4장 Review

- 하나의 반응변수에 대해 여러 개의 설명변수를 $X_1, X_2, \cdots, X_p$

자료의 순서	Y	X_1	X_2	변수값	X_j		X_p
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1p}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
i	y_i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nj}	$\cdots$	x_{np}

x_{ij} : j번째 설명변수에서 i번째 자료의 값

■ 다중선형회귀모형

단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$을 확장

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \cdots, n \qquad (4.7)$$

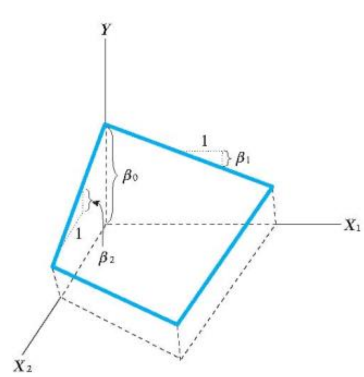
1

4장 Review

설명변수가 두 개($p=2$)이면 $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$

$p=2$인 경우의 선형회귀반응표면

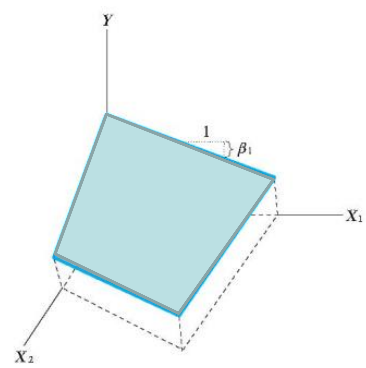
그림 4.1



2

4장 Review

설명변수가 두 개($p=2$)이면 $y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\epsilon_i$



3

4장 Review

- X_2로 인한 추가설명력은 다음과 같은 과정을 통하여 할 수 있음

(1) Y의 변화량 중에서 X_1으로 설명되지 않는 부분: $e_1(Y|X_1)$

- Y의 X_1에 관한 회귀모형을 구한 후 나오는 잔차

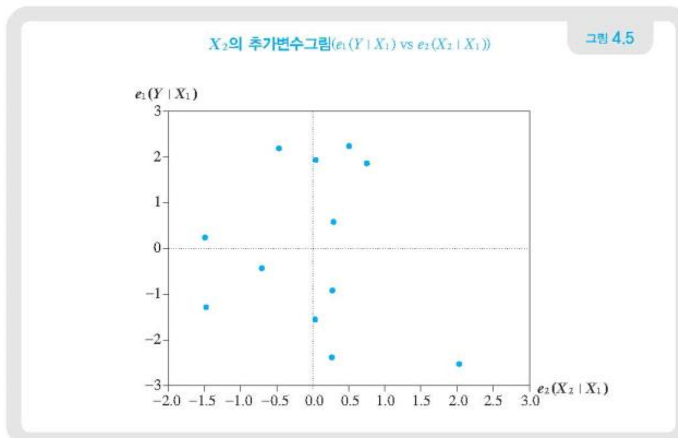
(2) X_2와 X_1에서 서로 중복되지 않는 부분: $e_2(X_2|X_1)$

- X_2의 X_1에 관한 회귀모형을 구한 후 나오는 잔차

(3) (1)에서 구한 잔차(세로축)와 (2)에서 구한 잔차(가로축)의 산점도
두 잔차에 상관성 / 기울기가 유의성으로 X_2의 추가적인 설명력 판단

4

4장 Review



5

4장 Review

■ 추가변수그림

- $e_1(Y|X_1)$(세로축)와 $e_2(X_2|X_1)$(가로축) 산점도([그림 4.5])

  X_1에 대한 X_2의 추가변수그림 / 편회귀그림

  (added variable plot / partial regression plot)

