Generalized R^2 in Linear Mixed Models

Julia Volaufova, Lynn R. LaMotte, and Ondrej Blaha

Biostatistics Program
LSU Health
New Orleans, Louisiana, USA

LinStat 2014
Linköping, Sweden, August 24 – 28, 2014

"Full model":

$(\boldsymbol{Y}, \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 I)$,

- $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{X}_1)$: known $n \times (p+1)$-model matrix;
- $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$ unknown fixed $p+1$-vector;
- $\sigma^2 > 0$: unknown variance parameter;

$\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} = P_{\boldsymbol{X}}\boldsymbol{Y}$: orthogonal projection of $\boldsymbol{Y}$ onto $R(\boldsymbol{X})$;

$$\hat{\sigma}^2 = \frac{1}{n - r(\boldsymbol{X})}(\boldsymbol{Y} - \widehat{\boldsymbol{X}\beta})'(\boldsymbol{Y} - \widehat{\boldsymbol{X}\beta}).$$

"Null model" – intercept only model:

$(\boldsymbol{Y}, \beta_0 \mathbf{1}, \sigma^2 I)$,

$\hat{\boldsymbol{Y}}_0 = \hat{\beta}_0 \mathbf{1} = \bar{Y} \mathbf{1}$;

$$\hat{\sigma^2}_0 = \frac{1}{n-1} (\boldsymbol{Y} - \bar{Y}\mathbf{1})'(\boldsymbol{Y} - \bar{Y}\mathbf{1}).$$

...measure of proportion of variability explained by the model;
...measure of goodness of fit;

$$R^2 = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)}.$$

Extension for
$\mathrm{cov}\,(\boldsymbol{Y}) = \sigma^2\boldsymbol{V}$, $\boldsymbol{V}$ known p.d. matrix,
transform $\boldsymbol{Y} \to \boldsymbol{V}^{-1/2}\boldsymbol{Y}$ the rest follows...

General form of R^2:

$$R^2 = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})/n}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)/n}.$$

General form of R^2:

$$R^2 = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})/n}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)/n}.$$

$$R^2_{adj} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})/(n - r(\boldsymbol{X}))}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)/(n-1)}.$$

General form of R^2:

$$R^2 = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})/n}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)/n}.$$

$$R^2_{adj} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})/(n - r(\boldsymbol{X}))}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)/(n-1)}.$$

Willett-Singer (1988), consider Euclidean distance:

$$R^2_{pseudo} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)};$$

Add:

$\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V})$,

If F_p is the F-statistic testing $H_0 : \boldsymbol{\beta}_1 = 0_p$,

$$R^2 = \frac{F_p p/(n - r(\boldsymbol{X}))}{1 + F_p p/(n - r(\boldsymbol{X}))}.$$

Add:

$\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{V})$,

If F_p is the F-statistic testing $H_0 : \boldsymbol{\beta}_1 = 0_p$,

$$R^2 = \frac{F_p p/(n - r(\boldsymbol{X}))}{1 + F_p p/(n - r(\boldsymbol{X}))}.$$

Alternatively,

$$R^2 = 1 - \left(\frac{L_0(\hat{\beta}_0, \hat{\sigma}_0^2)}{L(\hat{\beta}, \hat{\sigma}^2)} \right)^{2/n},$$

$L(., ..)$ - denotes the likelihood under the full, and $L_0(., ..)$ under the null model.

Desire to extend the definition of R^2 using the same principle - a measure of distance in the sample space;

– assess a model fit to data;

– express proportion of variability?

Desire to extend the definition of R^2 using the same principle - a measure of distance in the sample space;

– assess a model fit to data;

– express proportion of variability?

Notation:

Desire to extend the definition of R^2 using the same principle - a measure of distance in the sample space;

– assess a model fit to data;

– express proportion of variability?

