최소제곱법, 벡터미분, 정규방정식

시립대학교 통계학과

2019년 6월 3일

1 단순선형회귀분석

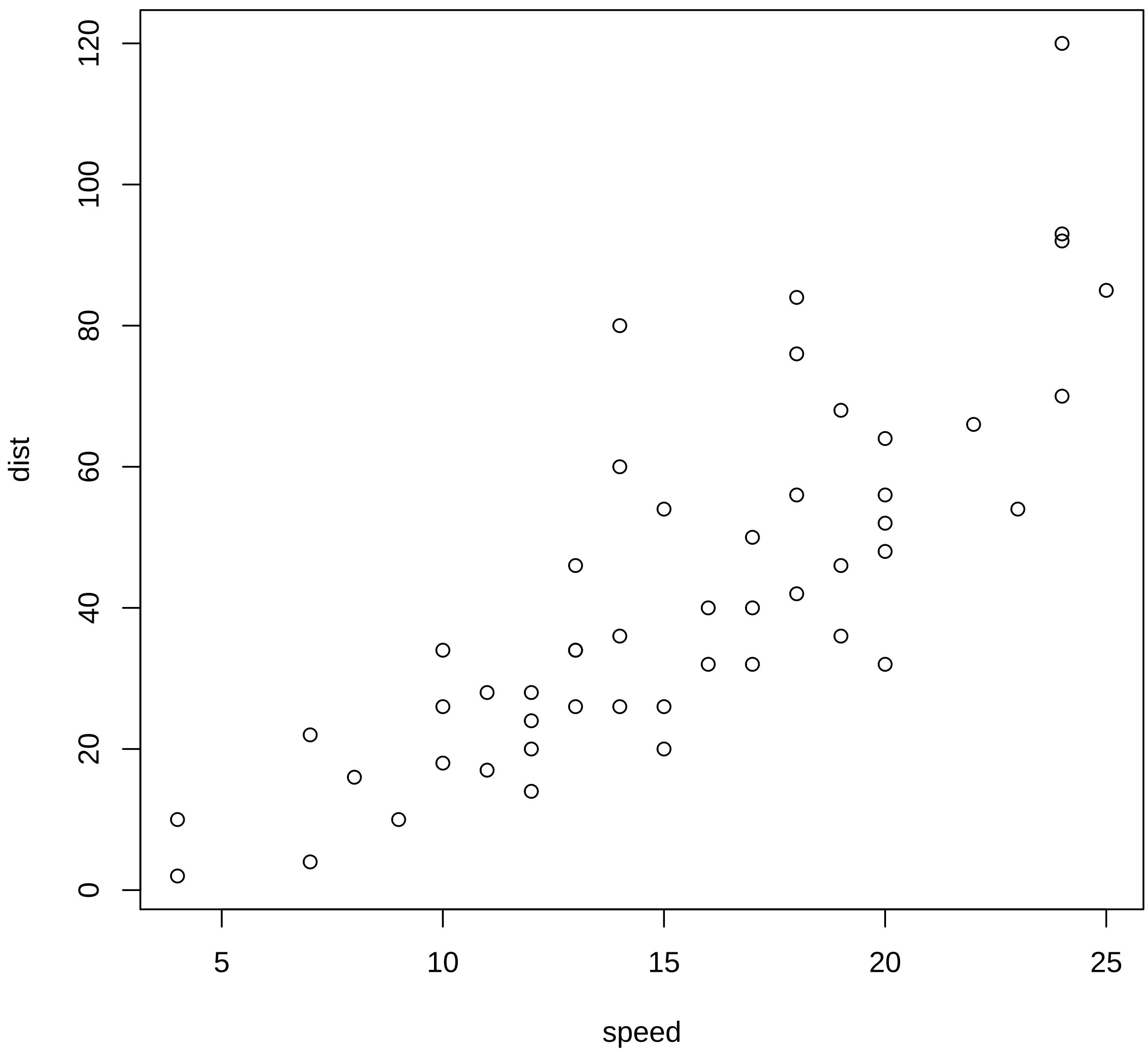
1.1 예제: 자동차의 속도와 제동거리

자동차가 달리는 속도(speed,mph)와 제동거리(dist,ft)의 관계를 알아보기 위하여 50대의 자동차로 실험한 결과 자료와 산포도가 아래와 같다.

```
head(cars)

##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10

plot(cars)
```



