

Multivariate Analysis for the Behavioral Sciences,
Second Edition (Chapman and Hall/CRC, 2019)

Examples of Chapter 15:
Exploratory Factor Analysis

Kimmo Vehkalahti and Brian S. Everitt

12 December 2018

Contents

Examples	2
Table 15.1: Maximum Likelihood Two-Factor Solution for Correlations of Six School Subjects	2
Figure 15.1	5
Figure 15.2	7
Table 15.2	8
Table 15.3	9
Table 15.4: Estimated Parameters for the Three-Factor Model Fitted to the Crime Rate Data by Maximum Likelihood	10
Table 15.5	11
Figure 15.3	12

Examples

Table 15.1: Maximum Likelihood Two-Factor Solution for Correlations of Six School Subjects

```
R <- c( 1.00, 0.44, 0.41, 0.29, 0.33, 0.25,
        0.44, 1.00, 0.35, 0.35, 0.32, 0.33,
        0.41, 0.35, 1.00, 0.16, 0.19, 0.18,
        0.29, 0.35, 0.16, 1.00, 0.59, 0.47,
        0.33, 0.32, 0.19, 0.59, 1.00, 0.46,
        0.25, 0.33, 0.18, 0.47, 0.46, 1.00)

school_subjects <- c("French", "English", "History", "Arithmetic", "Algebra", "Geometry")
R <- matrix(R, ncol = 6, byrow = TRUE, dimnames = list(school_subjects, school_subjects))
R

##           French English History Arithmetic Algebra Geometry
## French      1.00    0.44    0.41      0.29    0.33    0.25
## English     0.44    1.00    0.35      0.35    0.32    0.33
## History     0.41    0.35    1.00      0.16    0.19    0.18
## Arithmetic  0.29    0.35    0.16      1.00    0.59    0.47
## Algebra     0.33    0.32    0.19      0.59    1.00    0.46
## Geometry    0.25    0.33    0.18      0.47    0.46    1.00

Rfa1 <- factanal(covmat = R, factors = 1, n.obs = 220, rotation = "none")
Rfa2 <- factanal(covmat = R, factors = 2, n.obs = 220, rotation = "none")
Rfa3 <- factanal(covmat = R, factors = 3, n.obs = 220, rotation = "none")
Rfa1

##
## Call:
## factanal(factors = 1, covmat = R, n.obs = 220, rotation = "none")
##
## Uniquenesses:
##      French      English      History Arithmetic      Algebra      Geometry
##      0.750      0.710      0.878      0.473      0.468      0.621
##
## Loadings:
##           Factor1
## French      0.500
## English     0.539
## History     0.349
## Arithmetic  0.726
## Algebra     0.729
## Geometry    0.615
##
##           Factor1
## SS loadings      2.10
## Proportion Var   0.35
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 51.6 on 9 degrees of freedom.
## The p-value is 5.37e-08
```

Rfa2

```
##
## Call:
## factanal(factors = 2, covmat = R, n.obs = 220, rotation = "none")
##
## Uniquenesses:
##      French      English      History Arithmetic      Algebra      Geometry
##      0.508      0.595      0.644      0.377      0.440      0.628
##
## Loadings:
##      Factor1 Factor2
## French      0.558  0.425
## English      0.569  0.286
## History      0.392  0.450
## Arithmetic   0.738 -0.279
## Algebra      0.718 -0.209
## Geometry     0.595 -0.133
##
##      Factor1 Factor2
## SS loadings      2.204  0.603
## Proportion Var    0.367  0.101
## Cumulative Var    0.367  0.468
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 2.18 on 4 degrees of freedom.
## The p-value is 0.703
```

