A coefficient of determination (R^2) for linear mixed models and its use for variety trials

Hans-Peter Piepho

Biostatistics Unit
Universität Hohenheim
Germany

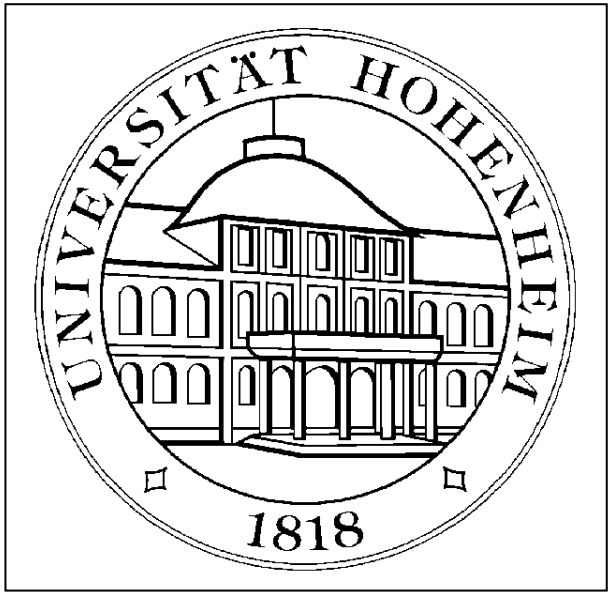


Table of contents

1. Coefficient of determination for linear model (LM)

$$y = X\beta + e\,, \qquad (1)$$

where

y = response vector of length n

β = fixed effects vector

X = design matrix, and

$e \sim N\left(0, V = I_n \sigma_e^2\right)$ = residual error vector

1. Coefficient of determination for linear model (LM)

Null model

$$y = 1_n \phi + e \ , \qquad (2)$$

where

1_n = a vector of ones

ϕ = intercept

$e \sim N\left(0, V_0 = I_n \sigma^2_{e0}\right)$ = residual error vector

1. Coefficient of determination for linear model (LM)

The standard procedure

Error sum of squares for full model:

$$SS_{error}^{full} = y^T P_\beta y \quad \text{where} \quad P_\beta = I_n - X\left(X^T X\right)^{-} X^T$$

Error sum of squares for full model:

$$SS_{error}^{null} = y^T P_\phi y \quad \text{where} \quad P_\phi = I_n - n^{-1} 1_n 1_n^T$$

1. Coefficient of determination for linear model (LM)

Coefficient of determination (LM)

$$R^2 = 1 - \frac{SS_{error}^{full}}{SS_{error}^{null}}$$

$$R^2_{adj} = 1 - \frac{(n-1)SS_{error}^{full}}{(n-p)SS_{error}^{null}} \qquad \text{where} \qquad p = rank(X)$$

1. Coefficient of determination for linear model (LM)

Coefficient of determination (LM)

$$R^2 = 1 - \frac{n^{-1} SS_{error}^{full}}{n^{-1} SS_{error}^{null}} = 1 - \frac{\hat{\sigma}^2_{e(ML)}}{\hat{\sigma}^2_{e0(ML)}}$$

$$R^2_{adj} = 1 - \frac{(n-p)^{-1} SS_{error}^{full}}{(n-1)^{-1} SS_{error}^{null}} = 1 - \frac{\hat{\sigma}^2_{e(REML)}}{\hat{\sigma}^2_{e0(REML)}}$$

1. Coefficient of determination for linear model (LM)

What does R^2 estimate?

$$\Omega_\beta = \frac{\Delta\theta(V, V_0)}{\theta(V_0)} \quad , \tag{3}$$

where

$\theta(V)$ = total variance implied by the variance-covariance structure V

$$\Delta\theta(V, V_0) = \theta(V_0) - \theta(V)$$

= variance explained by effects added in full model relative to null model

1. Coefficient of determination for linear model (LM)

For LM

$\theta(V_0) = \sigma^2_{e0}$,

$\theta(V) = \sigma^2_e$, and

$\Delta\theta(V, V_0) = \sigma^2_{e0} - \sigma^2_e$ and hence

$$\Omega_\beta = \frac{\sigma^2_{e0} - \sigma^2_e}{\sigma^2_{e0}} = 1 - \frac{\sigma^2_e}{\sigma^2_{e0}} \qquad (4)$$

2. Linear mixed models (LMM)

$$y = X\beta + Zu + e\ , \tag{5}$$

Z = design matrix

$u \sim N(0, G)$

$e \sim N(0, R)$

$y \sim N(X\beta, V)$ with

$$V = ZGZ^T + R \tag{6}$$

2. Linear mixed models (LMM)

Coefficient of determination for fixed effects in LMM

$$\Omega_\beta = \frac{\Delta\theta(V, V_0)}{\theta(V_0)}$$

V_0 is for the null model, in which $X\beta$ in (5) is replaced by $1_n\phi$.

2. Linear mixed models (LMM)

Requirements for definition of $\theta(V)$

$\Rightarrow$ allow for heterogeneity of variance

$\Rightarrow$ allow for covariance between observations

$\Rightarrow$ reduce to common R^2 for LM when random effects dropped and $R = I_n\sigma_e^2$

$\Rightarrow$ should be additive, i.e.,

$$\theta(V_1 + V_2) = \theta(V_1) + \theta(V_2) \tag{7}$$

2. Linear mixed models (LMM)

Marginal variance (mv)

$$mv(y_i) = v_{ii} \ , \tag{8}$$

y_i = *i*-th element of y

v_{ij} = *ij*-th element of V

Average marginal variance (AMV)

$$\theta^{AMV}(V) = \frac{1}{n}\sum_{i=1}^{n} mv(y_i) = \frac{1}{n} trace(V) \tag{9}$$

2. Linear mixed models (LMM)

Average marginal variance (AMV)

$$\theta^{AMV}(V)=\frac{1}{n}\sum_{i=1}^{n}mv(y_i)=\frac{1}{n}trace(V)$$

The trace of a variance-covariance matrix is a common measure of total variance in multivariate analysis.

