\ &= \frac{1 + (n - 1)t}{n} \sigma_y^2\end{aligned}$$

**n: number of observations  
per animal**

# Putting Results together

from before:  $b$  for a single record per animal:

$$b = h^2$$

$$\begin{aligned} b &= \frac{\text{cov}(u, \tilde{y})}{\text{var}(\tilde{y})} \\ &= \frac{n\sigma_u^2}{(1 + (n-1)t)\sigma_y^2} \\ &= \frac{nh^2}{1 + (n-1)t} \end{aligned}$$

for high values of  $t$ : additional information in the mean - corresponding to increasing then value of  $n$  - does not yield different predictions.

# Progeny Records

Important in livestock: Sires have many offspring and most of them are daughters with observations.

Problem with previous scenarios: Only animals with observations can get predicted breeding values. But most of the important traits in dairy cattle, can only be observed in female animals. But the selection of male parents is very important.

Solutions: Sires with many daughters ( $n$ ) and we use the mean of these observations from daughters as information of the sires.

Example: Sire  $i$  and we want to predict its breeding value

$$\hat{u}_i = b * (\bar{y}_i - \mu)$$

where

$$b = \frac{\text{cov}(u_i, \bar{y}_i)}{\text{var}(\bar{y}_i)}$$

## Covariance and Variance

$$\begin{aligned}\text{cov}(u_i, \bar{y}_i) &= \text{cov}\left(u_i, \frac{1}{2}u_i + \frac{1}{2}\frac{1}{n}\sum_{k=1}^n u_{d,i} + \frac{1}{n}\sum_{k=1}^n m_k + \frac{1}{n}\sum_{k=1}^n e_k\right) \\ &= \text{cov}\left(u_i, \frac{1}{2}u_i\right) \\ &= \frac{1}{2}\text{cov}(u_i, u_i) = \frac{1}{2}\sigma_u^2\end{aligned}$$

$$\text{var}(\bar{y}_i) = (t + (1 - t)/n)\sigma_y^2$$

with  $t = h^2/4$

## Intra-Class $t$

### ► Progeny mean

$$\begin{aligned}\bar{y}_i &= \frac{1}{n} \sum_{k=1}^n y_k = \frac{1}{n} \sum_{k=1}^n u_k + \frac{1}{n} \sum_{k=1}^n e_k \\ &= \frac{1}{n} \sum_{k=1}^n (1/2 u_i + 1/2 u_{d,k}) + \frac{1}{n} \sum_{k=1}^n e_k \\ &= \frac{1}{2} u_i + \frac{1}{n} \sum_{k=1}^n 1/2 u_{d,k} + \frac{1}{n} \sum_{k=1}^n e_k\end{aligned}$$

### ► Variance

$$\text{var}(\bar{y}_i) = \text{var}\left(\frac{1}{2} u_i\right) + \text{var}\left(\frac{1}{n} \sum_{k=1}^n 1/2 u_{d,k}\right) + \text{var}\left(\frac{1}{n} \sum_{k=1}^n e_k\right)$$

with  $\text{cov}(\cdot) = 0$ ,  $t = \text{var}(\frac{1}{2} u_i) / \text{var}(y) = h^2/4$

# Results

$$\begin{aligned} b &= \frac{1/2\sigma_u^2}{(t + (1 - t)/n)\sigma_y^2} \\ &= \frac{1/2h^2\sigma_y^2}{(\frac{1}{4}h^2 + (1 - \frac{1}{4}h^2)/n)\sigma_y^2} \\ &= \frac{2nh^2}{nh^2 + (4 - h^2)} \\ &= \frac{2n}{n + (4 - h^2)/h^2} \\ &= \frac{2n}{n + k} \end{aligned}$$

**n is the number of daughters  
per sire i**

with  $k = \frac{4-h^2}{h^2}$ .

•