Weighted Least Squares and Generalized Least Squares[1]

STA302 Fall 2020

[1] See last slide for copyright information.

Weighted and Generalized Least Squares

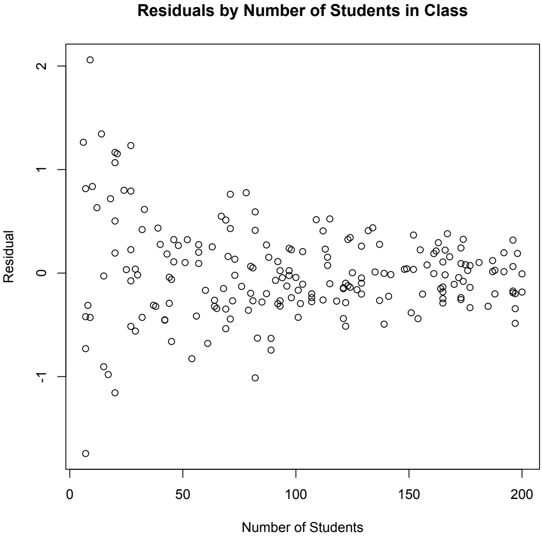
An antidote to unequal variance (of a certain kind)

Example: Aggregated data.Teaching evaluations. Have

- Mean ratings $\overline{y}_1, \ldots, \overline{y}_m$
- Number of students $n_1, \ldots, n_m$
- Lots of predictor variables.

$$Var(\overline{y}_i) = \frac{\sigma^2}{n_i}$$

Residual Plot



Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

$$
\begin{aligned}
cov(\boldsymbol{\epsilon}) &= \begin{pmatrix} \frac{\sigma^2}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma^2}{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma^2}{n_m} \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n_m} \end{pmatrix}
\end{aligned}
$$

Unknown σ^2 times a *known* matrix.

Generalize

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$
- $\mathbf{V}$ is a *known* symmetric positive definite matrix.
- A good estimate of $\mathbf{V}$ can be substituted and everything works out for large samples.

Generalized Least Squares

Transform the data.

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{ with } cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V} \\
\Longrightarrow \mathbf{V}^{-\frac{1}{2}}\mathbf{y} &= \mathbf{V}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\epsilon} \\
\mathbf{y}^* &= \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*
\end{aligned}
$$

Same $\boldsymbol{\beta}$.

$$
\begin{aligned}
cov(\boldsymbol{\epsilon}^*) &= cov(\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\epsilon}) \\
&= \mathbf{V}^{-\frac{1}{2}} cov(\boldsymbol{\epsilon}) \mathbf{V}^{-\frac{1}{2}\prime} \\
&= \mathbf{V}^{-\frac{1}{2}} (\sigma^2\mathbf{V}) \mathbf{V}^{-\frac{1}{2}} \\
&= \sigma^2\mathbf{I}
\end{aligned}
$$

Least Squares Estimate for the * Model is B.L.U.E.

$\mathbf{y}^* = \mathbf{X}^*\beta + \boldsymbol{\epsilon}^*$

- By Gauss-Markov, $\widehat{\boldsymbol{\beta}}^*$ will beat any other linear combination of the $\mathbf{y}^*$.
- $\mathbf{y}^* = \mathbf{V}^{-1/2}\mathbf{y}$.
- So any linear combination of the $\mathbf{y}$ is a linear combination of the $\mathbf{y}^*$.

$$\begin{aligned} \mathbf{c}'\mathbf{y} &= \mathbf{c}'\mathbf{V}^{1/2}\mathbf{V}^{-1/2}\mathbf{y} \\ &= \mathbf{c}'\mathbf{V}^{1/2}\mathbf{y}^* \\ &= \mathbf{c}_2'\mathbf{y}^* \end{aligned}$$

- And $\widehat{\boldsymbol{\beta}}^*$ beats it. It's B.L.U.E. for the original problem.

Generalized Least Squares

$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^* &= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* \\
&= \left((\mathbf{V}^{-\frac{1}{2}}\mathbf{X})'(\mathbf{V}^{-\frac{1}{2}}\mathbf{X})\right)^{-1}(\mathbf{V}^{-\frac{1}{2}}\mathbf{X})'\mathbf{V}^{-\frac{1}{2}}\mathbf{y}^* \\
&= (\mathbf{X}'\mathbf{V}^{-\frac{1}{2}\prime}\mathbf{V}^{-\frac{1}{2}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-\frac{1}{2}\prime}\mathbf{V}^{-\frac{1}{2}}\mathbf{y} \\
&= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}
\end{aligned}
$$

- So it is not necessary to literally transform the data.
- Convenient expressions for tests and confidence intervals are only a homework problem away.
- $\widehat{\boldsymbol{\beta}}^*$ is called the "generalized least squares" estimate of $\boldsymbol{\beta}$.
- If $\mathbf{V}$ is diagonal, it's called "weighted least squares."

