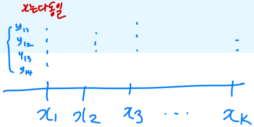
SLR : other topics

Test for Lack of Fit



■ 적합결여검정

- 두 변수 x와 y 사이의 함수관계가 단순회귀모형

$$y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

  으로 표현되는 것이 적합한가의 검정방법

- x의 각 수준(level)에서 반복측정(repeated observations)
  - ▹ x의 수준 : $x_1, x_2, \cdots, x_k$
  - ▹ 각 수준에서 $n_1, n_2, \cdots, n_k$개 반복 관측
  - ▹ $n = \sum_{i=1}^{k} n_i$

Test for Lack of Fit

$(x_1, y_{11}), \cdots, (x_1, y_{1n_1})$

$\vdots$

$(x_k, y_{1k}), \cdots, (x_k, y_k n_k)$

n개 $\to$ LSE $\hat{\beta}_0, \hat{\beta}_1$

■ 적합결여검정

$\hat{\beta}_0 + \hat{\beta}_1 x_1 = \hat{y}_{11} = \hat{y}_{12} = \hat{\beta}_0 + \hat{\beta}_1 x_1$

추정량 동일시

$$x_1 \to y_{11}, y_{12}, \cdots, y_{1n_1} \quad \to \hat{y}_1 \qquad n_1\text{개 관측} \Rightarrow \bar{y}_1 \quad \hat{y}_1$$

$y_{11} - \hat{y}_1 \quad y_{12} - \hat{y}_1 \quad \cdots$: 잔차

차이가 얼마 없어야 잘 적합된 회귀모형

$$x_2 \to y_{21}, y_{22}, \cdots, y_{2n_2} \quad \to \hat{y}_2 \qquad n_2\text{개 관측} \Rightarrow \bar{y}_2 \quad \hat{y}_2$$

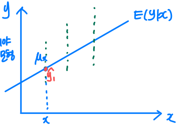
$$\vdots$$

$$x_k \to y_{k1}, y_{k2}, \cdots, y_{kn_k} \quad \to \hat{y}_k \qquad n_k\text{개 관측} \Rightarrow \bar{y}_k$$

n개

$$\sum_{j=1}^{n_1} (y_{1j} - \hat{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \hat{y}_2)^2 + \cdots + \sum_{j=1}^{n_k} (y_{kj} - \hat{y}_k)^2$$

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



$(y_{11}, y_{12}, \cdots, y_{1n})$

$\Rightarrow \hat{\mu}_{y|x_1} = \bar{y}_1$

모평균 $\mu \to x_1 \cdots x_n$, $\hat{\mu} = \bar{x}$

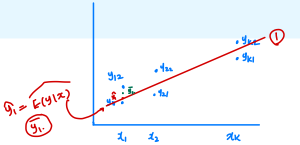
$y|x_1$, $\mu_{y|x_1} \to y_{11}, \cdots, y_{1n_1}$, $\hat{\mu}_{y|x_1} = \bar{y}_1$

- 최소제곱법으로 구한 회귀모형

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad i = 1, 2, \cdots, k$$

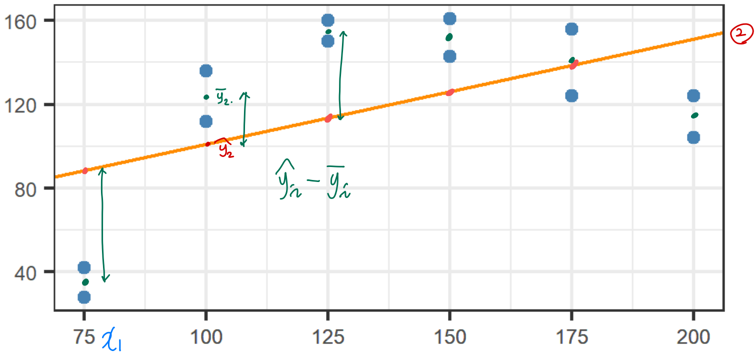
Test for Lack of Fit

- 가설



$$H_0 : \mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x$$

$$H_1 : \mathrm{E}(Y|X=x) \neq \beta_0 + \beta_1 x$$



Test for Lack of Fit

■ 제곱합분해

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

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \hat{y}_i)^2 = \sum_{i=1}^{k} n_i(\bar{y}_i - \hat{y}_i)^2$$

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

$$+2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)(\bar{y} - \hat{y}_i) = 2\sum_{i=1}^{k}\{(\bar{y}_i - \hat{y}_i)\underbrace{\boxed{\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)}}_{=0}\}$$

$$= \underbrace{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\underbrace{y_{ij}}_{n개} - \underbrace{\bar{y}_i}_{k개})^2}_{SSPE(\text{순오차제곱합})} + \underbrace{\sum_{i=1}^{k} n_i(\underbrace{\bar{y}_i}_{k개} - \underbrace{\hat{y}_i}_{2개})^2}_{SSLF(\text{적합결여제곱합})}$$

$\hat{\beta}_0, \hat{\beta}_1$ 추정시 제약조건 2개..

SSPE: Pure

H_0: 작아야 좋음

H_1: SSLF ↑

Test for Lack of Fit

검정

- 검정통계량

$F_0 \downarrow$: 모형적합이 잘되었다.

