# 24 Lecture 24: Dec 4

## Announcement

- HW8 due today.
- Course evaluation: `https://classeval.ncsu.edu/`
- No OH from next week. Reachable by email.

## Last time

- Variance component testing: eLRT, eRLRT, score test.

## Today

- Variance component model: testing with two or more variance components.
- (Last topic!) Prediction: BLP and BLUP

## Testing with two or more variance components

- First we extend the test with one variance component to a slightly more general case $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{b}, \boldsymbol{V})$ with

  $$\boldsymbol{V} = \sigma_0^2 \boldsymbol{V}_0 + \sigma_1^2 \boldsymbol{V}_1, \tag{5}$$

  where $\boldsymbol{V}_0 \in \mathbb{R}^{n \times n}$ is a known psd matrix. Let $r = \text{rank}(\boldsymbol{V}_0)$.
  Given eigen-decomposition $\boldsymbol{V}_0 = \boldsymbol{U}_r \boldsymbol{D}_r \boldsymbol{U}_r^T$, define $\boldsymbol{T} = \boldsymbol{D}_r^{-1/2} \boldsymbol{U}_r^T \in \mathbb{R}^{r \times n}$. Then

  $$\boldsymbol{T}\boldsymbol{Y} \sim N(\boldsymbol{T}\boldsymbol{X}\boldsymbol{b}, \sigma_0^2 \boldsymbol{I}_r + \sigma_1^2 \boldsymbol{T}\boldsymbol{V}_1 \boldsymbol{T}^T)$$

  and the eLRT, eRLRT or score test can be applied to $\boldsymbol{T}\boldsymbol{Y}$.

- Now we consider the linear model with two variance components

  $$\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{b}, \boldsymbol{V}),$$

  where

  $$\boldsymbol{V} = \sigma_0^2 \boldsymbol{I}_n + \sigma_1^2 \boldsymbol{V}_1 + \sigma_2^2 \boldsymbol{V}_2.$$

- We are interested in testing $H_0 : \sigma_2^2 = 0$ vs $H_A : \sigma_2^2 > 0$.
- The idea is to massage the problem to the case (5) above.
- First perform QR (Gram-Schmidt) on the matrix $(\boldsymbol{X}, \boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{I}_n)$ to obtain an orthonormal basis $(\boldsymbol{Q}_0, \boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{Q}_3)$ of $\mathbb{R}^n$, where
  - $\boldsymbol{Q}_0$ is an orthonormal basis of $\mathcal{C}(\boldsymbol{X})$
  - $\boldsymbol{Q}_1$ is an orthonormal basis of $\mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1) - \mathcal{C}(\boldsymbol{X})$
  - $\boldsymbol{Q}_2$ is an orthonormal basis of $\mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1, \boldsymbol{V}_2) - \mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1)$
  - $\boldsymbol{Q}_3$ is an orthonormal basis of $\mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1, \boldsymbol{V}_2)^{\perp}$.
- If $\text{rank}(\boldsymbol{Q}_2) > 0$, that is $\mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1) \subsetneq \mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1, \boldsymbol{V}_2)$, then

  $$\boldsymbol{Q}_2^T \boldsymbol{Y} \sim N(\mathbf{0}, \sigma_0^2 \boldsymbol{I} + \sigma_2^2 \boldsymbol{Q}_2^T \boldsymbol{V}_2 \boldsymbol{Q}_2)$$

  and the exact F test, eLRT, eRLRT and score test can be applied to $\boldsymbol{Q}_2^T \boldsymbol{y}$.