위와 같은 자료를 이용하여 자동차의 속도가 주어졌을 경우 제동거리의 평균에 대한 예측을 하려고 한다면 어떤 방법을 사용해야 할까? 회귀분석(regression analysis)는 여러 가지 변수들의 관계를 분석하는 통계적 방법이다. 일반적으로 한 개 또는 여러 가지의 설명변수들(explanatory variables)이 관심있는 종속변수(response variable)에 어떤 형태로 영향을 미치는지에 파악하고 설명변수와 종속변수의 관계를 통계적으로 추론하는 것이 회귀분석의 목적이다.

자동차의 속도를 x 라고 하고 제동거리를 y 라고 하면 다음과 같은 선형식으로 자동차의 속도와 제동거리의 관계를 나타내는 것을 단순선형회귀식(simple linear regression equation)이라고 한다.

$$E(y|x) = \beta_0 + \beta_1 x \tag{1}$$

식 (1)은 y의 평균이 x 의 선형식으로 나타나는 관계를 가정한 것이며 절편 β_0 와 기울기 β_1은 모르는 모수로서

자료를 통하여 추정해야 한다.

위에서 본 cars 예제와 같이 n 개의 자료를 독립적으로 추출하였다면 자료의 생성 과정을 다음과 같은 선형 회귀모형(linear regression model)로 나타낸다. 종속변수 y는 설명변수 x의 선형식으로 나타내어지는 결정적인 요인과 확률 변수로 나타내어지는 임의의 오차항 e의 합으로 나타내어진다.

$$y_i = E(y_i|x_i) + e_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \ldots, n \tag{2}$$

위에서 오차항 e_i 평균이 0이고 분산이 σ^2 인 임의의 확률분포를 따르며 서로 독립이다.

$$E(e_i) = 0, \quad V(e_i) = \sigma^2 \quad i = 1, 2, \ldots, n$$

1.2 최소제곱법

단순회귀식 (2)에서 모수 β_0와 β_1를 회귀계수(regression coefficient) 라고 하며 자료(observation; data)를 수집하여 추정해야 한다. n개의 자료를 이용하여 회귀계수 β_0와 β_1를 추정하려고 할 때 가장 쉽고 오래됐으며 또한 가장 유용한 방법인 최소제곱법(least square method)을 사용할 수 있다. 일단 위의 식 (2)에서 종속변수의 관측값 y_i을 대응하는 설명변수 $x = x_i$를 이용하여 예측한 값은 $\beta_0 + \beta_1 x_i$이다. 여기서 실제 관측하여 얻어진 값 y_i와 예측값 $\beta_0 + \beta_1 x_i$사이에는 차이가 발생한다. 그 차이를 잔차(residual)라고 하며 표현하면 다음과 같다.

$$r_i = y_i - E(y_i|x_i) = y_i - (\beta_0 + \beta_1 x_i)$$

잔차는 위에 식에서 알 수 있듯이 관측값과 회귀식을 통한 예측값의 차이를 나타낸 것이다. 그러면 자료를 가장 잘 설명할 수 있는 회귀직선을 얻기 위해서는 잔차 r_i를 가장 작게하는 회귀모형을 세워야 한다. 잔차들을 최소로 하는 방법들 중 하나인 최소제곱법은 잔차들의 제곱합을 최소로 하는 회귀계수를 추정하는 방법이다. 잔차들의 제곱합은 다음과 같이 표현된다.

