BAYESIAN STATISTICS

Chapter 11

Instructor: Seokho Lee (이석호)

Hankuk University of Foreign Studies

11. Regression Analysis

11.1. Introduction

Regression analysis analyzes the relation between the **independent variables** (causes) and the **dependent variable** (result).

- simple regression analysis: a single independent variable affects on the dependent variable.
- multiple regression analysis: a set of multiple independent variables affects on the dependent variable.

11.2. Simple linear regression

A simple linear regression model is given as

$$y_i|x_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

with

- y_i : dependent variable for the ith observation
- x_i : independent variable for the ith observation
- ϵ_i : error term for the ith observation. We assume that $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$
- $\beta_0,\ \beta_1$: regression coefficients. They are parameters in the regression model

The regression model is totally specified by the regression coefficients, β_0 and β_1. Thus, estimation of them is of most interest.

Least square method gives the estimators for β_0 and β_1 by minimizing the sum of squared residuals. The **residual** is defined as $e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

The sum of squared residuals is given as

$$Q = \sum_{i=1}^{n}(y - \hat{y}_i)^2 = \sum_{i=1}^{n} \left\{ y - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right\}^2 .$$

To find out the solution which minimizes Q, we first take the partial derivative of Q with respect to β_0 and β_1, and set them zero:

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i-1}^{n} \left\{ y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right\} = 0$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i-1}^{n} x_i \left\{ y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right\} = 0$$

These gives the **normal equation** as follows:

$$\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i$$
$$\sum_{i=1}^{n} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$$

By solving the normal equations, we get the estimates:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Example (11-1)

Suppose that y_i is the amount of crop yield in the ith year and x_i is the amount of fertilizer used in the same year. Using the data during 20 years, built the simple linear regression model and give the estimates of the regression coefficients.

year	yield y_i	fertilizer x_i	year	yield y_i	fertilizer x_i
1922	423	39	1932	372	22
1923	483	60	1933	381	17
1924	479	42	1934	419	27
1925	486	52	1935	449	33
1926	494	47	1936	511	48
1927	498	51	1937	520	51
1928	511	45	1938	477	33
1929	534	60	1939	517	46
1930	478	39	1940	548	54
1931	440	41	1941	629	100

(solution) From the table, the estimates are $\hat{\beta}_0 = 345$ and $\hat{\beta}_1 = 3.00$ (verify them). Thus, the estimated regression line is

$$\hat{y}_i = 345 + 3x_i.$$

11.3. Simple linear regression: Bayesian approach

For the simple regression model,

$$y_i|x_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

with $\epsilon \sim \mathrm{N}(0, \sigma^2)$, we have three parameters, β_0, β_1, and σ^2. There are three cases:

- case 1: β_0 and β_1 are unknown, σ^2 is known.
- case 2: β_0 and β_1 are known, σ^2 is unknown.
- case 3: β_0 and β_1 are unknown, σ^2 is unknown.

The likelihood from the data is

$$\ell(\mathbf{y}|\mathbf{x}, \beta_0, \beta_1, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

Known variance, unknown coefficients

Suppose the variance is known as $\sigma^2 = s^2$. Assuming no prior information, we may take a vague prior for the regression coefficients (β_0, β_1) as

$$g(\beta_0, \beta_1) = g(\beta_0)g(\beta_1) \propto \text{constant}, \qquad -\infty < \beta_0, \beta_1 < \infty$$

where $g(\beta_0) \propto$ constant and $g(\beta_1) \propto$ constant.

From the Bayes theorem, the joint posterior distribution of (β_0, β_1) becomes

$$h(\beta_0, \beta_1 | \mathbf{x}, \mathbf{y}) \propto \exp\left\{-\tfrac{1}{2s^2} \textstyle\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right\}.$$

To obtain the marginal posteriors, consider the least square estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. Note that

$$\begin{aligned} & h(\beta_0, \beta_1 | \mathbf{x}, \mathbf{y}) \\ \propto \quad & \exp\left[-\tfrac{1}{2s^2}\left\{n(\beta_0 - \hat{\beta}_0)^2 + (\beta_1 - \hat{\beta}_1)^2 \textstyle\sum_{i=1}^n x_i^2 + 2(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1)\sum_{i=1}^n x_i\right\}\right]. \end{aligned}$$

