

6장 다변량 분산분석

덕성여자대학교 정보통계학과

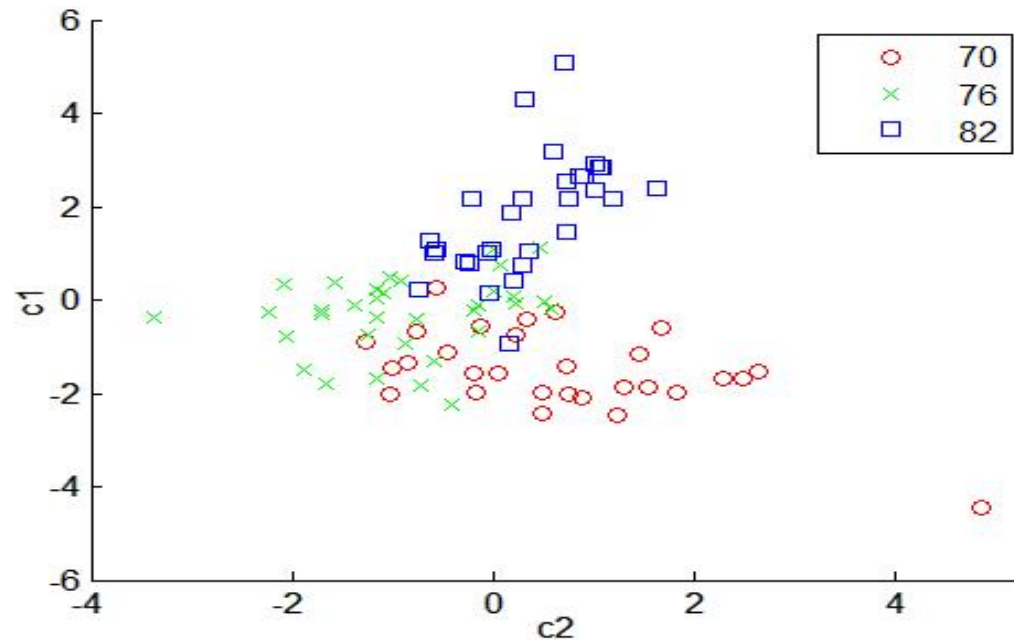
김 재 희



Copyright (c) 2008-2012 덕성여대 김재희 All rights reserved.

6.1 서론

- 한 개의 반응변수를 측정した場合 집단(처리, 그룹)간 평균비교=> 일변량 분산분석(ANOVA)
- 2개 이상의 반응변수를 측정した場合, 즉 관측값이 반응변수벡터로 주어지는 경우, 집단간 평균벡터에 대한 비교=> 다변량 분산분석(MANOVA: Multivariate ANalysis Of VAriance)



[문제 예]

(1) 소 품종(한우, Wagyu, Angus)에 대해 10개 등심 시료에서 아미노산측정.

품종에 따라 아미노산 평균 차이가 있는지 검정 => 일원배치 일변량 분산분석

(2) 소 품종(한우, Wagyu, Angus)에 대해 10개 등심 시료에서

(배장근면적, 아미노산) 측정.

품종에 따라 (배장근면적, 아미노산) 평균벡터 차이가 있는지 검정

=> 일원배치 다변량 분산분석

(3) 소 품종(한우, Wagyu, Angus) 각 9마리에 대해

각 3마리씩 사료 종류(건초, 생초, 곡류)를 먹인후

27개 등심 시료에서 아미노산측정.

품종과 사료종류에 따라 아미노산 평균 차이가 있는지 검정

=> 이원배치 일변량 분산분석

(4) 품종(한우, Wagyu, Angus) 각 9마리에 대해
각 3마리씩 사료 종류(건초, 생초, 곡류)를 먹인 후
27개 등심 시료에서 (배장근면적, 아미노산) 측정.

품종과 사료종류에 따라 (배장근면적, 아미노산) 평균벡터 차이가 있는지 검정
=> 이원배치 다변량 분산분석

6.2 일원배치 분산분석- 일변량의 경우

■ 분산이 동일한 일변량 정규분포를 따르는 g 개 집단에서의 각각 n_1, n_2, \dots, n_g 개의 확률표본에 대해 관측값이 얻어졌으며 이들은 서로 독립이라 가정한다.

$$\text{모집단 1} : X_{11}, X_{12}, \dots, X_{1, n_1} \sim iid N(\mu_1, \sigma^2)$$

$$\text{모집단 2} : X_{21}, X_{22}, \dots, X_{2, n_2} \sim iid N(\mu_2, \sigma^2)$$

$$\vdots$$

$$\text{모집단 } g : X_{g1}, X_{g2}, \dots, X_{g, n_g} \sim iid N(\mu_g, \sigma^2)$$

통계적 모형:

$$X_{lj} = \mu + \tau_l + \epsilon_{lj}, \quad l = 1, \dots, g, \quad j = 1, \dots, n_l$$

여기서 μ 는 전체평균,

$$\text{오차항 } \epsilon_{lj} \sim iid N(0, \sigma^2)$$

τ_l 는 l 번째 처리효과로 $\sum_{l=1}^g n_l \tau_l = 0$ 를 만족한다(고정효과).

데이터의 변동을 처리에 대한 변동과 오차에 의한 변동 분해

$$X_{lj} - \bar{X} = (\bar{X}_l - \bar{X}) + (X_{lj} - \bar{X}_l)$$

$$\sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X})^2 = \sum_{l=1}^g n_l (\bar{X}_l - \bar{X})^2 + \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X}_l)^2$$

총 제곱합 = (처리간변동제곱합) + (처리내변동제곱합)

$$TSS = SS_t + SS_e$$

[표 6.1] 일원배치 일변량 분산분석표

| 요인 | 제 곱 합 | 평균 제 곱 합 | F-통 계 량 | 자 유 도 |
|-----|--------|--|---------------------------|------------------------|
| 처 리 | SS_t | $MS_t = SS_t / (g - 1)$ | $F_t = \frac{MS_t}{MS_e}$ | $g - 1$ |
| 오 차 | SS_e | $MS_e = SS_e / (\sum_{i=1}^g n_i - g)$ | | $\sum_{i=1}^g n_i - g$ |
| 총 | TSS | | | $\sum_{i=1}^g n_i - 1$ |

(1) 통계적 가설

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0$ 에 대해 H_1 : 적어도 한 개의 $\tau_i \neq 0, i = 1, \cdots, g$

(2) 검정법 : 유의수준 α 에서

$F_t \geq F_{g-1, \sum n_i - g}(\alpha)$ 이면 H_0 를 기각한다.

6.3 일원배치 다변량 분산분석 (one-way MANOVA)

■ g 개 모집단에 대해 모평균벡터에 대한 비교를 위해.

공분산행렬이 같은 다변량 정규분포를 따르는 g 개 모집단에서

각각 n_1, n_2, \dots, n_g 개의 $p \times 1$ 관측벡터가 얻어졌으며 이들은 서로 독립이라 가정한다.

$$\text{모집단 1} : X_{11}, X_{12}, \dots, X_{1, n_1} \sim iid N_p(\mu_1, \Sigma)$$

$$\text{모집단 2} : X_{21}, X_{22}, \dots, X_{2, n_2} \sim iid N_p(\mu_2, \Sigma)$$

$$\vdots$$

$$\text{모집단 } g : X_{g1}, X_{g2}, \dots, X_{g, n_g} \sim iid N_p(\mu_g, \Sigma)$$

■ 통계적 모형: $X_{lj} = \mu + \tau_l + \epsilon_{lj}, \quad l = 1, \dots, g, \quad j = 1, \dots, n_l$

여기서 μ 는 전체평균,

오차벡터 $\epsilon_{lj} \sim iid N_p(0, \Sigma)$

τ_l 은 l 번째 처리효과이며 $\sum_{l=1}^g n_l \tau_l = 0$ 을 만족한다.

