A coefficient of determination (R^2) for linear mixed models in one go

Hans-Peter Piepho

Biostatistics Unit

Institute of Crop Science

University of Hohenheim

Table of contents

R^2 and adjusted R^2 for linear models (LM)

Full model

$$y = X\beta + e,$$

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

R^2 and adjusted R^2 for linear models (LM)

Null model

$$y = 1_n \lambda + e \quad ,$$

where

1_n = a vector of n ones

λ = intercept

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

R^2 and adjusted R^2 for linear models (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 **null model**:

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

R^2 and adjusted R^2 for linear models (LM)

Coefficient of determination (R^2)

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

Adjusted coefficient of determination (R_{adj}^2)

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

R^2 and adjusted R^2 for linear models (LM)

Coefficient of determination (R^2)

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

Adjusted coefficient of determination (R_{adj}^2)

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

R^2 and adjusted R^2 for linear models (LM)

What does R^2 estimate?

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

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

R^2 and adjusted R^2 for linear models (LM)

For LM

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

$\theta(V) = \sigma_e^2$, and

$\Delta\theta(V, V_0) = \sigma_{e0}^2 - \sigma_e^2$ and hence

$$\Omega_\beta = \frac{\sigma_{e0}^2 - \sigma_e^2}{\sigma_{e0}^2} = 1 - \frac{\sigma_e^2}{\sigma_{e0}^2}$$

R^2 and adjusted R^2 for linear models (LM)

Extensions of R^2

Generalized linear models (GLM): Zhang (2017)

Linear mixed models (LMM): Edwards et al. (2008), Liu et al. (2008),
Demidenko et al. (2012), Schreck & Wiesenfarth (2022)

Generalized linear mixed models (GLMM): Nagakawa and Schielzeth (2013),
Jaeger et al. (2017, 2018), Nakagawa et al. (2017),
Stoffel et al. (2017), Ives (2019), Piepho (2019), Zhang (2022)

⇒ No time to review in detail

⇒ None of these seemed general enough & easy to communicate

R^2 and adjusted R^2 for linear models (LM)

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2013, **4,** 133–142 doi: 10.1111/j.2041-210x.2012.00261.x

A general and simple method for obtaining R^2 from generalized linear mixed-effects models

Shinichi Nakagawa[1,2]* and Holger Schielzeth[3]

[1]*National Centre for Growth and Development, Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9054, New Zealand;* [2]*Department of Behavioral Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard-Gwinner-Straße, 82319 Seewiesen, Germany; and* [3]*Department of Evolutionary Biology, Bielefeld University, Morgenbreede 45, 33615, Bielefeld, Germany*

> 6900 citations and counting on SCOPUS !

> Two problems:

(i) covariance among observations ignored

(ii) bias in estimate of variance explained by fixed effects

The main idea

Data vector

$$y = (y_1, y_2, \ldots, y_n)^T$$ with

$$E(y) = \mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$$ and $$\mathrm{var}(y) = V = \{v_{ij}\}$$

Semivariance

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

(Webster & Oliver, 2007)

The main idea

Average semivariance (ASV)

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

where $P_\lambda = I_n - n^{-1} 1_n 1_n^T$

This is a discrete version of the double integral given in Webster & Oliver (2007), which integrates the semivariance over a defined spatial area:

The main idea

4.8 SUPPORT AND KRIGE'S RELATION

Spatial dependence within a finite region has both theoretical and practical consequences, which we now explore.

The variance of $Z(\mathbf{x})$ within a region R of area $|R|$ is the double integral of the variogram:

$$\sigma_R^2 = \overline{\gamma}(R,R) = \frac{1}{|R|^2}\int_R \int_R \gamma(\mathbf{x}-\mathbf{x}')\,\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}', \qquad (4.23)$$

where $\mathbf{x}$ and $\mathbf{x}'$ sweep independently over R. In geostatistics this variance is called the *dispersion variance* of $Z(\mathbf{x})$ in R. Unless the variogram is

(Webster & Oliver, 2007, *Geostatistics for environmental scientists*. Wiley, p.61)