Variance Proportional to x_i

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$$cov(\boldsymbol{\epsilon}) = \sigma^2 \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{pmatrix}$$

First-order Autoregressive Time Series

Estimate ρ with the first-order sample autocorrelation

$$cov(\boldsymbol{\epsilon}) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-3} \\ \rho^3 & \rho^2 & \rho & 1 & \cdots & \rho^{n-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 \end{pmatrix}$$

An amazing scalar example with no independent variables

- $y_{ij} \stackrel{i.i.d.}{\sim} ?(\mu, \sigma^2)$.
- Have $\overline{y}_1, \ldots, \overline{y}_m$ based on $n_1, \ldots, n_m$.
- $\overline{y}_j \stackrel{\cdot}{\sim} N(\mu, \frac{\sigma^2}{n_j})$ by the Central Limit Theorem.
- Want to estimate μ.
- A natural estimator is the mean of means: $\widehat{\mu}_1 = \frac{1}{m}\sum_{j=1}^m \overline{y}_j$.
- $E(\widehat{\mu}_1) = \mu$, so it's unbiased.
- $Var(\widehat{\mu}_1) = \frac{\sigma^2}{m^2}\sum_{j=1}^m \frac{1}{n_j}$. Can we do better?
- Noting that $\widehat{\mu}_1 = \sum_{j=1}^m \frac{1}{m}\overline{y}_j$ is a linear combination of the $\overline{y}_j$ with the weights adding to one ...

Try Weighted Least Squares

$\overline{y}_j = \mu + \epsilon_j$ with $E(\epsilon_j) = 0$ and $Var(\epsilon_j) = \frac{\sigma^2}{n_j}$

It's a regression with $\beta_0 = \mu$ and no explanatory variables.

$$cov(\boldsymbol{\epsilon}) = \sigma^2 \begin{pmatrix} \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n_m} \end{pmatrix} = \sigma^2 \mathbf{V}$$

Scalar Calculation

$$\begin{aligned} \overline{y}_j &= \mu + \epsilon_j \\ \Longrightarrow \sqrt{n_j}\,\overline{y}_j &= \sqrt{n_j}\,\mu + \sqrt{n_j}\,\epsilon_j \\ y_j^* &= x_j^* \beta_1^* + \epsilon_j^* \end{aligned}$$

- It's another regression model.
- This time there is no intercept, and μ is the slope.

$$\begin{aligned} Var(\epsilon_j^*) &= Var(\sqrt{n_j}\,\epsilon_j) \\ &= n_j Var(\epsilon_j) \\ &= n_j \frac{\sigma^2}{n_j} \\ &= \sigma^2 \end{aligned}$$

Least Squares for Simple Regression through the Origin

$y_j^* = x_j^* \beta_1^* + \epsilon_j^*$, with $\beta_1^* = \mu$, $y_j^* = \sqrt{n_j}\,\overline{y}_j$ and $x_j^* = \sqrt{n_j}$

$$
\begin{aligned}
\widehat{\beta}_1^{\,*} &= \frac{\sum_{j=1}^m x_j^* y_j^*}{\sum_{j=1}^m x_j^{*2}} \\
&= \frac{\sum_{j=1}^m \sqrt{n_j}\,\sqrt{n_j}\,\overline{y}_j}{\sum_{j=1}^m \sqrt{n_j}^2} \\
&= \frac{\sum_{j=1}^m n_j \overline{y}_j}{\sum_{j=1}^m n_j} \\
&= \sum_{j=1}^m \left(\frac{n_j}{\sum_{\ell=1}^m n_\ell} \right) \overline{y}_j
\end{aligned}
$$

- A linear combination of the $\overline{y}_j$; the weights add to one.
- B.L.U.E.

And not only that, but . . .

$$\begin{aligned}
\widehat{\beta}_1^{\,*} &= \frac{\sum_{j=1}^{m} n_j \overline{y}_j}{\sum_{j=1}^{m} n_j} \\
&= \frac{\sum_{j=1}^{m} n_j \frac{\sum_{i=1}^{n_j} y_{ij}}{n_j}}{\sum_{j=1}^{m} n_j} \\
&= \frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} y_{ij}}{\sum_{j=1}^{m} n_j} = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} y_{ij}}{n}
\end{aligned}$$

- So the B.L.U.E. of μ is just the sample mean of all the data.
- One more comment is that $\widehat{\boldsymbol{\beta}}^{*} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ yields the same expression.

Copyright Information