

HW3

1. 어떤 큰 공장에서 동일한 기계들의 정비기록에 관한 표본자료를 취하였다. 이는 기계의 사용연도(age of machines)와 정비비용(maintenance cost) 간에 어떤 관계가 있는가를 밝혀내기 위한 것이다. 그 자료는 다음과 같다 (표본의 크기 $n = 14$).

사용연도 X (단위:년)	정비비용 Y (단위 : 1,000원)	사용연도 X (단위:년)	정비비용 Y (단위 : 1,000원)
3	39	6	90
1	24	9	140
5	115	3	112
8	105	5	70
1	50	7	186
4	86	2	43
2	67	6	126

- (1) $X^T X$, $X^T \mathbf{y}$, $\mathbf{y}^T \mathbf{y}$ 와 $(X^T X)^{-1}$ 을 구하시오.
- (2) $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$ 를 구하고, 적합된 회귀선형을 써 보아라.
- (3) σ^2 을 MSE 로 추정할 경우, $\hat{\beta}$ 의 분산-공분산행렬의 추정은

$$\widehat{\text{Var}}(\hat{\beta}) = (X^T X)^{-1}(MSE)$$

이다. 먼저 분산분석하여 MSE 를 구하고 $\widehat{\text{Var}}(\hat{\beta})$ 을 구하시오.

2. 어떤 공정에서 나오는 제품의 강도(kg/cm²)가 그 공정의 온도와 압력에 어떠한 영향을 받는가를 조사하기 위하여 다음의 데이터를 얻었다.

공정온도 x_1 (단위 : °C)	공정압력 x_2 (단위 : psi)	강도 y (단위 : kg/cm ²)
195	57	81.4
179	61	122.2
205	60	101.7
204	62	175.6
201	61	150.3
184	54	64.8
210	58	92.1
209	61	113.8

$n=8$
 $p=2$

- (1) 선형회귀모형, $y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j$ 가 성립된다고 가정하고 데이터로부터 회귀모형을 추정하시오.
- (2) 오차분산이 $\sigma^2 = 3$ 이라 하면, $\text{Var}(\hat{\beta}_0)$, $\text{Var}(\hat{\beta}_1)$, $\text{Var}(\hat{\beta}_2)$ 와 $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$ 는 무엇인가?

$$\widehat{\mathbf{y}} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

$$\text{Var}(\beta) = \frac{1}{n} (X^T X)^{-1} \sigma^2 = \left(\right)$$

1. 어떤 큰 공장에서 동일한 기계들의 정비기록에 관한 표본자료를 취하였다. 이는 기계의 사용연도(age of machines)와 정비비용(maintenance cost) 간에 어떤 관계가 있는가를 밝혀내기 위한 것이다. 그 자료는 다음과 같다 (표본의 크기 $n = 14$).

사용연도 X (단위:년)	정비비용 Y (단위 : 1,000원)	사용연도 X (단위:년)	정비비용 Y (단위 : 1,000원)
3	39	6	90
1	24	9	140
5	115	3	112
8	105	5	70
1	50	7	186
4	86	2	43
2	67	6	126

$n=14$
 $p=1$

(1) $X^T X$, $X^T y$, $y^T y$ 와 $(X^T X)^{-1}$ 을 구하시오.

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad X = \begin{pmatrix} 1 & x_1 \end{pmatrix}_{14 \times 2}$$

$$X = \begin{pmatrix} 1 & 3 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 2 \\ 1 & 6 \end{pmatrix}_{14 \times 2}$$

$$\underline{y} = \begin{pmatrix} 39 \\ 24 \\ \vdots \\ 43 \\ 126 \end{pmatrix}$$