데이터의 변동을 처리에 대한 변동과 오차에 의한 변동으로 분해

$$X_{lj} - \bar{X} = (\bar{X}_l - \bar{X}) + (X_{lj} - \bar{X}_l)$$

제공합행렬:

$$\begin{aligned} \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X})(X_{lj} - \bar{X})' &= \sum_{l=1}^g n_l (\bar{X}_l - \bar{X})(\bar{X}_l - \bar{X})' + \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X}_l)(X_{lj} - \bar{X}_l)' \\ &= \quad \quad \quad B \quad \quad \quad + \quad \quad \quad E \\ &= (\text{처리간 변동 제공합행렬}) + (\text{처리내 변동 제공합행렬}) \end{aligned}$$

[표 6.2] 일원배치 다변량 분산분석표 ($N = \sum_{i=1}^g n_i$)

| 요인 | 제곱합과 교차곱 행렬 | 자유도 |
|----|---|------------------------|
| 처리 | $B = \sum_{l=1}^g n_l (\bar{X}_l - \bar{X})(\bar{X}_l - \bar{X})'$ | $g - 1$ |
| 오차 | $E = \sum_{l=1}^g \sum_{j=1}^{n_l} (X_{lj} - \bar{X}_l)(X_{lj} - \bar{X}_l)'$ | $\sum_{l=1}^g n_l - g$ |
| 총 | $B + E$ | $\sum_{l=1}^g n_l - 1$ |

(1) 통계적 가설

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0$ 에 대해 $H_1 : \text{적어도 한 개의 } \tau_l \neq 0, l = 1, \cdots, g$

(2) 검정법 : 유의수준 α 에서 Wilks lambda 통계량 $\Lambda = \frac{|E|}{|B + E|}$

$\Lambda \leq \Lambda_{p, g-1, N-g}(\alpha)$ 이면 H_0 를 기각한다.

6.4 많이 사용되는 4개의 통계량

행렬 $E^{-1}B$ 의 고유값을 순서대로 늘어놓은 것을 $\lambda_1 > \lambda_2 > \dots > \lambda_s$ 이라 할 때
여기서

p = 변수의 개수

$\nu_B = H_0$ 하에서 자유도(처리에 대한 자유도) = $g - 1$

ν_E = 오차에 대한 자유도 $s = \min(\nu_B, p)$

(1) Wilks lambda

$$\Lambda = \prod_{i=1}^s \frac{1}{1 + \lambda_i} = \frac{|E|}{|B + E|}$$

유의수준 α 에서 검정법은 $\Lambda \leq \Lambda_{p, g-1, \sum_{i=1}^g n_i - g}(\alpha)$ 이면 H_0 를 기각한다.

여기서 $\Lambda_{p, g-1, \sum_i n_i - g}(\alpha)$ 는 Wilks lambda 분포에서의 기각값.

(2) Pillai 검정

$$V^{(S)} = \sum_{i=1}^s \frac{\lambda_i}{1 + \lambda_i} = \text{tr}[(E+B)^{-1}B]$$

$V^{(S)}$ 가 클수록 H_0 를 기각한다. 유의수준 α 에서 검정법은

$$V^{(S)} \geq V_{m,N}(\alpha) \text{ 이면 } H_0 \text{ 를 기각한다.}$$

여기서 $m = \frac{1}{2}(|\nu_B - p| - 1), \quad N = \frac{1}{2}(\nu_E - p - 1)$

$V_{m,N}(\alpha)$ 는 Pillai 검정통계량의 분포에서 기각값

(3) Lawley - Hotelling 검정

Lawley - Hotelling 검정통계량은

$$U^{(S)} = \sum_{i=1}^s \lambda_i = \text{tr}(E^{-1}B)$$

로 정의되며 $U^{(S)}$ 가 클수록 H_0 를 기각한다. 유의수준 α 에서 검정법은

$$U^{(S)} \geq U_{\nu_B, \nu_E}(\alpha) \text{이면 } H_0 \text{ 를 기각한다.}$$

여기서 ν_B 는 처리에 대한 자유도이고, ν_E 는 오차에 대한 자유도이다.

$U_{\nu_B, \nu_E}(\alpha)$ 는 Lawley-Hotelling 검정통계량의 분포에서의 기각값으

(4) Roy의 최대근 검정

$$\theta = \frac{\lambda_1}{1 + \lambda_1}$$

이며 θ 가 클수록 H_0 를 기각한다. 유의수준 α 에서 검정법은

$\theta \geq R_{m,N}(\alpha)$ 이면 H_0 를 기각한다.

여기서 $m = \frac{1}{2}(|\nu_B - p| - 1)$, $N = \frac{1}{2}(\nu_E - p - 1)$

$R_{m,N}(\alpha)$ 는 Roy의 검정통계량 분포에서의 기각값

6.5 공분산행렬에 대한 동일성 검정

6.5.1 한 개 모집단의 경우

- 모집단이 기존의 공분산행렬 또는 특정한 공분산행렬을 갖는다고 할 수 있는지를 검정.
 p -변량 확률표본 $X_1, \dots, X_n \sim iid N_p(\mu, \Sigma)$ 을 얻었을 때,
- 관심 있는 가설

$$H_0 : \Sigma = \Sigma_0 \text{ 에 대해 } H_1 : \Sigma \neq \Sigma_0 .$$

- 자연로그 우도비를 이용하여 계산하면 검정통계량은

$$\Lambda^* = (n-1) \left\{ \ln |\Sigma_0| - \ln |\mathcal{S}_n| + \text{tr}(\mathcal{S}_n \Sigma_0^{-1}) - p \right\} .$$

귀무가설 H_0 하에서 근사적으로 Λ^* 는 자유도가 $p(p+1)/2$ 인 카이제곱분포 $\chi_{p(p+1)/2}^2$.

유의수준 α 에서 검정법은

$$\Lambda^* \geq \chi_{p(p+1)/2}^2(\alpha) \text{이면 } H_0 \text{를 기각한다.}$$

6.5.2 여러 개 모집단의 경우

- 일원배치 분산분석 문제에서 g 개 모집단이 있는 경우($g \geq 2$)를 평균벡터를 비교문제. 공분산행렬이 같고 다변량 정규분포를 따르는 g 개 집단에서 각각 n_1, n_2, \dots, n_g 개의 확률표본에 대해 $p \times 1$ 관측벡터가 얻어졌으며 이들은 서로 독립이라 한다.

$$\text{모집단 1 : } X_{11}, X_{12}, \dots, X_{1, n_1} \sim iid N_p(\mu_1, \Sigma_1)$$

$$\text{모집단 2 : } X_{21}, X_{22}, \dots, X_{2, n_2} \sim iid N_p(\mu_2, \Sigma_2)$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$\text{모집단 } g : X_{g1}, X_{g2}, \dots, X_{g, n_g} \sim iid N_p(\mu_g, \Sigma_g)$$

- 각 모집단의 공분산행렬이 같다고 할 수 있는지에 대한 검정.
관심 있는 가설:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g (\text{공분산행렬이 모두 같다})$$

$$H_1 : \text{적어도 한 개의 } \Sigma_i \text{ 는 서로 다르다.}$$

- 각각 n_1, n_2, \dots, n_g 개 씩의 서로 독립인 $N_p(\mu_i, \Sigma_i)$ 를 따르는 확률표본에 대해 우도비를 구하면

$$LR = \frac{|\hat{\Sigma}_1|^{\frac{n_1}{2}} |\hat{\Sigma}_2|^{\frac{n_2}{2}} \dots |\hat{\Sigma}_g|^{\frac{n_g}{2}}}{|\hat{\Sigma}_{pool}|^{\sum_{i=1}^g \frac{n_i}{2}}}$$

여기서

$$\hat{\Sigma}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})(X_{ij} - \bar{X}_{i.})' = \frac{1}{n_i} W_i$$

$$\hat{\Sigma}_{pool} = \frac{1}{\sum_{i=1}^g (n_i - 1)} \sum_{i=1}^g W_i$$

$\hat{\Sigma}_{pool}$ 는 $\hat{\Sigma}_i$ 들을 합동시켜 얻은 합동공분산행렬이다.

우도비 검정에 의해

$$\begin{aligned}
 M &= \frac{|S_1|^{\frac{\nu_1}{2}} |S_2|^{\frac{\nu_2}{2}} \dots |S_g|^{\frac{\nu_g}{2}}}{|S_{pl}|^{\sum_i \frac{n_i}{2}}} \\
 &= \left(\frac{|S_1|}{|S_{pl}|} \right)^{\frac{\nu_1}{2}} \left(\frac{|S_2|}{|S_{pl}|} \right)^{\frac{\nu_2}{2}} \dots \left(\frac{|S_g|}{|S_{pl}|} \right)^{\frac{\nu_g}{2}}
 \end{aligned}$$

여기서 $\nu_i = n_i - 1$, $S_i = \frac{W_i}{\nu_i}$

$$S_{pl} = \frac{\sum_{i=1}^g \nu_i S_i}{\sum_{i=1}^g \nu_i} \quad (\nu_i > p),$$

만약 $S_1 = S_2 = \dots = S_g = S_{pl}$ 이면 $M=1$ 이 되고, S_i 들 간에 차이가 클수록 0에 가까워진다.

