**데이터 애널리틱스 2차 과제**

* **학번: 201402167**
* **성명: 윤현우**

**※ 수업내용을 바탕으로 다음 통계분석을 수행하시오.**

**1.“DA\_AS02-01.csv”파일은 선발된 일반인 20명을 대상으로 다이어트 보조제 복용 전(x1)과 2개월간 복용 후(x2)의 몸무게를 개인별로 기록한 것이다. 다이어트 보조제 복용 후 체중이 줄었다고 할 수 있는가? 유의수준 0.05에서 검정하시오. (1점)**

**(R 코드는 “DA\_AS02.R”파일의 #01 아래에 작성하고, (분석 결과 설명에 필요하다면) 그래프를 포함하여 분석결과는 아래에 작성할 것.)**

**<풀이>**

**- 이 문제에 적용할 분석기법은 무엇이며, 해당 분석기법을 적용하려는 이유는 무엇인가?**

**: 쌍체비교. 각 관측치는 독립적인 데이터가 아닌 연관성이 있는 종속표본이기 때문이다.**

**- 가설설정**

**H0: μx1 ≤ μx2**

**H1: μx1 ＞ μx2**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> df01 <- read.csv("C:/DA2019/exercise2/DA\_AS02-01.csv")

> df01

x1 x2

1 82 75

2 54 50

3 74 74

4 75 71

5 71 69

6 76 73

7 70 68

8 62 62

9 77 68

10 75 72

11 72 70

12 83 77

13 78 71

14 74 74

15 68 67

16 76 73

17 75 77

18 75 71

19 75 76

20 71 74

> t.test(df01$x1, df01$x2, paired=TRUE, alternative = "greater")

Paired t-test

data: df01$x1 and df01$x2

t = 3.6355, df = 19, **p-value = 0.0008801**

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

1.337172 Inf

sample estimates:

mean of the differences

2.55

**t-검정 결과 p-value가 0.0008801로 유의수준 0.05보다 작으므로(p-value = 0. 0008801< α = 0.05) 귀무가설을 기각하고 대립가설을 채택한다. 따라서 다이어트 보조제 복용 후 체중의 평균이 다이어트 보조제 복용 전 체중의 평균보다 작다고 할 수 있다.**

**2. K社의 영업부서에서는 영업사원 교육을 위한 네 가지 교육프로그램의 효과에 차이가 있는지를 조사하기 위하여 실험을 실시하였다. 이 실험에서는 28명의 신입사원들을 무작위로 네 집단으로 나누어 교육프로그램 p1, p2, p3, p4로써 교육을 실시하였다. 교육 도중 5명이 중도에 탈락하고 교육을 마친 후 1주일간의 판매실적(단위: 백 만원)은 “DA\_AS02-02.csv”파일에 기록되어 있다. 유의수준 0.05에서 교육프로그램의 효과가 다른지 검정하고, 필요하다면 추가분석을 실시하시오. (1점)**

**(R 코드는 “DA\_AS02.R”파일의 #02 아래에 작성하고, (분석 결과 설명에 필요하다면 그래프를 포함하여) 분석결과는 아래에 작성할 것.)**

**<풀이>**

**- 이 문제에 적용할 분석기법은 무엇이며, 해당 분석기법을 적용하려는 이유는 무엇인가?**

**: 일원분산분석. 요인이 1개지만 요인 수준이 4개라 t-검정보다 anova분석을 하는 것이 좋다.**

**- 가설설정**

**H0: μ(p1) = μ(p2) = μ(p3) = μ(p4)**

**H1: 적어도 하나의 교육프로그램에서 판매실적은 같지 않다.**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> df02 <- read.csv("C:/DA2019/exercise2/DA\_AS02-02.csv")

> df02

program sales

1 p1 65

2 p1 87

3 p1 73

4 p1 79

5 p1 81

6 p1 69

7 p2 75

8 p2 69

9 p2 83

10 p2 81

11 p2 72

12 p2 79

13 p2 90

14 p3 59

15 p3 78

16 p3 67

17 p3 62

18 p3 83

19 p3 76

20 p4 94

21 p4 89

22 p4 80

23 p4 88

>

> fit02 <- aov(df02$sales ~ df02$program, data=df02)

> anova(fit02)

Analysis of Variance Table

Response: df02$sales

Df Sum Sq Mean Sq F value Pr(>F)

df02$program 3 712.59 237.529 3.7715 **0.02804 \***

Residuals 19 1196.63 62.981

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> TukeyHSD(fit02)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = df02$sales ~ df02$program, data = df02)

$`df02$program`

diff lwr upr p adj

p2-p1 2.761905 -9.652957 15.176766 0.9225816

p3-p1 -4.833333 -17.716844 8.050177 0.7200708

p4-p1 12.083333 -2.320870 26.487536 0.1197999

p3-p2 -7.595238 -20.010100 4.819623 0.3410918

p4-p2 9.321429 -4.665172 23.308029 0.2718203

**p4-p3 16.916667 2.512464 31.320870** **0.0179873**

> plot(TukeyHSD(fit02))

**분산분석 결과 p-value가 0.02804로 유의수준 0.05보다 작으므로(p-value = 0.02804 < α = 0.05) 귀무가설을 기각하고 대립가설을 채택한다. 즉, 교육프로그램에 따른 판매실적은 다르다고 할 수 있다.**

**추가분석 결과 교육프로그램 3과 4 사이에 유의한 차이(교육프로그램 4가 교육프로그램 3보다 16.916667만큼 크다.)가 있음을 알 수 있다.**

**3. “DA\_AS02-03.csv”파일은 H대학 경영학과 1학년 30명의 키(height, 단위: cm)와 몸무게(weight, 단위: kg)를 기록한 것이다. 키와 몸무게 각각의 평균, 분산을 구하고 키와 몸무게의 공분산과 상관계수를 구한 후 유의수준 0.05에서 상관계수의 유의성을 검정하시오. (1점)**

**(R 코드는 “DA\_AS02.R”파일의 #03 아래에 작성하고, (분석 결과 설명에 필요하다면 그래프를 포함하여) 분석결과는 아래에 작성할 것.)**

**<풀이>**

**- 이 문제에 적용할 분석기법은 무엇이며, 해당 분석기법을 적용하려는 이유는 무엇인가?**

**: 상관분석. 상관계수를 통해 각 변수 간의 관련성의 정도와 방향성을 알 수 있기 때문이다.**

**- 가설설정**

**H0: ρ = 0 (두 변수 사이에 상관관계가 없다.)**

**H1: ρ ≠ 0 (두 변수 사이에 상관관계가 있다.)**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> df03 <- read.csv("C:/DA2019/exercise2/DA\_AS02-03.csv")

> df03

weight height

1 72 176

2 72 172

3 70 182

4 43 160

5 48 163

6 54 165

7 51 168

8 52 163

9 73 182

10 45 148

11 60 170

12 62 166

13 64 172

14 47 160

15 51 163

16 74 170

17 88 182

18 64 174

19 56 164

20 56 160

21 62 178

22 70 175

23 73 173

24 82 188

25 75 180

26 65 170

27 56 168

28 65 166

29 54 170

30 77 182

>

> # 평균

> mean(df03$weight)

[1] 62.7

> mean(df03$height)

[1] 170.3333

>

> # 분산

> var(df03$weight)

[1] 132.7

> var(df03$height)

[1] 75.4023

>

> # 공분산

> cov(df03, use="complete.obs")

weight height

weight 132.70000 85.68966

height 85.68966 75.40230

>

> # 상관관계

> cor.test(df03$weight, df03$height, data = df03)

Pearson's product-moment correlation

data: df03$weight and df03$height

t = 8.7863, df = 28, **p-value = 1.544e-09**

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.7179513 0.9299195

sample estimates:

cor

**0.8566445**

**분석 결과, weight의 평균은 62.7, height의 평균은 170.3333, weight의 분산은 132.7, height의 분산은 75.4023으로 나타났다.**

**또한 weight와 height의 공분산은 85.68966이며, weight와 height 간의 상관관계는 0.8566445이다.**

**상관관계에 대한 p-value는 1.544e-09으로 유의수준 0.05보다 작으므로(p-value = 1.544e-09 < α = 0.05) 귀무가설을 기각하고 대립가설을 채택한다. 즉, 두 변수의 관련성은 매우 유의하다(상관관계가 있다)고 할 수 있다.**

**4. 한 방송국에서는 65세 이상 시청자들을 위한 TV 프로그램을 개발하기로 하였다. 이를 위한 기초정보를 획득하기 위하여 25명의 시청자들을 대상으로 설문조사를 하여 “DA\_AS02-04.csv”파일에 기록하였다. 파일의 자료는 아래의 네 가지 변수이다.**

**y = 1일 평균 TV 시청시간**

**x1 = 배우자와 동거여부(만약 배우자와 동거 중이면 x1 = 1, 아니면 x1 = 0)**

**x2 = 연령**

**x3 = 교육기간(단위: 년, 초등학교 입학부터 최종학력까지 교육기간)**

**1일 평균 TV 시청시간을 예측하기 위한 모델을 수립하시오. 1일 평균 TV 시청시간에 가장 크게 영향을 미치는 요인과 가장 작게 영향을 미치는 요인은 무엇인가? (1점)**

**(R 코드는 “DA\_AS02.R”파일의 #04 아래에 작성하고, (분석 결과 설명에 필요하다면 그래프를 포함하여) 분석결과는 아래에 작성할 것.)**

**<풀이>**

**- 이 문제에 적용할 분석기법은 무엇이며, 해당 분석기법을 적용하려는 이유는 무엇인가?**

**: 다중회귀분석. 다수의 독립변수가 종속변수에 미치는 영향력의 크기를 파악하고, 이를 통하여 독립변수의 일정한 값에 대응하는 종속변수 값을 예측하는 기법이기 때문이다.**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> df04 <- read.csv("C:/DA2019/exercise2/DA\_AS02-04.csv")

> df04

x1 x2 x3 y

1 1 73 14 0.5

2 1 66 16 0.5

3 0 65 15 0.7

4 0 65 16 0.8

5 1 68 9 0.8

6 1 69 10 0.9

7 1 82 12 1.1

8 1 83 12 1.6

9 1 81 12 1.6

10 0 72 10 2.0

11 1 69 8 2.5

12 0 71 16 2.8

13 0 71 12 2.8

14 0 80 9 3.0

15 0 73 6 3.0

16 0 75 6 3.0

17 0 76 10 3.2

18 0 78 6 3.2

19 1 79 6 3.3

20 0 79 4 3.3

21 1 78 6 3.4

22 0 76 9 3.5

23 0 65 12 3.6

24 0 72 12 3.7

25 0 80 6 3.7

>

> fit04 <- lm(y ~ x1 + x2 + x3, data = df04)

> summary(fit04)

Call:

lm(formula = y ~ x1 + x2 + x3, data = df04)

Residuals:

Min 1Q Median 3Q Max

-1.0306 -0.5707 -0.1094 0.4293 1.4125

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.49526 2.63720 0.567 0.57673

x1 -1.17573 0.31557 -3.726 **0.00125 \*\***

x2 0.03876 0.03193 1.214 0.23824

x3 -0.15228 0.05011 -3.039 **0.00624 \*\***

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7536 on 21 degrees of freedom

Multiple R-squared: 0.6256, **Adjusted R-squared: 0.5722**

F-statistic: 11.7 on 3 and 21 DF, **p-value: 0.0001016**

>

> coefficients(fit04)

(Intercept) x1 x2 x3

1.4952611 -1.1757282 0.0387618 -0.1522776

> confint(fit04, level = 0.95)

2.5 % 97.5 %

(Intercept) -3.98909538 6.97961767

x1 -1.83198259 -0.51947382

x2 -0.02763971 0.10516330

x3 -0.25647864 -0.04807665

> fitted(fit04)

1 2 3 4 5

1.0172570 0.4413691 1.7306132 1.5783356 1.5848363

6 7 8 9 10

1.4713204 1.6706685 1.7094303 1.6319067 2.7633340

11 12 13 14 15

1.7758757 1.8109063 2.4200169 3.2257060 3.4112064

16 17 18 19 20

3.4887300 2.9183812 3.6050154 2.4680490 3.9483325

21 22 23 24 25

2.4292872 3.0706588 2.1874461 2.4587787 3.6825390

> residuals(fit04)

1 2 3 4

-0.51725700 0.05863086 -1.03061320 -0.77833555

5 6 7 8

-0.78483626 -0.57132041 -0.57066846 -0.10943026

9 10 11 12

-0.03190667 -0.76333400 0.72412430 0.98909367

13 14 15 16

0.37998309 -0.22570602 -0.41120638 -0.48872998

17 18 19 20

0.28161881 -0.40501536 0.83195105 -0.64833245

21 22 23 24

0.97071284 0.42934117 1.41255386 1.24122129

25

0.01746104

> anova(fit04)

Analysis of Variance Table

Response: y

Df Sum Sq Mean Sq F value Pr(>F)

x1 1 8.6400 8.6400 15.2122 0.0008244 \*\*\*

x2 1 6.0468 6.0468 10.6465 0.0037181 \*\*

x3 1 5.2459 5.2459 9.2362 0.0062376 \*\*

Residuals 21 11.9273 0.5680

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> vcov(fit04)

(Intercept) x1 x2

(Intercept) 6.95482103 0.143270257 -0.0828688434

x1 0.14327026 0.099581585 -0.0021013300

x2 -0.08286884 -0.002101330 0.0010195076

x3 -0.08566684 -0.002750068 0.0008296198

x3

(Intercept) -0.0856668387

x1 -0.0027500676

x2 0.0008296198

x3 0.0025106046

> influence(fit04)

