

Regression

Gwangsu Kim

JBNU

Second semester of 2024

Normal Linear Model with Non-informative priors I

- Model

$$y_i = x_i^T \beta + \epsilon_i, i = 1, \dots, n, \epsilon_i \sim i.i.d. \mathcal{N}(0, \sigma^2),$$

where

$$x_i = (x_{i1}, \dots, x_{ik})^T \text{ and } \beta = (\beta_1, \dots, \beta_k).$$

Normal Linear Model with Non-informative priors II

With matrix notation, we can write

$$y = X\beta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I),$$

where

$$y = (y_1, \dots, y_n)^T,$$

$$\epsilon = (\epsilon_1, \dots, \epsilon_n)^T,$$

$$X = \begin{pmatrix} x_1^1 \\ x_2^T \\ \dots \\ x_n^T \end{pmatrix}.$$

Normal Linear Model with Non-informative priors III

- Non-informative prior

A widely accepted non-informative prior for the normal linear models is

$$\pi(\beta, \sigma^2 | X) \propto 1/\sigma^2.$$

- Posterior

$$\beta | \sigma^2, y \sim N(\hat{\beta}, V_{\beta}\sigma^2),$$

$$\sigma^{-2}|y \sim \mathcal{Gamma}((n - k)/2, (n - k)s^2/2),$$

Normal Linear Model with Non-informative priors IV

where

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

$$V_{\beta} = (X^T X)^{-1},$$

$$s^2 = \frac{1}{n - k} (y - X\hat{\beta})^T (y - X\hat{\beta}).$$

Normal Linear Model with informative priors I

- Prior

$$\beta_j \sim \mathcal{N}(\beta_{j0}, \sigma_{\beta_j}^2), j = 1, \dots, k.$$

- Posterior is proportional to

$$\prod_{i=1}^n \frac{\pi(\sigma^2)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(y_i - x_i^T \beta)^2}{\sigma^2}\right) \prod_{j=1}^k \frac{1}{\sqrt{2\pi\sigma_{\beta_j}^2}} \exp\left(-\frac{1}{2} \frac{(\beta_j - \beta_{j0})^2}{\sigma_{\beta_j}^2}\right).$$

- * The right term is interpreted as pseudo likelihood.

Example: incumbency in congressional elections I

- U.S. House of Representatives election data (from 1900 to 2000)
 - 1 There are 435 single-member districts, about 100 to 150 elections are uncontested.
 - 2 Incumbency effects: the advantage of incumbent candidate.
 - 3 For fixed year, we have the contested district (considering only two parties) i , labeling $R_i = 1$ when incumbent candidate runs, 0 when not.

Example: incumbency in congressional elections II

■ Modeling of incumbent effects

incumbency advantage $i = y_{complete,i}^I - y_{complete,i}^I$

- 1 $y_{complete,i}^I$: proportion of vote in the district i received by the incumbent legislator.
- 2 $y_{complete,i}^I$: proportion of vote in the district i received by the incumbent party, if the incumbent legislator does not run.

Example: incumbency in congressional elections III

■ Regression and adjusted variables

- 1 Using R_i as covariates.
- 2 Considering the confound effects such as incumbent party, vote proportion in 1986.

Example: incumbency in congressional elections IV

■ Results

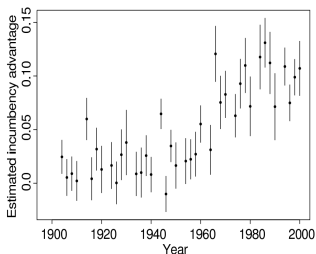


Figure 14.2 *Incumbency advantage over time: posterior median and 95% interval for each election year. The inference for each year is based on a separate regression. As an example, the results from the regression for 1988, based on the data in Figure 14.1, are displayed in Table 14.1.*

Example: incumbency in congressional elections V

■ Outliers, how to find?

- 1 Using predictive distribution of residual, $(y_{it} - x_{it}\beta)/s_t$.
 - 1 Draw (β, σ^2) from their posterior distribution.
 - 2 Draw a hypothetical replication, y^{rep} , from the predictive distribution, $y^{rep} \sim \mathcal{N}(X\beta, \sigma^2 I)$, given the drawn values of (β, σ^2) and existing vector.
 - 3 Run a regression of y^{rep} on X and save the residuals.
- 2 Observe quantile of predictive distribution and compare it with observed residuals, $(y_{it} - x_{it}\hat{\beta})/s_t$, where $\hat{\beta}$ is the Bayes estimate and s_t is the estimated standard deviation.

Example: incumbency in congressional elections VI

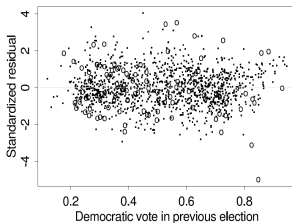


Figure 14.3 Standardized residuals, $(y_{it} - X_{it}\hat{\beta})/s_t$, from the incumbency advantage regressions for the 1980s, vs. Democratic vote in the previous election. (The subscript t indexes the election years.) Dots and circles indicate district elections with incumbents running and open seats, respectively.

Other issues I

- Sensitivity analysis
- Priors such as

$$\pi(1/\sigma_j^2) \sim \textit{Gamma}(\alpha_i, \beta_i)$$

can reduce the outliers.