This joint posterior is the bivariate normal distribution. Thus, the marginal posteriors are

$$h(\beta_0|\mathbf{x},\mathbf{y}) \propto \exp\left\{-\frac{(\beta_0-\hat{\beta}_0)^2}{2s_0^2}\right\}, \qquad s_0^2 = \frac{n^{-1}s^2\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i-\bar{x})^2},$$

$$h(\beta_1|\mathbf{x},\mathbf{y}) \propto \exp\left\{-\frac{(\beta_1-\hat{\beta}_1)^2}{2s_1^2}\right\}, \qquad s_1^2 = \frac{s^2}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

or

$$\frac{\beta_0-\hat{\beta}_0}{s_0}\Big|\mathbf{x},\mathbf{y} \sim \mathrm{N}(0,1), \qquad \frac{\beta_1-\hat{\beta}_1}{s_1}\Big|\mathbf{x},\mathbf{y} \sim \mathrm{N}(0,1).$$

Since the Bayes estimators are the marginal posterior means,

$$E(\beta_0|\mathbf{x},\mathbf{y}) = \hat{\beta}_0, \qquad E(\beta_1|\mathbf{x},\mathbf{y}) = \hat{\beta}_1$$

Example (11-2)

Suppose $s = 1.5$. *Inspect the relation between* x *and* y*:*

$$\{(1,6),\ (2.9),\ (3,12),\ (4,15),\ (5,12),\ (6,18),\ (7,18),\ (8,21),\ (9,24)\}$$

(solution) For the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, the Bayes estimates for the regression coefficients are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 2.05, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4.75.$$

The estimated regression line is

$$\hat{y}_i = 4.75 + 2.05 x_i.$$

Since $s = 1.5$, $s_0 = 1.090$ and $s_1 = 0.194$. The marginal posterior distributions are

$$(\beta_0 - 4.75)/1.090 \sim \mathrm{N}(0,1), \qquad (\beta_1 - 2.05)/0.194 \sim \mathrm{N}(0,1).$$

Thus, 95% EPD credible intervals are

$$4.75 \pm (1.96)(1.090) = (2.61, 6.89) \text{ for } \beta_0$$

and

$$2.05 \pm (1.96)(0.194) = (1.67, 2.43) \text{ for } \beta_1$$

11.3.2. Known coefficients, unknown variance

Suppose (β_0, β_1) is known as (b_0, b_1). The vague prior distribution for $\log \sigma$ is assumed to be

$$g(\sigma) \propto \frac{1}{\sigma}, \qquad 0 < \sigma < \infty.$$

Then the posterior for σ is

$$\begin{aligned} h(\sigma|\mathbf{x}, \mathbf{y}) &\propto \frac{1}{\sigma} \cdot \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2\right\} \\ &\propto \sigma^{-n-1} \exp\left\{-\frac{ns^2}{2\sigma^2}\right\}, \end{aligned}$$

where $ns^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$. This posterior is the inverse gamma distribution. The Bayes estimator for σ is the posterior mean which is

$$\hat{\sigma}_B = E(\sigma|\mathbf{x}, \mathbf{y}) = \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \left(\frac{n}{2}\right)^{1/2} s$$

Example (11-2, continue)

Suppose $b_0 = 5$ and $b_1 = 2$. Based on the same data, give the Bayes estimator $\hat{\sigma}_B$ and its 95% EPD credible interval.

(solution) From the data, $9s^2/\sigma^2 \sim \chi^2_9$, $s^2 = \sum_{i=1}^{9}(y_i - b_0 - b_1 x_i)^2/9 = 2$. The Bayes estimate is

$$\hat{\sigma}_B = \frac{\Gamma(4)}{\Gamma(4.5)}\sqrt{4.5}\sqrt{2} = 1.55.$$

Note that

$$\begin{aligned} & \Pr(2.70 = \chi^2_{0.026,9} \le 18/\sigma^2 \le \chi^2_{0.975,9} = 19.02) = 0.95 \\ \Rightarrow\quad & \Pr(0.9464 \le \sigma^2 \le 6.6) = 0.95 \\ \Rightarrow\quad & \Pr(0.98 \le \sigma \le 2.58) = 0.95. \end{aligned}$$

Thus, 95% EPD credible interval for σ is $(0.98, 2.58)$.