$$F_0 = \frac{MSLF}{MSPE} \sim_{H_0} F(n-2, n-k)$$

  - 순오차평균제곱(pure error mean square) : $MSPE = \dfrac{SSPE}{n-k}$

    고정되어있음
    회귀직선과 상관X

  - 적합결여평균제곱(lack-of-fit mean square) : $MSLF = \dfrac{SSLF}{k-2}$

- $f_0 = \dfrac{MSLF}{MSPE} > F_\alpha(n-2, n-k)$ 이면 귀무가설을 기각

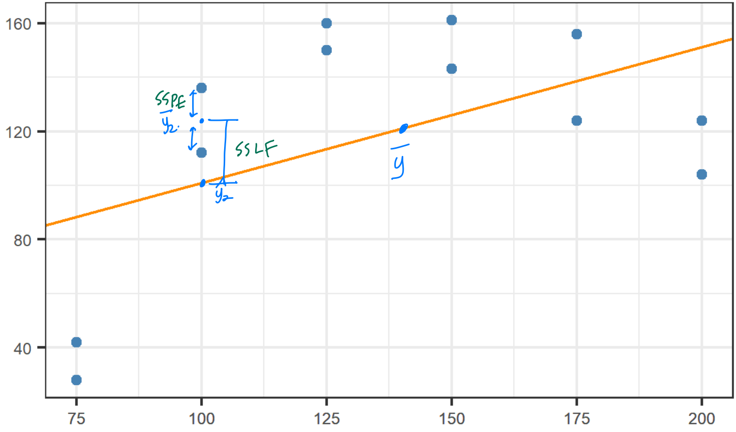
Test for Lack of Fit : Example

- 저축 예금자 자료

최저예금액 x (단위 : 천 원)	증가된 저축예금가입자 수 y		
	지점 A	지점 B	평 균
x_1 75	28	42	$\bar{y}_1 = 35$
x_2 100	112	136	$\bar{y}_2 = 124$
125	160	150	$\bar{y}_3 = 155$
150	143	161	$\bar{y}_4 = 152$
175	156	124	$\bar{y}_5 = 140$
x_6 200	124	104	$\bar{y}_6 = 114$

Test for Lack of Fit : Example

■ 저축 예금자 자료



Test for Lack of Fit : Example

■ 저축 예금자 자료

- 추정된 회귀 모형 : $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 50.857 + 0.503x$
- 분산분석표 $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

요 인	제 곱 합	자 유 도	평균제곱	F_0	$F_{0.05}(1, 10)$
회 귀	$SSR = 5,531.4$	1	5,531.4	3.54	4.94
잔 차	$SSE = 15,630.6$	10	1,563.1	기각안됨	
계	$SST = 21,162.0$	11		→ 회귀모형 부적합	

Test for Lack of Fit : Example

$H_0: \beta_0 + \beta_1 x$

vs $H_1: \neq \beta_0 + \beta_1 x$

- 저축 예금자 자료

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

$$= (28-35)^2 + (42-35)^2 + (112-124)^2 + \cdots = \underline{1,310.0}$$

$$SSLF = SSE - SSPE = 15,630.6 - 1310.0 = \underline{14,320.6}$$

$$F_0 = \frac{SSLF/4}{SSPE/6} = \frac{14,320.6/4}{1,310.0/6} = \frac{3,580.2}{218.3} = \underline{16.42}$$

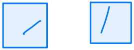
$$> F_{0.05}(4,6) = \underline{4.53}$$

H_0 기각 → 모형은 적합하지 않다.

두 회귀모형의 검정

$\mu_1 = \mu_2$?

두개가 같다면,



$(x.\ y)$ / n_1 / $y = \beta_{01} + \beta_{11}x + \varepsilon$

$(x.\ y)$ / n_2 / $y = \beta_{02} + \beta_{12}x + \varepsilon$

- 완전 모형 (full model) :

$$\begin{cases} y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \epsilon_{ij} \\ \quad i = 1, 2,\ \ j = 1, 2, \ldots, n_i \\ \quad \epsilon_{ij} \sim N(0, \sigma^2) \end{cases}$$

  - 모집단 1 : $\mathrm{E}(y_{1j} | x_{1j}) = \beta_{01} + \beta_{11} x_{1j}$
  - 모집단 2 : $\mathrm{E}(y_{2j} | x_{2j}) = \beta_{02} + \beta_{12} x_{2j}$

β_0 β_1

h_0 xy.. $\beta_{01} = \beta_{02}$ 아니야?

$\beta_{11} = \beta_{12}$ 아니야?

두 회귀모형의 검정

- 가설

$$H_0 : \beta_{01} = \beta_{02} \text{ and } \beta_{11} = \beta_{12}$$

$$H_1 : \beta_{01} \neq \beta_{02} \text{ or } \beta_{11} \neq \beta_{12}$$

- 축소 모형 (reduced model) :

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij},$$

$$i = 1, 2, \ j = 1, 2, \ldots, n_i$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

  - ▷ $\beta_{01} = \beta_{02} = \beta_0, \ \ \beta_{11} = \beta_{12} = \beta_1$

두 회귀모형의 검정

- (Step 1) 완전모형의 잔차제곱합 $SSE(F)$를 구한다.

$$SSE(F) = SSE_1 + SSE_2$$

$$SSE_i = \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_{ij})^2 = \sum_{j=1}^{n_i} (y_{ij} - \hat{\beta}_{0i} - \hat{\beta}_{1i} x_{ij})^2, \ i = 1, 2$$

- (Step 2) 축소모형의 잔차제곱합 $SSE(R)$을 구한다.

$$SSE(R) = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_{ij})^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{ij})^2$$

두 회귀모형의 검정

- (Step 3) 검정통계량

카이제곱분포

$$F_0 = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

자유도 SSE1 SSE2

- ▹ $df_R = (n_1 - 1) + (n_2 - 1),\ df_F = (n_1 - 2) + (n_2 - 2)$
- ▹ $F_0 \sim F(df_R - df_F, df_F) = F(2, n_1 + n_2 - 4)$