The major downside of this criterion is that it does not account for covariances $v_{ij}\ \ (i\neq j)$ (Mustonen, 1997; Johnson and Wichern, 2002, p.139).

2. Linear mixed models (LMM)

"semivariance" (*sv*)

$$sv(y_i, y_j) = \frac{1}{2}\text{var}(y_i - y_j) = \frac{1}{2}(v_{ii} + v_{jj}) - v_{ij} \quad (10)$$

Average semivariance (*ASV*)

$$\theta^{ASV}(V) = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j>i+1}^{n} sv(y_i, y_j) = \frac{1}{n-1} trace(VP_\phi) \quad (11)$$

2. Linear mixed models (LMM)

Average semivariance (ASV)

$$\theta^{ASV}(V) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i+1}^{n} sv(y_i, y_j) = \frac{1}{n-1} trace(VP_\phi) \qquad (11)$$

It is readily verified that

$$\theta^{AMV}(V) = \theta^{ASV}(V) = \sigma_e^2$$

when $V = R = I_n \sigma_e^2$ (i.e., for an LM) as required.

3. Balanced one-way model with random group effects

a groups

m observations per group

$$y = 1_n \phi + \left(I_a \otimes 1_m\right) u + e \,, \tag{12}$$

where

$u \sim N\left(0, I_a \sigma_u^2\right)$ represents a random group effects

$e \sim N\left(0, I_n \sigma_e^2\right)$ represents $n = am$ random deviations from the group means

3. Balanced one-way model with random group effects

The variance of y is

$$V = I_a \otimes \left(J_m \sigma_u^2 + I_m \sigma_e^2 \right) \qquad (13)$$

Intra-class correlation

$$\rho = \sigma_u^2 / \left(\sigma_u^2 + \sigma_e^2 \right)$$

3. Balanced one-way model with random group effects

Total variance

$$\theta^{AMV}(V) = \sigma_u^2 + \sigma_e^2$$

$$\theta^{ASV}(V) = (am-1)^{-1}(a-1)m\sigma_u^2 + \sigma_e^2$$

The marginal variance of $\sigma_u^2 + \sigma_e^2$ only occurs for observations between groups but not within groups, where the variance amounts to only σ_e^2

4. Coefficient of determination for random effects

$$\Omega_u = \frac{\theta\left(ZGZ^T\right)}{\theta\left(V_0\right)} \tag{14}$$

Variance explained jointly by fixed and random effects:

$$\Omega_{\beta u} = \Omega_\beta + \Omega_u = 1 - \frac{\theta\left(R\right)}{\theta\left(V_0\right)} \tag{15}$$

5. Example: Long-term trends in variety trials

Basic model for long-term MET data

$$y_{ijk} = \mu + G_i + L_j + Y_k + (LY)_{jk} + (GL)_{ij} + (GY)_{ik} + (GLY)_{ijk} \qquad (16)$$

y_{ijk} = mean yield of the *i*-th genotype in the *j*-th location and *k*-th year

μ = overall mean

G_i = main effect of the *i*-th genotype

L_j = main effect of the *j*-th location

Y_k = main effect of the *k*-th year

$(LY)_{jk}$ = *jk*-th location × year interaction

$(GL)_{ij}$ = *ij*-th genotype × location interaction

$(GY)_{ik}$ = *ik*-th genotype × year interaction

$(GLY)_{ijk}$ = residual comprising both genotype × location × year interaction as well as the error of a mean

5. Example: Long-term trends in variety trials

Mackay et al. (2011)

- Take G_i and Y_k as fixed (can't take random because of time trend)
- All other effects random (i.i.d. normal with constant variance)
- Adjusted means for G_i assess genetic trend

  $\Rightarrow$ Plotted against year in which variety entered trial
- Adjusted means for Y_k assess non-genetic trend

  $\Rightarrow$ Plotted against calendar year
- Estimate trend by linear regression based on adjusted means for G_i and Y_k
- Look at one intensity at a time

