Genetic Evaluation

Peter von Rohr

2022-10-07

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

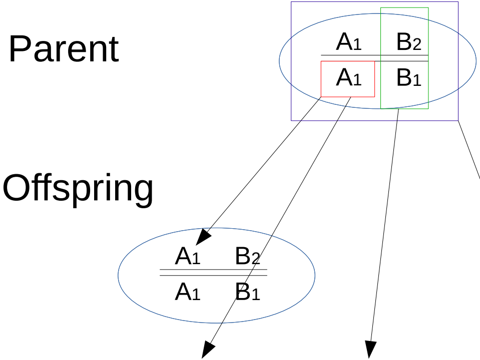
The Basic Model

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

where

y_{ij}	j^{th} record of animal i
μ	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



$$y = \mu + \textbf{Breeding value} + \textbf{Dominance} + \textbf{Epistasis} + e$$

$$y = \mu + \textbf{Breeding value} + e^*$$

New Model

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

where

y_{ij}	j^{th} record of animal i
μ_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}
- μ 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

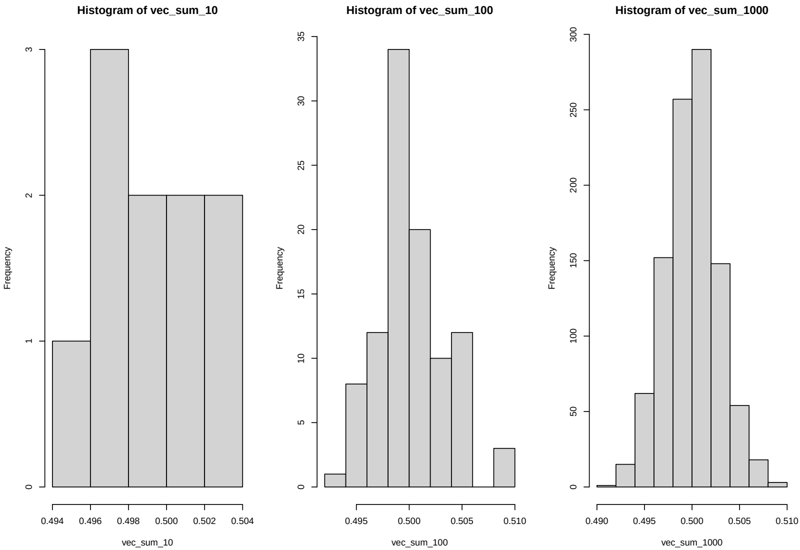
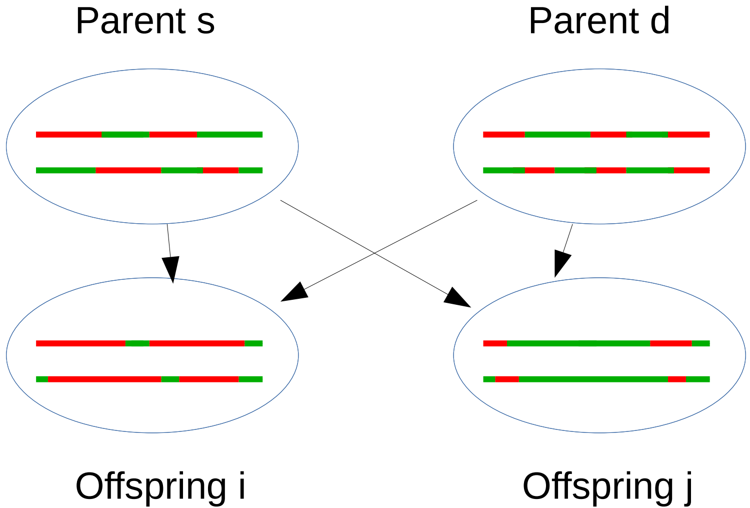


Figure 1: Distribution of Sums of Different Numbers of Components

Decomposition of Breeding Value



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

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 μ_i.
2. The corrected observations are weighted by a certain factor
   - factor reflects the amount of information available for prediction

Statistical Perspective

From a statistical point of view:

- Given phenotypic observation y as source of information
- Use best linear predictor ($\hat{u}$) for breeding value u
- Hence

$$\hat{u} = E(u) + b(y - E(y)) = E(u|y)$$

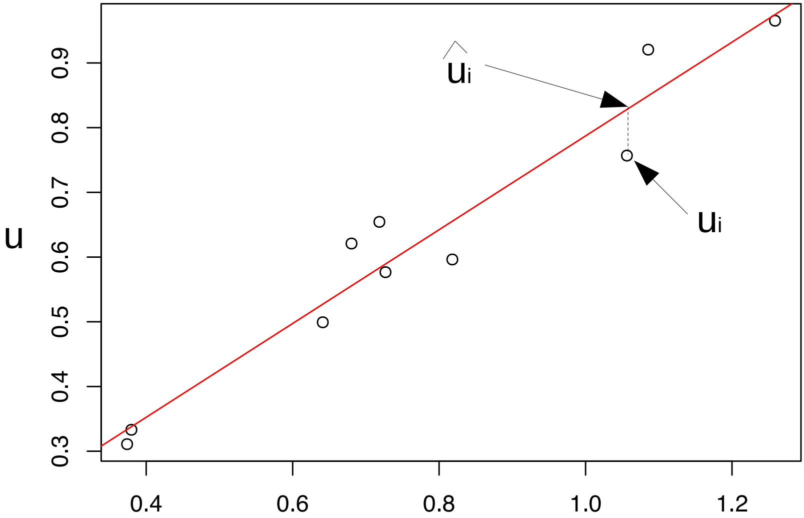
- with $E(u) = 0 \rightarrow \hat{u} = b(y - E(y))$
- b depends on the relationship between y and u
- examples will follow ...

Animal's Own Performance - Single Record

- 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

$\rightarrow$ plot

Plot u against y



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

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

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

$$
\begin{aligned}
r_{u,y} &= \frac{cov(u, y)}{\sigma_u \ \sigma_y} \\
&= \frac{\sigma_u^2}{\sigma_u \ \sigma_y} \\
&= \frac{\sigma_u}{\sigma_y} \\
&= h
\end{aligned}
$$

Response To Selection

- Why is the accuracy important?
- Response to selection depends on it
- Example of single record
- Breeders equation, quantifying the selection response per generation

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

Repeated Records

- Additional component of variation

$$var(y) = var(u) + var(pe) + var(te)$$

$$t = \frac{var(u) + var(pe)}{var(y)} = \frac{\sigma_u^2 + \sigma_{pe}^2}{\sigma_y^2}$$

- Predicted breeding value

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

Regression Coefficient

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

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

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

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

Putting Results together

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

Progeny Records

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

where

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

Note

$$\bar{y}_i = \frac{1}{n} \sum_{k=1}^{n} y_k$$

where y_k is the phenotypic record of progeny k of parent i

Covariance and Variance

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

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

with $t = h^2/4$

Intra-Class t

- Progeny mean

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

- Variance

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

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

Results

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

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