

Factor Analysis - 19BCE1460

```
library(corpcor)
library(GPArotation)
library(psych)
library(psy)
```

```
##
## Attaching package: 'psy'

## The following object is masked from 'package:psych':
##
##      wkappa
```

```
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##      %+%, alpha
```

```
library(MASS)
library(MVN)
```

```
## Warning: package 'MVN' was built under R version 4.1.3
```

```
require("datasets")
penguinsdata <- read.csv("S:/WIN SEM 21-22/Data Visualization/Lab/penguins_lter.csv")
penguinsdata <- na.omit(penguinsdata)
penguins.new<- penguinsdata[,c(10,11,12,13)]
penguinsmatrix <- cor(penguins.new)
round(penguinsmatrix,2)
```

```
##
##      CulmenLengthmm CulmenDepthmm FlipperLengthmm BodyMassg
## CulmenLengthmm      1.00         -0.23           0.65      0.59
## CulmenDepthmm       -0.23          1.00          -0.58     -0.47
## FlipperLengthmm      0.65         -0.58           1.00      0.87
## BodyMassg           0.59         -0.47           0.87      1.00
```

```

X <- cor(penguins.new)
iX <- ginv(X)
S2 <- diag(diag((iX^-1)))
AIS <- S2*iX*S2 # anti-image covariance matrix
IS <- X+AIS-2*S2 # image covariance matrix
Dai <- sqrt(diag(diag(AIS)))
IR <- ginv(Dai)%*%IS%*%ginv(Dai) # image correlation matrix
AIR <- ginv(Dai)%*%AIS%*%ginv(Dai) # anti-image correlation matrix

print(diag(AIR), row.names = FALSE)

```

```
## [1] 1 1 1 1
```

```

#Measure of Sampling Accuracy (MSA) using KMO Test
penguinsmatrix <- cor(penguins.new)
KMO(penguinsmatrix)

```

```

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = penguinsmatrix)
## Overall MSA = 0.68
## MSA for each item =
## CulmenLengthmm CulmenDepthmm FlipperLengthmm BodyMassg
## 0.79 0.69 0.62 0.70

```

```

#Bartlett's Test of Sphericity
penguinsmatrix <- cor(penguins.new)
cortest.bartlett(penguinsmatrix)

```

```
## Warning in cortest.bartlett(penguinsmatrix): n not specified, 100 used
```

```

## $chisq
## [1] 240.4193
##
## $p.value
## [1] 4.567573e-49
##
## $df
## [1] 6

```

```

#PCA
pc1 <- principal(penguins.new, nfactors = length(penguins.new), rotate = "none")
pc1

```

```

## Principal Components Analysis
## Call: principal(r = penguins.new, nfactors = length(penguins.new),
## rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
## PC1 PC2 PC3 PC4 h2 u2 com
## CulmenLengthmm 0.75 0.53 -0.39 -0.04 1 1.1e-16 2.4
## CulmenDepthmm -0.66 0.70 0.26 0.05 1 1.8e-15 2.3

```

```
## FlipperLengthmm 0.96 0.00 0.15 0.25 1 1.0e-15 1.2
## BodyMassg      0.91 0.07 0.36 -0.19 1 1.0e-15 1.4
##
##              PC1  PC2  PC3  PC4
## SS loadings    2.75 0.78 0.37 0.11
## Proportion Var 0.69 0.20 0.09 0.03
## Cumulative Var 0.69 0.88 0.97 1.00
## Proportion Explained 0.69 0.20 0.09 0.03
## Cumulative Proportion 0.69 0.88 0.97 1.00
##
## Mean item complexity = 1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0
## with the empirical chi square 0 with prob < NA
##
## Fit based upon off diagonal values = 1
```

```
cat("Eigen values\n")
```

```
## Eigen values
```

