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

적합결여검정

$$y_{11}, y_{12}, \cdots, y_{1n_1}$$

$$y_{21}, y_{22}, \cdots, y_{2n_2}$$

$$\vdots$$

$$y_{k1}, y_{k2}, \cdots, y_{kn_k}$$

- 최소제곱법으로 구한 회귀모형

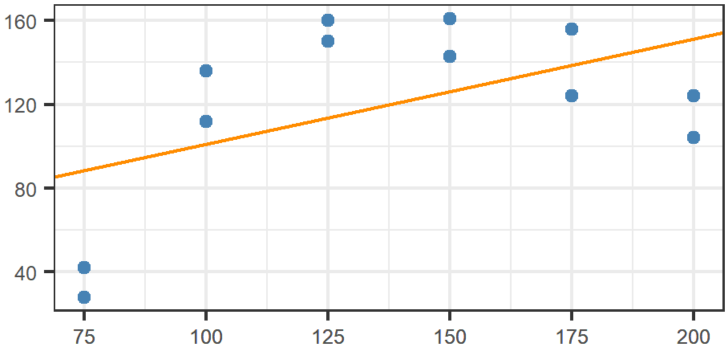
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad i = 1, 2, \cdots, k$$

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

$$
\begin{aligned}
SSE &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} \{(y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i)\}^2 \\
&= \underbrace{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{SSPE(\text{순오차제곱합})} + \underbrace{\sum_{i=1}^{k} n_i(\bar{y}_i - \hat{y}_i)^2}_{SSLF(\text{적합결여제곱합})}
\end{aligned}
$$

Test for Lack of Fit

검정

- 검정통계량

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

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

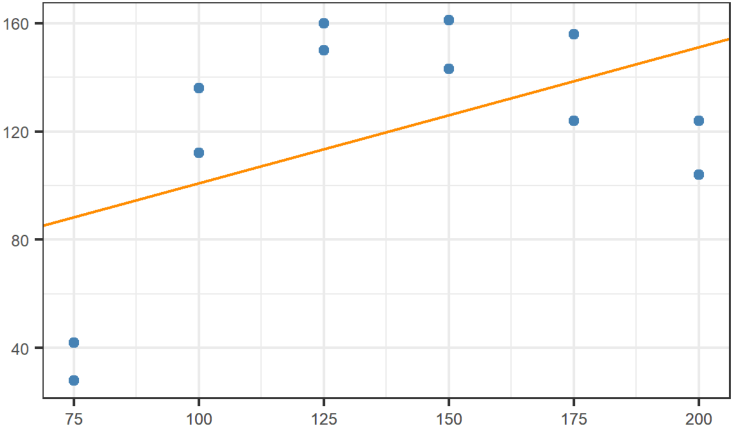
Test for Lack of Fit : Example

- 저축 예금자 자료

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

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		
계	$SST = 21,162.0$	11			

Test for Lack of Fit : Example

- 저축 예금자 자료

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

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

$$SSLF = SSE - SSPE = 15,630.6 - 1310.0 = 14,320.6$$

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

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

두 회귀모형의 검정

- 완전 모형 (full model) :

