On R^2 in linear mixed models

Julia Volaufova, Ondrej Blaha, and Lynn R. LaMotte

Biostatistics Program
LSU Health Sciences Center
New Orleans, Louisiana, USA

23rd IWMS
Ljubljana, Slovenia, August 8 – 12, 2014

Topics to talk about

- R^2 in a fixed effects linear model;

Topics to talk about

- R^2 in a fixed effects linear model;
- Null model specification and R^2 definitions in linear mixed model;

- R^2 in a fixed effects linear model;
- Null model specification and R^2 definitions in linear mixed model;
- New improved R^2?

"Full model":

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

- $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{X}_1)$: 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;
- $\boldsymbol{V}$: known p. d. matrix;

Notation:

the $\boldsymbol{V}$-inner product in R^n:

$$\langle x, y\rangle_{\boldsymbol{V}} = x'\boldsymbol{V}^{-1}y,$$

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

$$\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} = P_{\boldsymbol{X}}\boldsymbol{Y}.$$

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

"Null model" – intercept only model:

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

$\hat{\beta}_0 \mathbf{1} = P_{\mathbf{1}} \boldsymbol{Y} = \hat{\boldsymbol{Y}}_0$: (GLSE or weighted LS);

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

R^2 in linear fixed effects model

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

...measure of goodness of fit, etc.

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

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

...measure of goodness of fit, etc.

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

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

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

...measure of goodness of fit, etc.

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

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

Willett-Singer (1988), consider Euclidean distance:

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

For $V = I$ no controversy...

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 normal likelihood under the full, and $L_0(.,..)$ under the null model.

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

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

Notation partially from Edwards et al. (2008):

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

Notation partially from Edwards et al. (2008):

N sampling units, n_i observations on each,

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

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

$$\text{cov}\,(\boldsymbol{Y}_i) \equiv \boldsymbol{\Sigma}_i(\tau) = \boldsymbol{Z}_i\boldsymbol{\Sigma}_{\gamma_i}(\tau)\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{\epsilon_i}(\tau), \quad \tau = (\tau_\gamma', \tau_\epsilon')'.$$

$n = \sum_{i=1}^{N} n_i$;

Combine all vectors stacking them and combine the corresponding matrices appropriately:

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

$$\Sigma(\tau) \equiv \mathrm{cov}(\boldsymbol{Y}) = \mathrm{Diag}\{\boldsymbol{Z}_i \Sigma_{\gamma_i}(\tau)\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}(\tau)\}, \quad \tau = (\tau_{\boldsymbol{\gamma}}', \tau_{\boldsymbol{\epsilon}}')'.$$

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

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

Question: What to use for null model?

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

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

  Null model:

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

Question: What to use for null model?

- 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:

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

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

Question: What to use for null model?

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

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

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

$\widehat{\mathrm{cov}}(\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})$ in full model and $\widehat{\mathrm{cov}}(\boldsymbol{Y}_i - \beta_0 \mathbf{1}_{n_i})$ 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 $\boldsymbol{V}$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon)\} = \sigma^2 \boldsymbol{I}$?
    - $\boldsymbol{V} = \boldsymbol{\Sigma}(\hat{\tau})$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\hat{\tau}_\epsilon)\}$?

- 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 $\boldsymbol{V}$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon)\} = \sigma^2 \boldsymbol{I}$?
    - $\boldsymbol{V} = \boldsymbol{\Sigma}(\hat{\tau})$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\hat{\tau}_\epsilon)\}$?
  - 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 $\boldsymbol{V}$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\tau_\epsilon)\} = \sigma^2 I$?
    - $\boldsymbol{V} = \boldsymbol{\Sigma}(\hat{\tau})$?
    - $\boldsymbol{V} = \mathrm{Diag}\{\boldsymbol{\Sigma}_{\epsilon_i}(\hat{\tau}_\epsilon)\}$?
  - What to use for $\hat{\boldsymbol{Y}}$?
    - "Conditional model": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta} + \boldsymbol{Z}\hat{\gamma}$;
    - "Marginal model": $\hat{\boldsymbol{Y}} = \widehat{\boldsymbol{X}\beta}$.
  - Null model:

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

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

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

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

  Null models considered

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

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

- 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:

- Zheng (2000), under normality assumptions the same as Vonesh and Chinchilli (1997);
- Gelman and Pardoe (2006): Bayesian R^2, equivalent to Xu (2003);
- Magee (1990): R^2 based on log-likelihood, null model contains only fixed intercept;
- Orelien and Edwards (2008), 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:

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

Augmented model:

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

γ plays symbolically a double role;

$$\operatorname{cov}\begin{pmatrix} \epsilon \\ \gamma \end{pmatrix} = \sigma^2 \begin{pmatrix} I_n & 0 \\ 0 & \boldsymbol{G} \end{pmatrix};$$

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

Let

$$\Gamma = \begin{pmatrix} I_n & 0 \\ 0 & \Delta \end{pmatrix}.$$

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

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

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

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

Null model:

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

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

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

- $0 \leq R^2_{new} \leq 1$;
- R^2_{new} is increasing when adding columns into X matrix;

Disadvantage:

In the null model does not take into considerations dependencies between observations in $\boldsymbol{Y}$;

Choice of null model:

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

Suggested:

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

- Increasing with the number of fixed effect covariates;
- $0 \leq R^2 \leq 1$;
- Takes into consideration dependencies between observations also in the null model;
- Generalizable for unknown G: use the estimated variance-covariance components from the full model in both.

Still a lot has to be done – but what was the main question?

Still a lot has to be done – but what was the main question?

Thank you for your attention!