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \tag{3}$$

위의 잔차 제곱합 $S(\beta_0, \beta_1)$ 을 최소화하는 β_0와 β_1의 값을 구하는 방법은 잔차 제곱합이 β_0와 β_1의 미분 가능한 2차 함수이고 아래로 볼록한 함수(convex function)임을 이용한다. 각각의 회귀계수에 대해서 편미분을 하고 0으로 놓으면 아래와 같이 정리된다.

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{n} (-2)[y_i - (\beta_0 + \beta_1 x_i)] = 0 \tag{4}$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{n} (-2x_i)[y_i - (\beta_0 + \beta_1 x_i)] = 0 \tag{5}$$

위의 연립방정식을 행렬식으로 표시하면 다음과 같이 나타낼 수 있다.

$$\begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

위의 방정식을 풀어서 구한 회귀계수의 추정치를 $\hat{\beta}_0, \hat{\beta}_1$ 이라고 하면 다음과 같이 주어진다.

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}\end{aligned}$$

2 회귀식의 행렬형식

일반적으로 회귀모형에서 종속변수의 수는 하나인 경우가 많지만 설명변수의 수는 여러 개인 경우가 많다. 이런 경우 중선형회귀식(multiple linear regression)은 다음과 같이 표현할 수 있고, p개의 설명변수가 있다고 가정하고 $(x_1, x_2, \cdots, x_p)$ 표본의 크기 n인 자료가 얻어지면 선형회귀식을 행렬로 다음과 같이 표현할 수 있다.

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i \qquad (6) \\ &= \boldsymbol{x}_i^t \boldsymbol{\beta} + e_i \qquad (7) \end{aligned}$$

위의 식을 다시 표현하면 다음과 같이 쓸 수 있다.

$$y_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + e_i$$

n개의 관측치가 있을때 n개의 회귀식을 행렬식으로 표현하면 다음과 같다.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

위의 식을 벡터를 이용하여 표시하면 다음과 같다.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad (8)$$

위의 행렬식에서 각 벡터와 행렬의 차원은 다음과 같다.

- $\boldsymbol{y}$: $n \times 1$
- $\boldsymbol{X}$: $n \times (p+1)$
- $\boldsymbol{\beta}$: $(p+1) \times 1$
- $\boldsymbol{e}$: $n \times 1$

여기서 회귀분석의 오차항은 서로 독립이고 동일한 분산을 갖는다. 즉, 오차항은 다음의 분포를 따른다. 즉, $E(\boldsymbol{e}) = \mathbf{0}$ 이므로 관측값 벡터 $\boldsymbol{y}$의 평균을 보면

$$E(\boldsymbol{y}|\boldsymbol{X}) = E(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}) = \boldsymbol{X}\boldsymbol{\beta} + E(\boldsymbol{e}) = \boldsymbol{X}\boldsymbol{\beta} \tag{9}$$

3 벡터미분

3.1 스칼라미분

벡터미분(Vector diffrential) 또는 행렬미분(Matrix differential)은 벡터와 행렬의 미분식에 대한 표기법을 정의하는 방법이다. 보통 스칼라(scalar)에 대한 미분은 일분수 함수 $f : \Re^1 \to \Re^1$또는 다변수 함수(function of several variables) $f : \Re^p \to \Re^1$에서 쉽게 정의된다. 만약 $y = f(x)$ 또는 $y = f(\boldsymbol{x})$라고 하면 다음과 같이 미분이 주어진다.

$$\frac{\partial y}{\partial x} = \frac{\partial f(x)}{\partial x} = f'(x)$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_p} \right) = \nabla f(x)$$

함수가 다변수함수일 경우 함수의 값을 각 축의 변수로 미분한 것(partial derivative)을 벡터로 표시하는 것을 gradient 라고 한다.

3.2 벡터미분의 표기 방법

이제 다변량함수(multivariate function), $f : \Re^p \to \Re^q$에 대한 미분을 생각해보자. 앞 절에서 본것과 같이 스칼라 함수를 여러 변수로 미분하여 partial derivative를 구한 뒤 gradient를 만드는 경우 열벡터와 행벡터 중 하나를 선택해야 한다. 이러한 선택은 절대적인 것이 아니며 각 분야의 특성과 편의에 따라 다르게 선택 될 수 있다.

