

## CHAPTER 01.

### 1.

```
> mat <- read.csv("학력조사결과.csv")
> library(lattice)
> histogram(~사전점수|동아리,data=mat)
> boxplot(사전점수~동아리,data=mat,horizontal=TRUE)
```

### 2.

```
> tapply(mat$사전점수,mat$동아리,mean)
  무소속 인문계열 체육계열
183.5000 188.8655 182.6379
> tapply(mat$사전점수,mat$동아리,median)
  무소속 인문계열 체육계열
  185.5    186.5    180.0
> tapply(mat$사전점수,mat$동아리,sd)
  무소속 인문계열 체육계열
49.18550 46.30607 51.48008
```

### 3.

```
> m <- mat$수학[mat$성별=="M"]
> f <- mat$수학[mat$성별=="F"]
> t.test(m,f,var.equal=TRUE)
> effectd1(m,f,clevel=0.95)
      효과크기      신뢰수준      구간하한      구간상한      U3
0.02165224  0.95000000 -0.11695593  0.16024684  0.50863732
```

《결과 보고》 5% 수준에서 수학 평균에 성별 차이는 나타나지 않았다( $t(798) = 0.306$ , n.s.,  $d = 0.022(95\%CI [-0.117, 0.160])$ ).

### 4.

```
> goukei <- apply(mat[,c("사전점수","사후점수")],1,sum)
```

### 5.

```
> spre1 <- scale(mat$사전점수)
> plot(mat$사전점수,spre1)
> cor(mat$사전점수,spre1)
```

《분포에 대해 말할 수 있는 것》 표준화하면 분포의 원점과 단위가 변하지만, 데이터 간의 상대적 위치 관계는 변하지 않는다.

6.

```
> library(psych)
> matc <- mat[,c("사전점수","사후점수","국어","사회","영어")]
> partial.r(matc,c(3:5),c(1:2))
partial correlations
      국어  사회  영어
국어  1.00  0.02 -0.28
사회  0.02  1.00 -0.13
영어 -0.28 -0.13  1.00
```

7.

```
> library(polycor)
> kcat <- cut(mat$국어,breaks=c(-Inf,mean(mat$국어),Inf),right=FALSE,labels=c(0,1))
> scat <- cut(mat$사회,breaks=c(-Inf,mean(mat$사회),Inf),right=FALSE,labels=c(0,1))
> ecat <- cut(mat$영어,breaks=c(-Inf,mean(mat$영어),Inf),right=FALSE,labels=c(0,1))
> mat2 <- data.frame(kcat,scat,ecat)
> hetcor(mat2,ML=TRUE)
Correlations/Type of Correlation:
      kcat      scat      ecat
kcat      1 Polychoric Polychoric
scat 0.6654      1 Polychoric
ecat 0.5094      0.5474      1
```

8.

```
> effectv(mat$성별,mat$동아리,clevel=0.95)
효과크기V 카이제곱값 신뢰수준 구간하한 구간상한
0.1815144 26.3579743 0.9500000 0.1187906 0.2528534
```

## CHAPTER 02.

1.

```
> # 아래는 작업 디렉터리를 'POS폴더2'로 변환하는 것을  
> # 전제로 하고 있다  
> fname <- dir()  
> fname2 <- paste(fname, "/2013", sprintf("%02d", 1:12), ".csv", sep="")  
> tmpall <- lapply(fname2, read.csv, stringsAsFactors=FALSE)  
> posall2 <- do.call(rbind, tmpall)  
> locv <- c("고객ID", "점포", "상품카테고리")  
> posall2[, locv] <- lapply(posall2[, locv], as.factor)
```

2.

```
> loc2 <- (substr(posall2$구매일, 1, 6) == "201302")  
> loc5 <- (substr(posall2$구매일, 1, 6) == "201305")  
> pos02 <- posall2[loc2,]  
> pos05 <- posall2[loc5,]
```

