Variance Components Estimation

Peter von Rohr

06.05.2019

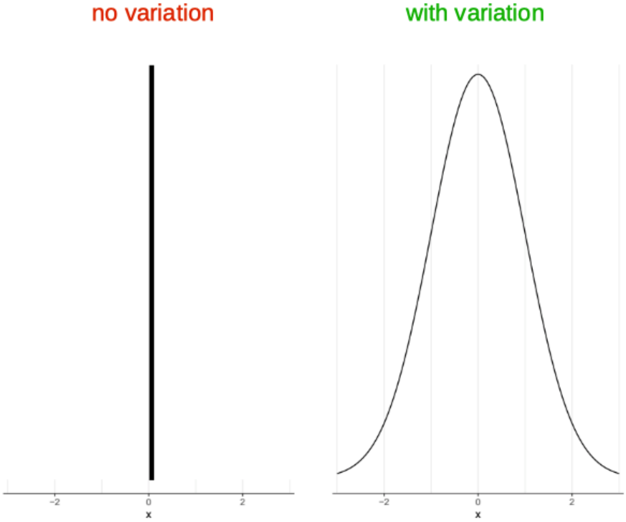
Genetic Variation

- Requirement for trait to be considered in breeding goal
- Breeding means improvement of next generation via selection and mating
- Only genetic (additive) components are passed to offspring
- Selection should be based on genetic component of trait
- Selection only possible with genetic variation

$\rightarrow$ genetic variation indicates how good characteristics are passed from parents to offspring

$\rightarrow$ measured by **heritability** $h^2 = \frac{\sigma_a^2}{\sigma_p^2}$

Two Traits



Problems

- Genetic components cannot be observed or measured
- Must be estimated from data
- Data are mostly phenotypic

$\rightarrow$ topic of variance components estimation

- Model based, that means connection between phenotypic measure and genetic component are based on certain model

$$p = g + e$$

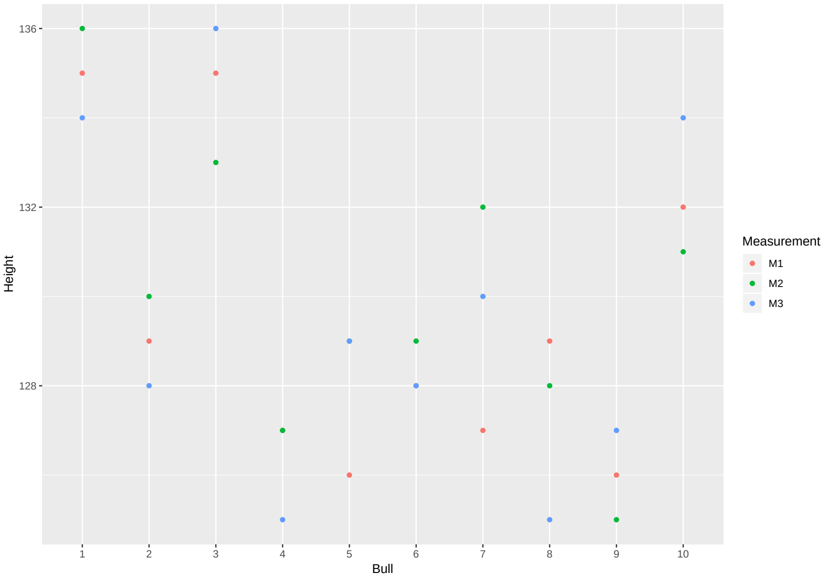
with $cov(g, e) = 0$

- **Goal**: separate variation due to g (σ_a^2) from phenotypic variation

Example of Variance Components Separation

- Estimation of repeatability
- Given repeated measurements of same trait at the same animal
- Repeatability means variation of measurements at the same animal is smaller than variation between measurements at different animals

Repeatability Plot



Model

$$y_{ij} = \mu + t_i + \epsilon_{ij}$$

where

y_{ij}	measurement j of animal i
μ	expected value of y
t_i	deviation of y_{ij} from μ attributed to animal i
ϵ_{ij}	measurement error

