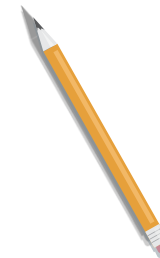


3강

연속형 자료 분석

정보통계학과
이태림 교수



목차

1. 단일 모평균 비교

평균 분산 비율비교

2. 두 집단의 평균 비교

두 모평균 비교

대응 평균 비교

3. 여러 집단의 평균비교

분산분석의 기본개념

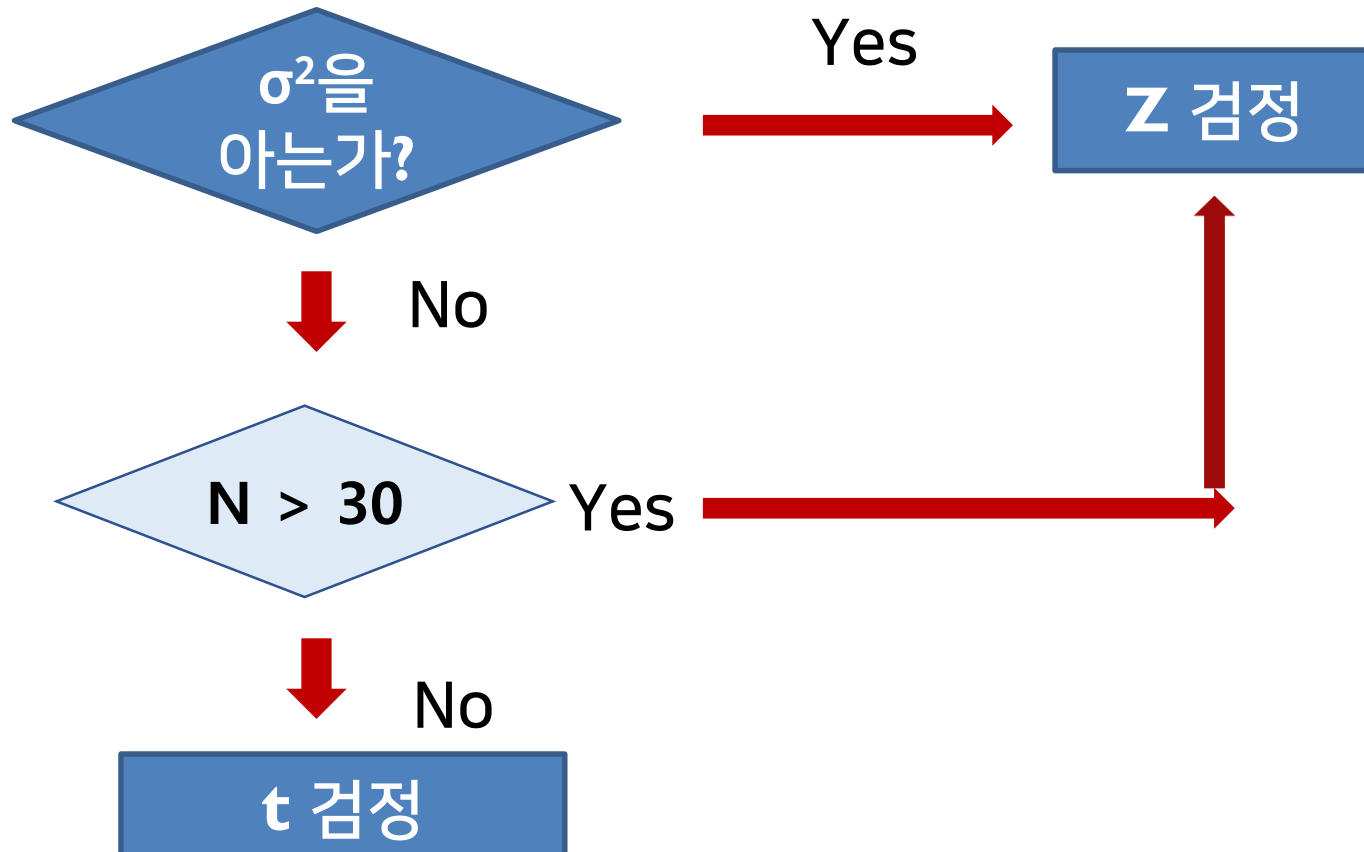
일원분산분석

이원분산분석

4. 통계패키지 수행

R 자료분석

조건에 따른 검정방법



I 단일 모평균 비교

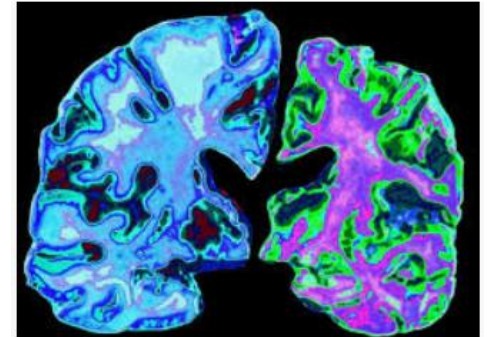
1. 단일 모평균 비교

알츠하이머 환자의 $^{14}\text{CO}_2$ 측정치

| | | | | | | | |
|------|------|------|------|------|------|------|------|
| 1009 | 1280 | 1180 | 1255 | 1547 | 2352 | 1956 | 1080 |
| 1776 | 1767 | 1680 | 2050 | 1452 | 2857 | 3100 | 1621 |

정상인의 기준 $^{14}\text{CO}_2$ 1500과 차이가 있는가?

| | |
|-------------|--------------------|
| $n = 16$ | $\bar{X} = 1747.6$ |
| $s = 604.7$ | $\mu_0 = 1500$ |



I 한 모평균 비교

1. 단일 모평균 비교

알츠하이머 환자의 $^{14}\text{CO}_2$ 측정치

가설

$$H_0: \mu = 1500 \quad H_1: \mu > 1500$$

유의수준

$$\alpha = 0.05$$

검정통계량

$$t_{\theta} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{1747.6 - 1400}{604.7/\sqrt{16}} = \frac{347.6}{151.2} = 2.3$$

검정

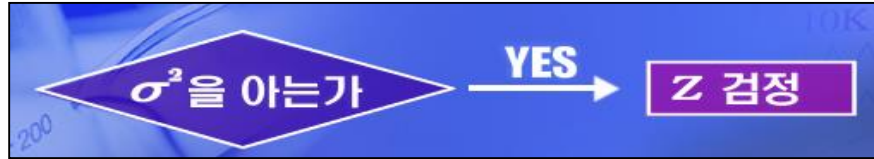
$$t_{\theta} = 2.3 > t(0.05, 15) = 1.753$$

이므로 H_0 를 기각한다

I 두 모평균 비교

2. 두 집단 모평균 비교

흡연에 따른 폐 파괴 현상



가설

$$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 \neq 0$$

유의수준

$$\alpha = 0.05$$

검정통계량

$$Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/n}} = \frac{12.9 - 9.9}{\sqrt{0.69/5 + 0.175/12}} = 7.68$$

검정

$$Z_0 = 7.68 > Z_{0.025} = 1.96$$

이므로 H_0 를 기각한다

I 한 모평균 비교

표 6.1 게의 체온 데이터



25 마리 게의 체온을 측정한 것이다. 과거 자료에서 게의 평균 체온은 24.3°C 로 알려져 있다. 이 데이터에서도 게의 평균 체온이 24.3°C 라고 해도 좋은지 알고 싶다. 이 경우 가설은 다음과 같이 설정한다.



| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 25.8 | 24.6 | 26.1 | 22.9 | 25.1 | 27.3 | 24.0 | 24.5 | 23.9 | 25.5 |
| 24.3 | 24.6 | 23.3 | 25.5 | 28.1 | 24.8 | 23.5 | 26.3 | 25.4 | 25.5 |
| 23.9 | 27.0 | 24.8 | 22.9 | 25.4 | | | | | |

$$H_0: \mu = 24.3$$

$$H_0: \mu - 24.3 = 0$$

\Leftrightarrow

$$H_1: \mu \neq 24.3$$

$$H_1: \mu - 24.3 \neq 0$$

I 한 모평균 비교

SAS 프로그램

```
DATA crab;  
  INPUT bodytemp @@;  
  hypotemp=bodytemp-24.3;  
  DATALINES;  
25.8 24.6 26.1 22.9 25.1  
27.3 24.0 24.5 23.9 26.2
```

출력 결과

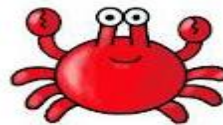
The TTEST Procedure

Statistics

| Variable | N | Lower CL | | Upper CL | | Std Err |
|----------|----|----------|-------|----------|------|---------|
| | | Mean | Mean | Mean | Mean | |
| bodytemp | 25 | 24.47 | 25.03 | 25.58 | | 0.268 |

T-Tests

| Variable | DF | t Value | Pr > t |
|----------|----|---------|---------|
| bodytemp | 24 | 2.71 | 0.0121 |