$hat

1 2 3 4 5

0.14156038 0.23132687 0.17946527 0.20472316 0.22274515

6 7 8 9 10

0.16996131 0.23455004 0.26657531 0.20611481 0.06903742

11 12 13 14 15

0.23037122 0.19905213 0.08095180 0.13497775 0.13742417

16 17 18 19 20

0.12018678 0.08130454 0.12125590 0.16596204 0.18213248

21 22 23 24 25

0.16582510 0.07695567 0.15673568 0.08088625 0.13991878

$coefficients

(Intercept) x1 x2 x3

1 0.15984562 -0.054056827 -0.0011336291 -7.737983e-03

2 0.03465877 0.008079625 -0.0005920769 8.739518e-04

3 -0.62659866 0.076443446 0.0091919904 -1.308625e-02

4 -0.34062890 0.064303659 0.0057328448 -1.452301e-02

5 -1.23028080 -0.133712827 0.0145375489 1.672472e-02

6 -0.63438472 -0.085266962 0.0076687038 7.352577e-03

7 0.95186584 -0.049278950 -0.0112688287 -1.278398e-02

8 0.21226766 -0.009310238 -0.0025230684 -2.776415e-03

9 0.04544961 -0.002805245 -0.0005353423 -6.304588e-04

10 -0.18997747 0.051287097 0.0016863204 1.195581e-03

11 1.15099356 0.125666426 -0.0132313234 -1.836853e-02

12 -0.65126817 -0.108550219 0.0060664098 2.914771e-02

13 0.03139740 -0.028335551 -0.0003846278 2.448411e-03

14 0.20475007 0.022780264 -0.0028291552 -1.515193e-03

15 -0.32851445 0.022349402 0.0029100655 8.427824e-03

16 -0.22070089 0.030152812 0.0013966945 8.197651e-03

17 -0.10787865 -0.023710644 0.0015705428 1.344066e-03

18 0.01862365 0.030133997 -0.0013231159 4.782040e-03

19 0.06577487 0.105984016 0.0009635605 -1.372223e-02

20 -0.09143936 0.047084051 -0.0013827665 1.407488e-02

21 0.24651904 0.127946194 -0.0009647328 -1.770811e-02

22 -0.09353439 -0.033725520 0.0017036688 -1.662116e-05

23 1.59363557 -0.077616663 -0.0195993035 -4.761048e-03

24 -0.09448449 -0.097548297 0.0011677802 9.969793e-03

25 -0.00674452 -0.001477550 0.0001311634 -1.513286e-04

$sigma

1 2 3 4 5

0.7620899 0.7721018 0.7291371 0.7471794 0.7461503

6 7 8 9 10

0.7594094 0.7583483 0.7717179 0.7722051 0.7517117

11 12 13 14 15

0.7498662 0.7316372 0.7671438 0.7703377 0.7658742

16 17 18 19 20

0.7634072 0.7694468 0.7661796 0.7448969 0.7554256

21 22 23 24 25

0.7347686 0.7657544 0.6914161 0.7159289 0.7722351

$wt.res

1 2 3 4

-0.51725700 0.05863086 -1.03061320 -0.77833555

5 6 7 8

-0.78483626 -0.57132041 -0.57066846 -0.10943026

9 10 11 12

-0.03190667 -0.76333400 0.72412430 0.98909367

13 14 15 16

0.37998309 -0.22570602 -0.41120638 -0.48872998

17 18 19 20

0.28161881 -0.40501536 0.83195105 -0.64833245

21 22 23 24

0.97071284 0.42934117 1.41255386 1.24122129

25

0.01746104

> layout(matrix(c(1, 2, 3, 4), 2, 2))

> plot(fit04)

>

> # 잔차의 정규성 검정 추가분석 : 샤피로-윌크스검정

> shapiro.test(rstandard(fit04))

Shapiro-Wilk normality test

data: rstandard(fit04)

W = 0.9367, p-value = 0.1242

>

> # 잔차의 등분산성

> install.packages("cars")

Warning in install.packages :

package ‘cars’ is not available (for R version 3.6.1)

Warning in install.packages :

Perhaps you meant ‘CARS’ ?

> library(car)

> ncvTest(fit04)

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 0.2107125, Df = 1, p = 0.64621

>

> # 다중공선성 검정

> vif(fit04)

x1 x2 x3

**1.051980 1.395372 1.376317**

>

> # gvlma패키지

> install.packages("gvlma")

Error in install.packages : Updating loaded packages

> library(gvlma)

> gvlma.result = gvlma::gvlma(fit04)

> summary(gvlma.result)

Call:

lm(formula = y ~ x1 + x2 + x3, data = df04)

Residuals:

Min 1Q Median 3Q Max

-1.0306 -0.5707 -0.1094 0.4293 1.4125

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.49526 2.63720 0.567 0.57673

x1 -1.17573 0.31557 -3.726 0.00125 \*\*

x2 0.03876 0.03193 1.214 0.23824

x3 -0.15228 0.05011 -3.039 0.00624 \*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7536 on 21 degrees of freedom

Multiple R-squared: 0.6256, Adjusted R-squared: 0.5722

F-statistic: 11.7 on 3 and 21 DF, p-value: 0.0001016

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS

USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:

Level of Significance = 0.05

Call:

gvlma::gvlma(x = fit04)

Value p-value

Global Stat 6.3062 0.17742

Skewness 1.0736 0.30012

Kurtosis 0.8921 0.34491

Link Function 3.4379 0.06372

Heteroscedasticity 0.9026 0.34210

Decision

Global Stat Assumptions acceptable.

Skewness Assumptions acceptable.

Kurtosis Assumptions acceptable.

Link Function Assumptions acceptable.

Heteroscedasticity Assumptions acceptable.

**추정된 회귀식은 Y = 1.49526 – 1.17573X1 + 0.03876 – 0.15228X3 이다.**

**추정된 회귀식 중 독립변수 X1의 p-value = 0.00125 < α = 0.05이므로 귀무가설(회귀선의 모든 독립변수의 기울기(b1, b2, b3)가 0이다.)을 기각하고 대립가설(회귀선의 독립 변수 중 적어도 하나의 기울기가 0이 아니다.)을 채택한다. 즉, X1은 통계적으로 유의하다.**

**추정된 회귀식 중 독립변수 X2의 p-value = 0.23824 > α = 0.05이므로 귀무가설을 기각할 수 없다. 즉, X2는 통계적으로 유의하지 못하다.**

**추정된 회귀식 중 독립변수 X3의 p-value = 0.00624 < α = 0.05이므로 귀무가설을 기각하고 대립가설을 채택한다. 즉, X3는 통계적으로 유의하다.**

**추정된 회귀모형이 조정된 결정계수는 0.5722로 나타났다. 따라서 추정된 회귀식에서 독립변수들은 종속변수 변동의 57.22%를 설명한다고 할 수 있으며, 세 독립변수로 구성된 회귀식의 p-value = 0.0001016 < α = 0.05이므로 이 회귀모형의 통계적 유의성은 있다고 할 수 있다.**

**회귀식의 기본 가정을 검토한 결과는 아래와 같다.**

**- 다중공선성 : X1과 X2, X3의 분산확대인자(VIF)값은 각각 1.051980, 1.395372, 1.376317로 나타나 세 독립변수 간에 다중공선성은 큰 문제가 없는 것으로 나타났다. 즉, 다중공선성에 문제가 없어야 하는 가정을 만족한다고 할 수 있다.**

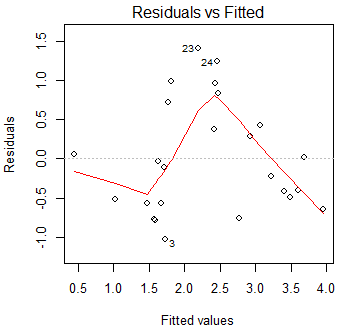


그림 1. 잔차의 산포도

**- 잔차의 독립성 : 산포도의 분포가 일정한 패턴을 보이지 않고 고르게 분포되어 있으므로 잔차의 독립성을 만족한다고 할 수 있다.**

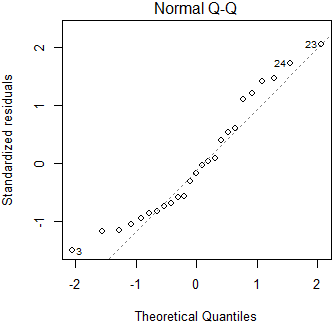


그림 2. 표준화된 잔차의 Normal Q-Q plot

**- 잔차의 정규성 : 누적확률분포와 정규분포의 누적확률분포의 산포를 그린 표준화된 잔차의 Normal Q-Q plot이 대각선 직선의 형태를 띄고 있으며 다른 그래프들은 특별한 형태를 보이지 않고 있어 잔차가 정규분포를 한다고 할 수 있으며, 샤피로-윌크스검정을 통해 추가분석한 결과도 p-value = 0.1242 > α = 0.05이므로 귀무가설(표준화 잔차는 정규분포이다)을 기각할 수 없다. 즉, 잔차의 정규성을 만족한다고 할 수 있다.**

**- 잔차의 등분산성 : fitted values에 따라 오차의 분산이 변하는지 검정해주는 ncvTest 함수를 통한 등분산성 검정 결과, p-value = 0.64621 > α = 0.05 이므로 귀무가설(분산은 동일하다)을 기각할 수 없다. 즉, 잔차의 등분산성을 만족한다고 할 수 있다.**

**최종적으로 선형모형 가정에 대한 전반적 검정과 함께 비대칭도(skewness), 첨도(kurtosis), heteroscedasticity 등에 대한 평가를 수행해주는 gvlma패키지를 이용해도 선형모형 가정에 대한 가정이 모두 만족함을 알 수 있다.**

**결과적으로 다중회귀모형은 회귀식의 기본 가정을 만족하여 올바른 회귀모형이라고 할 수 있다.**

**1일 평균 TV 시청시간에 가장 크게 영향을 미치는 요인은 배우자와 동거여부이며, 가장 작게 영향을 미치는 요인은 교육기간이다.‘연령’의 경우 coefficient가 가장 작지만 p-value = 0.23824 > α = 0.05 이므로 변수의 기여도가 없다고 판단했다.**

**5. “DA\_AS02-05.csv”파일은 A社의 부장급 이하 사원 70명을 대상으로 상급자가 평가한 사교성(x1, 100점 만점), 대학 평균 학점(x2), 경력(x3), 상급자가 평가한 직무성적(x4, 100점 만점), 승진 여부(y, 승진 = 1, 승진누락 = 0)를 기록한 것이다. 승진여부에 가장 큰 영향을 미치는 요인은 무엇인가? 그리고 추가 인원(“DA\_AS02-05test.csv”파일)들의 승진여부를 판단하시오. (2점)**

**(R 코드는 “DA\_AS02.R”파일의 #05 아래에 작성하고, (분석 결과 설명에 필요하다면 그래프를 포함하여) 분석결과는 아래에 작성할 것.)**

**<풀이>**

**- 이 문제에 적용할 분석기법은 무엇이며, 해당 분석기법을 적용하려는 이유는 무엇인가?**

**: 로지스틱 회귀분석.**

**분석에 이용되는 독립변수 x1~x4는 전부 비율척도로 이루어진 독립변수로서 판별분석의 전제조건  
을 만족한다. 로지스틱 회귀분석과 판별분석을 비교했을 때, 비교적 판별분석이 더욱 정확한 결과를 보여주기 때문에 판별분석을 이용하고자 했다. 다만, 판별분석의 기본 가정 중 하나인 ‘독립변수들 간의 등분산성’이 충족되지 않아 판별분석을 활용할 수 없었다.**

**그럼에도 불구하고 판별분석을 수행한 결과, 로지스틱 회귀분석의 분류 결과와 다르지 않았다.**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> #05

> ##### 로지스틱회귀 #####

> df05\_1 <- read.csv("C:/DA2019/exercise2/DA\_AS02-05.csv")

> df05\_1

x1 x2 x3 x4 y

1 23 2.28 3 57 0

2 48 3.40 7 93 1

3 57 2.28 8 76 1

4 40 3.86 7 88 1

5 45 3.16 6 82 1

6 42 3.45 5 85 1

7 58 3.47 6 86 1

8 47 2.45 3 74 0

9 32 2.47 7 64 1

10 36 2.11 3 72 1

11 51 3.48 7 88 1

12 40 2.12 8 77 1

13 65 3.73 8 88 1

14 49 2.70 7 73 0

15 74 3.55 6 85 1

16 46 3.87 9 98 1

17 38 3.07 6 71 1

18 57 3.84 10 88 1

19 37 2.19 7 76 0

20 47 2.67 4 74 1

21 64 3.24 8 76 1

22 63 3.50 7 86 1

23 30 2.29 7 67 0

24 42 2.58 5 80 0

25 50 3.52 7 86 1

26 59 3.41 8 77 1

27 49 2.79 5 58 0

28 28 2.46 8 61 0

29 45 3.76 6 87 1

30 78 3.81 9 93 1

31 43 3.85 7 84 0

32 55 2.54 8 73 1

33 46 3.29 5 86 1

34 52 3.29 7 77 1

35 42 2.55 4 88 1

36 63 3.34 6 93 1

37 36 3.20 6 77 1

38 44 3.67 8 57 0

39 35 3.89 7 64 0

40 69 3.74 7 87 1

41 52 2.77 4 56 1

42 47 2.97 4 74 0

43 44 3.16 4 73 1

44 71 2.58 4 70 1

45 43 3.62 5 79 0

46 48 3.12 3 61 0

47 55 3.20 6 84 1

48 51 2.22 7 82 1

49 46 3.50 5 69 0

50 32 3.90 7 96 1

51 37 3.43 7 77 1

52 47 2.09 4 62 0

53 47 2.71 8 53 0

54 50 2.95 4 52 1

55 42 3.95 8 64 0

56 68 3.77 9 74 1

57 68 3.36 4 75 0

58 50 2.82 5 72 1

59 46 3.44 6 71 0

60 47 3.33 4 75 0

61 38 2.06 3 74 0

62 47 3.79 7 70 0

63 42 2.08 7 57 0

64 32 2.01 8 76 1

65 63 2.38 7 77 1

66 71 3.65 7 67 0

67 72 3.23 5 64 1

68 41 2.68 4 56 0

69 54 2.53 5 80 1

70 49 2.19 4 85 0

> # 1단계 : 모든 독립변수를 적용한 로지스틱 회귀모형

> fit05\_full <-glm(y ~ x1 + x2 + x3 + x4, family = binomial(), data = df05\_1)

> summary(fit05\_full)

Call:

glm(formula = y ~ x1 + x2 + x3 + x4, family = binomial(), data = df05\_1)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.9823 -0.7873 0.3636 0.6837 2.2582

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -10.51236 2.92159 -3.598 0.000320 \*\*\*

x1 0.05696 0.02961 1.923 0.054430 .

x2 -0.65171 0.58540 -1.113 0.265593

x3 0.26544 0.19443 1.365 0.172189

x4 0.11648 0.03398 3.428 **0.000608 \*\*\***

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 93.351 on 69 degrees of freedom

Residual deviance: 67.012 on 65 degrees of freedom

AIC: 77.012

Number of Fisher Scoring iterations: 5

> confint(fit05\_full)

Waiting for profiling to be done...