  - 회귀계수 β_2에 대한 t-검정을 대체
  - 이상점이나 모형의 비선형성까지 시각적으로 확인

  - $e_2(X_2|X_1)$에 대한 p-값이 0.687 ([표 4.2] 참조)
  - X_2의 추가 여부에 대한 통계적인 유의성은 미미함

    추가변수그림의 선형관계가 시각적으로 불명확

6

4장 Review

■ 편상관계수

추가변수그림을 구성하고 있는 두 잔차에서 얻어지는 표본상관계수 (sample partial correlation coefficient)

두 번째 설명변수 X_2의 표본편상관계수 $r_{X_2|X_1}$의 제곱은 다음과 같음

$$\left(r_{X_2|X_1}\right)^2 = \frac{SSR(X_2|X_1)}{S_{yy} - SSR(X_1)} \qquad (4.9)$$

분모: X_1이 설명을 해주지 못하는 부분

분자: X_1 이후 X_2가 추가됨으로써 생기는 SS

$r_{X_2|X_1}$는 X_1에 대해 조정된 Y와 X_2의 표본편상관계수라 불린다.

7

4장 Review

예) 3개의 설명변수 X_1, X_2, X_3

- X_2의 표본편상관계수 $r_{X_2|X_1X_3}$

$$\left(r_{X_2|X_1X_3}\right)^2 = \frac{SSR(X_2|X_1, X_3)}{S_{yy} - SSR(X_1, X_3)}$$

- (X_1, X_3)에 대한 X_2의 추가변수그림

세로축 $e_1(Y|X_1, X_3)$

가로축 $e_2(X_2|X_1, X_3)$

8

4장 Review

■ 다중회귀모형

반응변수와 p개의 설명변수에 대한 n개의 자료값으로 구성

$$\boldsymbol{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix},\quad \boldsymbol{X}=\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},\quad \boldsymbol{\beta}=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},\quad \boldsymbol{\epsilon}=\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$\boldsymbol{y}$: 반응변수벡터, $n\times 1$ 벡터

$\boldsymbol{X}$: 설명변수행렬, $n\times(p+1)$ 행렬

$\boldsymbol{\beta}$: 회귀계수벡터, $(p+1)\times 1$ 벡터

$\boldsymbol{\epsilon}$: 오차벡터, $n\times 1$ 벡터

9

4장 Review

- 다중회귀모형

$$y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_px_{ip}+\epsilon_i,\quad i=1,\cdots,n \qquad (4.7)$$

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\epsilon} \qquad (4.10)$$

$$y_i=\boldsymbol{x}_i^T\boldsymbol{\beta}+\epsilon_i,\quad i=1,\cdots,n \qquad (4.11)$$

$\boldsymbol{x}_i^T=(1\ x_{i1}\ x_{i2}\ \cdots\ x_{ip})$: $\boldsymbol{X}$의 i번째 행벡터

$$\boldsymbol{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}\quad \boldsymbol{X\beta}=\begin{pmatrix} \beta_0+\beta_1x_{11}+\cdots+\beta_jx_{1j}+\cdots+\beta_px_{1p} \\ \beta_0+\beta_1x_{21}+\cdots+\beta_jx_{2j}+\cdots+\beta_px_{2p} \\ \vdots \\ \beta_0+\beta_1x_{i1}+\cdots+\beta_jx_{ij}+\cdots+\beta_px_{ip} \\ \vdots \\ \beta_0+\beta_1x_{n1}+\cdots+\beta_jx_{nj}+\cdots+\beta_px_{np} \end{pmatrix}\quad \boldsymbol{\epsilon}=\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\boldsymbol{x}_i^T\boldsymbol{\beta}=\begin{pmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_px_{ip}$$

10

4장 Review

가정: 각각의 ϵ_i가 정규분포 $N(0,\sigma^2)$를 따르고 서로 독립이라는 가정

$$\epsilon_i \sim N(0,\sigma^2), \quad Cov(\epsilon_i, \epsilon_j)=0, \ i \neq j$$