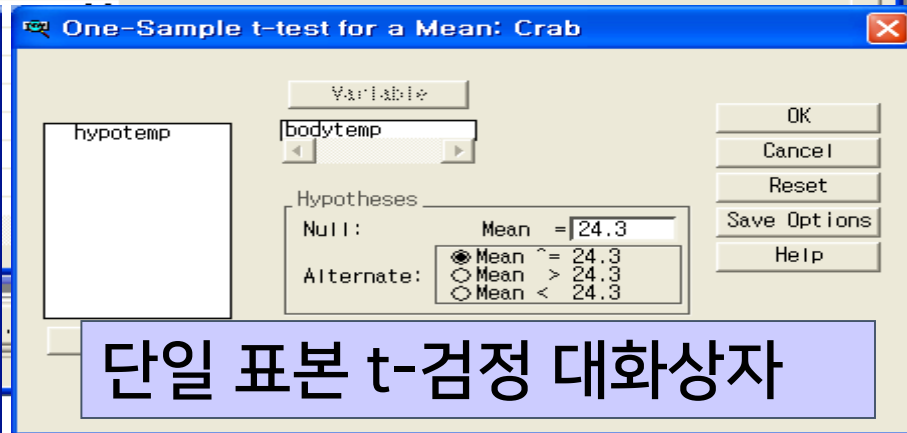
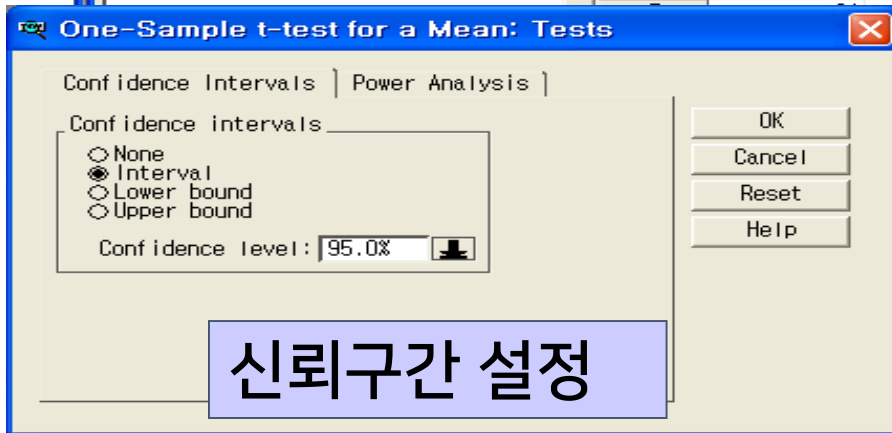
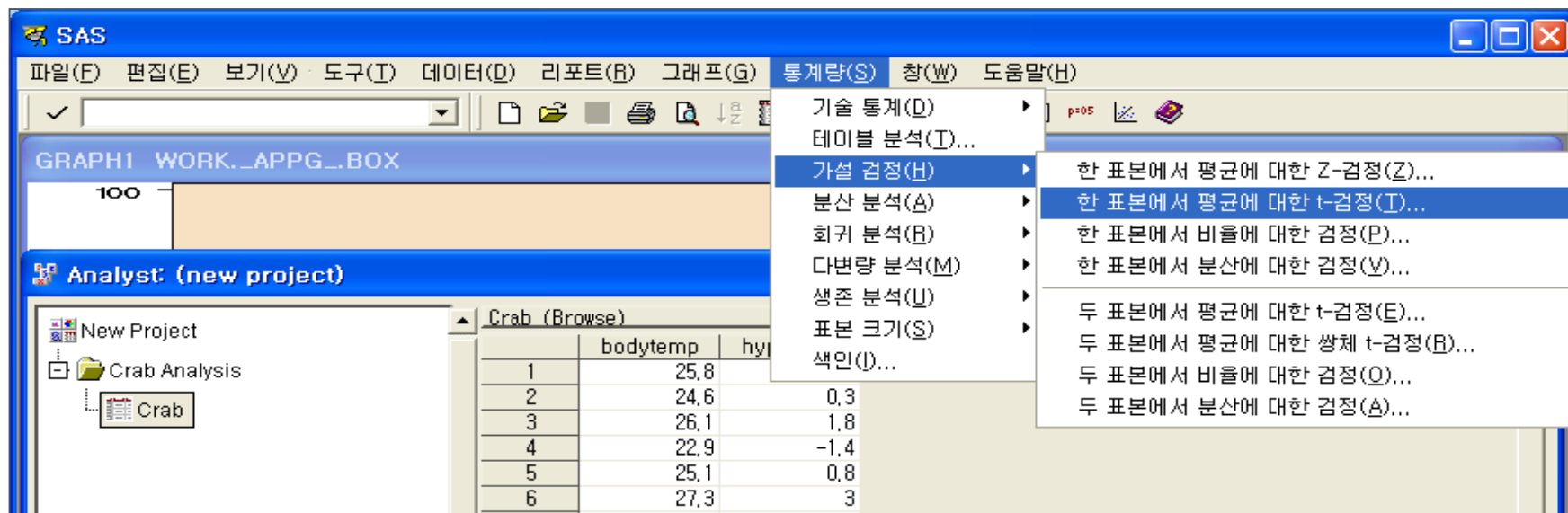


24.3

한모평균 비교

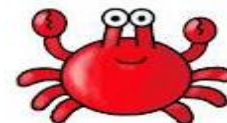
1. 한모평균 비교

SAS 프로그램



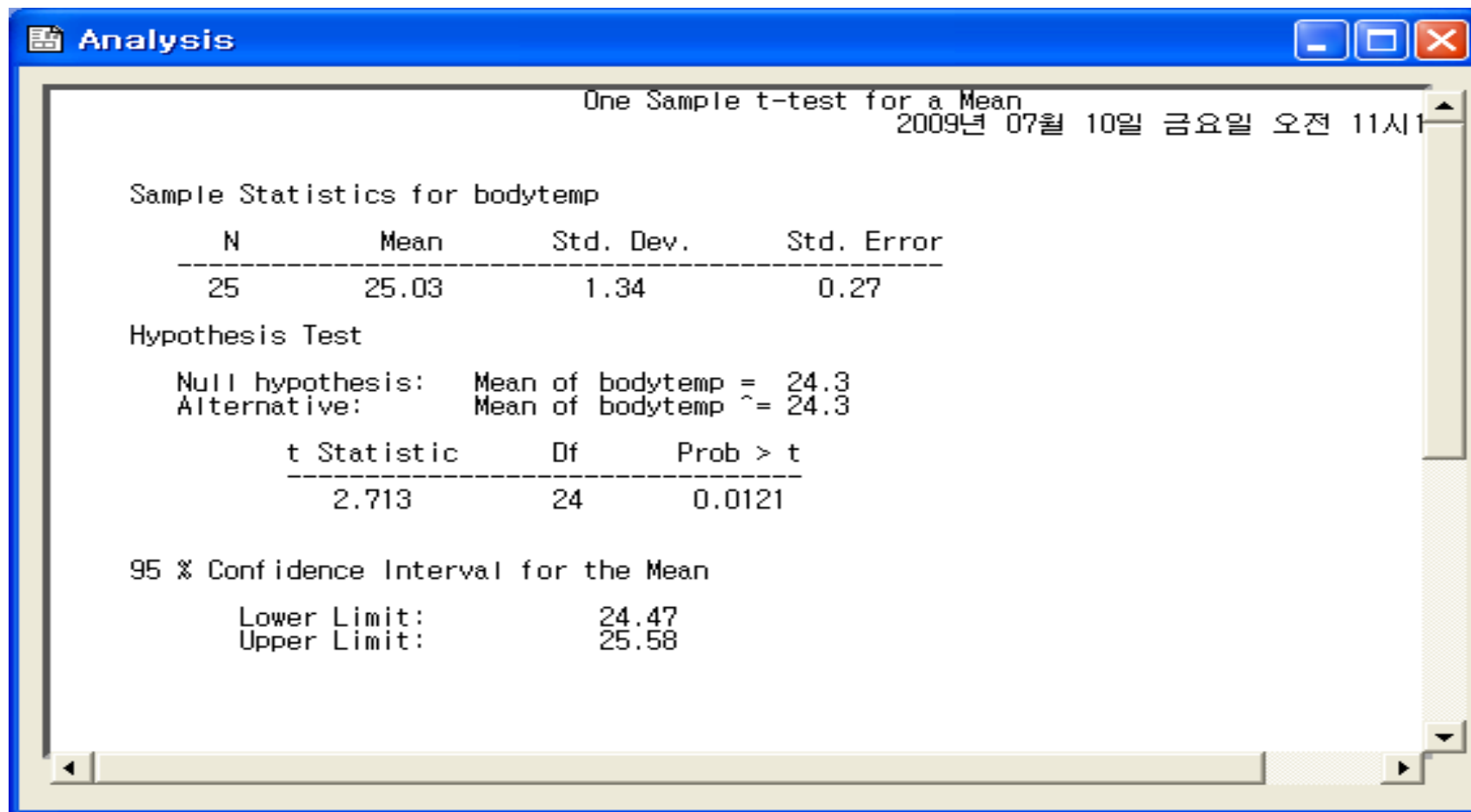
I 한 모평균 비교

1. 한 모평균 비교



출력 결과

단일표본 t-검정 결과



R 프로그램 수행

표 2.4 담즙의 과포화비율 자료

| | | | | | | | | | | | | |
|----|----|-----|-----|-----|-----|----|-----|-----|-----|-----|----|-----|
| 남자 | 40 | 86 | 111 | 86 | 106 | 66 | 123 | 90 | 112 | 52 | 88 | 137 |
| | 88 | 80 | 65 | 79 | 87 | 56 | 110 | 106 | 110 | 78 | 80 | 47 |
| | 74 | 58 | 88 | 73 | 118 | 67 | 57 | | | | | |
| 여자 | 65 | 86 | 76 | 89 | 142 | 58 | 98 | 146 | 80 | 66 | 52 | 35 |
| | 55 | 127 | 77 | 91 | 128 | 75 | 82 | 69 | 84 | 116 | 73 | 87 |
| | 76 | 107 | 84 | 120 | 123 | | | | | | | |



예 2.7

담즙과포화비율자료를 대상으로 다음과 같은 단측검정을 행하고 싶다고 하자.

H_0 : 담즙과포화비율의 모평균은 80보다 작거나 같다($\mu \leq 80$).

H_1 : 담즙과포화비율의 모평균은 80보다 크다($\mu > 80$).