11.3.3. Unknown coefficients, unknown variance

This is most common in practice. The popularly used joint prior distribution for $(\beta_0, \beta_1, \log\sigma)$ is a vague prior given as

$$g(\beta_0, \beta_1, \sigma) = g(\beta_0)g(\beta_1)g(\sigma) \propto \frac{1}{\sigma}, \qquad -\infty < \beta_0, \beta_1 < \infty, \quad 0 < \sigma < \infty.$$

The joint posterior distribution becomes

$$\begin{aligned}
h(\beta_0, \beta_1, \sigma|\mathbf{x}, \mathbf{y}) \quad &\propto \quad \tfrac{1}{\sigma} \cdot \tfrac{1}{\sigma^n} \exp\left\{-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right\} \\
&\propto \quad \tfrac{1}{\sigma^{n+1}} \exp\left\{-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right\} \\
&\propto \quad \tfrac{1}{\sigma^{n+1}} \exp\Big[-\tfrac{1}{2\sigma^2} \big\{s^2(n-2) + n(\beta_0 - \hat\beta_0)^2 + (\beta_1 - \hat\beta_1)^2 \textstyle\sum_{i=1}^n x_i^2 \\
&\qquad +2(\beta_0 - \hat\beta_0)(\beta_1 - \hat\beta_1) \textstyle\sum_{i=1}^n x_i\big\}\Big] \\
&\propto \quad \tfrac{1}{\sigma^{n-1}} \exp\left\{-\tfrac{s^2(n-2)}{2\sigma^2}\right\} \times \tfrac{1}{\sigma^2} \exp\Big[-\tfrac{1}{2\sigma^2}\big\{n(\beta_0 - \hat\beta_0)^2 \\
&\qquad (\beta_1 - \hat\beta_1)^2 \textstyle\sum_{i=1}^n x_i^2 + 2(\beta_0 - \hat\beta_1)(\beta_1 - \hat\beta_1) \textstyle\sum_{i=1}^n x_i\big\}\Big],
\end{aligned}$$

where $\hat\beta_0 = \bar{y} - \hat\beta_1 \bar{x}$, $\hat\beta_1 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (x_i - \bar{x})^2$ and $s^2 = \sum_{i=1}^n (y_i - \hat\beta_0 - \hat\beta_1 x_i)^2/(n-2)$.

Note that the joint posterior for (β_0, β_1) is

$$\begin{aligned} h_1(\beta_0, \beta_1 | \mathbf{y}, \mathbf{x}) \quad \propto \quad & \tfrac{1}{\sigma^2} \exp\Big[-\tfrac{1}{2\sigma^2}\big\{(\beta_0 - \hat{\beta}_0)^2 + (\beta_1 - \hat{\beta}_1)^2 \textstyle\sum_{i=1}^n x_i^2 \\ & +2(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) \textstyle\sum_{i=1}^n x_i\big\}\Big]. \end{aligned}$$

The marginal posterior distributions are

$$h(\beta_0|\mathbf{x}, \mathbf{y}) \propto \left\{(n-2) + \frac{(\beta_0 - \hat{\beta}_0)^2}{2s_0^2}\right\}^{-(n-1)/2}, \qquad s_0^2 = \frac{\frac{s^2}{n}\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$h(\beta_1|\mathbf{x}, \mathbf{y}) \propto \left\{(n-2) + \frac{(\beta_1 - \hat{\beta}_1)^2}{2s_1^2}\right\}^{-(n-1)/2}, \qquad s_1^2 = \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Thus,

$$\left.\frac{\beta_0 - \hat{\beta}_0}{s_0}\right| \mathbf{x}, \mathbf{y} \sim t(n-2), \qquad \left.\frac{\beta_1 - \hat{\beta}_1}{s_1}\right| \mathbf{x}, \mathbf{y} \sim t(n-2)$$

The associated $100(1-\alpha)\%$ EPD credible intervals are

- $100(1-\alpha)\%$ for β_0: $\hat{\beta}_0 \pm t_{n-2,\alpha/2} s_0$
- $100(1-\alpha)\%$ for β_1: $\hat{\beta}_1 \pm t_{n-2,\alpha/2} s_1$

Example (11-2, continue)

Find 95% EPD credible intervals for β_0 and β_1.