  Is this beneficial to add the $\boldsymbol{Q}_3^T \boldsymbol{Y}$ component???
- If $\text{rank}(\boldsymbol{Q}_2) = 0$, that is $\mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1) = \mathcal{C}(\boldsymbol{X}, \boldsymbol{V}_1, \boldsymbol{V}_2)$, then we would like to choose $\lambda$ and matrix $\boldsymbol{K}$ such that

  $$\boldsymbol{Q}_1^T \boldsymbol{Y} + \boldsymbol{K} \boldsymbol{Q}_3^T \boldsymbol{Y} \sim N(\mathbf{0}, (\sigma_1^2 + \sigma_0^2/\lambda) \boldsymbol{Q}_1^T \boldsymbol{V}_1 \boldsymbol{Q}_1 + \sigma_2^2 \boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1),$$

  which is of form (5). We consider following situations.
  - If $\boldsymbol{Q}_1^T \boldsymbol{V}_2 = \mathbf{0}$, e.g., when $\mathcal{C}(\boldsymbol{V}_2) \subset \mathcal{C}(\boldsymbol{X})$, then this test cannot be performed.
  - If $\boldsymbol{Q}_1^T \boldsymbol{V}_1 \boldsymbol{Q}_1 = \lambda \boldsymbol{I}$, note

    $$\begin{aligned} \boldsymbol{Q}_1^T \boldsymbol{Y} &\sim N(\mathbf{0}, \sigma_0^2 \boldsymbol{I} + \sigma_1^2 \boldsymbol{Q}_1^T \boldsymbol{V}_1 \boldsymbol{Q}_1 + \sigma_2^2 \boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1) \\ &= N(\mathbf{0}, (\sigma_0^2 + \lambda \sigma_1^2) \boldsymbol{I} + \sigma_2^2 \boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1). \end{aligned}$$

    Then the ordinary tests (F test, eLRT, eRLRT, score) for one variance component can be applied without using the $\boldsymbol{K} \boldsymbol{Q}_3^T \boldsymbol{y}$ piece as long as $\boldsymbol{Q}_1^T \boldsymbol{V}_2 \neq \mathbf{0}$.
  - In general, $\boldsymbol{Q}_1^T \boldsymbol{V}_1 \boldsymbol{Q}_1 \neq \lambda \boldsymbol{I}$, then the test requires the $\boldsymbol{K} \boldsymbol{Q}_3^T \boldsymbol{y}$ term, which has distribution

    $$\boldsymbol{K} \boldsymbol{Q}_3^T \boldsymbol{Y} \sim N(\mathbf{0}, \sigma_0^2 \boldsymbol{K} \boldsymbol{K}^T).$$

Note that

$$\boldsymbol{Q}_1^T \boldsymbol{V} \boldsymbol{Q}_3 = \boldsymbol{Q}_1^T (\sigma_0^2 \boldsymbol{I}_n + \sigma_1^2 \boldsymbol{V}_1 + \sigma_2^2 \boldsymbol{V}_2) \boldsymbol{Q}_3 = \boldsymbol{0}.$$

Therefore $\boldsymbol{Q}_1^T \boldsymbol{Y} \perp \boldsymbol{K} \boldsymbol{Q}_3^T \boldsymbol{Y}$. We simply pick $\boldsymbol{K}$ such that

$$\boldsymbol{K} \boldsymbol{Q}_3^T \boldsymbol{Y} \sim N(\boldsymbol{0}, \sigma_0^2 (\lambda^{-1} \boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1 - \boldsymbol{I})).$$

That is

$$\boldsymbol{K} \boldsymbol{K}^T = \lambda^{-1} \boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1 - \boldsymbol{I}.$$

Apparently we need to choose $\lambda$ such that $\lambda^{-1} \boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1 - \boldsymbol{I}$ is psd. Let

$$\boldsymbol{Q}_1^T \boldsymbol{V}_2 \boldsymbol{Q}_1 = \boldsymbol{W} \boldsymbol{\Lambda} \boldsymbol{W}^T = \boldsymbol{W} \text{diag}(\lambda_i) \boldsymbol{W}^T$$

be the eigendecomposition. Setting $\lambda$ to be the smallest eigenvalue yields

$$\boldsymbol{K} = \boldsymbol{W} \text{diag}(\sqrt{\lambda_i / \lambda - 1}).$$

## Best linear prediction (BLP)

- Given data $(y, x_1, \ldots, x_p)$, regression can be thought of one way to predict $y$ from $x_1, \ldots, x_p$.
- A reasonable criterion is to choose predictor $f(\boldsymbol{x})$ such that the mean squared error

$$\text{MSE} = \text{E}[y - f(\boldsymbol{x})]^2$$

  is minimized. Here the expectation is wrt the joint distribution of $(y, \boldsymbol{x})$.
- Let $m(\boldsymbol{x}) = \text{E}(y|\boldsymbol{x})$. Then for any other predictor $f(\boldsymbol{x})$,

$$\text{E}[y - m(\boldsymbol{x})]^2 \leq \text{E}[y - f(\boldsymbol{x})]^2.$$

  That is $m(\boldsymbol{x})$ is the best predictor of $y$.