한모평균 비교

1. 한모평균 비교

```
> #===== 2.5 추정과 가설검정  
> # [예 2.3]  
> # 단측 t-검정  
> t.test(담즙과포화비율, mu = 80, alternative = "greater")
```

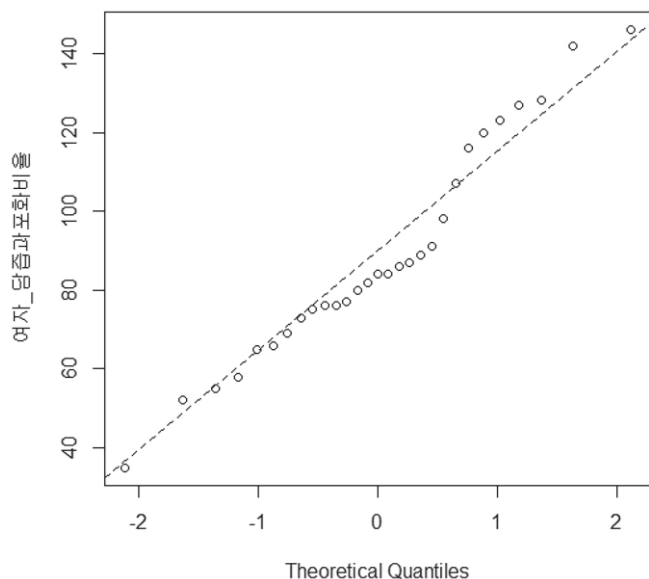
One Sample t-test

```
data: 담즙과포화비율  
t = 1.9389, df = 59, p-value = 0.02865  
alternative hypothesis: true mean is greater than 80  
95 percent confidence interval:  
 80.88621      Inf  
sample estimates:  
mean of x  
 86.41667
```

I 정규성검정: Shapiro-Wilk test

P75

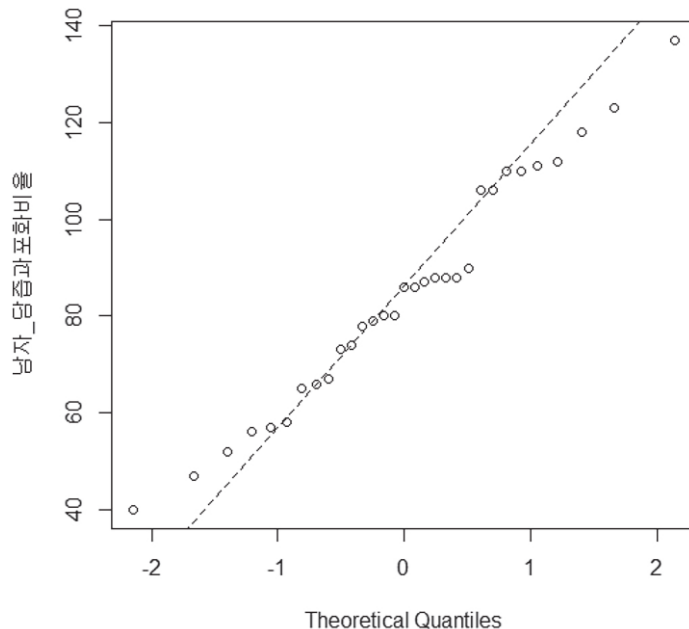
```
> # 정규확률플롯  
> qqnorm(남자_담즙과포화비율, ylab="남자_담즙과포화비율", main="")  
> qqline(남자_담즙과포화비율, lty=2)  
> # Shapiro-Wilk test  
> shapiro.test(남자_담즙과포화비율)
```



```
R Console  
➤ shapiro.test(남자_담즙과포화비율)  
  
Shapiro-Wilk normality test  
data: 남자_담즙과포화비율  
W = 0.97685, p-value = 0.7207
```

I 정규성검정: Shapiro-Wilk test

```
> # 정규확률플롯  
> qqnorm(여자_담즙과포화비율, ylab="여자_담즙과포화비율", main=" ")  
> qqline(여자_담즙과포화비율, lty=2)  
> # Shapiro-Wilk test  
> shapiro.test(여자_담즙과포화비율)
```



R Console

```
➤ shapiro.test(여자_담즙과포화비율)
```

```
Shapiro-Wilk normality test  
data: 여자_담즙과포화비율  
W = 0.95679, p-value = 0.2733
```

I 두 모평균 비교

양측 검정

σ^2 을
아는가?

↓ No

$n \geq 30$

↓ No

t 검정

폐기종 발병 이전의 흡연여부에 따른 폐 파괴지수 측정치가 같은가?

자료

| 흡연자 | | | | | | 비흡연자 | | | | |
|------|------|------|------|------|------|------|-----|------|------|-----|
| 16.6 | 13.9 | 11.3 | 26.5 | 17.4 | 15.3 | 18.1 | 6.0 | 10.8 | 11.0 | 7.7 |
| 15.8 | 12.3 | 18.6 | 12.0 | 24.1 | 16.5 | 17.9 | 8.5 | 13.0 | 18.9 | |
| 21.8 | 16.3 | 23.4 | 18.8 | | | | | | | |

$$n_1 = 16$$

$$\bar{X}_1 = 1747.6$$

$$\bar{X}_2 = 1747.6$$

$$n_2 = 16$$

$$s_1 = 604.7$$

$$s_2 = 4.85$$

P85

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 4.61$$

I 두 모평균 비교

1. 한 모평균 비교: 양측검정

양측검정

σ²을
아는가?

↓ No

n ≥ 30

↓ No

t 검정

가설

$$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 \neq 0$$

유의수준

$$\alpha = 0.05$$

검정통계량

$$t_{\theta} = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{17.54 - 12.43}{4.61 \sqrt{\frac{1}{16} + \frac{1}{9}}} = 2.66$$

검정

$$t_{\theta} = 2.66 > t_{0.025} = 2.069$$

이므로 H_0 를 기각하여 흡연여부에 따라 폐기종 환자의 폐파괴지수는 다르다고 결론짓는다.

I 두 모평균 비교

1. 한 모평균 비교: 단측검정

단측검정

폐기종 발병 이전의 흡연자의 폐 파괴지수 측정치가 더 큰가?

가설

$$H_0: \mu_1 - \mu_1 = 0 \quad H_1: \mu_1 - \mu_1 > 0$$

유의수준

$$\alpha = 0.05$$

검정통계량

$$t_{\theta} = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{17.54 - 12.43}{4.61 \sqrt{\frac{1}{16} + \frac{1}{9}}} = 2.66$$

검정

$$t_{\theta} = 2.66 > t_{(0.05, 23)} = 1.714 \text{ 이므로}$$

단측검정의 가설 흡연자의 폐파괴지수가 더 높다고
할 수 있다고 결론 짓는다.



<표3.1> 폐 파괴지수 자료의 t-검정

제 3장 연속형 자료의 분석 P85

I 병렬 상자그림, 병렬 바이올린

> #===== 3.1.1 독립표본의 평균비교

> # 자료 입력

> 흡연자

=c(16.6,13.9,11.3,26.5,17.4,15.3,15.8,12.3,18.6,12.0,24.1,16.5,21.8,16.3,23.4,18.8)

> 흡연자

[1] 16.6 13.9 11.3 26.5 17.4 15.3 15.8 12.3 18.6 12.0 24.1 16.5 21.8 16.3

[15] 23.4 18.8

Chapter 3. 연속형 자료의 분석 _ 89

> 비흡연자=c(18.1,6.0,10.8,11.0,7.7,17.9,8.5,13.0,18.9)

> 비흡연자

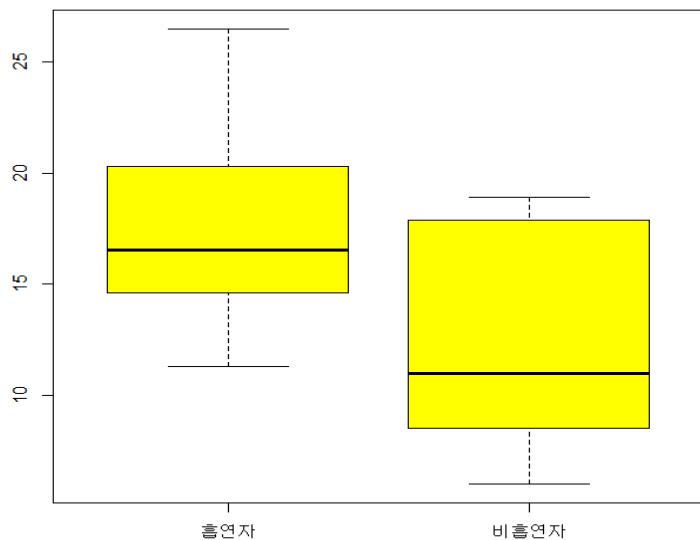
[1] 18.1 6.0 10.8 11.0 7.7 17.9 8.5 13.0 18.9

흡연자가2

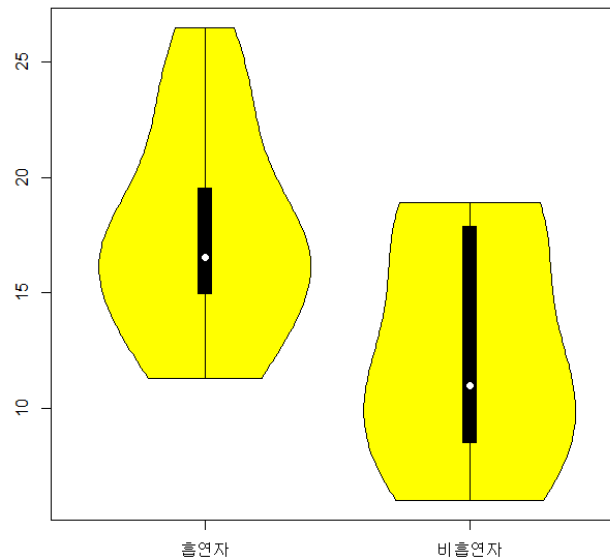
I 병렬 상자그림, 병렬 바이올린

```
> # 병렬상자그림  
> boxplot(흡연자,비흡연자,col="yellow",names=c("흡연자","비흡연자"))  
> points(c(mean(흡연자),mean(비흡연자)),pch=10,cex=2)  
> # 병렬바이올린그림  
> install.packages("vioplot")  
> library(vioplot)
```

vi
>
나



v",
흡연



연자","비흡연자"))

I 두 모분산 비교(양측검정)

```
> # 대립가설의 형태: alternative = c("two.sided",  
"less", "greater")  
> sd(흡연자);sd(비흡연자)  
> var.test(흡연자,비흡연자)
```

R Console

```
> # 두 모분산 비교(양측검정)  
> # 대립가설의 형태: alternative = c("two.sided", "less", "greater")  
> sd(흡연자);sd(비흡연자)  
[1] 4.475247  
[1] 4.849227  
> var.test(흡연자,비흡연자)  
  
      F test to compare two variances  
  
data:  흡연자 and 비흡연자  
F = 0.8517, num df = 15, denom df = 8, p-value = 0.7498  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.2076714 2.7243799  
sample estimates:  
ratio of variances  
      0.8517046
```

