

## Ch8. Appendix

- 단순선형회귀모형에서 회귀계수의 최소제곱 추정량의 유도

$SS(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$ 를 최소로 하는  $\hat{\alpha}$ 과  $\hat{\beta}$ 을 찾는 문제.

$$\begin{cases} \frac{\partial SS}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 & (1) \\ \frac{\partial SS}{\partial \hat{\beta}} = -2 \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 & (2) \end{cases}$$

(1)을 정리하면  $\hat{\alpha}$ 에 관한 식을 얻을 수 있음

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) &= 0 \\ \sum_{i=1}^n y_i - n\hat{\alpha} - \hat{\beta} \sum_{i=1}^n x_i &= 0 \\ \therefore \hat{\alpha} &= \frac{\sum_{i=1}^n y_i}{n} - \hat{\beta} \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - \hat{\beta}\bar{x} \end{aligned}$$

$\hat{\alpha}$ 에 관한 식을 (2)에 대입한 뒤  $\hat{\beta}$ 에 대해 정리해 주면 아래와 같음.

$$\begin{aligned} \sum_{i=1}^n x_i (y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i) &= 0 \\ \sum_{i=1}^n x_i (y_i - \bar{y}) &= \hat{\beta} \sum_{i=1}^n x_i (x_i - \bar{x}) \\ \therefore \hat{\beta} &= \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} \\ \times \sum_{i=1}^n x_i (y_i - \bar{y}) &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^n (x_i - \bar{x})y_i \text{이고,} \\ \sum_{i=1}^n x_i (x_i - \bar{x}) &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \text{이므로,} \\ \hat{\beta} \text{의 분자는 } \sum_{i=1}^n (x_i - \bar{x})y_i \text{로, 분모는 } \sum_{i=1}^n (x_i - \bar{x})^2 \text{로 표현할 수도 있음.} \end{aligned}$$

따라서 최소제곱 추정량은  $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ ,  $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 가 됨.

- $\hat{\beta}$ 의 분포

회귀 모형의 가정

$$y_i \sim iid \quad Normal [\alpha + \beta x_i, \sigma^2]$$

임을 이용하면,

$$\hat{\beta} \sim Normal \left[ \beta, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

를 따르는 것을 보일 수 있다.

1)  $\hat{\beta} \sim Normal$  을 따른다.

최소제곱추정량  $\hat{\beta}$ 은  $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 로 구해진다.

최소제곱추정량  $\hat{\beta}$ 에서  $(x_i - \bar{x}), i = 1, \dots, n$ 은 주어진 상수로,  $y_i, i = 1, \dots, n$ 만 확률변수에

해당한다. 여기서 확률변수인  $y_i$ 가 모두 서로 독립인 정규 확률변수이며,  $\hat{\beta}$ 은 그런 각각의  $y_i$ 에 상수  $\frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 를 곱하고 더한 형태이므로,  $\hat{\beta}$ 도 정규분포를 따르게 된다 (정규 분포의 선형불변성).

2)  $\hat{\beta}$ 의 기대값  $E[\hat{\beta}]$ 은  $\beta$ 가 된다.

$$\begin{aligned} E[\hat{\beta}] &= E\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x}) E[y_i]}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\because (x_i - \bar{x}) \text{와 } \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{는 상수}) \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x}) (\alpha + \beta x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \alpha \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta \frac{\sum_{i=1}^n x_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \beta \quad (\because \sum_{i=1}^n (x_i - \bar{x}) = 0 \text{ 이고, } \sum_{i=1}^n x_i (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2) \end{aligned}$$

3)  $\hat{\beta}$ 의 분산  $V[\hat{\beta}]$ 은  $\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 가 된다.

$$\begin{aligned} V[\hat{\beta}] &= V\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 V(y_i)}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \quad (\because (x_i - \bar{x}) \text{와 } \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{는 상수, } y_i \text{는 서로 독립}) \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \quad (\because V(y_i) = \sigma^2) \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

## ● $\hat{\alpha}$ 의 분포

$$y_i \sim iid \quad Normal [\alpha + \beta x_i, \sigma^2]$$

임을 이용하면,

$$\hat{\alpha} \sim Normal \left[ \alpha, \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right]$$

를 따르는 것을 보일 수 있다.

1)  $\hat{\alpha} \sim Normal$ 을 따른다.

최소제곱추정량  $\hat{\alpha}$ 은

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta} \bar{x} \\ &= \frac{1}{n} \sum_{i=1}^n y_i - \hat{\beta} \bar{x} \\ &= \frac{1}{n} \sum_{i=1}^n y_i - \frac{\sum_{i=1}^n (x_i - \bar{x}) \bar{x} y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

이므로,

$\hat{\alpha}$ 도 독립인 정규확률변수인  $y_i$ 의 선형결합 형태다. 따라서  $\hat{\alpha}$ 도 정규분포를 따르게 된다 (정규분포의 선형불변성).