3.

```
> store02 <- tapply(pos02$구매금액, list(pos02$고객ID, pos02$점포), sum)  
> store05 <- tapply(pos05$구매금액, list(pos05$고객ID, pos05$점포), sum)  
> store02[is.na(store02)] <- 0  
> store05[is.na(store05)] <- 0
```

4.

```
> #고객 ID를 변수로 포함하는 데이터프레임 작성  
> dat02 <- data.frame(rownames(store02), store02)  
> dat05 <- data.frame(rownames(store05), store05)  
  
> #변수명 변경  
> colnames(dat02) <- c("고객ID", "2월점포A", "2월점포B", "2월점포C")  
> colnames(dat05) <- c("고객ID", "5월점포A", "5월점포B", "5월점포C")  
  
> #고객 ID를 이용하여 두 데이터 세트 합병  
> mdat <- merge(dat02, dat05, by="고객ID")
```

5.

```
> #200원 이상이었던 월과 점포의 정보를 추출하는 함수 정의
```

```

> get200 <- function(x)
+ {
+   loc <- which(x>=200)
+   return(names(x[loc]))
+ }

> #직접 정의한 자작함수를 이용하여 고객 ID별로 정보 추출
> apply(mdat[,2:7],1,get200)

```

## 6.

```

> ptime <- factor(round(posall2$구매시간),level=seq(9,21,1))
> table(posall2$상품카테고리,ptime)

```

## 7.

```

> #년월일의 정보를 yyyy-mm-dd 형식으로 변환
> tmpdate2 <- paste(substr(posall2$구매일,1,4), "-",
+   substr(posall2$구매일,5,6), "-", substr(posall2$구매일,7,8), sep="")

> ndate2 <- as.Date(tmpdate2) #문자열을 date 형식으로 합병한다
> restime3 <- tapply(ndate2,posall2$고객ID,diff) #고객별 내점간격
> restime4 <- lapply(restime3,as.numeric) #리스트의 요소를 수치화해둔다

> #벡터의 요소가 50이상인 경우 TRUE를, 아니면 FALSE를 반환하는 함수 정의
> x50 <- function(x)
+ {
+   res <- sum(x>=50)
+   return(ifelse(res>=1,TRUE,FALSE))
+ }

> sid <- sapply(restime4,x50)
> names(sid[sid==TRUE])

```

## 8.

```

> fmat <- readLines("항목반응고정길이2.txt") #데이터 입력
> sp <- c(1,7:106) #시점의 자릿수
> ep <- c(6,7:106) #종점의 자릿수
> fmat2 <- sapply(fmat,substring,sp,ep) #행렬 형식으로 변환

```

```
> #행 이름 변경
> dimnames(fmat2)[[2]] <- paste("ID",sprintf("%04d",1:1000),sep="")
> fmat3 <- t(fmat2)    #행렬의 전치
> key2 <- read.csv("key2.txt")    #정답 키 입력
> sweep(fmat3[,-1],2,key2[,1],FUN=="=")*1    #정오 반응 데이터 생성
```

## CHAPTER 03.

1.

```
> kwamokdat <- read.csv("과목시험결과.csv")
> dim(kwamokdat)
> colnames(kwamokdat)
```

2.

```
> retest <- lm(final~t1+t2+t3+t4+t5,data=kwamokdat)
```

3.

```
> library(car)
> vif(retest)
      t1      t2      t3      t4      t5
1.567852 57.690405 1.801130 58.275192 1.630464
```

《다중공선성에 관한 판단》 변수2와 4에는 다중공선성의 가능성이 있다.