Rfa3

```
##
## Call:
## factanal(factors = 3, covmat = R, n.obs = 220, rotation = "none")
##
## Uniquenesses:
##      French      English      History Arithmetic      Algebra      Geometry
##      0.448      0.497      0.679      0.411      0.376      0.611
##
## Loadings:
##      Factor1 Factor2 Factor3
## French      0.582  0.432 -0.162
## English      0.594  0.306  0.238
## History      0.398  0.402
## Arithmetic   0.713 -0.281
## Algebra      0.731 -0.269 -0.130
## Geometry     0.591 -0.158  0.123
##
##      Factor1 Factor2 Factor3
## SS loadings      2.241  0.617  0.118
## Proportion Var    0.373  0.103  0.020
## Cumulative Var    0.373  0.476  0.496
##
## The degrees of freedom for the model is 0 and the fit was 0.001
```

```

#install.packages("psych")
library(psych)

RfaNone <- fa(R, nfactors = 2, fm = "ml", rotate = "none")
print.psych(RfaNone)

## Factor Analysis using method = ml
## Call: fa(r = R, nfactors = 2, rotate = "none", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           ML1    ML2    h2    u2 com
## French      0.56   0.42  0.49  0.51 1.9
## English     0.57   0.29  0.41  0.59 1.5
## History     0.39   0.45  0.36  0.64 2.0
## Arithmetic  0.74  -0.28  0.62  0.38 1.3
## Algebra     0.72  -0.21  0.56  0.44 1.2
## Geometry    0.59  -0.13  0.37  0.63 1.1
##
##           ML1    ML2
## SS loadings      2.20 0.60
## Proportion Var    0.37 0.10
## Cumulative Var    0.37 0.47
## Proportion Explained 0.79 0.21
## Cumulative Proportion 0.79 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 15 and the objective function was 1.43
## The degrees of freedom for the model are 4 and the objective function was 0.01
##
## The root mean square of the residuals (RMSR) is 0.01
## The df corrected root mean square of the residuals is 0.03
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
##           ML1    ML2
## Correlation of (regression) scores with factors 0.91 0.73
## Multiple R square of scores with factors         0.82 0.53
## Minimum correlation of possible factor scores    0.64 0.07

Correlation matrix of the six school subjects implied by the fitted two-factor model:

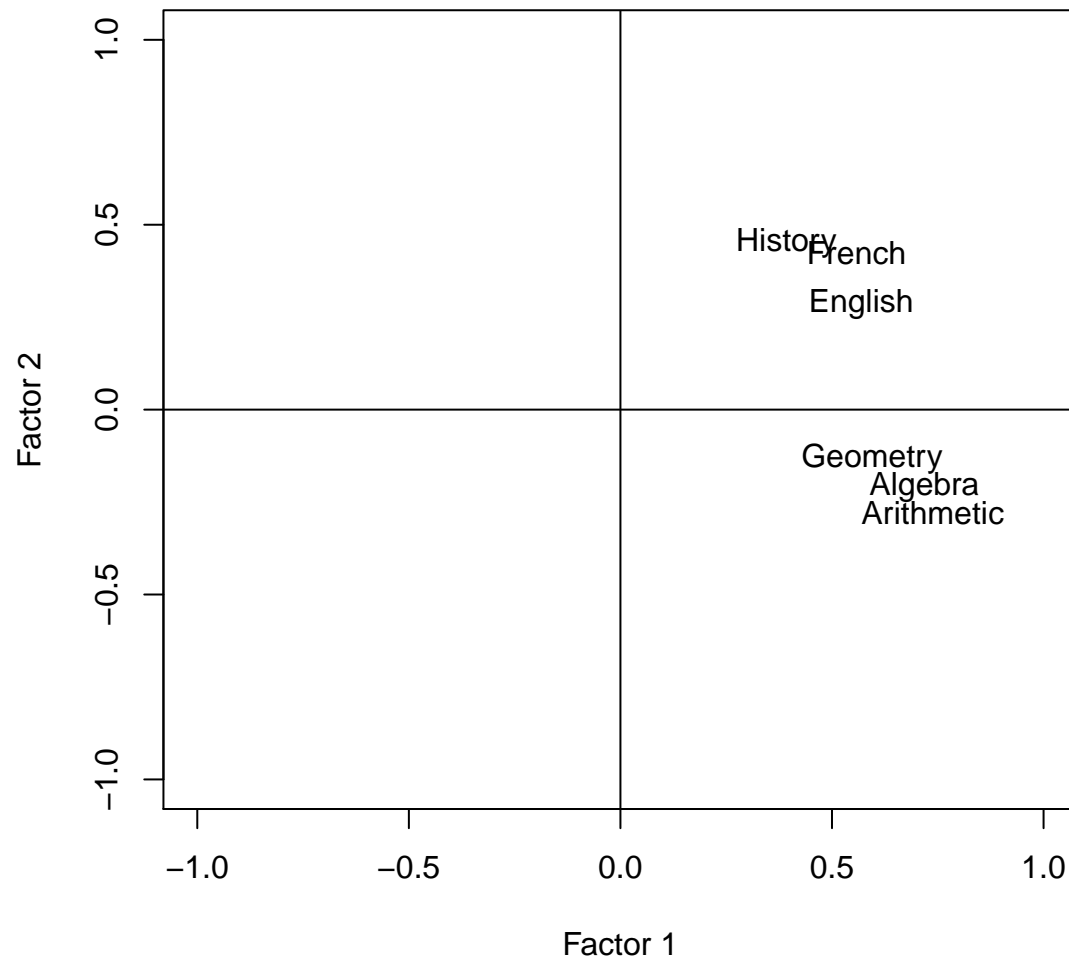
lambdaHat <- as.matrix(RfaNone$loadings)
Rhat <- lambdaHat %*% t(lambdaHat) + RfaNone$residual
Rhat

##           French English History Arithmetic Algebra Geometry
## French      1.00    0.44    0.41    0.29    0.33    0.25
## English     0.44    1.00    0.35    0.35    0.32    0.33
## History     0.41    0.35    1.00    0.16    0.19    0.18
## Arithmetic  0.29    0.35    0.16    1.00    0.59    0.47
## Algebra     0.33    0.32    0.19    0.59    1.00    0.46
## Geometry    0.25    0.33    0.18    0.47    0.46    1.00