2)  $\hat{\alpha}$ 의 기대값  $E[\hat{\alpha}]$ 은  $\alpha$ 가 된다.

$$E[\hat{\alpha}] = E\left[\frac{1}{n} \sum_{i=1}^n y_i - \hat{\beta} \bar{x}\right]$$

$$\begin{aligned}
&= \frac{\sum_{i=1}^n E[y_i]}{n} - \bar{x} E[\hat{\beta}] \\
&= \frac{\sum_{i=1}^n (\alpha + \beta x_i)}{n} - \bar{x} \beta \\
&= \frac{\sum_{i=1}^n \alpha}{n} + \beta \frac{\sum_{i=1}^n x_i}{n} - \bar{x} \beta \\
&= \alpha + \beta \bar{x} - \beta \bar{x} = \alpha
\end{aligned}$$

3)  $\hat{\alpha}$ 의 분산  $V[\hat{\alpha}]$  는  $\sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$ 가 된다.

$$\begin{aligned}
V[\hat{\alpha}] &= V \left( \frac{1}{n} \sum_{i=1}^n y_i - \frac{\sum_{i=1}^n (x_i - \bar{x}) \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} y_i \right) \\
&= V \left( \sum_{i=1}^n \left( \frac{1}{n} - \frac{(x_i - \bar{x}) \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) y_i \right) \\
&= \sum_{i=1}^n \left( \frac{1}{n} - \frac{(x_i - \bar{x}) \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 V(y_i) \\
&= \sum_{i=1}^n \left( \frac{1}{n} - \frac{(x_i - \bar{x}) \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^2 \sigma^2 \\
&= \sigma^2 \sum_{i=1}^n \left( \left( \frac{1}{n} \right)^2 + \frac{(x_i - \bar{x})^2 \bar{x}^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} - 2 \left( \frac{1}{n} \right) \left( \frac{(x_i - \bar{x}) \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right) \\
&= \sigma^2 \left( \sum_{i=1}^n \left( \frac{1}{n} \right)^2 + \frac{\bar{x}^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} - 2 \left( \frac{1}{n} \right) \left( \frac{\bar{x} \sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \right) \\
&= \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - 0 \right) \quad (\because \sum_{i=1}^n (x_i - \bar{x}) = 0) \\
&= \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)
\end{aligned}$$

### ● y의 변동성의 분해

$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 로 분해된다.

$$\begin{aligned}
\sum_{i=1}^n (y_i - \bar{y})^2 &\stackrel{A}{=} \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\
&= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n e_i (\hat{y}_i - \bar{y}) \\
&= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 (\sum_{i=1}^n e_i \hat{y}_i + \sum_{i=1}^n e_i \bar{y}) \\
&= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2
\end{aligned}$$

$$\sum (A+B)^2$$

$$= \sum A^2 + \sum B^2 + \frac{2 \sum AB}{0}$$

마지막 식을 정리하는 과정에서  $\sum_{i=1}^n e_i \bar{y}$ 와  $\sum_{i=1}^n e_i \hat{y}_i$ 가 모두 0인 이유는 다음과 같다.

최소제곱추정량  $\hat{\alpha}$ 와  $\hat{\beta}$ 는 아래 식으로 구해지므로,

최소제곱법에 의한 잔차( $e_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$ )는  $\sum_{i=1}^n e_i = 0$ ,  $\sum_{i=1}^n x_i e_i = 0$ 를 만족함.

$$\begin{cases} \frac{\partial SS}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i) = -2 \sum_{i=1}^n e_i = 0 & (1) \\ \frac{\partial SS}{\partial \hat{\beta}} = -2 \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta} x_i) = -2 \sum_{i=1}^n x_i e_i = 0 & (2) \end{cases}$$

따라서  $\sum_{i=1}^n e_i \bar{y}$ 와  $\sum_{i=1}^n e_i \hat{y}_i$ 는 모두 0이 됨.

$$\sum_{i=1}^n e_i \bar{y} = \bar{y} \sum_{i=1}^n e_i = 0$$

$$\sum_{i=1}^n e_i \hat{y}_i = \sum_{i=1}^n e_i (\hat{\alpha} + \hat{\beta} x_i) = \hat{\alpha} (\sum_{i=1}^n e_i) + \hat{\beta} (\sum_{i=1}^n e_i x_i) = 0$$