Assignment 7

Jennifer M. Wanat

Introduction

The correlation matrix was re-created from a study by Stoetzel (1960) to explain consumer preferences for nine liquors. Originally the preferences were described in terms of their sweetness, price and regional popularity. Factor analysis models containing different numbers of factors will be generated utilizing the VARIMAX orthogonal rotation technique and maximum likelihood factor analysis. Model goodness-of-fit will be assessed. An additional model utilizing the PROMAX oblique rotation technique and maximum likelihood factor analysis for comparison.

Data Preparation

Factor analysis was used for the liquor preference of French consumers in February 1956 in a publication by Jean Stoetzel. Personal interviews were conducted with 2,014 adult men and women who were selected by quota sampling methods to represent a cross-section of the French population. A total of 1,442 completed interviews were used for the factor analysis. Respondents were asked to rank liquors by their personal preference. The names of nine liquors were provided: Armagnac, Calvados, Cognac, Kirsch, Marc, Mirabelle, Rum, Whiskey, Liqueurs. The correlation coefficients were computed and are provided below. The correlation coefficients were placed into a matrix called `cor.matrix`. A positive correlation between liquors indicates that a respondent liked both liquors, while a negative correlation between liquors indicates that one liquor was preferred over the other.

Table 1: Correlation Table

	Armagnac	Calvados	Cognac	Kirsch	Marc	Mirabelle	Rum	Whiskey	Liqueurs
Armagnac	1.00	0.21	0.37	-0.32	0.00	-0.31	-0.26	0.09	-0.38
Calvados	0.21	1.00	0.09	-0.29	0.12	-0.30	-0.14	0.01	-0.39
Cognac	0.37	0.09	1.00	-0.31	-0.04	-0.30	-0.11	0.12	-0.39
Kirsch	-0.32	-0.29	-0.31	1.00	-0.16	0.25	-0.13	-0.14	0.90
Marc	0.00	0.12	-0.04	-0.16	1.00	-0.20	-0.03	-0.08	-0.38
Mirabelle	-0.31	-0.30	-0.30	0.25	-0.20	1.00	-0.24	-0.16	0.18
Rum	-0.26	-0.14	-0.11	-0.13	-0.03	-0.24	1.00	-0.20	0.04
Whiskey	0.09	0.01	0.12	-0.14	-0.08	-0.16	-0.20	1.00	-0.24
Liqueurs	-0.38	-0.39	-0.39	0.90	-0.38	0.18	0.04	-0.24	1.00

Section 1: Data Check

The data matrix was checked to ensure that it was a matrix object that that it was symmetric. Data checks were confirmed to be true.

```
[1] "Is the cor.matrix a matrix:  TRUE"
```

```
[1] "What is the type of the matrix:  numeric"
```

```
[1] "Is the matrix symmetric:  TRUE"
```

The correlation matrix was constructed into a graphical display. The correlation coefficient is proportional to the color and size of the circle. Positive correlation is blue and negative correlation is red. The large dark blue circle visualized diagonally across the plot represents a correlation of 1, and it is the correlation of the liquor against itself. Kirsch and Liqueurs have a high positive correlation between each other. Kirsch, Mirabelle, and Liqueurs have a moderately negative correlation with Armagnac, Calvados, and Cognac. Marc and Liqueurs also have a moderate negative correlation. Cognac and Armagnac have a moderate positive correlation.