5. Example: Long-term trends in variety trials

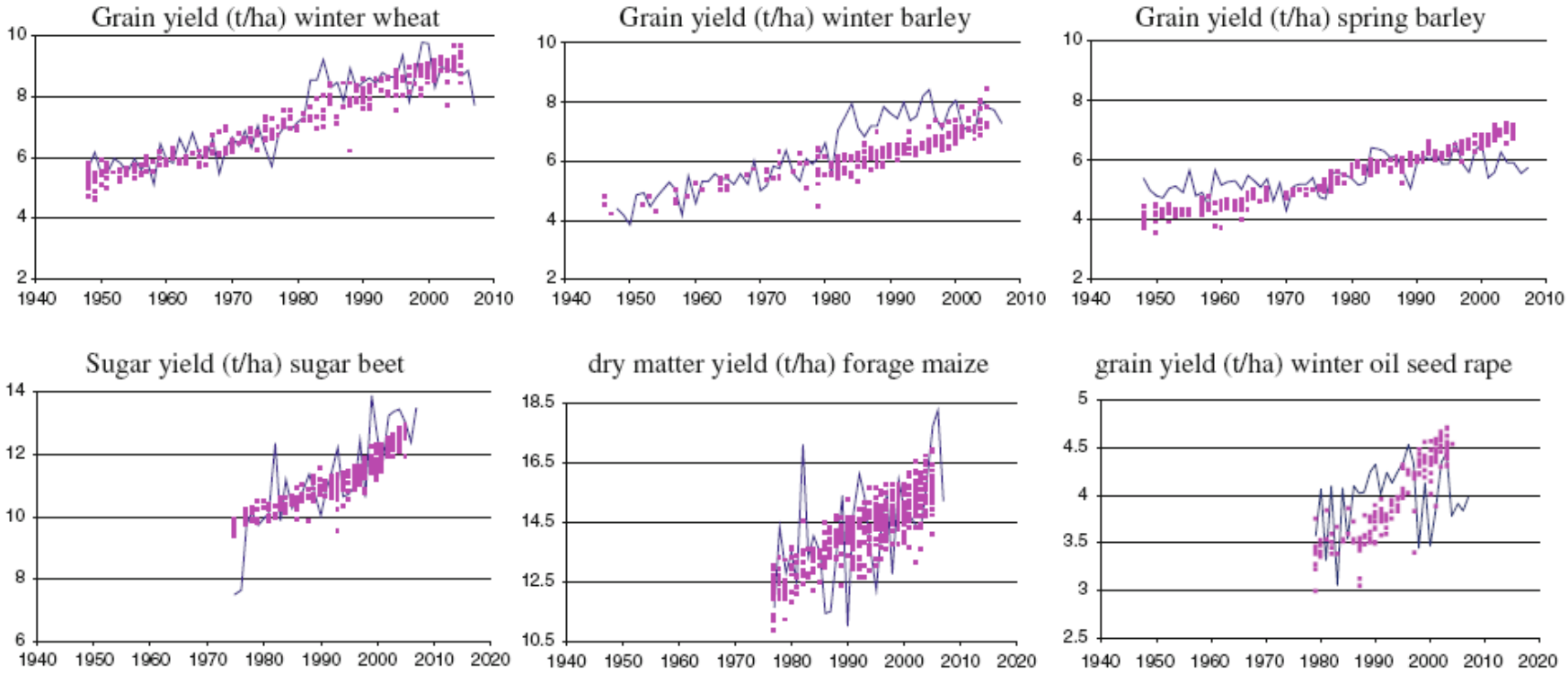


Fig. 1 Trends in variety and year effect for yield (t/ha) from 1948 to 2007. Ordinate and abscissa are on the same scale for all crops except oil seed rape. Variety and year means were estimated as described in "Materials and methods" section. Variety effects (*squares*) are plotted against the year in which the variety first entered the trial. Year means are plotted as a *line*

(Mackay et al., 2011)