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}) \qquad (4.12)$$

$$E(\boldsymbol{\epsilon}) = \begin{pmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0} \qquad (4.13)$$

$$\begin{aligned} Var(\boldsymbol{\epsilon}) &= E[(\boldsymbol{\epsilon}-E(\boldsymbol{\epsilon}))(\boldsymbol{\epsilon}-E(\boldsymbol{\epsilon}))^T] \qquad (4.14) \\ &= \begin{pmatrix} Var(\epsilon_1) & Cov(\epsilon_1,\epsilon_2) & \cdots & Cov(\epsilon_1,\epsilon_n) \\ Cov(\epsilon_2,\epsilon_1) & Var(\epsilon_2) & \cdots & Cov(\epsilon_2,\epsilon_n) \\ \vdots & & \ddots & \vdots \\ Cov(\epsilon_n,\epsilon_1) & & \cdots & Var(\epsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \sigma^2 \end{pmatrix} \\ &= \sigma^2 \boldsymbol{I} \end{aligned}$$

11

4장 Review

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \qquad E(\boldsymbol{y}) = \boldsymbol{X\beta}$$

$$Var(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I} \qquad Var(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}$$

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}) \qquad \boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \cdots, n$$

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i$$

$$\epsilon_i \sim N(0,\sigma^2), \quad Cov(\epsilon_i, \epsilon_j)=0, \ i \neq j$$

$$y_i \sim N(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2), \quad Cov(y_i, y_j)=0, \ i \neq j$$

12

4장 Review

함수 $Q(\beta)$를 최소화하는 β

$$Q(\beta) = (\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})$$

$$\frac{\partial Q}{\partial \beta} = -2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X\beta} = \boldsymbol{0}$$

정규방정식 : $Q(\beta)$를 벡터 β에 대해 편미분한 결과를 0벡터

$$\boldsymbol{X}^T\boldsymbol{X\beta} = \boldsymbol{X}^T\boldsymbol{y}$$

$\boldsymbol{X}^T\boldsymbol{X}$의 역행렬이 존재하면

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

13

4장 Review

1) 적합값벡터(vector of fitted value)

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

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}$$

2) 잔차벡터(residual vector)

$$\begin{aligned} \boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} &= \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} \end{aligned}$$

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_p x_{ip}$$

14

4장 Review

$H = X(X^TX)^{-1}X^T$: 햇행렬

$$H^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T((X^TX)^{-1})^TX^T = X(X^TX)^{-1}X^T$$

$$HH = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T$$

$$(I-H)^T = I-H$$

$$(I-H)(I-H) = I-H-H+HH = I-H$$

$$Hy = X(X^TX)^{-1}X^Ty = X\hat{\beta} = \hat{y}$$

$$(I-H)y = y - \hat{y} = e$$

15

4장 Review

3) 잔차제곱합 SSE

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

$$= y^Ty - \hat{\beta}^TX^Ty$$

$$= y^Ty - \hat{\beta}^TX^TX\hat{\beta}$$

$$SSE = y^T(I-H)y$$

4) σ^2의 추정량 $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{SSE}{n-p'}$$

5) $\hat{\beta}$의 분산-공분산행렬 (대각원소의 제곱근을 취하면 표준오차)

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

16

4장 Review

■ SS의 분포

$$S_{yy}/\sigma^2 \sim \chi^2(n-1) \qquad (4.29)$$

$$SSR/\sigma^2 \sim \chi^2(p)$$

$$SSE/\sigma^2 \sim \chi^2(n-p')$$

[증명] 보론 A.4.7 참조

SSR/σ^2의 분포는 $\beta_1 = ... = \beta_p = 0$인 경우 성립되며 이 때

$$F^* = \frac{\frac{SSR/\sigma^2}{p}}{\frac{SSE/\sigma^2}{n-p'}} = \frac{SSR/p}{SSE/(n-p')} = \frac{MSR}{MSE} \qquad (4.30)$$

의 분포는 $F(p, n-p')$

17

4장 Review

0) 모든 회귀계수에 대한 검정

$H_0 : \boldsymbol{\beta} = \mathbf{0}$ $\qquad H_0 : Y = \beta_0 + \epsilon$

$H_1 : \boldsymbol{\beta} \neq \mathbf{0}$ $\qquad H_1 : Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$