■ Box(1949, 1950)의 M -검정법

H_0 하에서 $n_i \rightarrow \infty$ 일 때

$$u = -2(1 - c_1) \ln M$$

는 근사적으로 $\chi^2_{(g-1)p(p+1)/2}$ 분포를 따른다.

여기서

$$c_1 = \left[\sum_{i=1}^g \frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^g \nu_i} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right]$$

유의수준 α 에서 Box의 M -검정법은

$$u \geq \chi^2_{(g-1)p(p+1)/2}(\alpha) \text{이면 } H_0 \text{를 기각한다.}$$

《예 6.5》 [표 6.4] 데이터에 대해 3개 그룹의 공분산행렬이 같은지 검정하고자한다.

각 그룹의 표본의 크기는 $n_1 = n_2 = n_3 = 8$ 이며 각 표본공분산행렬은

$$\mathbf{S}_1 = \begin{pmatrix} 0.339 & 0.202 \\ 0.202 & 0.199 \end{pmatrix}, \quad \mathbf{S}_2 = \begin{pmatrix} 0.342 & 0.259 \\ 0.259 & 0.307 \end{pmatrix}, \quad \mathbf{S}_3 = \begin{pmatrix} 0.685 & 0.314 \\ 0.314 & 0.154 \end{pmatrix}.$$

합동공분산행렬:

$$\mathbf{S}_{pl} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + (n_3 - 1)\mathbf{S}_3}{n_1 + n_2 + n_3 - 3} = \frac{7(\mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3)}{21} = \frac{1}{3} \begin{pmatrix} 1.366 & 0.775 \\ 0.775 & 0.66 \end{pmatrix}$$

$$|\mathbf{S}_1| = 0.027, \quad |\mathbf{S}_2| = 0.038, \quad |\mathbf{S}_3| = 0.007, \quad |\mathbf{S}_{pl}| = 0.1000 \text{이므로}$$

$$\begin{aligned}
M &= \left(\frac{|S_1|}{|S_{pl}|} \right)^{\frac{\nu_1}{2}} \left(\frac{|S_2|}{|S_{pl}|} \right)^{\frac{\nu_2}{2}} \left(\frac{|S_3|}{|S_{pl}|} \right)^{\frac{\nu_3}{2}} \\
&= \left(\frac{0.027}{0.1} \right)^{7/2} \cdot \left(\frac{0.038}{0.1} \right)^{7/2} \cdot \left(\frac{0.007}{0.1} \right)^{7/2} = 0.000000031
\end{aligned}$$

$$c_1 = \left[\sum_{i=1}^3 \frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^3 \nu_i} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(3-1)} \right] = \left(\frac{3}{7} - \frac{1}{21} \right) \cdot 0.361 = 0.1375$$

$u = -2(1 - c_1) \ln M = 29.8136$ 으로 계산된다.

유의수준 $\alpha = 0.05$ 에서 Box의 M -검정법은 $u = 29.8136 \geq \chi_4^2(0.05) = 9.49$ 이므로 H_0 를 기각한다.

그러므로 세 집단의 공분산행렬에 통계적으로 유의한 차이가 있다고 할 수 있다.

6.6 이원배치 다변량 분산분석

반응에 영향을 미치는 요인이 두 개인 경우의 평균벡터 비교에 대한 문제.
요인이 두 개이므로 두 요인간의 상호작용을 고려한다.

■ 이원배치 다변량 분산분석(two-way MANOVA)에 대한 통계적 모형:

$$\mathbf{X}_{lkr} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_{lk} + \mathbf{e}_{lkr}, \quad l = 1, 2, \dots, g, \quad k = 1, 2, \dots, b, \quad r = 1, 2, \dots, n$$

여기서

$\boldsymbol{\tau}_l$ 는 첫 번째 요인1 효과

$\boldsymbol{\beta}_k$ 는 두 번째 요인2 효과

$\boldsymbol{\gamma}_{lk}$ 는 요인1과 요인2의 상호작용효과(interaction effect)

$$\sum_{l=1}^g \boldsymbol{\tau}_l = \sum_{k=1}^b \boldsymbol{\beta}_k = \sum_{l=1}^g \boldsymbol{\gamma}_{lk} = \sum_{k=1}^b \boldsymbol{\gamma}_{lk} = \mathbf{0}$$

오차항 $\mathbf{e}_{lkr} \sim iid N_p(\mathbf{0}, \boldsymbol{\Sigma})$ 인 확률 변수이다.

- 관측벡터 X_{lkr} 을 요인에 의한 편차로 분해:

$$X_{lkr} - \bar{X} = (\bar{X}_{l..} - \bar{X}) + (\bar{X}_{.k.} - \bar{X}) + (\bar{X}_{lk.} - \bar{X}_{l..} - \bar{X}_{.k.} + \bar{X}) + (X_{lkr} - \bar{X}_{lk.})$$

요인1 효과
요인 2 효과
상호작용효과
처리내 잔차

- 변동을 제곱합과 교차제곱합으로 분해

$$\begin{aligned}
& \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (X_{lkr} - \bar{X})(X_{lkr} - \bar{X})' \\
&= \sum_{l=1}^g bn (\bar{X}_{l..} - \bar{X})(\bar{X}_{l..} - \bar{X})' \\
&+ \sum_{k=1}^b gn (\bar{X}_{.k.} - \bar{X})(\bar{X}_{.k.} - \bar{X})' \\
&+ \sum_{l=1}^g \sum_{k=1}^b n (\bar{X}_{lk.} - \bar{X}_{l..} - \bar{X}_{.k.} + \bar{X})(\bar{X}_{lk.} - \bar{X}_{l..} - \bar{X}_{.k.} + \bar{X})' \\
&+ \sum_{l=1}^g (\bar{X}_{lkr} - \bar{X}_{lk.})(\bar{X}_{lkr} - \bar{X}_{lk.})' \\
&= SSP_A + SSP_B + SSP_{int} + SSP_E
\end{aligned}$$

여기서

\bar{X} 는 전체평균

$\bar{X}_{l..}$ 는 요인1의 l 번째 수준에서 관측벡터의 평균

$\bar{X}_{.k.}$ 는 요인2의 k 번째 수준에서 관측벡터의 평균

$\bar{X}_{lk.}$ 는 요인1의 l 번째 수준과 요인2의 k 번째 수준에서 관측벡터의 평균

► 표 6.6 이원배치 다변량 분산분석표

| 요인 | 제곱합과 교차곱행렬 | 자유도 |
|------|-------------|--------------|
| 요인1 | SSP_A | $g-1$ |
| 요인2 | SSP_B | $b-1$ |
| 상호작용 | SSP_{int} | $(g-1)(b-1)$ |
| 오차 | SSP_E | $gb(n-1)$ |
| 총 | SSP_{cor} | $gbn-1$ |

전체평균에 의해 수정된(corrected) 제곱합행렬 :

$$SSP_{cor} = \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (X_{lkr} - \bar{X})(X_{lkr} - \bar{X})'$$

요인1에 의한 변동을 나타내는 제곱합과 교차곱행렬:

$$SSP_A = \sum_{l=1}^g bn(\bar{X}_{l..} - \bar{X})(\bar{X}_{l..} - \bar{X})'$$

요인2에 의한 변동을 나타내는 제곱합과 교차곱행렬:

$$SSP_B = \sum_{k=1}^b gn(\bar{X}_{.k.} - \bar{X})(\bar{X}_{.k.} - \bar{X})'$$

요인1과 요인2에 의한 상호작용에 대한 변동을 나타내는 제곱합과 교차곱행렬

$$SSP_{int} = \sum_{l=1}^g \sum_{k=1}^b n(\bar{X}_{lk.} - \bar{X}_{l..} - \bar{X}_{.k.} + \bar{X}) \times (\bar{X}_{lk.} - \bar{X}_{l..} - \bar{X}_{.k.} + \bar{X})'$$

잔차 변동을 나타내는 제곱합과 교차곱행렬:

$$SSP_E = \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (X_{lkr} - \bar{X}_{lk.})(X_{lkr} - \bar{X}_{lk.})'$$

■ 관심있는 가설들에 대해 검정을 해보자.