Notation:

N sampling units, n_i observations on each, $n = \sum_{i=1}^{N} n_i$;

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\gamma_i + \epsilon_i, i = 1, 2, \ldots, N;$$

$$\begin{pmatrix} \gamma_i \\ \epsilon_i \end{pmatrix} \sim N_{m+n_i}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{\gamma_i}(\tau_\gamma) & 0 \\ 0 & \boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon) \end{pmatrix}\right),$$

Combine all vectors stacking them one below the other, combine the corresponding matrices appropriately:

$\boldsymbol{Y}, \quad \boldsymbol{X}, \quad \boldsymbol{Z}, \quad \gamma, \quad \epsilon;$

$\tau = (\tau'_{\boldsymbol{\gamma}}, \tau'_{\boldsymbol{\epsilon}})';$

$\boldsymbol{\Sigma}(\tau) \equiv \text{cov}\,(\boldsymbol{Y}) = \text{Diag}\,\{\boldsymbol{Z}_i \Sigma_{\gamma_i}(\tau) \boldsymbol{Z}'_i + \boldsymbol{\Sigma}_{\epsilon_i}(\tau)\}.$

A lot of good suggestions...

- Snijders and Bosker (1994), express the proportion of "modeled variance" as opposed to "explained":

$$\boldsymbol{\Sigma}_{\epsilon_i}(\tau_{\epsilon}) = \sigma^2 I_{n_i};$$

A lot of good suggestions...

- Snijders and Bosker (1994), express the proportion of "modeled variance" as opposed to "explained":

$$\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon) = \sigma^2 I_{n_i};$$

Null model:

$$\boldsymbol{Y}_i = \beta_0 \mathbf{1}_{n_i} + \gamma_{i0} \mathbf{1}_{n_i} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N;$$

A lot of good suggestions...

- Snijders and Bosker (1994), express the proportion of "modeled variance" as opposed to "explained":

$$\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon) = \sigma^2 I_{n_i};$$

Null model:

$$\boldsymbol{Y}_i = \beta_0 \mathbf{1}_{n_i} + \gamma_{i0}\mathbf{1}_{n_i} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N;$$

$$\mathrm{cov}_0(\boldsymbol{Y}_i) = \tau_{\gamma_{i0}} \mathbf{1}_{n_i}\mathbf{1}'_{n_i} + \sigma^2 I_{n_i}.$$

A lot of good suggestions...

- Snijders and Bosker (1994), express the proportion of "modeled variance" as opposed to "explained":

  $$\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon) = \sigma^2 I_{n_i};$$

  Null model:

  $$\boldsymbol{Y}_i = \beta_0 \mathbf{1}_{n_i} + \gamma_{i0}\mathbf{1}_{n_i} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, N;$$

  $$\mathrm{cov}_0(\boldsymbol{Y}_i) = \tau_{\gamma_{i0}} \mathbf{1}_{n_i}\mathbf{1}'_{n_i} + \sigma^2 I_{n_i}.$$

  ... R^2 defined, based on comparison of

  $\hat{\mathrm{cov}}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$ in full model and $\hat{\mathrm{cov}}(\boldsymbol{Y}_i - \beta_0\mathbf{1}_{n_i})$ in null model,

  averaged across observations on the sampling unit.

- Vonesh and Chinchilli (1997):

$$R^2_{VC} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)},$$

$\hat{\boldsymbol{Y}}_0$: predicted $\boldsymbol{Y}$ under null model;
$\boldsymbol{V}$ some p.d. matrix;

- Vonesh and Chinchilli (1997):

$$R^2_{VC} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)},$$

$\hat{\boldsymbol{Y}}_0$: predicted $\boldsymbol{Y}$ under null model;
$\boldsymbol{V}$ some p.d. matrix;

  - What to choose for $\boldsymbol{V}$?
    - $\boldsymbol{V} = I$?
    - $\boldsymbol{V} = \boldsymbol{\Sigma}(\hat{\tau})$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}(\hat{\tau}_{\boldsymbol{\epsilon}})\}$?