이제 간단한 예제를 고려해 보자. 두 열벡터 $\boldsymbol{x} = (x_1, x_2)^t \in \Re_2$, $\boldsymbol{y} = (y_1, y_2, y_3)^t \in \Re^3$를 고려하고 다음과 같은 함수로 두 벡터의 관계가 정의된다고 하자.

$$y_1 = x_1^2 + x_2, \quad y_2 = \exp(x_1) + 3x_2, \quad y_3 = \sin(x_1) + x_2^3$$

일단 각각의 partial derivative $\partial y_i / \partial x_j$를 구해야 하며 이는 scalar 미분으로 쉽게 구해진다.

$$\begin{array}{lll}
\dfrac{\partial y_1}{\partial x_1} = 2x_1, & \dfrac{\partial y_2}{\partial x_1} = \exp(x_1), & \dfrac{\partial y_3}{\partial x_1} = \cos(x_1) \\
\dfrac{\partial y_1}{\partial x_2} = 1, & \dfrac{\partial y_2}{\partial x_1} = 3, & \dfrac{\partial y_3}{\partial x_1} = 3x_2^2
\end{array}$$

통계학에서는 벡터 $\boldsymbol{y}$를 벡터 $\boldsymbol{x}$로 미분하려면 다음과 같이 분모 표기법 (Denominator layout)을 사용하여 표기한다.

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \equiv \frac{\partial \boldsymbol{y}^t}{\partial \boldsymbol{x}} \underset{def}{\equiv} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \frac{\partial y_3}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_3}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 & \exp(x_1) & \cos(x_1) \\ 1 & 3 & 3x_2^2 \end{bmatrix}$$

즉 분모표기법은 분모를 열벡터로, 분자를 행벡터로 보고 각각 위치에 있는 변수들에 대하여 미분을 표기하는 방법이다.

3.3 핵심공식

디음은 분모표기법을 이용한 가장 기본적이고 핵심적인 미분 공식들이다. 공식을 유도하는 경우 분모표기법에서는 $\partial \boldsymbol{y}/\partial \boldsymbol{x} \equiv \partial \boldsymbol{y}^t/\partial \boldsymbol{x}$ 임을 이용한다. 변환이 있거나 여러가지 곱이 있는 경우 미분할 대상 벡터를 가장 왼쪽에 전치형태(즉, 행벡터의 형태로)로 놓는 것이 필요하다. 예를 들어

$$\frac{\partial \boldsymbol{a}^t \boldsymbol{V} \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})^t \boldsymbol{V}^t \boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})^t}{\partial \boldsymbol{x}} \boldsymbol{V}^t \boldsymbol{a} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{V}^t \boldsymbol{a}$$

또한 행렬은 교환법칙이 성립하지 않기 때문에 연산의 순서를 유지해야 하는 것을 유념하자.

1. 기본행렬 미분 벡터 $\boldsymbol{c}$를 상수벡터하고 하자.

$$\frac{\partial \boldsymbol{c}}{\partial \boldsymbol{x}} = \boldsymbol{0}, \quad \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{I}$$

2. 벡터-스칼라 미분

   이 경우는 $\boldsymbol{x} \in \Re^1$, $\boldsymbol{y} \in \Re^q$ 인 경우이며 결과는 다음과 같이 행벡터로 결과가 주어진다.

$$\frac{\partial \boldsymbol{y}}{\partial x} \underset{def}{\equiv} \frac{\partial \boldsymbol{y}^t}{\partial x} = \left[\frac{\partial y_1}{\partial x}, \frac{\partial y_2}{\partial x}, \cdots, \frac{\partial y_q}{\partial x}\right]$$

3. 스칼라-벡터 미분

   이 경우는 $\boldsymbol{x} \in \Re^p$, $\boldsymbol{y} \in \Re^1$ 인 경우이며 결과는 다음과 같이 열벡터로 결과가 주어진다.