4.

```
> #t4를 삭제하는 경우
> retest2 <- lm(final~t1+t2+t3+t5,data=kwamokdat)
> #t2를 삭제하는 경우
> retest2 <- lm(final~t1+t3+t4+t5,data=kwamokdat)
> summary(retest2)
```

Call:

```
lm(formula = final ~ t1 + t3 + t4 + t5, data = kwamokdat)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-13.2023	-2.5509	0.1511	2.5912	8.9968

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-8.20262	1.85638	-4.419	1.65e-05 ***
t1	0.32520	0.03603	9.027	< 2e-16 ***

```

t3          0.14469    0.03860    3.748 0.000234 ***
t4          0.39303    0.03564   11.028 < 2e-16 ***
t5          0.30113    0.03674    8.196 3.25e-14 ***

```

---

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

Residual standard error: 4.061 on 195 degrees of freedom

Multiple R-squared: 0.8384, Adjusted R-squared: 0.8351

F-statistic: 252.9 on 4 and 195 DF, p-value: < 2.2e-16

## 5.

```
> confint(restest2)
```

```

                2.5 %      97.5 %
(Intercept) -11.86378472 -4.5414549
t1           0.25415065  0.3962541
t3           0.06856489  0.2208219
t4           0.32274238  0.4633168
t5           0.22866465  0.3735894

```

## 6.

```
> restest3 <- lm(final~t1+t2,data=kwamokdat)
```

```
> extractAIC(restest2)
```

```
[1] 5.0000 565.4909
```

```
> extractAIC(restest3)
```

```
[1] 3.0000 648.3233
```

《모형 적합에 관한 판단》 4개의 설명변수를 포함한 모형 쪽이 적합이 잘 된다.

## 7.

```
> round(cor(kwamokdat),2)
```

```

      t1  t2  t3  t4  t5 final
t1    1.00 0.37 0.54 0.38 0.50 0.70
t2    0.37 1.00 0.53 0.99 0.48 0.73
t3    0.54 0.53 1.00 0.53 0.51 0.68
t4    0.38 0.99 0.53 1.00 0.49 0.74

```

t5	0.50	0.48	0.51	0.49	1.00	0.73
final	0.70	0.73	0.68	0.74	0.73	1.00

《편회귀계수의 해석》 다중공선성이 생기는 변수를 제외해도, 설명변수 사이에 0.37~0.54와 같은 무시할 수 없는 상관계수가 관측된다. 특정 설명변수의 영향을 편회귀계수로 분석하는 것은 실질적으로 불가능하므로, 설명변수군에 의한 목적변수의 설명력에 관한 결정계수를 바탕으로 분석해야 한다.



## CHAPTER 04.

1.

```
> ssk <- read.csv("성적.csv") #데이터 입력
> res1 <- lm(시험성적~성별+학원유무+자신감, data=ssk)
> summary(res1)
> res2 <- lm(시험성적~성별+학원유무+자신감+열정, data=ssk)
> summary(res2)
```

2.

```
> anova(res1, res2)
Analysis of Variance Table
```

Model 1: 시험성적 ~ 성별 + 학원유무 + 자신감

Model 2: 시험성적 ~ 성별 + 학원유무 + 자신감 + 열정

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	496	105462				
2	495	103139	1	2322.7	11.148	0.0009048 ***

《결정계수의 유의성》 결정계수의 증분은 통계적으로 유의하다.

3.

```
> extractAIC(res1)
[1] 4.000 2683.749
```

```
> extractAIC(res2)
[1] 5.000 2674.613
```

《AIC를 이용한 모형 비교》 AIC 관점에서는 단계 2의 모형 쪽이 데이터에 적합하다.

4.

```
> ssk$열정.c <- ssk$열정-mean(ssk$열정)
> res3 <- lm(시험성적~성별+학원유무+자신감+열정.c
+ +학원유무*열정.c, data=ssk)
> summary(res3)
```

5.