- Vonesh and Chinchilli (1997):

$$R^2_{VC} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)},$$

$\hat{\boldsymbol{Y}}_0$: predicted $\boldsymbol{Y}$ under null model;
$\boldsymbol{V}$ some p.d. matrix;

  - What to choose for $\boldsymbol{V}$?
    - $\boldsymbol{V} = I$?
    - $\boldsymbol{V} = \boldsymbol{\Sigma}(\hat{\tau})$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}(\hat{\tau}_{\boldsymbol{\epsilon}})\}$?
  - What to use for $\hat{\boldsymbol{Y}}$?
    - "Conditional model": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X\beta}} + \boldsymbol{Z}\hat{\gamma}$;
    - "Marginal model": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X\beta}}$.

- Vonesh and Chinchilli (1997):

$$R^2_{VC} = 1 - \frac{(\boldsymbol{Y} - \hat{\boldsymbol{Y}})'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})}{(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)'\boldsymbol{V}^{-1}(\boldsymbol{Y} - \hat{\boldsymbol{Y}}_0)},$$

  $\hat{\boldsymbol{Y}}_0$: predicted $\boldsymbol{Y}$ under null model;
  $\boldsymbol{V}$ some p.d. matrix;
  - What to choose for $\boldsymbol{V}$?
    - $\boldsymbol{V} = I$?
    - $\boldsymbol{V} = \boldsymbol{\Sigma}(\hat{\tau})$?
    - $\boldsymbol{V} = \mathrm{Diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}(\hat{\tau}_{\boldsymbol{\epsilon}})\right\}$?
  - What to use for $\hat{\boldsymbol{Y}}$?
    - "Conditional model": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\boldsymbol{\beta}} + \boldsymbol{Z}\hat{\gamma}$;
    - "Marginal model": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\boldsymbol{\beta}}$.
  - Null model:

$$\boldsymbol{Y} = \beta_0\mathbf{1} + \boldsymbol{\epsilon};$$

  - If $\boldsymbol{V} = \mathrm{Diag}\left\{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}(\hat{\tau}_{\boldsymbol{\epsilon}})\right\}$, R^2_{VC} identical to R^2 suggested by Kramer (2005).

- Xu (2003): proportional reduction in conditional residual variance explained by the model;

  $\mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon)\} = \sigma^2 \boldsymbol{I};$

  Null models considered

  - $\boldsymbol{Y} = \beta_0 \mathbf{1} + \boldsymbol{\epsilon}$ – the same as Vonesh and Chinchilli (1997);
  - $\boldsymbol{Y} = \beta_0 \mathbf{1} + \mathrm{Diag}\{\mathbf{1}_{n_i}\}\mathrm{Col}\{\gamma_{i0}\} + \boldsymbol{\epsilon}$ – the same as Snijders and Bosker (1994);

  Compares conditional variances
  $\mathrm{var}(Y_{ij}|\boldsymbol{X}, \boldsymbol{\gamma})$ and $\mathrm{var}(Y_{ij})$ (or $\mathrm{var}(Y_{ij}|\gamma_{i0})$).

- Edwards et al. (2008): Null model differs from full only in fixed effects:

$$\boldsymbol{Y} = \beta_0 \mathbf{1} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

Let $C = (0_p, I_p)$, $\quad H_0 : C\boldsymbol{\beta} \equiv \boldsymbol{\beta}_1 = 0_p$.

$$F_p = \frac{1}{p} C\hat{\boldsymbol{\beta}}' \left[\widehat{\mathrm{cov}}\, C\hat{\boldsymbol{\beta}} \right]^{-1} C\hat{\boldsymbol{\beta}},$$

the basis for the approximate F-test of H_0;

- Edwards et al. (2008): Null model differs from full only in fixed effects:

$$\boldsymbol{Y} = \beta_0 \mathbf{1} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

