

KT하이텔 자료의 분석

이시아, 김도연
동국대학교 응용통계학과

분석목적

- 영역별 시험 성적과 영업 수익성과의 관계를 규명

자료 설명

- 1 KT하이텔 직원 50명을 대상으로 4가지 영역 시험의 성적과 영업 수익성 평가 지수
- 2 변수

변수명	변수설명	변수속성	범위
x1	창의력 평가 점수	연속형	
x2	단순추론 평가 점수	연속형	
x3	복합추론 평가 점수	연속형	
x4	계량능력시험 점수	연속형	
y	6개월간의 영업 수익성 평가 지수	연속형	

자료 설명

```
#library(XLConnect)
data_file <- "/var/www/data/multivariate/KT하이텔.csv"
KT <- read.csv(data_file)
#wb <- loadWorkbook(data_file)
#KT <- readWorksheet(wb, sheet=1, header=T)
names(KT) <- c("사원번호", "창의력", "단순추론", "복합추론", "계량능력", "영업수익성")
head(KT)
```

##	사원번호	창의력	단순추론	복합추론	계량능력	영업수익성
## 1	1	9	12	9	20	96.0
## 2	2	10	15	12	32	107.8
## 3	3	10	17	13	31	108.3
## 4	4	16	17	11	34	112.5
## 5	5	11	12	11	32	105.3
## 6	6	10	12	7	15	95.3

기초분석

```
kable(apply(KT[, -1], 2, summary))
```

	창의력	단순추론	복합추론	계량능력	영업수익성
Min.	1.00	5.00	4.00	9.00	87.3
1st Qu.	9.00	12.00	9.00	21.50	99.5
Median	10.00	15.00	11.00	31.50	106.2
Mean	11.28	14.18	10.36	29.76	106.6
3rd Qu.	14.00	17.00	12.00	37.00	114.8
Max.	18.00	20.00	15.00	51.00	122.3

상관관계

```
cor(KT[, -1])
```

##		창의력	단순추론	복합추론	계량능력	영업수익성
##	창의력	1.0000000	0.6010113	0.1032396	0.3937306	0.5310078
##	단순추론	0.6010113	1.0000000	0.4208298	0.5745533	0.7459097
##	복합추론	0.1032396	0.4208298	1.0000000	0.5477429	0.4982234
##	계량능력	0.3937306	0.5745533	0.5477429	1.0000000	0.9442960
##	영업수익성	0.5310078	0.7459097	0.4982234	0.9442960	1.0000000

회귀분석

```
m <- lm(영업수익성~., data=KT[, -1]); summary(m)
```

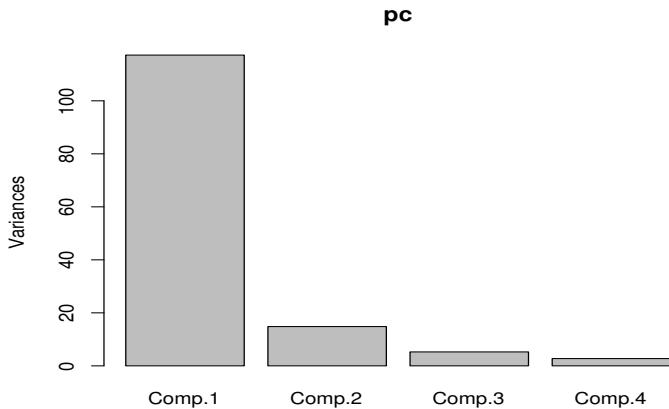
```
##
## Call:
## lm(formula = 영업수익성 ~ ., data = KT[, -1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.614 -1.264 -0.224  1.058  4.059
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 73.15526    1.68258  43.478 < 2e-16 ***
## 장의력      0.14245    0.10157   1.403  0.168
## 단순추론    0.84501    0.13186   6.408 7.72e-08 ***
## 복합추론   -0.27220    0.16825  -1.618  0.113
## 계량능력    0.76269    0.03949  19.312 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.138 on 45 degrees of freedom
## Multiple R-squared:  0.9591, Adjusted R-squared:  0.9554
```

주성분 분석 (PCA)

```
pc <- princomp(KT[, -c(1,6)])
```


주성분 분석 (PCA)

```
screplot(pc)
```



PCA를 이용한 회귀분석

```
pc$loadings
```

```
##  
## Loadings:  
##          Comp.1 Comp.2 Comp.3 Comp.4  
## 창의력   -0.165  0.848  0.401 -0.303  
## 단순추론 -0.197  0.469 -0.738  0.443  
## 복합추론 -0.117          -0.523 -0.841  
## 계량능력 -0.959 -0.233  0.147  
##  
##          Comp.1 Comp.2 Comp.3 Comp.4  
## SS loadings      1.00  1.00  1.00  1.00  
## Proportion Var   0.25  0.25  0.25  0.25  
## Cumulative Var   0.25  0.50  0.75  1.00
```

PCA를 이용한 회귀분석

```
head(pc$scores)
```

```
##           Comp.1      Comp.2      Comp.3      Comp.4
## [1,] 10.327716 -0.5775465 -0.02467923  0.2492278
## [2,] -2.291029 -1.3493765 -1.64716053 -0.4850043
## [3,] -1.843349 -0.2551249 -3.79235790 -0.5029902
## [4,] -5.477452  4.2897929  0.09792906 -0.4512250
## [5,] -1.747175 -1.8309493  1.49043537 -1.2773437
## [6,] 15.192768  1.5896735  0.68872556  1.3101148
```

PCA를 이용한 회귀분석

```
m2 <- lm(KT$영업수익성 ~ pc$scores[, 1:2]); summary(m2)
```

```
##  
## Call:  
## lm(formula = KT$영업수익성 ~ pc$scores[, 1:2])  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -6.3733 -1.5674  0.0801  1.6210  4.1693   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    106.62200    0.34655 307.665 < 2e-16 ***  
## pc$scores[, 1:2]Comp.1  -0.89003    0.03200 -27.810 < 2e-16 ***  
## pc$scores[, 1:2]Comp.2   0.36018    0.09002   4.001 0.000222 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 2.45 on 47 degrees of freedom  
## Multiple R-squared:  0.9438, Adjusted R-squared:  0.9414  
## F-statistic: 394.7 on 2 and 47 DF,  p-value: < 2.2e-16
```