—— year means

■ variety means

5. Example: Long-term trends in variety trials

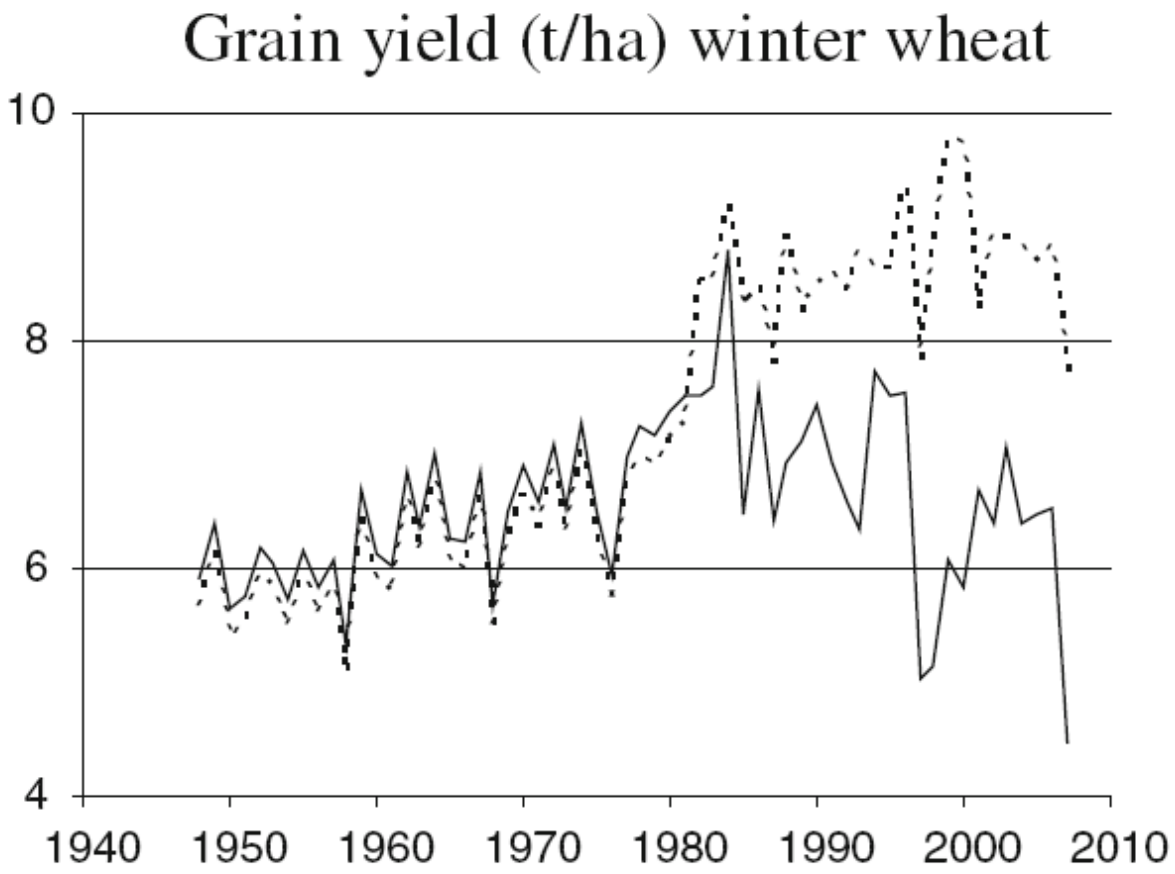


Fig. 3: Trends in year effects in UK variety trials (Mackay et al. 2011)
– – – - *Dotted line*: analysis with treated trials only (from 1982 onwards).
——— *Solid line*: analysis with untreated trials only (from 1982 onwards).

5. Example: Long-term trends in variety trials

Genetic trend

$$G_i = \beta r_i + H_i \qquad (17)$$

β = fixed regression coefficient for genetic trend

r_i = year of first trial for *i*-th variety

$H_i \sim N\left(0, \sigma_H^2\right)$

Non-genetic trend

$$Y_k = \gamma t_k + Z_k \qquad (18)$$

γ = fixed regression coefficient for agronomic trend

t_k = calendar year

$Z_k \sim N\left(0, \sigma_Z^2\right)$

5. Example: Long-term trends in variety trials

The fixed part of the model so far

$$\eta_{ik} = \mu + \beta r_i + \gamma t_k \qquad (19)$$

η_{ik} = expected response of the *i*-th genotype in the *k*-th year

5. Example: Long-term trends in variety trials

Variance component estimates (rye, Germany, 2001 to 2016)

Random effect	Full model	Null model
G	19.5	44.7
L	43.4	43.4
Y	20.7	19.3
L•Y	87.7	87.7
G•Y	2.4	2.5
G•L	1.8	1.8
G•L•Y	18.2	18.2

5. Example: Long-term trends in variety trials

Regression coefficient estimates

Fixed effect	Estimate	Standard error
μ	-1464	568
β	0.8875	0.0906
γ	-0.1183	0.2769

5. Example: Long-term trends in variety trials

Coefficient of determination

Coefficient of determination	θ^{AMV}	θ^{ASV}
Ω_{β}	0.0818	0.0816
Ω_{u}	0.8329	0.8315
$\Omega_{\beta u}$	0.9147	0.9131

6. And now for something completely different?

Broad-sense heritability

$$H^2 = \sigma_g^2 \big/ \sigma_p^2$$

σ_g^2 = genotypic variance

Narrow-sense heritability

$$h^2 = \sigma_a^2 \big/ \sigma_p^2$$

σ_a^2 = additive genetic variance

σ_p^2 = phenotypic variance

Uses of heritability

- Descriptive measure of precision of trial
- Compute response to selection ($R = h^2 S$, where S = selection differential)
- Compute predictive accuracy in genomic prediction

6. And now for something completely different?

What is the 'phenotype' here?

- Single plot observation?
- Genotype mean in a trial?
- Genotype mean in multi-environment trial (MET)?
- BLUP of genotypic value?

$\Rightarrow$ How to estimate the phenotypic variance σ_p^2 ?

7. Heritability for balanced data

Basic model for MET data

$$y_{ijk} = \phi + g_i + e_j + (ge)_{ij} + b_{jk} + \varepsilon_{ijk}$$

y_{ijk} = yield of the i^{th} genotype in the j^{th} location and k^{th} replicate

ϕ = overall mean

g_i = main effect of the i^{th} genotype; $\sim N(0, \sigma_g^2)$

e_j = main effect of the j^{th} environment

$(ge)_{ij}$ = ij^{th} genotype × environment interaction effect; $\sim N(0, \sigma_{ge}^2)$

b_{jk} = k^{th} replicate effect in j^{th} environment

ε_{ijk} = residual comprising both genotype × location × year interaction as well as the error of a mean; $\sim N(0, \sigma^2)$

Model assumes a randomized complete block design (RCBD) per environment

7. Heritability for balanced data

The phenotype

Assume balanced data from RCBD or CRD & complete G x E classification $\Rightarrow$

$$\bar{y}_{i\bullet\bullet} = \phi + g_i + \bar{e}_{\bullet} + \overline{(ge)}_{i\bullet} + \bar{b}_{\bullet\bullet} + \bar{\varepsilon}_{i\bullet\bullet}$$