```

Figure 15.1

```
plot(RfaNone$loadings, type = "n",  
     xlim = c(-1, +1), ylim = c(-1, +1),  
     xlab = "Factor 1", ylab = "Factor 2")  
text(RfaNone$loadings, labels = school_subjects)  
abline(h = 0, v = 0)
```



```
# same figure with ggplot (overkill, perhaps) - just testing:
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
## The following objects are masked from 'package:psych':
##
##      %+%, alpha
RfaNoneLoadings <- as.matrix(RfaNone$loadings)
class(RfaNoneLoadings) <- "matrix"
RfaNoneLoadings <- as.data.frame(RfaNoneLoadings)
RfaNoneLoadings$Subject <- row.names(RfaNoneLoadings)

p1 <- ggplot(RfaNoneLoadings, aes(x = ML1, y = ML2)) + geom_point()
p2 <- p1 + geom_text(aes(x = ML1+0.03, y = ML2+0.00,
                        label = RfaNoneLoadings$Subject), size=4, hjust=0)
p3 <- p2 + theme_bw() + theme(panel.grid.major = element_blank(),
                             panel.grid.minor = element_blank())
p4 <- p3 + scale_x_continuous(limits = c(-1, +1), name = "Factor 1")
p5 <- p4 + scale_y_continuous(limits = c(-1, +1), name = "Factor 2")
p6 <- p5 + geom_vline(xintercept=0) + geom_hline(yintercept=0)
p6
```