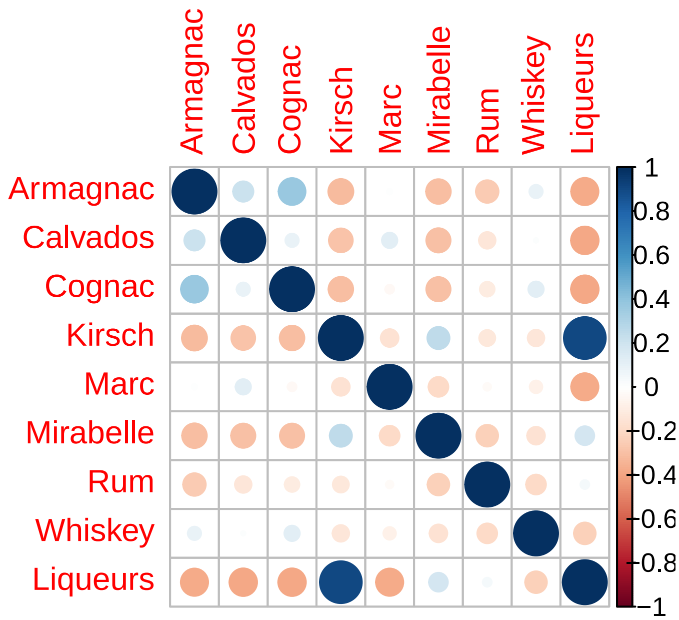

Figure 1: Graphical representation of liquor correlation.

Section 2: Factor Analysis - VARIMAX Rotation

Section 2.1: Three Factor Model

Since the publication by Stoetzel utilized a three factor model, an estimate of a three factor model with a VARIMAX rotation using maximum likelihood factor analysis was generated. The VARIMAX rotation is an orthogonal rotation that will result in uncorrelated factors. The maximum likelihood factor analysis is a statistical approach to factor analysis and allows for formal inference of factor loadings and goodness-of-fit assessment.

The three factor model results are:

```
Call:
factanal(factors = 3, covmat = cor.matrix, n.obs = 1442, rotation = "varimax")

Uniquenesses:
 Armagnac  Calvados    Cognac    Kirsch      Marc Mirabelle       Rum
    0.759     0.792     0.739     0.134     0.005     0.005     0.933
  Whiskey  Liqueurs
    0.890     0.005

Loadings:
          Factor1 Factor2 Factor3
Armagnac  -0.450  -0.023   0.193
Calvados  -0.411   0.100   0.172
```

```
Cognac    -0.473  -0.064   0.183
Kirsch     0.921  -0.121   0.043
Marc      -0.044   0.996   0.020
Mirabelle  0.293  -0.169  -0.938
Rum       -0.017  -0.036   0.256
Whiskey   -0.305  -0.094   0.091
Liqueurs   0.923  -0.344   0.158

               Factor1 Factor2 Factor3
SS loadings      2.477   1.179   1.082
Proportion Var   0.275   0.131   0.120
Cumulative Var   0.275   0.406   0.527

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 1820.72 on 12 degrees of freedom.
The p-value is 0
```

The factor loadings were plotted against each other. Marc loads on the second factor, while Kirsch and Liqueurs load on the first factor.