(1) 요인1의 처리효과가 동일한지에 대한 검정

$H_0 : \tau_1 = \tau_2 = \dots = \tau_g = 0$ 에 대해 $H_1 : \text{적어도 한 개의 } \tau_l \neq 0 \text{ 이다.}$

에 대해 검정통계량

$$\Lambda_1 = \frac{|SSP_E|}{|SSP_A + SSP_E|}$$

을 사용한다. 검정법은 Λ_1 이 작을수록 H_0 를 기각한다.

유의수준 α 에서 검정법으로 $\Lambda_1 \leq \Lambda_{p, g-1, gb(n-1)}(\alpha)$ 이면 H_0 를 기각한다.

근사적으로, 유의수준 α 에서

$$-\left[gb(n-1) - \frac{p+1-(g-1)}{2} \right] \ln \Lambda_1 \geq \chi^2_{(g-1)p}(\alpha) \quad \text{이면 } H_0 \text{를 기각한다.}$$

(2) 요인2의 처리효과가 동일한지에 대한 검정

$H_0 : \beta_1 = \beta_2 = \dots = \beta_g = 0$ 에 대해 H_1 : 적어도 한 개의 $\beta_j \neq 0$ 이다.

에 대해 검정통계량

$$\Lambda_2 = \frac{|SSP_E|}{|SSP_B + SSP_E|}$$

을 사용한다. 검정법은 Λ_2 가 작을수록 H_0 를 기각한다.

유의수준 α 에서 검정법으로 $\Lambda_2 \leq \Lambda_{p,b-1,gb(n-1)}(\alpha)$ 이면 H_0 를 기각한다.

근사적으로, 유의수준 α 에서

$$-\left[gb(n-1) - \frac{p+1-(b-1)}{2}\right] \log \Lambda_2 \geq \chi^2_{(b-1)p}(\alpha) \quad \text{이면 } H_0 \text{를 기각한다.}$$

(3) 요인1과 요인2의 상호작용효과가 동일한지에 대한 검정

$$H_0 : \gamma_{11} = \gamma_{12} = \dots = \gamma_{gb} = 0 \quad (\text{상호작용 효과 없다})$$

에 대해 H_1 : 적어도 한개의 $\gamma_{lk} \neq 0$ 이다.

의 가설에 대하여 검정통계량

$$\Lambda_{12} = \frac{|SSP_E|}{|SSP_{int} + SSP_E|}$$

을 사용한다. 검정법으로 Λ_{12} 가 작을수록 H_0 를 기각한다.

유의수준 α 에서 $\Lambda_{12} \leq \Lambda_{p, (g-1)(b-1), gb(n-1)}(\alpha)$ 이면 H_0 를 기각한다.

근사적으로, 유의수준 α 에서

$$-\left[gb(n-1) - \frac{p+1-(g-1)(b-1)}{2} \right] \ln \Lambda_{12} \geq \chi^2_{(g-1)(b-1)p}(\alpha) \text{이면 } H_0 \text{를 기각한다.}$$

《예제 6.6》 회전 속도(rotational velocity)와 윤활유(lubricant)의 2개 요인이 대해 4번 반복한 2×4 요인 실험 설계로 실험하여 철강조각에 대해 힘(torque)과 긴장력(strain)을 측정하여 [표 6.7]과 같은 데이터를 얻었다. 각 요인의 효과와 상호작용효과에 대해 검정하고자한다.

▶ 표 6.7 회전속도와 윤활유에 따른 철강 데이터

| 회전속도 | 윤활유 | | | | | | | |
|------|-------|-------|-------|-------|-------|-------|-------|-------|
| | B1 | | B2 | | B3 | | B4 | |
| | X_1 | X_2 | X_1 | X_2 | X_1 | X_2 | X_1 | X_2 |
| A1 | 7.80 | 90.4 | 9.00 | 82.5 | 7.28 | 79.6 | 7.60 | 94.1 |
| | 7.10 | 88.9 | 8.43 | 92.4 | 8.96 | 95.1 | 7.00 | 86.6 |
| | 7.89 | 85.9 | 7.65 | 82.4 | 7.75 | 90.2 | 7.82 | 85.9 |
| | 7.82 | 88.8 | 7.70 | 87.4 | 7.80 | 88.0 | 7.80 | 88.8 |
| A2 | 7.12 | 85.1 | 8.19 | 66.0 | 7.15 | 81.2 | 7.06 | 81.2 |
| | 7.06 | 89.0 | 8.25 | 74.5 | 7.15 | 72.0 | 7.04 | 79.9 |
| | 7.45 | 75.9 | 7.45 | 83.1 | 7.70 | 79.9 | 7.52 | 86.4 |
| | 7.45 | 77.9 | 7.45 | 86.4 | 7.45 | 71.9 | 7.70 | 76.4 |

통계적 모형:

$$X_{lkr} = \mu + \tau_l + \beta_k + \gamma_{lk} + e_{lkr}$$

$$l = 1, 2, \quad k = 1, 2, 3, 4, \quad r = 1, 2, 3, 4$$

이고 $g = 2, \quad b = 4, \quad n = 4$ 이다.

여기서

τ_l 은 첫 번째 요인인 회전속도 효과

β_k 는 두 번째 요인인 윤활유 효과

γ_{lk} 는 회전속도와 윤활유에 따른 상호작용효과

$$\sum_{l=1}^2 \tau_l = \sum_{k=1}^4 \beta_k = \sum_{l=1}^2 \gamma_{lk} = \sum_{k=1}^4 \gamma_{lk} = 0$$

오차항 $e_{lkr} \sim iid \ N_2(0, \Sigma)$ 인 확률변수이다.

▶ 표 6.8 철강데이터에 대한 이원배치 다변량 분산분석표

| 요인 | 제 곱 합과 교차 곱 행렬 | 자유도 |
|-----------|---|----------------------|
| 요인1(회전속도) | $SSP_A = \begin{pmatrix} 1.205 & 27.208 \\ 27.208 & 614.251 \end{pmatrix}$ | $g - 1 = 1$ |
| 요인2(윤활유) | $SSP_B = \begin{pmatrix} 1.694 & -9.862 \\ -9.862 & 74.874 \end{pmatrix}$ | $b - 1 = 3$ |
| 상호작용 | $SSP_{int} = \begin{pmatrix} 0.132 & 1.585 \\ 1.585 & 32.244 \end{pmatrix}$ | $(g - 1)(b - 1) = 3$ |
| 오차 | $SSP_E = \begin{pmatrix} 4.897 & -1.890 \\ -1.890 & 736.390 \end{pmatrix}$ | $gb(n - 1) = 24$ |
| 총 | $SSP_{cor} = \begin{pmatrix} 7.928 & 17.041 \\ 17.041 & 1457.759 \end{pmatrix}$ | $gbn - 1 = 31$ |

■ 이제 관심있는 가설들에 대해 검정을 해보자.

(1) 요인1(회전속도)의 효과가 동일한지에 대한 검정

$$H_0 : \tau_1 = \tau_2 = 0 \quad \text{에 대해} \quad H_1 : \tau_1 \neq \tau_2$$

검정통계량은

$$\begin{aligned} \Lambda_1 &= \frac{|SSP_E|}{|SSP_A + SSP_E|} = \frac{\left| \begin{pmatrix} 4.897 & -1.890 \\ -1.890 & 736.390 \end{pmatrix} \right|}{\left| \begin{pmatrix} 6.102 & 25.318 \\ 25.318 & 1350.641 \end{pmatrix} \right|} \\ &= \frac{4.897(736.390) - (1.890)^2}{6.102(1350.641) - 25.318^2} = \frac{3602.5397}{7600.6103} = 0.474 \end{aligned}$$

유의수준 $\alpha = 0.05$ 에서 $\Lambda_1 = 0.474 \leq \Lambda_{2,1,24}(0.05) = 0.771$ 이므로 H_0 를 기각한다.

그러므로 회전 속도가 철강의 반응에 유의한 영향을 미친다.

(2) 요인2(윤희유)의 효과가 동일한지에 대한 검정

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ 에 대해 H_1 : 적어도 한개의 $\beta_k \neq 0$ 이다.