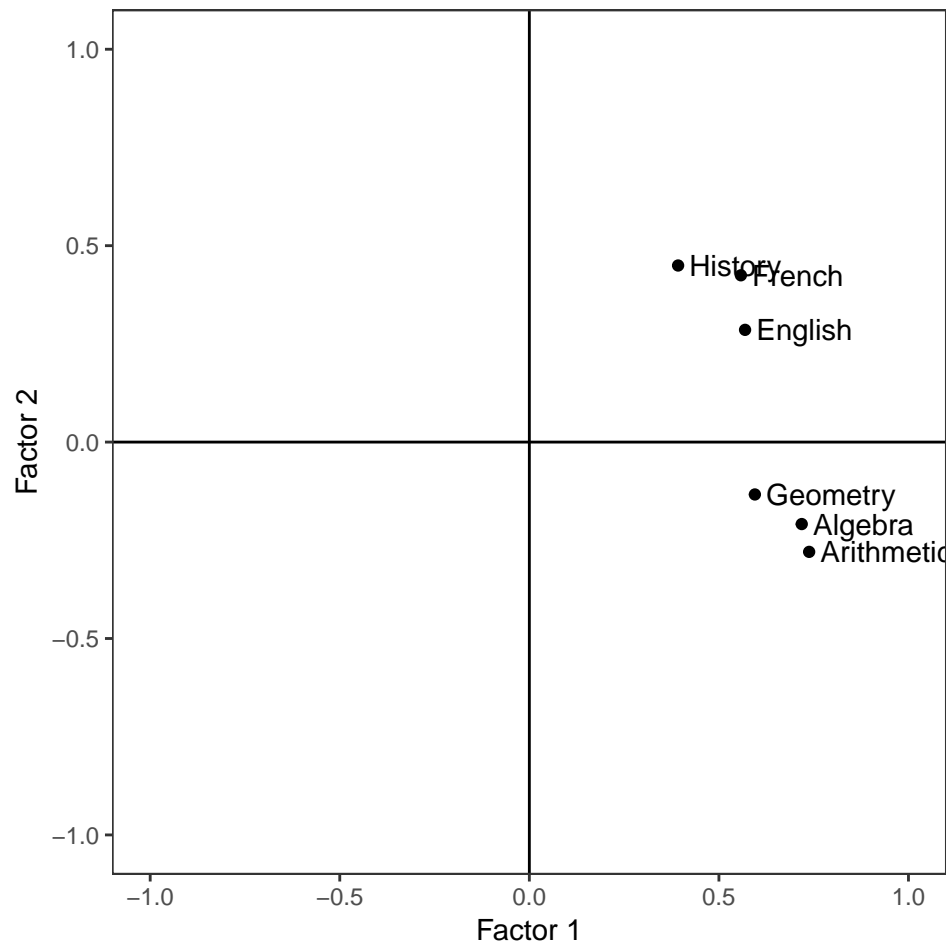


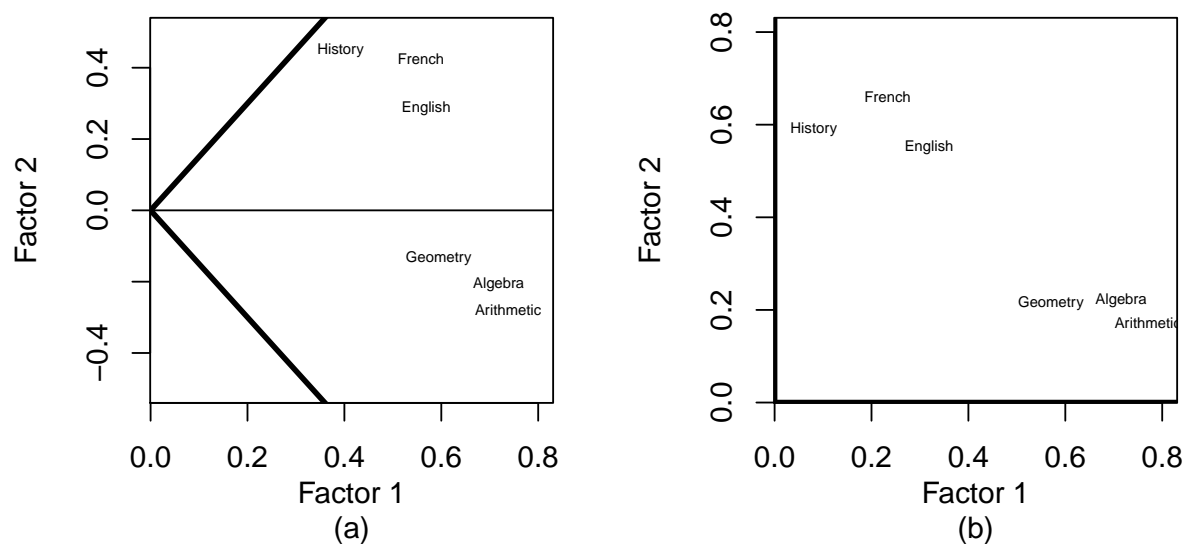
Figure 15.2

```
par(mfrow=c(1,2))

plot(RfaNone$loadings, type = "n", xlim = c(0.03, 0.8), ylim = c(-0.5, +0.5),
     xlab = "Factor 1\n(a)", ylab = "Factor 2")
text(RfaNone$loadings, labels = school_subjects, cex=0.5)
abline(h = 0)
abline(0, +1.5, lwd = 3)
abline(0, -1.5, lwd = 3)

RfaVarimax <- fa(R, nfactors = 2, fm = "ml", rotate = "varimax") # See the note below!

plot(RfaVarimax$loadings, type = "n", xlim = c(0.03, 0.8), ylim = c(0.03, 0.8),
     xlab = "Factor 1\n(b)", ylab = "Factor 2")
text(RfaVarimax$loadings, labels = school_subjects, cex=0.5)
abline(h = 0, v = 0, lwd = 3)
```



Above, the rotation was achieved using the varimax, because, as far as we know, no function for graphical rotation exists in R. For an inspiration to program such a function, you should take a good look at **Survo**: *an environment for creative processing of text and numerical data*, created by professor **Seppo Mustonen** (see www.survo.fi/presentation/history.html), cf. the example of *classical music composers* in **Chapter 14**.

Interactive graphical rotation has been available in various generations of Survo since the 1970s:

- A short demo on YouTube shows how an **oblique graphical rotation** is applied in **SURVO MM**: <https://youtu.be/-yZRUR6gpAg> (see www.survo.fi/demos/#ex38 for more details and links).
- A historical example of using the graphical rotation in **SURVO 76** appears on pp. 29-30 of the report: Mustonen, S. (1977). SURVO 76, a statistical data processing system. Research Report No. 6. Dept of Statistics, University of Helsinki. www.survo.fi/publications/Research_Report_6_Mustonen_1977.pdf