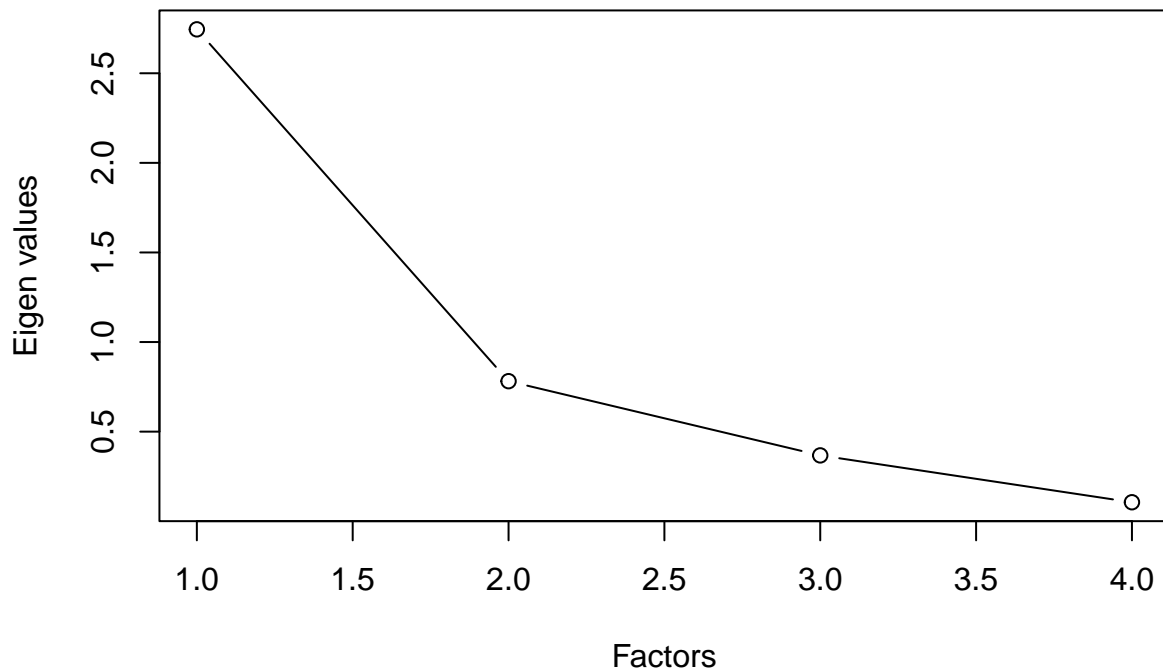
```
print(pc1$values)
```

```
## [1] 2.7451026 0.7815942 0.3673524 0.1059508
```

```
#Scree plot
```

```
plot(pc1$values, type = "b", xlab = "Factors", ylab = "Eigen values", main = "SCREE PLOT")
```

SCREE PLOT



```
#Principal axis factoring
```

```
penguinsmatrix <- cor(penguins.new)
```

```
solution <- fa(r=penguinsmatrix, nfactors = 2, rotate = "none", fm = "pa")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :  
## The estimated weights for the factor scores are probably incorrect. Try a  
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An  
## ultra-Heywood case was detected. Examine the results carefully
```

```
print(solution)
```

```
## Factor Analysis using method = pa
```

```
## Call: fa(r = penguinsmatrix, nfactors = 2, rotate = "none", fm = "pa")
```

```
## Standardized loadings (pattern matrix) based upon correlation matrix
```

```
##          PA1    PA2    h2    u2 com  
## CulmenLengthmm  0.66  0.35  0.56  0.444 1.5  
## CulmenDepthmm  -0.56  0.41  0.48  0.519 1.8  
## FlipperLengthmm 1.01 -0.04  1.01 -0.013 1.0  
## BodyMassg       0.87  0.05  0.76  0.236 1.0
```

```
##
```

```
##          PA1    PA2
```

```
## SS loadings      2.52  0.29
```

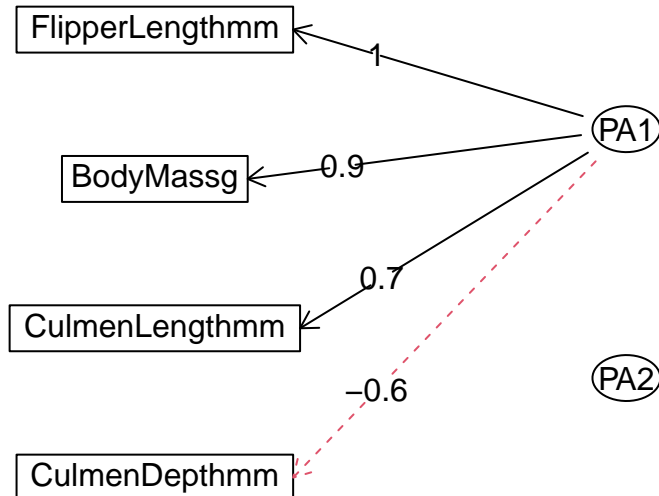
```
## Proportion Var      0.63 0.07
## Cumulative Var      0.63 0.70
## Proportion Explained 0.90 0.10
## Cumulative Proportion 0.90 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 6 and the objective function was 2.48
## The degrees of freedom for the model are -1 and the objective function was 0
##
## The root mean square of the residuals (RMSR) is 0
## The df corrected root mean square of the residuals is NA
##
## Fit based upon off diagonal values = 1
```

```
print(solution$loadings)
```

```
##
## Loadings:
##          PA1    PA2
## CulmenLengthmm  0.660 0.348
## CulmenDepthmm  -0.559 0.410
## FlipperLengthmm 1.006
## BodyMassg      0.873
##
##          PA1    PA2
## SS loadings  2.521 0.294
## Proportion Var 0.630 0.073
## Cumulative Var 0.630 0.704
```

```
fa.diagram(solution)
```

Factor Analysis



#Orthogonal varimax rotation

```
penguinsmatrix <- cor(penguins.new)
solution1 <- fa(r=penguinsmatrix, nfactors = 2, rotate = "varimax", fm = "pa")
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```

