GLS and LME

Peter Hoff
Duke STA 610

LME - Is it worth it?

```
fitOLS<-lm(y.nels ~ flp.nels + ses.nels + flp.nels*ses.nels)
summary(fitOLS)

##
## Call:
## lm(formula = y.nels ~ flp.nels + ses.nels + flp.nels * ses.nels)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.107  -5.758   0.142   5.977  33.538
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        54.8442     0.2280  240.50   <2e-16 ***
## flp.nels           -2.0809     0.1075  -19.36   <2e-16 ***
## ses.nels            4.9058     0.2810   17.46   <2e-16 ***
## flp.nels:ses.nels  -0.1279     0.1361   -0.94    0.347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.754 on 12970 degrees of freedom
## Multiple R-squared:  0.2028, Adjusted R-squared:  0.2026
## F-statistic:  1100 on 3 and 12970 DF,  p-value: < 2.2e-16
```

LME - Is it worth it?

```
fitLME<-lmer(y.nels ~ flp.nels + ses.nels + flp.nels:ses.nels + (ses.nels|g.nels)  )
summary(fitLME)

## Linear mixed model fit by REML ['lmerMod']
## Formula: y.nels ~ flp.nels + ses.nels + flp.nels:ses.nels + (ses.nels |
##     g.nels)
##
## REML criterion at convergence: 92388.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9769 -0.6415  0.0198  0.6659  4.5206
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  g.nels   (Intercept)  9.056   3.009
##           ses.nels     1.602   1.266    0.06
##  Residual             67.258   8.201
## Number of obs: 12974, groups:  g.nels, 684
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)        55.3989     0.3866 143.285
## flp.nels           -2.4070     0.1822 -13.212
## ses.nels            4.4899     0.3333  13.472
## flp.nels:ses.nels  -0.1931     0.1590  -1.214
##
## Correlation of Fixed Effects:
##             (Intr) flp.nl ss.nls
## flp.nels    -0.930
## ses.nels    -0.157  0.088
## flp.nls:ss.  0.085 -0.007 -0.926
```

LME - Is it worth it?

For the mixed effects model

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j,$$

the OLS estimate is still *unbiased*. However,

- it is no longer the BLUE;
- its variance is no longer $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.

For this model, the BLUE is (approximately) the MLE returned by `lmer`.

LME - Is it worth it?

For the mixed effects model

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j,$$

the OLS estimate is still *unbiased*. However,

- it is no longer the BLUE;
- its variance is no longer $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.

For this model, the BLUE is (approximately) the MLE returned by `lmer`.

Linear unbiased estimators

Model:

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $\mathrm{E}[\epsilon|\mathbf{X}] = \mathbf{0}$, $\mathrm{Var}[\epsilon|\mathbf{X}] = \sigma^2\mathbf{I}$.

or equivalently,

- $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$
- $\mathrm{Var}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{I}$

OLS Estimator: $\hat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$

Linear unbiased estimators: $\check{\beta} = [(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{H}^\top]\mathbf{y}$

Exercise: Show that $\check{\beta}$, and hence $\hat{\beta}$, are unbiased.

Linear unbiased estimators

Model:

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $\mathrm{E}[\epsilon|\mathbf{X}] = \mathbf{0}, \mathrm{Var}[\epsilon|\mathbf{X}] = \sigma^2\mathbf{I}$.

or equivalently,

- $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$
- $\mathrm{Var}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{I}$

OLS Estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$

Linear unbiased estimators: $\check{\beta} = [(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{H}^\top]\mathbf{y}$

Exercise: Show that $\check{\beta}$, and hence $\hat{\beta}$, are unbiased.

Linear unbiased estimators

Model:

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $\mathrm{E}[\epsilon|\mathbf{X}] = \mathbf{0}, \mathrm{Var}[\epsilon|\mathbf{X}] = \sigma^2\mathbf{I}$.

or equivalently,

- $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$
- $\mathrm{Var}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{I}$

OLS Estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$

Linear unbiased estimators: $\check{\boldsymbol{\beta}} = [(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{H}^\top]\mathbf{y}$

Exercise: Show that $\check{\beta}$, and hence $\hat{\beta}$, are unbiased.

Linear unbiased estimators

Model:

- $\mathbf{y} = \mathbf{X}\beta + \epsilon$
- $\mathsf{E}[\epsilon|\mathbf{X}] = \mathbf{0}$, $\mathsf{Var}[\epsilon|\mathbf{X}] = \sigma^2\mathbf{I}$.

or equivalently,

- $\mathsf{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta$
- $\mathsf{Var}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{I}$

OLS Estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$

Linear unbiased estimators: $\check{\boldsymbol{\beta}} = [(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{H}^\top]\mathbf{y}$

Exercise: Show that $\check{\boldsymbol{\beta}}$, and hence $\hat{\boldsymbol{\beta}}$, are unbiased.