The main idea

- ASV only captures the total variance in the random-effects part.
- Also need to capture the fixed-effects part

To do so, we here use the expected value of $\frac{1}{2}\left(y_i - y_j\right)^2$, which may be denoted as **expected semi-squared difference**:

$$essd\left(y_i, y_j\right) = \frac{1}{2} E\left[\left(y_i - y_j\right)^2\right] = ssb\left(y_i, y_j\right) + sv\left(y_i, y_j\right)$$

where

$ssb\left(y_i, y_j\right) = \frac{1}{2}\left(\mu_i - \mu_j\right)^2$ is the **semi-squared bias**

The main idea

Average semi-squared bias (ASSB):

$$\theta^{ASSB}(\mu) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ssb(y_i, y_j) = \frac{1}{n-1} \mathrm{trace}(\mu^T P_\lambda \mu)$$

Average expected semi-squared difference (AESSD):

$$\theta^{AESSD}(V, \mu) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} essd(y_i, y_j) = \theta^{ASSB}(\mu) + \theta^{ASV}(V)$$

The main idea

The **average expected semi-squared difference**, $\theta^{AESSD}(V,\mu)$, can be related to the sample variance

$$s_y^2 = \frac{1}{n-1} y^T P_\lambda y$$

It follows from results on quadratic forms that

$$E\left(s_y^2\right) = \theta^{AESSD}(V,\mu)$$

The LM case

The **candidate model** can be written as

$$y = X\beta + e$$

where $e \sim N\left(0, V = I\sigma_e^2\right)$. The error variance can be unbiasedly estimated by

$$\hat{\sigma}_e^2 = \frac{1}{n-p} y^T P_\beta y \quad ,$$

where $p = rank\left(X\right)$, and $P_\beta = I - X\left(X^T X\right)^- X^T$.

The LM case

Average expected semi-squared difference:

$$\theta^{AESSD}\left(V = I\sigma_e^2, \mu = X\beta\right) = \theta^{ASSB}\left(X\beta\right) + \theta^{ASV}\left(I\sigma_e^2\right)$$

The proposed coefficient of determination:

$$\Omega_\beta = \frac{\theta^{ASSB}\left(X\beta\right)}{\theta^{ASSB}\left(X\beta\right) + \theta^{ASV}\left(I\sigma_e^2\right)}$$

We find that $\theta^{ASV}\left(I\sigma_e^2\right) = \sigma_e^2$, which is estimated by

$$\hat{\theta}_{LM}^{ASV}\left(I\sigma_e^2\right) = \hat{\sigma}_e^2$$

The LM case

Naïve estimator of $\theta^{ASSB}(X\beta)$:

$$\frac{1}{n-1}\hat{\beta}^T X^T P_\lambda X\hat{\beta} \qquad \text{with} \qquad \hat{\beta} = \left(X^T X\right)^{-} X^T y$$

$$E\left[\frac{1}{n-1}\hat{\beta}^T X^T P_\lambda X\hat{\beta}\right] = \theta^{ASSB}(X\beta) + \frac{1}{n-1} trace\left[X^T P_\lambda X \operatorname{var}\left(\hat{\beta}\right)\right]$$

$\Rightarrow$ unbiased estimator:

$$\hat{\theta}_{LM}^{ASSB}(X\beta) = \frac{1}{n-1}\hat{\beta}^T X^T P_\lambda X\hat{\beta} - \frac{1}{n-1} trace\left[X^T P_\lambda X \operatorname{var}\left(\hat{\beta}\right)\right]$$

The LM case

Putting it all together:

$$\hat{\theta}_{LM}^{ASSB}\left(X\beta\right)+\hat{\theta}_{LM}^{ASV}\left(I\sigma_e^2\right)=\frac{1}{n-1}\hat{\beta}^T X^T P_\lambda X\hat{\beta}-\frac{\hat{\sigma}_e^2}{n-1}\left(p-1\right)+\hat{\sigma}_e^2=\hat{\sigma}_{e0}^2$$

i.e., the estimator of the residual variance under the null model ⇒

$$\hat{\Omega}_\beta=\frac{\hat{\theta}_{LM}^{ASSB}\left(X\beta\right)}{\hat{\theta}_{LM}^{ASSB}\left(X\beta\right)+\hat{\theta}_{LM}^{ASV}\left(I\sigma_e^2\right)}=1-\frac{\hat{\sigma}_e^2}{\hat{\sigma}_{e0}^2}$$

which is identical with the adjusted R^2 for LM (Draper & Smith, 1998)