$$\sigma_p^2 = \sigma_g^2 + \underbrace{\sigma_{ge}^2 / m + \sigma^2 / (rm)}$$

variance of a mean = $\frac{1}{2}$ variance of a difference

m = number of environments

r = number of replicates (complete blocks) per environment

8. Heritability for unbalanced & correlated data

The phenotype : mostly from unbalanced data / designs

***Crop variety trials and plant breeding trials*:**

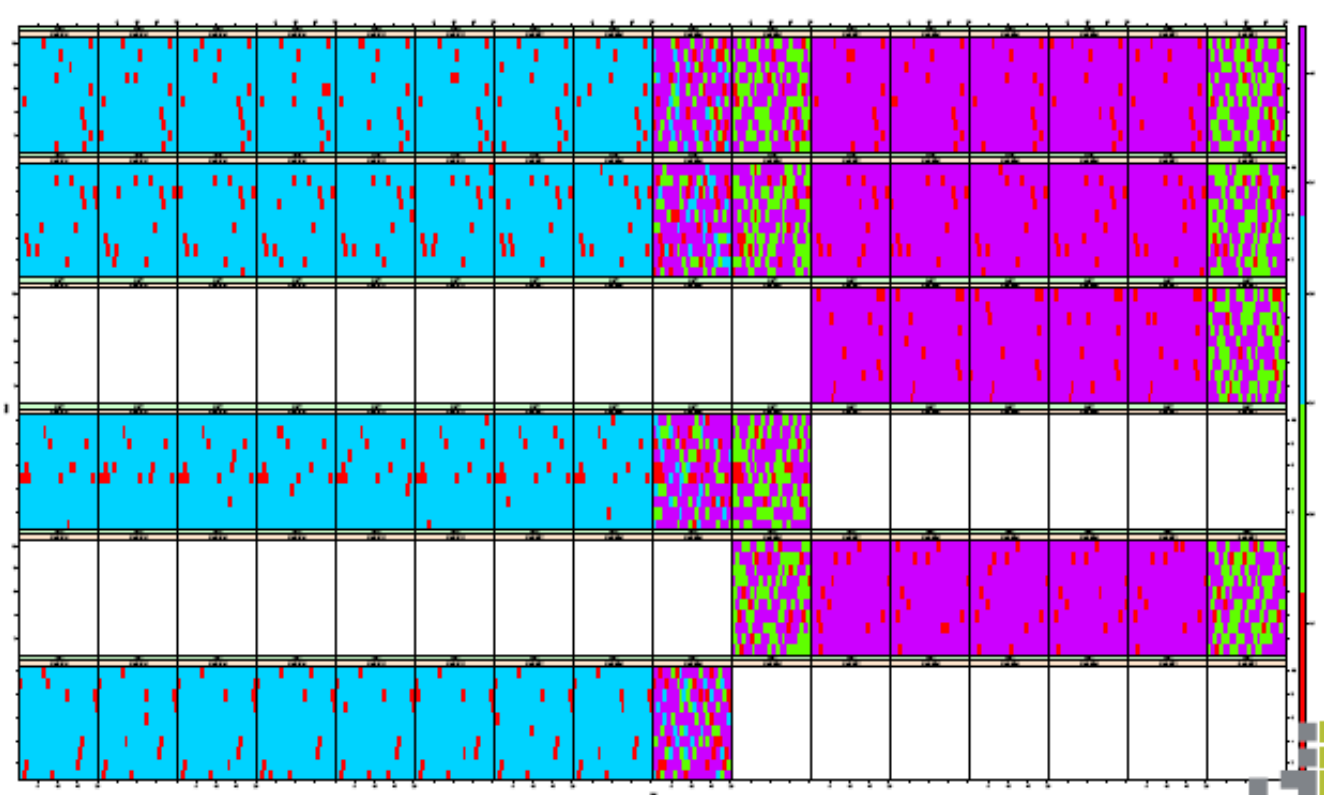
- Test performance for target region
- Trials in large number of environments (ideally random sample from target)

***Standard trial designs for large number of treatments*:**

- Lattice designs, α-designs, row-column designs (Williams and John, 1995)
- Designs with spatial analysis in mind (Cullis et al., 2006; Williams et al., 2006)
- Unreplicated designs with checks, p-rep designs, augmented p-rep designs (Cullis et al., 2006; Williams et al., 2011, 2013)

8. Heritability for unbalanced & correlated data

A multi-location trial in a maize breeding programme (KWS)



6 locations

16 series (laid out as 10 x 10 lattices)

(T. Albrecht, TUM)

8. Heritability for unbalanced & correlated data

Holland et al. (2003, p.64)

Balanced case:

$$\sigma_p^2 = \sigma_g^2 + \sigma_{ge}^2 / m + \sigma^2 / (rm)$$

Divisor of σ_{ge}^2 = no. of environments m

Divisor of σ^2 = no. of plots $p = rm$

Expected genetic gain (EGG):

$$EGG = i \sigma_g H$$

8. Heritability for unbalanced & correlated data

Unbalanced (used with all kinds of incomplete block design):

$$\sigma_p^2 = \sigma_g^2 + \sigma_{ge}^2 / m_h + \sigma^2 / p_h$$

$$m_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{m_i}} \quad , \quad p_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{p_i}}$$

m_i = no. of environments for i^{th} genotype

p_i = no. of plots for i^{th} genotype

n = no. of genotypes

(Holland et al. 2003)

8. Heritability for unbalanced & correlated data

Problems:

- $\bar{y}_{i\bullet\bullet}$ is not the best 'phenotype' (BLUE of $\mu_i = \phi + g_i$) with unbalanced data
- $\sigma^2_{ge} / m_h + \sigma^2 / p_h$ is not ½ variance of a difference,

  no matter if $\bar{y}_{i\bullet\bullet}$ or the BLUE of $\mu_i = \phi + g_i$ is used

  $\Leftrightarrow$ some or many genotype-environment combinations may be missing!