$\beta_{01} = \beta_{02}$
$\beta_{11} = \beta_{12}$

- (Step 4) 유의수준 α에서, $F_0 > F_\alpha(2, n_1 + n_2 - 4)$ 이면 H_0 기각

Example

■ 회귀모형 비교

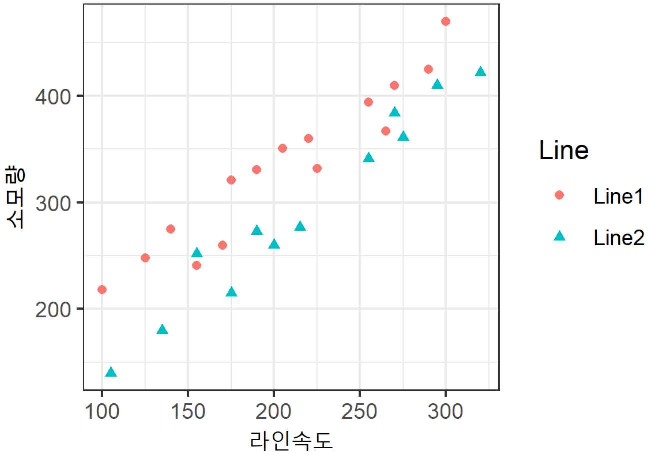
- (예제) 맥주를 생산하는 어느 맥주회사에 두 개의 생산라인(production line)이 있다. 이 라인을 움직이는 라인 속도(line speed)와 하루 동안에 라인으로부터 흘러 나와서 못 쓰게 되는 맥주의 양 간에 는 관계가 있는 것으로 판명되었다. 그런데, 라인속도와 흘러 나오는 소모량 간의 관계가 생산라인이 다름에 따라 차이가 있는가를 알기 위해서 다음의 실험 자료를 얻었다. 두 회귀모형의 동일성 여부를검정하시오.

Example

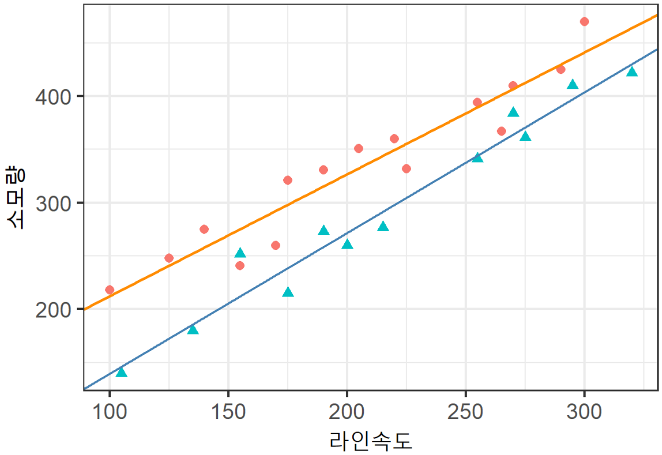
Table: 맥주생산라인의 자료

생 산 라 인 1			생 산 라 인 2		
no.	라인속도(x_{1j})	소모량(y_{1j})	no.	라인속도(x_{2j})	소모량(y_{2j})
1	100	218	1	105	140
2	125	248	2	215	277
3	220	360	3	270	384
4	205	351	4	255	341
5	300	470	5	175	215
6	255	394	6	135	180
7	225	332	7	200	260
8	175	321	8	275	361
9	270	410	9	155	252
10	170	260	10	320	422
11	155	241	11	190	273
12	190	331	12	295	410
13	140	275			
14	290	425			
15	265	367			

Example

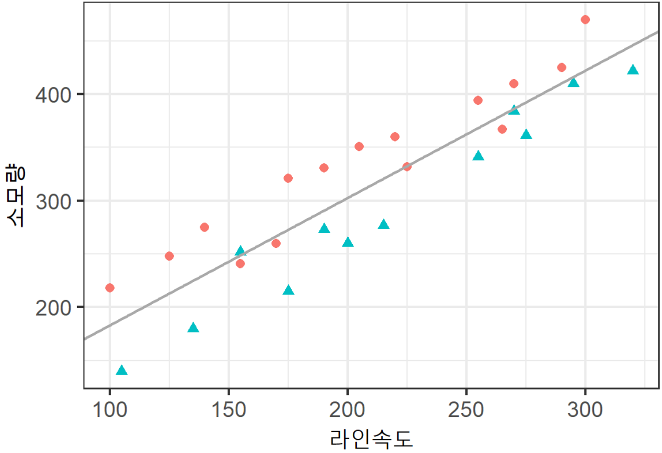


Example

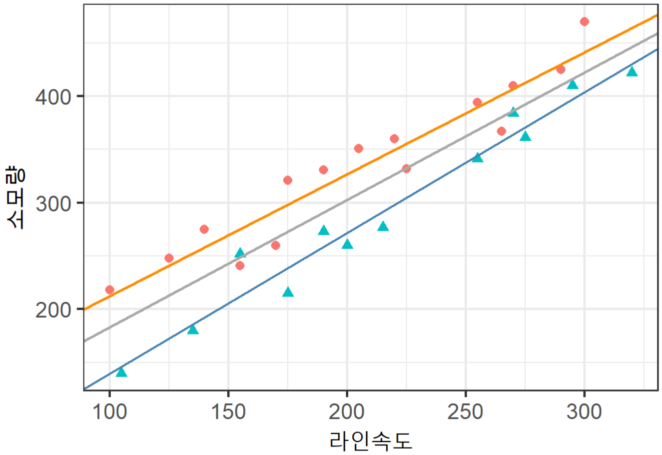


Example

Reduced



Example



Example

Table: 각 생산라인의 회귀분석 자료

생산라인 1

회귀모형 : $\hat{y}_{1j} = 97.965 + 1.145x_{1j}$

분산분석표

요인	제곱합	자유도
회 귀	$SSR_1 = 70,441$	1
잔 차	$SSE_1 = \ \ 6,403$	13
계	$SST_1 = 76,844$	14

생산라인 2

회귀모형 : $\hat{y}_{2j} = 7.574 + 1.322x_{2j}$

분산분석표

요인	제곱합	자유도
회 귀	$SSR_2 = 87,726$	1
잔 차	$SSE_2 = \ \ 3,501$	10
계	$SST_2 = 91,227$	11

$SSE(F) = SSE_1 + SSE_2 = 6,403 + 3,501 = 9,904$

$df_F = (n_1 - 2) + (n_2 - 2) = 13 + 10 = 23$

Example

Table: 축소모형의 회귀분석 자료

회 귀 직 선 :	$\hat{y}_{ij} = 64.036 + 1.196x_{ij}$		
분산분석표 :	요인	제곱합	자유도
	회 귀	$SSR(R) = 149,661$	1
	잔 차	$SSE(R) = 29,408$	25
	계	$SST(R) = 179,069$	26