I 두 모평균 비교(양측검정)-등분산 가정

```
>t.test(흡연자,비흡연자,var.equal=T)
```

R R Console

```
> # 두 모평균 비교(양측검정)-등분산 가정  
> t.test(흡연자,비흡연자,var.equal=T)
```

Two Sample t-test

data: 흡연자 and 비흡연자

t = 2.658, df = 23, p-value = 0.01405

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1.131680 9.076653

sample estimates:

mean of x mean of y

17.53750 12.43333

I 두 모평균 비교(양측검정)-등분산 가정

```
> t.test(흡연자,비흡연자)
```

R Console

```
> # 두 모평균 비교(양측검정)-이분산 가정  
> t.test(흡연자,비흡연자)
```

Welch Two Sample t-test

data: 흡연자 and 비흡연자

t = 2.5964, df = 15.593, p-value = 0.01978

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.9279143 9.2804190

sample estimates:

mean of x mean of y

17.53750 12.43333

<표3.6> 혈압 비교 자료의 짝지은 t-검정

제 3장 연속형 자료의 분석 P94

I 짝지은 표본의 평균비교

표 3.3 투약 전후의 혈압비교에 관한 자료

p94

| 환자번호 | 복용 전 혈압 | 복용 후 혈압 | 차이 |
|------|---------|---------|----|
| 1 | 90 | 72 | 18 |
| 2 | 56 | 55 | 1 |
| 3 | 49 | 56 | -7 |
| 4 | 64 | 57 | 7 |
| 5 | 65 | 62 | 3 |
| 6 | 88 | 79 | 9 |
| 7 | 62 | 55 | 7 |
| 8 | 91 | 72 | 19 |
| 9 | 74 | 73 | 1 |
| 10 | 93 | 74 | 19 |
| 11 | 55 | 58 | -3 |
| 12 | 71 | 59 | 12 |
| 13 | 54 | 58 | -4 |
| 14 | 64 | 71 | -7 |
| 15 | 54 | 61 | -7 |

I 자료 입력

```
> 복용전혈압=c(90,56,49,64,65,88,62,91,74,93,55,71,54,64,54)
> 복용후혈압=c(72,55,56,57,62,79,55,72,73,74,58,59,58,71,61)
> 차이=복용전혈압-복용후혈압;차이
```

R R Console

```
> # 자료 입력
> 복용전혈압=c(90,56,49,64,65,88,62,91,74,93,55,71,54,64,54)
> 복용후혈압=c(72,55,56,57,62,79,55,72,73,74,58,59,58,71,61)
> 차이=복용전혈압-복용후혈압;차이
[1] 18  1 -7  7  3  9  7 19  1 19 -3 12 -4 -7 -7
>
```

I 정규성검정: Shapiro-Wilk test

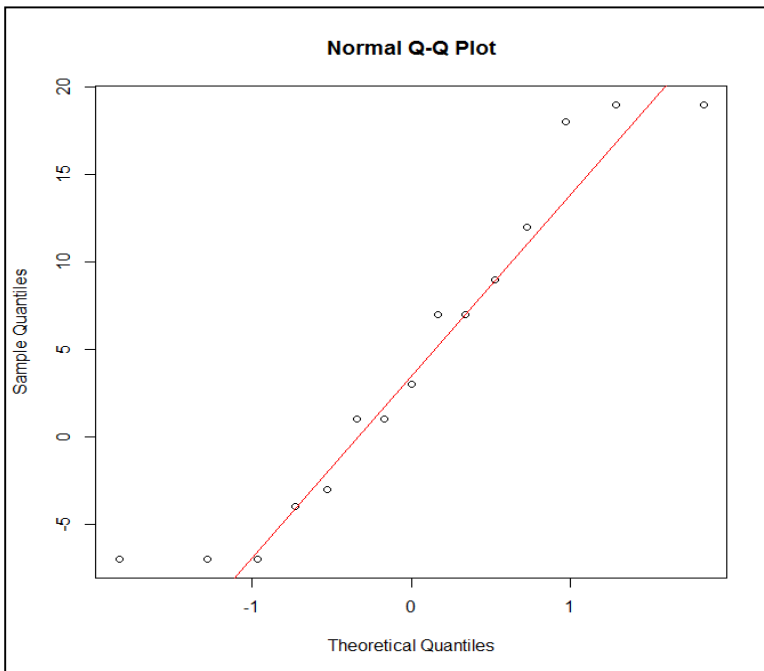
```
> qqnorm(차이)
> qqline(차이,col="red")
> shapiro.test(차이)
```

R Console

```
> # 정규성검정: Shapiro-Wilk test
> qqnorm(차이)
> qqline(차이,col="red")
> shapiro.test(차이)
```

Shapiro-Wilk normality test

```
data: 차이
W = 0.9098, p-value = 0.1345
```



I 짝지은 표본의 평균비교

```
> mean(차이);sd(차이)
> t.test(복용전혈압,복용후혈압
        ,paired=T,alternative="greater")
```

R R Console

```
>
> # 짝지은 표본의 평균비교
> mean(차이);sd(차이)
[1] 4.533333
[1] 9.425396
> t.test(복용전혈압, 복용후혈압, paired=T, alternative="greater")

        Paired t-test

data:  복용전혈압 and 복용후혈압
t = 1.8628, df = 14, p-value = 0.0418
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.2469617      Inf
sample estimates:
mean of the differences
      4.533333
```

여러 모평균 비교

3. 여러 집단의 평균비교

분산분석의 기본개념

일원분산분석(1-way ANOVA)

이원분산분석(2-way ANOVA)

반복이 없을 때

반복이 있을 때



I 분산분석(Analysis of variance)

세군 이상의 모평균 차이를 비교 검정하는 방법



I 일원분산분석(one-way ANOVA)

예

비만연구를 위해 30명의 비만 남성을 선정하고 임의로 세가지 체중감량 프로그램 중 하나에 참여시킨 다음 1년 후 체중 변화량을 측정하여 프로그램의 종류에 따라 유의한 차이가 있는지를 알아본다.



자료

| D1 | D2 | D3 |
|----|----|----|
| 10 | 2 | 8 |
| 7 | 3 | 10 |
| 3 | 4 | 9 |
| 5 | 2 | 10 |

자료구조식

$$y_i = \mu + a_i + \epsilon_i \quad (i = 1, 2, 3, \dots, k)$$

일원분산분석(one-way ANOVA)

가설

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots \mu_k$$
$$H_1: \mu_i \text{는 모두 같지 않다.}$$

유의수준

$$\alpha = 0.05$$

검정통계량

$$F_{\theta} = \frac{SS_{\text{trt}} / (k-1)}{SSE / (n-k)} = \frac{MS_{\text{trt}}}{MSE}$$

검정

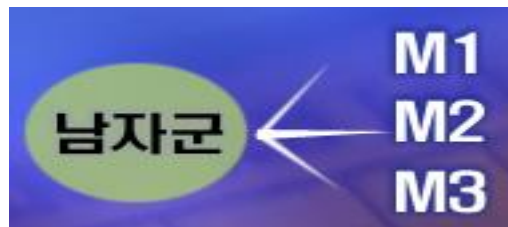
$$F_{\theta} = 2.66 > F_{\alpha(0.05, 23)} \text{ 이면 귀무가설 } H_0 \text{ 를 기각}$$

I 이원분산분석(two-way ANOVA)

예

세 종류의 호르몬 처리와 성별에 따라 혈액 내 칼슘량에 유의한 차이가 있는지 알아보기 위해 남녀 각 15명씩 선정하여 세 그룹으로 나누고 세 가지 호르몬 처리를 한 후 칼슘량을 측정한다.

자료



| | M1 | M2 | M3 |
|---|----|----|----|
| M | 25 | 18 | 17 |
| F | 28 | 16 | 18 |

자료구조식

$$y_{ij} = \mu + a_i + b_j + \epsilon_{ij} \quad (i, j = 1, 2, 3, \dots, k)$$

I 분산분석표(ANOVA Table)

$$\sum_i \sum_j (y - \bar{y})^2 = \sum_i n_i (\bar{y}_i - \bar{y})^2 + \sum_i \sum_j (\bar{y}_{ij} - \bar{y}_i)^2$$

$$\begin{aligned} \text{SST} &= \text{SS}_{\text{trt}} + \text{SSE} \\ \text{총제곱합} &= \text{처리제곱합} + \text{오차제곱합} \end{aligned}$$

$$\rightarrow F = \frac{\frac{SS_{\text{trt}}/(k-1)}{\frac{SSE}{n-k}}} = \frac{MS_{\text{trt}}}{MSE}$$

분산분석표

| | 제곱합 | 자유도 | 평균제곱합 | F |
|----|-------------------|-------|---|-----------------------|
| 처리 | SS_{trt} | $k-1$ | $MS_{\text{trt}} = \frac{SS_{\text{trt}}}{(k-1)}$ | $F = \frac{MSA}{MSE}$ |
| 오차 | SSE | $n-k$ | $MSE = \frac{SSE}{(n-k)}$ | |
| 계 | SST | $n-1$ | | |

I 일원분산분석 예제

p99

예

면역이상 연구를 위한 자폐아, 정상아, 지진아에 대한
혈청 항원농도를 측정한 자료.

자료

| 자폐아 | | 정상아 | | 지진아 | |
|-----|-----|-----|-----|-----|-----|
| 755 | 365 | 165 | 390 | 250 | 670 |
| 820 | 900 | 290 | 435 | 460 | 230 |
| 170 | 300 | 235 | 345 | | |
| ⋮ | | ⋮ | | | |

자료구조식

$$y_i = \mu + a_i + \epsilon_i \quad (i = 1, 2, 3, \dots, k)$$

일원분산분석(one-way ANOVA)

분산분석표

| | 제곱합 | 자유도 | 평균제곱합 | F | P-value |
|----|-----------|-----|---------|------------|---------|
| 처리 | 1375932.6 | 2 | 70617.3 | $F_0=3.49$ | 0.00871 |
| 오차 | 141234.6 | 68 | 20234.3 | | |
| 계 | 1517167.2 | 70 | | | |

가설

$H_0: \mu_1 = \mu_2 = \mu_3$ $H_1: \mu_i$ 는 모두 같지 않다.

유의수준

$$\alpha = 0.05$$

검정통계량

$$F_\theta = \frac{SS_{\text{trt}} / (k-1)}{SSE / (n-k)} = \frac{MS_{\text{trt}}}{MSE} = \frac{70619.3}{2234.3} = 3.49$$

검정결과

$F_\theta = 3.49 > F_{\alpha(0.05, 2, 68)} = 2.39$ 이면
귀무가설 H_0 를 기각

I 일원분산분석(one-way ANOVA)

분산분석표

| | 제곱합 | 자유도 | 평균제곱합 | F | P-value |
|----|-----------|-----|---------|------------|---------|
| 처리 | 1375932.6 | 2 | 70617.3 | $F_0=3.49$ | 0.00871 |
| 오차 | 141234.6 | 68 | 20234.3 | | |
| 계 | 1517167.2 | 70 | | | |

검정결과



자폐아 정성아 지진아 군에 따라 혈청항원의 농도평균이 모두 동일하지 않다.