Variance of linear unbiased estimators

$$
\begin{aligned}
\mathrm{Var}[\check{\beta}] &= (\mathbf{X}^{+} + \mathbf{H}^{\top})\mathrm{Var}[\epsilon](\mathbf{X}^{+} + \mathbf{H}^{\top})^{\top} \\
&= \sigma^2 (\mathbf{X}^{+} + \mathbf{H}^{\top})(\mathbf{X}^{+} + \mathbf{H}^{\top})^{\top} \\
&= \sigma^2 \left(\mathbf{X}^{+}(\mathbf{X}^{+})^{\top} + \mathbf{X}^{+}\mathbf{H} + \mathbf{H}^{\top}(\mathbf{X}^{+})^{\top} + \mathbf{H}^{\top}\mathbf{H} \right).
\end{aligned}
$$

Now calculate the individual terms:

$$
\begin{aligned}
\mathbf{X}^{+}(\mathbf{X}^{+})^{\top} &= (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{X}(\mathbf{X}^{\top}\mathbf{X})^{-1} \\
&= (\mathbf{X}^{\top}\mathbf{X})^{-1}, \\
\mathbf{H}^{\top}(\mathbf{X}^{+})^{\top} &= \mathbf{H}^{\top}\mathbf{X}(\mathbf{X}^{\top}\mathbf{X})^{-1} \\
&= \mathbf{0}.
\end{aligned}
$$

So

$$
\begin{aligned}
\mathrm{Var}[\check{\beta}] &= \sigma^2(\mathbf{X}^{\top}\mathbf{X})^{-1} + \sigma^2\mathbf{H}^{\top}\mathbf{H} \\
&= \mathrm{Var}[\hat{\beta}] + \sigma^2\mathbf{H}^{\top}\mathbf{H}.
\end{aligned}
$$

Gauss-Markov theorem

Definition (Loewner order)

For two positive semidefinite matrices Σ_1 and Σ_2 of the same size, we say that $\Sigma_1 > \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive definite, and that $\Sigma_1 \geq \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive semidefinite.

Theorem

Let $\check{\beta}$ be a linear unbiased estimator of β in a linear model where $E[\mathbf{y}] = \mathbf{X}\beta, \beta \in \mathbb{R}^p$ and $Var[\mathbf{y}] = \sigma^2\mathbf{I}$, $\sigma^2 > 0$. Then

$$Var[\check{\beta}] \geq Var[\hat{\beta}],$$

where $\hat{\beta}$ is the OLS estimator.

The OLS estimator is the BLUE in this case.

Gauss-Markov theorem

Definition (Loewner order)

For two positive semidefinite matrices Σ_1 and Σ_2 of the same size, we say that $\Sigma_1 > \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive definite, and that $\Sigma_1 \geq \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive semidefinite.

Theorem

Let $\check{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear model where $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathbb{R}^p$ and $Var[\mathbf{y}] = \sigma^2\mathbf{I}$, $\sigma^2 > 0$. Then

$$Var[\check{\boldsymbol{\beta}}] \geq Var[\hat{\boldsymbol{\beta}}],$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator.

The OLS estimator is the BLUE in this case.

Gauss-Markov theorem

Definition (Loewner order)

For two positive semidefinite matrices Σ_1 and Σ_2 of the same size, we say that $\Sigma_1 > \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive definite, and that $\Sigma_1 \geq \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive semidefinite.

Theorem

Let $\check{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear model where $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathbb{R}^p$ and $Var[\mathbf{y}] = \sigma^2\mathbf{I}$, $\sigma^2 > 0$. Then

$$Var[\check{\boldsymbol{\beta}}] \geq Var[\hat{\boldsymbol{\beta}}],$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator.

The OLS estimator is the BLUE in this case.

Non-isotropic variance

What if $\text{Var}[\mathbf{y}] \neq \sigma^2 \mathbf{I}$?

- Heteroscedasticity: $\text{Var}[\mathbf{y}_i] = w_i \sigma^2$ for some known $w_1, \ldots, w_n$.
- Time series: $\text{Var}[\mathbf{y}] = \sigma^2 \mathbf{A}$, where $a_{i,j} = \rho^{|i-j|}$.

LME models and can be viewed as models for correlated data. Let

$$\mathbf{y}_j = \mathbf{X}_j \beta + \mathbf{Z}_j \mathbf{a}_j + \epsilon_j$$

where

- $\text{E}[\mathbf{a}_j] = \mathbf{0}, \text{Var}[\mathbf{a}_j] = \Psi$.
- $\text{E}[\epsilon_j] = \mathbf{0}, \text{Var}[\epsilon_j] = \sigma^2 \mathbf{I}$.