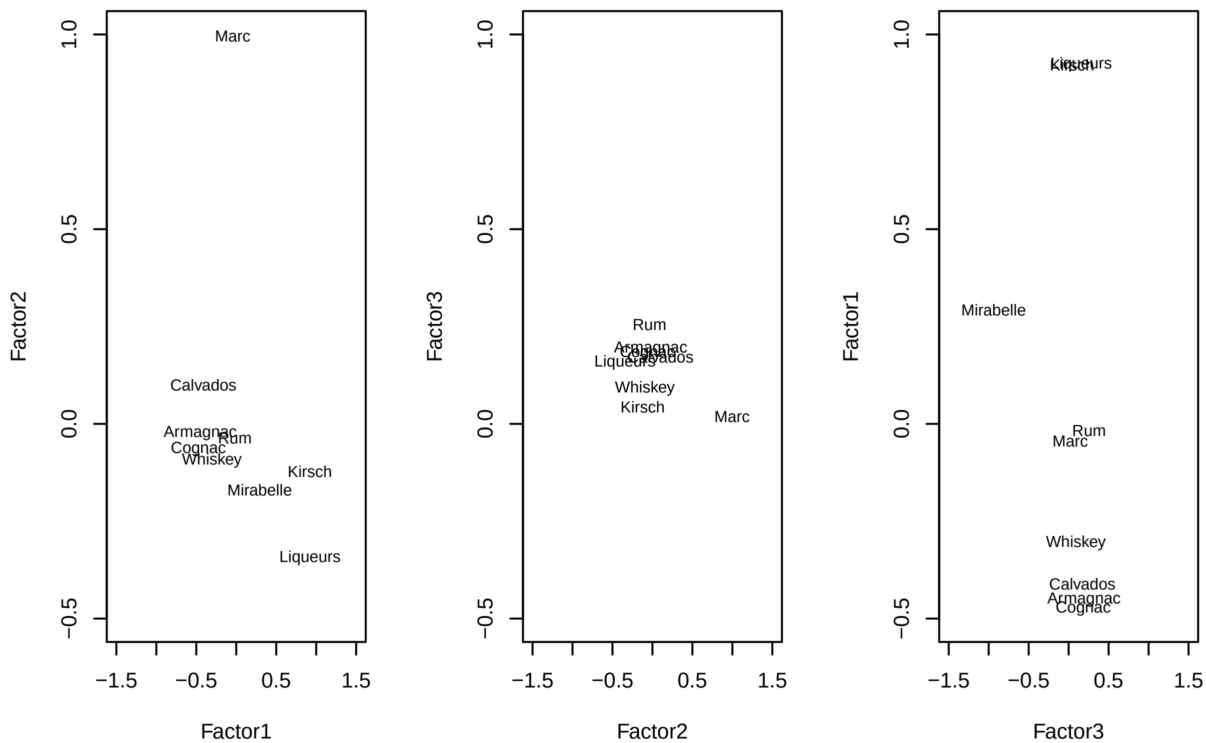

Figure 2: Loading plots for the three factor model with VARIMAX rotation.

The same factor loadings presented by Stoetzel were not obtained. However, the results were qualitatively the same and the same factor interpretations could be made.

The Uniquenesses is the amount of variance not accounted for by the components. Armagnac, Calvados, Cognac, Rum and Whiskey all have uniquenesses greater than 0.75. These variables are not well represented by the three factor model.

The output row labeled SS loadings contains the eigenvalues associated with the factors. The Proportion Var is the amount of variance accounted for by each factor. The three factors account for 52.7% of the variance in the nine liquors. The chi-square statistic is 1820.72 on 12 degrees of freedom with a p-value of 0. This indicates that the null hypothesis of a perfect model fit (that the three factors are sufficient) is rejected.

Section 2.2: One, Two, Four, and Five Factor Models

Additional factor models were generated utilizing one, two, four, five and six factors in an effort to find the correct number of factors needed to describe the correlation matrix.

The one factor model results are:

```
Call:
factanal(factors = 1, covmat = cor.matrix, n.obs = 1442, rotation = "varimax")

Uniquenesses:
 Armagnac  Calvados    Cognac    Kirsch      Marc Mirabelle       Rum 
    0.855     0.848     0.848     0.187     0.859     0.966     0.999 
  Whiskey  Liqueurs 
    0.943     0.005 

Loadings:
          Factor1
Armagnac  -0.381 
Calvados  -0.390 
Cognac    -0.390 
Kirsch     0.902 
Marc      -0.376 
Mirabelle  0.184 
Rum        0.037 
Whiskey   -0.238 
Liqueurs   0.998 

               Factor1
SS loadings      2.490
Proportion Var   0.277

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 2972.05 on 27 degrees of freedom.
The p-value is 0
```

The Uniquenesses for Kirsch is 0.187 and Liqueurs is 0.005. All other liquors have uniquenesses greater than 0.848. The proportion of variance accounted for the single factor is 27.7%. The chi-square statistic is 2972.05 with a p-value of 0. This indicates that the null hypothesis of a perfect model fit (that the one factor is sufficient) is rejected.