Publications related to Survo since 1963: www.survo.fi/publications

Table 15.2

```
RfaVarimax <- fa(R, nfactors = 2, fm = "ml", rotate = "varimax")
print.psych(RfaVarimax, sort = TRUE)
```

```
## Factor Analysis using method = ml
## Call: fa(r = R, nfactors = 2, rotate = "varimax", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           item ML1 ML2 h2 u2 com
## Arithmetic    4 0.77 0.17 0.62 0.38 1.1
## Algebra        5 0.72 0.22 0.56 0.44 1.2
## Geometry       6 0.57 0.22 0.37 0.63 1.3
## French         1 0.23 0.66 0.49 0.51 1.2
## History        3 0.08 0.59 0.36 0.64 1.0
## English        2 0.32 0.55 0.41 0.59 1.6
##
##           ML1 ML2
## SS loadings      1.59 1.21
## Proportion Var    0.27 0.20
## Cumulative Var    0.27 0.47
## Proportion Explained 0.57 0.43
## Cumulative Proportion 0.57 1.00
##
## Mean item complexity = 1.2
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 15 and the objective function was 1.43
## The degrees of freedom for the model are 4 and the objective function was 0.01
##
## The root mean square of the residuals (RMSR) is 0.01
## The df corrected root mean square of the residuals is 0.03
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##           ML1 ML2
## Correlation of (regression) scores with factors 0.86 0.79
## Multiple R square of scores with factors        0.73 0.62
## Minimum correlation of possible factor scores    0.47 0.24
```