  *Proof.*

$$\begin{aligned}
& \text{E}[y - f(\boldsymbol{x})]^2 \\
= \; & \text{E}[y - m(\boldsymbol{x}) + m(\boldsymbol{x}) - f(\boldsymbol{x})]^2 \\
= \; & \text{E}[y - m(\boldsymbol{x})]^2 + \text{E}[m(\boldsymbol{x}) - f(\boldsymbol{x})]^2 + 2\text{E}[y - m(\boldsymbol{x})][m(\boldsymbol{x}) - f(\boldsymbol{x})].
\end{aligned}$$

But the cross term vanishes

$$\begin{aligned}
& \mathrm{E}[y - m(\boldsymbol{x})][m(\boldsymbol{x}) - f(\boldsymbol{x})] \\
= \; & \mathrm{E}\left\{\mathrm{E}[y - m(\boldsymbol{x})][m(\boldsymbol{x}) - f(\boldsymbol{x})] \mid \boldsymbol{x}\right\} \\
= \; & \mathrm{E}\left\{[m(\boldsymbol{x}) - f(\boldsymbol{x})]\mathrm{E}[y - m(\boldsymbol{x}) \mid \boldsymbol{x}]\right\} \\
= \; & \mathrm{E}\left\{[m(\boldsymbol{x}) - f(\boldsymbol{x})]0\right\} \\
= \; & 0.
\end{aligned}$$

Therefore $\mathrm{E}[y - f(\boldsymbol{x})]^2 \geq \mathrm{E}[y - m(\boldsymbol{x})]^2$ for any $f(\boldsymbol{x})$. □

- In order to use this result, we need to know the joint distribution of $(y, \boldsymbol{x})$. This is often unrealistic ☹ If only the first two moments (means, variances, and covariances) are known, then we can find the *best linear predictor* (BLP) of $y$.

- Assume

$$\mathrm{E}\begin{pmatrix} y \\ \boldsymbol{x} \end{pmatrix} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_{\boldsymbol{x}} \end{pmatrix}, \quad \mathrm{Cov}\begin{pmatrix} \sigma_{yy} & \boldsymbol{\Sigma}_{\boldsymbol{x}y}^T \\ \boldsymbol{\Sigma}_{\boldsymbol{x}y} & \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}} \end{pmatrix}.$$

Let $\boldsymbol{\beta}^*$ be a solution of $\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}y}$. Then

$$\widehat{\mathrm{E}}(y|\boldsymbol{x}) := \mu_y + (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})^T \boldsymbol{\beta}^*$$

is *the* best linear predictor (BLP) of $y$.

*Proof.* Let $f(\boldsymbol{x}) = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}$ be an arbitrary linear predictor. Then we find $\alpha, \boldsymbol{\beta}$ by minimizing the MSE

$$\begin{aligned}
& \mathrm{E}[y - f(\boldsymbol{x})]^2 \\
= \; & \mathrm{E}(y - \alpha - \boldsymbol{x}^T\boldsymbol{\beta})^2 \\
= \; & \mathrm{E}(y - \alpha)^2 + \boldsymbol{\beta}^T\mathrm{E}(\boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{\beta} - 2\mathrm{E}[(y - \alpha)(\boldsymbol{x}^T\boldsymbol{\beta})] \\
= \; & \mathrm{E}(y - \alpha)^2 + \boldsymbol{\beta}^T\mathrm{E}(\boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{\beta} - 2\mathrm{E}(y\boldsymbol{x}^T)\boldsymbol{\beta} + 2\alpha\mathrm{E}(\boldsymbol{x}^T)\boldsymbol{\beta} \\
= \; & \mathrm{E}(y - \alpha)^2 + \boldsymbol{\beta}^T\mathrm{E}(\boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{\beta} - 2\mathrm{E}(y\boldsymbol{x})^T\boldsymbol{\beta} + 2\alpha\boldsymbol{\mu}_{\boldsymbol{x}}^T\boldsymbol{\beta}.
\end{aligned}$$