The two factor model results are:

```
Call:
factanal(factors = 2, covmat = cor.matrix, n.obs = 1442, rotation = "varimax")

Uniquenesses:
 Armagnac  Calvados    Cognac    Kirsch      Marc Mirabelle       Rum 
    0.831     0.848     0.807     0.148     0.005     0.946     0.998 
  Whiskey  Liqueurs 
    0.910     0.005 
```

```
Loadings:
          Factor1 Factor2
Armagnac  -0.404   0.074 
Calvados  -0.343   0.185 
Cognac    -0.437   0.039 
Kirsch     0.866  -0.321 
Marc       0.178   0.981 
Mirabelle  0.082  -0.217 
Rum        0.021  -0.035 
Whiskey   -0.298  -0.026 
Liqueurs   0.839  -0.540 

               Factor1 Factor2
SS loadings      2.053   1.448
Proportion Var   0.228   0.161
Cumulative Var   0.228   0.389

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 2448.36 on 19 degrees of freedom.
The p-value is 0
```

The Uniquenesses for Kirsch is 0.148 and 0.005 for both Marc and Liqueurs. All other liquors have uniquenesses greater than 0.807. The proportion of variance accounted for the two factors is 38.9%. The chi-square statistic is 2448.36 with a p-value of 0. This indicates that the null hypothesis of a perfect model fit (that the two factors are sufficient) is rejected.

The four factor model results are:

```
Call:
factanal(factors = 4, covmat = cor.matrix, n.obs = 1442, rotation = "varimax")

Uniquenesses:
 Armagnac  Calvados    Cognac    Kirsch      Marc Mirabelle       Rum 
    0.696     0.727     0.639     0.005     0.005     0.005     0.781 
  Whiskey  Liqueurs 
    0.765     0.005 

Loadings:
          Factor1 Factor2 Factor3 Factor4
Armagnac  -0.189   0.515          -0.057 
Calvados  -0.172   0.477   0.121  -0.032 
Cognac    -0.176   0.573  -0.037  -0.014 
Kirsch     0.943  -0.252  -0.028   0.206 
Marc      -0.151  -0.051   0.985  -0.016 
Mirabelle -0.061  -0.536  -0.227   0.808 
Rum       -0.092  -0.161  -0.059  -0.426 
Whiskey   -0.048   0.469  -0.061   0.094 
Liqueurs   0.848  -0.438  -0.280  -0.082 

               Factor1 Factor2 Factor3 Factor4
SS loadings      1.740   1.611   1.123   0.897
Proportion Var   0.193   0.179   0.125   0.100
Cumulative Var   0.193   0.372   0.497   0.597
```

```
Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 968.7 on 6 degrees of freedom.
The p-value is 5.26e-206
```

The Uniquenesses for Kirsch, Marc and Liqueurs is 0.005. All other liquors have uniquenesses greater than 0.639. The proportion of variance accounted for the four factors is 59.7%. The chi-square statistic is 968.7 with a p-value of 5.26e-206. This indicates that the null hypothesis of a perfect model fit (that the four factors are sufficient) is rejected.

The five factor model results are:

```
Call:
factanal(factors = 5, covmat = cor.matrix, n.obs = 1442, rotation = "varimax")

Uniquenesses:
 Armagnac  Calvados    Cognac    Kirsch      Marc Mirabelle       Rum
    0.660     0.664     0.603     0.005     0.005     0.005     0.768
  Whiskey  Liqueurs
    0.005     0.005

Loadings:
          Factor1 Factor2 Factor3 Factor4 Factor5
Armagnac  -0.164   0.555  -0.022   0.047  -0.047
Calvados  -0.141   0.552   0.101  -0.027  -0.023
Cognac    -0.150   0.605  -0.058   0.071
Kirsch     0.937  -0.279  -0.025  -0.042   0.193
Marc      -0.142           0.986  -0.046  -0.019
Mirabelle -0.064  -0.497  -0.207  -0.228   0.806
Rum       -0.092  -0.096  -0.059  -0.171  -0.426
Whiskey   -0.110   0.053  -0.049   0.985   0.091
Liqueurs   0.834  -0.446  -0.273  -0.132  -0.093

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      1.687   1.515   1.110   1.080   0.888
Proportion Var   0.187   0.168   0.123   0.120   0.099
Cumulative Var   0.187   0.356   0.479   0.599   0.698

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 674.73 on 1 degree of freedom.
The p-value is 9.34e-149
```