2.5 % 97.5 %

(Intercept) -16.914533096 -5.3118605

x1 0.001763715 0.1202457

x2 -1.875272322 0.4513948

x3 -0.104683598 0.6684865

x4 0.055275120 0.1905347

> layout(matrix(c(1, 2, 3, 4), 2, 2))

> plot(fit05\_full)

> # 2단계 : 최적의 로지스틱 회귀모형 설정

> ## (1). 수동으로 조정해보기 : full model에서 p-value < 0.05 인 독립변수만 적용하기

> fit05\_reduced1 <- glm(y ~ x4, family = binomial(), data = df05\_1)

> summary(fit05\_reduced1)

Call:

glm(formula = y ~ x4, family = binomial(), data = df05\_1)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.9615 -0.8530 0.4745 0.8609 2.1518

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -8.4780 2.3791 -3.564 0.000366 \*\*\*

x4 0.1205 0.0321 3.754 **0.000174 \*\*\***

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 93.351 on 69 degrees of freedom

Residual deviance: 72.977 on 68 degrees of freedom

**AIC: 76.977**

Number of Fisher Scoring iterations: 4

> ## (2). 자동으로 조정해보기 : step 함수를 이용하여 최적의 독립변수 적용하기

> fit05\_reduced2 = step(fit05\_full)

Start: AIC=77.01

y ~ x1 + x2 + x3 + x4

Df Deviance AIC

- x2 1 68.318 76.318

- x3 1 68.966 76.966

<none> 67.012 77.012

- x1 1 71.111 79.111

- x4 1 82.660 90.660

Step: AIC=76.32

y ~ x1 + x3 + x4

Df Deviance AIC

- x3 1 69.654 75.654

<none> 68.318 76.318

- x1 1 71.502 77.502

- x4 1 82.675 88.675

Step: AIC=75.65

y ~ x1 + x4

Df Deviance AIC

<none> 69.654 75.654

- x1 1 72.977 76.977

- x4 1 86.007 90.007

> summary(fit05\_reduced2)

Call:

glm(formula = y ~ x1 + x4, family = binomial(), data = df05\_1)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.9409 -0.7673 0.4019 0.7203 2.0355

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -10.19544 2.70415 -3.770 0.000163 \*\*\*

x1 0.04844 0.02792 1.735 0.082699 .

x4 0.11224 0.03271 3.431 **0.000600 \*\*\***

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 93.351 on 69 degrees of freedom

Residual deviance: 69.654 on 67 degrees of freedom

**AIC: 75.654**

Number of Fisher Scoring iterations: 4

> ## step 함수를 이용하여 얻은 fit05\_reduced2 모델의 AIC가 수동으로 조정한 fit05\_reduced1 모델의 AIC보다 작기 때문에 fit05\_reduced2를 이용하고자 함.

> # 3단계 : fit05\_full 과 fit05\_reduced2 간의 차이검정 실시 후 log(odds)구하기

> anova(fit05\_reduced2, fit05\_full, test ="Chisq")

Analysis of Deviance Table

Model 1: y ~ x1 + x4

Model 2: y ~ x1 + x2 + x3 + x4

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 67 69.654

2 65 67.012 2 2.6426 **0.2668**

> # p-value = 0.2668 > a = 0.05 이므로, 전체 독립변수가 투입된 모델과 유의한 독립변수만 투입된 모델 간에 차이가 없음. 따라서 fit05\_reduced2를 계속 이용.

> coef(fit05\_reduced2)

(Intercept) x1 x4

-10.19543827 0.04844199 0.11223917

> exp(coef(fit05\_reduced2))

(Intercept) x1 x4

3.734027e-05 **1.049634e+00** **1.118780e+00**

> # 다른 변수의 값을 일정하게 놓고, 사교성이 1단위 증가 시 승진할 확률은 승진하지 않을 확률보다 약 1.0496배 증가

> # 다른 변수의 값을 일정하게 놓고, 직무성적이 1단위 증가 시 승진할 확률은 승진하지 않을 확률보다 약 1.1187배 증가

> # 4단계 : test 데이터로 예측하기

> testdata\_1 <-read.csv("C:/DA2019/exercise2/DA\_AS02-05test.csv")

> testdata\_1

x1 x2 x3 x4

1 43 3.30 2 55

2 44 2.95 5 90

3 79 3.19 4 74

4 45 3.55 7 85

5 80 4.12 8 80

> testdata\_1$prob <-predict(fit05\_reduced2, newdata = testdata\_1, type = "response")

> testdata\_1

x1 x2 x3 x4 prob

1 43 3.30 2 55 **0.1257272**

2 44 2.95 5 90 **0.8846820**

3 79 3.19 4 74 **0.8740418**

4 45 3.55 7 85 **0.8212414**

5 80 4.12 8 80 **0.9345669**

> # 5단계 : 과대산포 검정

> ## (1) 교수님께서 알려주신 과대산포 검정 방법 : 1.039616 (> 1) 의 결과값을 가지므로 과대산포가 있다고 판단됨.

> deviance(fit05\_reduced2)/df.residual(fit05\_reduced2)

[1] 1.039616

> ## (2) 카이제곱을 이용한 과대산포 검정 방법 : 0.4573546 (> 0.05) 의 결과값을 가지므로 과대산포가 없다고 판단됨.

> fit05\_reduced2.od <- glm(y ~ x1 + x4, family = quasibinomial(), data = df05\_1)

> pchisq(summary(fit05\_reduced.od)$dispersion \* fit05\_reduced$df.residual, fit05\_reduced$df.residual, lower = F)

[1] 0.4573546

> # 6단계 : 로지스틱회귀모형 시각화

> library(ggplot2)

> ggplot(df05\_1, aes(x = x1 + x4, y = y)) + geom\_point() + stat\_smooth(method = "glm", method.args = list(family = "binomial"), color = "black", se = FALSE)

> # 7단계 : 판별분석이 가능한지 알아보기 위해 독립변수에 대한 기본 가정 파악

> ## (1) 독립변수의 정규성 검정

> library(car)

필요한 패키지를 로딩중입니다: carData

> scatterplotMatrix(df05\_1[1:5])

> # x1 ~ x4 모두 peak가 있기 때문에 정규분포를 따른다고 할 수 있다.

> # 또한, 각 독립변수의 관측치가 70개이므로 중심극한정리에 의해 정규분포에 근접

> ## (2) 다중공선성 검정

> vif\_test <- lm(y ~ x1 + x2 + x3 + x4, data = df05\_1)

> vif(vif\_test)

x1 x2 x3 x4

**1.140741 1.351017 1.179971 1.227988**

> ## 3. 독립변수들간의 등분산 검정

> install.packages("lawstat")

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/lawstat\_3.3.zip'

Content type 'application/zip' length 264440 bytes (258 KB)

downloaded 258 KB

package ‘lawstat’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\user\AppData\Local\Temp\RtmpiE1jWd\downloaded\_packages

> library(lawstat)

다음의 패키지를 부착합니다: ‘lawstat’

The following object is masked from ‘package:car’:

levene.test

> tmp <- c(df05\_1$x1, df05\_1$x2,df05\_1$x3, df05\_1$x4)

> group <- c(rep(1, 70), rep(2, 70), rep(3, 70), rep(4, 70))

> dat<-data.frame(group,tmp)

> levene.test(tmp,group)

Modified robust Brown-Forsythe Levene-type test based on the absolute deviations from the median

data: tmp

Test Statistic = 52.863, **p-value < 2.2e-16**

> bartlett.test(tmp~group,dat)

Bartlett test of homogeneity of variances

data: tmp by group

Bartlett's K-squared = 478.62, df = 3, **p-value < 2.2e-16**

> # 등분산이 아니다. 즉, 판별분석의 기본 가정에 위배되어 판별분석은 불가능할 것으로 보인다.

> ##### 그럼에도 불구하고 판별분석 시도 #####

> df05\_2 <- read.csv("C:/DA2019/exercise2/DA\_AS02-05.csv")

> head(df05\_2)

x1 x2 x3 x4 y

1 23 2.28 3 57 0

2 48 3.40 7 93 1

3 57 2.28 8 76 1

4 40 3.86 7 88 1

5 45 3.16 6 82 1

6 42 3.45 5 85 1

> df05\_2$y[df05\_2$y == 1] <- 2

> df05\_2$y[df05\_2$y == 0] <- 1

> head(df05\_2)

x1 x2 x3 x4 y

1 23 2.28 3 57 1

2 48 3.40 7 93 2

3 57 2.28 8 76 2

4 40 3.86 7 88 2

5 45 3.16 6 82 2

6 42 3.45 5 85 2

> install.packages("MASS")

Error in install.packages : Updating loaded packages

> install.packages("MASS")

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/MASS\_7.3-51.4.zip'

Content type 'application/zip' length 1185835 bytes (1.1 MB)

downloaded 1.1 MB

package ‘MASS’ successfully unpacked and MD5 sums checked

Warning in install.packages :

cannot remove prior installation of package ‘MASS’

Warning in install.packages :

problem copying C:\Program Files\R\R-3.6.1\library\00LOCK\MASS\libs\x64\MASS.dll to C:\Program Files\R\R-3.6.1\library\MASS\libs\x64\MASS.dll: Permission denied

Warning in install.packages :

restored ‘MASS’

The downloaded binary packages are in

C:\Users\user\AppData\Local\Temp\RtmpiE1jWd\downloaded\_packages

> library(MASS)

> Mda.lda <-lda(y ~ ., data = df05\_2)

> Mda.lda

Call:

lda(y ~ ., data = df05\_2)

Prior probabilities of groups:

1 2

0.3857143 0.6142857

Group means:

x1 x2 x3 x4

1 44.29630 2.951481 5.555556 68.25926

2 51.83721 3.131860 6.441860 79.86047

Coefficients of linear discriminants:

LD1

x1 0.03915148

x2 -0.55375987

x3 0.16730711

x4 0.09137212

> Mda.lda.values <- predict(Mda.lda)

> ldahist(data = Mda.lda.values$x[,1], g = df05\_2$y)

> layout(matrix(c(1), 1, 1))

> plot(Mda.lda.values$x[,1])

> text(Mda.lda.values$x[,1], cex = 0.7, pos = 4, col = "red")

> testdata\_2 <-read.csv("C:/DA2019/exercise2/DA\_AS02-05test.csv")

> testdata\_2

x1 x2 x3 x4

1 43 3.30 2 55

2 44 2.95 5 90

3 79 3.19 4 74

4 45 3.55 7 85

5 80 4.12 8 80

> predict(Mda.lda, newdata = testdata\_2)

$class

[1] **1 2 2 2 2**

Levels: 1 2

$posterior

1 2

1 **0.96762578** 0.03237422

2 0.10688120 **0.89311880**

3 0.17179485 **0.82820515**

4 0.17653686 **0.82346314**

5 0.06824671 **0.93175329**

$x

LD1

1 -2.9123940

2 1.0205189

3 0.6286572

4 0.6051681

5 1.3702732

**Ⅰ. 최적의 로지스틱 회귀모형 결정 과정**

**로지스틱 회귀분석의 결과, full 모델의 경우 ‘직무성적(x4)’만 유의한 변수로 밝혀졌다. 이후 p-value가 유의수준 0.05 아래인 독립변수 x4만 적용한 reduced1 모델을 수동으로 만들었고, step 함수를 적용하여 reduced2 모델을 자동으로 만들었는데 이때는 독립변수 x1과 x4가 적용되었다.**

**분석 결과, reduced2 모델의 AIC가 더 낮으므로 reduced2 모델을 이용하기로 결정했고, full모델과 redcued2 모델을 anova분석한 결과 두 모델의 차이가 없다고 밝혀져 reduced2 모델을 계속 이용하기로 결정했다.**