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_p} \end{bmatrix}$$

4. 상수벡터와 내적에 대한 미분

   열벡터 $\boldsymbol{a}$를 $p \times 1$ 상수벡터이라고 하고 $y = \boldsymbol{a}^t \boldsymbol{x} = \boldsymbol{x}^t \boldsymbol{a}$라 하자.

$$\frac{\partial y}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{a}^t \boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^t \boldsymbol{a}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial \boldsymbol{a}^t \boldsymbol{x}}{\partial x_1} \\ \frac{\partial \boldsymbol{a}^t \boldsymbol{x}}{\partial x_2} \\ \vdots \\ \frac{\partial \boldsymbol{a}^t \boldsymbol{x}}{\partial x_p} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \boldsymbol{a}$$

5. 선형변환에 대한 미분

   행렬 $\boldsymbol{A}$를 $q \times p$ 행렬이라고 하고 $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$라 하자. 여기서 행렬 $\boldsymbol{A}$를 다음과 같이 나타내자.

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^t \\ \boldsymbol{a}_2^t \\ \vdots \\ \boldsymbol{a}_q^t \end{bmatrix} \text{ or } \boldsymbol{A}^t = [\boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \cdots \ \boldsymbol{a}_q]$$

   위의 내적에 대한 미분 결과를 이용하면 다음은 결과를 얻는다.

$$\begin{aligned} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} \\ &\equiv \frac{\partial \boldsymbol{x}^t \boldsymbol{A}^t}{\partial \boldsymbol{x}} \\ &= \frac{\partial}{\partial \boldsymbol{x}}[\boldsymbol{x}^t \boldsymbol{a}_1 \ \boldsymbol{x}^t \boldsymbol{a}_2 \ \cdots \ \boldsymbol{x}^t \boldsymbol{a}_q] \\ &= [\frac{\partial \boldsymbol{x}^t \boldsymbol{a}_1}{\partial \boldsymbol{x}} \ \frac{\partial \boldsymbol{x}^t \boldsymbol{a}_2}{\partial \boldsymbol{x}} \ \cdots \ \frac{\partial \boldsymbol{x}^t \boldsymbol{a}_q}{\partial \boldsymbol{x}}] \\ &= [\boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \cdots \ \boldsymbol{a}_q] \\ &= \boldsymbol{A}^t \end{aligned}$$

   위의 결과를 응용하면 다음의 결과를 얻는다.

$$\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^t \quad \text{and} \quad \frac{\partial \boldsymbol{x}^t \boldsymbol{A}}{\partial \boldsymbol{x}} = \boldsymbol{A}$$

6. 이차형식

$$\frac{\partial \boldsymbol{x}^t \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^t}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{x} + \frac{\partial \boldsymbol{x}^t \boldsymbol{A}^t}{\partial \boldsymbol{x}}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^t \boldsymbol{x}$$

   만약 행렬 $\boldsymbol{A}$가 대칭이면

$$\frac{\partial \boldsymbol{x}^t \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$$

