R^2 as a goodness of fit statistic for mixed models

Matt Kramer

kramerm@ba.ars.usda.gov

Biometrical Consulting Service, ARS/BARC/USDA

Outline

- Introduction
- (Desirable) properties of R^2
- Philosophies for extension into mixed models
- R^2 estimates for examples of mixed models data
- Conclusion

Introduction

- R^2 is often quoted as a measure of goodness of fit, typically as the proportion of variance in the dependent variable that is explained by the model
- It is natural to ask how R^2 changes when adding random effects or spatially correlated residuals
- Current packages don't provide an R^2 statistic for an estimated mixed model

(Desirable) properties of R^2

Kvålseth (1985, Am. Statistician 39, 279–285) proposed the following requirements for R^2

- 1. R^2 must possess utility as a measure of goodness of fit and have an intuitively reasonable interpretation
- 2. R^2 ought to be dimensionless
- 3. $0 \leq R^2 \leq 1$, where $\mathrm{R}^2 = 1$ corresponds to perfect fit, and $R^2 \geq 0$ for any reasonable model specification
- 4. Applicable to (a) any type of model, (b) whether effects are fixed or random, and (c) regardless of the statistical properties of the model variables
- 5. R^2 should not be confined to any specific model-fitting technique

(Desirable) properties of R^2

- 6. Values for different models fit to the same data set are directly comparable
- 7. Generally compatible with other acceptable measures of fit
- 8. Positive and negative residuals weighted equally

(Desirable) properties of R^2

Under the usual regression model, various definitions yield the same numeric result, e.g.,

- $1 - \sum(y-\hat{y})^2 / \sum(y-\bar{y})^2$
- $\sum(\hat{y}-\bar{y})^2 / \sum(y-\bar{y})^2$
- $(\sigma_y^2 - \sigma_{y|x}^2)/\sigma_y^2$
- $1 - \sum(e-\bar{e})^2 / \sum(y-\bar{y})^2$, e is a model residual
- Squared multiple correlation coefficient between the regressand and the regressors
- Squared correlation coefficient between y and $\hat{y}$
- Different definitions of R^2 may yield different quantities when the usual regression model is generalized

(Desirable) properties of R^2

Cameron and Windmeijer (1996, JBES 14, 209–220), to extend the definition to count data, suggest

- 1. $0 \le R^2 \le 1$
- 2. R^2 does not decrease as regressors are added
- 3. R^2 based on residual SS coincides with R^2 based on explained SS
- 4. There is a correspondence between R^2 and a significance test on all slope parameters and between changes in R^2 as regressors are added and significance tests
- 5. R^2 has an interpretation in terms of information content of the data

Philosophies for extension into mixed models

Philosophy 1: R^2 is a measure of between variable effects and should be free of contamination of within variable effects (e.g., autocorrelation due to repeated measures or geographic proximity), otherwise part of the variance of y is explainable by its own past or its neighbors.

Pierce (1979, JASA 74: 901-910) suggests the following form: $R_*^2 = (\sigma_{y|y_*}^2 - \sigma_{y|y_*,x}^2)/\sigma_{y|y_*}^2$, where y_* denotes past or neighboring y. This is similar to the expression for R^2, $(\sigma_y^2 - \sigma_{y|x}^2)/\sigma_y^2$, except that we are now also conditioning on y_*.

Philosophies for extension into mixed models

Philosophy 2: How much better than the mean is a model that predicts y when conditioned on the set of x variables and on past and neighboring values of y?

Magee (1990, Am. Statistician 44: 250–253) suggests developing general R^2 measures based on Wald and likelihood ratio test statistics.

Wald R^2

Wald test: Buse (1973, Am. Statistician 27: 106–108) modifies R^2 as $1 - \frac{\hat{\boldsymbol{u}}'\boldsymbol{V}^{-1}\hat{\boldsymbol{u}}}{(\boldsymbol{Y}-\bar{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y}-\bar{\boldsymbol{Y}})}$, where $\hat{\boldsymbol{u}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$ (i.e. the spatially correlated residuals), $\boldsymbol{V}$ is the variance-covariance matrix of the residuals, and $\bar{\boldsymbol{Y}} = \bar{y}\boldsymbol{1}$.

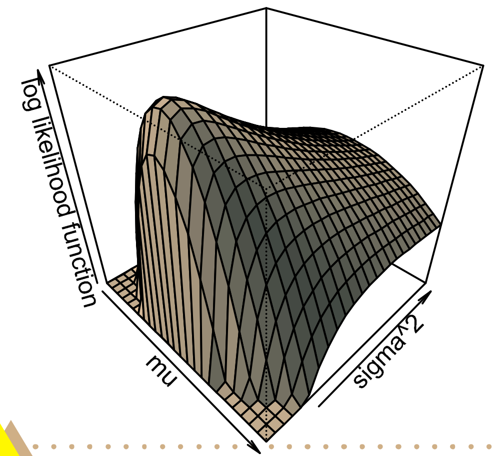
The inverse of $\boldsymbol{V}$ “undoes” the correlation between residuals.

One problem, we don’t have $\boldsymbol{V}$, we only have an estimate of it, and it may not be a very good estimate.

A second problem is that software packages don’t have this expression pre-programmed, to calculate this R^2 would require some work.