**🡪 reduced2 모델에 기반한 추정된 회귀식은 y = -10.19544 + 0.04844x1 + 0.11224x4이다.**

**Ⅱ. 로지스틱 회귀모형에서 비표준화계수 추출 과정**

**reduced2 모델에서 비표준화계수 추출을 한 결과, 다른 변수의 값을 일정하게 놓았다고 가정했을 때 사교성이 1단위 증가할 경우 승진할 확률은 승진하지 않을 확률보다 약 1.0496배 증가한다.**

**또한 다른 변수의 값을 일정하게 놓았다고 가정했을 때 직무성적이 1단위 증가할 경우 승진할 확률은 승진하지 않을 확률보다 약 1.1186배 증가한다.**

**Ⅲ. 로지스틱 회귀모형의 과대산포 검정**

**교안에 나와있는 과대산포 검정 방법을 이용할 경우, 1.039616 ( > 1) 이라는 결과값이 나와 과대산포가 있다고 판단된다.**

**다만, 카이제곱을 이용한 과대산포 검정 방법을 이용할 경우, 0.4573546 ( > 0.05) 이라는 결과값이 나와 과대산포가 없다고 판단된다.**

**교안에 나와있는 검정 방법에서 1이랑 0.04 정도의 차이만 보이고, 카이제곱을 이용한 과대산포 검정 방법에서 0.05와 상당히 큰 차이를 보이므로 과대산포가 크게 없다고 판단했다.**

**Ⅳ. 로지스틱회귀모형 시각화**

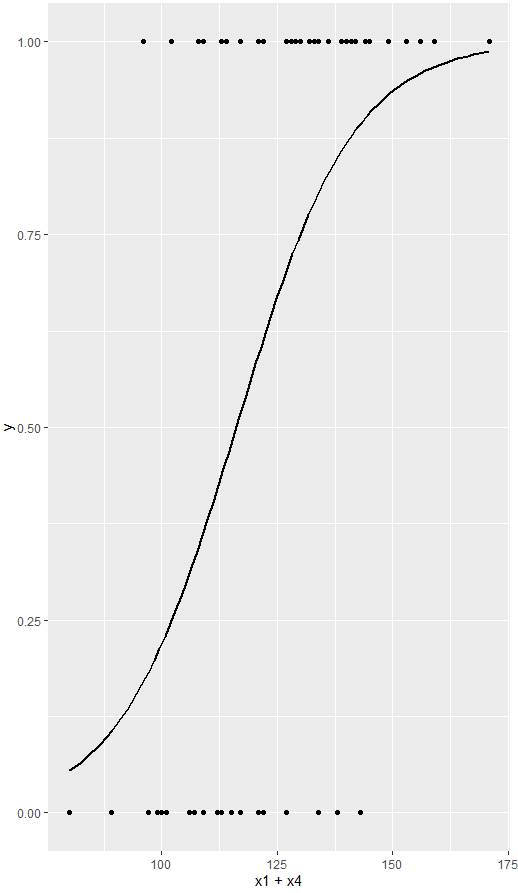


그림 3. 로지스틱 회귀모형 'reduced2 '의 시각화 결과

**Ⅴ. 판별분석이 가능한지 알아보기 위한 독립변수에 대한 기본 가정 검정**

**(1) 독립변수의 정규성 검정**

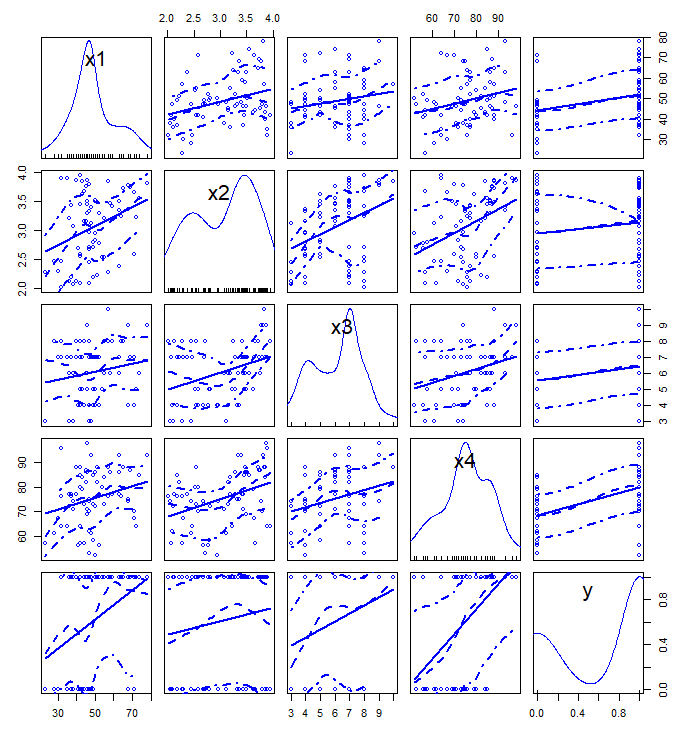


그림 4. 각 변수의 산포도 그래프

**x1~x4 모두 peak가 존재하며, 각 독립변수의 관측치가 70개이므로 중심극한정리에 의해 정규분포를 따른다고 할 수 있다.**

**(2) 다중공선성 검정**

**다중공선성 검정 결과, x1~x4의 다중공선성이 각각 1.140741, 1.351017, 1.179971, 1.227988 으로 밝혀져 다중공선성에도 문제가 없다고 할 수 있다.**

**(3) 등분산 검정**

**Levene 검정과 Bartlett 검정을 모두 이용한 결과, 독립변수 간의 분산은 동일하지 않다는 결론이 나왔다. 따라서 다른 기본 가정을 만족했더라도 ‘등분산’의 기본 가정이 충족되지 않았기 때문에 판별분석을 쓰기는 힘들다고 판단했다. 그럼에도 불구하고 혹시나 하는 마음에 판별분석을 시행해봤다.**

**(판별분석을 할 때 lda 함수에 로지스틱 회귀모형에서 적용한 x1과 x4만 독립변수로 넣는 것보다는 모든 독립변수를 투입하는 것이 더 좋은 결과를 보여준다는 결론을 얻었었다. 이러한 시행착오 과정은 소스코드에서 제외했다.)**

**판별분석 결과, 각 변수별 선형함수로 다음의 선형판별식을 구할 수 있었다.**

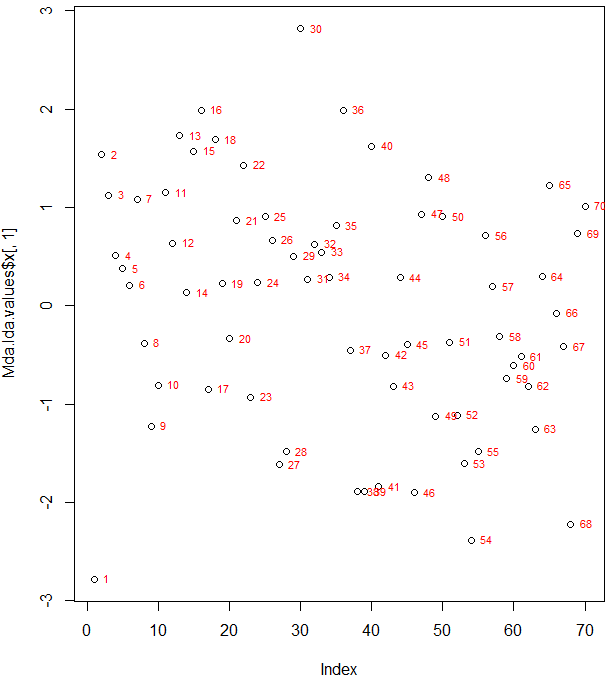
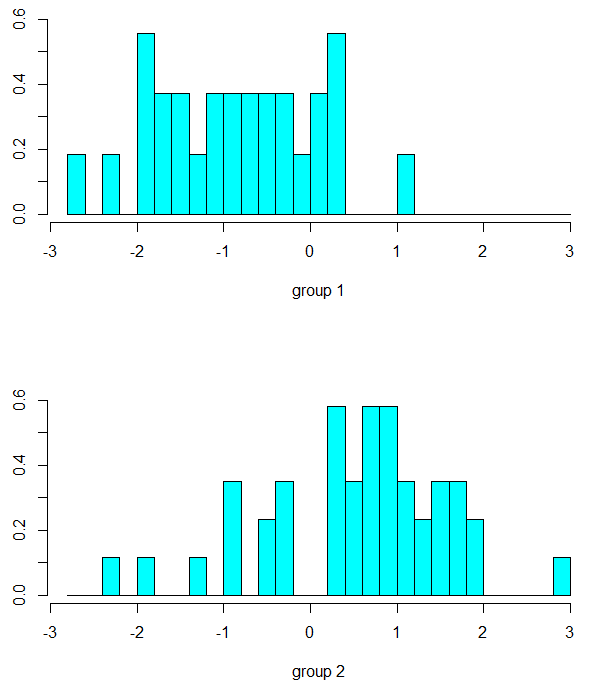
**D = 0.03915148x1 - 0.55375987x2 + 0.16730711x3 + 0.09137212x4**

그림 5. 집단 별 선형판별함수 누적 히스토그램

그림 6. 개체 별 산포도

**집단 별 선형판별함수 누적 히스토그램을 확인해본 결과, 잘못 판별되는 관측치가 약간 있음을 알 수 있다. 개체 별 산포도를 확인해본 결과, 각 관측치가 명확하게 구분될 수 있는 surface를 찾기 힘들어서 그런 결과가 나왔다고 판단한다.**

**승진여부에 가장 큰 영향을 미치는 요인은 ‘직무성적(x4)’인 것으로 밝혀졌다. Full model에서도, step 함수를 거친 다음의 reduced 모델에서도 유일하게 p-value가 유의수준 0.05보다 낮았으며 유의한 독립변수들 중 회귀계수의 절댓값도 가장 크기 때문이다.**

**추가 인원들의 승진여부를 판단하기 위해 로지스틱 회귀분석과 판별분석 모두 이용했는데 동일한 분류 결과를 보였으며, 그 결과는 다음과 같다.**

**1번 사원 : 승진누락(n)  
2번 사원 ~ 5번 사원 : 승진(y)**

**6. “DA\_AS02-06.csv”파일은 스마트워치 브랜드 대한 고객의 인식을 조사한 자료이다. 총 13개 설문에 대해 스마트워치를 구매한 고객 210명으로부터 설문 조사한 결과이다. 응답자는 각 질문에 대해 1(매우 그렇지 않다.)부터 9(매우 그렇다.)로 응답을 하였다. 질문의 내용은 다음과 같다.**

**x1: 가격이 저렴하다.**

**x2: 제조사가 다양한 이벤트를 제공한다.**

**x3: 디자인이 다양하다.**

**x4: A/S가 신속하게 이루어진다.**

**x5: 언제 어디서든지 쉽게 구입할 수 있다.**

**x6: 다양한 방법으로 구입할 수 있다.**

**x7: 내구성이 높다.**

**x8: 액정 화질이 좋다.**

**x9: 다양한 시계줄이 제공된다.**

**x10: 판매, 유통하는 기업의 이미지가 좋다.**

**x11: 제품 자체의 이미지가 좋다.**

**x12: 제품의 색상이 마음에 든다.**

**x13: 제품의 무게가 가볍다.**

**이 열 세 개 문항을 서로 연관성이 높은 요인들로 축약하시오. (요인분석은 주성분분석과 공통요인분석을 모두 수행하시오.) (2점)**

**(R 코드는 “DA\_AS02.R”파일의 #06 아래에 작성하고, (분석 결과 설명에 필요하다면 그래프를 포함하여) 분석결과는 아래에 작성할 것.)1**

**<풀이>**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> #06

> ##### 주성분분석 #####

> df06 <- read.csv("C:/DA2019/exercise2/DA\_AS02-06.csv")

> head(df06)

x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

1 5 5 5 5 5 5 5 5 5 5 5 5 5

2 9 1 1 1 9 5 5 5 5 5 9 7 8

3 7 5 5 5 5 5 6 5 5 1 1 1 1

4 9 9 9 5 5 5 5 5 5 5 5 5 1

5 9 1 1 1 1 9 9 1 1 1 1 5 5

6 4 3 8 8 9 9 8 8 8 6 7 8 9

> mydata <-na.omit(df06)

> head(mydata)

x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

1 5 5 5 5 5 5 5 5 5 5 5 5 5

2 9 1 1 1 9 5 5 5 5 5 9 7 8

3 7 5 5 5 5 5 6 5 5 1 1 1 1

4 9 9 9 5 5 5 5 5 5 5 5 5 1

5 9 1 1 1 1 9 9 1 1 1 1 5 5

6 4 3 8 8 9 9 8 8 8 6 7 8 9

> cor(mydata)

x1 x2 x3 x4 x5 x6 x7 x8

x1 1.00000000 0.07427578 -0.05333158 -0.04260321 -0.17427929 -0.19513271 -0.09113597 -0.04707152

x2 0.07427578 1.00000000 0.32782477 0.30653565 0.04777474 0.02549129 0.07049537 0.22522176

x3 -0.05333158 0.32782477 1.00000000 0.25535632 0.21847409 0.23965277 0.24875131 0.27415774

x4 -0.04260321 0.30653565 0.25535632 1.00000000 0.27596774 0.16767499 0.42887316 0.41036125

x5 -0.17427929 0.04777474 0.21847409 0.27596774 1.00000000 0.66090425 0.40070143 0.35387622

x6 -0.19513271 0.02549129 0.23965277 0.16767499 0.66090425 1.00000000 0.45557138 0.32772490

x7 -0.09113597 0.07049537 0.24875131 0.42887316 0.40070143 0.45557138 1.00000000 0.63669110

x8 -0.04707152 0.22522176 0.27415774 0.41036125 0.35387622 0.32772490 0.63669110 1.00000000

x9 -0.02179109 0.33713589 0.33250637 0.35077096 0.20998501 0.22714885 0.32810994 0.53929503

x10 -0.13411727 0.12816025 0.25073197 0.29914333 0.35497679 0.33475502 0.39430122 0.52930900

x11 -0.01919338 0.12764913 0.26461207 0.31947469 0.29529931 0.23584622 0.38978233 0.48207204

x12 -0.01212293 0.15349115 0.37509751 0.27089128 0.22527761 0.20410855 0.27883110 0.30624531

x13 -0.08123592 0.25126478 0.40494239 0.34440038 0.26429447 0.27573048 0.27127606 0.34811724

x9 x10 x11 x12 x13

x1 -0.02179109 -0.1341173 -0.01919338 -0.01212293 -0.08123592

x2 0.33713589 0.1281603 0.12764913 0.15349115 0.25126478

x3 0.33250637 0.2507320 0.26461207 0.37509751 0.40494239

x4 0.35077096 0.2991433 0.31947469 0.27089128 0.34440038

x5 0.20998501 0.3549768 0.29529931 0.22527761 0.26429447

x6 0.22714885 0.3347550 0.23584622 0.20410855 0.27573048

x7 0.32810994 0.3943012 0.38978233 0.27883110 0.27127606

x8 0.53929503 0.5293090 0.48207204 0.30624531 0.34811724

x9 1.00000000 0.4845413 0.52984131 0.33783837 0.43567319

x10 0.48454130 1.0000000 0.63679861 0.38201903 0.33886097

x11 0.52984131 0.6367986 1.00000000 0.51930973 0.42333609

x12 0.33783837 0.3820190 0.51930973 1.00000000 0.57344784

x13 0.43567319 0.3388610 0.42333609 0.57344784 1.00000000

> fit <-princomp(mydata, cor = TRUE)

> summary(fit)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7

Standard deviation 2.178639 1.2396944 1.03794466 1.02454127 0.94221592 0.86640117 0.80062055

Proportion of Variance 0.365113 0.1182186 0.08287147 0.08074499 0.06829007 0.05774238 0.04930717

Cumulative Proportion 0.365113 0.4833316 0.56620310 **0.64694809 0.71523815 0.77298054** 0.82228771

Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13

Standard deviation 0.73826811 0.69169088 0.61242700 0.58929388 0.56569240 0.4944104

Proportion of Variance 0.04192614 0.03680279 0.02885129 0.02671287 0.02461599 0.0188032

Cumulative Proportion 0.86421385 0.90101664 0.92986794 0.95658080 0.98119680 1.0000000

> # 4개(누적분산 64.69%) - 6개(누적분산 77.30%) 가 적당.