3.4 합성함수에 대한 미분공식

두 개의 다변량 함수를 고려하고

$$g : \Re^p \to \Re^q, \quad f : \Re^q \to \Re^r$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{f}^t(\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}} \\
&= \left[\frac{\partial f_1(\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}}, \frac{\partial f_2(\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}}, \cdots, \frac{\partial f_r(\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}}\right] \\
&= \begin{bmatrix} \frac{\partial f_1(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_1} & \frac{\partial f_2(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_1} & \cdots & \frac{\partial f_r(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_1} \\ \frac{\partial f_1(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_2} & \frac{\partial f_2(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_2} & \cdots & \frac{\partial f_r(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_2} \\ & & \vdots & \\ \frac{\partial f_1(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_p} & \frac{\partial f_2(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_p} & \cdots & \frac{\partial f_r(\boldsymbol{g}(\boldsymbol{x}))}{\partial x_p} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=1}^q \frac{\partial f_1}{\partial g_k}\frac{\partial g_k}{\partial x_1} & \sum_{k=1}^q \frac{\partial f_2}{\partial g_k}\frac{\partial g_k}{\partial x_1} & \cdots & \sum_{k=1}^q \frac{\partial f_r}{\partial g_k}\frac{\partial g_k}{\partial x_1} \\ \sum_{k=1}^q \frac{\partial f_1}{\partial g_k}\frac{\partial g_k}{\partial x_2} & \sum_{k=1}^q \frac{\partial f_2}{\partial g_k}\frac{\partial g_k}{\partial x_2} & \cdots & \sum_{k=1}^q \frac{\partial f_r}{\partial g_k}\frac{\partial g_k}{\partial x_2} \\ & & \vdots & \\ \sum_{k=1}^q \frac{\partial f_1}{\partial g_k}\frac{\partial g_k}{\partial x_p} & \sum_{k=1}^q \frac{\partial f_2}{\partial g_k}\frac{\partial g_k}{\partial x_p} & \cdots & \sum_{k=1}^q \frac{\partial f_r}{\partial g_k}\frac{\partial g_k}{\partial x_p} \end{bmatrix} \\
&= \sum_{k=1}^q \begin{bmatrix} \frac{\partial f_1}{\partial g_k}\frac{\partial g_k}{\partial x_1} & \frac{\partial f_2}{\partial g_k}\frac{\partial g_k}{\partial x_1} & \cdots & \frac{\partial f_r}{\partial g_k}\frac{\partial g_k}{\partial x_1} \\ \frac{\partial f_1}{\partial g_k}\frac{\partial g_k}{\partial x_2} & \frac{\partial f_2}{\partial g_k}\frac{\partial g_k}{\partial x_2} & \cdots & \frac{\partial f_r}{\partial g_k}\frac{\partial g_k}{\partial x_2} \\ & & \vdots & \\ \frac{\partial f_1}{\partial g_k}\frac{\partial g_k}{\partial x_p} & \frac{\partial f_2}{\partial g_k}\frac{\partial g_k}{\partial x_p} & \cdots & \frac{\partial f_r}{\partial g_k}\frac{\partial g_k}{\partial x_p} \end{bmatrix} \\
&= \sum_{k=1}^q \begin{bmatrix} \frac{\partial g_k}{\partial x_1} \\ \frac{\partial g_k}{\partial x_2} \\ \vdots \\ \frac{\partial g_k}{\partial x_p} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial g_k} & \frac{\partial f_2}{\partial g_k} & \cdots & \frac{\partial f_r}{\partial g_k} \end{bmatrix} \\
&= \sum_{k=1}^q \frac{\partial g_k}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{f}}{\partial g_k} \\
&= \left[\frac{\partial g_1}{\partial \boldsymbol{x}} \frac{\partial g_2}{\partial \boldsymbol{x}} \cdots \frac{\partial g_q}{\partial \boldsymbol{x}}\right] \begin{bmatrix} \partial \boldsymbol{f}/\partial g_1 \\ \partial \boldsymbol{f}/\partial g_2 \\ \vdots \\ \partial \boldsymbol{f}/\partial g_q \end{bmatrix} \\
&= \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{g}} \quad (p \times q)(q \times r)
\end{aligned}
$$

특별히 f가 일변량인 경우($r = 1$),

$$\frac{\partial f(\boldsymbol{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \frac{\partial f}{\partial \boldsymbol{g}} \quad (p \times q)(q \times 1)$$

더 나아가 다음도 보일 수 있다.

$$\frac{\partial \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{u}(\boldsymbol{x})))}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{g}}$$

4 최소제곱추정

최소제곱추정법(least square estimation)은 자료의 관계을 잘 반영하는 회귀식을 구한 다음 실제 관측값 y_i과 예측값 $\boldsymbol{x}_i^t\boldsymbol{\beta}$ 간에 차이인 잔차를 가장 작게 만드는 것이 목적이다. 모든 잔차항의 제곱의 합을 최소화하는 방법을 최소제곱법이라고 하며 이를 이용하여 회귀계수의 추정량을 찾는다.