```
> ssk$열정.h <- ssk$열정.c-sd(ssk$열정.c)
> res3.h <- lm(시험성적~성별+학원유무+자신감+열정.h
```

```
+ +학원유무*열정.h, data=ssk)
```

```
> summary(res3.h)
```

-출력의 일부-

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	38.7563	2.4449	15.852	< 2e-16 ***
성별	-4.8309	1.2772	-3.782	0.000174 ***
학원유무	9.9728	1.9025	5.242	2.36e-07 ***
자신감	5.3702	0.6028	8.909	< 2e-16 ***
열정.h	-1.1814	1.0799	-1.094	0.274484
학원유무:열정.h	6.1626	1.3613	4.527	7.51e-06 ***

```
> ssk$열정.l <- ssk$열정.c+sd(ssk$열정.c)
```

```
> res3.l <- lm(시험성적~성별+학원유무+자신감+열정.l
```

```
+ +학원유무*열정.l, data=ssk)
```

```
> summary(res3.l)
```

-출력의 일부-

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	41.0058	2.0913	19.608	< 2e-16 ***
성별	-4.8309	1.2772	-3.782	0.000174 ***
학원유무	-1.7619	1.8319	-0.962	0.336630
자신감	5.3702	0.6028	8.909	< 2e-16 ***
열정.l	-1.1814	1.0799	-1.094	0.274484
학원유무:열정.l	6.1626	1.3613	4.527	7.51e-06 ***

《결과 해석》 학원 수강 효과는 열정이 낮은 중학생에게서는 보이지 않고, 열정이 높은 중학생에게만 보인다.

## 6.

```
> library(MASS) #패키지 입력
```

```
> base <- lm(시험성적~1, data=ssk)
```

```
> step.res <- stepAIC(base, direction="both", scope=list(upper=~성별+학원유무
```

```
+ +자신감+열정+평일공부시간+휴일공부시간+뉴스시청+독서+실외놀이))
```

```
> summary(step.res)
```

## CHAPTER 05.

1.

```
> kch <- read.csv("가치.csv") #데이터 입력
> library(ICC)
> ICCest(as.factor(학생), 흥미, data=kch, alpha=0.05, CI.type="Smith")
```

2.

```
> kch$가치.cwc <- kch$가치-ave(kch$가치, kch$학생) #집단평균중심화
> kch$기대.c <- kch$기대-mean(kch$기대) #전체평균중심화
```

3.

```
> library(lmerTest) #패키지 입력
> model1 <- lmer(흥미~가치.cwc+(1|학생), data=kch, REML=FALSE)
> summary(model1)
```

4.

```
> model2 <- lmer(흥미~가치.cwc+(1+가치.cwc|학생), data=kch, REML=FALSE)
> summary(model2)
```

5.

```
> model3 <- lmer(흥미~가치.cwc+기대.c+가치.cwc*기대.c+(1+가치.cwc|학생),
+ data=kch, REML=FALSE)
> summary(model3)
```

-출력의 일부-

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t )
(Intercept)	6.77333	0.13360	99.99275	50.697	< 2e-16 ***
가치.cwc	0.37750	0.04028	100.08245	9.371	2.35e-15 ***
기대.c	0.07774	0.04064	99.99275	1.913	0.0586 .
가치.cwc:기대.c	-0.02469	0.01124	82.46769	-2.197	0.0308 *

《교차 수준의 상호작용효과》 유의한 교차 수준의 상호작용효과가 나타났으며, 가치의 인지와 흥미의 관계는 대학생의 기대에 따라 다르다는 것을 시사한다.

6.

```
> anova(model1, model2, model3)
Data: kch
Models:
model1: 흥미 ~ 가치.cwc + (1 | 학생)
model2: 흥미 ~ 가치.cwc + (1 + 가치.cwc | 학생)
```

```

model3: 흥미 ~ 가치.cwc + 기대.c + 가치.cwc * 기대.c + (1 + 가치.cwc |
model3: 학생)
      Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
model1  4 5887.7 5909.0 -2939.9  5879.7
model2  6 5867.5 5899.4 -2927.8  5855.5 24.2383      2 5.454e-06 ***
model3  8 5865.1 5907.6 -2924.6  5849.1  6.3787      2  0.0412 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

《가장 바람직한 모형》 AIC와 우도비 검정 결과에서는 교차 수준의 상호작용효과를 더한 모형의 적합이 가장 좋았다.

