STAT 410 - Linear Regression

Lecture 11

Meng Li

Department of Statistics

Oct. 24, 2017

Generalized Least Squares

- What is a linear regression models does not have a constant variance? That is, the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

  where

$$\mathrm{E}(\boldsymbol{\varepsilon}) = 0, \quad \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}.$$

- Is the ordinary least square $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ still appropriate?
- If $\mathbf{V}$ is positive definite and symmetric, then there exists a positive definite and symmetric matrix $\mathbf{K}$ such that $\mathbf{K}\mathbf{K} = \mathbf{V}$.
- $\mathbf{K}$ is called the *principal* square root of $\mathbf{V}$.
- This allows us to transfer the original model to our familiar context where the error has constant variance:

$$\mathbf{K}^{-1}\mathbf{y} = \mathbf{K}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}^{-1}\boldsymbol{\varepsilon}.$$

Results for GLS

- After the transformation, all we have learned from OLS apply - we just use the transformed response, design matrix and error term.
  - Loss function: $S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.
  - Normal equations:

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

  - Solution:

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

  - Properties:

$$\mathrm{E}\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}, \quad \mathrm{Var}\hat{\boldsymbol{\beta}} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

  - $\hat{\beta}$ is the best linear unbiased estimators of $\boldsymbol{\beta}$ (BLUE).

A special case: weighted least squares

- Consider the case where the errors $\boldsymbol{\varepsilon}$ are uncorrected but the variances are unequal, i.e.,

$$\mathbf{V} = \mathrm{diag}(1/w_1, 1/w_2, \ldots, 1/w_n);$$

  $\mathrm{diag}(\boldsymbol{a})$ means a diagonal matrix with diagonal vector $\boldsymbol{a}$.
- The weights w_i's have to be positive because $\mathbf{V}$ is a covariance matrix.
- It follows that $\mathbf{K} = \mathrm{diag}(1/\sqrt{w_1}, 1/\sqrt{w_2}, \ldots, 1/\sqrt{w_n})$, and consequently

$$\mathbf{K}^{-1}\mathbf{y} = \begin{pmatrix} y_1\sqrt{w_1} \\ y_2\sqrt{w_2} \\ \vdots \\ y_n\sqrt{w_n} \end{pmatrix}, \quad \mathbf{K}^{-1}\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1\sqrt{w_1} \\ \varepsilon_2\sqrt{w_2} \\ \vdots \\ \varepsilon_n\sqrt{w_n} \end{pmatrix}.$$

- Similarly,

$$\mathbf{K}^{-1}\mathbf{X} = \begin{pmatrix} 1\sqrt{w_1} & x_{11}\sqrt{w_1} & \cdots & x_{1k}\sqrt{w_1} \\ 1\sqrt{w_2} & x_{21}\sqrt{w_2} & \cdots & x_{2k}\sqrt{w_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1\sqrt{w_n} & x_{n1}\sqrt{w_n} & \cdots & x_{nk}\sqrt{w_n} \end{pmatrix}$$

- The loss functions simplifies to

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i e_i^2,$$

  where e_i is the ith element of $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ with a slight abuse of notation.
- Practical motivations:
  - IF model diagnostics indicate $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_{ij}$ for some j, then we can let $w_i = 1/x_{ij}$ and refit the model using weighted least squares.
  - From the experimental design point of view, if y_i is an average of n_i i.i.d. observations, we then have $\mathrm{Var}(\varepsilon_i) = \sigma^2/n_i$ and thus $w_i = n_i$.

Application of GLS: mixed models

- SLR or MLR considers one source of variability using ε_i.
- However, many important experimental designs require the use of multiple sources of variability.
- Paper helicopters example: consider an experiment to determine the effect of the length of the helicopter's wings to the typical flight time.
  - There often is quite some error associated with measuring the time for a specific flight, especially when the people who are timing have no prior experience.
    - A popular protocol has three people timing each flight.
  - Helicopters vary as they may be made in a corporate short course where the students have never made these helicopters before.
  - As a result, this particular experiment has two sources of variability: within each specific helicopter and between the various helicopters used in the study.

SLR

- A simple linear regression model would be

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij},$$

  where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, r_i$.
  - m: number of helicopters
  - r_i number of measured flight times for the ith helicopter
  - y_{ij}: flight time for the jth flight of the ith helicopter
  - x_i: length of the wings for the ith helicopter
  - ε_{ij}: error term associated with the jth flight of the ith helicopter
- Does it make sense to assume ε_{ij} uncorrelated or even independent of each other?
- Random effects allow the analyst to take into account multiple sources of variability.

Mixed models

- A more sensible model for the helicopter example is:

$$y_{ij} = \beta_0 + \beta_1 x_i + \delta_i + \varepsilon_{ij},$$

  where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, r_i$.
  - m: number of helicopters
  - r_i number of measured flight times for the ith helicopter
  - y_{ij}: flight time for the jth flight of the ith helicopter
  - x_i: length of the wings for the ith helicopter
  - δ_i: error term associated with the ith helicopter
  - ε_{ij}: error term associated with the jth flight of the ith helicopter
- This is called a mixed model: fixed effects via x_i and random effects via δ_i.
- Random effect is viewed as a random sample from a population, thus its variability rather than itself is of our interest.

- The sample size is $n = \sum_{i=1}^{m} r_i$.
- We assume $\delta_i \sim (0, \sigma_\delta^2)$ and $\varepsilon_{ij} \sim (0, \sigma^2)$ and each is uncorrelated across i (or j).
- We can rewrite the mixed model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\delta + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_{11}, y_{12}, \ldots, y_{1r_1}, \ldots, y_{m1}, y_{m2}, \ldots, y_{mr_m})$, and

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_{r_1} & 0 & \ldots & 0 \\ 0 & \mathbf{1}_{r_2} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \mathbf{1}_{r_m} \end{pmatrix}$$

- It follows that $\mathrm{Var}(\mathbf{y}) = \sigma^2\mathbf{I} + \sigma_\delta^2\mathbf{Z}\mathbf{Z}'$.
- Therefore, the mixed model is equivalent to

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

where $\boldsymbol{\varepsilon}^*$ has mean $\mathbf{0}$ and covariance $\mathbf{V}$ where $\mathbf{V} = \sigma^2\mathbf{I} + \sigma_\delta^2\mathbf{Z}\mathbf{Z}'$.