컴퓨터 수행결과 p-value가 0.008 이므로
유의수준 0.05 보다 작아 귀무가설을 기각.

일원분산분석(one-way ANOVA)

<표3.12> 혈청 항원의 농도면역이상에 대한 일원분산분석

제 3장 연속형 자료의 분석 P99

I 자료입력(일원분산분석)

> 자폐아

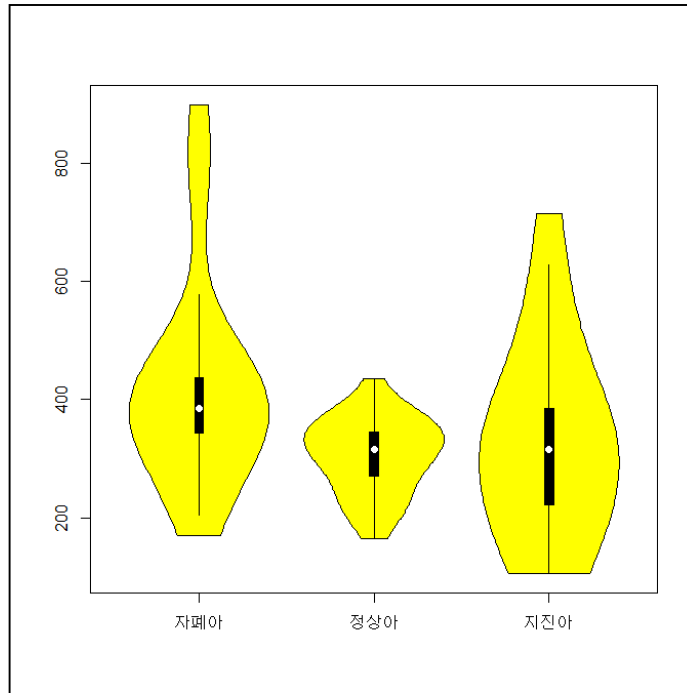
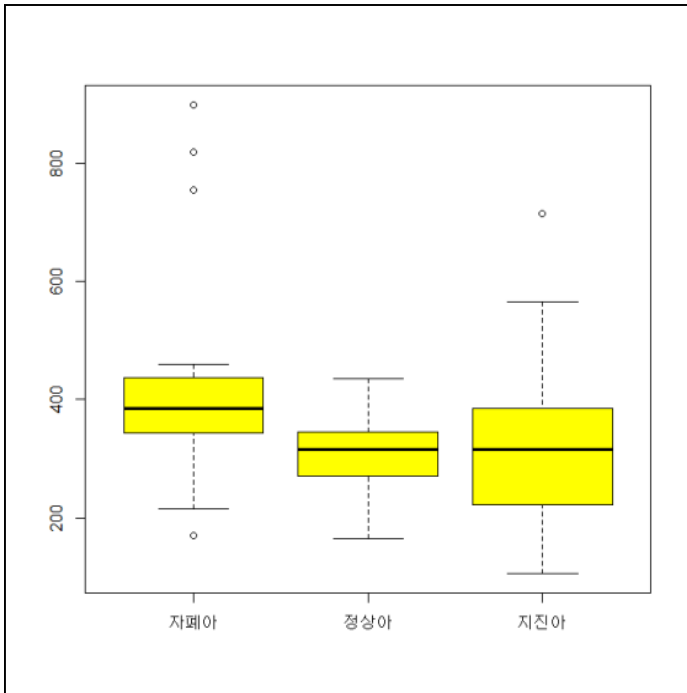
```
=c(755,365,820,900,170,300,325,385,380,215,400,343,415,  
345, 410,460,225,440,400,360,435,450,360)
```

> 정상아

```
R R Console  
2  
3  
5  
> 자폐아=c(755,365,820,900,170,300,325,385,380,215,400,343,415,345,  
+ 410,460,225,440,400,360,435,450,360)  
> 정상아=c(165,390,290,435,235,345,320,330,205,375,345,305,220,270,  
+ 355,360,335,305,325,245,285,370,345,345,230,370,285,315,195,270,305,375,220)  
> 지진아=c(380,510,315,565,715,380,390,245,155,335,295,200,105,105,245)  
>
```

I 병렬상자그림 병렬바이올린그림

```
> boxplot(자폐아,정상아,지진아  
,col="yellow",names=c("자폐아","정상아","지진아"))  
> library(vioplot)  
> violplot(자폐아,정상아,지진아  
,col="yellow",names=c("자폐아","정상아","지진아"))
```



I 각 그룹의 평균

```
> sapply(list(자폐아,정상아,지진아),mean)
```

```
> # 각 그룹의 평균
> sapply(list(자폐아, 정상아, 지진아), mean)
[1] 419.9130 305.0000 329.3333
```

> **apply**

```
> # 각 그룹의 분산
> sapply(list(자폐아, 정상아, 지진아), var)
[1] 31693 356 4071 875 28224 524
```

> 혈청항원농도

> 그 결과 =

| | | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| [1] | 755 | 365 | 820 | 900 | 170 | 300 | 325 | 385 | 380 | 215 | 400 | 343 | 415 | 345 | 410 | 460 | 225 | 440 | 400 | 360 | 435 | 450 | 360 |
| [24] | 165 | 390 | 290 | 435 | 235 | 345 | 320 | 330 | 205 | 375 | 345 | 305 | 220 | 270 | 355 | 360 | 335 | 305 | 325 | 245 | 285 | 370 | 345 |
| [47] | 345 | 230 | 370 | 285 | 315 | 195 | 270 | 305 | 375 | 220 | 380 | 510 | 315 | 565 | 715 | 380 | 390 | 245 | 155 | 335 | 295 | 200 | 105 |

72 [70] 105 245

```
> group = factor(rep(1:3, c(23, 33, 15)))
```

[illegible]

I 등분산성검정: Fligner-Killeen test

```
> fligner.test(혈청항원농도~그룹)
```

R Console

```
>  
> # 등분산성검정: Fligner-Killeen test  
> fligner.test(혈청항원농도~그룹)
```

Fligner-Killeen test of homogeneity of variances

data: 혈청항원농도 by 그룹

Fligner-Killeen:med chi-squared = 6.8506, df = 2, p-value = 0.03254

One-way ANOVA

- > 일원분산분석= `aov`(혈청항원농도~그룹)
- > 일원분산분석
- > `summary`(일원분산분석)

R Console

```
> # One-way ANOVA
> 일원분산분석=aov(혈청항원농도~그룹)
> 일원분산분석
Call:
  aov(formula = 혈청항원농도 ~ 그룹)

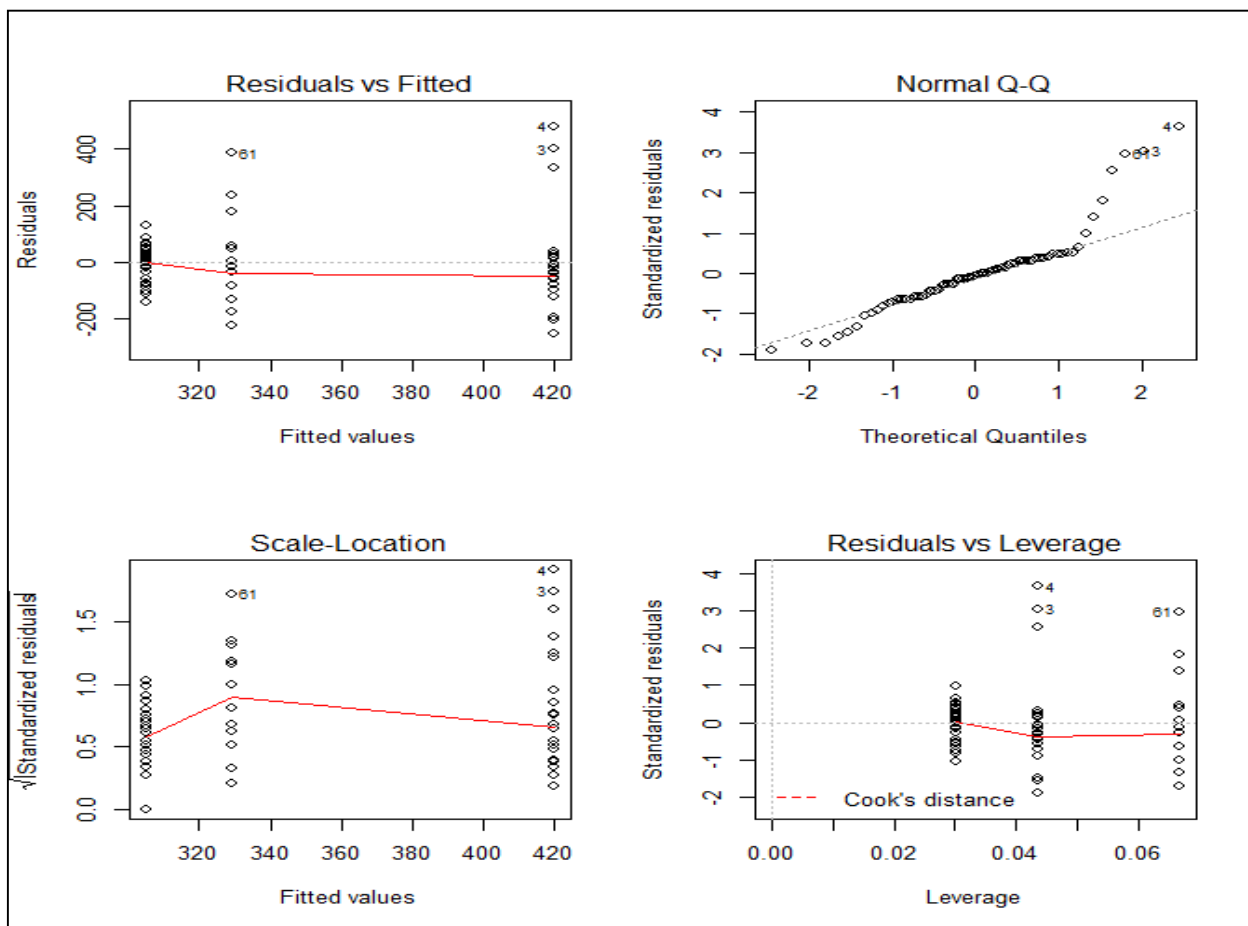
Terms:
              그룹 Residuals
Sum of Squares  185159.3 1236697.2
Deg. of Freedom      2      68

Residual standard error: 134.8582
Estimated effects may be unbalanced
> summary(일원분산분석)
              Df Sum Sq Mean Sq F value Pr(>F)
그룹           2  185159    92580   5.091 0.00871 **
Residuals      68 1236697    18187
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

모형적합성 검토=오차 검토