$SSE(R) = 29,408$

$df_R = (n_1 - 1) + (n_2 - 1) = 14 + 11 = 25$

Example

- 가설 :

$$H_0 : \beta_{01} = \beta_{02} \text{ and } \beta_{11} = \beta_{12}$$

$$H_1 : \underbrace{\beta_{01} \neq \beta_{02}}_{\text{절편}} \text{ or } \underbrace{\beta_{11} \neq \beta_{12}}_{\text{기울기}}$$

- 검정통계량

$$\begin{aligned} F_0 &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{29408 - 9904}{2} \div \frac{9904}{23} = 22.65 > F_{0.05}(2, 23) = 3.42 \end{aligned}$$

- 귀무가설 기각. 두 라인의 회귀계수는 동일하지 않다.

두 기울기의 비교

$\hat{\beta}_{11} \sim N\left(\beta_{01}, \frac{\sigma^2}{S_{xx}}\right)$ — 15개..

$\hat{\beta}_{12} \sim N\left(\beta_{02}, \frac{\sigma^2}{S_{xx}'}\right)$ — 12개..

- 기울기 비교에 대한 가설 : $H_0 : \beta_{11} = \beta_{12} \;\; vs. \;\; H_1 : \beta_{11} \neq \beta_{12}$
- 검정통계량

2가 맞어

2 : 오타   2 : 오타

$$t_0 = \frac{\hat{\beta}_{11} - \hat{\beta}_{12}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_{11} - \hat{\beta}_{12})}} \sim_{H_0} t((n_1 - 1) + (n_2 - 1))$$

- 두 표본이 독립이라고 가정하면

$$\mathrm{Var}(\hat{\beta}_{11} - \hat{\beta}_{12}) = \mathrm{Var}(\hat{\beta}_{11}) + \mathrm{Var}(\hat{\beta}_{12})$$

$$= \frac{\sigma^2}{\sum(x_{1j} - \bar{x}_1)^2} + \frac{\sigma^2}{\sum(x_{2j} - \bar{x}_2)^2}$$

$\frac{\sigma^2}{S_{xx}}$ $\frac{\sigma^2}{S_{xx}'}$

$$\widehat{\mathrm{Var}}(\hat{\beta}_{11} - \hat{\beta}_{12}) = MSE(F)\left[\frac{1}{\sum(x_{1j} - \bar{x}_1)^2} + \frac{1}{\sum(x_{2j} - \bar{x}_2)^2}\right]$$

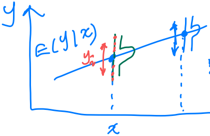
Example

$H_0 : \beta_{11} = \beta_{12}$

- $\hat{\beta}_{11} = 1.1454, \;\; \hat{\beta}_{12} = 1.3221$
- $SSE(F) = 9904, \;\; MSE(F) = \dfrac{SSE(F)}{df_F} = \dfrac{9904}{23} = 430.6$
- $\widehat{\mathrm{Var}}(\hat{\beta}_{11} - \hat{\beta}_{12}) = 430.6 \left[\dfrac{1}{53,693} + \dfrac{1}{50,192} \right] = 0.0166$
- 검정통계량 : $t_0 = \dfrac{1.145 - 1.322}{\sqrt{0.0166}} = -1.374$
- $|t_0| < t_{0.025}(23) = 2.069$ 이므로 유의수준 5%에서 귀무가설 기각하지 못함.

Weighted Regression

$\epsilon_i \sim N(0, \sigma^2)$. Indep



$E(y|X=x)$

등분산성 ↕↕ 비슷함!

$x_i \uparrow$ $\sigma_i \uparrow$

$E \uparrow$

① 변수변환

$y = \beta_0 + \beta_1 x$

$\Rightarrow \log y = \beta_0 + \beta_1 x$

- 가중회귀모형, $i = 1, \ldots, n$

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

$\rightarrow E(MSE) = \sigma^2$

$$y_i = \beta_0 + \beta_1 x_1 + \epsilon_i, \quad \mathrm{Var}(\epsilon_i) = \sigma_i^2 = \frac{\sigma^2}{w_i}$$

$N(0, \sigma_i^2)$

$\frac{1}{w_i} \propto x_i^2$

w_i: 알려진값

- 오차제곱합

$S = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

$$Q = \sum_{i=1}^{n} w_i \left\{ y_i - (\beta_0 + \beta_1 x_i) \right\}^2$$

- 가중회귀최소추정량(WLSE)

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1 \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} w_i \left\{ y_i - (\beta_0 + \beta_1 x_i) \right\}^2$$

WLSE

$\hat{\beta}_1 = \dfrac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$

- 정규방정식

$$\begin{cases} \hat{\beta}_0 \sum w_i + \hat{\beta}_1 \sum w_i x_i = \sum w_i y_i \Leftarrow \dfrac{\partial Q}{\partial \beta_0} \\ \hat{\beta}_0 \sum w_i x_i + \hat{\beta}_1 \sum w_i x_i^2 = \sum w_i x_i y_i \Leftarrow \dfrac{\partial Q}{\partial \beta_1} \end{cases}$$

- 가중최소제곱추정량 (WLSE)

$$\hat{\beta}_1 = \frac{\sum w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum w_i (x_i - \bar{x}_w)^2}, \quad \hat{\beta}_0 = \bar{y}_w - \hat{\beta}_1 \bar{x}_w$$

- $\bar{x}_w, \bar{y}_w$: 가중평균

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}, \quad \bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}$$

Quadratic form

벡터: 열벡터
고딕체

$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$
$n \times 1$
row column

■ $\boldsymbol{y}$의 이차형식(quadratic form)

$$\underbrace{\underbrace{\underset{1\times n}{\boldsymbol{y}^\top} \underset{n\times n}{A}}_{1\times n} \underset{n\times 1}{\boldsymbol{y}}}_{1\times 1} = \sum_{i=1}^{n} a_{ij} y_i y_j = \sum_{i=1}^{n} a_{ii} {y_i}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ii} y_i y_j$$