> biplot(fit)

> e\_value = eigen(cor(df06))

> e\_value # **4개일 때 1.0496848, 5개일 때는 0.8877708. 고유값에 의하면 4개가 적당.**

eigen() decomposition

$values

[1] 4.7464690 1.5368423 1.0773291 **1.0496848** **0.8877708** 0.7506510 0.6409933 0.5450398 0.4784363

[10] 0.3750668 0.3472673 0.3200079 0.2444417

$vectors

[,1] [,2] [,3] [,4] [,5] [,6] [,7]

[1,] 0.06346489 -0.36156120 0.32115883 -0.22642840 0.81834333 -0.0968500641 -0.042068389

[2,] -0.15678930 -0.42811177 -0.34910715 -0.43109160 -0.13851446 -0.3241471540 -0.252094785

[3,] -0.24507399 -0.20173101 -0.48419701 0.02714145 0.09795844 -0.1024586049 0.631741046

[4,] -0.26685531 -0.11313593 -0.02727232 -0.39135324 -0.13870994 0.5805795822 -0.362131330

[5,] -0.26283209 0.44855265 -0.20491371 -0.08198975 0.26228488 -0.1467816109 -0.341971678

[6,] -0.25341721 0.48193178 -0.24415327 -0.05459490 0.27513616 -0.2314762338 -0.073685653

[7,] -0.30889629 0.23790595 0.18834376 -0.29338537 0.06852121 0.3332101351 0.391347637

[8,] -0.34344670 0.03589420 0.28896417 -0.26621559 -0.09066086 0.0009238242 0.282816419

[9,] -0.31835409 -0.22346237 0.15042978 -0.05156418 -0.19947831 -0.3469138031 -0.007890168

[10,] -0.32826560 0.05505285 0.31189453 0.17767521 -0.17664298 -0.2930510178 -0.082770785

[11,] -0.33335344 -0.09655101 0.34543534 0.29405788 -0.02158643 -0.1057494381 -0.120887774

[12,] -0.28582624 -0.20105750 -0.08585951 0.48034843 0.21593303 0.2900198706 -0.043806839

[13,] -0.30251277 -0.19606338 -0.26946581 0.30152209 0.08244052 0.2242761249 -0.159637610

[,8] [,9] [,10] [,11] [,12] [,13]

[1,] 0.0336225719 0.104509003 -0.11754254 0.03966553 0.04341192 -0.025895706

[2,] -0.0005803048 -0.524625698 0.03464236 0.08436024 -0.10513919 -0.044819545

[3,] 0.4060512803 0.266766144 -0.03975803 -0.05668286 -0.05203246 0.036424197

[4,] 0.3182751338 0.335210830 0.00939908 0.09834503 0.15043597 0.169052482

[5,] 0.1475620165 0.074888867 0.08165288 -0.55941196 -0.14974383 -0.322970580

[6,] -0.1543764357 0.005428634 0.05468240 0.50318199 0.18707449 0.437239782

[7,] -0.1209587723 -0.263063884 0.11196569 0.29660716 -0.19079952 -0.486629383

[8,] -0.2562869294 -0.168335113 -0.19429187 -0.51468761 0.09542458 0.482080236

[9,] -0.3826240237 0.499282300 0.40303269 0.02972367 0.21712293 -0.238912223

[10,] 0.3813003938 -0.088802880 -0.53098441 0.16455342 0.33818999 -0.244460363

[11,] 0.2591199011 -0.022714603 0.26738838 0.11583621 -0.64494465 0.281012014

[12,] 0.0150328537 -0.388742269 0.36927788 -0.12462692 0.45145760 -0.007524305

[13,] -0.4976058599 0.135398921 -0.52070060 0.05001079 -0.27840619 -0.082781261

> plot(fit, type = "lines") # 위 고유값을 scree 도표로 표시

> loadings(fit)

Loadings:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13

x1 0.362 0.321 0.226 0.818 0.105 0.118

x2 -0.157 0.428 -0.349 0.431 -0.139 0.324 0.252 -0.525 -0.105

x3 -0.245 0.202 -0.484 0.102 -0.632 0.406 0.267

x4 -0.267 0.113 0.391 -0.139 -0.581 0.362 0.318 0.335 0.150 -0.169

x5 -0.263 -0.449 -0.205 0.262 0.147 0.342 0.148 -0.559 -0.150 0.323

x6 -0.253 -0.482 -0.244 0.275 0.231 -0.154 0.503 0.187 -0.437

x7 -0.309 -0.238 0.188 0.293 -0.333 -0.391 -0.121 -0.263 -0.112 0.297 -0.191 0.487

x8 -0.343 0.289 0.266 -0.283 -0.256 -0.168 0.194 -0.515 -0.482

x9 -0.318 0.223 0.150 -0.199 0.347 -0.383 0.499 -0.403 0.217 0.239

x10 -0.328 0.312 -0.178 -0.177 0.293 0.381 0.531 0.165 0.338 0.244

x11 -0.333 0.345 -0.294 0.106 0.121 0.259 -0.267 0.116 -0.645 -0.281

x12 -0.286 0.201 -0.480 0.216 -0.290 -0.389 -0.369 -0.125 0.451

x13 -0.303 0.196 -0.269 -0.302 -0.224 0.160 -0.498 0.135 0.521 -0.278

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11

SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

Proportion Var 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077

Cumulative Var 0.077 0.154 0.231 0.308 0.385 0.462 0.538 0.615 0.692 0.769 0.846

Comp.12 Comp.13

SS loadings 1.000 1.000

Proportion Var 0.077 0.077

Cumulative Var 0.923 1.000

> fit$scores

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7

1 0.95603019 0.524416254 -0.620787203 0.288221121 -0.7367251678 0.052497956 0.182217730

2 0.47806116 -0.317502135 1.681211916 -2.211649243 2.2463770638 0.357640415 1.559839340

3 3.39482455 -0.132052104 -0.733428042 3.296886338 -0.0440647642 0.284454076 -0.811045064

4 0.88146460 2.223258809 -1.073492819 2.222258004 0.8392193587 1.591909982 -0.835558922

5 4.49996729 -1.280776945 -0.028879402 0.096573667 2.6141790732 -1.614910969 -0.854340800

6 -3.81414053 -0.682619601 -1.251114414 -0.287929348 -0.0286496392 -0.944289952 -0.279302022

7 0.38475404 0.119171525 -0.524695150 0.585753895 -1.6212604876 -0.936920200 1.769417458

8 -1.12924027 -0.705724623 0.313693754 -0.882674416 -1.0356775749 -1.168750769 -1.392476115

9 -0.61999189 0.014402152 0.574513887 0.108700404 -1.3699642343 -1.460646410 0.453515310

10 -0.90274218 -2.026150575 -0.962195773 0.813852923 0.3802131617 0.097304373 -0.820713890

11 3.00897927 -2.668129912 0.302420654 1.430579075 1.2469291718 -0.145677625 -0.385732739

12 -1.19784232 -3.214313194 1.496499098 -0.598722766 0.6981810229 0.338573210 -1.026971103

13 0.91263631 -1.147414181 0.159155827 -0.070075650 -0.1075272663 0.054312969 0.155529231

14 3.29499554 1.651323421 -1.110279613 -1.284025227 -0.9138040157 -0.250477692 1.157241288

15 -0.65740278 0.467313321 0.931329193 -1.251949742 1.2609274175 -0.658867273 -0.627107340

16 5.76804109 0.741127915 0.347939059 -0.154135530 -1.9813084251 -1.849244900 1.174455531

17 0.14511081 -0.382395925 -0.716723353 0.276663358 -0.4037155613 0.608998497 0.912094689

18 -3.29205555 -2.511654070 0.682739798 0.907535606 -1.8638308034 -0.501187789 0.059356294

19 -1.20433473 -3.410719594 2.611101480 -0.456521417 -2.7369971547 2.195298419 -0.184679069

20 -1.76318679 0.084877885 1.133131671 0.465058454 1.0643467399 0.281310705 -0.086629684

21 -0.46521276 0.615074654 1.660353175 1.891971927 0.9086891140 -0.206170320 -0.975791761

22 -0.43934249 3.117581071 -1.280108544 1.730128658 0.5594932372 1.915417519 -0.496364218

23 -0.86495323 1.634125375 -0.044739312 -0.259486116 0.8040863382 0.913092636 0.536042156

24 -1.75167436 0.569579006 -0.053156011 0.376691037 0.3516503742 0.400997006 1.137188928

25 1.73386019 -1.099192516 3.106107074 0.921123175 -0.4154456309 0.338920897 0.423095652

26 4.53464350 -2.347471796 -1.622059773 -0.377309977 0.3598861189 0.218535072 -0.424660002

27 -3.30361056 -1.110746891 -0.553094834 -0.522079237 0.1381534694 -0.705089282 -0.065565334

28 -1.30632644 0.539551462 -1.168820036 0.041369215 -0.0021915787 0.132581489 0.277762430

29 -0.60139863 -1.217361016 0.151676419 -2.230557985 0.5071937901 1.039111462 0.243815320

30 2.99570718 0.478065572 -0.750621012 -0.426122469 1.4433186246 -0.907307646 0.461257771

31 -1.07998616 0.876195729 0.162253629 -0.054537897 0.4201695465 -0.080756299 1.199716842

32 6.83023675 1.290699822 0.734396500 -1.028805194 0.9023565388 -0.130019474 0.153474303

33 -1.07195332 1.083460462 -0.584112473 -0.969950349 -0.5495258050 0.101233654 0.370306141

34 0.68980151 -1.239999674 2.031537450 0.581346888 -1.1445660867 0.312471843 -1.182891078

35 0.70558768 0.097007826 -0.816041533 0.366345977 -0.4868040804 0.192360462 0.508069282

36 -0.31386750 -1.268350962 0.158818110 -1.612562609 -0.3469772515 1.129753462 0.280248330

37 0.13556495 -2.419418753 -0.153106013 -0.597507104 0.3753882476 0.277062941 0.898026847

38 -0.25106948 0.183746421 -0.533602592 0.155150230 -0.4222534959 0.095679051 -0.054744521

39 -0.37905112 -1.392092938 0.459582552 -0.327963776 0.2084666372 -1.375280393 1.917155106

40 0.50612404 0.087699796 -1.109113562 0.611736294 -0.4275395557 0.466734021 0.671275017

41 -1.29562225 -0.257702667 -0.155520453 0.483706424 -0.2490718919 -0.584918628 0.412928576

42 0.91008287 -0.345536082 -0.513629210 -0.258622136 -0.4336493470 0.846875595 -0.081088396

43 1.67332361 -0.255924781 1.045050588 1.123687823 -1.9522101461 -2.119687437 -1.332371191

44 0.90259974 0.363631181 -0.299990056 -0.255905531 -0.7701764211 0.887549770 -0.087828262

45 1.95541451 -0.639257441 -1.033450259 -3.960416285 0.7196548783 0.967032366 0.742881838

46 0.34686810 1.798081471 -0.041793956 0.126208371 -0.1614925200 -0.359011404 0.355578909

47 -0.60652713 0.414564073 -0.275913470 -0.656876941 -0.4937922131 0.105481670 0.354852371

48 -3.84003655 -0.912111842 1.226544248 -0.650436379 -0.0238171027 -0.445118781 -1.139261239

49 -0.76587368 0.369394249 -0.315919390 0.238701160 1.0420046083 -0.259667601 -0.803308813

50 -0.36468474 0.661616059 1.039430175 -1.090938893 -0.8124416993 0.647118358 0.362174693

51 0.26639199 0.724159995 0.093845907 -0.320122822 -0.7813829443 0.271271317 0.432309126

52 5.56515252 -1.780620518 -2.224284414 -2.160113771 -1.3535737854 -0.725184986 1.207072434

53 0.49097864 0.446423271 -0.178928440 0.036509910 -0.9869740203 0.467661250 0.299478513

54 0.61121109 0.624288125 -0.263470648 -0.015950851 -0.7590540560 0.161884637 0.307263428

55 1.31272821 -1.428599976 -3.553981106 -1.102048960 -0.7048950712 -1.308244629 -2.031918190

56 0.95603019 0.524416254 -0.620787203 0.288221121 -0.7367251678 0.052497956 0.182217730

57 4.80584462 2.189126269 1.697227845 -2.239798387 0.4474208317 0.537060764 1.027696526

58 6.05811742 -1.696473591 -1.703728205 0.419155264 1.9219161023 1.381783956 -0.210183543

59 1.63655379 2.289108558 2.828060249 -0.006312572 -3.2862443424 3.818445647 -1.010595795

60 -1.31331632 3.729617158 0.343840446 -0.616678718 -0.4956031793 0.625406769 0.127451316

61 -1.42108659 -1.691593419 -0.338309455 -0.492778059 -0.9808536217 -0.098745977 2.300811012

62 1.31807496 1.739073858 1.114273184 1.319874247 0.6821786041 -1.440780740 0.756863587

63 -0.88966074 1.125330164 2.444120323 0.807703443 0.3353090543 -2.513156981 0.707016051

64 3.87308487 2.978781241 -2.508052737 0.990064198 -0.8195910601 -0.081116200 0.386710794

65 -1.73938285 0.722572765 -0.958571927 -1.262185956 -0.1147545898 0.647043211 0.432450290

66 -2.75787853 -0.205773510 0.136208744 0.126399665 0.9617852411 0.186820902 0.572646427

67 -4.76087618 0.433228399 -0.783503547 -0.690721987 0.7417582632 0.527566112 -0.418325853

68 -0.10364356 -0.740422032 0.498162738 -1.381456987 -0.1998433086 0.085488780 -0.584778908

69 -2.97960875 0.267632696 0.365563231 -0.621840238 -0.4972329092 0.427318058 -1.375980281

70 -1.29580506 -1.594772940 -0.589091813 -0.929883354 -0.6676004898 0.351088134 0.477189072

71 -1.07260981 2.164878915 0.074648429 -0.809408391 0.1487026343 0.043408770 -1.014952523

72 -0.96721901 0.460871091 -0.579744880 0.970821399 -0.1699305831 0.181506970 0.591255277

73 -3.75895186 -0.694081997 0.084912540 -2.065557201 -0.6592546974 -0.773611109 -1.582648438

74 -0.42361114 -0.276690912 -1.810763338 0.890945540 -0.5026852638 -0.032891844 -0.379029448

75 -0.83660045 2.617200579 -1.137712232 1.348880256 1.0051022140 0.468482284 -0.576047132