Log-likelihood function for a two parm. model (mean and variance)

100 normally distributed samples were generated ($\mu = 0.5, \sigma^2 = 0.025$) and the log-likelihood function plotted for $\hat{\mu} = [0,1]$ and $\hat{\sigma}^2 = [0.05, 2]$



Log likelihood R^2

Likelihood ratio: $R^2_{LR} = 1 - \exp(-\frac{2}{n}(\log L_M - \log L_0))$, where n is the number of observations, $\log L_M$ is the log-likelihood of the model of interest, and $\log L_0$ is the log-likelihood of the intercept-only model.

What is the log-likelihood? The log-likelihood of a statistical model is a function of the data collected and the parameters of the model; the form of this model is assumed known.

It is a special function, the value of the log-likelihood function increases as we reduce the difference between the data and our model for them (we change the value of the log-likelihood function by varying the parameters of the model).

The maximum log-likelihood occurs at those parameter values where this difference is minimized.

Philosophies for extension into mixed models

R^2 based on the likelihood ratio test possesses many desirable properties for a goodness-of-fit statistic

- produces the usual R^2 for ordinary regression (like others)
- since it is based on likelihoods, there is a direct relationship with Kullback-Liebler distance, "information", and information gain $\mathrm{IG} = -\log(1 - R^2_{LR})$ (note that IG is not a linear function of R^2_{LR}) (see Kent (1983), Biometrika 70: 163–174)
- it is easily calculated using output from mixed models software

R^2 examples

Ex. 1: RCBD + covariate (random coefficients, 3 treatments, 4 blocks, 2 obs/block-trt combination, $\sigma^2_{\beta_0} = 4$, $\sigma^2_{\beta_1} = 1$, $\sigma_{\beta_0,\beta_1} = 0$, $\sigma^2 = 1$)

blk 1

A	B
C	A
B	C

blk 2

B	C
B	A
A	C

blk 3

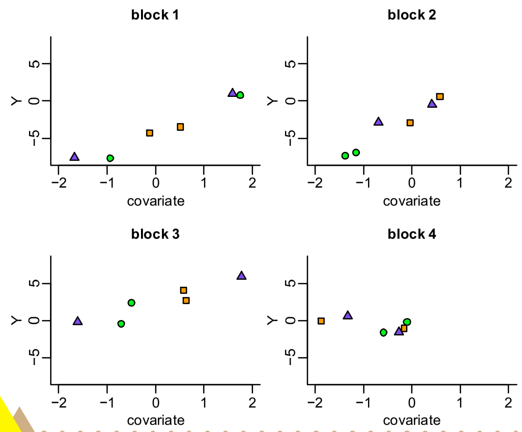
B	C
A	C
A	B

blk 4

B	A
C	C
A	B

R^2 examples

Ex. 1: RCBD + covariate (random coefficients, 3 treatments, 4 blocks, 2 obs/block-trt combination, $\sigma^2_{\beta_0} = 4$, $\sigma^2_{\beta_1} = 1$, $\sigma_{\beta_0,\beta_1} = 0$, $\sigma^2 = 1$)



R^2 examples

R program used for simulating data and estimating the maximum log likelihood (with *nmle* package by Bates and Pinheiro)

model	parms	log likelihood	R^2_{LR}	R^2_W
intercept only	2	-64.45	0	0
trt	4	-63.55	0.07	0.07
trt + cov (f)	5	-59.10	0.36	0.36
trt + blk (r)	5	-60.60	0.27	0.32
trt + blk (r) + cov (r)	7	-42.74	0.84	0.93

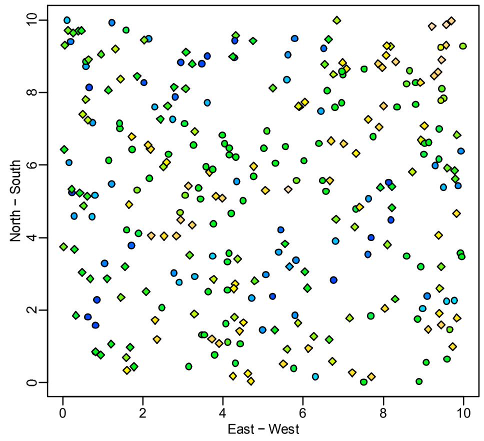
f = fixed
r = random

Spatial exponential correlation

Ex. 2: $\rho = \exp(-d_{i,j}/2)$, $\sigma^2 = 1$, level effect $= 2$, $d =$ distance between i and j

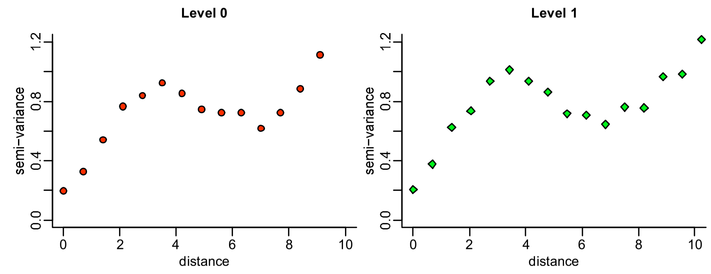
circle = level 0
diamond = level 1

topograhic colors
(blue = lowest values, light brown = highest values)



R^2 examples

Example 2: Semi-variograms



R^2 examples

model	log likelihood	R^2_{LR}
intercept only	-495.94	0
level	-389.68	0.51
level + corr. resid.	-225.27	0.67

Conclusions

- there are various R^2's that can be developed for mixed models, all produce the same value for ordinary regression
- an R^2 based on the likelihood ratio test is easy to calculate from standard mixed models output and has a connection to information theory
- examples were shown demonstrating increases in R^2 when adding random effects or correlated errors to the model