Extension to linear mixed models (LMM)

A LMM can be written as

$$y = X\beta + Zu + e$$

with $\mathrm{var}(u) = G$, $\mathrm{var}(e) = R$ and $\mathrm{cov}(u, e) = 0$, such that

$$\mathrm{var}(y) = V = ZGZ^T + R$$

The fixed effects are estimated by

$$\hat{\beta} = \left(X^T V^{-1} X\right)^{-} X^T V^{-1} y$$

Extension to linear mixed models (LMM)

Average expected semi-squared difference, $\theta^{AESSD}(V,\mu)$**:**

$$\theta^{AESSD}(V,X\beta)=\theta^{ASSB}(X\beta)+\theta^{ASV}(V)$$

Coefficient of determination:

$$\Omega_{\beta}=\frac{\theta^{ASSB}(X\beta)}{\theta^{ASSB}(X\beta)+\theta^{ASV}(V)}$$

Extension to linear mixed models (LMM)

The unbiased estimator of $\theta^{ASSB}(X\beta)$ is

$$\hat{\theta}_{LMM}^{ASSB}(X\beta) = \frac{1}{n-1}\hat{\beta}^T X^T P_\lambda X\hat{\beta} - \frac{1}{n-1} trace\left[X^T P_\lambda X \operatorname{var}\left(\hat{\beta}\right)\right]$$

where $\operatorname{var}\left(\hat{\beta}\right) = \left(X^T V^{-1} X\right)^-$.

Need to replace V by $\hat{V}$, its residual maximum likelihood (REML) estimator

$\Rightarrow$ consistent

Extension to linear mixed models (LMM)

Variance explained jointly by random effects u and residual e:

$$\hat{\theta}_{LMM}^{ASV}(V) = \frac{1}{n-1} trace\left(P_\lambda \hat{V}\right)$$

Consistent estimator of the coefficient of determination for LMM:

$$\hat{\Omega}_\beta = \frac{\hat{\theta}_{LMM}^{ASSB}(X\beta)}{\hat{\theta}_{LMM}^{ASSB}(X\beta) + \hat{\theta}_{LMM}^{ASV}(V)}$$

Extension to linear mixed models (LMM)

Coefficient of determination for the variance explained by random effects:

$$\Omega_u = \frac{\theta^{ASV}\left(ZGZ^T\right)}{\theta^{ASSB}\left(X\beta\right) + \theta^{ASV}\left(V\right)}$$

This may be motivated by the partition

$$\theta^{ASV}\left(V\right) = \frac{1}{n-1} trace\left(P_\lambda V\right) = \theta^{ASV}\left(ZGZ^T\right) + \theta^{ASV}\left(R\right)$$

The estimator of $\theta^{ASV}\left(ZGZ^T\right)$ is simply $\theta^{ASV}\left(Z\hat{G}Z^T\right)$.

Extension to linear mixed models (LMM)

The variance explained by both fixed and random effects:

$$\Omega_{\beta u} = \frac{\theta^{ASSB}\left(X\beta\right) + \theta^{ASV}\left(ZGZ^{T}\right)}{\theta^{ASSB}\left(X\beta\right) + \theta^{ASV}\left(V\right)}$$

Extension to linear mixed models (LMM)