The Uniquenesses for Kirsch, Marc, Mirabelle, Whiskey and Liqueurs is 0.005. All other liquors have uniquenesses greater than 0.603. The proportion of variance accounted for the four factors is 69.8%. The chi-square statistic is 674.73 with a p-value of 9.34e-149. This indicates that the null hypothesis of a perfect model fit (that the five factors are sufficient) is rejected.

An error is received when a six factor model is attempted to be generated. There are not enough degrees of freedom for six factors. None of the factor models generated with the VARIMAX factor rotation are sufficient to describe the correlation matrix. An alternative rotation technique, oblique versus orthogonal, will be used to determine if an improved model can be generated.

Section 3: Factor Analysis - PROMAX Rotation

An estimate of a three factor model with a PROMAX rotation using maximum likelihood factor analysis was generated. The PROMAX rotation is an oblique rotation that will allow correlated factors.

The three factor model results with PROMAX rotation are:

```
Call:
factanal(factors = 3, covmat = cor.matrix, n.obs = 1442, rotation = "promax")

Uniquenesses:
 Armagnac  Calvados    Cognac    Kirsch      Marc Mirabelle       Rum
    0.759     0.792     0.739     0.134     0.005     0.005     0.933
  Whiskey  Liqueurs
    0.890     0.005

Loadings:
          Factor1 Factor2 Factor3
Armagnac  -0.347   0.202  -0.102
Calvados  -0.331   0.193   0.026
Cognac    -0.371   0.186  -0.146
Kirsch     0.961   0.068   0.035
Marc      -0.139   0.123   0.977
Mirabelle -0.186  -1.075  -0.101
Rum        0.123   0.287  -0.043
Whiskey   -0.248   0.084  -0.146
Liqueurs   1.047   0.175  -0.186

               Factor1 Factor2 Factor3
SS loadings      2.518   1.407   1.057
Proportion Var   0.280   0.156   0.117
Cumulative Var   0.280   0.436   0.554

Factor Correlations:
        Factor1 Factor2 Factor3
Factor1  1.0000 -0.0591  0.0147
Factor2 -0.0591  1.0000  0.4787
Factor3  0.0147  0.4787  1.0000

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 1820.72 on 12 degrees of freedom.
The p-value is 0
```

The Uniquenesses for Kirsch is 0.134, and 0.005 for Marc, Mirabelle,and Liqueurs. All other liquors have uniquenesses greater than 0.739. The proportion of variance accounted for the four factors is 55.4%. The chi-square statistic is 1820.72 with a p-value of 0. This indicates that the null hypothesis of a perfect model fit (that the three factors are sufficient) is rejected.

The results for the PROMAX three factor model are similar to the VARIMAX three factor model, although the loadings are represented by difference factors. The cumulative variances is within 3% for the PROMAX and VARIMAX rotation techniques. The uniquenesses and the chi-square statistic (1820.72 on 12 degrees of freedom) are the same for both models.