Setting derivatives to 0 gives

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\mathrm{E}[y - f(\boldsymbol{x})]^2 &= 2(\alpha - \mu_y) + 2\boldsymbol{\mu}_{\boldsymbol{x}}^T\boldsymbol{\beta} = 0 \\
\nabla_{\boldsymbol{\beta}}\mathrm{E}[y - f(\boldsymbol{x})]^2 &= 2\mathrm{E}(\boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{\beta} - 2\mathrm{E}(\boldsymbol{x}y) + 2\alpha\boldsymbol{\mu}_{\boldsymbol{x}} = \mathbf{0}_p.
\end{aligned}$$

From the first equation, $\alpha = \mu_y - \boldsymbol{\mu}_{\boldsymbol{x}}^T \boldsymbol{\beta}$. Substitution into the second equation yields

$$(\mathrm{E}\boldsymbol{x}\boldsymbol{x}^T - \boldsymbol{\mu}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{x}}^T)\boldsymbol{\beta} = \mathrm{E}(\boldsymbol{x}y) - \boldsymbol{\mu}_{\boldsymbol{x}}\mu_y.$$

That is

$$\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}y}.$$

Therefore the optimal $\alpha$ is $\alpha^* = \mu_y - \boldsymbol{\mu}_{\boldsymbol{x}}^T \beta^*$, where $\boldsymbol{\beta}^*$ is any solution to above equation. And the BLP is

$$\alpha^* + \boldsymbol{x}^T \boldsymbol{\beta}^* = \mu_y + (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})^T \boldsymbol{\beta}^*.$$

Because the criterion function is a convex function, the stationary condition is both necessary and sufficient for the global minima. Therefore any BLP must be of this form. □

## Best linear unbiased prediction (BLUP)

- Consider random variables $y_0, y_1, \ldots, y_n$, and we are interested predicting $y_0$ given data $y_1, \ldots, y_n$. If we know the mean, variances, and covariances, then we can use above theory to find the BLP of $y_0$.

- In practice, we don't know the means $\mu_0 = \mathrm{E}y_0, \boldsymbol{\mu}_{\boldsymbol{y}} = \mathrm{E}\boldsymbol{y}$ most of time ☹
  Let's impose a linear (Aitken) model for the means $\mu_i$

$$\mathrm{E}\begin{pmatrix} \boldsymbol{y} \\ y_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}\boldsymbol{b} \\ \boldsymbol{x}_0^T \boldsymbol{b} \end{pmatrix}, \quad \mathrm{Cov}\begin{pmatrix} \boldsymbol{y} \\ y_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{V} & \boldsymbol{V}_{\boldsymbol{y}y_0} \\ \boldsymbol{V}_{\boldsymbol{y}y_0}^T & V_{y_0 y_0} \end{pmatrix}. \tag{6}$$

- If $\boldsymbol{b}$ is known, then the BLP of $y_0$ is

$$\boldsymbol{x}_0^T \boldsymbol{b} + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})^T \boldsymbol{\beta}^*,$$

  where $\boldsymbol{\beta}^*$ is a solution to $\boldsymbol{V}\boldsymbol{\beta} = \boldsymbol{V}_{\boldsymbol{y}y_0}$.

- If $\boldsymbol{b}$ is unknown, then the hope is to find the *best linear unbiased predictor* (BLUP).

