Advanced Data Analysis

Whiteboard Lecture

Longitudinal Data Analysis:

Whiteboard Interlude III

Here we are just considering the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2).$$

Note that

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma_\varepsilon^2)$$

so the likelihood is

$$\mathcal{L}(\beta_0, \beta_1, \sigma_\varepsilon^2) = \prod_{i=1}^{n} \left\{ (2\pi\sigma_\varepsilon^2)^{-1/2} e^{\frac{-(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma_\varepsilon^2}} \right\}.$$

Straightforward algebra then leads to the matrix algebraic expression for the log-likelihood:

$$\ell(\boldsymbol{\beta}, \sigma_\varepsilon^2) = -\frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) - \frac{n}{2}\ln(2\pi\sigma_\varepsilon^2)$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix}.$$

For all $\sigma_\varepsilon^2 > 0$, the maximimim likelihood estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

(as shown earlier in the course).

The profile log-likelihood for σ_ε^2 is

$$\begin{aligned}\ell_P(\sigma_\varepsilon^2) &= -\frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) - \frac{n}{2}\ln(2\pi\sigma_\varepsilon^2) \\ &= -\tfrac{1}{2}(\sigma_\varepsilon^2)^{-1}\text{RSS} - \frac{n}{2}\ln(\sigma_\varepsilon^2) - \frac{n}{2}\ln(2\pi)\end{aligned}$$

where

$$\begin{aligned}\text{RSS} &= (\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \boldsymbol{y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}\end{aligned}$$

and

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T.$$

$$\frac{\partial}{\partial\sigma_\varepsilon^2}\ell_P(\sigma_\varepsilon^2) = \tfrac{1}{2}(\sigma_\varepsilon^2)^{-2}\text{RSS} - \frac{n}{2}(\sigma_\varepsilon^2)^{-1}$$

which is zero if and only if

$$\sigma_\varepsilon^2 = \frac{\text{RSS}}{n}.$$

This leads to

$$\widehat{\sigma}^2_{\varepsilon,\text{ML}} = \frac{\text{RSS}}{n}.$$

Results for expectations of quadratic forms

lead to

$$E(\text{RSS}) = (n-2)\sigma_\varepsilon^2.$$

$$E(\widehat{\sigma}^2_{\varepsilon,\text{ML}}) = \left(\frac{n-2}{n}\right)\sigma_\varepsilon^2$$

so the maximum likelihood estimator of σ_ε^2 is slightly biased.

It is common to instead use

$$\frac{\mathsf{RSS}}{n-2}$$

to estimate σ_ε^2.

This corresponds to maximising the restricted log-likelihood.

i.e. the restricted maximum likelihood (REML) estimator of σ_ε^2 is:

$$\widehat{\sigma}^2_{\varepsilon,\mathsf{REML}} = \frac{\mathsf{RSS}}{n-2}.$$

The actual restricted log-likelihood is:

$$\ell_R(\sigma_\varepsilon^2) = \ell_P(\sigma_\varepsilon^2) - \tfrac{1}{2}\log|(1/\sigma_\varepsilon^2)\boldsymbol{X}^T\boldsymbol{X}|.$$

The expression for $\ell_R(\sigma_\varepsilon^2)$ can be derived using

- "contrast" arguments that account for estimation of the regression coefficients vector $\boldsymbol{\beta}$. One of the course reference books, McCulloch & Searle (2000), has the details.
- the theory of modified profile likelihood. (e.g. look up via Google or Wikipedia).

♦ ♦ ♦