Let $C = (0_p, I_p), \quad H_0 : C\boldsymbol{\beta} \equiv \boldsymbol{\beta}_1 = 0_p$.

$$F_p = \frac{1}{p} C\hat{\boldsymbol{\beta}}' \left[\widehat{\mathrm{cov}}\, C\hat{\boldsymbol{\beta}} \right]^{-1} C\hat{\boldsymbol{\beta}},$$

the basis for the approximate F-test of H_0; Extension from linear fixed effects model R^2:

$$R_E^2 = \frac{p/\nu F_p}{1 + p/\nu F_p}.$$

ν: denominator degrees of freedom (Satterthwaite, Kenward-Roger, etc.).

Property:

$0 \leq R_E^2 \leq 1$;

But - ν depends on estimated variance components.

Several others:

- Gelman and Pardoe (2006): Bayesian R^2;
- Magee (1990): R^2 based on log-likelihood, null model contains only fixed intercept;
- Zheng (2000) for generalized linear models based on proportions of deviances;
- etc.

Hodges (1998), Vaida and Blanchard (2005), Arendacká and Puntanen (2014):

Assume:

- $\boldsymbol{\Sigma}_{\epsilon_i}(\tau_{\epsilon}) = \sigma^2 I_{n_i}$, $i = 1, \ldots, N$;
- $\boldsymbol{\Sigma}_{\gamma_i}(\tau_{\gamma}) = \sigma^2 \boldsymbol{G}_i$, $\boldsymbol{G}_i$ known p.d. matrix.

Hodges (1998), Vaida and Blanchard (2005), Arendacká and Puntanen (2014):

Assume:

- $\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon) = \sigma^2 I_{n_i}$, $i = 1, \ldots, N$;
- $\boldsymbol{\Sigma}_{\gamma_i}(\tau_\gamma) = \sigma^2 \boldsymbol{G}_i$, $\boldsymbol{G}_i$ known p.d. matrix.

Augmented model:

$$\boldsymbol{Y}^* \equiv \begin{pmatrix} \boldsymbol{Y} \\ 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ 0 & -I_{Nm} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} + \begin{pmatrix} \epsilon \\ \gamma \end{pmatrix},$$

$$\text{cov}\left(\begin{array}{c}\epsilon\\ \gamma\end{array}\right)=\sigma^2\left(\begin{array}{cc}I_n & 0\\ 0 & \boldsymbol{G}\end{array}\right);$$

$$\text{diag}\left\{\boldsymbol{G}_i\right\}=\boldsymbol{G}=(\Delta'\Delta)^{-1}.$$

Let

$$\Gamma=\left(\begin{array}{cc}I_n & 0\\ 0 & \Delta\end{array}\right).$$

Following Hodges (1998), Vaida and Blanchard (2005), Arendacká and Puntanen (2014):

$$\Gamma \boldsymbol{Y}^* = \boldsymbol{Y}^* = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ 0 & -\Delta \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \epsilon \\ \Delta\gamma \end{pmatrix},$$

$$\operatorname{cov} \begin{pmatrix} \epsilon \\ \Delta\gamma \end{pmatrix} = \sigma^2 I.$$

LS solutions result in $\boldsymbol{X}\hat{\beta}$ (BLUE) and $\boldsymbol{Z}\hat{\gamma}$ (BLUP) (in the sense of Harville (1977));

Null model:

$$\boldsymbol{Y}^* = \begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \boldsymbol{\epsilon}^*, \quad \operatorname{cov}(\boldsymbol{\epsilon}^*) = \sigma^2 I;$$

Define R^2_{aug} as in a fixed effects model:

$$R^2_{aug} = 1 - \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta} - \boldsymbol{Z}\hat{\gamma})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta} - \boldsymbol{Z}\hat{\gamma}) + \hat{\gamma}'\boldsymbol{G}^{-1}\hat{\gamma}}{(\boldsymbol{Y} - \bar{Y}\boldsymbol{1})'(\boldsymbol{Y} - \bar{Y}\boldsymbol{1})}.$$