Table 15.3

```
#install.packages("GPArotation")
library(GPArotation)
RfaOblimin <- fa(R, nfactors = 2, fm = "ml", rotate = "oblimin")
print.psych(RfaOblimin, sort = TRUE)

## Factor Analysis using method = ml
## Call: fa(r = R, nfactors = 2, rotate = "oblimin", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           item  ML1  ML2  h2  u2 com
## Arithmetic    4  0.81 -0.04 0.62 0.38 1.0
## Algebra        5  0.73  0.03 0.56 0.44 1.0
## Geometry       6  0.57  0.07 0.37 0.63 1.0
## French         1  0.04  0.68 0.49 0.51 1.0
## History        3 -0.11  0.65 0.36 0.64 1.1
## English        2  0.17  0.53 0.41 0.59 1.2
##
##           ML1  ML2
## SS loadings    1.60 1.21
## Proportion Var    0.27 0.20
## Cumulative Var    0.27 0.47
## Proportion Explained 0.57 0.43
## Cumulative Proportion 0.57 1.00
##
## With factor correlations of
##           ML1  ML2
## ML1 1.00 0.52
## ML2 0.52 1.00
##
## Mean item complexity = 1.1
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 15 and the objective function was 1.43
## The degrees of freedom for the model are 4 and the objective function was 0.01
##
## The root mean square of the residuals (RMSR) is 0.01
## The df corrected root mean square of the residuals is 0.03
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
##           ML1  ML2
## Correlation of (regression) scores with factors 0.89 0.84
## Multiple R square of scores with factors        0.79 0.70
## Minimum correlation of possible factor scores    0.58 0.40
```

Table 15.4: Estimated Parameters for the Three-Factor Model Fitted to the Crime Rate Data by Maximum Likelihood

see Chapter 13, where this data set was used for the first time:

```
crime <- read.table("data/crime.txt", sep = '\t')
```

```
crime_fa1 <- factanal(crime, factors = 1, rotation = "none")
```

```
crime_fa2 <- factanal(crime, factors = 2, rotation = "none")
```

```
crime_fa3 <- factanal(crime, factors = 3, rotation = "none")
```

```
crime_fa3None <- fa(crime, nfactors = 3, fm = "ml", rotate = "none")
```

```
print.psych(crime_fa3None, digits = 3)
```

```
## Factor Analysis using method = ml
```

```
## Call: fa(r = crime, nfactors = 3, rotate = "none", fm = "ml")
```