8. Heritability for unbalanced & correlated data

Piepho & Möhring (2007)

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + 0.5\overline{vd}}$$

Rationale:

- $\overline{vd}$ is the average variance of a difference between adjusted means (BLUE of $\mu_i = \phi + g_i$) based on an analysis of MET with fixed genotype effects
  $\Rightarrow$ proportional to *effective error mean square*

8. Heritability for unbalanced & correlated data

Rationale continued

$$\hat{\mu} = 1\phi + g + e$$

ϕ = general mean

$\hat{\mu}$ = vector of adjusted genotype means

g = genotypic effects; $g \sim N(0, G)$; $G = I\sigma_g^2$

e = errors of $\hat{\mu}$; $e \sim N(0, R)$

8. Heritability for unbalanced & correlated data

Assume balanced data from RCBD or CRD & complete G x E classification $\Rightarrow$

$$\bar{y}_{i\bullet\bullet} = \phi + g_i + \bar{e}_{\bullet} + \overline{(ge)}_{i\bullet} + \bar{b}_{\bullet\bullet} + \bar{\varepsilon}_{i\bullet\bullet}$$

$$\sigma_p^2 = \sigma_g^2 + \underbrace{\sigma_{ge}^2 / m + \sigma^2 / (rm)}$$

variance of a mean = ½ variance of a difference

m = number of environments

r = number of replicates per environment

8. Heritability for unbalanced & correlated data

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + 0.5\overline{vd}} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2/m + \sigma^2/(rm)}$$

What has all this got to do with R^2?

Now explicitly take effects for environments and blocks random!

$$\Omega_u = \frac{\theta^{ASV}\left(ZGZ^T\right)}{\theta^{ASV}\left(V_0\right)} = H^2$$

$$\Omega_u = \frac{\theta^{AMV}\left(ZGZ^T\right)}{\theta^{AMV}\left(V_0\right)} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/m + \sigma_{ge}^2/m + \sigma_b^2/(rm) + \sigma^2/(rm)} \neq H^2$$

8. Heritability for unbalanced & correlated data

When genotypic effects are correlated

- So far, genotypic effects were assumed to be i.i.d.
- Often, use pedigree or markers to model genetic covariance

$$\hat{\mu} = 1\phi + g + e$$

ϕ = general mean

$\hat{\mu}$ = vector of adjusted genotype means

g = genotypic effects; $g \sim N(0, G)$; $G = K\sigma_g^2$, K = kinship matrix

e = errors of $\hat{\mu}$; $e \sim N(0, I\sigma_e^2)$

8. Heritability for unbalanced & correlated data

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$?

True when $K = I$!

$$\Rightarrow \text{var}\left(\hat{\mu}_1 - \hat{\mu}_2\right) = 2\sigma_g^2 + 2\sigma_e^2$$

8. Heritability for unbalanced & correlated data

But what if

$$K = \begin{pmatrix} 1 & 0.90 & 0.85 & \cdots & 0.86 \\ & 1 & 0.77 & & 0.95 \\ & & 1 & & 0.98 \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix} \quad ?$$

$$\Rightarrow \operatorname{var}(\hat{\mu}_1 - \hat{\mu}_2) = \operatorname{var}(\hat{\mu}_1) + \operatorname{var}(\hat{\mu}_2) - 2\operatorname{cov}(\hat{\mu}_1, \hat{\mu}_2) = 0.2\sigma_g^2 + 2\sigma_e^2$$

$$\Rightarrow h^2 = \frac{0.1\sigma_g^2}{0.1\sigma_g^2 + \sigma_e^2}$$ for this pair of genotypes!

8. Heritability for unbalanced & correlated data

The next question

In general, both numerator and denominator of pairwise heritabilities vary between pairs

How do we average over pairs?
- Average the pairwise h^2?
- Average numerators and denominator separately

$\Rightarrow$ ongoing work (Paul Schmidt)

8. Heritability for unbalanced & correlated data

Squared correlation between true genotype mean and it's estimator

$$\rho_{M(i)} = \frac{\mathrm{cov}(\mu_i, \hat{\mu}_i)}{\sqrt{\mathrm{var}(\mu_i)\mathrm{var}(\hat{\mu}_i)}}$$

Balanced data, independent genotypes

$$\hat{\mu}_i = \bar{y}_{i\bullet\bullet} = \phi + g_i + \bar{e}_{\bullet} + \overline{(ge)}_{i\bullet} + \bar{b}_{\bullet\bullet} + \bar{\varepsilon}_{i\bullet\bullet}$$

$$\rho_{M(i)} = \frac{\sigma_g^2}{\sqrt{\sigma_g^2\left[\sigma_g^2 + \sigma_e^2/m + \sigma_{ge}^2/m + \sigma_b^2/(rm) + \sigma^2/(rm)\right]}}$$

$$\rho_{M(i)}^2 \neq H^2$$

8. Heritability for unbalanced & correlated data

Squared correlation between true genotype difference and it's estimator

$$\rho_{D(i,i')} = \frac{\mathrm{cov}\left(\mu_i - \mu_{i'}, \hat{\mu}_i - \hat{\mu}_{i'}\right)}{\sqrt{\mathrm{var}\left(\mu_i - \mu_{i'}\right)\mathrm{var}\left(\hat{\mu}_i - \hat{\mu}_{i'}\right)}}$$

Balanced data, independent genotypes

$$\hat{\mu}_i - \hat{\mu}_{i'} = \bar{y}_{i\bullet\bullet} - \bar{y}_{i'\bullet\bullet} = g_i - g_{i'} + \overline{(ge)}_{i\bullet} - \overline{(ge)}_{i'\bullet} + \bar{\varepsilon}_{i\bullet\bullet} - \bar{\varepsilon}_{i'\bullet\bullet}$$

$$\rho_{D(i,i')} = \frac{2\sigma_g^2}{\sqrt{2\sigma_g^2\left[2\sigma_g^2 + 2\sigma_{ge}^2/m + 2\sigma^2/(rm)\right]}}$$

$$\rho_{D(i,i')}^2 = H^2$$

8. Heritability for unbalanced & correlated data

Unbalanced data, correlated genotypes

$\rho^2_{D(i,i')}$ has numerator and denominator varying between pairs (i, i')

$\Rightarrow$ how to average across pairs?