## CHAPTER 06.

1.

```
> hyk <- read.csv("수업평가.csv")
> hyk.model <- '
+ 흥미~난이도+유용성
+ 학습행동~흥미
+ 성적~학습행동
+ '
```

2.

```
> hyk.fit <- sem(hyk.model, data=hyk)
```

3.

```
> summary(hyk.fit, standardized=TRUE, rsquare=TRUE, fit.measures=TRUE)
```

《적합도 여부》 CFI=0.813 , TLI=0.664 , RMSEA=0.113 , SRMR=0.079으로 적합이 잘 되지 않았다.

4.

```
> modindices(hyk.fit)
```

-출력의 일부-

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
11	흥미	~~	학습행동	6.480	-21.617	-21.617	-0.633	-0.633
12	흥미	~~	성적	1.311	3.011	3.011	0.069	0.069
13	학습행동	~~	성적	2.497	-39.469	-39.469	-0.325	-0.325
14	흥미	~	학습행동	6.480	-0.228	-0.228	-0.641	-0.641
15	흥미	~	성적	0.815	0.015	0.015	0.057	0.057
16	학습행동	~	성적	2.497	-0.253	-0.253	-0.336	-0.336
17	학습행동	~	난이도	15.640	-0.971	-0.971	-0.223	-0.096
18	학습행동	~	유용성	0.133	0.082	0.082	0.021	0.008
19	성적	~	흥미	2.497	0.329	0.329	0.088	0.088
20	성적	~	난이도	3.588	-0.585	-0.585	-0.102	-0.043
21	성적	~	유용성	1.865	0.386	0.386	0.073	0.029
23	난이도	~	학습행동	17.390	-0.052	-0.052	-0.227	-0.227

《수정 지표로 생각할 수 있는 수정안》 ‘난이도 → 학습행동’ 혹은 ‘학습행동 → 난이도’라는 경로를 추가하면, 적합도가 개선되는 것으로 나타났다. 학습 내용이 어려워 이해하기 어려울 때는 학습행동에 방해받을 가능성이 있기 때문에 ‘난이도 → 학습행동’이라는 경로를 추가하는 수정안을 생각할 수 있다.

5.

```
> hyk.model2 <- '  
+ 흥미~난이도+유용성  
+ 학습행동~흥미+난이도  
+ 성적~학습행동  
+ '  
> hyk.fit2 <- sem(hyk.model2, data=hyk)  
> summary(hyk.fit2, standardized=TRUE, rsquare=TRUE, fit.measures=TRUE)
```

《적합도 개선》 새로운 ‘난이도 → 학습행동’이라는 경로를 추가하여 분석한 결과, CFI=0.961, TLI=0.913, RMSEA=0.057, SRMR=0.040으로 적합도가 개선된다.

6.

```
> summary(hyk.fit2, standardized=TRUE, rsquare=TRUE, fit.measures=TRUE,  
+ ci=TRUE)
```

## CHAPTER 07.

1.

```
> skk <- read.csv("성격.csv")
> cor.skk <- cor(skk)
> eigen(cor.skk)
eigen() decomposition
$values
[1] 3.4003439 1.2941737 0.8562578 0.6638816 0.5443059 -생략-
```

《카이저 기준의 결과》 카이저 기준으로는 2인자 해를 암시한다.

2.

```
> library(psych)
> VSS.scree(skk)
```

《스크리 테스트의 결과》 스크리 테스트에서는 2인자 해를 암시한다.

3.

```
> fa.parallel(skk, fm="ml", fa="pc", n.iter=100)
```

《평행분석의 결과》 평행분석의 결과에서는 2인자 해를 암시한다.

