#### Phần code:

## 8.1

invest = read.table('D:\\PTDL\\MultivariateStatisticalAnalysis\\data\\investment.txt',header = T)

head(invest)

rownames(invest) = invest[, 1]

invest <- invest[,-1]

print("Giá trị riêng")

eigen(cov(invest))$value

print("Vecto riêng")

eigen(cov(invest))$vecto

# Xác nhận căn của giá trị riêng của phương sai của các thành phần chính giống với giá trị nhận được từ princomp

sqrt(eigen(cor(invest))$values)

princomp(invest, cor = TRUE)

## 8.2

# a) Perform a factor analysis of this data

library(datasets)

data <- Harman23.cor

factanal(factors = 2, covmat = Harman23.cor)

# b)Vary the number of factors to find an adequate fit of the model and interpret the resulting factor loadings.

factanal(factors = 3, covmat = Harman23.cor)

factanal(factors = 4, covmat = Harman23.cor)

## Ta thấy với phân tích 4 nhân tố là mô hình phù hợp nhất do chỉ có 4 nhân tố là nhận được giá trị chi-square = 0.0988 > 0.05

# c) Kết quả không đổi khi thêm cor = True

factanal(factors = 4, covmat = Harman23.cor, cor = TRUE)

## 8.5

# a) Examine the pair-wise scatterplot for this data (with the pairs command) to reveal that some variables are very highly correlated

pairs(USJudgeRatings)

# b)Perform a principal components analysis for this data. The first two components explain 94 % of the variability. The second component is almost entirely the number of contacts, and the first component is essentially all other variables, all given the same weight.Interpret this result.

pc = princomp(USJudgeRatings)

summary(pc)

pc$loadings

## Từ kết quả cho thấy thành phần chính thứ 2 giải thích sự biến thiên chủ yếu bởi số lần liên hệ(contact), còn thành phần chính thứ nhất giải thích sự biến thiên của các biến khác và trọng số của chúng gần giống nhau cho thấy chúng có ảnh hưởng ngang nhau

## 8.6

# (a) Perform a factor analysis on the covariance matrix. Use the loadings to identify those variables that group together within the first two factors. Interpret these factors

library(datasets)

factanal(factors = 2, covmat = ability.cov)

## Kết quả phân tích thấy reading và vocab thành nhóm đại diện cho factor 1, blocks, picture và general đại diện cho factor 2

# (b)

pc<-princomp(ability.cor)

summary(pc)

pc$loadings

## Ở factor 1 reading và vocab đóng góp lớn nhất (0.559 và 0.584), ở factor 2 picture và maze có đóng góp lớn nhất

# c) Do you think it is more appropriate to examine the covariance or the correlation in a principal components analysis of this data?

# Ma trận tương quan phù hợp hơn

# d)

ability.cor <- cov2cor(ability.cov$cov)

princomp(ability.cor)

princomp(ability.cor)$loadings

## Sự khác biệt: phần a sử dụng ma trận hiệp phương sai còn đây sử dụng ma trận tương quan

## Sự giống nhau: đều thực hiện phương pháp pca để giảm số chiều dữ liệu và cho phép xác định các biến đóng góp nhiều nhất vào các thành phần chính

## 8.9

# (a) Draw a pairs plot to identify strong correlations and any outliers

pairs(trees)

# b)Perform a principal components analysis of this data. Use the biplot to show how the first principal component captures the strong correlation in part a. Again, identify any outliers in this data

pc <- princomp(trees)

summary(pc)

pc$loadings

biplot(pc, col = c(2, 3), cex = c(.5, 1),

xlim = c( -.45, .45),

xlab = "First principal component",

ylab = "Second principal component",

main = "Biplot for trees")

## Một số outliers là cây 20, 18

##### Phần chạy chương trình:

> ## 8.1

> invest = read.table('D:\\PTDL\\MultivariateStatisticalAnalysis\\data\\investment.txt',header = T)

> head(invest)

Manager S.US S.Non.US S.Dev. B.US B.Non.US B.Dev. A.native A.Cash

1 AllianceBernstein 45 3 17 35 0 0 0 0

2 AtlanticTrust 28 6 9 30 3 0 24 0

3 BankofAmerica 53 9 3 28 1 1 0 5

4 BNYMellon 26 9 10 30 0 0 25 0

5 Bessemer 19 9 3 20 4 5 34 6

6 BrownAdvisory 29 13 12 19 3 0 20 4

> rownames(invest) = invest[, 1]

