CSSS/POLS 510 MLE

Lab 3. Heteroskedastic Normal

Minji Jeong

October 10, 2025

Housekeeping

- All Problem Set answers should be submitted on Canvas.
  - Submit your file in pdf form and make sure you present the R code you wrote (echo=TRUE).
  - Project proposal, final paper, and poster should be sent to Chris's email.
- PS 1 grade will be released next week.
- For office hours, please contact me 24 hours before the time you want to have a zoom meeting.
- Make sure you install Chris's simcf package.

Least Squares

- Linear homoskedastic:

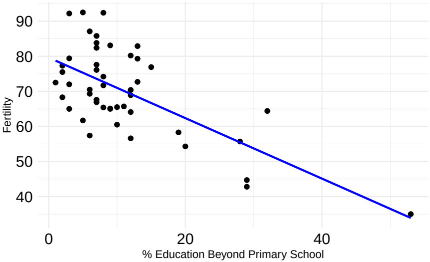
$$Y_i = x_i\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

- Estimating the slope:

$$\hat{\beta}_j = \frac{\text{Cov}(X_j, Y)}{\text{Var}(X_j)}$$

- Matrix Algebra:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$



Least Squares

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (1)$$

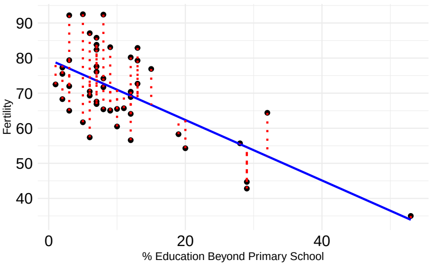
$$\hat{e}_i = y_i - \hat{y}_i \quad (2)$$

$$\sum \hat{e}_i{}^2 = \sum (y_i - \hat{y}_i)^2 \quad (3)$$

$$\sum \hat{e}_i{}^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (4)$$

Note: (3) and (4) are equivalent, and provide the *Residual Sum of Squares* (RSS) or *Sum of Squared Residuals* (SSR).

Choosing the best combination of $\hat{\beta}_j$ that minimizes the SSR provide the best fit for the model.



Maximum Likelihood Estimation

- How do we estimate the MLE?
  1. Define a probability model (PDF): $Y_i \sim N(\mu_i, \sigma^2)$.
  2. Derive the log-likelihood function.
  3. Reduce to sufficient statistics and substitute systematic component.
  4. Use `optim()` or any other function to find the maxima.

Normal homoskedastic

Two **different** notations for the **same** model.

LS notation:

$$\begin{aligned} \varepsilon_i &\sim N(0, \sigma^2) && (\textit{stochastic}) \\ Y_i &= x_i\beta && (\textit{systematic}) \\ Y_i &= x_i\beta + \varepsilon && (\textit{stochastic} + \textit{systematic}) \end{aligned}$$

MLE notation:

$$\begin{aligned} Y_i &\sim N(\mu_i, \sigma^2) && (\textit{stochastic}) \\ \mu_i &= x_i\beta && (\textit{systematic}) \\ Y_i &\sim N(x_i\beta, \sigma^2) && (\textit{stochastic} + \textit{systematic}) \end{aligned}$$

MLE general notation

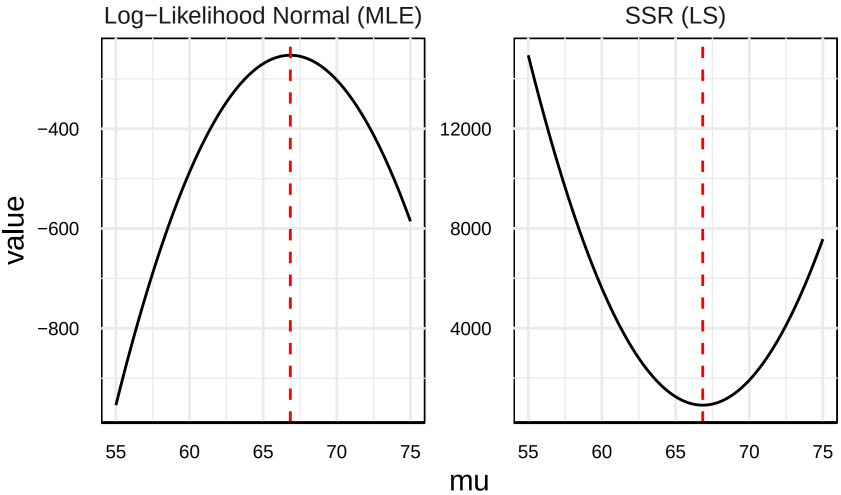
Systematic–Stochastic Decomposition:

$$
\begin{aligned}
Y_i &\sim f(\theta_i, \alpha) \quad \text{(stochastic)} \\
\theta_i &= g(\boldsymbol{x}_i \boldsymbol{\beta}) \quad \text{(systematic)}
\end{aligned}
$$

where

- Y_i is a outcome random variable.
- $f()$ is a probability density function.
- θ_i is a systematic feature of the PDF that varies over i.
- α is an ancillary parameter (feature of f that we treat as constant).
- $g()$ functional form for reparametrization of the data model.
- x_i explanatory variables vector.
- β vector of effect parameters.

MLE - Homoskedastic normal



2. Heteroskedastic normal

- Steps
  - The full `R` code can be found here from Chris' website
    - Generate Data
    - Fit OLS - `lm()`
    - Fit MLE - `optim()`

FIN