Define R^2_{aug} as in a fixed effects model:

$$R^2_{aug} = 1 - \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta} - \boldsymbol{Z}\hat{\gamma})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta} - \boldsymbol{Z}\hat{\gamma}) + \hat{\gamma}'\boldsymbol{G}^{-1}\hat{\gamma}}{(\boldsymbol{Y} - \bar{Y}\mathbf{1})'(\boldsymbol{Y} - \bar{Y}\mathbf{1})}.$$

Properties

- Under normality, with estimated G coincides with R^2 in Zheng (2000);
- $0 \leq R^2_{aug} \leq 1$;
- R^2_{aug} is increasing when adding columns into X or Z matrices;

Define R^2_{aug} as in a fixed effects model:

$$R^2_{aug} = 1 - \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta} - \boldsymbol{Z}\hat{\gamma})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta} - \boldsymbol{Z}\hat{\gamma}) + \hat{\gamma}'\boldsymbol{G}^{-1}\hat{\gamma}}{(\boldsymbol{Y} - \bar{Y}\boldsymbol{1})'(\boldsymbol{Y} - \bar{Y}\boldsymbol{1})}.$$

Properties

- Under normality, with estimated G coincides with R^2 in Zheng (2000);
- $0 \leq R^2_{aug} \leq 1$;
- R^2_{aug} is increasing when adding columns into X or Z matrices;

Question: is R^2_{aug} for estimated G also monotone set function?

Alternative choice of null model:

$$Y^* = \begin{pmatrix} \mathbf{1} & 0 & Z \\ 0 & 0 & -\Delta \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} + \boldsymbol{\epsilon}^*, \quad \mathrm{cov}\,(\boldsymbol{\epsilon}^*) = \sigma^2 I.$$

Suggested:

$$R^2_{aug2} = 1 - \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\gamma})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\gamma}) + \hat{\gamma}'\boldsymbol{G}^{-1}\hat{\gamma}}{(\boldsymbol{Y} - \mathbf{1}\hat{\beta}_0 - \boldsymbol{Z}\hat{\gamma}_0)'(\boldsymbol{Y} - \mathbf{1}\hat{\beta}_0 - \boldsymbol{Z}\hat{\gamma}_0) + \hat{\gamma}_0'\boldsymbol{G}^{-1}\hat{\gamma}_0}.$$

- Monotone set function in X;
- $0 \le R^2_{aug2} \le 1$;
- Takes into consideration dependencies between observations also in the null model;
- For unknown G, we recommend the estimated variance-covariance components from the full model in both, numerator and denominator.

Orelien and Edwards (2008): compare model fit for fixed effects only - only models and sub-models compared;

Orelien and Edwards (2008): compare model fit for fixed effects only - only models and sub-models compared;

Goal 1:

- Investigate monotonicity of R^2 with increasing number of fixed effects;
- Among models with 2 fixed effect variables, identify the "true" model (with the highest R^2);

Setting:

Data generated from:

- balanced design with respect to sample 2 groups ("trt");
- unequal number of time points per sampling unit (from 2 up to 8);
- random "intercept" and "time" coefficients model;
- "trt" additional dichotomous fixed effects variable;

- $n = 64$, $\sigma^2 \in \{3, 6, 9, 12, 15, 45\}$;
- G matrix unstructured;
- REML used to estimate G and σ^2 in full and null models;
- 10000 simulations for different configurations;
- Important: in all 10000 cases convergence was achieved and the estimated G was n.n.d.
- SAS version 9.4 used for all calculations.