> invest <- invest[,-1]

> print("Giá trị riêng")

[1] "Giá trị riêng"

> eigen(cov(invest))$value

[1] 232.91494484 58.04655915 41.92708537 12.34356035 7.28608180 5.18669915 2.22869539 0.06637395

> print("Vecto riêng")

[1] "Vecto riêng"

> eigen(cov(invest))$vecto

[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]

[1,] 0.705757774 0.52109931 -0.18159496 -0.15449336 0.03947393 -0.1881193 -0.09875614 -0.3561133

[2,] -0.051588920 -0.09809623 0.73437401 -0.24857494 0.47573011 -0.1956688 -0.02997825 -0.3480186

[3,] 0.051310949 -0.18753530 -0.13472067 0.84021553 0.28618859 -0.1629341 0.02924154 -0.3588585

[4,] 0.114476704 -0.71558390 -0.41296306 -0.39815245 -0.01911032 -0.1233086 -0.07744345 -0.3524247

[5,] -0.109077202 -0.01220358 0.31460354 0.13639820 -0.79963774 -0.2206973 -0.22355206 -0.3636895

[6,] -0.004402053 0.04215308 0.03412030 -0.05620262 -0.16064178 0.2571653 0.88688787 -0.3396630

[7,] -0.686304659 0.41022809 -0.37194014 -0.16592673 0.14949289 -0.2078608 -0.06704510 -0.3532399

[8,] -0.023989980 0.03728176 0.02347104 0.02366242 0.04774025 0.8523643 -0.37610546 -0.3558994

> # Xác nhận căn của giá trị riêng của phương sai của các thành phần chính giống với giá trị nhận được từ princomp

> sqrt(eigen(cor(invest))$values)

[1] 1.61880031 1.19121764 1.11581035 0.92824190 0.87350565 0.82640234 0.63750201 0.03819113

> princomp(invest, cor = TRUE)

Call:

princomp(x = invest, cor = TRUE)

Standard deviations:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8

1.61880031 1.19121764 1.11581035 0.92824190 0.87350565 0.82640234 0.63750201 0.03819113

8 variables and 27 observations.

> ## 8.2

> # a) Perform a factor analysis of this data

> library(datasets)

> data <- Harman23.cor

> factanal(factors = 2, covmat = Harman23.cor)

Call:

factanal(factors = 2, covmat = Harman23.cor)

Uniquenesses:

height arm.span forearm lower.leg weight bitro.diameter chest.girth chest.width

0.170 0.107 0.166 0.199 0.089 0.364 0.416 0.537

Loadings:

Factor1 Factor2

height 0.865 0.287

arm.span 0.927 0.181

forearm 0.895 0.179

lower.leg 0.859 0.252

weight 0.233 0.925

bitro.diameter 0.194 0.774

chest.girth 0.134 0.752

chest.width 0.278 0.621

Factor1 Factor2

SS loadings 3.335 2.617

Proportion Var 0.417 0.327

Cumulative Var 0.417 0.744

Test of the hypothesis that 2 factors are sufficient.

The chi square statistic is 75.74 on 13 degrees of freedom.

The p-value is 6.94e-11

> # b)Vary the number of factors to find an adequate fit of the model and interpret the resulting factor loadings.

> factanal(factors = 3, covmat = Harman23.cor)

Call:

factanal(factors = 3, covmat = Harman23.cor)

Uniquenesses:

height arm.span forearm lower.leg weight bitro.diameter chest.girth chest.width

0.127 0.005 0.193 0.157 0.090 0.359 0.411 0.490

Loadings:

Factor1 Factor2 Factor3

height 0.886 0.267 -0.130

arm.span 0.937 0.195 0.280

forearm 0.874 0.188

lower.leg 0.877 0.230 -0.145

weight 0.242 0.916 -0.106

bitro.diameter 0.193 0.777

chest.girth 0.137 0.755

chest.width 0.261 0.646 0.159

Factor1 Factor2 Factor3

SS loadings 3.379 2.628 0.162

Proportion Var 0.422 0.329 0.020

Cumulative Var 0.422 0.751 0.771

Test of the hypothesis that 3 factors are sufficient.

The chi square statistic is 22.81 on 7 degrees of freedom.