Also, $\theta^{ASV}(V)$ can be partitioned according to the component random effects:

$$Zu = Z_1 u_1 + Z_2 u_2$$

with $\mathrm{var}(u_1) = G_1$, $\mathrm{var}(u_2) = G_2$ and $\mathrm{cov}(u_1, u_2) = 0 \quad \Rightarrow$

$$\begin{aligned}
\theta^{ASV}(V) &= \frac{1}{n-1} trace(P_\lambda V) \\
&= \frac{1}{n-1} trace(P_\lambda Z_1 G_1 Z_1^T) + \frac{1}{n-1} trace(P_\lambda Z_2 G_2 Z_2^T) + \frac{1}{n-1} trace(P_\lambda R) \\
&= \theta^{ASV}(Z_1 G_1 Z_1^T) + \theta^{ASV}(Z_2 G_2 Z_2^T) + \theta^{ASV}(R)
\end{aligned}$$

Extension to generalized linear mixed models (GLMM)

A GLMM has linear predictor

$$\eta = X\beta + Zu + f$$

$$\mathrm{var}(f) = R_f$$

The residual random effect f associated with the n units in the linear predictor is optional and may be added to account for overdispersion.

Extension to generalized linear mixed models (GLMM)

The observed data have conditional expectation

$$E(y|\eta) = \mu = g^{-1}(\eta)$$

where $g(.)$ is the link function.

Extension to generalized linear mixed models (GLMM)

The variance takes the general form

$$\mathrm{var}\left(y \middle| \mu\right) = A_\mu^{1/2} R A_\mu^{1/2}$$

A_μ = a diagonal matrix with evaluations of the variance function $\mathrm{var}\left(y_i \mid \mu_i\right)$ on the diagonal

R = a correlation matrix or a covariance matrix if overdispersion needs to be modelled (Wolfinger & O'Connell, 1993; Stroup, 2015)

Extension to generalized linear mixed models (GLMM)

Challenge with GLMMs

- The random model terms occur both on the *linear predictor scale* (via the random effects Zu) and on the *observed scale* (via the conditional distribution of y for given value of the linear predictor η)

- In defining a coefficient of determination, a choice needs to be made as to the scale on which variance is to be assessed

- In either case, the variance occurring on the one scale needs to be projected onto the other scale in order to have a common scale on which to define the coefficient of determination $\Rightarrow$ I am projecting onto the linear predictor scale

Extension to generalized linear mixed models (GLMM)

Extending the linear predictor for the projection:

$$\eta_h = X\beta + Zu + f + h$$

$$\mathrm{var}(h) = R_h$$

h = auxiliary residual term to take up the projection of the residual from the original scale (Nakagawa & Schielzeth, 2013)

$$V_h = ZGZ^T + R_f + R_h$$

Extension to generalized linear mixed models (GLMM)

Use the Taylor-series expansion approach of Foulley et al. (1987) to project the residual variance from the original scale onto h on the linear predictor scale:

$$R_h = D_\eta^{-1} A_\mu^{1/2} R A_\mu^{1/2} D_\eta^{-1}$$

where $D_\eta = diag\left[\partial g^{-1}(\eta)/\partial\eta\right]$

$\Rightarrow$ Particularly easy to compute when model is fitted using pseudo-likelihood (Wolfinger & O'Connell, 1993)

Extension to generalized linear mixed models (GLMM)

Table: Values of diagonal elements of D_η and A_μ for a few examples (m = sample size of binomial distribution).

Link function	D_η	Distribution	A_μ
Logit	$m\mu(1-\mu)$	Binomial	$m\mu(1-\mu)$
Probit[§]	$m\varphi(\eta)$	Binomial	$m\mu(1-\mu)$
Complementary log-log	$m\exp[\eta-\exp(\eta)]$	Binomial	$m\mu(1-\mu)$
Log	μ	Poisson	μ

[§] $\varphi(.)$ is the probability density function of the standard normal distribution

Extension to generalized linear mixed models (GLMM)

Exact results for the binary distribution (binomial distribution with $m = 1$):

Link function	Implied c.d.f.	Variance of h_i
Logit	standard logistic	$\text{var}(h_i) = \pi^2 / 3$
Probit	standard normal	$\text{var}(h_i) = 1$
Complementary log-log	standard extreme value	$\text{var}(h_i) = \pi^2 / 6$