검정통계량은

$$\begin{aligned}\Lambda_2 &= \frac{|SSP_E|}{|SSP_B + SSP_E|} = \frac{3602.5397}{\left| \begin{pmatrix} 6.591 & -11.752 \\ -11.752 & 811.264 \end{pmatrix} \right|} \\ &= \frac{3602.5397}{6.591(811.264) - 11.752^2} = \frac{3602.5397}{5208.9315} = 0.691\end{aligned}$$

유의수준 $\alpha = 0.05$ 에서 기각값은 $\Lambda_{p, b-1, gb(n-1)}(\alpha) = \Lambda_{2, 3, 24}(0.05) = 0.591$.

$\Lambda_2 = 0.685 \not\leq \Lambda_{2, 3, 24}(0.05) = 0.591$ 이므로 H_0 를 기각할 수 없다.

즉 윤희유간의 차이는 유의하지 않다.

(3) 요인1과 요인2의 상호작용효과가 존재하는지에 대한 검정

$$H_0 : \gamma_{11} = \gamma_{12} = \dots = \gamma_{24} = 0 \quad (\text{상호작용 효과 없다})$$

에 대해 H_1 : 적어도 한개의 $\gamma_{lk} \neq 0$ 이다.

검정통계량은

$$\begin{aligned} \Lambda_{12} &= \frac{|SSP_E|}{|SSP_{int} + SSP_E|} = \frac{3602.5397}{\left| \begin{pmatrix} 5.029 & -0.305 \\ -0.305 & 768.634 \end{pmatrix} \right|} \\ &= \frac{3602.5397}{5.029(768.634) - 0.305^2} = \frac{3602.5397}{3865.3673} = 0.932, \end{aligned}$$

유의수준 $\alpha = 0.05$ 에서 기각값은 $\Lambda_{2,3,24}(0.05) = 0.591$ 이고

$\Lambda_{12} = 0.932 \not\leq \Lambda_{2,3,24}(0.05) = 0.591$ 이므로 H_0 를 기각할 수 없다.

즉 회전속도와 윤활유간의 상호작용효과는 유의하지 않다.

6.7 R을 이용한 다변량분산분석

《예제 6.7》 3 그룹의 토끼에 대해 결핵균을 투입하여 몸 상태에 따른 결핵균 성장에 관련한 측정값으로 [표 6.6]의 자료를 얻었다. 각 그룹은

G_1 : 대조 그룹(control)

G_2 : 신진 대사 기능 저하 그룹(metabolic depression)

G_3 : 신진 대사 활성 그룹(heightened metabolic activity)

이며 측정 변수는

X_1 = 결핵당 간상균 수(number of bacilli)

X_2 = 결핵 크기(tubercle size in milimeters)

▶ 표 6.9 토끼의 결핵에 대한 신진대사 효과 비교자료

| G_1 | | G_2 | | G_3 | |
|-------|-------|-------|-------|-------|-------|
| X_1 | X_2 | X_1 | X_2 | X_1 | X_2 |
| 24.0 | 3.5 | 7.4 | 3.5 | 16.4 | 3.2 |
| 13.3 | 3.5 | 13.2 | 3.0 | 24.0 | 2.5 |
| 12.2 | 4.0 | 8.5 | 3.0 | 53.0 | 1.5 |
| 14.0 | 4.0 | 10.1 | 3.0 | 32.7 | 2.6 |
| 22.2 | 3.6 | 9.3 | 2.0 | 42.8 | 2.0 |
| 16.1 | 4.3 | 8.5 | 2.5 | | |
| 27.9 | 5.2 | 4.3 | 1.5 | | |

● aov() 함수를 이용하여 일변량 분산분석, manova() 함수를 이용하여 다변량 분산분석을 할 수 있으며 plotmeans() 함수를 사용하여 그룹별 평균 그림을 그릴 수 있다.

그룹을 나타내는 변수는 factor() 함수를 사용하여 범주형 변수로 정해 놓은 후 분산분석을 한다.

[프로그램 6.1] 일변량 일원배치 분산분석

```
rabbit=read.csv("C:/data/ header=T)
rabbit ; attach(rabbit)
group <- factor(group) # factor 지정 필요

tapply(x1, group, mean) # groupwise mean for x1
tapply(x1, group, var)  # groupwise variance
tapply(x2, group, mean) # groupwise mean for x2
tapply(x2, group, var)  # groupwise variance

## oneway ANOVA
ax1=aov(x1~group)
summary(ax1)          # display Type I ANOVA table

TukeyHSD(ax1)          # Tukey Honestly Significant Differences
pairwise.t.test(x1, group) # pairwise multiple comparison
```

```
# Plot Means with Error Bars
library(gplots)
group <- factor(group)
plotmeans(x1~group,xlab="█ylab="x1",
          main="Mean Plot \n with 95% CI")
plot(ax1, which=1)  # Residual plot
plot(ax1, which=2)  # Normal Q-Q plot for residuals

ax2=aov(x2~group) ; summary(ax2) ; TukeyHSD(ax2)
```