Table 2: Comparison of Cumulative Variance by Rotation Technique

	Cumulative Variance
PROMAX	0.554
VARIMAX	0.527

Table 3: Comparison of Uniquenesses by Rotation Technique

	PROMAX	VARIMAX
Armagnac	0.7592201	0.7592201
Calvados	0.7916741	0.7916741
Cognac	0.7386008	0.7386008
Kirsch	0.1344403	0.1344403
Marc	0.0050000	0.0050000
Mirabelle	0.0050000	0.0050000
Rum	0.9328050	0.9328050
Whiskey	0.8897497	0.8897497
Liqueurs	0.0050000	0.0050000

Section 4: Correlation Matrix Approximation

The factor loadings and specific variances (uniquenesses) can be used to approximate the correlation matrix. The factor loadings will be multiplied by the transpose of the matrix and then the specific variances will be added. This will be the approximated correlation matrix. The Mean Absolute Error (MAE) will be calculated by subtracting the correlation matrix from the approximated correlation matrix, and calculating the mean value of the absolute error. The MAE will be calculated for both the VARIMAX and PROMAX three factor models and compared. A smaller MAE value indicates less error in the model. The VARIMAX model has a smaller MAE compared to the PROMAX model.

Table 4: Comparison of MAE by Rotation Technique

	MAE
PROMAX	0.554
VARIMAX	0.527

Conclusion

There was no factor analysis model that explained the correlation matrix of liquor preferences of French consumers from a statistically significant assessment. A comparison of the factor rotation techniques, orthogonal (VARIMAX) and oblique (PROMAX), for a three factor model yielded essentially the same model. The factors were uncorrelated with the VARIMAX rotation and the PROMAX rotation allowed for the factors to be correlated. A smaller MAE value was observed with the VARIMAX model. The VARIMAX model better approximates the correlation matrix.

References

Everitt, B. (2010). Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Fox, J. and Weisberg, S. (2011). An R Companion to Applied Regression, Second Edition. Thousand Oaks, CA: Sage Publications, Inc.

Kabacoff, R. (2015). R in Action, Second Edition. Shelter Island, NY: Manning Publications Co.

Stoezel, J. (1960). A Factor Analysis of the Liquor Preferences of French Consumers. *Journal of Advertising Research, 1*, 7-11.

