Assignment 2 STAT 315-463: Multivariable Statistical Methods and Applications

Lisa Lu 31088272

Due date: Friday 24 March 2023

- Your assignment needs to show the R code you used, and your well discussed answers to the questions.
- Submit your assignments on Learn.

Background

In the dataset, `USJudgeRatings.csv`, you are presented with ratings of State Judges on the Superior Court on 12 variables provided by 43 Lawyers in 1977.

CONT	Number of contacts of lawyer with judge	INTG	Judicial integrity	DMNR	Demeanour
DILG	Diligence	CFMG	Case flow managing	DECI	Prompt decisions
PREP	Preparation for trial	FAMI	Familiarity with law	ORAL	Sound oral rulings
WRIT	Sound written rulings	PHYS	Physical ability	RTEN	Worthy of retention

```
# Read in the data
dataset <- read.csv("USJudgeRatings.csv")
```

Principal Component Analysis of the Rating Data.

Perform a PCA on the standardised ratings. Note, you will need to standardise the ratings yourself. Then answer the following questions.

```
# Standardise the ratings
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
dataset_std <- dataset %>% mutate(across(where(is.numeric),scale))
summary(dataset_std)
```

```
##     Lawyer              CONT.V1           INTG.V1
## Length:43        Min.   :-1.846373   Min.   :-2.7539372
## Class :character  1st Qu.:-0.624109   1st Qu.:-0.6114828
## Mode  :character  Median :-0.145831   Median : 0.1026687
##                   Mean   : 0.000000   Mean   : 0.0000000
##                   3rd Qu.: 0.491872   3rd Qu.: 0.6869745
##                   Max.   : 3.361535   Max.   : 1.5309717
##        DMNR.V1            DILG.V1           CFMG.V1
## Min.   :-2.8121570  Min.   :-2.8782657  Min.   :-2.4172132
## 1st Qu.:-0.5388442  1st Qu.:-0.6027579  1st Qu.:-0.5569865
## Median : 0.1606366  Median : 0.1187446  Median : 0.1405985
## Mean   : 0.0000000  Mean   : 0.0000000  Mean   : 0.0000000
## 3rd Qu.: 0.7289648  3rd Qu.: 0.8402471  3rd Qu.: 0.6637873
## Max.   : 1.2972929  Max.   : 1.4507492  Max.   : 1.4195044
##        DECI.V1            PREP.V1           FAMI.V1
## Min.   :-2.3228698  Min.   :-2.7979078  Min.   :-2.5167602
## 1st Qu.:-0.5792693  1st Qu.:-0.5951957  1st Qu.:-0.5673125
## Median : 0.1679881  Median : 0.2439327  Median : 0.1176285
## Mean   : 0.0000000  Mean   : 0.0000000  Mean   : 0.0000000
## 3rd Qu.: 0.7284311  3rd Qu.: 0.7683879  3rd Qu.: 0.8025696
## Max.   : 1.5379600  Max.   : 1.7124074  Max.   : 1.6982618
##        ORAL.V1            WRIT.V1           PHYS.V1
## Min.   :-2.5672387  Min.   :-2.5841599  Min.   :-3.442921
## 1st Qu.:-0.4386179  1st Qu.:-0.5032821  1st Qu.:-0.249989
## Median : 0.2049186  Median : 0.2250252  Median : 0.175735
## Mean   : 0.0000000  Mean   : 0.0000000  Mean   : 0.000000
## 3rd Qu.: 0.6999467  3rd Qu.: 0.6932227  3rd Qu.: 0.601459
## Max.   : 1.5909973  Max.   : 1.6816397  Max.   : 1.240046
##        RTEN.V1
## Min.   :-2.5453217
## 1st Qu.:-0.4108424
## Median : 0.1795455
## Mean   : 0.0000000
## 3rd Qu.: 0.5882756
## Max.   : 1.4511502
```

1. How many principal components do you believe should be retained. Justify your answer by looking at the variation in the data explained by each component.

```
sapply(dataset[,-1], var)
```

```
##     CONT      INTG      DMNR      DILG      CFMG      DECI      PREP      FAMI
## 0.8852492 0.5931229 1.3080620 0.8116168 0.7397896 0.6447065 0.9089147 0.9005759
##     ORAL      WRIT      PHYS      RTEN
## 1.0201883 0.9237763 0.8828018 1.2121373
```

2. In your own words, describe what you believe the first principal component is measuring.

3. What do you think the second principal component represents?

4. You are told *Judicial Integrity* and *Demeanour* are particularly important traits, and should be given 5 times the weight of the other variables. Re-run the Principal Component Analysis such that Integrity and Demeanour is given 5 times the weight of all other variables. What impact does this have?

Factor Analysis for the Rating Data

Perform Factor Analysis on the standardised Ratings Data.

1. What happens when you try to fit a 3 and a 4 factor solution with no rotation. Hint: For the three factor solution, you may need to add `control=list(nstart=100)` as an additional argument in the `factanal` function.

2. Which variables are grouped by the first two factors? (e.g. threshold $|\text{loading}| \geq 0.25$)

3. Compare the factor loadings for the first two factors to the first two principal components of the standardised data found in the previous section. Comment on any similarities and/or differences.

4. Comment on the observed variable specific variances (the uniquenesses). Do you believe all observed variables are explained by the factors discovered.

5. Re-fit the 3 factor solution with a varimax rotation. How does this change the interpretation of the factors? In this case, do you find the rotated or non-rotated solution easier to interpret. Explain why or why not?