76 -0.05806265 -0.112821101 -1.063745139 1.197109188 -0.5036456230 -0.784329602 1.530360529

Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13

1 -0.2386990114 -0.0963873447 0.119831838 0.107178943 -0.0833736666 8.484356e-03

2 -1.5308154736 -0.1113044507 0.065305464 -0.934720241 -1.5046900089 5.424512e-01

3 -0.5549428744 0.6598204200 -0.508581659 -0.064894270 0.1355331477 2.436771e-01

4 1.6752258039 -0.6550617688 -0.655860107 0.149709344 0.2524571807 -8.443022e-02

5 -2.3282219333 -1.2608802127 0.243888599 3.065799227 0.2750242080 5.898712e-01

6 -0.9467473226 1.1436250163 -0.261629889 -0.217503107 0.0640054039 -9.074146e-02

7 -0.2859779825 -1.4194673976 -0.720423190 0.187111259 -0.8939692495 -1.926509e-01

8 -0.2622119601 -0.6659872841 -0.477202105 -0.162737976 0.1475267595 -4.102639e-01

9 0.9372469871 0.6439800217 -0.354977923 0.834582921 -1.0972573243 -3.029197e-01

10 -0.8756091861 1.1407399776 1.055373982 -0.475088210 -0.7435155605 -4.353746e-01

11 -0.2535882843 0.0106966601 0.261894194 -0.250109246 -0.4913247855 2.568383e-01

12 0.2221761627 -1.7321711988 0.327062847 -0.481401743 0.2807342122 -1.121995e-01

13 -1.0295021462 -0.3772386615 -0.104639092 0.018500800 -0.3332328151 8.040825e-01

14 0.4880417452 1.4411794721 -0.393989937 -0.684400614 -0.4216036233 4.683242e-01

15 0.0093798952 0.6026434358 -0.081917962 0.652821431 0.4214866589 4.026026e-01

16 0.3187715311 0.9479054627 -0.882848792 -0.120755389 0.6687206138 -9.202653e-01

17 0.4026369413 -0.0375408889 0.465718674 -0.803490550 -0.0493362765 8.548601e-01

18 -0.1399524565 1.0447745628 -0.098680749 0.065728076 0.4703240195 2.224611e-02

19 -1.8704708120 0.3417702897 -0.228408976 -0.081922700 -1.0231794405 2.818590e-01

20 0.1747470142 -1.1253351736 0.236460598 0.093192255 -0.1790263488 -6.801564e-02

21 0.1970390770 -1.1580065667 0.282891106 -1.174585451 0.3116596909 1.027059e+00

22 -0.1821906250 0.7334566402 -0.507685261 0.316872868 0.1768764683 6.162320e-01

23 0.2491729241 -0.1955604403 -0.572273832 0.263657117 0.5554346019 2.227291e-01

24 0.7978439394 0.6534978471 -0.186299995 -0.074190394 0.7111672387 -9.320447e-02

25 0.1740707850 0.8086780531 -0.721972677 0.801527542 -0.8187983783 1.947567e-01

26 -0.2805626228 -0.5273951738 -0.418574379 -1.515933098 0.9665585726 -1.248907e+00

27 0.5397705431 -0.7401977251 -0.632612988 -0.188034559 -0.2408188295 -6.644870e-01

28 -0.3233506334 -1.3237735671 -0.502347934 -0.316389254 -0.2636833346 -5.316841e-01

29 0.2621844576 -0.4571781429 -0.271537035 -0.040854061 0.1244573980 -5.098805e-02

30 0.3168983505 0.6336418462 0.216922884 -0.038671601 -0.5796281463 -8.079939e-01

31 0.6229048858 -0.0983575755 -0.376632950 -0.729059879 0.4530813399 1.746838e-01

32 0.2383876105 -0.0659595318 0.744474458 0.028551050 0.5498802909 1.927032e-02

33 -0.4116181123 -0.0552922374 0.586876548 -0.220809740 -0.2896292617 1.028538e-01

34 1.1592873758 0.5078731548 0.093392474 0.122927774 -0.0998381283 -1.576581e-01

35 -0.0980928876 -0.0250286473 0.042027975 -0.425863024 -0.2260587552 3.162305e-01

36 -0.3104230401 -0.8197178198 0.102239898 0.019147748 0.2825266893 -2.840196e-01

37 0.2967815949 0.5833260413 0.888507730 0.064717642 0.5832782578 -6.103439e-02

38 0.2348978023 -0.3843717782 0.012330264 -0.485004518 -0.2385866865 2.938557e-01

39 0.2043609559 0.0535693312 -1.311534239 -0.327130181 0.3681826606 -9.490893e-02

40 -0.1712368411 -0.2897473065 -0.001425162 -0.145438845 -0.1913457753 1.327367e-01

41 0.1292270076 -0.6495637570 -0.369781752 0.006088038 -0.0359197821 -3.815937e-01

42 -0.3208894689 -0.1853224462 0.721600320 0.036061546 0.4954414380 4.719757e-01

43 2.0253112624 0.5477620672 0.499912302 -0.012454760 0.8106448452 1.005820e-01

44 -0.1846237183 -0.5025917979 0.630364409 0.168026340 0.0913770153 4.019261e-01

45 1.6423884467 -0.1946441532 1.087181796 0.561194737 -0.5099665347 -6.979817e-01

46 -0.5548596635 -0.0006667352 1.147372360 -0.405075681 0.3484551228 -5.537141e-01

47 0.5522520784 0.2576301830 -0.312241834 0.209759999 0.2961271989 -8.502445e-02

48 -0.2043822522 0.8217758520 -1.310873551 0.257205571 0.8150608519 -3.964221e-01

49 0.5017085375 -1.1321133912 -0.638123634 -0.023403361 -0.0104325708 -6.681845e-01

50 2.0525806409 -1.4275977818 -1.280702766 0.286416132 0.7618459214 -2.615208e-01

51 0.2973656013 -0.1433790848 -0.433338510 0.346819714 -1.4176286001 -5.728704e-01

52 -1.3687363030 -0.4267827717 0.164764812 -0.442544921 0.2896620105 -2.434793e-01

53 0.3014865462 -0.2221937533 0.872073682 0.340300605 0.3957376844 3.548096e-01

54 0.0293332949 -0.1198832148 -0.156753336 0.226999329 -0.7505011333 -2.821930e-01

55 -0.4219395195 -0.0829549030 -1.000268386 -0.451742269 0.8333154918 1.219892e+00

56 -0.2386990114 -0.0963873447 0.119831838 0.107178943 -0.0833736666 8.484356e-03

57 -0.8075904621 0.7867043083 0.496167194 0.467293309 0.2050235527 4.959863e-01

58 0.2784986051 1.1121769572 0.354239166 0.714585582 0.2855033240 -1.226128e+00

59 0.2720622184 0.8156796192 0.346052908 -0.894413283 -0.4504236830 -1.295247e+00

60 1.4297674844 1.9196250659 0.173995751 -0.129550060 0.5899293681 2.676487e-01

61 0.7586981050 -0.1973421705 0.018875373 0.399473432 0.7218907221 -2.800002e-01

62 0.1854760335 -0.7091022308 0.228897698 0.018865422 0.0856846766 -1.302120e+00

63 -0.6586649070 -0.3028631461 -0.406096244 -0.622752620 -0.9760447904 -6.779055e-01

64 -0.3242820352 0.0148562748 0.071245354 0.176319069 -0.1315767830 -2.650167e-05

65 0.0330967430 0.1976758985 -0.408974489 0.206399781 -0.0873678772 2.343991e-02

66 -0.6892621424 -0.4000022157 -0.011791736 0.547186585 0.4660728063 3.248842e-01

67 -0.5839351378 0.5013459591 -0.337153601 0.013023457 -0.1477437725 2.052917e-01

68 0.5656884587 0.4813130469 0.363946480 0.279460574 -0.0002057186 2.567662e-01

69 -0.6789326744 0.8640313099 -0.200354619 0.236584405 0.0512412694 2.501140e-01

70 -0.1416011284 0.2624702110 -1.045810488 0.893812653 0.2387032280 5.729549e-01

71 0.4101233095 -0.2756684036 -0.049720334 0.670827043 0.0050395121 -3.222926e-01

72 -0.4157697041 -0.7082775809 0.918188487 0.801149894 -0.4459482153 1.895090e-01

73 0.2905768367 0.4106382061 0.021671619 -1.038407049 -0.4799320754 1.053953e+00

74 -0.3971415227 -0.3051067235 -0.077647739 -0.480459838 -0.8030386964 4.529431e-01

75 -1.0652090271 -0.3257938079 1.838196930 -0.720045029 -0.5926735816 -7.454741e-01

76 0.6546100602 0.6909970783 -0.053851907 -0.769770902 -0.0794300730 2.988602e-01

[ reached getOption("max.print") -- omitted 134 rows ]

> install.packages("FactoMineR")

> library(FactoMineR)

> result <-PCA(mydata)

> library(psych)

> # 주성분분석 이후 적용 (4요인)

> refa1 <-data.frame(mydata$x3, mydata$x7, mydata$x8, mydata$x9, mydata$x12, mydata$x13)

> alpha(refa1, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa1, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.79** 0.79 0.79 0.38 3.7 0.023 5.5 1.3 0.34

lower alpha upper 95% confidence boundaries

0.74 0.79 0.83

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x3 0.77 0.77 0.78 0.41 3.4 0.025 0.0176 0.34

mydata.x7 0.76 0.76 0.74 0.39 3.2 0.026 0.0096 0.36

mydata.x8 0.74 0.74 0.71 0.36 2.8 0.028 0.0092 0.34

mydata.x9 0.75 0.75 0.74 0.37 3.0 0.028 0.0178 0.33

mydata.x12 0.75 0.76 0.75 0.38 3.1 0.027 0.0157 0.34

mydata.x13 0.74 0.74 0.73 0.37 2.9 0.028 0.0156 0.33

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x3 210 0.64 0.63 0.50 0.45 5.3 1.9

mydata.x7 210 0.66 0.66 0.59 0.49 5.9 1.8

mydata.x8 210 0.74 0.74 0.71 0.60 5.8 2.0

mydata.x9 210 0.71 0.71 0.63 0.56 5.2 1.8

mydata.x12 210 0.69 0.69 0.60 0.53 5.9 1.8

mydata.x13 210 0.73 0.73 0.66 0.58 4.9 2.0

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x3 0.07 0.04 0.06 0.03 0.41 0.15 0.14 0.05 0.06 0

mydata.x7 0.04 0.01 0.03 0.04 0.34 0.12 0.24 0.12 0.06 0

mydata.x8 0.04 0.02 0.05 0.07 0.29 0.17 0.14 0.14 0.08 0

mydata.x9 0.04 0.02 0.10 0.08 0.44 0.10 0.09 0.07 0.05 0

mydata.x12 0.02 0.03 0.04 0.07 0.31 0.14 0.16 0.15 0.08 0

mydata.x13 0.10 0.05 0.05 0.09 0.40 0.12 0.09 0.05 0.04 0

> # 크론바하 알파계수 = 0.79

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa2 <-data.frame(mydata$x5, mydata$x6)

> alpha(refa2, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa2, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.8** 0.8 0.66 0.66 3.9 0.028 6 1.9 0.66

lower alpha upper 95% confidence boundaries

0.74 0.8 0.85

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x5 0.66 0.66 0.44 0.66 NA NA 0.66 0.66

mydata.x6 0.44 0.66 NA NA NA NA 0.44 0.66

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x5 210 0.91 0.91 0.74 0.66 6.2 2.1

mydata.x6 210 0.91 0.91 0.74 0.66 5.9 2.1

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x5 0.05 0.01 0.04 0.05 0.25 0.11 0.19 0.14 0.16 0

mydata.x6 0.05 0.02 0.05 0.06 0.30 0.10 0.16 0.10 0.15 0

> # 크론바하 알파계수 = 0.8

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa3 <-data.frame(mydata$x10, mydata$x11)

> alpha(refa3, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa3, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.78** 0.78 0.64 0.64 3.5 0.031 6 1.8 0.64

lower alpha upper 95% confidence boundaries

0.72 0.78 0.84

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x10 0.64 0.64 0.41 0.64 NA NA 0.64 0.64

mydata.x11 0.41 0.64 NA NA NA NA 0.41 0.64

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x10 210 0.91 0.9 0.72 0.64 6.0 2.1

mydata.x11 210 0.90 0.9 0.72 0.64 6.1 1.9

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x10 0.05 0.04 0.04 0.04 0.24 0.10 0.23 0.15 0.10 0

mydata.x11 0.04 0.02 0.02 0.04 0.29 0.12 0.23 0.12 0.11 0

> # 크론바하 알파계수 = 0.78

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa4 <-data.frame(mydata$x2, mydata$x4)

> alpha(refa4, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa4, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.47** 0.47 0.31 0.31 0.88 0.073 4.4 1.6 0.31

lower alpha upper 95% confidence boundaries

0.33 0.47 0.61

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x2 0.307 0.31 0.094 0.31 NA NA 0.307 0.31

mydata.x4 0.094 0.31 NA NA NA NA 0.094 0.31

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x2 210 0.80 0.81 0.45 0.31 3.7 2.0

mydata.x4 210 0.82 0.81 0.45 0.31 5.1 2.1

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x2 0.18 0.12 0.18 0.08 0.31 0.04 0.04 0.03 0.01 0

mydata.x4 0.09 0.05 0.06 0.07 0.39 0.10 0.11 0.09 0.05 0

> # 크론바하 알파계수 = 0.47

> # 크론바하 알파계수가 다소 낮다. 개선할 필요가 보인다.

> ##### 공통요인분석 #####

> # 공통요인분석 이후 적용 (4요인)

> fit2 <-factanal(df06, 4, rotation = "varimax")

> print(fit2, digits = 2, cutoff = .3, sort = TRUE)

Call:

factanal(x = df06, factors = 4, rotation = "varimax")

Uniquenesses:

x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

0.94 0.77 0.66 0.70 0.44 0.21 0.46 0.20 0.52 0.47 0.13 0.50 0.41

Loadings:

Factor1 Factor2 Factor3 Factor4

x3 0.53

x12 0.55 0.43

x13 0.69

x7 0.59 0.36

x8 0.81

x10 0.37 0.56

x11 0.87

x5 0.69

x6 0.86

x1

x2 0.44

x4 0.36 0.37

x9 0.40 0.41 0.38

Factor1 Factor2 Factor3 Factor4

SS loadings 1.73 1.67 1.64 1.57

Proportion Var 0.13 0.13 0.13 0.12

Cumulative Var 0.13 0.26 0.39 0.51

Test of the hypothesis that 4 factors are sufficient.