8. Heritability for unbalanced & correlated data

So far all discussion in terms of BLUE

Can also use BLUP, with similar questions arising

When genotypic effects are correlated

⇒ Generalized heritability (Oakey *et al.*, 2006)
⇒ Generalized heritability (Rodríguez-Álvarez *et al.*, 2018)
⇒ Simulation (Piepho & Möhring, 2007)

8. Heritability for unbalanced & correlated data

Generalized heritability (Oakey et al., 2006)

$$y = X\beta + Z_g g + \ldots$$

$y \sim MVN(X\beta, V)$ (observed data vector)

$g = (g_1, g_2, \ldots, g_n)^T \sim MVN(0, G)$

Consider contrast $c^T g$ where c is a contrast vector

$$H^2 = \frac{\left[\mathrm{cov}(c^T g, c^T \hat{g})\right]^2}{\mathrm{var}(c^T g)\,\mathrm{var}(c^T \hat{g})} \qquad \text{where} \quad \hat{g} = BLUP(g)$$

$\Rightarrow$ Find c such that H^2 is maximized

8. Heritability for unbalanced & correlated data

$$H^2 = \frac{\left[\text{cov}\left(c^T g, c^T \hat{g}\right)\right]^2}{\text{var}\left(c^T g\right)\text{var}\left(c^T \hat{g}\right)} = \frac{c^T G Z_g^T P_v Z_g G c}{c^T G c}$$

$$P_v = V^{-1} - V^{-1} X \left(X^T V^{-1} X\right)^{-1} X^T V^{-1}$$

Constraint: $c^T G c = 1$

Method of Lagrange multipliers:

Maximize $c^T G Z_g^T P_v Z_g G c - \lambda\left(c^T G c - 1\right)$

λ = first eigenvalue of $Z_g^T P_v Z_g G$

c = corresponding eigenvector

8. Heritability for unbalanced & correlated data

$$\max_{\boldsymbol{c}} H_c^2 = \lambda$$

is a component of the full heritability

Full set of non-zero eigenvalues: $\lambda_1, \lambda_2, \ldots, \lambda_s$

$\Rightarrow$ full heritability

$\Rightarrow$ eigenvectors $c_1, c_2, \ldots, c_s$ = full set of orthogonal genotype contrasts

Generalized heritability

$$H^2 = \frac{\sum_{h=1}^{s} \lambda_h}{s}$$ (Oakey *et al.*, 2006)

8. Heritability for unbalanced & correlated data

Questions

- Would a breeder select for $c^T g$?
- Why would we allow contrast vector c to be determined by the data?
- What information is contained in the s orthogonal contrasts corresponding to $\lambda_1, \lambda_2, ..., \lambda_s$?

$\Rightarrow$ average of $H^2 = \dfrac{\left[\mathrm{cov}\left(c^T g, c^T \hat{g}\right)\right]^2}{\mathrm{var}\left(c^T g\right)\mathrm{var}\left(c^T \hat{g}\right)}$ across all orthogonal contrasts?

9. Summary

Coefficient of determination for LMM can be defined based on pairwise differences and semivariances

Heritability can also be defined based on pairwise differences and semivariances

In both cases, there are several proposals in the market. Not all of them are easy to interpret and communicate.

I think that pairwise differences are easy to communicate and they make sense

Thanks!

Monte-Carlo simulation

- Can estimate variance and covariance of $\boldsymbol{g}$ and $\hat{\boldsymbol{g}} = BLUP(\boldsymbol{g})$
- From this can simulate many realizations of $(\boldsymbol{g}, \hat{\boldsymbol{g}})$
- Simulate H^2 = squared correlation of $\boldsymbol{g}$ and $\hat{\boldsymbol{g}}$
- Simulate response to selection!
- Simulate anything else you would want to use H^2 for in the balanced case!

(Piepho & Möhring, 2007)

Advantages

- Completely flexible
- Can handle any covariance structure
- Can directly simulate any statistic of interest

Example

- Rapeseed variety trials in Germany
- 120 cultivars (G) tested in 4 years (Y) and at 4 locations (L)
- At some locations, several trials (T) were performed
- The series was rather unbalanced
- Trials were laid out in randomized complete blocks
- Trial means were analyzed by the variance components model

L.Y.T : G + G.L + G.Y + G.L.Y + G.L.Y.T

TABLE 3

Variance component estimates (REML) for example 2 based on model (21)

Term	Variance
G	3.8299
$G \cdot L$	0.2094
$G \cdot Y$	1.9245
$G \cdot L \cdot Y$	5.8699
$G \cdot L \cdot Y \cdot T$	1.8617

TABLE 5

Response to selection for example 2 computed by simulation based on (16), based on an *ad hoc* approach using $\overline{H}^2$, and based on $\overline{H}^2_C$ (CULLIS *et al.* 2006) for different selection fractions $p = n/I$ with $I = 120$

	Response to selection		
n	Based on simulation	*Ad hoc* $\overline{H}^2$	$\overline{H}^2_C$
1	3.364	2.632	2.979
2	3.056	2.455	2.777
3	2.850	2.327	2.635
4	2.693	2.227	2.522
5	2.565	2.145	2.428
10	2.153	1.858	2.103
15	1.899	1.669	1.887
20	1.711	1.520	1.720
30	1.433	1.291	1.461
40	1.221	1.108	1.254
50	1.045	0.952	1.077
60	0.889	0.811	0.918