Then

$$\text{E}[\mathbf{y}_j] = \mathbf{X}_j \beta$$

$$\text{Var}[\mathbf{y}_j] = \mathbf{Z}_j \Psi \mathbf{Z}_j^\top + \sigma^2 \mathbf{I}.$$

Non-isotropic variance

What if $\text{Var}[\mathbf{y}] \neq \sigma^2 \mathbf{I}$?

- Heteroscedasticity: $\text{Var}[\mathbf{y}_i] = w_i \sigma^2$ for some known $w_1, \ldots, w_n$.
- Time series: $\text{Var}[\mathbf{y}] = \sigma^2 \mathbf{A}$, where $a_{i,j} = \rho^{|i-j|}$.

LME models and can be viewed as models for correlated data. Let

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{a}_j + \boldsymbol{\epsilon}_j$$

where

- $\text{E}[\mathbf{a}_j] = \mathbf{0}, \text{Var}[\mathbf{a}_j] = \Psi$.
- $\text{E}[\boldsymbol{\epsilon}_j] = \mathbf{0}, \text{Var}[\boldsymbol{\epsilon}_j] = \sigma^2 \mathbf{I}$.

Then

$$\text{E}[\mathbf{y}_j] = \mathbf{X}_j \boldsymbol{\beta}$$
$$\text{Var}[\mathbf{y}_j] = \mathbf{Z}_j \Psi \mathbf{Z}_j^\top + \sigma^2 \mathbf{I}.$$

OLS with dependent data

The OLS estimator is still unbiased when data are correlated:

$$\begin{aligned} \mathsf{E}[\hat{\beta}] = \mathsf{E}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{E}[\mathbf{X}^\top\boldsymbol{\beta} + \boldsymbol{\epsilon}] \\ &= \boldsymbol{\beta} + \mathbf{0} = \boldsymbol{\beta}. \end{aligned}$$

However, its variance in this case is complicated:

$$\begin{aligned} \mathsf{Var}[\hat{\beta}] = \mathsf{Var}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{Var}[\mathbf{y}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}. \end{aligned}$$

This is quite messy, and not equal to $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$ unless $\mathsf{Var}[\mathbf{y}]$ is special.

OLS with dependent data

The OLS estimator is still unbiased when data are correlated:

$$\begin{aligned} \mathsf{E}[\hat{\boldsymbol{\beta}}] = \mathsf{E}[(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}] &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathsf{E}[\mathbf{X}^\top \boldsymbol{\beta} + \boldsymbol{\epsilon}] \\ &= \boldsymbol{\beta} + \mathbf{0} = \boldsymbol{\beta}. \end{aligned}$$

However, its variance in this case is complicated:

$$\begin{aligned} \mathsf{Var}[\hat{\boldsymbol{\beta}}] = \mathsf{Var}[(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}] &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathsf{Var}[\mathbf{y}]\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}. \end{aligned}$$

This is quite messy, and not equal to $\sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$ unless $\mathsf{Var}[\mathbf{y}]$ is special.

OLS with dependent data

The OLS estimator is still unbiased when data are correlated:

$$\begin{aligned} \mathsf{E}[\hat{\boldsymbol\beta}] = \mathsf{E}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{E}[\mathbf{X}^\top\boldsymbol\beta + \boldsymbol\epsilon] \\ &= \boldsymbol\beta + \mathbf{0} = \boldsymbol\beta. \end{aligned}$$

However, its variance in this case is complicated:

$$\begin{aligned} \mathsf{Var}[\hat{\boldsymbol\beta}] = \mathsf{Var}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathsf{Var}[\mathbf{y}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}. \end{aligned}$$

This is quite messy, and not equal to $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$ unless $\mathsf{Var}[\mathbf{y}]$ is special.

GLS estimator

Let $\text{Var}[\mathbf{y}] = \text{Var}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$. We define the *symmetric square root* $\mathbf{V}^{1/2}$ of $\mathbf{V}$ as

$$\mathbf{V}^{1/2} = \mathbf{E}\boldsymbol{\Lambda}^{1/2}\mathbf{E}^\top.$$

where $(\mathbf{E}, \boldsymbol{\Lambda})$ are the eigenvectors and values of Σ. Note that$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V}$.