1) 각각의 회귀계수에 대한 검정

$H_0 : \beta_k = 0$ $\qquad H_0 : Y = \beta_0 + \beta_1 X_1 + ... + \beta_{k-1} X_{k-1} + \beta_{k+1} X_{k+1} + ... + \beta_p X_p + \epsilon$

$H_1 : \beta_k \neq 0$ $\qquad H_1 : Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \epsilon$

2) 일부의 회귀계수에 대한 검정

H_0 : k개 β_j =0 $\qquad H_0$: p개의 설명변수 중 k개의 X_j가 제거된 모형

H_1 : not H_0 $\qquad H_1 : Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \epsilon$

(k개의 β_j중에서 일부는 0이 아님)

18

4장 Review

■ 가설검정

0) 모든 회귀계수에 대한 검정

$$H_0: Y = \beta_0 + \epsilon$$

$$H_1: Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

$$\begin{cases} H_0: \boldsymbol{\beta} = \mathbf{0} & (\beta_1 = \dots = \beta_p = 0) \\ H_1: \boldsymbol{\beta} \neq \mathbf{0} \end{cases}$$

- 검정통계량 : $F^* = \dfrac{SSR/p}{SSE/(n-p')} = \dfrac{MSR}{MSE}$

  - 검정통계량 F^*는 귀무가설 H_0에서 $F(p, n-p')$를 따름
  - F^*을 기각값인 $F(\alpha; p, n-p')$과 비교하여 H_0의 기각여부 결정
  - p-값$= P(F(p, n-p') > F^*)$ (p-value, 유의확률)

19

4장 Review

■ 가설검정

1) 각각의 회귀계수에 대한 검정

$$H_0: Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{k-1} X_{k-1} + \beta_{k+1} X_{k+1} + \dots + \beta_p X_p + \epsilon$$

$$H_1: Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon$$

$$\begin{cases} H_0: \beta_k = 0 \\ H_1: \beta_k \neq 0 \end{cases}$$

- 검정통계량 : $t^* = \dfrac{\hat{\beta}_k}{s.e.(\hat{\beta}_k)}$

  - 검정통계량 t^*는 H_0에서 분포 $t(n-p')$를 따름
  - $|t^*|$을 $t(\alpha/2, n-p')$과 비교하여 H_0의 기각여부 결정
  - p-값$= 2P(t(n-2) > |t^*|)$

20

4장 Review

2) 일부의 회귀계수에 대한 검정

H_0 : p개의 설명변수 중 k개의 X_j가 제거된 모형

H_1 : $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon$

H_0 : k개의 β_j가 0

H_1 : not H_0 (k개의 β_j중에서 일부는 0이 아님)

H_0 : (reduced model) $SSR(r)$, $SSE(r)$, $df(r) = n - p' + k$

H_1 : (full model) $SSR(f)$, $SSE(f)$, $df(f) = n - p'$

21

4장 Review

- 검정통계량

$$F^* = \frac{\frac{SSE(r) - SSE(f)}{df(r) - df(f)}}{MSE(f)} = \frac{\frac{SSR(f) - SSR(r)}{k}}{MSE(f)} \qquad (4.34)$$

- 검정통계량 F^*는 H_0에서 분포 $F(k, n-p')$를 따름
- F^*을 $F(\alpha; k, n-p')$과 비교하여 H_0의 기각여부 결정
- p-값$= P(F(k, n-p') > F^*)$
- $k=1$일 때 F^*는 1)에서의 $(t^*)^2$

22

4장 Review

예) 설명변수 X_1, X_2, X_3, X_4

$$H_0 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$
$$H_1 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

$$H_0 : \beta_3 = \beta_4 = 0$$
$H_1 : \beta_3$와 β_4가 모두 0은 아님

$$F^* = \frac{\frac{SSE(X_1X_2) - SSE(X_1X_2X_3X_4)}{(n-3)-(n-5)}}{MSE(X_1X_2X_3X_4)} = \frac{\frac{SSR(X_1X_2X_3X_4) - SSR(X_1X_2)}{2}}{MSE(X_1X_2X_3X_4)}$$