```
> par(mfrow=c(2,2))
```

```
> plot(일원분산분석)
```





<표3.14> 다중비교를 하기 위한 R 프로그램

제 3장 연속형 자료의 분석
P104

다중비교(multiple comparison): Tukey ' s HSD(honest significant differences)

```
> hsd=TukeyHSD
```

```
> hsd
```

```
> plot(hsd)
```

R Console

```
> # 다중비교(multiple comparison): Tukey's honest  
significant differences
```

```
> hsd=TukeyHSD(일원분산분석)
```

```
> hsd
```

Tukey multiple comparisons of means

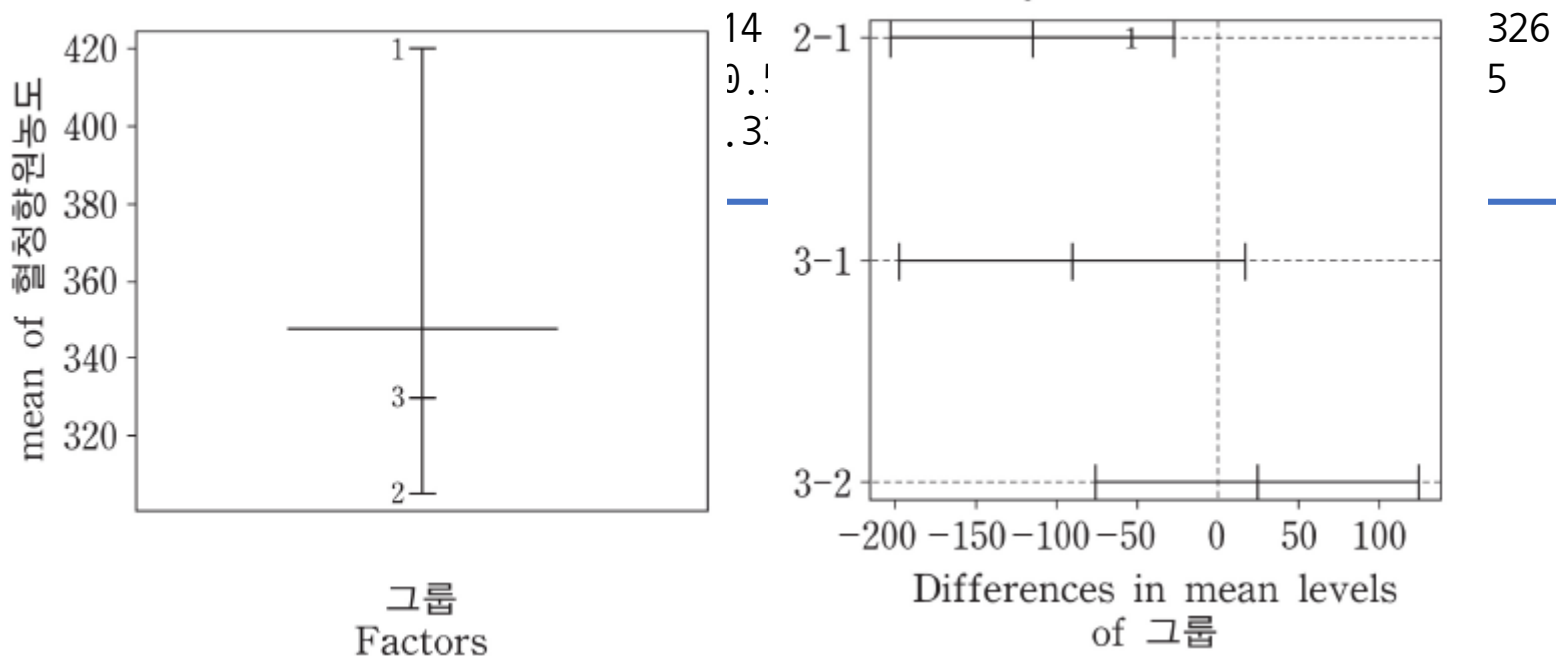
95% family-wise confidence level

Fit: aov(formula=혈청항원농도 ~ 그룹)

\$`그룹`

diff lwr upr p adj

95% family-wise confidence level



I 다중비교: LSD test(최소유의차검정)

> `pairwise.t.test`(혈청항원농도,그룹)

R Console

```
>
> # 다중비교: LSD test(최소유의차검정)
> pairwise.t.test(혈청항원농도, 그룹)

Pairwise comparisons using t tests with pooled SD

data: 혈청항원농도 and 그룹

   1      2
2 0.0076 -
3 0.0938 0.5642

P value adjustment method: holm
>
```

I 이원분산분석(two-way ANOVA)

자료

반복이 없는 경우
반복이 있는 경우

| | M1 | M2 | M3 |
|---|----|----|----|
| M | 25 | 18 | 17 |
| F | 28 | 16 | 18 |

자료구조식

$$y_{ij} = \mu + a_i + b_j + ab_{ij} + \epsilon_{ij} \quad (i = 1, 2, 3, \dots, k)$$

이원분산분석표

| | 제곱합 | 자유도 | 평균제곱합 | F비 |
|------|-------------------|--------------|-------------------|--|
| 요인A | SS_A | $I-1$ | MS_A | $F_A = MS_A / MSE$ |
| 요인B | SS_B | $J-1$ | MS_B | $F_B = MS_B / MSE$ |
| 교호작용 | $SS_{A \times B}$ | $(I-1)(J-1)$ | $MS_{A \times B}$ | $F_{A \times B} = MS_{A \times B} / MSE$ |
| 오차 | SSE | $IJ(n-1)$ | MSE | |
| 계 | SST | $IJn-1$ | | |

I 이원분산분석(two-way ANOVA)

가설

| | 귀무가설(H_0) | 검정 |
|--------------|----------------------------|--|
| A | $a_1 = a_2 \cdots a_i = 0$ | $F_{A0} > F_{\alpha}(I-1, IJ(n-1))$ |
| B | $b_1 = b_2 \cdots b_j = 0$ | $F_{B0} > F_{\alpha}(J-1, IJ(n-1))$ |
| A X B | $(a b)_{ij} = 0$ | $F_{A \times B} > F_{\alpha}((I-1)(J-1), IJ(n-1))$ |

대립가설(H_1) : H_0 가 아니다.

I 이원분산분석(two-way ANOVA)

자료

세 종류의 호르몬 처리와 성별에 따른
혈액 칼슘값이 차이가 있는가의 검정

| | 처리 1 | 처리 2 | 처리 3 |
|---|-------|-------|-------|
| 남 | 16.87 | 19.07 | 32.45 |
| | 16.18 | 18.77 | 28.71 |
| 여 | 15.86 | 17.20 | 30.54 |
| | 14.92 | 17.64 | 32.41 |

이원분산분석표

| | SS | df | MS | F | P-value |
|------|----------|----|--------|--------|---------|
| 성별 | 4.063 | 1 | 4.063 | 1.28 | 0.2094 |
| 처리 | 1146.6 | 2 | 573.32 | 180.35 | 0.0001 |
| 교호작용 | 3.845 | 2 | 1.923 | 0.60 | 0.5543 |
| 오차 | 76.292 | 24 | 3.180 | | |
| 계 | 1230.843 | 29 | | | |

I 이원분산분석(two-way ANOVA)

이원분산분석표

| | SS | df | MS | F | P-value |
|------|----------|----|--------|--------|---------|
| 성별 | 4.063 | 1 | 4.063 | 1.28 | 0.2094 |
| 처리 | 1146.6 | 2 | 573.32 | 180.35 | 0.0001 |
| 교호작용 | 3.845 | 2 | 1.923 | 0.60 | 0.5543 |
| 오차 | 76.292 | 24 | 3.180 | | |
| 계 | 1230.843 | 29 | | | |

기각치

$$F_{A0.05}(1, 24) = 4.26 > F_{0A} = 1.28 \quad H_{0A} \text{ 수용}$$

$$F_{B0.05}(2, 24) = 3.40 < F_{0B} = 180.35 \quad H_{0B} \text{ 기각}$$

$$F_{A \times B 0.05}(2, 24) = 3.40 < F_{0A \times B} = 0.60$$

$$H_{0A \times B} \text{ 수용}$$

I 이원분산분석(two-way ANOVA)

이원분산분석표

| | SS | df | MS | F | P-value |
|------|----------|----|--------|--------|---------|
| 성별 | 4.063 | 1 | 4063 | 1.28 | 0.2094 |
| 처리 | 1146.6 | 2 | 573.32 | 180.35 | 0.0001 |
| 교호작용 | 3.845 | 2 | 1.923 | 0.60 | 0.5543 |
| 오차 | 76.292 | 24 | 3.180 | | |
| 계 | 1230.843 | 29 | | | |



검정결과

→ 처리에 의해서는 칼슘값이 유의한 차이가 있으나 성별과 교호작용에 의해서는 유의한 차이가 없다.

→ P-value에 의해서도 $P_A = 0.2094$, $P_B = 0.0001$, $P_{A \times B} = 0.5543$ 을 $\alpha = 0.05$ 와 비교하여 같은 결과를 얻을 수 있다.

I 이원분산분석(two-way ANOVA)

| | SS | df | MS | F | P-value |
|------|----------|----|--------|--------|---------|
| 성별 | 4.063 | 1 | 4.063 | 1.28 | 0.2094 |
| 처리 | 1146.6 | 2 | 573.32 | 180.35 | 0.0001 |
| 교호작용 | 3.845 | 2 | 1.923 | 0.60 | 0.5543 |
| 오차 | 76.292 | 24 | 3.180 | | |
| 계 | 1230.843 | 29 | | | |

검정결과

→ 처리에 의해서는 칼슘값이 유의한 차이가 있으나 성별과 교호작용에 의해서는 유의한 차이가 없다.

→ P-value에 의해서도 $P_A = 0.2094$, $P_B = 0.0001$, $P_{A \times B} = 0.5543$ 을 $\alpha = 0.05$ 와 비교하여 같은 결과를 얻을 수 있다.

<표3.19>**의보장구 사용 숙지시간의
이원분산분석**

제 3장 연속형 자료의 분석 P95

I 데이터파일 읽기

> 의보장구사용숙지시간자자료<-

`read.table("C:\\Users\\knou\\Desktop\\R 실습자료\\의보장구사용숙지시간자자료.txt",header=T)`

> 의보장구사용숙지시간자자료

> `attach(의보장구사용숙지시간자자료)`