Variables unrelated to the response: “genr” (dichotomous), x_5, and x_6 (transformed uniform);

Compared models:

Differ in fixed effects only:

- “full”: time, trt, genr, x_5, x_6
- “true”: time, trt;
- “genr”: time, genr;
- “x_5”: time, x_5;
- “x_6”: time, x_6;
- “reduced”: time;

Compared R^2s:

- "VC": Vonesh and Chinchilli (1997), $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} + \boldsymbol{Z}\hat{\gamma}$;
- "VCm": the same but $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta}$;
- "R^2_{aug}": (same as Zheng (2000)): $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} + \boldsymbol{Z}\hat{\gamma}$;
- "R^2_{aug}m": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta}$;
- "R^2_{aug2}";
- "R^2_{aug2}m";

R^2	$\sigma^2 = 3$	$\sigma^2 = 12$	$\sigma^2 = 45$
"VC"	0.003	0.12	0.19
"VCm"	0.86	0.92	0.91
"R^2_{aug}"	0.50	0.64	0.59
"R^2_{aug}m"	0.87	0.93	0.94
"R^2_{aug2}"	0.40	0.34	0.40
"R^2_{aug2}m"	0.76	0.59	0.63

Table : Proportion of R^2 from "full" higher than all others

Correct model: proportion of times the R^2 for the correct model is the highest among all (except full);

R^2	$\sigma^2 = 3$	$\sigma^2 = 12$	$\sigma^2 = 45$
"VC"	0.002	0.11	0.34
"VCm"	1.00	1.00	1.00
"R^2_{aug}"	0.48	0.61	0.69
"R^2_{aug}m"	1.00	1.00	1.00
"R^2_{aug2}"	0.85	0.67	0.72
"R^2_{aug2}m"	1.00	0.90	0.90

Table : Proportion of R^2 from "true" higher than all others (except "full")

Goal 2:

- Among models with varying random effects identify the "true" model (with the highest R^2);

Data generated from:

- The same G; balanced treatment groups, generated unequal time points as above;
- "intercept" and "time" fixed effects,
- random coefficients "intercept" and of "time";

Compared models:

Differ in random effects only:

- “true”: int, t;
- “int”: int;
- “t2”: int, t^2;

Compared R^2s:

- “VC”: $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} + \boldsymbol{Z}\hat{\gamma}$;
- “VCm”: $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta}$;
- “R^2_{aug}”: (same as Zheng (2000)): $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} + \boldsymbol{Z}\hat{\gamma}$;
- “R^2_{aug}m”: $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta}$;

Results – random effects – true model identification

R^2	$\sigma^2 = 3$	$\sigma^2 = 12$	$\sigma^2 = 45$
"VC"	1.00	0.83	0.51
"VCm"	0.30	0.30	0.27
"R^2_{aug}"	1.00	0.75	0.52
"R^2_{aug}m"	0.32	0.42	0.32

Table : Proportion of R^2 from "true" higher than all others (except "full")

R^2	$\sigma^2 = 3$	$\sigma^2 = 12$	$\sigma^2 = 45$
"VC"	1.00	0.83	0.51
"VCm"	0.30	0.30	0.27
"R^2_{aug}"	1.00	0.75	0.52
"R^2_{aug}m"	0.32	0.42	0.32
"AIC"	0.76	0.55	0.32

Table : Proportion of R^2 from "true" higher than all others (except "full")

Conclusions

- For identifying model fit in models differing in fixed effects only, “VCmu”, “R^2_{aug}mu”, and “R^2_{aug2}” performed better than “VC” and “R^2_{aug}”;
- On the other hand, to identify model fit with respect to random effects, “VC” and “R^2_{aug}” had higher proportion of correct picks;
- In models in which $\widehat{\boldsymbol{X}\beta}$ coincides between models, R^2 with $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta}$ instead of $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} + \boldsymbol{Z}\hat{\gamma}$ does not differentiate models.

Still a lot has to be investigated ...

Still a lot has to be investigated ...

Thank you for your attention!