The p-value is 0.00184

> factanal(factors = 4, covmat = Harman23.cor)

Call:

factanal(factors = 4, covmat = Harman23.cor)

Uniquenesses:

height arm.span forearm lower.leg weight bitro.diameter chest.girth chest.width

0.137 0.005 0.191 0.116 0.138 0.283 0.178 0.488

Loadings:

Factor1 Factor2 Factor3 Factor4

height 0.879 0.277 -0.115

arm.span 0.937 0.194 0.277

forearm 0.875 0.191

lower.leg 0.887 0.209 0.135 -0.188

weight 0.246 0.882 0.111 -0.109

bitro.diameter 0.187 0.822

chest.girth 0.117 0.729 0.526

chest.width 0.263 0.644 0.141

Factor1 Factor2 Factor3 Factor4

SS loadings 3.382 2.595 0.323 0.165

Proportion Var 0.423 0.324 0.040 0.021

Cumulative Var 0.423 0.747 0.787 0.808

Test of the hypothesis that 4 factors are sufficient.

The chi square statistic is 4.63 on 2 degrees of freedom.

The p-value is 0.0988

> ## Ta thấy với phân tích 4 nhân tố là mô hình phù hợp nhất do chỉ có 4 nhân tố là nhận được giá trị chi-square = 0.0988 > 0.05

> # c) Kết quả không đổi khi thêm cor = True

> factanal(factors = 4, covmat = Harman23.cor, cor = TRUE)

Call:

factanal(factors = 4, covmat = Harman23.cor, cor = TRUE)

Uniquenesses:

height arm.span forearm lower.leg weight bitro.diameter chest.girth chest.width

0.137 0.005 0.191 0.116 0.138 0.283 0.178 0.488

Loadings:

Factor1 Factor2 Factor3 Factor4

height 0.879 0.277 -0.115

arm.span 0.937 0.194 0.277

forearm 0.875 0.191

lower.leg 0.887 0.209 0.135 -0.188

weight 0.246 0.882 0.111 -0.109

bitro.diameter 0.187 0.822

chest.girth 0.117 0.729 0.526

chest.width 0.263 0.644 0.141

Factor1 Factor2 Factor3 Factor4

SS loadings 3.382 2.595 0.323 0.165

Proportion Var 0.423 0.324 0.040 0.021

Cumulative Var 0.423 0.747 0.787 0.808

Test of the hypothesis that 4 factors are sufficient.

The chi square statistic is 4.63 on 2 degrees of freedom.

The p-value is 0.0988

> ## 8.5

> # a) Examine the pair-wise scatterplot for this data (with the pairs command) to reveal that some variables are very highly correlated

> pairs(USJudgeRatings)

> # b)Perform a principal components analysis for this data. The first two components explain 94 % of the variability. The second component is almost entirely the number of contacts, and the first component is essentially all other variables, all given the same weight.Interpret this result.

> pc = princomp(USJudgeRatings)

> summary(pc)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8

Standard deviation 2.9944812 0.97631616 0.54821237 0.46436892 0.260897305 0.172098999 0.127444176 0.107411902

Proportion of Variance 0.8476102 0.09010189 0.02840865 0.02038352 0.006434164 0.002799688 0.001535299 0.001090581

Cumulative Proportion 0.8476102 0.93771205 0.96612070 0.98650423 0.992938391 0.995738079 0.997273378 0.998363959

Comp.9 Comp.10 Comp.11 Comp.12

Standard deviation 0.0862216855 0.0703843063 0.0550575991 0.0434546479

Proportion of Variance 0.0007027259 0.0004682789 0.0002865415 0.0001784947

Cumulative Proportion 0.9990666848 0.9995349638 0.9998215053 1.0000000000

> pc$loadings

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

CONT 0.933 0.320 0.113

INTG -0.235 -0.139 0.370 0.252 0.463 -0.366 -0.418 0.377 -0.180 -0.160

DMNR -0.348 -0.232 0.663 -0.194 -0.361 0.394 0.167 0.123 0.113

DILG -0.287 -0.224 0.273 -0.376 0.564 0.255 0.283 0.416

CFMG -0.272 0.163 -0.189 -0.480 -0.170 0.109 -0.680 -0.269 -0.132 0.194

DECI -0.253 0.118 -0.249 -0.420 -0.369 -0.483 0.318 0.408 -0.187

PREP -0.309 -0.217 0.191 0.146 0.384 0.169 -0.641 -0.340 0.293

FAMI -0.305 -0.267 0.169 0.471 -0.108 0.229 -0.123 0.535 -0.468

ORAL -0.332 0.253 -0.142 -0.117 -0.272 0.355 -0.637 -0.430

WRIT -0.314 -0.115 0.142 0.295 -0.227 -0.102 -0.142 0.435 0.106 0.703

PHYS -0.278 -0.859 0.241 0.159 0.275

RTEN -0.359 0.196 -0.153 0.164 -0.466 0.294 -0.625 -0.155 0.245

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

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

Proportion Var 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083

Cumulative Var 0.083 0.167 0.250 0.333 0.417 0.500 0.583 0.667 0.750 0.833 0.917 1.000