4.

```
> library(GPArotation)
> fa.skk <- fa(skk, nfactors=2, fm="ml", rotate="promax")
> print(fa.skk, sort=TRUE, digits=3)
```

5.

```
> skk2 <- skk[,c("쾌활", "적극적", "외향적", "사교적")]
> alpha(skk2)

> skk3 <- skk[,c("협조적", "온화", "솔직", "친절")]
> alpha(skk3)
```

6.

```
> omega(skk2, nfactors=1)
```

```
> omega(skk3, nfactors=1)
```



## CHAPTER 08.

1.

```
> skk <- read.csv("성격.csv")
> skk.model1 <- '
+ F1=~온화+쾌활+외향적+친절+사교적+협조적+적극적+솔직
+ '
> skk.fit1 <- cfa(skk.model1, data=skk, std.lv=TRUE)
> summary(skk.fit1, fit.measures=TRUE, standardized=TRUE)
```

2.

```
> skk.model2 <- '
+ F1=~쾌활+외향적+사교적+적극적
+ F2=~온화+친절+협조적+솔직
+ '
> skk.fit2 <- cfa(skk.model2, data=skk, std.lv=TRUE)
> summary(skk.fit2, fit.measures=TRUE, standardized=TRUE)
```

《인자 모형 비교》 AIC와 BIC의 값은 2인자 모형이 더 작아, 2인자 모형이 더 적합하다는 것을 시사한다.

## CHAPTER 09.

1.

‘인간 관계의 양호함’과 ‘마음의 건강’

2.

자유모수가 27개(이 중에 계수는 11개, 분산은 16개)이므로, 78-27에 의해 자유도는 51

3.

```
> sws <- read.csv("행복조사.csv")
> sws.model<-"
+ f1 =~ E1 + E2 + 1 * E3
+ f2 =~ R1 + R2 + 1 * R3
+ f3 =~ M1 + M2 + 1 * M3
+ f4 =~ H1 + H2 + 1 * H3
+ f1 ~ f3
+ f4 ~ f2 + f3
+ "
> sws.fit <- lavaan(model = sws.model, data = sws, auto.var = TRUE)
```

4.

```
> summary(sws.fit, fit.measures = TRUE, standardized = TRUE, ci = TRUE)
```

《적합도 평가》CFI=0.966 , TLI=0.956 , RMSEA=0.068(90%CI [0.051. 0:086]) , SRMR =0.120으로 모형이 데이터에 잘 적합한다고 말할 수 없다.

5.

Std.all을 보면, ‘인간 관계의 양호함’의 값은 0.350, ‘마음의 건강’의 값은 0.330으로 ‘인간 관계의 양호함’으로 인한 영향 쪽이 강한 것 같다.

## CHAPTER 10.

1.

```
> bdat2 <- read.csv("자전거데이터연습1.csv")
> bdat2$연령<- factor(bdat2$연령,levels=c("30대","20대","40대"))
> bdat2$성별<-factor(bdat2$성별,levels=c("M","F"))
> bdat2$제조사<-factor(bdat2$제조사,levels=c("엘파마","디엠",
+ "스마트"))
```

2.

```
> fullmodel2 <- glm(도수~연령*성별*제조사,data=bdat2,family="poisson")
> indmodel2 <- glm(도수~연령+성별+제조사,data=bdat2,family="poisson")
```

3.

```
> anova(indmodel2,fullmodel2,test="Chisq")
```

《검정 결과와 적합도 비교》 검정 결과는 유의하다. 포화모형의 적합이 상대적으로 좋다.

4.

```
> summary(fullmodel2)
```

《유의한 상호작용효과》 ‘연령 20대: 성별 F: 제조사 디엠’의 상호작용효과가 유의하다.

5.

```
> xtabs(fullmodel2$fitted.values~연령+제조사+성별,data=bdat2)
```

6.

```
> (frate <- ((837/1266)/(649/442))) #여성 비율
[1] 0.4502662
```

```
> (mrage <- ((744/888)/(626/432))) #남성 비율
[1] 0.5781884
```

```
> log(frate/mrate)
[1] -0.2500609
```

7.