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^t\boldsymbol{\beta})^2 = \min_{\beta} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{10}$$

4.1 방법 1

$\hat{\boldsymbol{\beta}}$는 잔차의 제곱합 (10) 을 최소로 하는 최소제곱 추정량이다. 잔차의 제곱합을 $S(\boldsymbol{\beta})$이라고 하면

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}^t\boldsymbol{y} - \boldsymbol{y}^t\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{y} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{y}^t\boldsymbol{y} - 2\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{y} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} \end{aligned} \tag{11}$$

여기서 $S(\boldsymbol{\beta})$를 최소로 하는 회귀계수벡터의 값을 구하기 위하여 $S(\boldsymbol{\beta})$를 회귀계수벡터 $\boldsymbol{\beta}$로 미분한후 $\mathbf{0}$ 으로 놓고 선형 방정식을 풀어야 한다.

앞 절에 나오는 벡터미분을 이용하면

$$\begin{aligned} \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{y}^t\boldsymbol{y} - 2\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{y} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}) \\ &= \mathbf{0} - 2\boldsymbol{X}^t\boldsymbol{y} + 2\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} \\ &= \mathbf{0} \end{aligned}$$

최소제곱 추정량을 구하기 위한 정규방정식은 다음과 같이 쓸 수 있다.

$$\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^t\boldsymbol{y} \tag{12}$$

방정식 (12)를 정규방정식(normal equation)이라고 한다. 만약 $\boldsymbol{X}^t\boldsymbol{X}$가 정칙행렬일 경우 최소제곱법에 의한 회귀계수 추정량 $\hat{\boldsymbol{\beta}}$ 다음과 같다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y} \tag{13}$$

예측값 벡터 $\hat{\boldsymbol{y}}$ 는 $E(\boldsymbol{y}|\boldsymbol{X})$의 추정치로서 다음과 같다.

$$\hat{E}(\boldsymbol{y}|\boldsymbol{X}) = \hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t y$$

만약 $\boldsymbol{X}^t\boldsymbol{X}$가 정칙행렬이 아닐 경우 최소제곱법에 의한 회귀계수 추정량 $\hat{\boldsymbol{\beta}}$은 $\boldsymbol{X}^t\boldsymbol{X}$의 일반화 역행렬 $(\boldsymbol{X}^t\boldsymbol{X})^-$를 이용하여 다음과 같이 구한다. 이 경우 일반화 역행렬이 유일하지 않기 때문에 회귀계수 추정량도 유일하지 않다.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-}\boldsymbol{X}^t\boldsymbol{y}$$

4.2 방법 2

식 (10)에서 나오는 오차벡터를 정의하고 $\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ 오차벡터를 모수벡터 $\boldsymbol{\beta}$로 미분하면 다음과 같은 결과를 얻는다.

$$\frac{\partial \boldsymbol{e}}{\partial \boldsymbol{\beta}} = \frac{\partial (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{\partial \boldsymbol{X}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} \equiv -\frac{\partial \boldsymbol{\beta}^t\boldsymbol{X}^t}{\partial \boldsymbol{\beta}} = -\boldsymbol{X}^t$$

이제 오차제곱합 $S(\boldsymbol{\beta}) = \boldsymbol{e}^t\boldsymbol{e}$ 를 모수벡터로 미분하면 이차형식의 미분공식과 합성함수 미분공식을 차례로 적용하면 된다.

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{e}^t\boldsymbol{e}}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{e}}{\partial \boldsymbol{\beta}}\frac{\partial \boldsymbol{e}^t\boldsymbol{e}}{\partial \boldsymbol{e}} = -\boldsymbol{X}^t\,(2\boldsymbol{e}) = -2\boldsymbol{X}^t(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

위의 방정식을 $\mathbf{0}$으로 놓으면 최소제곱 추정량 (열)벡터를 구한다.

$$\boldsymbol{X}^t\boldsymbol{y} - \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta} = \mathbf{0} \quad \rightarrow \quad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$