12.1. Clustered Data

A Motivating Example: Let

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_K \end{pmatrix},$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ is a vector of responses on the ith cluster (patient, household, school, etc.). Assuming *clusters* are independent,

$$\text{cov}(\mathbf{Y}) = \begin{pmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_K \end{pmatrix},$$

where we might assume a common variance σ^2 and common pairwise correlation ρ within a cluster, i.e., an *exchangeable* correlation structure:

$$cov(\mathbf{Y}_i) = \sigma^2 \mathbf{V}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}_{n_i \times n_i}.$$

12.2. Generalized Least Squares Estimates

If we relax the assumption $\text{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$ what are the implications for estimation of $\boldsymbol{\beta}$? In general, let

$$\text{cov}(\mathbf{Y}) = \sigma^2\mathbf{V}, \quad \text{for some } \mathit{known}\mathbf{V}.$$

In practice, we do not know $\mathbf{V}$ and have to estimate it (e.g., estimate the correlation parameter ρ in the exchangeable case).

Question: Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, where $\text{rank}(\mathbf{X}_{n\times p}) = p$, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, with known p.d. $\mathbf{V}$. How to estimate $\boldsymbol{\beta}$?

Solution: Transform $\mathbf{Y}$ to a new response vector that has covariance matrix $\sigma^2\mathbf{I}$. Apply our knowledge of least squares estimates to transformed $\mathbf{Y}$.

Details: Since $\mathbf{V}$ is p.d., $\mathbf{V} = \mathbf{K}\mathbf{K}'$ for non-singular $\mathbf{K}$ (e.g., there is orthogonal $\mathbf{T}$, diagonal $\boldsymbol{\Lambda}$, such that $\mathbf{V} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}' = (\mathbf{T}\boldsymbol{\Lambda}^{1/2})(\mathbf{T}\boldsymbol{\Lambda}^{1/2})')$.

For the transformed response $\mathbf{Z} = \mathbf{K}^{-1}\mathbf{Y}$:

$$\begin{aligned}\mathbf{K}^{-1}\mathbf{Y} &= \mathbf{K}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}^{-1}\boldsymbol{\varepsilon} \\ \mathbf{Z} &= \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\eta},\end{aligned}$$

where

$$E[\boldsymbol{\eta}] = E[\mathbf{K}^{-1}\boldsymbol{\varepsilon}] = \mathbf{K}^{-1}E[\boldsymbol{\varepsilon}] = \mathbf{0}$$

$$\mathrm{cov}(\boldsymbol{\eta}) = \mathrm{cov}(\mathbf{K}^{-1}\boldsymbol{\varepsilon}) = \mathbf{K}^{-1}\sigma^2\mathbf{V}(\mathbf{K}^{-1})' = \sigma^2\mathbf{K}^{-1}\mathbf{K}\mathbf{K}'(\mathbf{K}^{-1})' = \sigma^2\mathbf{I}.$$

Note that $\boldsymbol{\beta}$ is not transformed.

Apply the usual least-squares formula:

$$\begin{aligned}\boldsymbol{\beta}^* &= (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Z} \\ &= [(\mathbf{K}^{-1}\mathbf{X})'\mathbf{K}^{-1}\mathbf{X}]^{-1}(\mathbf{K}^{-1}\mathbf{X})'\mathbf{K}^{-1}\mathbf{Y} \\ &= [\mathbf{X}'(\mathbf{K}\mathbf{K}')^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{K}\mathbf{K}')^{-1}\mathbf{Y} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}\end{aligned}$$

$\boldsymbol{\beta}^*$ is the *Generalized Least Squares* (GLS) estimate.

12.3. Properties of the GLS Estimate $\boldsymbol{\beta}^*$

1. Unbiased:

$$E[\boldsymbol{\beta}^*] = E[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}]$$
$$= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E[\mathbf{Y}] = \boldsymbol{\beta}$$

2.
$$\mathrm{cov}(\boldsymbol{\beta}^*) = \sigma^2(\mathbf{B}'\mathbf{B})^{-1} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1},$$

3.

$$\begin{aligned} RSS &= (\mathbf{Z} - \mathbf{B}\boldsymbol{\beta}^*)'(\mathbf{Z} - \mathbf{B}\boldsymbol{\beta}^*) \\ &= [\mathbf{K}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)]'[\mathbf{K}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)] \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)'(\mathbf{K}\mathbf{K}')^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*) \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*) \end{aligned}$$

12.4. GLS versus OLS Estimates

Suppose $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$.

GLS (Generalized Least Squares):

$$\boldsymbol{\beta}^* = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}.$$

OLS (Ordinary Least Squares):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Both estimates are unbiased.

We saw $\mathrm{cov}(\boldsymbol{\beta}^*) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$.

We see that $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$

What if we go ahead and used the OLS estimates, even when $\mathrm{cov}(\mathbf{Y}) = \sigma^2\mathbf{V}$? Which is more efficient, GLS or OLS?

Theorem: (Optimality of GLS estimates). If $E[Y] = \mathbf{X}\boldsymbol{\beta}$ and $\text{cov}(\mathbf{Y}) = \sigma^2\mathbf{V}$, then for any constant vector $\mathbf{a}$, $\mathbf{a}'\boldsymbol{\beta}^*$ is the BLUE of $\mathbf{a}'\boldsymbol{\beta}$.