Example 1

- Beetle larvae sampled from 12 populations (Nakagawa and Schielzeth, 2013)
- Within each population, larvae obtained from two microhabitats
- Larvae distinguished as male and female
- Sexed pupae were reared in containers, each holding eight animals

There are three responses:

(i) body length (Gaussian distribution)

(ii) frequency of two male colour morphs (binary distribution)

(iii) the number of eggs laid by each female (Poisson distribution)

Example 1

Linear predictor

Fixed effects: habitat

Random effects: population and container

Distribution, link function and unit variance

Morph frequency:

$\Rightarrow$ binomial, logit link, $\mathrm{var}(h_i) = \pi^2/3$ and $\mathrm{cov}(h_i, h_j) = 0 \ (i \neq j)$

Egg counts:

$\Rightarrow$ Poisson, log link, $\mathrm{var}(h_i) = \mu_i^{-1}$ and $\mathrm{cov}(h_i, h_j) = 0 \ (i \neq j)$

Example 1

Table: Coefficients of determination (%) the beetle data in Nakagawa & Schielzeth (2013)

Trait	Variance parameter estimation method	Coefficients of determination (%)							
		Approach of this paper			Piepho (2019)[§]			[%]Nakagawa & Schielzeth (2013)	
		Ω_{β}	Ω_{u}	$\Omega_{\beta u}$	Ω_{β}	Ω_{u}	$\Omega_{\beta u}$	$R^2_{GLMM(m)}$	$R^2_{GLMM(c)}$
Body length	REML	40.09	33.30	73.39	40.09	33.30	73.39	39.16	74.09
Egg count	Pseudo-likelihood	8.72	[&]43.45	52.17	5.78	[&]44.85	50.63	-	-
Egg count	Laplace	9.13	[&]41.80	50.93	[$]7.21	[$,&]42.68	[$]49.89	9.76	[a]57.23
Colour morph	Laplace	7.46	21.99	29.46	−3.77	24.67	20.89	7.77	31.13

Example 2

Beitler & Landis (1985)

- clinical trial with two treatments (control versus intervention)
- eight clinics, 273 patients
- clinics are regarded as a random sample from a larger target population
- Linear predictor:

  fixed effect: treatment

  random effects: clinic + clinic.treatment
- binomial count y_i of the number of patients responding favourably out of the total number of patients m_i allocated to a treatment in a given clinic
- logit, probit and complementary log-log link
- Gaussian quadrature

Example 2

Patient-level analysis

The rows of the relevant vector and matrices (η, X, Z, R_h, V_h) need to be expanded from the binomial model for grouped data (y_i, m_i) with 16 clinic $\times$ treatment combinations, to represent the binary patient-level response y_{ij}.

$\Rightarrow$ binary inflation

Example 2

	Binomial link function					
	Logit		Probit		Complementary log-log	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Fixed effects:						
Intercept	-0.4574	0.5529	-0.2638	0.3190	-0.8568	0.4208
treatment (control)	-0.7460	0.3247	-0.4434	0.1897	-0.4906	0.2100
Variance components:						
Clinic	1.9632	1.1973	0.6614	0.3900	1.1293	0.6989
Clinic × treatment	0.01102	0.1593	0.003433	0.05692	-	-
Information criteria:						
AIC	82.07		82.31		81.27	
BIC	82.39		82.63		81.51	

Example 2

Coefficient of determination (%):

Level:	Logit Group§	Logit Patient	Probit Group§	Probit Patient	Complementary log-log Group§	Complementary log-log Patient
Ω_{β}	4.66	2.23	5.03	2.52	3.44	1.86
Ω_{u}	71.34	32.76	72.84	34.92	69.21	35.87
$\Omega_{\beta u}$	75.99	34.99	77.87	37.44	72.66	37.74

(Binomial link function: Logit, Probit, Complementary log-log)

§ Group = Clinic × treatment combination

Example 3

Gilmour et al. (1987)