```
print(solution1)
```

```
## Factor Analysis using method = pa
## Call: fa(r = penguinsmatrix, nfactors = 2, rotate = "varimax", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##              PA1  PA2  h2    u2 com
## CulmenLengthmm  0.73  0.15 0.56  0.444 1.1
## CulmenDepthmm  -0.18 -0.67 0.48  0.519 1.1
## FlipperLengthmm 0.75  0.67 1.01 -0.013 2.0
## BodyMassg       0.71  0.51 0.76  0.236 1.8
##
##              PA1  PA2
## SS loadings    1.64 1.18
## Proportion Var  0.41 0.29
## Cumulative Var  0.41 0.70
## Proportion Explained 0.58 0.42
## Cumulative Proportion 0.58 1.00
##
```

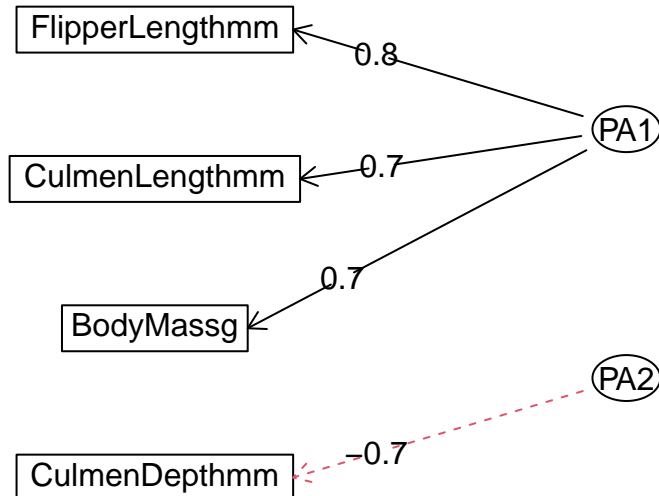
```
## Mean item complexity = 1.5
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 6 and the objective function was 2.48
## The degrees of freedom for the model are -1 and the objective function was 0
##
## The root mean square of the residuals (RMSR) is 0
## The df corrected root mean square of the residuals is NA
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors    PA1  PA2
## Multiple R square of scores with factors          0.86 0.82
## Minimum correlation of possible factor scores      0.74 0.67
## Minimum correlation of possible factor scores      0.47 0.34
```

```
print(solution1$loadings)
```

```
##
## Loadings:
##          PA1    PA2
## CulmenLengthmm  0.732 0.146
## CulmenDepthmm  -0.176 -0.671
## FlipperLengthmm 0.753 0.668
## BodyMassg      0.709 0.511
##
##          PA1    PA2
## SS loadings  1.636 1.179
## Proportion Var 0.409 0.295
## Cumulative Var 0.409 0.704
```

```
fa.diagram(solution1)
```

Factor Analysis



#Oblique rotation (oblimin)

```
penguinsmatrix <- cor(penguins.new)
```

```
solution2 <- fa(r=penguinsmatrix, nfactors = 2, rotate = "oblimin", fm = "pa")
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## An ultra-Heywood case was detected. Examine the results carefully
```

```
print(solution2)
```

```
## Factor Analysis using method = pa
```

```
## Call: fa(r = penguinsmatrix, nfactors = 2, rotate = "oblimin", fm = "pa")
```

```
## Standardized loadings (pattern matrix) based upon correlation matrix
```

```
##          PA1  PA2  h2    u2 com
## CulmenLengthmm  0.74 0.31 0.56 0.444 1.3
## CulmenDepthmm -0.44 0.45 0.48 0.519 2.0
## FlipperLengthmm 0.98 -0.11 1.01 -0.013 1.0
## BodyMassg      0.87 -0.01 0.76 0.236 1.0
```

```
##
```

```
##          PA1  PA2
```

```
## SS loadings      2.48 0.33
```



```
## Proportion Var      0.62 0.08
## Cumulative Var      0.62 0.70
## Proportion Explained 0.88 0.12
## Cumulative Proportion 0.88 1.00
##
## With factor correlations of
##      PA1  PA2
## PA1  1.0 -0.2
## PA2 -0.2  1.0
##
## Mean item complexity = 1.3
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 6 and the objective function was 2.48
## The degrees of freedom for the model are -1 and the objective function was 0
##
## The root mean square of the residuals (RMSR) is 0
## The df corrected root mean square of the residuals is NA
##
## Fit based upon off diagonal values = 1
```

```
print(solution2$loadings)
```

```
##
## Loadings:
##      PA1    PA2
## CulmenLengthmm  0.742  0.311
## CulmenDepthmm  -0.443  0.454
## FlipperLengthmm 0.979 -0.110
## BodyMassg      0.873
##
##      PA1    PA2
## SS loadings  2.467 0.315
## Proportion Var 0.617 0.079
## Cumulative Var 0.617 0.695
```

```
cronbach(penguins.new)
```

```
## $sample.size
## [1] 330
##
## $number.of.items
## [1] 4
##
## $alpha
## [1] 0.04652234
```

```
#Orthogonal Rotation (varimax)
fa.diagram(solution1)
```

```
#Oblique Rotation (oblimin)
fa.diagram(solution2)
```

Factor Analysis