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \underline{3} & 1 & \cdots & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ \vdots & \vdots \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} 14 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} = \begin{pmatrix} 14 & 62 \\ 62 & 360 \end{pmatrix}$$

$$(X^T X)^{-1} = \frac{1}{14 \times 360 - 62 \times 62} \begin{pmatrix} 360 & -62 \\ -62 & 14 \end{pmatrix} = \begin{pmatrix} 0.301 & -0.052 \\ -0.052 & 0.012 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \underline{3} & 1 & \cdots & 2 & 6 \end{pmatrix} \begin{pmatrix} 39 \\ 24 \\ \vdots \\ 43 \\ 126 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} = \begin{pmatrix} 1253 \\ 6714 \end{pmatrix}$$

$$y^T y = \sum y_i^2 = 138197.$$

(2) $\hat{\beta} = (X^T X)^{-1} X^T y$ 를 구하고, 적합한 회귀선형을 써 보아라.

$$\hat{\beta} = \begin{pmatrix} 29.107 \\ 13.637 \end{pmatrix} \quad \downarrow \quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 29.107 + 13.637 x.$$

(3) σ^2 을 MSE 로 추정할 경우, $\hat{\beta}$ 의 분산-공분산행렬의 추정은

$$\widehat{\text{Var}}(\hat{\beta}) = (X^T X)^{-1} (MSE)$$

$$\text{Var}(\hat{\beta}) = (X^T X)^{-1} \sigma^2$$

$$\widehat{\text{Var}}(\hat{\beta}) = (X^T X)^{-1} MSE.$$

이다. 먼저 분산분석하여 MSE 를 구하고 $\widehat{\text{Var}}(\hat{\beta})$ 을 구시오.

$$\begin{cases} SST = y^T y - n(\bar{y})^2 \\ SSR = SST - SSE \\ SSE = y^T (I - H) y, \quad H = X(X^T X)^{-1} X^T \quad I_n \end{cases}$$

* 분산분석표

제곱합		자유도	평균제곱합	F	기각역
회귀 (SSR)	15887.25	1	$MSR = 15887.25$	$\frac{MSR}{MSE} = 18.75$	$F_{0.05}(1, 12) = 4.74$
잔차 (SSE)	10166.25	12	$MSE = 847.19$		
총 (SST)	26053.5	13			

$$\widehat{\text{Var}}(\hat{\beta}) = (X^T X)^{-1} \cdot MSE = \begin{pmatrix} 255.01 & -43.92 \\ -43.92 & 9.92 \end{pmatrix}$$

$$2. (2) \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad \underline{\text{Var}(\underset{\sim}{\hat{\beta}})} = \begin{pmatrix} \underbrace{\text{Var}(\hat{\beta}_0)} & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ & \underbrace{\text{Var}(\hat{\beta}_1)} & \underbrace{\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)} \\ & & \underbrace{\text{Var}(\hat{\beta}_2)} \end{pmatrix}$$

$$(X^T X)^{-1} X^T y \quad y^T y$$

$$\Rightarrow \hat{\beta}$$

$$\hat{\mu}_0 = E(y | x_1 = 200, x_2 = 59)$$

$$= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = \hat{\beta}^T x_0$$

$$x_0 = \begin{pmatrix} 1 \\ 200 \\ 59 \end{pmatrix}$$

(3) $x_1 = 200^\circ\text{C}$ 이고 $x_2 = 59\text{psi}$ 에서 평균 제품의 강도의 추정값 \hat{y} 는 얼마인가? 이의 분산을 $\sigma^2 = 3$ 이라 가정하고 구하시오.

(4) 추정된 회귀계수 $\hat{\beta}_1, \hat{\beta}_2$ 의 의미는 무엇인가?

(5) 분산분석표를 작성하고 $\alpha = 0.05$ 로 F -검정을 행하시오.

(6) 결정계수 R^2 을 구하시오.

(7) 수정된 결정계수 R_{adj}^2 을 구하시오.

(8) σ^2 의 추정값 MSE 를 구하시오.

(9) MSR 의 기대값을 σ^2 과 $\beta_0, \beta_1, \beta_2$ 의 함수로 표시하여라.

3. 어떤 공장에서 물의 소비량을 조사하기 위하여 매달의 물소비량(y), 평균온도(x_1), 작업일수(x_2)와 작업량(x_3)에 관한 데이터를 얻었다.