Estimation Of Variance Components

- $E(t_i) = 0$
- $\sigma_t^2 = E(t_i^2)$: variance component of total variance (σ_y^2) which can be attributed to the t-effects
- $E(\epsilon_{ij}) = 0$
- $\sigma_\epsilon^2 = E(\epsilon_{ij}^2)$: variance component attributed to ϵ-effects
- $\sigma_y^2 = \sigma_t^2 + \sigma_\epsilon^2$
- Repeatability w defined as:

$$w = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\epsilon^2}$$

$\rightarrow$ estimate of σ_t^2 needed

Analysis Of Variance (ANOVA)

Effect	df	Sum Sq	Mean Sq	$E(Mean\ Sq)$
Bull (t)	$r-1$	$SSQ(t)$	$SSQ(t)/(r-1)$	$\sigma_\epsilon^2 + n * \sigma_t^2$
Residual (ϵ)	$N-r$	$SSQ(\epsilon)$	$SSQ(\epsilon)/(N-r)$	σ_ϵ^2

where

$$SSQ(t) = \left[\frac{1}{n}\sum_{i=1}^{r}\left(\sum_{j=1}^{n} y_{ij}\right)^2\right] - \left(\sum_{i=1}^{r}\sum_{j=1}^{n} y_{ij}\right)^2 / N$$

$$SSQ(\epsilon) = \sum_{i=1}^{r}\sum_{j=1}^{n} y_{ij}^2 - \left[\frac{1}{n}\sum_{i=1}^{r}\left(\sum_{j=1}^{n} y_{ij}\right)^2\right]$$

Zahlenbeispiel

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Bull          9  286.7   31.85   13.85 8.74e-07 ***
## Residuals    20   46.0    2.30                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Setting expected values of `Mean Sq` equal to estimates of variance components

$$\hat{\sigma}_{\epsilon}^2 = 2.3 \text{ and } \hat{\sigma}_t^2 = \frac{31.85 - 2.3}{3} = 9.85$$

Repeatability

$$\hat{w} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_{\epsilon}^2} = 0.81$$

Same Strategy for Sire Model

- Sire model is a mixed linear effects model with sire effects s as random components

$$y = Xb + Zs + e$$

- In case where sires are not related, \$var(s) $= I * \sigma_s^2$
- From σ_s^2, we get genetic additive variance as $\sigma_a^2 = 4 * \sigma_s^2$

ANOVA

Effect	Degrees of Freedom	Sum Sq	Mean Sq	$E(Mean\ Sq)$
Sire $(s\|b)$	$r-1$	$SSQ(s\|b)$	$SSQ(s\|b)/(r-1)$	$\sigma_e^2 + k * \sigma_s^2$
Residual (e)	$N-r$	$SSQ(e)$	$SSQ(e)/(N-r)$	σ_e^2

with

$$k = \frac{1}{r-1}\left[N - \frac{\sum_{i=1}^{r} n_i^2}{N}\right]$$

Maximum Likelihood (ML)

- Likelihood

$$L(\theta) = f(y|\theta)$$

- Normal distribution

$$L(\theta) = (2\pi)^{-1/2n}\sigma^{-n}|H|^{-1/2} * exp\left\{-\frac{1}{2\sigma^2}(y - Xb)^T H^{-1}(y - Xb)\right\}$$

with $var(y) = H * \sigma^2$ and $\theta^T = \left[\begin{array}{cc} b & \sigma^2 \end{array}\right]$

Maximisation of Likelihood

- Set $\lambda = logL$
- Compute partial derivatives of λ with respect to all unknowns

$$\frac{\partial \lambda}{\partial b}$$

$$\frac{\partial \lambda}{\partial \sigma^2}$$

- Set partial derivatives to 0 and solve for unknowns
- Use solutions as estimates

Restricted Maximum Likelihood (REML)

- Problem with ML: estimate of σ^2 depends on $b \rightarrow$ undesirable
- Do transformations Sy and Qy

1. The matrix S has rank $n - t$ and the matrix Q has rank t
2. The result of the two transformations are independent, that means $cov(Sy, Qy) = 0$ which is met when $SHQ^T = 0$
3. The matrix S is chosen such that $E(Sy) = 0$ which means $SX = 0$
4. The matrix QX is of rank t, so that every linear function of the elements of Qy estimate a linear function of b.

REML II

- From (i) and (ii) it follows that the likelihood L of y is the product of the likelihoods of Sy ($L*$) and Qy ($L**$) that means

$$\lambda = \lambda * + \lambda **$$

- Variance components are estimated from $\lambda*$ which will then be independent of b