Example

- Sugar beet
- 26 breeding trials (6 x 6 simple lattices)
- Connected by checks
- 825 entries
- 33 crosses
- Pedigree data available (ad hoc measures do not apply)

T.R + C + Ts : T.R.B + X.G

T = trial, R = replicate, B = block
C = factor separating individual checks from entries
Ts = tester
G = genotype, X = dummy variable (1 for entries, 0 for checks)

TABLE 6

Model fit for two genotypic variance–covariance structures (example 3)

Term	Form of var(g)	
	$\mathbf{I}\sigma_g^2$	$\mathbf{A}\sigma_a^2$
$T \cdot R \cdot B$	901.90	858.88
$X \cdot G$ (independent)	2,745.84	—
$X \cdot G$ (pedigree)	—	1,010.46
Residual	3,134.02	3,626.05
$r^2(g_i, \hat{g}_i)$	0.5970[a]	0.5328
−2 restricted log-likelihood	20,969.4	20,997.2

[a] Up to the fourth decimal place, the same estimate was obtained for $\overline{H}_C^2$ (Cullis *et al.* 2006).

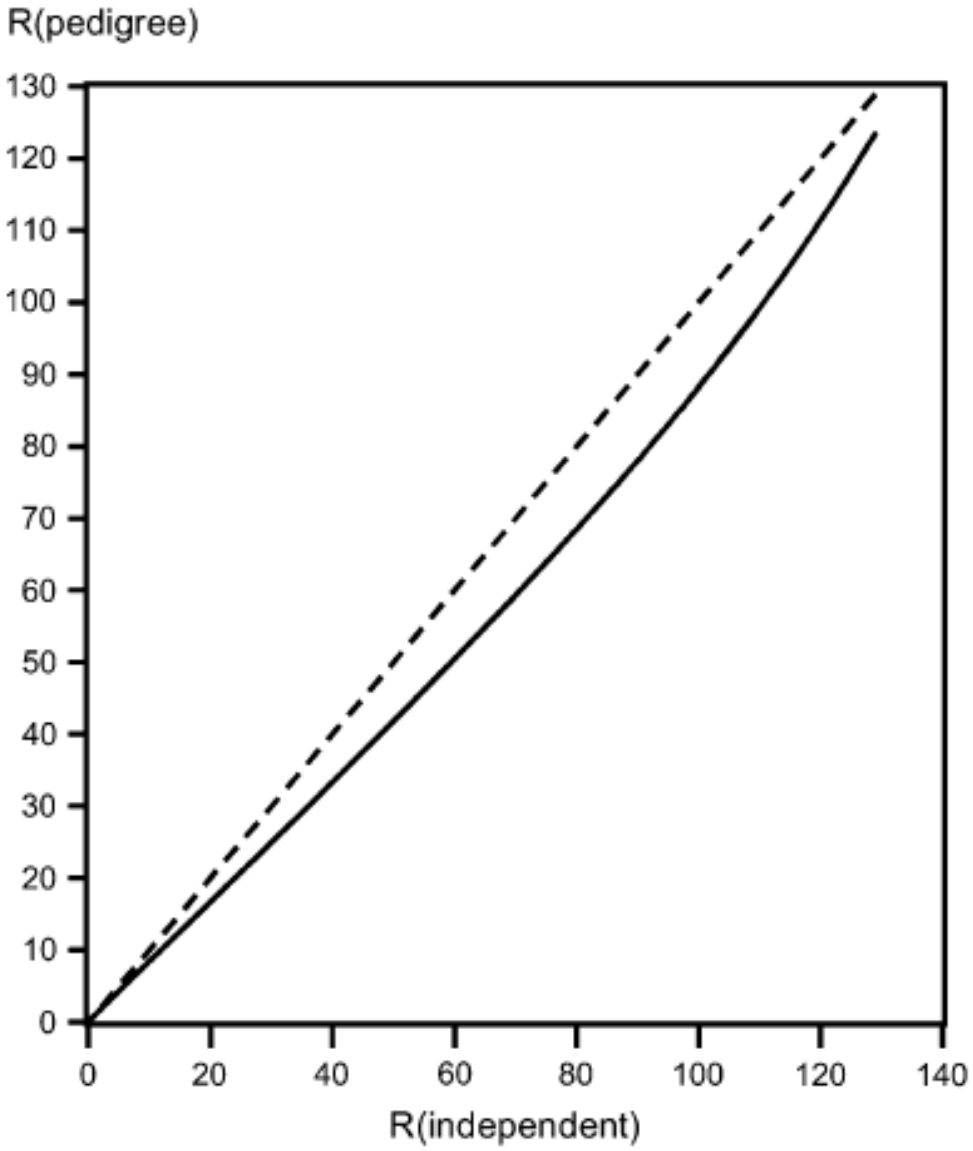


FIGURE 1.—Plot of simulated response to selection under pedigree-based [R(pedigree)] and independent [R(independent)] models for genotypic variance and for all selection fractions. The dashed diagonal indicates equal response to selection (example 3).

References

Bennewitz, J., Böglein, S., Stratz, P., Rodehutscord, M., Piepho, H.P., Kjaer, W., Bessei, W. (2014): Genetic parameters for feather pecking and aggressive behaviour in a large F2-cross of laying hens using generalized linear mixed models. *Poultry Science* **93**, 810-817.

Piepho, H.P. (2018): A coefficient of determination (R^2) for linear mixed models. *Biometrics* (submitted)

Piepho, H.P., Möhring, J. (2007): Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* **177**, 1881-1888.