```
R Console
> # 데이터파일 읽기
> 의보장구사용숙지시간자자료<- read.table("C:\\U
> 의보장구사용숙지시간자자료
    나이 방법 숙지시간
1  20세미만   A        7
2    20-29   A        8
3    30-39   A        9
4    40-49   A       10
5   50이상   A       11
6  20세미만   B        9
7    20-29   B        9
8    30-39   B        9
9    40-49   B        9
10   50이상   B       12
11  20세미만   C       10
12    20-29   C       10
13    30-39   C       12
14    40-49   C       12
15   50이상   C       14
> attach(의보장구사용숙지시간자자료)
```

Two-way ANOVA

- > 이원분산분석=aov(숙지시간~나이+방법)
- > 이원분산분석
- > summary(이원분산분석)

R Console

```
>
> # Two-way ANOVA
> 이원분산분석=aov(숙지시간~나이+방법)
> 이원분산분석
Call:
  aov(formula = 숙지시간 ~ 나이 + 방법)

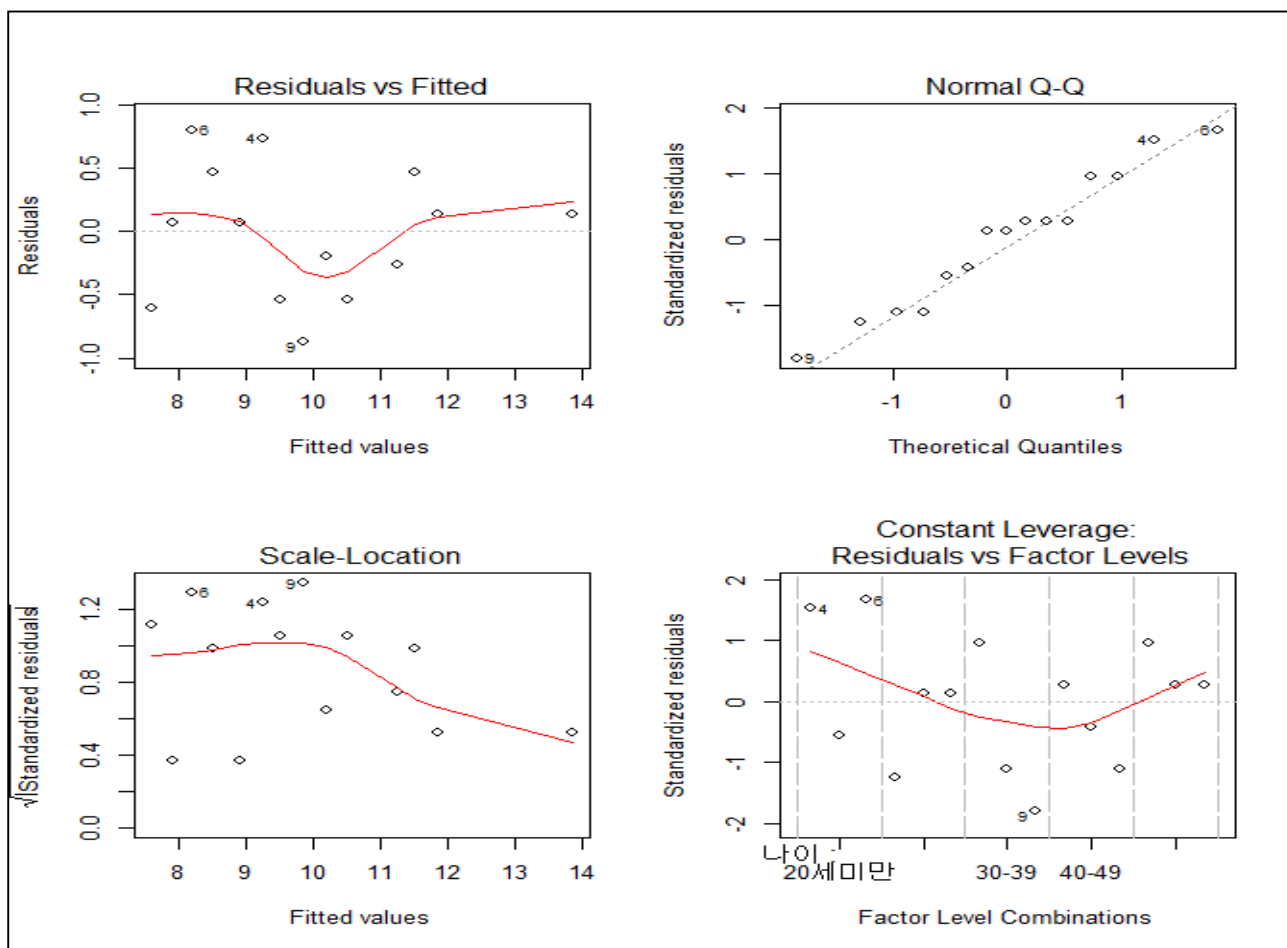
Terms:
              나이              방법 Residuals
Sum of Squares 24.933333 18.533333  3.466667
Deg. of Freedom      4          2      8

Residual standard error: 0.6582806
Estimated effects may be unbalanced
> summary(이원분산분석)
              Df Sum Sq Mean Sq F value    Pr(>F)
나이           4  24.933     6.233   14.38 0.001002 **
방법           2  18.533     9.267   21.39 0.000617 ***
Residuals       8   3.467     0.433
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```


모형적합성 검토=오차 검토

```
> par(mfrow=c(2,2))
```

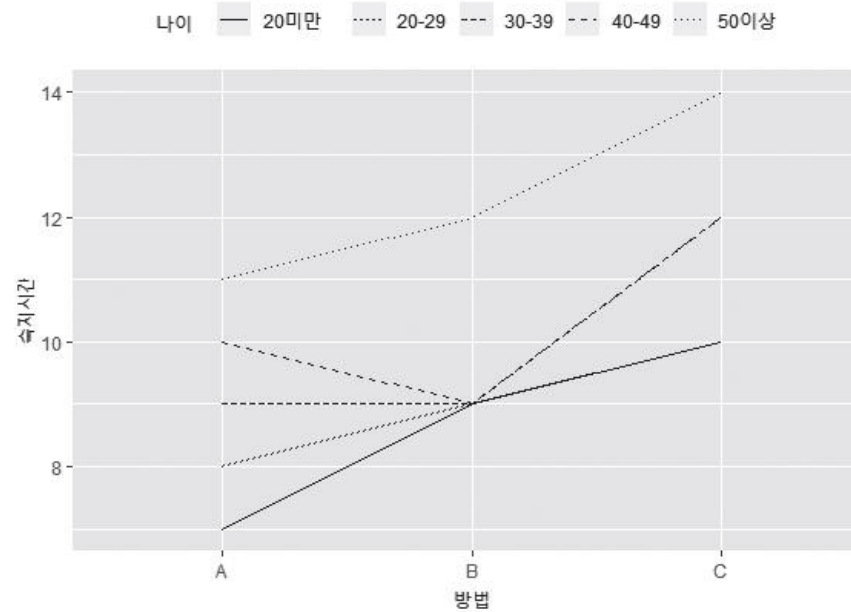
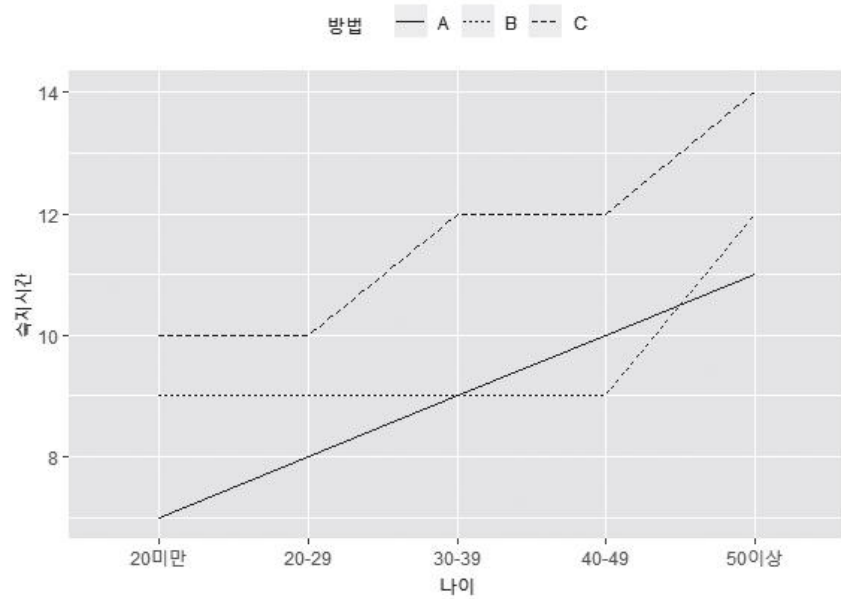
```
> plot(이원분산분석)
```



I 다중비교:LSD test

```
> # average response profile plot(평균반응프로파일그림)-효과 크기를 알 수 있는그림
> install.packages("ggplot2")
> library(ggplot2)
> ggplot(의보장구사용숙지시간자료,aes(x=나이,y=숙지시간,group=방법,linetype=방법))+geom_line()+theme(legend.position="top",legend.direction="horizontal")
> ggplot(의보장구사용숙지시간자료,aes(x=방법,y=숙지시간,group=나이,linetype=나이))+geom_line()+theme(legend.position="top",legend.direction="horizontal")
```

다중비교: LSD test



I 다중비교: LSD test

```
> lsd1=LSD.test(이원분산분석, "나이")
```

```
> lsd1$groups
```

숙지시간 groups

50이상 12.333333 a

40-49 10.333333 b

30-39 10.000000 bc

20-29 9.000000 cd

20미만 8.666667 d

```
> lsd2=LSD.test(이원분산분석, "방법")
```

```
> lsd2$groups
```

숙지시간 groups

| | | |
|---|------|---|
| C | 11.6 | a |
|---|------|---|

| | | |
|---|-----|---|
| B | 9.6 | b |
|---|-----|---|

| | | |
|---|-----|---|
| A | 9.0 | b |
|---|-----|---|

<표3.13> 호르몬 처리 후 혈액 칼슘 측정치의 이원분산분석

제 3장 연속형 자료의 분석 P115

<표3.13> 호르몬 처리 후 혈액 칼슘 측정치의 이원분산분석

표 3.13 세 가지 호르몬 처리와 남녀의 혈액 칼슘 측정결과

| 성별 \ 처리 | 처리 1 | 처리 2 | 처리 3 |
|---------|---|---|---|
| 남 | 16.87 16.18 17.12 16.83 17.19 | 19.07 18.77 17.63 16.99 18.04 | 32.45 28.71 34.65 28.79 24.46 |
| 여 | 15.86 14.92 15.63 15.24 | 17.20 17.64 17.89 16.78 | 30.54 32.41 28.97 28.46 |