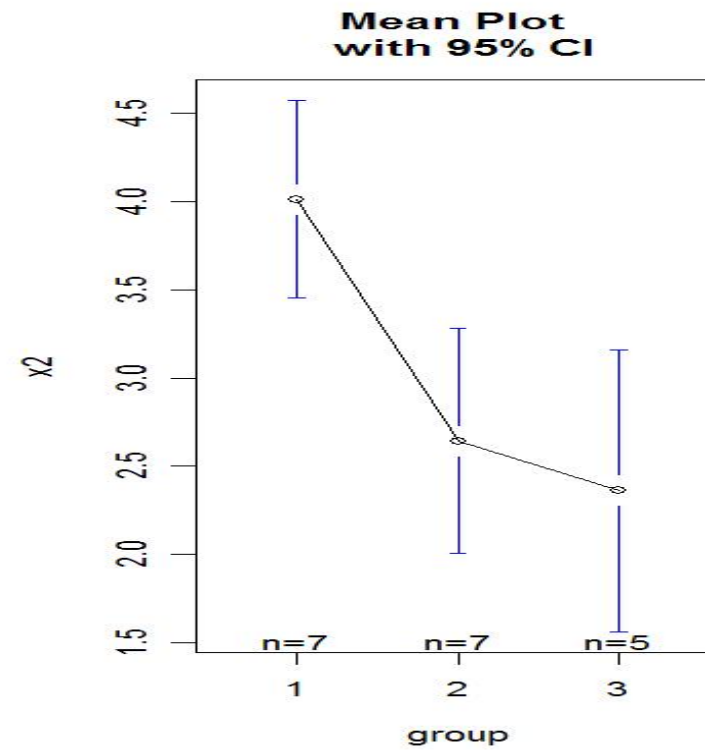
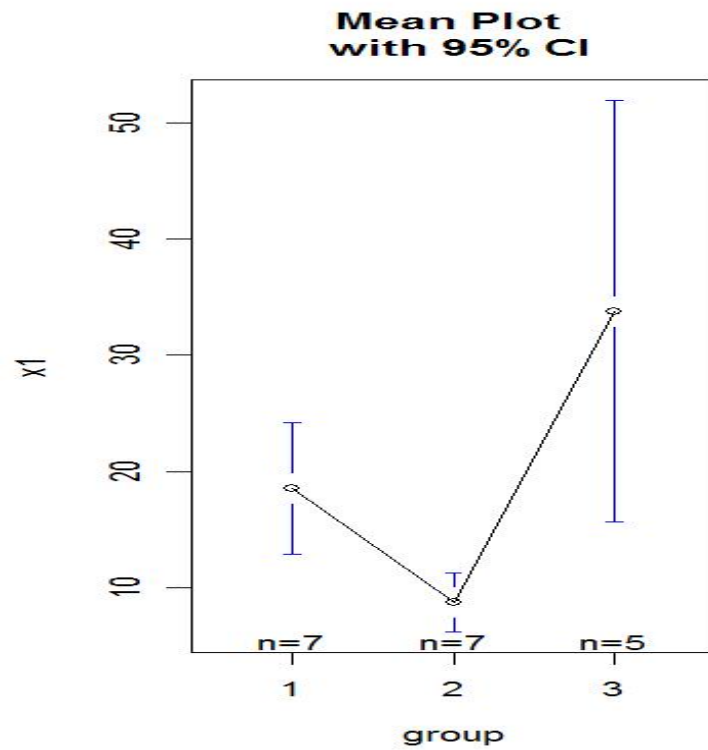
▶ 표 6.7 그룹별 평균과 분산

```
> group = factor(group) # factor 지정 필요
> tapply(x1, group, mean) # groupwise mean for x1
      1      2      3
18.528571  8.757143 33.780000

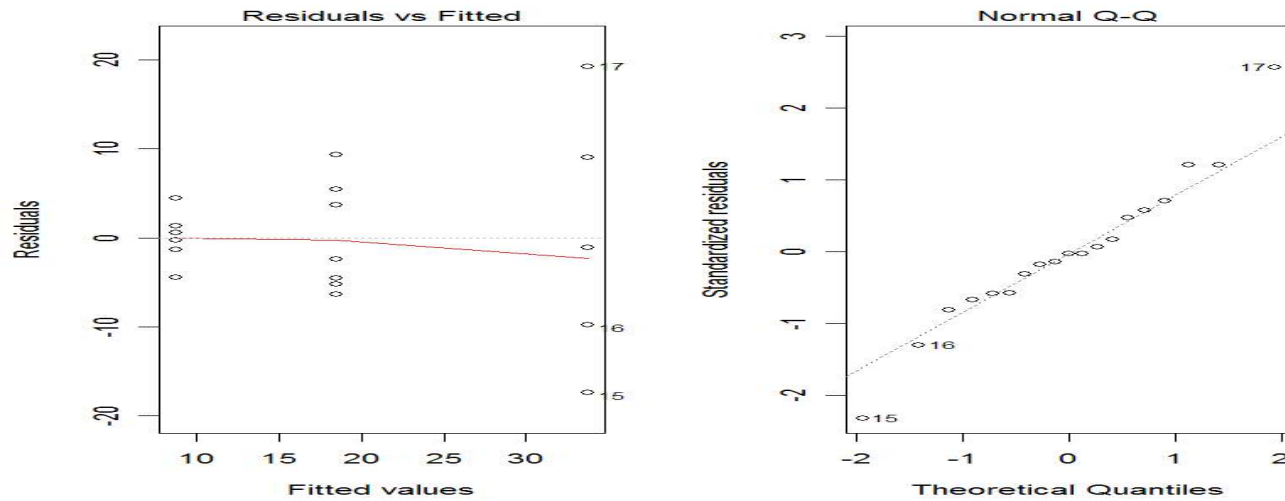
> tapply(x1, group, var) # groupwise variance
      1      2      3
37.505714  7.279524 212.412000

> tapply(x2, group, mean) # groupwise mean for x2
      1      2      3
4.014286 2.642857 2.360000

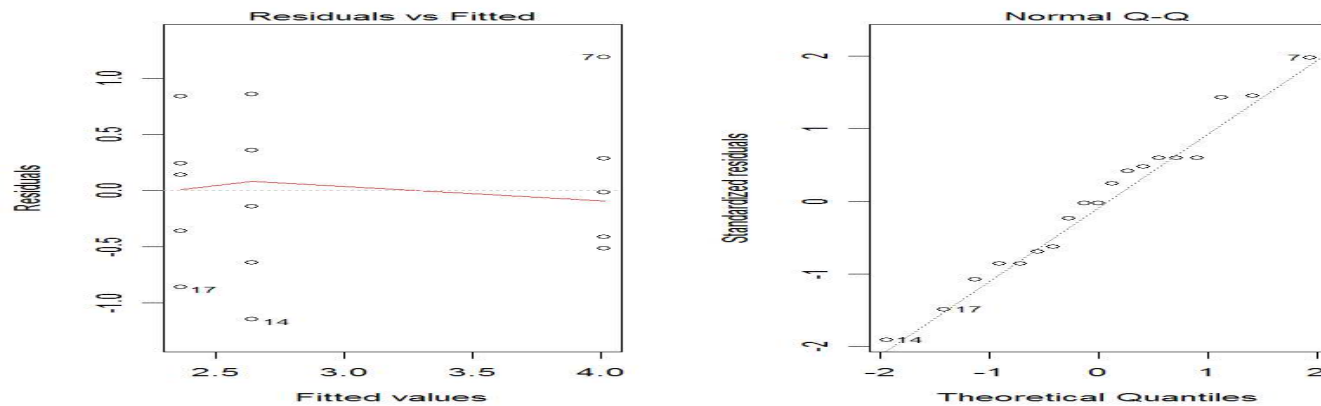
> tapply(x2, group, var) # groupwise variance
      1      2      3
0.3647619 0.4761905 0.4130000
```

[그림 6.2] 그룹별 x1과 x2 에 대한 평균 그림



[그림 6.3](1) x_1 변수에 대한 잔차 그림과 잔차에 대한 정규성 Q-Q 그림



[그림 6.3](2) x_2 변수에 대한 잔차 그림과 잔차에 대한 정규성 Q-Q 그림

[결과 6.1] 일변량 분산분석 결과

```
> ax1=aov(x1~group)
> summary(ax1)          # display Type I ANOVA table
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|---------------|
| group | 2 | 1828.15 | 914.07 | 13.077 | 0.0004307 *** |
| Residuals | 16 | 1118.36 | 69.90 | | |

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> TukeyHSD(ax1) # where fit ax1 comes from aov()
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = x1 ~ group)
$group
```

| | diff | lwr | upr | p adj |
|-----|-----------|------------|-----------|-----------|
| 2-1 | -9.771429 | -21.302564 | 1.759707 | 0.1041078 |
| 3-1 | 15.251429 | 2.619702 | 27.883155 | 0.0173045 |
| 3-2 | 25.022857 | 12.391131 | 37.654583 | 0.0002918 |

```

> ax2=aov(x2~group) ; summary(ax2)
              Df  Sum Sq Mean Sq F value    Pr(>F)
group          2 10.0391   5.0196  11.991 0.0006577 ***
Residuals     16   6.6977   0.4186
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(ax2)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = x2 ~ group)
$group
              diff          lwr          upr          p adj
2-1 -1.3714286 -2.263798 -0.4790591 0.0030071
3-1 -1.6542857 -2.631827 -0.6767440 0.0013155
3-2 -0.2828571 -1.260399  0.6946846 0.7399202

```

[프로그램 6.2] 다변량 일원배치 분산분석

```
# groupwise mean vector and covariance matrix
mean( rabbit[(group==1), 2:3])
cov( rabbit[(group==1),2:3])
mean( rabbit[(group==2), 2:3])
cov( rabbit[(group==2),2:3])
mean( rabbit[(group==3), 2:3])
cov( rabbit[(group==3),2:3])

group.num=as.numeric(group)
plot( x1,x2, pch=group.num, main="scatterplot with specified group")
  text( x1,x2, labels=group.num, adj=0, cex=1)  # 그림 6.4

## oneway MANOVA with 2 Dependent Variables ##
group <- factor(group)  # factor 지정 필요
y=cbind(x1,x2)          # response vector
fit=manova( y ~ group)
summary(fit, test="Wilks")      # Wilks' lambda
summary(fit, test="Pillai")     # Pillai's trace
summary(fit, test="Roy")       # Roy's greatest root
summary(fit, test="Hotelling") # Hotelling-Lawley trace
```