The chi square statistic is 57.1 on 32 degrees of freedom.

**The p-value is 0.00412**

> # p-value = 0.00412 < 0.05가 되어 'H0 : 4요인으로 구성된 모형은 적합하다'를 기각하게 됨.

> # 4요인은 불가능하여 5요인으로 넘어감

> # 공통요인분석 이후 적용 (5요인)

> fit2 <-factanal(df06, 5, rotation = "varimax")

> print(fit2, digits = 2, cutoff = .3, sort = TRUE)

Call:

factanal(x = df06, factors = 5, rotation = "varimax")

Uniquenesses:

x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

0.94 0.57 0.66 0.66 0.47 0.16 0.00 0.38 0.43 0.37 0.31 0.26 0.47

Loadings:

Factor1 Factor2 Factor3 Factor4 Factor5

x9 0.54 0.46

x10 0.70

x11 0.71 0.37

x5 0.67

x6 0.88

x12 0.80

x13 0.57 0.35

x7 0.92

x2 0.64

x1

x3 0.36 0.41

x4 0.35 0.37

x8 0.49 0.49 0.30

Factor1 Factor2 Factor3 Factor4 Factor5

SS loadings 1.82 1.55 1.38 1.36 1.20

Proportion Var 0.14 0.12 0.11 0.10 0.09

Cumulative Var 0.14 0.26 0.37 0.47 0.56

Test of the hypothesis that 5 factors are sufficient.

The chi square statistic is 22.05 on 23 degrees of freedom.

**The p-value is 0.517**

> # p-value = 0.517 > 0.05가 되어 'H0 : 5요인으로 구성된 모형은 적합하다'을 기각할 수 없게 됨

> refa1 <-data.frame(mydata$x9, mydata$x10, mydata$x11)

> alpha(refa1, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa1, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.79** 0.79 0.72 0.55 3.7 0.025 5.7 1.6 0.53

lower alpha upper 95% confidence boundaries

0.74 0.79 0.83

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x9 0.78 0.78 0.64 0.64 3.5 0.031 NA 0.64

mydata.x10 0.69 0.69 0.53 0.53 2.3 0.042 NA 0.53

mydata.x11 0.65 0.65 0.48 0.48 1.9 0.048 NA 0.48

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x9 210 0.78 0.80 0.63 0.56 5.2 1.8

mydata.x10 210 0.86 0.85 0.73 0.64 6.0 2.1

mydata.x11 210 0.86 0.86 0.77 0.68 6.1 1.9

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x9 0.04 0.02 0.10 0.08 0.44 0.10 0.09 0.07 0.05 0

mydata.x10 0.05 0.04 0.04 0.04 0.24 0.10 0.23 0.15 0.10 0

mydata.x11 0.04 0.02 0.02 0.04 0.29 0.12 0.23 0.12 0.11 0

> # 크론바하 알파계수 = 0.79

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa2 <-data.frame(mydata$x5, mydata$x6)

> alpha(refa2, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa2, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.8** 0.8 0.66 0.66 3.9 0.028 6 1.9 0.66

lower alpha upper 95% confidence boundaries

0.74 0.8 0.85

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x5 0.66 0.66 0.44 0.66 NA NA 0.66 0.66

mydata.x6 0.44 0.66 NA NA NA NA 0.44 0.66

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x5 210 0.91 0.91 0.74 0.66 6.2 2.1

mydata.x6 210 0.91 0.91 0.74 0.66 5.9 2.1

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x5 0.05 0.01 0.04 0.05 0.25 0.11 0.19 0.14 0.16 0

mydata.x6 0.05 0.02 0.05 0.06 0.30 0.10 0.16 0.10 0.15 0

> # 크론바하 알파계수 = 0.8

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa3 <-data.frame(mydata$x12, mydata$x13)

> alpha(refa3, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa3, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.73** 0.73 0.57 0.57 2.7 0.037 5.4 1.7 0.57

lower alpha upper 95% confidence boundaries

0.65 0.73 0.8

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x12 0.57 0.57 0.33 0.57 NA NA 0.57 0.57

mydata.x13 0.33 0.57 NA NA NA NA 0.33 0.57

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x12 210 0.88 0.89 0.67 0.57 5.9 1.8

mydata.x13 210 0.90 0.89 0.67 0.57 4.9 2.0

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x12 0.02 0.03 0.04 0.07 0.31 0.14 0.16 0.15 0.08 0

mydata.x13 0.10 0.05 0.05 0.09 0.40 0.12 0.09 0.05 0.04 0

> # 크론바하 알파계수 = 0.73

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa4 <-data.frame(mydata$x7, mydata$x8)

> alpha(refa4, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa4, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.78** 0.78 0.64 0.64 3.5 0.031 5.8 1.7 0.64

lower alpha upper 95% confidence boundaries

0.72 0.78 0.84

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x7 0.64 0.64 0.41 0.64 NA NA 0.64 0.64

mydata.x8 0.41 0.64 NA NA NA NA 0.41 0.64

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x7 210 0.90 0.9 0.72 0.64 5.9 1.8

mydata.x8 210 0.91 0.9 0.72 0.64 5.8 2.0

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x7 0.04 0.01 0.03 0.04 0.34 0.12 0.24 0.12 0.06 0

mydata.x8 0.04 0.02 0.05 0.07 0.29 0.17 0.14 0.14 0.08 0

> # 크론바하 알파계수 = 0.78

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> refa5 <-data.frame(mydata$x2, mydata$x3, mydata$x4)

> alpha(refa5, na.rm = TRUE)

Reliability analysis

Call: alpha(x = refa5, na.rm = TRUE)

raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r

**0.56** 0.56 0.46 0.3 1.3 0.053 4.7 1.5 0.31

lower alpha upper 95% confidence boundaries

0.45 0.56 0.66

Reliability if an item is dropped:

raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r

mydata.x2 0.41 0.41 0.26 0.26 0.69 0.082 NA 0.26

mydata.x3 0.47 0.47 0.31 0.31 0.88 0.073 NA 0.31

mydata.x4 0.49 0.49 0.33 0.33 0.98 0.070 NA 0.33

Item statistics

n raw.r std.r r.cor r.drop mean sd

mydata.x2 210 0.74 0.75 0.54 0.40 3.7 2.0

mydata.x3 210 0.71 0.72 0.48 0.36 5.3 1.9

mydata.x4 210 0.73 0.71 0.46 0.34 5.1 2.1

Non missing response frequency for each item

1 2 3 4 5 6 7 8 9 miss

mydata.x2 0.18 0.12 0.18 0.08 0.31 0.04 0.04 0.03 0.01 0

mydata.x3 0.07 0.04 0.06 0.03 0.41 0.15 0.14 0.05 0.06 0

mydata.x4 0.09 0.05 0.06 0.07 0.39 0.10 0.11 0.09 0.05 0

> # 크론바하 알파계수 = 0.56

> # 어떤 변수를 제거하더라도 크론바하 알파계수가 증가하지 않으므로 멈춘다.

> # 0.6에 근접하여 괜찮은 수준으로 판단한다.

> # 공통요인분석의 파라미터를 6요인으로 설정하고 분석한 결과

> # 요인 6에는 실질적으로 어떠한 변수도 포함되지 않는 것을 알 수 있다.

> # 소스코드에서는 생략하도록 한다.

**주성분분석과 공통요인분석을 한 결과, 각 분석에서 차이점이 존재했고 그에 따른 장단점이 존재했다.**

**주성분분석의 경우, scree도표와 고유값에 근거했을 때 4개의 요인으로 축소시키는 것이 최적이었으며 그에 따라 요인을 구분한 결과는 다음과 같다.**

**요인 1 : x3, x7, x8, x9, x12, x13  
요인 2 : x5, x6  
요인 3 : x10, x11  
요인 4 : x2, x4**

**변수 x1은 그 어느 요인에도 속하지 않는 결과를 보였다.   
또한, 주성분분석의 경우 요인4의 크론바하 알파계수가 0.47로 다소 낮은 결과를 보여줬다.**

**공통요인분석의 경우, 4개 요인으로 축소시킬 경우 카이제곱 통계량이 57.1이고 자유도는 32이며 p-value가 0.00412 > α = 0.05 이므로 귀무가설 (4요인으로 구성된 모형은 적합하다.)를 기각하게 된다. 즉, 4요인은 적합하지 않다는 결과가 나온다.**

**5개 요인부터는 p-value가 α = 0.05보다 크게 나와 공통요인분석이 통계적으로 유의하다는 결과를 보여준다. 다만, 6개 요인으로 공통요인분석을 하게 될 경우 요인6에는 어떠한 변수도 포함되지 않아 5개 요인으로 파라미터를 설정하여 공통요인분석을 실시했다. 그에 따라 요인을 구분한 결과는 다음과 같다.**

**요인 1 : x9, x10, x11  
요인 2 : x5, x6  
요인 3 : x12, x13  
요인 4 : x7, x8  
요인 5 : x2, x3, x4**

**주성분분석과 마찬가지로 변수 x1은 그 어느 요인에도 속하지 않는 결과를 보였다.  
또한, 요인5의 크론바하 알파계수가 0.56으로 다소 낮은 결과를 보여줬다.**

**두 요인분석의 결과를 봤을 때, 주성분분석의 결과가 더욱 의미있다고 판단된다. 공통요인분석으로 구분된 요인들은 속해 있는 변수의 공통성을 찾기 어렵기 때문이다.**

**결론적으로 주성분분석을 통한 요인 4개를 의미있다고 판단하고, 각 요인에 이름을 붙인 결과는 다음과 같다.**

**요인 1 – 제품 내부 요인 : x3(제품의 다양한 디자인), x7(제품의 내구성), x8(제품의 액정 화질), x9(제품 관련 구성품의 다양성), x12(제품의 색상), x13(제품의 무게)  
요인 2 – 구입 용이성 요인: x5(구입 장소와 시간의 다양성), x6(구입 방법의 다양성)  
요인 3 – 이미지 요인: x10(판매/유통사의 이미지), x11(제품의 이미지)  
요인 4 – 제품 외부 요인: x2(이벤트의 다양성), x4(A/S의 신속성)**

**7. 위 6번 자료를 이용해 응답자들에 대해 계층적 군집분석과 비계층적 군집분석을 수행하시오. (2점)**

**(R 코드는 “DA\_AS02.R”파일의 #07 아래에 작성하고, (분석 결과 설명에 필요하다면 그래프를 포함하여) 분석결과는 아래에 작성할 것.)**

**<풀이>**

**- R을 이용한 분석 결과(R 출력) 및 결론**

> ##### 계층적 군집분석 #####

> df07 <- read.csv("C:/DA2019/exercise2/DA\_AS02-06.csv")

> head(df07)

x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

1 5 5 5 5 5 5 5 5 5 5 5 5 5

2 9 1 1 1 9 5 5 5 5 5 9 7 8

3 7 5 5 5 5 5 6 5 5 1 1 1 1

4 9 9 9 5 5 5 5 5 5 5 5 5 1

5 9 1 1 1 1 9 9 1 1 1 1 5 5

6 4 3 8 8 9 9 8 8 8 6 7 8 9

> mydata07 <-na.omit(df07)

> d <-dist(mydata07, method = "euclidean")

> fit0701\_1 <-hclust(d, method = "ward.D")

> plot(fit0701\_1)

> groups <-cutree(fit0701\_1, k = 3)

> rect.hclust(fit0701\_1, k = 3, border = "red")

> install.packages("psych")

Error in install.packages : Updating loaded packages

> install.packages("psych")

Warning in install.packages :

package ‘psych’ is in use and will not be installed

> library(psych)

> describeBy(mydata07, groups)