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \epsilon_{ij}$$

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

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

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

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

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

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

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

Example

■ 회귀모형 비교

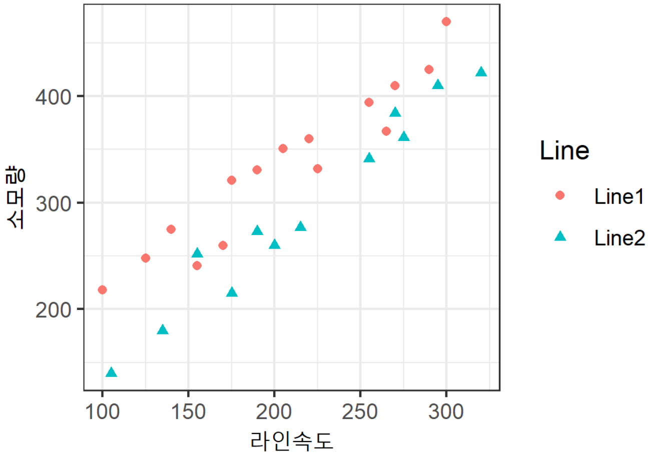
- (예제) 맥주를 생산하는 어느 맥주회사에 두 개의 생산라인(production line)이 있다. 이 라인을 움직이는 라인속도(line speed)와 하루 동안에 라인으로부터 흘러 나와서 못 쓰게 되는 맥주의 양 간에 는 관계가 있는 것으로 판명되었다. 그런데, 라인속도와 흘러 나오는 소모량 간의 관계가 생산라인이 다름에 따라 차이가 있는가를 알기 위해서 다음의 실험 자료를 얻었다. 두 회귀모형의 동일성 여부를검정하시오.

Example

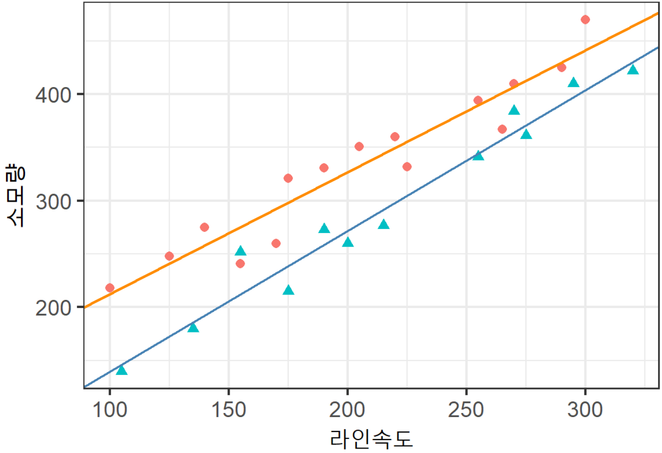
Table: 맥주생산라인의 자료

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			

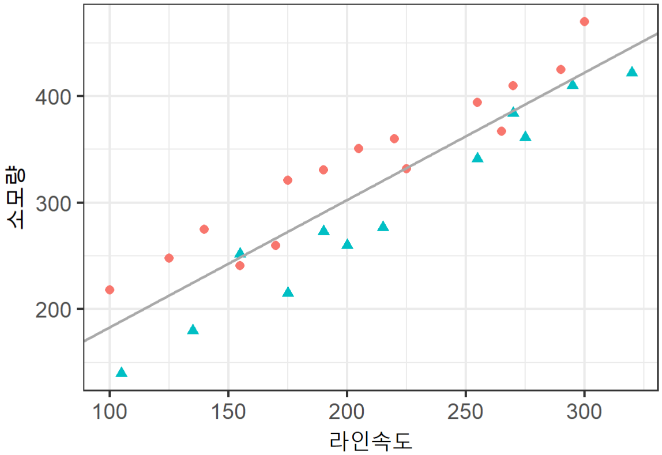
Example



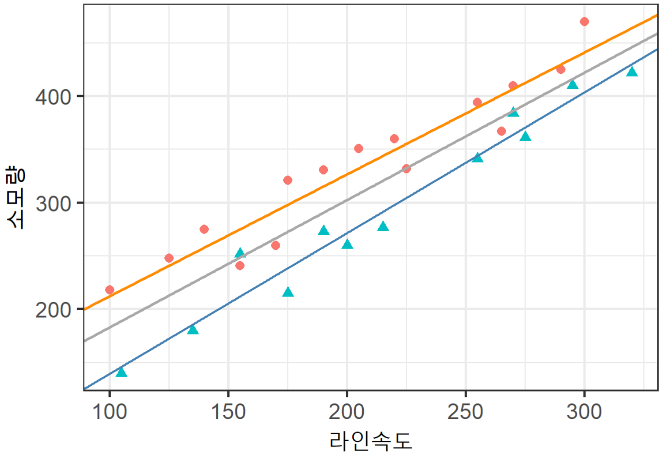
Example



Example



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 :

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

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

- 검정통계량

$$\begin{aligned} F_0 &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{9408 - 9904}{2} \div \frac{9904}{23} = 22.65 > F_{0.05}(2, 23) = 3.42 \end{aligned}$$

- 귀무가설 기각.