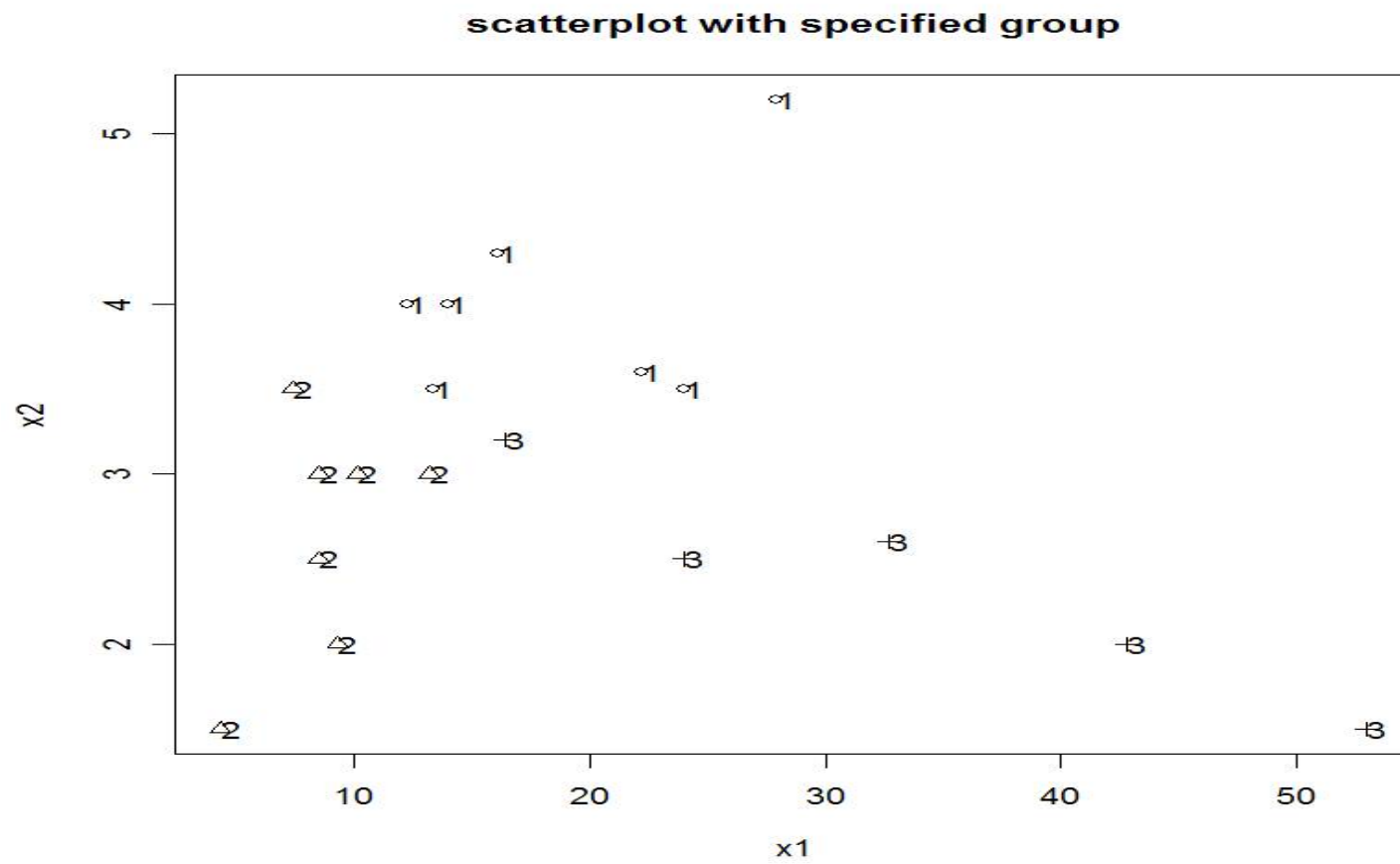
```

# Anova SSCP matrix
n=length(x1)          # number of samples
E=(n-1)* cov(fit$residuals)  # SSCP matrix for error
E
B=(n-1)* cov(fit$fitted.values) # SSCP matrix for treatments
B
test=solve(E)%*%B      #  $E^{-1}B$ 
eigen(test)

```

▶ 표 6.9 그룹별 평균벡터와 공분산 행렬

```
> # groupwise mean vector and covariance matrix
> mean( rabbit[(group==1), 2:3])
      x1      x2
18.528571  4.014286
> cov( rabbit[(group==1), 2:3])
      x1      x2
x1 37.505714 1.4878571
x2  1.487857 0.3647619
> mean( rabbit[(group==2), 2:3])
      x1      x2
8.757143 2.642857
> cov( rabbit[(group==2), 2:3])
      x1      x2
x1 7.2795238 0.9321429
x2 0.9321429 0.4761905
> mean( rabbit[(group==3), 2:3])
      x1      x2
33.78  2.36
> cov( rabbit[(group==3), 2:3])
      x1      x2
x1 212.412 -9.001
x2 -9.001  0.413
```



[그림 6.4] 그룹이 표시된 산점도

[결과 6.2] 일원배치 다변량 분산분석 결과

```
> fit=manova( y ~ group)

> summary(fit, test="Wilks")      # Wilks' lambda
      Df  Wilks approx F num Df den Df    Pr(>F)
group    2  0.149   11.927     4    30 6.44e-06 ***
Residuals 16
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(fit, test="Pillai")    # Pillai's trace
      Df  Pillai approx F num Df den Df    Pr(>F)
group    2  1.2269   12.6965     4    32 2.686e-06 ***
Residuals 16
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

> summary(fit, test="Roy")          # Roy's greatest root
      Df      Roy approx F num Df den Df      Pr(>F)
group    2  1.7215  13.7717      2    16 0.0003323 ***
Residuals 16
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(fit, test="Hotelling") # Hotelling-Lawley trace
      Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
group    2           3.1868  11.1537      4    28 1.542e-05 ***
Residuals 16
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.

```

▶ 표 6.10 오차제곱합행렬 E 와 처리제곱합행렬 B

```

> # Anova SSCP matrix
> n=length(x1)      # number of samples
> E=(n-1)* cov(fit$residuals)    # SSCP matrix for error
> E
      x1      x2
x1 1118.359 -21.484000
x2 -21.484   6.697714
> B=(n-1)* cov(fit$fitted.values) # SSCP matrix for treatments
> B
      x1      x2
x1 1828.14689 -24.95495
x2 -24.95495  10.03913
> test=solve(E)%*%B      #  $E^{-1}B$ 
> eigen(test)
$values
[1] 1.721459 1.465316
$vectors
      [,1]      [,2]
[1,] 0.1229872 -0.03443585
[2,] 0.9924083  0.99940691

```

《예제 6.8》 [표 6.4] 데이터에 대해 3개 그룹의 공분산행렬이 같은지 R 프로그램을 활용하여 검정하고자 한다.

[프로그램 6.3]은 일원배치 다변량 데이터에 대해 공분산행렬에 대한 동일성 검정 프로그램이며 아랫 부분에는 MANOVA 검정 프로그램도 보여준다. [결과 6.3]은 Box의 M 검정법에 대한 검정통계량 값과 p-값을 보여준다.

[프로그램 6.3] 다변량 일원배치 공분산행렬에 대한 동일성 검정

```
rootstock=read.csv("C:/data/roo[REDACTED]header=T)
attach(rootstock)
p=2    # number of variables
g=3    # number of groups

g1= rootstock[(rootstock$group==1),[REDACTED]]
s1= cov(g1) ;    s1=round(s1, digits=3)          # group1 cov
g2= rootstock[(rootstock$group==2),[REDACTED]]
s2= cov(g2) ;    s2=round(s2, digits=3)          # group2 cov
g3= rootstock[(rootstock$group==3),[REDACTED]]
s3= cov(g3) ;    s3=round(s3, digits=3)          # group3 cov

n1=nrow(g1) ; n2=nrow(g2) ; n3=nrow(g3)
spool=( (n1-1)*s1 + (n2-1)*s2 + (n3-1)*s3 )/(n1+n2+n3-3)
spool=round(spool, digits=3)                      # pooled cov
```

```

M1= ( det(s1)/det(spool) )^{(n1-1)/2}
M2= ( det(s2)/det(spool) )^{(n2-1)/2}
M3= ( det(s3)/det(spool) )^{(n3-1)/2}
M=M1*M2*M3                                # Box's M test

v1=n1-1 ; v2=n2-1 ; v3=n3-1 ;   v=c(v1,v2,v3)
temp1= sum(1/v) ; temp2=(2*p^2 + 3* p -1)/(6*(p+1)*(g-1))
c1= (temp1-1/sum(v))*temp2
u= -2*(1-c1)*log(M)                       # test statistic
df= (g-1)*p*(p+1)/2
p.value=pchisq(u, df, lower.tail=FALSE)    # p-값
s1 ; s2; s3; spool
u ; p.value

# MANOVA test #
group <- factor(rootstock$group) # factor 지정 필요
y=cbind(x1,x2)                      # response vector
fit=manova( y ~ group)
summary(fit, test="Wilks")          # Wilks' lambda

```

[결과 6.3] 일원배치 공분산행렬에 대한 동일성 검정

```
> s1 ; s2; s3; spool
      x1    x2
x1 0.339 0.202
x2 0.202 0.199
      x1    x2
x1 0.342 0.259
x2 0.259 0.307
      x1    x2
x1 0.685 0.314
x2 0.314 0.154
      x1    x2
x1 0.455 0.258
x2 0.258 0.220
> u
[1] 10.19572
> p.value
[1] 0.1166482
```