  We call a linear predictor $f(\boldsymbol{y}) = a_0 + \boldsymbol{a}^T \boldsymbol{y}$ of $y_0$ BLUP if

1. it is unbiased, i.e.,

$$\mathrm{E}\, f(\boldsymbol{y}) = a_0 + \boldsymbol{a}^T \boldsymbol{X}\boldsymbol{b} = \boldsymbol{x}_0^T \boldsymbol{b} = \mathrm{E} y_0$$

   for all $\boldsymbol{b}$, and

2. for any other linear unbiased predictor $b_0 + \boldsymbol{b}^T \boldsymbol{y}$,

$$\mathrm{E}(y_0 - a_0 - \boldsymbol{a}^T \boldsymbol{y})^2 \le \mathrm{E}(y_0 - b_0 - \boldsymbol{b}^T \boldsymbol{y})^2.$$

- Theorem: Under the Aitken model (6) and assume $\boldsymbol{V}_{\boldsymbol{y}y_0} \in \mathcal{C}(\boldsymbol{V}, \boldsymbol{X})$ and $\boldsymbol{x}_0 \in \mathcal{C}(\boldsymbol{X}^T)$, the BLUP of $y_0$ is

$$\boldsymbol{x}_0^T \hat{\boldsymbol{b}}_{\mathrm{GLS}} + (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}_{\mathrm{GLS}})^T \boldsymbol{\beta}_*,$$

where $\boldsymbol{\beta}_*$ is a solution of $(\boldsymbol{V} + \boldsymbol{X}\boldsymbol{X}^T)\boldsymbol{\beta} = \boldsymbol{V}_{\boldsymbol{y}y_0}$ and $\boldsymbol{X}\hat{\boldsymbol{b}}_{\mathrm{GLS}}$ is the BLUE of $\boldsymbol{X}\boldsymbol{b}$.

Remark 1: We don't assume $\boldsymbol{V}$ is nonsingular. For nonsingular $\boldsymbol{V}$, we can take $\boldsymbol{\beta}_* = \boldsymbol{V}^{-1}\boldsymbol{V}_{\boldsymbol{y}y_0}$.

Remark 2: Both $\hat{\boldsymbol{b}}_{\mathrm{GLS}}$ and $\boldsymbol{\beta}^*$ depend crucially on $\boldsymbol{V}$.

*Proof.* Let $a_0 + \boldsymbol{a}^T \boldsymbol{y}$ be arbitrary linear predictor of $y_0$. Unbiasedness requires

$$a_0 + \boldsymbol{a}^T \boldsymbol{X}\boldsymbol{b} = \boldsymbol{x}_0^T \boldsymbol{b}$$

for all $\boldsymbol{b}$. Thus $a_0 = 0$ and $\boldsymbol{a}^T \boldsymbol{X} = \boldsymbol{x}_0^T$. We need to solve the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathrm{E}(y_0 - \boldsymbol{a}^T \boldsymbol{y})^2 = \frac{1}{2}\boldsymbol{a}^T \mathrm{E}(\boldsymbol{y}\boldsymbol{y}^T)\boldsymbol{a} - \mathrm{E}(y_0 \boldsymbol{a}^T \boldsymbol{y}) + \frac{1}{2}\mathrm{E} y_0^2 \\ \text{subject to} & \boldsymbol{X}^T \boldsymbol{a} = \boldsymbol{x}_0. \end{array}$$

Setting the gradient of the Lagrangian

$$L(\boldsymbol{a}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{a}^T \mathrm{E}(\boldsymbol{y}\boldsymbol{y}^T)\boldsymbol{a} - \mathrm{E}(y_0 \boldsymbol{a}^T \boldsymbol{y}) + \frac{1}{2}\mathrm{E} y_0^2 + \boldsymbol{\lambda}^T (\boldsymbol{X}^T \boldsymbol{a} - \boldsymbol{x}_0)$$

to zero yields equations

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{y}\boldsymbol{y}^T)\boldsymbol{a} - \boldsymbol{E}(\boldsymbol{y} y_0) + \boldsymbol{X}\boldsymbol{\lambda} &= \boldsymbol{0}_n \\ \boldsymbol{X}^T \boldsymbol{a} &= \boldsymbol{x}_0. \end{aligned}$$