두 기울기의 비교

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

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

- 두 표본이 독립이라고 가정하면

$$\begin{aligned} \text{Var}(\hat{\beta}_{11} - \hat{\beta}_{12}) &= \text{Var}(\hat{\beta}_{11}) + \text{Var}(\hat{\beta}_{12}) \\ &= \frac{\sigma^2}{\sum (x_{1j} - \bar{x}_1)^2} + \frac{\sigma^2}{\sum (x_{2j} - \bar{x}_2)^2} \end{aligned}$$

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

Example

- $\hat{\beta}_{11} = 1.1454, \;\; \hat{\beta}_{12} = 1.3221$
- $SSE(F) = 9904, \;\; MSE(F) = \dfrac{SSE(F)}{df_F} = \dfrac{9904}{23} = 430.6$
- $\widehat{\mathrm{Var}}(\hat{\beta}_{11} - \hat{\beta}_{12}) = 430.6 \left[\dfrac{1}{53,693} + \dfrac{1}{50,192} \right] = 0.0166$
- 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

- 가중회귀모형, $i = 1, \ldots, n$

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

- 오차제곱합

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

- 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

- 정규방정식

$$\begin{cases} \hat{\beta}_0 \sum w_i + \hat{\beta}_1 \sum w_i x_i = \sum w_i y_i \\ \hat{\beta}_0 \sum w_i x_i + \hat{\beta}_1 \sum w_i x_i^2 = \sum w_i x_i y_i \end{cases}$$

- 가중최소제곱추정량 (WLSE)

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

- $\bar{x}_w, \bar{y}_w$: 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

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

$$\boldsymbol{y}^\top A\boldsymbol{y} = \sum_{i=1}^{n} a_{ij}y_iy_j = \sum_{i=1}^{n} a_{ii}{y_i}^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ii}y_iy_j$$

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

Quadratic form

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

Quadratic form : SST

$$\begin{aligned}\sum (y_i - \bar{y})^2 &= \sum y_i^2 - \frac{(\sum y_i)^2}{n} \\ &= \boldsymbol{y}^\top I_n \mathbf{y} - \frac{1}{n}\boldsymbol{y}^\top \mathbf{1}_n \mathbf{1}_n^\top \boldsymbol{y} = \boldsymbol{y}^\top \left(I_n - \frac{1}{n} J_n \right) \boldsymbol{y}\end{aligned}$$

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

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

- $\boldsymbol{y}^\top = (y_1, y_2, \cdots, y_n)$: 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

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

Multivariate normal distribution

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

- 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

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

distribution of quadratic forms

- 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_k는 대각 원소 중 k개가 1, 나머지는 0인 대각행렬을 의미한다.

⟨정리 1.31⟩ $\lambda_1, \ldots, \lambda_n$: $n \times n$ 행렬 A의 고유값

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

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

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

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

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

distribution of quadratic forms

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

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

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

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

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

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

distribution of quadratic forms

⟨정리 5.4⟩

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

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

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

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

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

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

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

distribution of quadratic forms

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

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

$$BVA = O_n$$

⟨정리 5.6 ⟩ $\boldsymbol{y} \sim N(\boldsymbol{\mu}, V)$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

$\langle$ 정리 5.7 $\rangle$ 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 rank$(A_i) = k_i$

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

⟨정리 5.7⟩ 계속

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

(a) A_iV는 모든 i에 대하여 멱등행렬이다.

(b) 모든 $i < j$에 대하여, $A_iVA_j = O_n$

(c) AV는 멱등행렬이다.

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

distribution of quadratic forms

⟨정리 5.8 ⟩ (코크란의 정리)

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

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

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

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

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

이다.