> ## 8.6

> # (a) Perform a factor analysis on the covariance matrix. Use the loadings to identify those variables that group together within the first two factors. Interpret these factors

> library(datasets)

> factanal(factors = 2, covmat = ability.cov)

Call:

factanal(factors = 2, covmat = ability.cov)

Uniquenesses:

general picture blocks maze reading vocab

0.455 0.589 0.218 0.769 0.052 0.334

Loadings:

Factor1 Factor2

general 0.499 0.543

picture 0.156 0.622

blocks 0.206 0.860

maze 0.109 0.468

reading 0.956 0.182

vocab 0.785 0.225

Factor1 Factor2

SS loadings 1.858 1.724

Proportion Var 0.310 0.287

Cumulative Var 0.310 0.597

Test of the hypothesis that 2 factors are sufficient.

The chi square statistic is 6.11 on 4 degrees of freedom.

The p-value is 0.191

> # (b)

> pc<-princomp(ability.cor)

> summary(pc)

Importance of components:

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

Standard deviation 0.4776241 0.3448660 0.19810024 0.15709576 0.08183170 0

Proportion of Variance 0.5461757 0.2847480 0.09395716 0.05908658 0.01603257 0

Cumulative Proportion 0.5461757 0.8309237 0.92488085 0.98396743 1.00000000 1

> pc$loadings

Loadings:

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

general 0.136 0.226 0.794 0.522 0.161

picture -0.280 0.683 -0.284 0.157 0.589

blocks -0.263 0.252 0.499 -0.783

maze -0.366 -0.645 0.179 0.644

reading 0.599 -0.708 0.358

vocab 0.584 -0.283 0.683 0.327

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

SS loadings 1.000 1.000 1.000 1.000 1.000 1.000

Proportion Var 0.167 0.167 0.167 0.167 0.167 0.167

Cumulative Var 0.167 0.333 0.500 0.667 0.833 1.000

> # d)

> ability.cor <- cov2cor(ability.cov$cov)

> princomp(ability.cor)

Call:

princomp(x = ability.cor)

Standard deviations:

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

0.4776241 0.3448660 0.1981002 0.1570958 0.0818317 0.0000000

6 variables and 6 observations.

> princomp(ability.cor)$loadings

Loadings:

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

general 0.136 0.226 0.794 0.522 0.161

picture -0.280 0.683 -0.284 0.157 0.589

blocks -0.263 0.252 0.499 -0.783

maze -0.366 -0.645 0.179 0.644

reading 0.599 -0.708 0.358

vocab 0.584 -0.283 0.683 0.327

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

SS loadings 1.000 1.000 1.000 1.000 1.000 1.000

Proportion Var 0.167 0.167 0.167 0.167 0.167 0.167

Cumulative Var 0.167 0.333 0.500 0.667 0.833 1.000

> ## 8.9

> # (a) Draw a pairs plot to identify strong correlations and any outliers

> pairs(trees)

> # b)Perform a principal components analysis of this data. Use the biplot to show how the first principal component captures the strong correlation in part a. Again, identify any outliers in this data

> pc <- princomp(trees)

> summary(pc)

Importance of components:

Comp.1 Comp.2 Comp.3

Standard deviation 16.9039975 4.90098994 0.736412856

Proportion of Variance 0.9208461 0.07740622 0.001747636

Cumulative Proportion 0.9208461 0.99825236 1.000000000

> pc$loadings

Loadings:

Comp.1 Comp.2 Comp.3

Girth 0.176 0.980

Height 0.242 -0.969

Volume 0.954 0.229 -0.192

Comp.1 Comp.2 Comp.3

SS loadings 1.000 1.000 1.000

Proportion Var 0.333 0.333 0.333

Cumulative Var 0.333 0.667 1.000

> biplot(pc, col = c(2, 3), cex = c(.5, 1),

+ xlim = c( -.45, .45),

+ xlab = "First principal component",

+ ylab = "Second principal component",

+ main = "Biplot for trees")