- $\boldsymbol{y}^\top = (y_1, y_2, \cdots, y_n)$: $n \times 1$ vector
- $A = (a_{ij})$: 이차형식 $\boldsymbol{y}^\top A \boldsymbol{y}$의 계수, $n \times n$ symmetric matrix

대칭 $\left(\diagdown \right)$ 대각선기준 대칭행렬

$a_{12} = a_{21}$
$\vdots$
$a_{ij} = a_{ji}$ $(i \neq j)$

예)

$$y^TAy = (y_1\ \ y_2\ \ y_3)\begin{pmatrix}3&5&1\\5&13&0\\1&0&1\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix} = (3y_1+5y_2+y_3,\ 5y_1+13y_2,\ y_1+y_3)\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$$

$$= 3y_1^2+5y_1y_2+y_1y_3+5y_1y_2+13y_1y_2+y_1y_3+y_3^2$$

$$= 3y_1^2+13y_2^2+y_3^2+10y_1y_2+2y_1y_3$$

$$= (y_1+2y_2)^2+(y_1+3y_2)^2+(y_1+y_3)^2 \geq 0$$

$$= a^2 + b^2 + c^2$$

$y^TAy \geq 0.$

$y^TAy = 0 \Leftrightarrow a=b=c=0 \Leftrightarrow \begin{cases} y_1+2y_2=0 \\ y_2+3y_3=0 \\ y_1+y_3=0 \end{cases} \Leftrightarrow y_2=0,\ y_1=0,\ y_3=0$

$y^TAy > 0.\ (y \neq 0)$ ⇒ A: positive def matrix (pdm) 양정치행렬

Quadratic form

ax^2
$a>0 \to ax^2 \geq 0$
$x \neq 0 \to ax^2 > 0$

- $\mathbf{0}_n$ 이 아닌 모든 벡터 $\boldsymbol{y}$에 대하여
  - $\boldsymbol{y}^\top A\boldsymbol{y} > 0 \;\Rightarrow\; A$: 양정치(positive definite)행렬 pd
  - $\boldsymbol{y}^\top A\boldsymbol{y} \geq 0 \;\Rightarrow\; A$: 양반정치(positive semidefinite)행렬 psd
  - $\boldsymbol{y}^\top A\boldsymbol{y} < 0 \;\Rightarrow\; A$: 음정치(negative definite)행렬
  - $\boldsymbol{y}^\top A\boldsymbol{y} \leq 0 \;\Rightarrow\; A$: 음반정치(negative semidefinite)행렬

Quadratic form : SST

$\sum y_i^2 - 2\sum y_i \bar{y} + \sum (\bar{y})^2 = \sum y_i^2 - n(\bar{y})^2 = \frac{(\sum y_i)^2}{n}$

$\Rightarrow 2\bar{y}\sum y_i = 2n(\bar{y})^2$, $\quad \sum(\bar{y})^2 \to n(\bar{y})^2$

$\bar{y} = \frac{1}{n}\sum y_i \Rightarrow \sum y_i = n\bar{y}$

$y^T y = (y_1, \cdots, y_n)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = y_1^2 + y_2^2 + \cdots + y_n^2 = \sum y_i^2$

SST

$$\sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$

$$= \boldsymbol{y}^\top I_n \boldsymbol{y} - \frac{1}{n}\boldsymbol{y}^\top \underbrace{\mathbf{1}_n \mathbf{1}_n^\top}_{=J_n} \boldsymbol{y} = \boldsymbol{y}^\top \overbrace{\left(I_n - \frac{1}{n}J_n\right)}^{A} \boldsymbol{y} = y^T A y$$

묶어주기위해 항등행렬 넣어주기

y의 계수

$\sum y_i = (1, \cdots, 1)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = y_1 + \cdots + y_n = 1_n^T y$, $\quad (\sum y_i)^2 = y^T 1_n 1_n^T y$

where

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \; J_n = \mathbf{1}_n \mathbf{1}_n^\top = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

Multivariate normal distribution

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$$

$$V = \begin{pmatrix} Var(y_1) & cov(y_1, y_2) & \cdots & Cov(y_1, y_n) \\ & \ddots & & Cov(y_i, y_j) \\ & & \ddots & \\ & & & Var(y_n) \end{pmatrix}$$

symm

- $\boldsymbol{y}^\top = (y_1, y_2, \cdots, y_n)$: random vector
- $\boldsymbol{\mu}^\top = (\mu_1, \mu_2, \cdots, \mu_n)$: mean vector of y
- V : variance-covariance matrix of $\boldsymbol{y}$ (positive definite)
- probability density function

다변량정규분포를 따른다.

$$f(y_1, y_2, \cdots, y_n) = \frac{e^{-\frac{1}{2}(\underset{}{\overset{1\times n}{\boldsymbol{y}-\boldsymbol{\mu}}})^\top \underset{n\times n}{V^{-1}} \underset{n\times 1}{(\boldsymbol{y}-\boldsymbol{\mu})}}}{(2\pi)^{\frac{1}{2}n} \, |\underline{V}|^{\frac{1}{2}}}$$

determinant
판별식

Multivariate normal distribution

$0_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_{n\times 1}$ $\quad$ $\mathrm{cov}(y_i, y_j) = 0 \Rightarrow y_i \perp y_j$ 서로 독립

- $\boldsymbol{y} \sim N(\mathbf{0}_n, I_n) \Rightarrow \underbrace{\boldsymbol{y}^\top \boldsymbol{y}}^{I_n} = \sum_{i=1}^{n} y_i^2 \sim \chi^2(n)$
- $Q_1 \sim \chi^2(n_1),\ Q_2 \sim \chi^2(n_2)$: 서로 독립

$$\Rightarrow \frac{Q_1/n_1}{Q_2/n_2} \sim F(n_1, n_2)$$

- $y \sim N(0,1),\ Q \sim \chi^2(n)$: 서로 독립

$$\Rightarrow \frac{y}{\sqrt{Q/n}} \sim t(n)$$

비중심 χ^2-분포

χ^2 분포와 비슷한데, 중심위치가 조금다른 분포...