20대 여성의 디엠 사용자 오즈비는 0.450이다. 이것은 기준이 되는 동성인 20대 엘마파 사용자 오즈의 0.45배로, 기준과 비교해 디엠 사용자가 상대적으로 적다는 것을 시사한다. 한편, 20대 남성의 디엠 사용자의 오즈비는 0.578로, 여성과 마찬가지로 기준과 비교해 디엠 사용자가 적다는 것을 알 수 있다. 양쪽 오즈비의 비는  $0.77854(=0.450/0.578)$ 이기 때문에 동일 기준으로 보면 남성보다 여성이 20대의 디엠 사용자가 상대적으로 적음을 알 수 있다.

## CHAPTER 11.

1.

```
> sks <- read.csv("자격시험.csv")
> sks$시험결과01 <- ifelse(sks$시험결과 == "합격", 1, 0)
> sks$기원01 <- ifelse(sks$기원 == "있음", 1, 0)
```

2.

```
> sks.out <- glm(시험결과01 ~ 공부시간 + 기원01 + 연령, family = "binomial", data = sks)
> summary(sks.out)
```

《변수의 계수에 관한 유의성》 공부시간 및 기원01의 계수가 유의하다.

3.

```
> exp(sks.out$coefficients)
```

《절편과 계수의 지수 변환값 해석》 절편의 변환값은 0.193이며, ‘공부시간이 0’ ‘합격기원을 하지 않는다’ ‘연령이 0’일 때, 합격의 오즈는 0.193(합격할 확률은 불합격 확률의 0.193배)이 된다. 계수의 변환값은 1.55이며, 공부시간이 1[시간] 증가하면, 합격의 오즈는 증가 전의 1.55배가 된다.

4.

```
> LRastdcoef(sks.out, c("공부시간", "연령"))
> exp(LRastdcoef(sks.out, c("공부시간", "연령")))
```

《계수의 지수 변환값 해석》 계수의 변환값은 1.36이고, 공부시간이  $z$ 점수로 1 단위 증가하면, 합격 오즈는 증가 전의 1.36배가 된다.

5.

```
> library(ResourceSelection)
> hoslem.test(x = sks.out$y, y = fitted(sks.out))
```

《모형의 적합》 검정 결과는 유의하지 않아 모형은 적합하다고 판단된다.

6.

```
> library(car)
> vif(sks.out)
```

《다중공선성 평가》 다중공선성이 생겼다고는 생각되지 않는다.

## CHAPTER 12.

1.

```
> tsks <- read.csv("도시의 기상.csv", row.names = 1)
```

2.

```
> tsks.stdz <- scale(tsks)
```

3.

```
> D0.stdz <- dist(tsks.stdz, method = "euclidean")
> D.stdz <- (1 / 2) * D0.stdz ^ 2
> tsks.stdz.out <- hclust(d = D.stdz, method = "ward.D")
> plot(as.dendrogram(tsks.stdz.out), xlim = c(100, 0), xlab = "비유사도",
+ horiz = TRUE)
```

4.

```
> from <- 1; to <- 11
> clabel <- function(x){factor(cutree(tsks.stdz.out, k = x))}
> clusters <- data.frame(lapply(from:to, clabel))
> names(clusters) <- from:to
> CNvalidity(dat = tsks.stdz, clusters = clusters)
```

《5개 군집의 타당성》 검토 대상인 2~10개의 군집수에서 5개 군집의 CH와 H는 중간이며, diffH가 비교적 크고, KL은 최대이다. 타당성을 강하게 지지하지는 않지만 상대적으로 볼 때 큰 문제는 없다고 할 수 있다.