[프로그램 6.4] 다변량 이원배치 분산분석

```
iron<- read.csv("C:/data/iron_2manova. header=T) ; attach(iron)
rot= factor(rot) # factor 지정 필요
lub= factor(lub)
## Two-way ANOVA with interaction
tx1=aov(x1~ rot + lub + rot:lub )
summary(tx1)
tx2=aov(x2~ rot + lub + rot:lub)
summary(tx2)
# Two-way Interaction Plot for x1
par(mfrow=c(1,2))
interaction.plot( rot, lub, x1, type="b", col=c(1:3),
  leg.bty="o", leg.bg="beige", lwd=2, pch=c(18,24,22),
  xlab="Rotation Speed", ylab="Lubricant", main="Interaction Plot : x1")
interaction.plot( lub, rot, x1, type="b", col=c(1:3),
  leg.bty="o", leg.bg="beige", lwd=2, pch=c(18,24,22),
  ylab="Rotation Speed", xlab="Lubricant", main="Interaction Plot : x1")
```



```
## Two-way MANOVA with interaction
x=cbind(x1, x2)    # response vector
fit = manova(x ~ rot + lub + rot:lub)    # manova(x ~ rot*lub)
summary(fit, test="Wilks")    # Wilks' lambda
summary(fit, test="Pillai")    # Pillai's trace
summary(fit, test="Roy")    # Roy's greatest root
summary(fit, test="Hotelling")    # Hotelling-Lawley trace
```

[결과 6.4](1) 이원배치 일변량 분산분석 결과

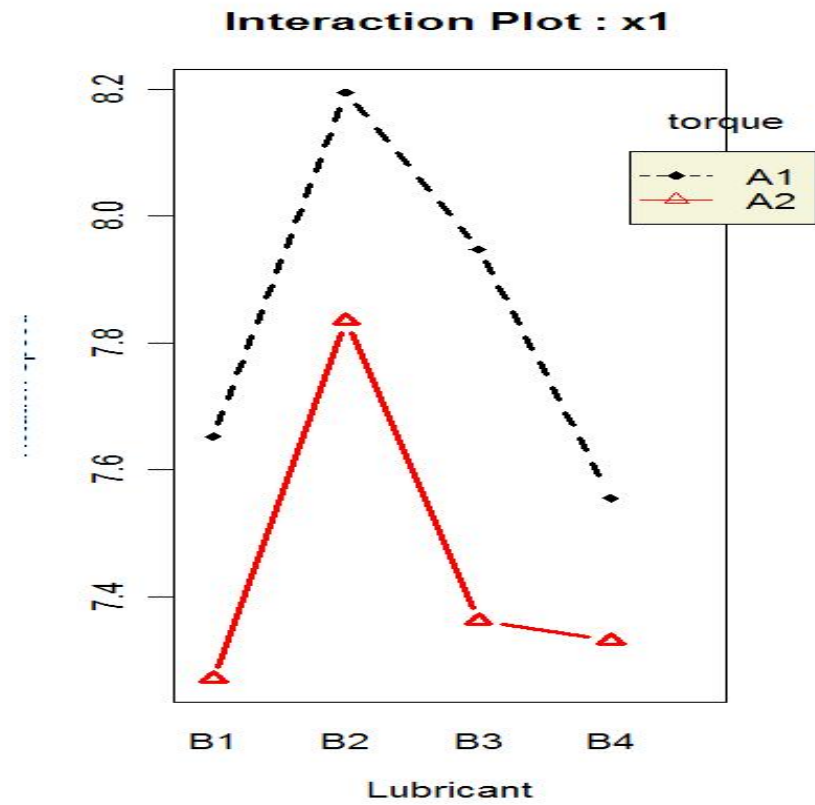
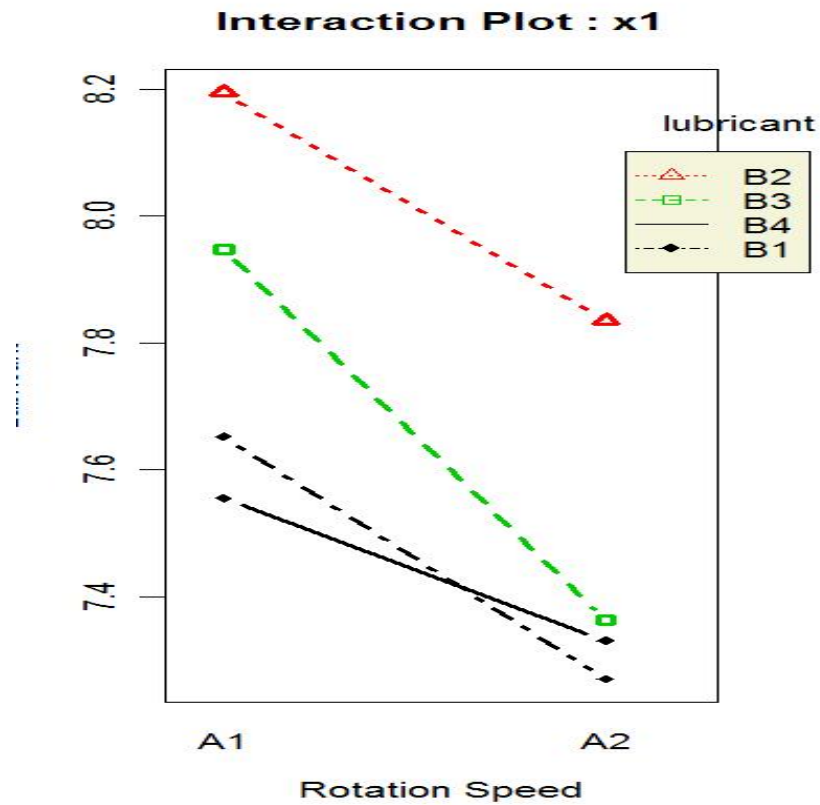
```
> ## Two-way ANOVA with interaction
> tx1=aov(x1~ rot + lub + rot:lub )
> summary(tx1)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|-----------|
| rot | 1 | 1.2051 | 1.2051 | 5.9069 | 0.02293 * |
| lub | 3 | 1.6941 | 0.5647 | 2.7678 | 0.06368 . |
| rot:lub | 3 | 0.1324 | 0.0441 | 0.2163 | 0.88412 |
| Residuals | 24 | 4.8965 | 0.2040 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> tx2=aov(x2~ rot + lub + rot:lub)
> summary(tx2)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|---------------|
| rot | 1 | 614.25 | 614.25 | 20.0193 | 0.0001581 *** |
| lub | 3 | 74.87 | 24.96 | 0.8134 | 0.4990176 |
| rot:lub | 3 | 32.24 | 10.75 | 0.3503 | 0.7892768 |
| Residuals | 24 | 736.39 | 30.68 | | |



[그림 6.5] x1에 대한 회전속도와 윤활유의 상호작용 그림

[결과 6.4](2) 이원배치 다변량 분산분석 결과

```
> ## Two-way MANOVA with interaction
> rot= factor(rot) # factor 지정 필요
> lub= factor(lub)
> x=cbind(x1, x2) # response vector
> fit <- manova(x ~ rot + lub + rot:lub) # manova(x ~ rot*lub)
> summary(fit, test="Wilks") # Wilks' lambda
```

| | Df | Wilks | approx F | num Df | den Df | Pr(>F) |
|-----------|----|--------|----------|--------|--------|---------------|
| rot | 1 | 0.4740 | 12.7636 | 2 | 23 | 0.0001867 *** |
| lub | 3 | 0.6916 | 1.5524 | 6 | 46 | 0.1827862 |
| rot:lub | 3 | 0.9319 | 0.2751 | 6 | 46 | 0.9458304 |
| Residuals | 24 | | | | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

> summary(fit, test="Pillai")      # Pillai's trace
      Df  Pillai approx F num Df den Df    Pr(>F)
rot      1  0.5260  12.7636      2    23 0.0001867 ***
lub      3  0.3141   1.4905      6    48 0.2015936
rot:lub   3  0.0685   0.2838      6    48 0.9418287
Residuals 24
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(fit, test="Roy")        # Roy's greatest root
      Df      Roy approx F num Df den Df    Pr(>F)
rot      1  1.1099  12.7636      2    23 0.0001867 ***
lub      3  0.4181   3.3448      3    24 0.0359080 *
rot:lub   3  0.0651   0.5204      3    24 0.6722819
Residuals 24
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```