$X \sim \chi^2(n)$

$\Rightarrow E(X) = n$

$= \sum (1+\mu_i)^2 = n + \sum \mu_i^2 = n + \mu^\top \mu$

$E(\sum y_i^2) = \sum E(y_i^2)$

$\Rightarrow E(y_i - \mu_i + \mu_i)^2$

$\Rightarrow \underbrace{E(y_i - \mu_i)^2}_{\text{분산}} + 2\underbrace{E(y_i - \mu_i)}_{=0}\mu_i + \mu_i^2$

$= 1 + \mu_i^2$

- 비중심 χ^2-분포

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, I_n) \;\Rightarrow\; \boldsymbol{y}^\top \boldsymbol{y} \sim \chi^2(n, \lambda),\;\; \lambda = \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\mu}$$

- 비중심 F-분포 : $Q_1 \sim \chi^2(n_1, \lambda),\;\; Q_2 \sim \chi^2(n_2, \lambda)$, 서로 독립

$$\Rightarrow \frac{Q_1/n_1}{Q_2/n_2} \sim F(n_1, n_2, \lambda)$$

$V \sim \chi^2(n, \lambda)$

$\Rightarrow E(V) = n + 2\lambda$

distribution of quadratic forms

* $f(\lambda) = |A - \lambda I_n| = 0$ $\quad A : n \times n$

λ : 고유값

$$\begin{pmatrix} a_{11}-\lambda & a_{12} & \cdots \\ & \ddots & \\ & & a_{nn}-\lambda \end{pmatrix}$$

$\underline{Ax = \lambda x}$

x : 고유벡터

- A : 멱등행렬(idempotent matrix) $\Leftrightarrow$

$$AA = A$$

⟨정리 1.29⟩ 멱등행렬의 고유값은 0 또는 1 이다.

⟨정리 1.30⟩ 행렬 A가 rank$(A) = k$ 인 멱등행렬일 때는

$$P^\top AP = E_k$$

를 만족하는 직교행렬 P가 존재한다. 여기서 E_k 는 대각 원소 중 k 개가 1, 나머지는 0 인 대각행렬을 의미한다.

(eigenvector)

〈정리 1.31〉 $\lambda_1, \ldots, \lambda_n$: $n \times n$ 행렬 A의 고유값

대각행렬 다 더한것

$$tr(A) = \sum_{i=1}^{n} \lambda_i$$

$$tr(A^\top A) = \sum_{i=1}^{n} \lambda_i^2$$

$$tr(A^{-1}) = \sum_{i=1}^{n} \lambda_i^{-1}$$

$$|A| = \prod_{i=1}^{n} \lambda_i^{-1}$$

만약 A가 멱등행렬이면, $tr(A) = \text{rank}(A)$.

$A = \left(\quad \right)_{n \times n}$

제1장 확인

$\Sigma a_i c_i = 0 \Leftrightarrow c=0 \ldots$

rank(A) : linearly independent 수

ex) $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$ $\times 2$, $\times 3$: independent 하지않다.

rank(A)=1

$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 8 \end{pmatrix}$: rank(A)=2

distribution of quadratic forms

$y \sim N(\mu, I_n) \Rightarrow y^T y \quad E(y^T y) = tr(I_n \cdot I_n) + \mu^T I_n \mu$

$= n + \mu^T \mu$

⟨정리 5.1⟩ 만약 $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$ 이면,

$$\mathrm{E}(\boldsymbol{y}^\top A\boldsymbol{y}) = \boxed{\mathrm{tr}(AV) + \boldsymbol{\mu}^\top A\boldsymbol{\mu}}, \quad \mathrm{Cov}(\boldsymbol{y},\ \boldsymbol{y}^\top A\boldsymbol{y}) = 2VA\boldsymbol{\mu}$$

⟨정리 5.2⟩ 만약 $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$ 이면,

$$\mathrm{Var}(\boldsymbol{y}^\top A\boldsymbol{y}) = 2\mathrm{tr}(AV)^2 + 4\boldsymbol{\mu}^\top AVA\boldsymbol{\mu}$$

⟨정리 5.3⟩ 만약 $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$ 이면,

$$\boldsymbol{y}^\top A\boldsymbol{y} \sim \chi^2\left(r(A), \frac{1}{2}\boldsymbol{\mu}^\top A\boldsymbol{\mu}\right) \Leftrightarrow AV : \text{idempotent matrix}$$

멱등행렬

$AVAV = AV$

$r(I_n) = n$

$A = I_n$

$V = I_n$

$AV = I_n$

distribution of quadratic forms

⟨정리 5.4⟩

(1) $\boldsymbol{y} \sim N(\mathbf{0}_n, I_n)$ 이면

$$\boldsymbol{y}^\top A\boldsymbol{y} \sim \chi^2(p) \Leftrightarrow A : \text{idempotent matrix with } \text{rank}(A) = p$$

(2) $\boldsymbol{y} \sim N(\boldsymbol{\mu}, I_n\sigma^2)$ 이면

비중심카이제곱분포

$$\boldsymbol{y}^\top \boldsymbol{y}/\sigma^2 \sim \chi^2(n, \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\mu}/\sigma^2)$$

(3) $\boldsymbol{y} \sim N(\boldsymbol{\mu}, I_n)$ 이면

$$\boldsymbol{y}^\top A\boldsymbol{y} \sim \chi^2(p, \frac{1}{2}\boldsymbol{\mu}^\top A\boldsymbol{\mu}) \Leftrightarrow$$

$$A : \text{idempotent matrix with } \text{rank}(A) = p$$

$$SST = \sum (y_i - \bar{y})^2 = y^T \left(I_n - \frac{1}{n}J\right) y$$

$$\frac{SST}{\sigma^2} = y^T \left(I_n - \frac{1}{n}J\right)/_{\sigma^2} \cdot y \quad : \quad \underline{A = \frac{1}{\sigma^2}\left(I_n - \frac{1}{n}J\right)}$$