Descriptive statistics by group

group: 1

vars n mean sd median trimmed mad min max range skew kurtosis se

x1 1 144 5.72 1.80 5 5.70 1.48 1 9 8 0.07 -0.18 0.15

x2 2 144 4.13 1.93 5 4.09 1.48 1 9 8 0.08 -0.51 0.16

x3 3 144 5.14 1.72 5 5.22 1.48 1 9 8 -0.41 0.74 0.14

x4 4 144 5.26 1.70 5 5.25 1.48 1 9 8 0.03 0.40 0.14

x5 5 144 6.23 1.70 6 6.26 1.48 1 9 8 -0.22 -0.31 0.14

x6 6 144 5.84 1.69 5 5.84 1.48 1 9 8 -0.01 -0.16 0.14

x7 7 144 5.94 1.31 6 5.87 1.48 1 9 8 0.10 0.47 0.11

x8 8 144 5.94 1.45 6 5.86 1.48 3 9 6 0.52 -0.33 0.12

x9 9 144 5.38 1.39 5 5.28 0.00 2 9 7 0.69 0.72 0.12

x10 10 144 **6.18** 1.64 6 6.26 1.48 1 9 8 -0.48 0.20 0.14

x11 11 144 **6.20** 1.53 6 6.20 1.48 1 9 8 -0.17 0.06 0.13

x12 12 144 5.85 1.68 6 5.94 1.48 1 9 8 -0.33 -0.10 0.14

x13 13 144 4.94 1.69 5 5.03 1.48 1 9 8 -0.40 0.81 0.14

---------------------------------------------------------------------

group: 2

vars n mean sd median trimmed mad min max range skew kurtosis se

x1 1 33 **6.52** 1.97 7 6.63 2.97 1 9 8 -0.45 -0.28 0.34

x2 2 33 1.97 1.61 1 1.63 0.00 1 8 7 2.01 3.96 0.28

x3 3 33 4.15 2.35 5 4.00 2.97 1 9 8 0.19 -0.81 0.41

x4 4 33 2.58 1.89 2 2.37 1.48 1 7 6 0.76 -0.95 0.33

x5 5 33 4.03 2.44 5 3.89 2.97 1 9 8 0.18 -1.16 0.43

x6 6 33 4.00 2.47 4 3.81 2.97 1 9 8 0.30 -0.95 0.43

x7 7 33 3.82 2.16 4 3.70 1.48 1 9 8 0.17 -0.84 0.38

x8 8 33 3.12 1.83 3 3.00 2.97 1 7 6 0.36 -1.16 0.32

x9 9 33 2.94 1.46 3 2.93 1.48 1 5 4 0.04 -1.33 0.25

x10 10 33 3.21 2.09 3 3.00 2.97 1 8 7 0.55 -0.90 0.36

x11 11 33 3.82 2.10 5 3.74 2.97 1 8 7 -0.06 -1.25 0.37

x12 12 33 4.61 1.84 5 4.63 1.48 1 9 8 -0.02 0.15 0.32

x13 13 33 2.91 1.77 3 2.81 2.97 1 6 5 0.30 -1.51 0.31

---------------------------------------------------------------------

group: 3

vars n mean sd median trimmed mad min max range skew kurtosis se

x1 1 33 5.82 1.49 5 5.74 0.00 3 9 6 0.63 -0.56 0.26

x2 2 33 3.76 1.48 3 3.70 1.48 1 8 7 0.63 0.75 0.26

x3 3 33 6.94 1.32 7 6.93 1.48 5 9 4 -0.05 -1.03 0.23

x4 4 33 6.82 1.49 7 6.81 1.48 4 9 5 -0.14 -1.38 0.26

x5 5 33 **8.03** 1.31 9 8.26 0.00 4 9 5 -1.35 1.14 0.23

x6 6 33 **8.27** 1.01 9 8.44 0.00 5 9 4 -1.43 1.60 0.18

x7 7 33 7.82 0.85 8 7.85 1.48 6 9 3 -0.27 -0.66 0.15

x8 8 33 7.76 0.83 8 7.81 0.00 6 9 3 -0.50 -0.26 0.14

x9 9 33 6.76 1.52 7 6.74 1.48 4 9 5 -0.01 -1.37 0.26

x10 10 33 7.76 1.30 8 7.93 1.48 5 9 4 -0.97 -0.15 0.23

x11 11 33 7.73 1.26 8 7.85 1.48 5 9 4 -0.59 -0.82 0.22

x12 12 33 7.33 1.55 8 7.48 1.48 3 9 6 -0.69 -0.20 0.27

x13 13 33 6.73 1.55 7 6.74 1.48 4 9 5 0.00 -1.17 0.27

> ##### 비계층적 군집분석 #####

> fit0702\_2 <-kmeans(mydata07, 3)

> library(cluster)

> clusplot(mydata07, fit0702\_2$cluster, color = TRUE, xhade = TRUE, labels = 2, lines = 0)

> install.packages("fpc")

> library(fpc)

> plotcluster(mydata07, fit0702\_2$cluster)

> fit0702\_2 <-factanal(df07, 5, rotation = "varimax")

> print(fit0702\_2, digits = 2, cutoff = 0.3, sort = TRUE)

Call:

factanal(x = df07, factors = 5, rotation = "varimax")

Uniquenesses:

x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

0.94 0.57 0.66 0.66 0.47 0.16 0.00 0.38 0.43 0.37 0.31 0.26 0.47

Loadings:

Factor1 Factor2 Factor3 Factor4 Factor5

x9 0.54 0.46

x10 0.70

x11 0.71 0.37

x5 0.67

x6 0.88

x12 0.80

x13 0.57 0.35

x7 0.92

x2 0.64

x1

x3 0.36 0.41

x4 0.35 0.37

x8 0.49 0.49 0.30

Factor1 Factor2 Factor3 Factor4 Factor5

SS loadings 1.82 1.55 1.38 1.36 1.20

Proportion Var 0.14 0.12 0.11 0.10 0.09

Cumulative Var 0.14 0.26 0.37 0.47 0.56

Test of the hypothesis that 5 factors are sufficient.

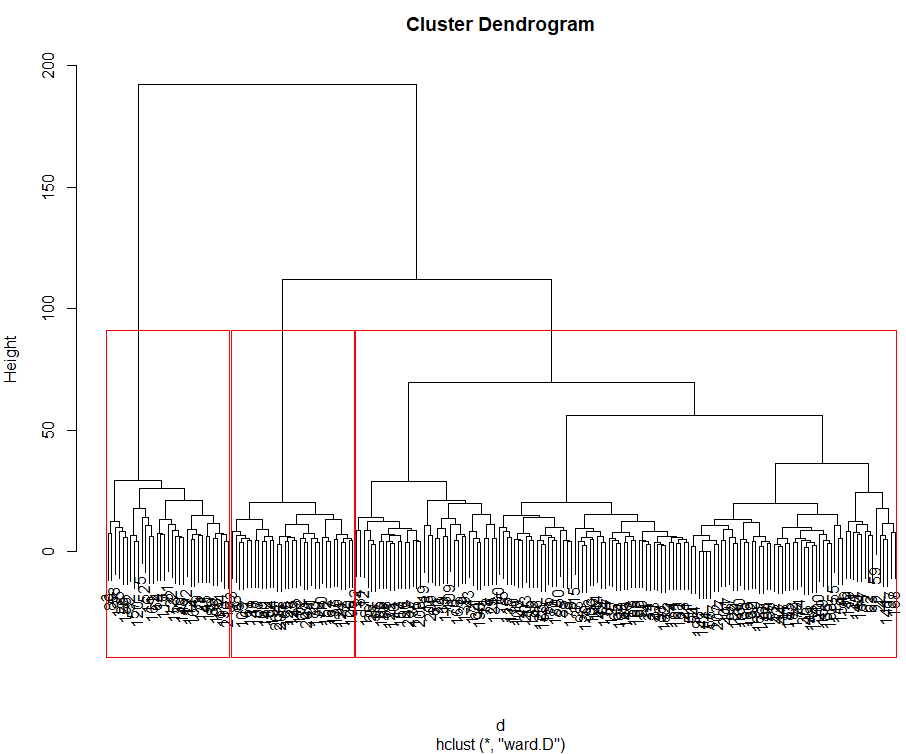
The chi square statistic is 22.05 on 23 degrees of freedom.

**The p-value is 0.517**

> load <-fit0702\_2$loadings[,1:2]

> plot(load, type = "n")

> text(load, labels = names(df07), cex = .7)

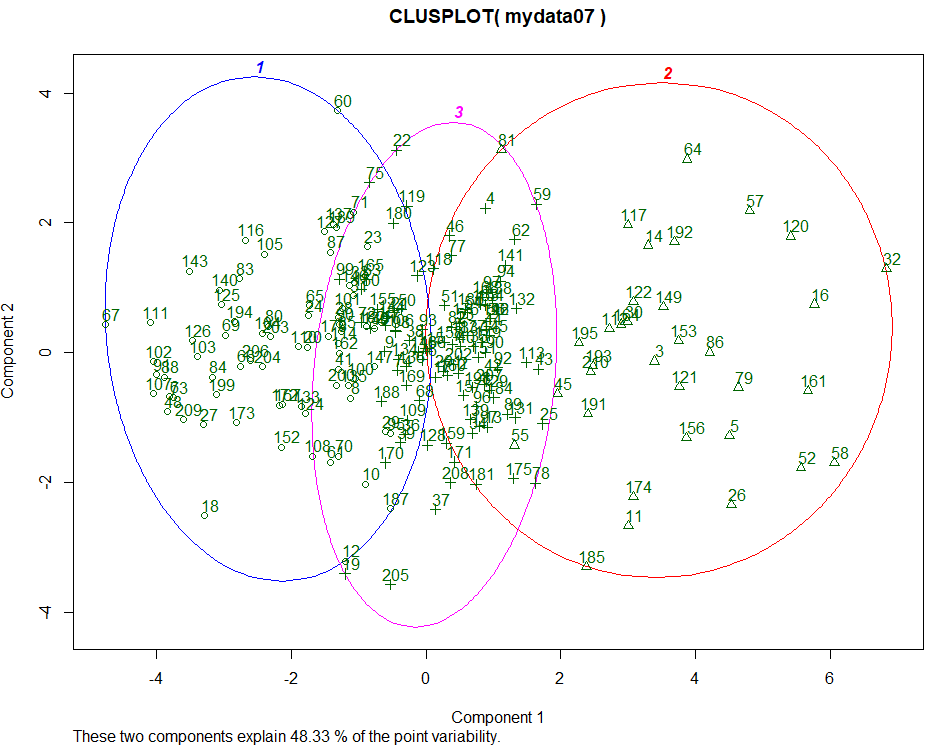


**계층적 군집분석으로 고객을 구분한 결과 위와 같은 그래프로 보여진다.**

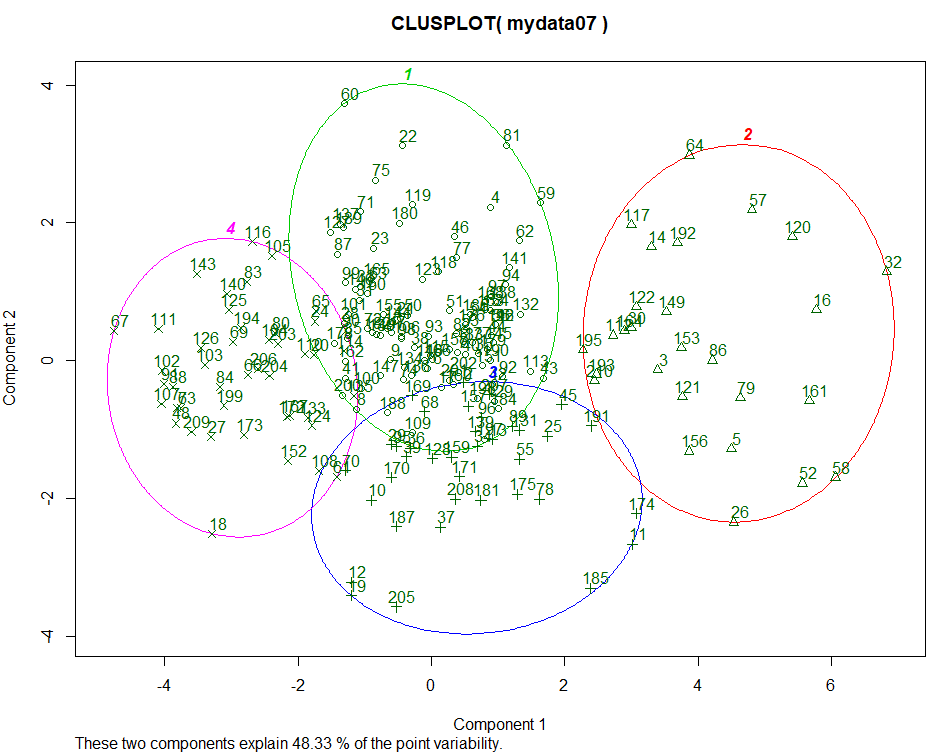
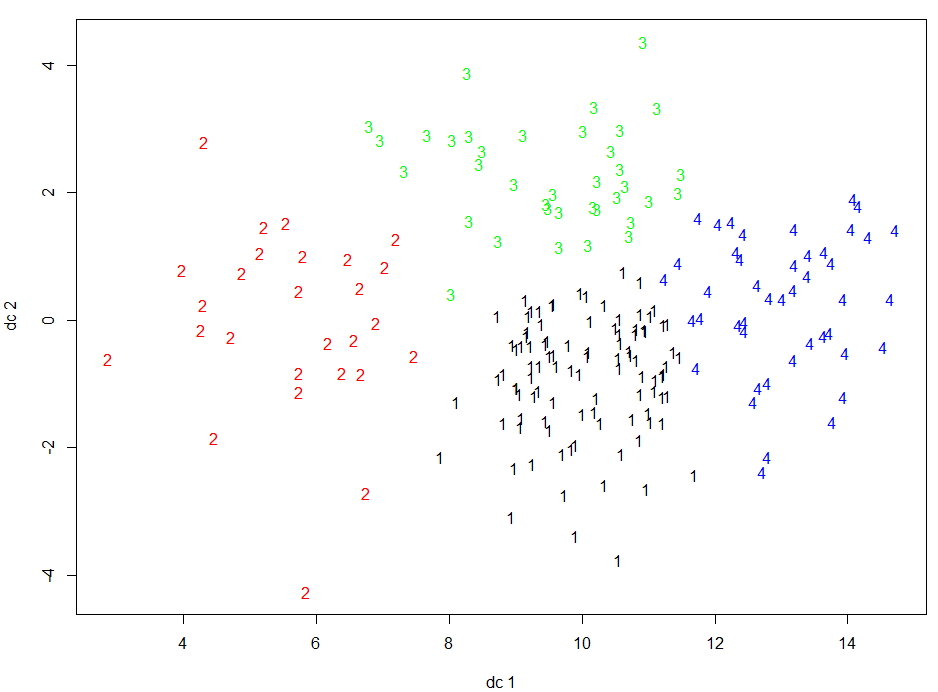
**군집1(144명) : x10(판매/유통사의 이미지), x11(제품의 이미지)의 평균은 각각 6.18, 6.20으로 다른 변수들에 비해 높은 편이다. 🡪 군집 1은 “이미지 중시고객군”으로 명명할 수 있다.**

**군집2(33명) : x1(제품의 가격)의 평균이 6.25이고 다른 변수들에 비해 월등히 높은 편이다. 🡪 군집 2는 “경제적 소비자군”으로 명명할 수 있다.**

**군집3(33명) : x5(구입 장소와 시간의 다양성), x6(구입 방법의 다양성)의 평균이 각각 8.03, 8.27로 다른 변수들에 비해 높은 편이다. 🡪 군집3은 “구입 용이성 추구 고객군”으로 명명할 수 있다.**



**비계층적 군집분석의 k를 3으로 설정했을 때 위와 같은 결과를 얻을 수 있다. 군집3은 군집1과 군집 2와 다수 겹치는 모습을 볼 수 있다. 그래서 k = 4로 설정한 다음 다시 시각화를 하니 다음과 같은 결과를 얻을 수 있었다.**

**이로써 비계층적 군집분석을 하게 될 경우, k=3보다 k=4로 설정하는 것이 집단 간 이질성을 멀어지게 하는 좋은 방법이라고 결론을 내렸다.**

**그에 따라 계층적 군집분석도 k=4로 구분하는 것이 좋다는 생각을 하게 되었지만 시간이 부족하여 추가적인 분석을 하지 못하게 되었다.**