23

4장 Review

$n = 17$

X_1, X_2 (r)	$SSR(X_1X_2) = 20,$	$SSE(X_1X_2) = 80,$	$df = 17 - 3 = 14$
X_1, X_2, X_3, X_4 (f)	$SSR(X_1X_2X_3X_4) = 40,$	$SSE(X_1X_2X_3X_4) = 60,$	$df = 17 - 5 = 12$

$$F^* = \frac{(40-20)/2}{60/(17-5)} = \frac{(50-30)/2}{60/(17-5)} = \frac{10}{5} = 2$$

$$SSR(X_1X_2) = 20 \qquad SSR(X_1X_2X_3X_4) = 40$$
$$SSE(X_1X_2) = 80 \qquad SSE(X_1X_2X_3X_4) = 60$$
$$SST(X_1X_2) = 100 \qquad SST(X_1X_2X_3X_4) = 100$$

24

4장 Review

■ 결정계수

$$R^2 = \frac{SSr}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}$$

– 설명변수가 추가되면 R^2은 증가

■ 수정결정계수

$$\overline{R}^2 = 1 - \frac{SSE/(n-p-1)}{S_{yy}/(n-1)} = 1 - \frac{(n-1)}{(n-p-1)}(1-R^2)$$

– 설명변수가 추가되다라도 $\overline{R}^2$은 계속 증가하지 않음

25

4장 Review

■ 기대값과 새로운 관측값

- **단순회귀**

설명변수의 값이 $X = x_0$로 주어졌을 때의

반응변수의 기대값: $E(Y|X = x_0)$

반응변수의 새로운 관측(예측)값: $y_0 = (y|X = x_0)$

- **다중회귀**

설명변수의 값이 $\boldsymbol{x}_0 = (1, x_{01}, x_{02}, \dots, x_{0p})^T$로 주어졌을 때의

$$(X_1 = x_{01}, X_2 = x_{02}, \dots, X_p = x_{0p})$$

반응변수의 기대값: $E(Y|\boldsymbol{x} = \boldsymbol{x}_0)$

반응변수의 새로운 관측(예측)값: $y_0 = (y|\boldsymbol{x} = \boldsymbol{x}_0)$

26

4장 Review

■ 기대값과 새로운 관측값

1) 설명변수의 값이 $\boldsymbol{x}_0 = (1, x_{01}, x_{02}, \dots, x_{0p})^T$일때 반응변수의 기대값: $E(Y|\boldsymbol{x}=\boldsymbol{x}_0)$

- 점추정량:

$$\hat{E}(Y|\boldsymbol{x}=\boldsymbol{x}_0) = \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} = \hat{y}_0$$

- 분산:

$$Var\left[\hat{E}(Y|\boldsymbol{x}=\boldsymbol{x}_0)\right] = \sigma^2\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0 \qquad (4.35)$$

σ^2을 모르면 $\hat{\sigma}^2 = MSE$ 로 대체하며 제곱근을 취한 값이 표준오차

$$\frac{\hat{E}(Y|\boldsymbol{x}=\boldsymbol{x}_0) - E(Y|\boldsymbol{x}=\boldsymbol{x}_0)}{s.e[\hat{E}(Y|\boldsymbol{x}=\boldsymbol{x}_0)]} \sim t(n-p')$$

27

4장 Review

2) 설명변수의 값이 $\boldsymbol{x}_0 = (1, x_{01}, x_{02}, \dots, x_{0p})^T$일때 새로운 관측(예측)값: $y_0 = (y|\boldsymbol{x}=\boldsymbol{x}_0)$

- 점추정량:

$$\hat{E}(Y|\boldsymbol{x}=\boldsymbol{x}_0) = \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} = \hat{y}_0$$