Proof: $\mathbf{a}'\boldsymbol{\beta}^*$ is unbiased because $\boldsymbol{\beta}^*$ is unbiased, and $\mathbf{a}'\boldsymbol{\beta}^*$ is linear in $\mathbf{Y}$ because

$$\mathbf{a}'\boldsymbol{\beta}^* = \mathbf{a}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \mathbf{b}'\mathbf{Y},$$

for $\mathbf{b}' = \mathbf{a}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$.

Let $\mathbf{b}_1'\mathbf{Y}$ be any linear unbiased estimate of $\mathbf{a}'\boldsymbol{\beta}$. Then

$$\mathbf{b}_1'\mathbf{Y} = \mathbf{b}_1'\mathbf{K}\mathbf{Z} = (\mathbf{K}'\mathbf{b}_1)'\mathbf{Z}$$

is also linear in $\mathbf{Z}$.

For $\mathbf{Z}$ we know that the least-squares estimate is the BLUE, $\mathbf{a}'\boldsymbol{\beta}^* = \mathbf{a}'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Z}$ has minimum variance among all unbiased estimates of $\mathbf{a}'\boldsymbol{\beta}$ that are linear in $\mathbf{Z}$, i.e.,

$$\text{var}(\mathbf{a}'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Z}) \leq \text{var}([\mathbf{K}'\mathbf{b}_1]'\mathbf{Z}),$$

with equality if and only if

$$(\mathbf{K}'\mathbf{b}_1)' = \mathbf{a}'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$$

i.e., $\mathbf{b}_1'\mathbf{K} = \mathbf{a}'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$

$$\begin{aligned}
\mathbf{b}_1' &= \mathbf{a}'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{K}^{-1} \\
&= \mathbf{a}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{K}^{-1})'\mathbf{K}^{-1} \\
&= \mathbf{a}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}.
\end{aligned}$$

Therefore, we have shown that

$$\text{var}(\mathbf{a}'\boldsymbol{\beta}^*) \leq \text{var}(\mathbf{b}_1'\mathbf{Y}),$$

with equality if and only if $\mathbf{b}_1'\mathbf{Y} = \mathbf{a}'\boldsymbol{\beta}^*$.

In-Class Exercise: (Heteroscedasticity and Weighted least squares). Let $Y_1, \ldots, Y_n$ be independent, $E[Y_i] = \beta x_i$, and $\text{var}(Y_i) = \sigma^2 w_i^{-1}$. Find the GLS and OLS estimates of β and their variances.

The GLS estimate of β is

$$\boldsymbol{\beta}^* = \frac{\sum_{i=1}^n w_i x_i Y_i}{\sum_{i=1}^n w_i x_i^2}.$$

The OLS estimate is

$$\hat{\boldsymbol{\beta}} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

The variances are

$$\text{var}(\boldsymbol{\beta}^*) = \frac{\sigma^2}{\sum_{i=1}^n w_i x_i^2},$$

$$\text{var}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2 \sum_{i=1}^n \frac{x_i^2}{w_i}}{(\sum_{i=1}^n x_i^2)^2}.$$

12.5. When are GLS and OLS Estimates are Equivalent?

Theorem: A necessary and sufficient condition for the GLS estimate ($\boldsymbol{\beta}^*$) and the OLS estimate ($\hat{\boldsymbol{\beta}}$) to be equal is $\mathcal{R}(\mathbf{V}^{-1}\mathbf{X}) = \mathcal{R}(\mathbf{X})$.

Proof: Let $\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2$, where $\mathbf{Y}_1 \in \mathcal{R}(\mathbf{X})$, and $\mathbf{Y}_2 \in \mathcal{R}(\mathbf{X})^\perp$, Then $\mathbf{Y}_1 = \mathbf{X}\mathbf{a}$ for some $\mathbf{a}$ and $\mathbf{X}'\mathbf{Y}_2 = \mathbf{0}$. Then

$$\begin{aligned}\boldsymbol{\beta}^* &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{Y}_1 + \mathbf{Y}_2)\\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{a} + (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}_2\\ &= \mathbf{a} + (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}_2,\end{aligned}$$

and

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y}_1 + \mathbf{Y}_2)\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{a} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_2\\ &= \mathbf{a} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_2 = \mathbf{a}.\end{aligned}$$

Therefore, $\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}$ if and only if

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}_2 = \mathbf{0}$$

if and only if

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}_2 = \mathbf{0} = (\mathbf{V}^{-1}\mathbf{X})'\mathbf{Y}_2,$$

That is, $\mathbf{Y}_2$, which is in $\mathcal{R}(\mathbf{X})^\perp$, must also be orthogonal to $\mathcal{R}(\mathbf{V}^{-1}\mathbf{X})$. So the the estimates are equal if and only if $\mathcal{R}(\mathbf{X})^\perp \subset \mathcal{R}(\mathbf{V}^{-1}\mathbf{X})^\perp$. But the two spaces have the same dimension, so $\mathcal{R}(\mathbf{X})^\perp = \mathcal{R}(\mathbf{V}^{-1}\mathbf{X})^\perp$ and $\mathcal{R}(\mathbf{X}) = \mathcal{R}(\mathbf{V}^{-1}\mathbf{X})$.

Corollary: The GLS and OLS estimates are equal if and only if $\mathcal{R}(\mathbf{V}\mathbf{X}) = \mathcal{R}(\mathbf{X})$.

Proof: Exercise.