(solution) Since $(\beta_0 - 4.75)/1.160 \sim t(7)$ and $(\beta_1 - 2.05)/0.206 \sim t(7)$, using $t_{0.025,7} = 2.365$, for β_0

$$4.75 \pm (2.365)(1.160) = (2.01, 7.49)$$

and for β_1,

$$2.05 \pm (2.365)(0.206) = (1.56, 2.54)$$

Now, we consider the marginal posterior for σ, which is obtained by

$$\begin{aligned} h(\sigma|\mathbf{x},\mathbf{y}) &= \int\int h(\beta_0,\beta_1,\sigma|\mathbf{x},\mathbf{y})d\beta_0 d\beta_1 \\ &= \int\int \frac{1}{\sigma^{n+1}} e^{-\frac{q(\beta_0,\beta_1)}{2\sigma^2}} d\beta_0 d\beta_1, \end{aligned}$$

where

$$q(\beta_0,\beta_1) = (n-2)s^2 + n(\beta_0-\hat{\beta}_0)^2 + (\beta_1-\hat{\beta}_1)^2 \textstyle\sum_{i=1}^n x_i^2 + 2(\beta_0-\hat{\beta}_0)(\beta_1-\hat{\beta}_1)\sum_{i=1}^n x_i.$$

After integration, the posterior is the inverse gamma distribution:

$$h(\sigma|\mathbf{x},\mathbf{y}) \propto \sigma^{-n-1} e^{-\frac{(n-2)s^2}{2\sigma^2}}, \quad 0<\sigma<\infty,\ n>2.$$

Thus, the Bayes estimator is

$$\hat{\sigma}_B = E(\sigma|\mathbf{x},\mathbf{y}) = \frac{\Gamma(\frac{n-3}{2})}{\Gamma(\frac{n-2}{2})}\left(\frac{n-2}{2}\right)^{1/2} s$$

and $100(1-\alpha)\%$ credible interval is

$$\left(\left[\frac{(n-2)s^2}{\chi^2_{n-2,\alpha/2}}\right]^{1/2}, \left[\frac{(n-2)s^2}{\chi^2_{n-2,1-\alpha/2}}\right]^{1/2}\right)$$

Example (11-2, continue)

Find the 95% credible interval for σ.

(solution) Since $(7)(2.55)/\sigma^2 \sim \chi^2_7$,

$$\Pr\left(1.69 \leq \frac{17.85}{\sigma^2} \leq 16.01\right) = 0.95 \quad \Rightarrow \quad \Pr(1.06 \leq \sigma \leq 3.25) = 0.95.$$

Thus, $(1.06, 3.25)$.

11.4. Predictive distribution

Consider the example 11-1 where the relation between y and x is $\hat{y}_i = 340 + 3.0x_i$. If $x = x^*$ then we predict y as

$$y^* = 340 + 3.0x^*.$$

The predictive distribution for y^* is

$$p(y^*|\mathbf{y}) = \int f(y^*, \theta|\mathbf{y})d\theta = \int \ell(y^*|\theta, \mathbf{y})h(\theta|\mathbf{y})d\theta = \int \ell(y^*|\theta)h(\theta|\mathbf{y})d\theta$$

Thus, in the regression problem, the predictive distribution for y^* is

$$p(y^*|x^*, \mathbf{x}, \mathbf{y}) = \int\int\int \ell(y^*|x^*, \beta_0, \beta_1, \sigma)h(\beta_0, \beta_1, \sigma|\mathbf{x}, \mathbf{y})d\beta_0 d\beta_1 d\sigma$$

which gives

$$\left.\frac{y^* - \hat{\beta}_0 - \hat{\beta}_1 x^*}{\hat{\sigma}\left\{1 + \frac{1}{n} + \frac{x^* - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\}^{1/2}}\right| \mathbf{x}, \mathbf{y} \sim t_{n-2}.$$

The corresponding $100(1-\alpha)\%$ EPD predictive interval for y^* at $x = x^*$ is

$$(\hat{\beta}_0 + \hat{\beta}_1 x^*) \pm t_{n-2,\alpha/2}\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{x^* - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Example (11-1, continue)

Suppose $x^* = 100$. *Find the 95% EPD predictive interval for* y^*.

(solution) When $x^* = 100$, $\frac{y^* - 645}{32.28} | \mathbf{x}, \mathbf{y} \sim t_{18}$. And, since $t_{0.025,18} = 2.101$,

$$
\begin{aligned}
& \Pr(-t_{0.025,18} \le \tfrac{y^* - 645}{32.28} \le t_{0.025,18}) = 0.95 \\
\Rightarrow \quad & \Pr(645 - (2.101)(32.28) \le y^* \le 645 + (2.101)(32.28)) = 0.95 \\
\Rightarrow \quad & \Pr(577.18 \le y^* \le 712.82) = 0.95.
\end{aligned}
$$

Thus, 95% EPD predictive interval for y^* at $x^* = 100$ is $(577.18, 712.82)$.