5.

```
(cluster <- factor(cutree(tsks.stdz.out, k = 5)))
by(tsks.stdz, INDICES = cluster, FUN = function(x){apply(x, 2, mean)})
```

《군집 1과 군집 4 비교》 군집 1은 삿포로와 아오모리라는 2개 대상으로 구성되어 있다. 47개 도시 가운데 최저기온과 최고기온이 매우 낮고 강수량이 적고 강설량이 극단적으로 많다. 여름은 시원하고 겨울은 춥고 폭설이 내리며 장마철 비가 적다는 특징을 지닌 도시군이라 할 수 있다. 한편, 군집 4는 고치, 사가, 나가사키, 구마모토, 미야자키, 가고시마라는 6개 대상으로 구성되어 있다. 47개 도시 중에서는 흐린 날씨와 강수량이 많고 강설량이 적다. 가을에는 쾌청한 날씨인 경우가 많고, 장마철에 비가 많이 내리며, 겨울에는 눈이 내리기 어렵다는 특징을 지닌 도시군이라 할 수 있다.

## CHAPTER 13.

1.

```
> exdat <- read.csv("자전거데이터연습2.csv")
```

2.

```
> library(dummies)
```

```
> dexdat <- dummy.data.frame(exdat,sep=":")
```

3.

```
> library(FactoMineR)
```

```
> rdexdat <- CA(dexdat)
```

```
> summary(rdexdat)
```

《제2축까지의 누적기여율》 제2축까지의 누적기여율은 43.615.

4.

```
> rexdatt <- MCA(exdat)
```

```
> summary(rexdatt)
```

《제2축까지의 누적기여율》 제2축까지의 누적기여율은 43.615이고, [3.]에서 함수 CA의 출력과 일치한다.

5.

```
> dfexdat <- as.data.frame(xtabs(~.,data=exdat))
```

```
> dfexdat2 <- dfexdat[which(dfexdat$Freq>=1),]
```

6.

```
> rdfexdat2 <- MCA(dfexdat2,quanti.sup=7,row.w=dfexdat2$Freq)
```

```
> summary(rdfexdat2)
```

《제2축까지의 누적기여율》 제2축까지의 누적기여율은 43.615이고, [3.]과 [4.]의 출력과 일치한다.



## CHAPTER 14.

1.

```
> rtk <- read.csv("6도시의 기상.csv",
+ colClasses = c(rep("factor", 6), rep("numeric", 3)))
> rtk$도시 <- factor(rtk$도시, levels = c("가고시마", "다카마쓰", "가나자와", "나가노",
+ "요코하마", "아오모리"))
> rtk$월 <- factor(rtk$월, levels = as.character(1:12))
> rtk$일 <- factor(rtk$일, levels = as.character(1:31))
> rtk$계절 <- factor(rtk$계절, levels = c("봄", "여름", "가을", "겨울"))
> rtk$오전날씨 <- factor(rtk$오전날씨, levels = c("쾌청", "맑음", "다소흐림",
+ "흐림", "엷은안개", "안개", "안개비", "비", "진눈깨비", "눈", "우박", "천둥번개"))
> rtk$오후날씨 <- factor(rtk$오후날씨, levels = c("쾌청", "맑음", "다소흐림",
+ "흐림", "엷은안개", "안개", "안개비", "비", "진눈깨비", "눈", "우박", "천둥번개"))
```

2.

```
> library(ggplot2)
> ggplot(data = rtk, aes(x = 오전날씨)) + geom_bar()
```

3.

```
> ggplot(data = rtk, aes(x = 계절, group = 1))
+ + stat_summary(aes(y = 기온), fun = mean, geom = "line")
+ + stat_summary(aes(y = 기온), fun = mean, geom = "point")
```

4.

```
> library(dplyr)
> rtk2 <- rtk %>%
+ group_by(계절) %>%
+ summarise(평균기온 = mean(기온)) %>%
+ as.data.frame()
```

5.

```
> ggplot(data = rtk2, aes(x = 계절, y = 평균기온, group = 1))
+ + geom_line(stat = "identity") + geom_point(stat = "identity")
```

6.

```
> ggplot(data = rtk, aes(x = 오후날씨)) + geom_bar(aes(fill = 계절))  
+ + facet_grid( ~ 도시)
```