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

```
##           ML1      ML2      ML3      h2      u2      com
```

```
## Murder    0.654    0.727   -0.114  0.970  0.0301  2.03
```

```
## Rape      0.611    0.307    0.415  0.640  0.3601  2.30
```

```
## Robbery   0.828    0.344   -0.103  0.815  0.1854  1.37
```

```
## Assault   0.697    0.479    0.181  0.747  0.2526  1.93
```

```
## Burglary  0.625    0.330    0.604  0.864  0.1364  2.51
```

```
## Theft     0.422    0.231    0.700  0.721  0.2787  1.90
```

```
## Vehicle   0.992   -0.106   -0.010  0.995  0.0050  1.02
```

```
##
```

```
##           ML1      ML2      ML3
```

```
## SS loadings          3.523  1.145  1.083
```

```
## Proportion Var       0.503  0.164  0.155
```

```
## Cumulative Var       0.503  0.667  0.822
```

```
## Proportion Explained 0.613  0.199  0.188
```

```
## Cumulative Proportion 0.613  0.812  1.000
```

```
##
```

```
## Mean item complexity = 1.9
```

```
## Test of the hypothesis that 3 factors are sufficient.
```

```
##
```

```
## The degrees of freedom for the null model are 21 and the objective function was 5.818 with Chi Sq
```

```
## The degrees of freedom for the model are 3 and the objective function was 0.109
```

```
##
```

```
## The root mean square of the residuals (RMSR) is 0.015
```

```
## The df corrected root mean square of the residuals is 0.04
```

```
##
```

```
## The harmonic number of observations is 51 with the empirical chi square 0.5 with prob < 0.919
```

```
## The total number of observations was 51 with Likelihood Chi Square = 4.898 with prob < 0.179
```

```
##
```

```
## Tucker Lewis Index of factoring reliability = 0.9446
```

```
## RMSEA index = 0.1281 and the 90 % confidence intervals are 0 0.2852
```

```
## BIC = -6.897
```

```
## Fit based upon off diagonal values = 0.999
```

```
## Measures of factor score adequacy
```

```
##
```

```
ML1      ML2      ML3
```

```
## Correlation of (regression) scores with factors 0.998 0.979 0.921
```

```
## Multiple R square of scores with factors 0.995 0.958 0.847
```

```
## Minimum correlation of possible factor scores 0.991 0.915 0.695
```

Table 15.5

```
crime_fa3varimax <- fa(crime, nfactors = 3, fm = "ml", rotate = "varimax")
print.psych(crime_fa3varimax, digits = 3, sort = TRUE)
```

```
## Factor Analysis using method = ml
## Call: fa(r = crime, nfactors = 3, rotate = "varimax", fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      item   ML3   ML2   ML1   h2   u2   com
## Theft      6 0.831 0.132 0.120 0.721 0.2787 1.09
## Burglary    5 0.828 0.330 0.263 0.864 0.1364 1.53
## Rape        2 0.645 0.369 0.297 0.640 0.3601 2.06
## Murder      1 0.259 0.922 0.229 0.970 0.0301 1.29
## Robbery     3 0.243 0.664 0.561 0.815 0.1854 2.24
## Assault     4 0.492 0.629 0.331 0.747 0.2526 2.46
## Vehicle     7 0.285 0.317 0.902 0.995 0.0050 1.46
##
##
##      ML3   ML2   ML1
## SS loadings      2.24 2.049 1.463
## Proportion Var    0.32 0.293 0.209
## Cumulative Var    0.32 0.613 0.822
## Proportion Explained 0.39 0.356 0.254
## Cumulative Proportion 0.39 0.746 1.000
##
## Mean item complexity = 1.7
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are 21 and the objective function was 5.818 with Chi Sq
## The degrees of freedom for the model are 3 and the objective function was 0.109
##
## The root mean square of the residuals (RMSR) is 0.015
## The df corrected root mean square of the residuals is 0.04
##
## The harmonic number of observations is 51 with the empirical chi square 0.5 with prob < 0.919
## The total number of observations was 51 with Likelihood Chi Square = 4.898 with prob < 0.179
##
## Tucker Lewis Index of factoring reliability = 0.9446
## RMSEA index = 0.1281 and the 90 % confidence intervals are 0 0.2852
## BIC = -6.897
## Fit based upon off diagonal values = 0.999
## Measures of factor score adequacy
##
##      ML3   ML2   ML1
## Correlation of (regression) scores with factors 0.931 0.975 0.991
## Multiple R square of scores with factors      0.868 0.951 0.982
## Minimum correlation of possible factor scores  0.735 0.903 0.963
```

Figure 15.3

```
crime_fa3 <- factanal(crime, factors = 3, method = "mle", rotation = "varimax",
  scores = "regression")
pairs(crime_fa3$scores,
  panel = function(x,y) text(x, y, labels = row.names(crime), cex=0.8))
```