물소비량 y (단위 : 1,000 톤)	평균온도 x_1 (단위 : $^\circ\text{C}$)	작업일수 x_2 (단위 : 일)	작업량 x_3 (단위 : 1,000 톤)
2.8	10	27	64
3.9	24	26	72
3.9	25	28	80
4.4	28	26	88
3.1	15	30	81
3.1	18	24	45
3.5	22	27	46
3.6	22	25	69
3.0	12	27	54
3.3	15	25	39

(1) 데이터로부터 회귀모형, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$ 을 구하여라. 그리고 어쩌서 이 모형이 선택되었는가에 대하여 토의하시오.

(2) $\hat{\beta}_1, \hat{\beta}_2$ 와 $\hat{\beta}_3$ 의 의미는 무엇인가?

(3) $\text{Var}(\hat{\beta}_3)$ 을 구하시오. σ^2 을 MSE 로 추정하면 $\widehat{\text{Var}}(\hat{\beta}_3)$ 은 무엇인가?

(4) 분산분석표를 작성하고, 결정계수 R^2 을 구하시오.

(5) $x_1 = 20, x_2 = 27, x_3 = 60$ 에서 평균 물소비량을 추정하시오. 이 추정된 소비량의 표준편차의 추정값을 구하시오.

(6) $x_1 = 20, x_2 = 27, x_3 = 60$ 에서 어느 한 달의 물소비량 y_s 를 예측하여라. 이 예측된 물소비량의 표준편차의 추정값을 구하시오.

(7) $y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j$ 라 가정하고 회귀제곱합 SSR 의 기대값을 σ^2 과 $\beta_i, i = 0, 1, 2, 3$ 의 함수로 표시하여라.

(8) β_3 의 95% 신뢰구간을 구하시오. 이 신뢰구간의 의미를 해석하시오.

(9) β_1 의 99% 신뢰구간을 구하시오. 이 신뢰구간의 의미를 해석하시오.

(10) $x_1 = 20, x_2 = 27, x_3 = 60$ 에서 평균 물소비량, $E(y)$ 의 95% 신뢰구간을 구하시오.

(11) 가설 $H_0 : \beta_1 = 0, H_1 : \beta_1 > 0$ 을 $\alpha = 0.05$ 로 검정하시오.

(12) 가설 $H_0 : \beta_1 = \beta_2 = \beta_3$ 을 $\alpha = 0.05$ 로 검정하시오.

(13) 가설 $H_0 : \beta_1 = \beta_2 + 3$ 을 $\alpha = 0.05$ 로 검정하시오.

(14) $x_1 = 20, x_2 = 27, x_3 = 60$ 에서 $E(y)$ 에 관한 가설 $H_0 : E(y) = 3.5$, $H_1 : E(y) \neq 3.5$ 를 $\alpha = 0.05$ 로 검정하시오.

4. 중회귀모형, $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}_n, \mathbf{I}\sigma^2)$ 에서 X 가 $n \times (p+1)$ 행렬이고 rank가 $p+1$ 이라면 적합된 모형 $\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$ 에 대하여

$$\sum_{j=1}^n \text{Var}(\hat{y}_j) = (p+1)\sigma^2$$

이 됨을 증명하여라.

5. 중회귀모형에서 다음을 증명하시오.

(1) $\text{Cov}(\mathbf{e}, \mathbf{y}) = \sigma^2[I_n - X(X^\top X)^{-1}X^\top]$ (2) $\text{Cov}(\mathbf{e}, \hat{\mathbf{y}}) = \mathbf{0}_n$

(3) $\text{Cov}(\mathbf{e}, \hat{\boldsymbol{\beta}}) = \mathbf{0}_{n \times (k+1)}$ (4) $\text{Cov}(\boldsymbol{\epsilon}, \hat{\boldsymbol{\beta}}) = \sigma^2 X(X^\top X)^{-1}$

(5) $\sum_{j=1}^n e_j y_j = SSE$ (6) $\sum_{j=1}^n e_j \hat{y}_j = 0$