- deformities in the feet of 2,513 lambs
- scored in three ordered categories, denoted as K1, K2 and K3
- lambs represent 34 sires
- Linear predictor:

  random effect: sire

  fixed effects: four contrasts denoted as YR (year), B1, B2 and B3 (breeds)
- binomial model with a probit link, merging either K2 & K3 or K1 & K2
- multinomial, cumulative probits (threshold model)

Example 3

	Binomial model				Multinomial model (proportional odds)	
	K1 vs. K2 & K3		K1 & K2 vs. K3			
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
Fixed effects:						
Intercept	0.3823	0.04968	-	-	0.3781	0.04907
Intercept 2	-	-	1.7558	0.08054	1.6435	0.05930
YR	0.1118	0.04966	0.3462	0.07912	0.1422	0.04834
B1	0.3719	0.07348	0.6069	0.1396	0.3781	0.07154
B2	0.3066	0.1014	0.3617	0.1146	0.3157	0.09709
B3	-0.05571	0.06683	-0.2222	0.07379	-0.09887	0.06508
Sire variance:	0.04974	0.01755	0.04160	0.02528	0.04849	0.01673

Example 3

Coefficient of determination (%):

	Binomial model				Multinomial model (proportional odds)		
	K1 vs. K2 & K3		K1 & K2 vs. K3		Sire [§,$]		Lamb[$]
	Sire[$]	Lamb[$]	Sire[$]	Lamb[$]	K1 vs. K2 & K3	K1 & K2 vs. K3	
Ω_{β}	50.06	5.54	59.54	20.19	54.19	40.83	6.75
Ω_{u}	32.72	4.34	9.06	3.09	29.69	22.37	4.18
$\Omega_{\beta u}$	82.78	9.88	68.60	23.28	83.89	63.20	10.93

Example 3

YR	B1	B2	B3	Ω_{β}	Ω_{u}	$\Omega_{\beta u}$	AIC	BIC
-	-	-	-	0.00	11.58	11.58	3904.63	3912.20
+	-	-	-	1.23	9.99	11.21	3902.56	3908.66
-	+	-	-	6.47	8.25	14.72	3897.08	3903.19
-	-	+	-	3.23	10.25	13.49	3903.50	3909.60
-	-	-	+	-0. 54	11.50	10.97	3906.29	3912.39
+	+	-	-	5.25	6.27	11.53	3891.32	3898.95
+	-	+	-	1.19	9.49	10.68	3902.51	3910.14
+	-	-	+	1.89	9.53	11.42	3903.36	3910.99
-	+	+	-	13.57	5.35	18.92	3888.65	3896.28
-	+	-	+	6.45	8.14	14.59	3898.73	3906.36
-	-	+	+	3.45	10.06	13.51	3905.01	3912.64
+	+	+	-	6.26	4.58	10.84	3884.36	3893.51
+	+	-	+	6.10	5.83	11.93	3891.57	3900.73
+	-	+	+	1.70	9.06	10.76	3903.25	3912.41
-	+	+	+	13.60	5.17	18.76	3889.89	3899.05
+	+	+	+	6.75	4.18	10.92	3884.12	3894.81

Simulation

Single covariate

Simulation scenarios for random-coefficient regression as described in Xu (2003).

$$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i}) z_{1ij} + e_{ij}$$

$(i = 1, \ldots, n; j = 1, \ldots, n_i)$ with $u_{i0} \sim N(0, \tau_0^2)$, $u_{i1} \sim N(0, \tau_1^2)$ and $e_{ij} \sim N(0, \sigma^2)$

The covariate values z_{1ij} were simulated once from a standard normal distribution and this one set of values used in all 1,000 simulation runs for a scenario.