$y = (y_1 \cdots y_n)^T$ $n \times 1$, $\quad y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \Rightarrow y \sim N(\mu, \underline{I_n \sigma^2})$ $\quad V$

$$\mu = \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{pmatrix}_{n\times 1} = \underbrace{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}_{n \times 2} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}_{2\times 1}}_{n \times 1} \quad \leftarrow (*) \quad \mu = (1 \quad x)\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\mu^T = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}^T (1 \quad x)^T = (\beta_0 \quad \beta_1)\begin{pmatrix} 1^T \\ x^T \end{pmatrix}$$

$y^T A y \quad : \quad AV = \frac{1}{\cancel{\sigma^2}}\left(I_n - \frac{1}{n}J\right) \cdot I_n \cancel{\sigma^2} = A, \quad J = 1 1^T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}(1 \cdots 1)$

정리 5.30 이용하여 멱등행렬인것 보이기

$$AA = \left(I_n - \frac{1}{n} 1 1^T\right)\left(I_n - \frac{1}{n} 1 1^T\right) \qquad (1 \cdots 1)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n$$

$$= I_n - \frac{1}{n} 1 1^T - \frac{1}{n} 1 1^T + \frac{1}{n^2} 1 \underbrace{1^T 1}_{=n} 1^T$$

$$= I_n - \frac{1}{n} 1 1^T = A \quad \Rightarrow A : \text{멱등} \Rightarrow AV : \text{멱등}$$

$$y^T A y \sim \chi^2\left(r(A), \tfrac{1}{2}\mu^T A \mu\right)$$

$$r(A) = tr(A) = tr\left(I_n - \frac{1}{n} 1 1^T\right) = tr(I_n) - tr\left(\frac{1}{n} 1 1^T\right) = n -$$

$$tr(A+B) = tr(A) + tr(B)$$

$$tr(AB) = tr(BA)$$

$$\mu^T A \mu = (\beta_0 \quad \beta_1)\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix}\left(I_n - \frac{1}{n}J\right)\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$= (\beta_0 \quad \beta_1)\underbrace{\underbrace{\begin{pmatrix} 1^T \\ x^T \end{pmatrix}_{2\times n} \underline{\left(I_n - \frac{1}{n} 1 1^T\right)}_{n \times n}}_{2 \times n} (1 \quad x)_{n \times 2}\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

교수님이

날리심ㅠㅠ

계산이복잡해서..

$$= \begin{pmatrix} 1^T - \frac{1}{n} 1^T 1 1^T \\ x^T - \frac{1}{n} x^T 1 1^T \end{pmatrix} = \begin{pmatrix} 1^T - 1^T \\ x^T - \frac{1}{n} x^T 1 1^T \end{pmatrix} = \begin{pmatrix} 0 \\ x^T - \frac{1}{n} x^T 1 1^T \end{pmatrix}(1 \quad x) = \begin{pmatrix} 0 & 0 \\ x^T 1 - \frac{1}{n} x^T \underline{1 1^T}_{n} & x^T x - \frac{1}{n} x^T 1 1^T x \end{pmatrix}$$

$$\Rightarrow (\beta_0 \quad \beta_1)\begin{pmatrix} 0 & 0 \\ x^T 1 - \frac{1}{n} x^T 1 1^T & x^T x - \frac{1}{n} x^T 1 1^T x \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} =$$

$$1^T A = 1^T\left(I_n - \frac{1}{n} 1 1^T\right) = 1^T - \frac{1}{n}\underbrace{1^T 1}_{n} 1^T = 0$$

$$= \begin{pmatrix} 1^T \\ x^T \end{pmatrix} A (1 \quad x) = \begin{pmatrix} 1^T A \\ x^T A \end{pmatrix}(1 \quad x) = \begin{pmatrix} \underbrace{1^T A 1}_{=0} & \underbrace{1^T A x}_{=0} \\ \underbrace{x^T A 1}_{=0} & \underbrace{x^T A x}_{= \sum (x_i - \bar{x})^2 = S_{(xx)}} \end{pmatrix}$$

$$\Rightarrow (\beta_0 \quad \beta_1)\begin{pmatrix} 0 & 0 \\ 0 & S_{(xx)} \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$= (0 \quad \beta_1 S_{(xx)})\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \beta_1^2 S_{(xx)}/\sigma^2 \qquad A = \left(I_n - \frac{1}{n}J\right)/\sigma^2$$

$$y^T A y \sim \chi^2\left(n-1, \frac{\beta_1^2 S_{(xx)}}{2\sigma^2}\right), \quad E(y^T A y) = (n-1) + \frac{\beta_1^2 S_{(xx)}}{\sigma^2} \qquad E(SST) = (n-1)\sigma^2 + \beta_1^2 S_{(xx)}$$

$$E\left(\frac{SST}{\sigma^2}\right) = (n-1) + \frac{\beta_1^2 S_{(xx)}}{\sigma^2}$$

$$E(SST) = (n-1)\sigma^2 + \beta_1^2 S_{(xx)}$$

불편추정량

$$E(SSE) \Rightarrow \quad MSE = \frac{SSE}{n-2} = \hat{\sigma}^2 \Rightarrow E(MSE) = \sigma^2$$