- 분산:

$$Var\left[\hat{y}_0|\boldsymbol{x}=\boldsymbol{x}_0\right] = \sigma^2(1+\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0) \qquad (4.36)$$

m개의 새로운 관측값에 대한 평균의 추정

$$Var\left[\hat{\bar{y}}_0|\boldsymbol{x}=\boldsymbol{x}_0)\right] = \sigma^2(1/m+\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0) \qquad (4.37)$$

σ^2을 모르면 $\hat{\sigma}^2 = MSE$ 로 대체하며 제곱근을 취한 값이 표준오차

앞으로는 간단히 $h_{00} = \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0$로 표기

28

4장 Review

$$\boldsymbol{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0p})^T \qquad h_{00} = \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0$$

$$\begin{aligned} \boldsymbol{x}_1 &= (1\, x_{11}\ \ x_{12}\ \ldots x_{1p})^T & h_{11} &= \boldsymbol{x}_1^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_1 \\ \boldsymbol{x}_2 &= (1\, x_{21}\ \ x_{22}\ \ldots x_{2p})^T & h_{22} &= \boldsymbol{x}_2^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_2 \\ &\vdots & &\vdots \\ \boldsymbol{x}_n &= (1\, x_{n1}\ \ x_{n2}\ \ldots x_{np})^T & h_{nn} &= \boldsymbol{x}_n^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_n \end{aligned}$$

$h_{11}, h_{22}, \ldots, h_{nn}$은 $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$의 대각원소

29

4장 Review

■ 외삽의 위험성

$h_{00} = \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0$ 이 커지면 식(4.35)(4.36)(4.37)의 분산도 커짐

신뢰구간폭이 증가하며 추정은 부정확하게 됨

단순회귀:

$$h_{00} = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}$$

30

4장 Review

- 적합결여검정(lack of fit test)

H_0: 주어진 모형이 적절함　　H_0: $E(Y) = \beta_0 + \beta_1 X$

H_1: 주어진 모형이 적절하지 않음　　H_1: $E(Y) \neq \beta_0 + \beta_1 X$

- σ^2의 값이 알려진 경우

$$\chi^2 = \frac{SSE}{\sigma^2} \sim \chi^2(n-2)$$

- H_0이 사실이면 $\chi^2 \sim \chi(n-2)$
- $\chi^2 \geq \chi^2(\alpha, n-2)$이면 H_0 기각

31

4장 Review

- σ^2의 값을 모르는 경우

모형에 의존하지 않는 σ^2의 추정방법이 필요

− y_{ij}: i번째 그룹의 j번째 반응변수의 값

$$i = 1, \cdots, m, \quad \sum_{i=1}^{m} n_i = n$$

$$SD_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \,/\, (n_i - 1), \quad i = 1, \cdots, m$$

- 순수제곱합 SS_{PE}

$$SS_{PE} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{m} (n_i - 1) SD_i^2$$

$$d.f = \sum_{i=i}^{m} (n_i - 1) = \sum_{i=i}^{m} n_i - \sum_{i=i}^{m} 1 = n - m$$

32

4장 Review

- $SSE = SS_{PE} + SS_{LOF}$ ($S_{yy} = SSE + SSR$ 과 유사)

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\bar{y}_i-\hat{y}_i)^2$$

$$\begin{array}{lccccc} & SSE & = & SS_{PE} & + & SS_{LOF} \\ df: & (n-2) & = & (n-m) & + & (m-2) \end{array}$$

- 적합결여에 따른 제곱합

$$SS_{LOF} = SSE - SS_{PE} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\bar{y}_i-\hat{y}_i)^2 = \sum_{i=1}^{m} n_i(\bar{y}_i-\hat{y}_i)^2$$

- 적합결여검정 : SS_{LOF}의 크기에 따라 H_0의 기각 여부 결정

$$F^* = \frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)} \sim F(m-2, n-m)$$

33

34