This matrix is a *whitening matrix* for $\mathbf{y}$:

$$\begin{aligned}\text{Var}[\mathbf{V}^{-1/2}\mathbf{y}] &= \mathbf{V}^{-1/2}\text{Var}[\mathbf{y}]\mathbf{V}^{-\top/2} \\ &= \mathbf{V}^{-1/2}(\sigma^2\mathbf{V})\mathbf{V}^{-1/2} = \sigma^2\mathbf{I}.\end{aligned}$$

GLS estimator

Let $\text{Var}[\mathbf{y}] = \text{Var}[\epsilon] = \sigma^2\mathbf{V}$. We define the *symmetric square root* $\mathbf{V}^{1/2}$ of $\mathbf{V}$ as

$$\mathbf{V}^{1/2} = \mathbf{E}\boldsymbol{\Lambda}^{1/2}\mathbf{E}^\top.$$

where $(\mathbf{E}, \boldsymbol{\Lambda})$ are the eigenvectors and values of Σ. Note that$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V}$.

This matrix is a *whitening matrix* for $\mathbf{y}$:

$$\begin{aligned}\text{Var}[\mathbf{V}^{-1/2}\mathbf{y}] &= \mathbf{V}^{-1/2}\text{Var}[\mathbf{y}]\mathbf{V}^{-\top/2} \\ &= \mathbf{V}^{-1/2}(\sigma^2\mathbf{V})\mathbf{V}^{-1/2} = \sigma^2\mathbf{I}.\end{aligned}$$

OLS with whitened data

Let $\tilde{\mathbf{y}} = \mathbf{V}^{-1/2}\mathbf{y}$. The linear model for $\tilde{\mathbf{y}}$ is then

$$\mathbf{V}^{-1/2}\mathbf{y} = \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\epsilon}$$
$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}},$$

where $\mathrm{E}[\tilde{\epsilon}] = \mathbf{0}$ and

$$\mathrm{Var}[\tilde{\epsilon}] = \sigma^2\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \sigma^2\mathbf{I}.$$

The BLUE based on $\tilde{\mathbf{y}}$, $\tilde{\mathbf{X}}$ is

$$\hat{\boldsymbol{\beta}}_V = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

GLS via whitened OLS

$\hat{\boldsymbol{\beta}}_V$ is linear in $\mathbf{y}$! So $\hat{\boldsymbol{\beta}}_V$ is the BLUE of $\boldsymbol{\beta}$, based on either $\tilde{\mathbf{y}}$ or $\mathbf{y}$.

On the original scale of the data, we have

$$\begin{aligned}\hat{\beta}_V &= (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}} \\ &= (\mathbf{X}^\top \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.\end{aligned}$$

This estimator is the *generalized least squares* (GLS) estimator of β. Its variance is

$$\mathrm{Var}[\hat{\beta}_V] = \sigma^2 (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}.$$

GLS via whitened OLS

$\hat{\boldsymbol{\beta}}_V$ is linear in $\mathbf{y}$! So $\hat{\boldsymbol{\beta}}_V$ is the BLUE of $\boldsymbol{\beta}$, based on either $\tilde{\mathbf{y}}$ or $\mathbf{y}$.

On the original scale of the data, we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_V &= (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}} \\ &= (\mathbf{X}^\top \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.\end{aligned}$$

This estimator is the *generalized least squares* (GLS) estimator of $\boldsymbol{\beta}$. Its variance is

$$\text{Var}[\hat{\boldsymbol{\beta}}_V] = \sigma^2 (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}.$$

Gauss-Markov-Aitkin theorem

Theorem

Let $\check{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear model with $E[\mathbf{y}] = \mathbf{X}\beta,\ Var[\mathbf{y}] = \sigma^2\mathbf{V}$ for $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}^+$ with $\mathbf{X}$ and $\mathbf{V}$ known. Then

$$Var[\check{\beta}] \geq \sigma^2(\mathbf{X}\mathbf{V}^{-1}\mathbf{X}^\top)^{-1} = Var[\hat{\beta}_V],$$

where $\hat{\boldsymbol{\beta}}_V = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y}$.

LME model as a GLM

Within-groups model: $\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j$.

- $\mathrm{E}[\mathbf{y}_j] = \mathbf{X}_j\boldsymbol{\beta}$;
- $\mathrm{Var}[\mathbf{y}_j] = \mathbf{Z}_j\Psi\mathbf{Z}_j^\top + \sigma^2 I_{n_j}$.

Let

- $\mathbf{y} = \mathbf{y}_1, \dots, \mathbf{y}_m) \in \mathbb{R}^{\sum n_j}$;
- $\mathbf{X} = (X_1^\top \cdots X_m^\top)^\top \in \mathbb{R}^{\sum n_j \times p}$.

Then

$$\mathrm{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$$

$$\mathrm{Var}[\mathbf{y}] = \begin{pmatrix} \mathbf{Z}_1\Psi\mathbf{Z}_1^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2\Psi\mathbf{Z}_2^\top & \cdots & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Z}_m\Psi\mathbf{Z}_m^\top \end{pmatrix}$$

Numerical comparison