Simulation

β_1	τ_0	τ_1		$n = 50, n_i = 5$				
			$\hat{\Omega}_\beta$	$\hat{\Omega}_u$	$\hat{\Omega}_{\beta u}$	Ω_β	Ω_u	$\Omega_{\beta u}$
0.7	0	0	0.531 (0.036)	0.019 (0.021)	0.551 (0.040)	0.533	0	0.533
	0	0.5	0.417 (0.072)	0.225 (0.057)	0.642 (0.056)	0.419	0.213	0.633
	0	1	0.255 (0.093)	0.525 (0.083)	0.780 (0.045)	0.256	0.521	0.776
	1	0	0.284 (0.042)	0.473 (0.065)	0.757 (0.036)	0.280	0.474	0.754
	1	0.5	0.243 (0.055)	0.539 (0.063)	0.782 (0.035)	0.245	0.540	0.785
	1	1	0.181 (0.072)	0.661 (0.070)	0.843 (0.027)	0.179	0.665	0.844
0.5	1	0	0.167 (0.034)	0.550 (0.062)	0.716 (0.044)	0.166	0.549	0.715
	1	0.5	0.141 (0.048)	0.611 (0.061)	0.752 (0.041)	0.142	0.613	0.756
	1	1	0.097 (0.058)	0.729 (0.060)	0.825 (0.031)	0.100	0.729	0.829

Simulation

Two covariates

$$y_{ij} = \beta_0 + u_{0i} + \left(\beta_1 + u_{1i}\right) z_{1ij} + \left(\beta_2 + u_{2i}\right) z_{2ij} + e_{ij}$$

where $u_{i2} \sim N\left(0, \tau_2^2\right)$

Simulation

β_1	β_2	τ_1	τ_2		$n = 50$, $n_i = 5$				
				$\hat{\Omega}_\beta$	$\hat{\Omega}_u$	$\hat{\Omega}_{\beta u}$	Ω_β	Ω_u	$\Omega_{\beta u}$
2	2	2	2	0.375 (0.063)	0.437 (0.061)	0.812 (0.032)	0.378	0.439	0.817
2	1	1	1	0.429 (0.053)	0.246 (0.054)	0.676 (0.047)	0.432	0.250	0.681
1	1	1	1	0.222 (0.054)	0.335 (0.067)	0.557 (0.063)	0.224	0.341	0.565
0.5	1	1	1	0.147 (0.049)	0.367 (0.071)	0.514 (0.068)	0.148	0.374	0.522
1	1	0.5	1	0.246 (0.051)	0.268 (0.065)	0.514 (0.064)	0.248	0.272	0.520
0.5	0.5	1	1	0.066 (0.039)	0.402 (0.075)	0.468 (0.073)	0.067	0.410	0.477
1	1	0.5	0.5	0.271 (0.047)	0.199 (0.060)	0.470 (0.062)	0.272	0.201	0.473
0.5	0.5	0.5	0.5	0.084 (0.036)	0.249 (0.073)	0.334 (0.074)	0.085	0.252	0.337
0.5	0.5	1	0.5	0.073 (0.038)	0.341 (0.074)	0.414 (0.074)	0.075	0.346	0.421

Summary

- Average semivariance (ASV) is a natural metric for total variance
- Average semi-squared bias (ASSB) is a natural extension of ASV that also includes fixed effects
- ASV and ASSB account for covariance among observations
- It is important to remove bias in the estimation of ASSB
- A coefficient of determination based on ASSB coincides with the adjusted R^2 for LM
- Extension to LMM and GLMM is straightforward
- In GLMM, total variance is assessed on the linear predictor scale
- Simulation shows that estimates of variance explained are accurate

Thanks!

References

Feldmann, M.J., Piepho, H.P., Bridges, W.C., Knapp, S.J. (2021): Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. *PLoS Genetics* **17**, e1009762.

Feldmann, M.J., Piepho, H.P., Knapp, S.J. (2022): Average semivariance directly yields accurate estimates of genomic variance of complex, quantitative traits. *G3* **12(6)**, jkac080.

Feldmann, M.J., Covarrubias-Pazaran, E., Piepho, H.P. (2023): Complex traits and candidate genes: Estimation of genetic variance components across modes of inheritance. *G3* **13**, jkad148.

Piepho, H.P. (2019): A coefficient of determination (R^2) for generalized linear mixed models. *Biometrical Journal* **61**, 860-872.

Piepho, H.P. (2023): A coefficient of determination (R^2) for linear mixed models in one go. *Biometrical Journal* **65**, 2200290.