표 3.14 처리조합별 평균

| 성별 \ 처리 | 처리 1 | 처리 2 | 처리 3 | 전체 |
|---------|-------------------------|-------------------------|-------------------------|-------------------------|
| 남 | $\bar{y}_{11.} = 16.84$ | $\bar{y}_{12.} = 18.1$ | $\bar{y}_{13.} = 29.81$ | $\bar{y}_{1..} = 21.58$ |
| 여 | $\bar{y}_{21.} = 15.29$ | $\bar{y}_{22.} = 17.25$ | $\bar{y}_{23.} = 30.01$ | $\bar{y}_{2..} = 20.85$ |
| 전체 | $\bar{y}_{.1.} = 16.06$ | $\bar{y}_{.2.} = 17.67$ | $\bar{y}_{.3.} = 29.91$ | $\bar{y}_{...} = 21.22$ |

I 데이터 읽기

> 혈액칼슘자료<-

`read.table("C:\\Users\\knou\\Desktop\\R 실습자료\\혈액칼슘자료.txt",header=T)`

> 혈액칼슘자료

> `attach(혈액칼슘자료)`

```
R Console
> 혈액칼슘자료<- read.table("C:\\Users\\knou\\Desktop\\R 실습자료\\혈액칼슘자료.txt",header=T)
> 혈액칼슘자료
  성별 처리 혈액칼슘
1   A   A    16.87
2   A   A    16.18
3   A   A    17.12
4   A   A    16.83
5   A   A    17.19
6   B   B    19.07
7   B   B    18.77
8   B   B    17.63
9   B   B    16.99
10  B   B    18.04
11  C   C    32.45
12  C   C    28.71
13  C   C    34.65
14  C   C    28.79
15  C   C    24.46
16  A   A    15.86
17  A   A    14.92
18  A   A    15.63
19  A   A    15.24
20  A   A    14.80
21  B   B    17.20
22  B   B    17.64
23  B   B    17.89
24  B   B    16.78
25  B   B    16.72
26  C   C    30.54
27  C   C    32.41
28  C   C    28.97
29  C   C    28.46
30  C   C    29.65
> attach(혈액칼슘자료)
```

Two-way ANOVA

- > 이원분산분석2=aov(혈액칼슘~성별*처리)
- > 이원분산분석2
- > summary(이원분산분석2)

```
R Console
>
> # Two-way ANOVA
> 이원분산분석2=aov(혈액칼슘~성별*처리)
> 이원분산분석2
Call:
aov(formula = 혈액칼슘 ~ 성별 * 처리)

Terms:
          성별          처리  성별:처리 Residuals
Sum of Squares    4.0627 1146.6420    3.8454    76.2924
Deg. of Freedom         1         2         2         24

Residual standard error: 1.782933
Estimated effects may be unbalanced
> summary(이원분산분석2)
          Df Sum Sq Mean Sq F value    Pr(>F)
성별       1    4.1      4.1    1.278    0.269
처리       2 1146.6    573.3  180.355 3.47e-15 ***
성별:처리   2     3.8      1.9    0.605    0.554
Residuals  24    76.3      3.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```


I 다중비교: LSD test

- > pairwise.t.test(혈액칼슘,성별)
- > pairwise.t.test(혈액칼슘,처리)

R Console

```
> # 다중비교: LSD test
> pairwise.t.test(혈액칼슘,성별)

Pairwise comparisons using t tests with pooled SD

data: 혈액칼슘 and 성별

    남
여 0.76

P value adjustment method: holm
> pairwise.t.test(혈액칼슘,처리)

Pairwise comparisons using t tests with pooled SD

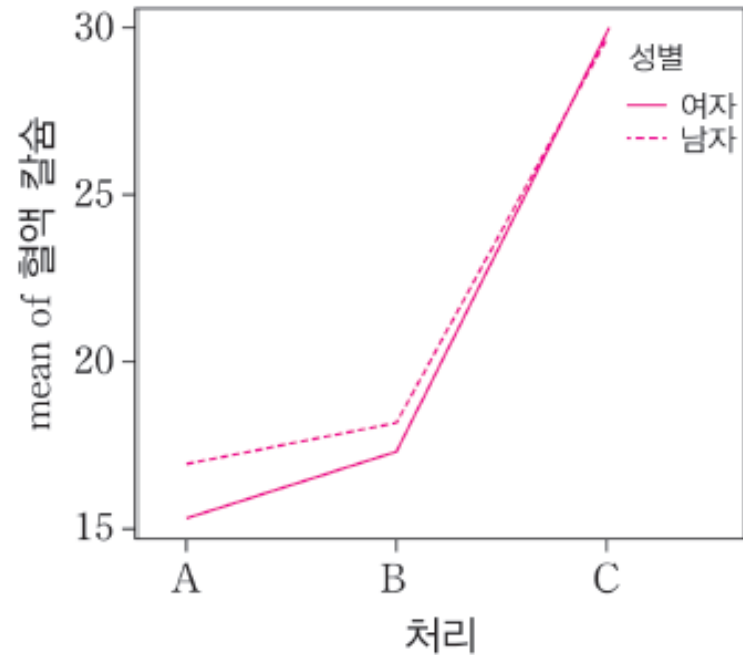
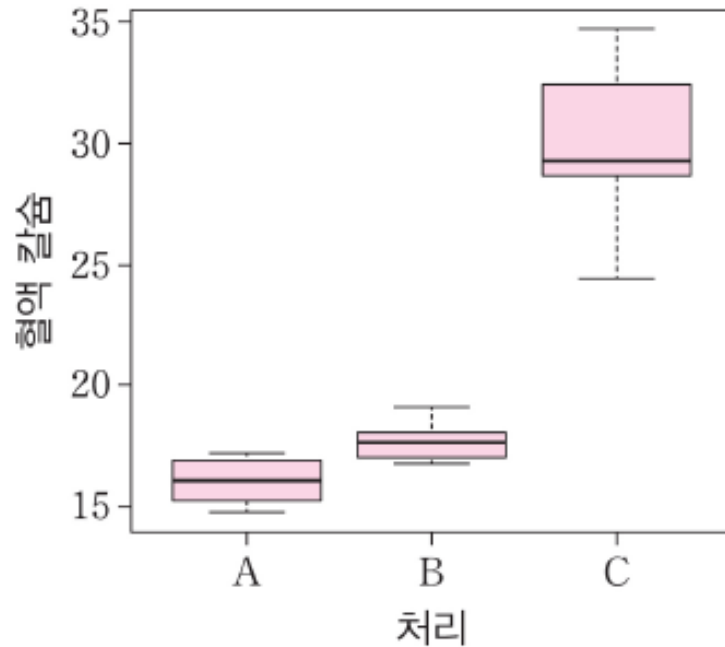
data: 혈액칼슘 and 처리

      A      B
B 0.052  -
C 8.4e-16 1.2e-14

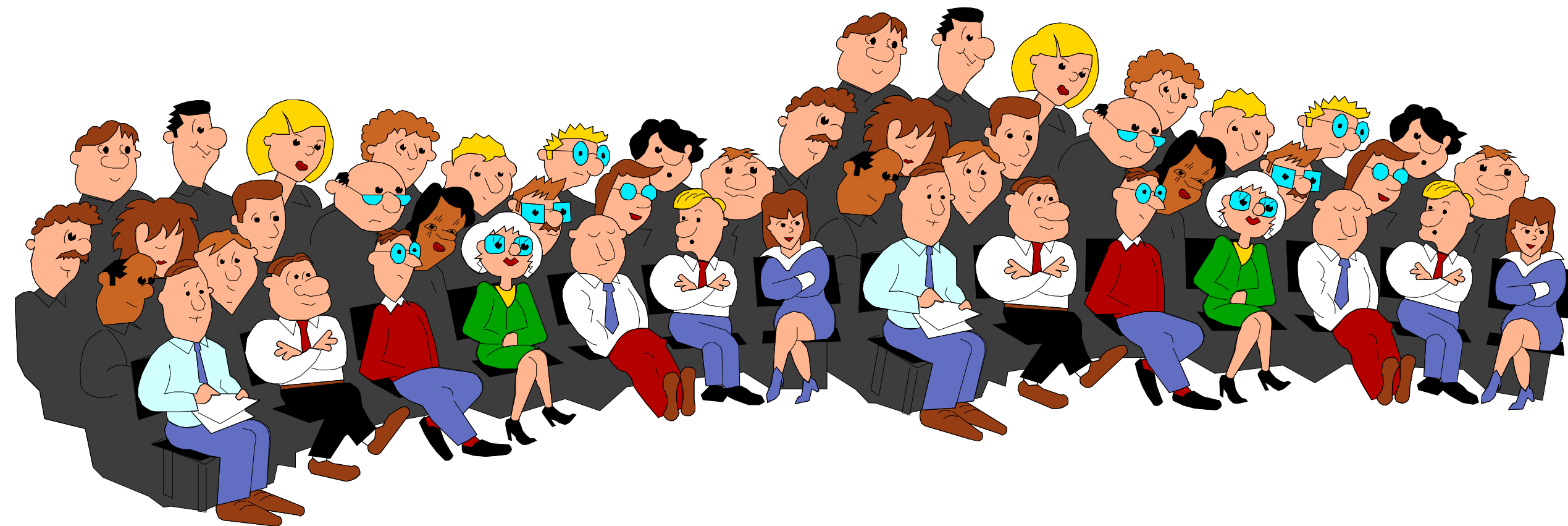
P value adjustment method: holm
>
```

다중비교:LSD test

- > # 병렬상자그림과 교호작용도 (interaction plot)
- > plot(혈액칼슘 ~ 처리, 혈액칼슘자료)
- > with(혈액칼슘자료, interaction.plot(처리, 성별, 혈액칼슘))



Q & A



다음시간에는

▶ 3강 연속형자료의 분석

4강 범주형자료 분석

5강 공분산 분석