Adding and subtracting $(\mathrm{E}\boldsymbol{y}\mathrm{E}\boldsymbol{y}^T)\boldsymbol{a} = \boldsymbol{X}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{X}^T\boldsymbol{a} = \boldsymbol{X}\boldsymbol{b}\boldsymbol{b}^T\boldsymbol{x}_0 = \mathrm{E}\boldsymbol{y}\mathrm{E}y_0$ to the first equation shows $\boldsymbol{V}\boldsymbol{a} - \boldsymbol{V}_{\boldsymbol{y}y_0} + \boldsymbol{X}\boldsymbol{\lambda} = \boldsymbol{0}_n$. In matrix notation,

$$\begin{pmatrix} \boldsymbol{V} & \boldsymbol{X} \\ \boldsymbol{X}^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}_{\boldsymbol{y}y_0} \\ \boldsymbol{x}_0 \end{pmatrix}.$$

By HW4 5(d), solution for the optimal $\boldsymbol{a}$ is

$$\boldsymbol{a}_* = [\boldsymbol{V}_0^- - \boldsymbol{V}_0^-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}_0^-\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{V}_0^-]\boldsymbol{V}_{\boldsymbol{y}y_0} + \boldsymbol{V}_0^-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}_0^-\boldsymbol{X})^-\boldsymbol{x}_0,$$

where $\boldsymbol{V}_0 = \boldsymbol{V} + \boldsymbol{X}\boldsymbol{X}^T$. Thus the BLUP is

$$\begin{aligned} \boldsymbol{a}_*^T\boldsymbol{y} &= \boldsymbol{V}_{\boldsymbol{y}y_0}^T[\boldsymbol{V}_0^- - \boldsymbol{V}_0^-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}_0^-\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{V}_0^-]\boldsymbol{y} + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{V}_0^-\boldsymbol{X})^-\boldsymbol{X}^T\boldsymbol{V}_0^-\boldsymbol{y} \\ &= \boldsymbol{V}_{\boldsymbol{y}y_0}^T\boldsymbol{V}_0^-(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}_{\mathrm{GLS}}) + \boldsymbol{x}_0^T\hat{\boldsymbol{b}}_{\mathrm{GLS}} \\ &= \boldsymbol{\beta}_*^T(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}_{\mathrm{GLS}}) + \boldsymbol{x}_0^T\hat{\boldsymbol{b}}_{\mathrm{GLS}}. \end{aligned}$$

□

- The prediction variance of BLUP is (TODO)

$$\mathrm{E}(y_0 - \boldsymbol{a}_*^T\boldsymbol{y})^2 = V_{y_0y_0} - 2\boldsymbol{a}_*^T\boldsymbol{V}_{\boldsymbol{y}y_0} + \boldsymbol{a}_*^T\boldsymbol{V}\boldsymbol{a}_*. \quad ???$$

- Example (BLUP in Gauss-Markov linear model): $\boldsymbol{V} = \sigma^2\boldsymbol{I}$ and $\boldsymbol{V}_{\boldsymbol{y}y_0} = \boldsymbol{0}_p$. Thus $\boldsymbol{\beta}_* = \boldsymbol{0}_p$ and the BLUP for $y_0$ is $\boldsymbol{x}_0^T\hat{\boldsymbol{b}}$, which is also the BLUE of $\boldsymbol{x}_0^T\boldsymbol{b}$.

## Mixed model equation (MME)

- Consider the mixed effects model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

  - $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is a design matrix for fixed effects $\boldsymbol{b} \in \mathbb{R}^p$.
  - $\boldsymbol{Z} \in \mathbb{R}^{n\times q}$ is a design matrix for random effects $\boldsymbol{u} \in \mathbb{R}^q$.
  - The most general assumption is $\boldsymbol{e} \in N(\boldsymbol{0}_n, \boldsymbol{R})$, $\boldsymbol{u} \in N(\boldsymbol{0}_q, \boldsymbol{G})$, and $\boldsymbol{e}$ is independent of $\boldsymbol{u}$. That is

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim N\left(\boldsymbol{0}_{q+n}, \begin{pmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{pmatrix}\right).$$

    Assume $\boldsymbol{G}$ and $\boldsymbol{R}$ are nonsingular.