##Code

```
knitr::opts_chunk$set(echo = FALSE)
library(corrplot)
library(knitr)
#Load correlation matrix into R
cor.values <- c(1.000,0.210,0.370,-0.32,0.000,-0.31,-0.26,0.090,-0.38,
                0.210,1.000,0.090,-0.29,0.120,-0.30,-0.14,0.010,-0.39,
                0.370,0.090,1.000,-0.31,-0.04,-0.30,-0.11,0.120,-0.39,
                -0.32,-0.29,-0.31,1.00,-0.16,0.25,-0.13,-0.14,0.900,
                0.00,0.120,-0.04,-0.16,1.000,-0.20,-0.03,-0.08,-0.38,
                -0.31,-0.30,-0.30,0.25,-0.20,1.000,-0.24,-0.16,0.180,
                -0.26,-0.14,-0.11,-0.13,-0.03,-0.24,1.000,-0.20,0.040,
                0.090,0.010,0.120,-0.14,-0.08,-0.16,-0.20,1.000,-0.24,
                -0.38,-0.39,-0.39,0.900,-0.38,0.180,0.040,-0.24,1.000
);

# How do we put these correlation values into a correlation matrix?;
help(matrix)
cor.matrix <- matrix(cor.values,nrow=9,ncol=9,byrow=TRUE);

#Provide row and column names for matrix
liquor <- c('Armagnac', 'Calvados', 'Cognac', 'Kirsch', 'Marc', 'Mirabelle', 'Rum', 'Whiskey', 'Liqueurs
rownames(cor.matrix) <- liquor
colnames(cor.matrix) <- liquor

cor.table <- as.data.frame(cor.matrix)
kable(cor.table, caption = 'Correlation Table')

# Check that object is a matrix object;
print(paste('Is the cor.matrix a matrix: ', is.matrix(cor.matrix)))
print(paste('What is the type of the matrix: ', mode(cor.matrix)))

# Check that matrix is symmetric;
# This check helps check for typos;
print(paste('Is the matrix symmetric: ', isSymmetric(cor.matrix)))


par(oma=c(0,0,0,1))
corrplot(cor.matrix)
f.3 <- factanal(covmat=cor.matrix, n.obs=1442, factors=3, rotation='varimax');
print(f.3, digits = 3, cutoff=0.01)

liquor <- c('Armagnac', 'Calvados', 'Cognac', 'Kirsch', 'Marc', 'Mirabelle', 'Rum', 'Whiskey', 'Liqueurs
par(mfrow = c(1,3))
f3.load1.2 <- f.3$loadings[,1:2]
plot(f3.load1.2, type='n', xlim = c(-1.5, 1.5), ylim = c(-0.5, 1.0))
text(f3.load1.2, labels=liquor, cex=.7)
f3.load2.3 <- f.3$loadings[,2:3]
plot(f3.load2.3, type='n', xlim = c(-1.5, 1.5), ylim = c(-0.5, 1.0))
text(f3.load2.3, labels=liquor, cex=.7)
f3.load3.1 <- cbind(f.3$loadings[,3], f.3$loadings[,1])
plot(f3.load3.1, type='n', xlim = c(-1.5, 1.5), ylim = c(-0.5, 1.0), xlab='Factor3', ylab='Factor1')
text(f3.load3.1, labels=liquor, cex=.7)
```

```r
par(mfrow = c(1,1))
#Fit factor models for k=1 through 6.
#k=1
f.1 <- factanal(covmat=cor.matrix, n.obs=1442, factors=1, rotation='varimax');
print(f.1, digits = 3, cutoff=0.01)
#k=2
f.2 <- factanal(covmat=cor.matrix, n.obs=1442, factors=2, rotation='varimax');
print(f.2, digits = 3, cutoff=0.01)
#k=4
f.4 <- factanal(covmat=cor.matrix, n.obs=1442, factors=4, rotation='varimax');
print(f.4, digits = 3, cutoff=0.01)
#k=5
f.5 <- factanal(covmat=cor.matrix, n.obs=1442, factors=5, rotation='varimax');
print(f.5, digits = 3, cutoff=0.01)
#k=6
#f.6 <- factanal(covmat=cor.matrix, n.obs=1442, factors=6, rotation='varimax');
#print(f.6, digits = 3, cutoff=0.01)
g.3 <- factanal(covmat=cor.matrix, n.obs=1442, factors=3, rotation='promax');
print(g.3, digits = 3, cutoff=0.01)
promax.cumvar <- 0.554
varimax.cumvar <- 0.527
overview.cumvar.all <- rbind(promax.cumvar, varimax.cumvar)
colnames(overview.cumvar.all) <- c("Cumulative Variance")
rownames(overview.cumvar.all) <- c("PROMAX", "VARIMAX")
kable(overview.cumvar.all, caption = 'Comparison of Cumulative Variance by Rotation Technique')
promax.unique <- g.3$uniquenesses
varimax.unique <- f.3$uniquenesses
overview.unique.all <- cbind(promax.unique, varimax.unique)
colnames(overview.unique.all) <- c("PROMAX", "VARIMAX")
kable(overview.unique.all, caption = 'Comparison of Uniquenesses by Rotation Technique')
#VARIMAX MAE
gamma.f3 <- f.3$loadings;
approx.f3 <- gamma.f3%*%t(gamma.f3) + diag(f.3$uniqueness);
mae.f3 <- mean(abs(approx.f3-cor.matrix))

#PROMAX MAE
gamma.g3 <- g.3$loadings;
approx.g3 <- gamma.g3%*%t(gamma.g3) + diag(g.3$uniqueness);
mae.g3 <- mean(abs(approx.g3-cor.matrix))

#table of MAE
overview.mae <- rbind(mae.g3, mae.f3)
colnames(overview.cumvar.all) <- c("MAE")
rownames(overview.cumvar.all) <- c("PROMAX", "VARIMAX")
kable(overview.cumvar.all, caption = 'Comparison of MAE by Rotation Technique')
```