$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2$$

$$E(e_i) = E(y_i - \hat{y}_i) = E(y_i) - E(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= \beta_0 + \beta_1 x_i - (\beta_0 + \beta_1 x_i) = 0$$

$$e_i = y_i - \hat{y}_i = y_i - \underbrace{\hat{\beta}_0}_{=\bar{y} - \hat{\beta}_1 \bar{x}} - \hat{\beta}_1 x_i = y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x}) \qquad \hat{\beta}_1 = \frac{(x_j - \bar{x})}{S_{xx}}$$

$$= y_i - \sum_{j=1}^{n} \left\{ \frac{1}{n} y_j - \frac{(x_j - \bar{x})}{S_{xx}} (x_i - \bar{x}) y_j \right\}$$

$$= y_i - \sum_{j=1}^{n} \left\{ \frac{1}{n} - \frac{(x_j - \bar{x})}{S_{(xx)}} (x_i - \bar{x}) \right\} y_j$$

$$Var(e_i) = \underbrace{Var(y_i)}_{= \sigma^2} + \underbrace{Var\left(\sum_{j=1}^{n} \{ \quad \} y_j\right)}_{*} - 2\underbrace{Cov\left(y_i, \sum_{j=1}^{n} \{ \quad \} y_j\right)}_{**}$$

$$Cov(X, Y+Z) = Cov(X, Y) + Cov(X, Z)$$

$$(*): \sum_{j=1}^{n} \{ \quad \}^2 \sigma^2 = \sum_{j=1}^{n} \left\{ \frac{1}{n^2} - \frac{2}{n}\frac{(x_j - \bar{x})}{S_{(xx)}}(x_i - \bar{x}) + \frac{(x_j - \bar{x})^2}{S_{(xx)}^2}(x_i - \bar{x})^2 \right\} \sigma^2$$

$$= \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \quad = 0$$

$$(**): Cov(y_i, (y_1 + y_2 + \cdots + y_n))$$

$$\Rightarrow Cov\left(y_i, \left\{\frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right\} y_i\right) = \left\{\frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right\} \sigma^2$$

$$Var(e_i) = Var(y_i) + Var\left(\sum_{j=1}^{n} \{ \quad \} y_i\right) - 2Cov\left(y_i, \sum_{j=1}^{n} \{ \quad \} y_i\right)$$

$$= \sigma^2 + \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right)\sigma^2 - 2\sigma^2\left(\frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right)$$

$$\boxed{Var(e_i) = \left\{ 1 - \left(\frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \right\} \sigma^2}$$

$$E(SSE) = E\left(\sum e_i\right)^2 \qquad E(e_i) = 0$$

$$= \sum \left(E(e_i^2)\right)$$

$$= \sum Var(e_i) = \sum_{i=1}^{n} \left\{ 1 - \left(\frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right) \right\} \sigma^2$$

$$= \sigma^2 \{ n - 1 - 1 \} = \sigma^2 (n-2)$$

$$\Rightarrow E\left(\frac{SSE}{n-2}\right) = \sigma^2$$

$$\Rightarrow E(MSE) = \sigma^2$$

$$\Rightarrow E(SST) = (n-1)\sigma^2 + \beta_1^2 S_{xx}$$

$$E(SSE) = (n-2)\sigma^2$$

$$E(SSR) = E(SST) - E(SSE) = \sigma^2 + \beta_1^2 S_{(xx)}$$

$$H_0: \boxed{\beta_1 = 0} \quad vs \quad H_1: \beta_1 = 0$$

$$F_0 = \frac{MSR}{MSE}$$

여기가0이되므..

$$\Rightarrow \frac{E(MSR)}{E(MSE)} = \frac{\sigma^2 + \beta_1^2 S_{xx}}{\sigma^2} = 1 + \beta^2 \frac{S_{xx}}{\sigma_2}$$

distribution of quadratic forms

$\langle$ 정리 5.5 $\rangle$ $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$ 일 때,

$\boldsymbol{y}^\top A\boldsymbol{y}$와 $B\boldsymbol{y}$가 독립적으로 분포(distributed independently) $\Leftrightarrow$

$$BVA = O_n$$

$\langle$ 정리 5.6 $\rangle$ $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$ 일 때,

두 이차형식, $\boldsymbol{y}^\top A\boldsymbol{y}$와 $\boldsymbol{y}^\top B\boldsymbol{y}$가 독립적으로 분포 $\Leftrightarrow$

$$AVB = O_n(\text{또는} BVA = O_n)$$

distribution of quadratic forms

⟨ 정리 5.7 ⟩ 다음이 성립하기 위한 필요충분조건은 다음의 I 또는 II 이다.

$$\boldsymbol{y}^\top A_i \boldsymbol{y} \sim \chi^2(k_i, \frac{1}{2}\boldsymbol{\mu}^\top A_i \boldsymbol{\mu})$$

$\boldsymbol{y}^\top A_i \boldsymbol{y}$: mutually independent

$$\boldsymbol{y}^\top A \boldsymbol{y} \sim \chi^2(k, \frac{1}{2}\boldsymbol{\mu}^\top A \boldsymbol{\mu})$$

이 때, $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$, $i = 1, 2, \cdots, p$ 에 대하여

A_i : symmetric matrix with $\text{rank}(A_i) = k_i$

$A = \sum_{i=1}^{p} A_i$: symmetric matrix with $\text{rank}(A) = k$

$\langle$정리 5.7$\rangle$ 계속

I : 다음의 (a), (b), (c) 중 두 개 조건만 성립하면 된다.

- (a) $A_i V$는 모든 i에 대하여 멱등행렬이다.
- (b) 모든 $i < j$에 대하여, $A_i V A_j = O_n$
- (c) AV는 멱등행렬이다.

II : I의 (c)가 옳고, 또한 $k = \sum_{i=1}^{p} k_i$가 성립한다.

distribution of quadratic forms

$\langle$정리 5.8 $\rangle$ (코크란의 정리)

$$\boldsymbol{y} \sim N(\mathbf{0}_n, I_n),$$

$$A_i : \text{대칭행렬},\ \ i = 1, 2, \cdots, p$$

$$\textstyle\sum_{i=1}^{p} A_i = I_n$$

이면, $\boldsymbol{y}^\top A_i \boldsymbol{y} \sim \chi^2(k_i)$ 이며 서로 독립적으로 분포되기 위한 필요충분조건은

$$\textstyle\sum_{i=1}^{p} k_i = n$$

이다.