- We already know that the BLUE of $\boldsymbol{Xb}$ is

$$\boldsymbol{X}\hat{\boldsymbol{b}}_{\text{GLS}} = \boldsymbol{X}[\boldsymbol{X}^T(\boldsymbol{R}+\boldsymbol{ZGZ}^T)^{-1}\boldsymbol{X}]^-\boldsymbol{X}^T(\boldsymbol{R}+\boldsymbol{ZGZ}^T)^{-1}\boldsymbol{y}.$$

- We can apply the previous theorem to derive the BLUP of $\boldsymbol{u}$. Note

$$\mathrm{E}\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{u}\end{pmatrix} = \begin{pmatrix}\boldsymbol{Xb}\\ \boldsymbol{0}_q\end{pmatrix}, \quad \mathrm{Cov}\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{u}\end{pmatrix} = \begin{pmatrix}\boldsymbol{R}+\boldsymbol{ZGZ}^T & \boldsymbol{ZG}\\ \boldsymbol{GZ}^T & \boldsymbol{R}\end{pmatrix}.$$

  Therefore the BLUP for $\boldsymbol{u}$ is

$$\boldsymbol{GZ}^T(\boldsymbol{R}+\boldsymbol{ZGZ}^T)^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{b}}_{\text{GLS}}).$$

- It turns out both BLUE of $\boldsymbol{Xb}$ and BLUP of $\boldsymbol{u}$ can be obtained simultaneously by solving a so-called mixed model equation (MME).

  *Mixed model equation* (MME) defined as

$$\begin{pmatrix}\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}\\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{G}^{-1}+\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\boldsymbol{b}\\ \boldsymbol{u}\end{pmatrix} = \begin{pmatrix}\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y}\\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y}\end{pmatrix}$$

  is a generalization of the normal equation for fixed effects model.

- Theorem: Let $(\hat{\boldsymbol{b}}, \hat{\boldsymbol{u}})$ be a solution to MME. Then $\boldsymbol{X}\hat{\boldsymbol{b}}$ is the BLUE of $\boldsymbol{Xb}$ and $\hat{\boldsymbol{u}}$ is the BLUP of $\boldsymbol{u}$.

  *Proof.* Let

$$\boldsymbol{V} = \mathrm{Cov}(\boldsymbol{y}) = \boldsymbol{R}+\boldsymbol{ZGZ}^T.$$

  Then $\boldsymbol{X}\hat{\boldsymbol{b}}$ is a BLUE of $\boldsymbol{Xb}$ if $\hat{\boldsymbol{b}}$ is a solution to

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Xb} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}.$$

  By the binomial inversion formula (HW1), we have

$$\boldsymbol{V}^{-1} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{G}^{-1}+\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}.$$

  The MME says

$$\begin{aligned}\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Xb}+\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Zu} &= \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y}\\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Xb}+(\boldsymbol{G}^{-1}+\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})\boldsymbol{u} &= \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y}.\end{aligned}$$

From the second equation we solve for

$$\boldsymbol{u} = (\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$$

and substituting into the first equation shows

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}.$$

Thus $\hat{\boldsymbol{b}}$ is a generalized least squares solution and $\boldsymbol{X}\hat{\boldsymbol{b}}$ is the BLUE of $\boldsymbol{X}\boldsymbol{b}$.

To show that $\hat{\boldsymbol{u}}$ is BLUP, only need to show that

$$(\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1} \quad = \quad \boldsymbol{G}\boldsymbol{Z}^T(\boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T)^{-1},$$

or equivalently

$$\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T) = (\boldsymbol{G}^{-1} + \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})\boldsymbol{G}\boldsymbol{Z}^T,$$

which is obvious. □

# References


Casella, G. and Berger, R. (2001). *Statistical Inference*. Duxbury Resource Center.

Christensen, R. (1996). Exact tests for variance components. *Biometrics*, 52(1):309–314.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(1):165–185.

Davies, R. B. (1980). The distribution of linear combinations of $\chi^2$ random variables. *Applied Statistics*, 29:323–333.

Öfversten, J. (1993). Exact tests for variance components in unbalanced mixed linear models. *Biometrics*, 49(1):45–57.

Seely, J. F. and El Bassiouni, Y. (1983). Applying Wald's variance component test. *Ann. Statist.*, 11(1):197–201.

Teets, D. and Whitehead, K. (1999). The discovery of Ceres: how Gauss became famous. *Math. Mag.*, 72(2):83–93.