Assignment #7

Factor Analysis

Name: Young, Brent

Predict 410 Section #: 57

Quarter: Summer 2017

Introduction

Context

The data for this assignment consists of a 9 by 9 correlation matrix input from the Stoetzel article called 'A Factor Analysis of Liquor Preference' by Stoetzel (Journal of Marketing Research). The correlation matrix represents the correlations between rankings of the nine liquors. The sample population consists of a total of 1,442 completed interviews out of a total of 2,014 French men and women. The participants were asked to rank the 9 liquors from best to worst.

Objectives/Purpose

The overall purpose/objective of assignment 7 is to apply factor analysis using Stoetzel's liquor preference application. This example is also useful in the context of marketing segmentation. First, we will prep the data by loading the correlation matrix into R. Second, we will estimate a three factor model with a VARIMAX rotation using maximum likelihood factor analysis and compare it to Stoetzel's estimated three factor model. We will then discuss whether the results are the same numerically (i.e., same factor loadings), qualitatively (i.e. same factor interpretations), and whether if three factors were the correct number of factors to describe this correlation matrix (in addition to stating the null and alternative hypothesis). Third, we will find the correct number of factors to describe this correlation matrix by fitting factor models for k=1 through 6. Fourth, we will fit a three factor model with a PROMAX rotation using maximum likelihood factor analysis. We will then discuss whether or not this model has better interpretability than the three factor model with the VARIMAX rotation, while also addressing the correct number of factors to describe this correlation matrix. A discussion of this topic will be interwoven with statistical inference for this maximum likelihood factor analysis and how the factor rotation affects the statistical inference for the number of factors. Lastly, we will use the factor loadings and the specific (or unique) variances to approximate the correlation matrix. We will then measure the fit of these approximations using the Mean Absolute Error of the residual matrix so that we can see which factor model better approximates the correlation matrix.

Section 1: Data Prep

Figure 1: Correlation Matrix

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.00	0.21	0.37	-0.32	0.00	-0.31	-0.26	0.09	-0.38
[2,]	0.21	1.00	0.09	-0.29	0.12	-0.30	-0.14	0.01	-0.39
[3,]	0.37	0.09	1.00	-0.31	-0.04	-0.30	-0.11	0.12	-0.39
[4,]	-0.32	-0.29	-0.31	1.00	-0.16	0.25	-0.13	-0.14	0.90
[5,]	0.00	0.12	-0.04	-0.16	1.00	-0.20	-0.03	-0.08	-0.38
[6,]	-0.31	-0.30	-0.30	0.25	-0.20	1.00	-0.24	-0.16	0.18
[7,]	-0.26	-0.14	-0.11	-0.13	-0.03	-0.24	1.00	-0.20	0.04
[8,]	0.09	0.01	0.12	-0.14	-0.08	-0.16	-0.20	1.00	-0.24
[9,]	-0.38	-0.39	-0.39	0.90	-0.38	0.18	0.04	-0.24	1.00

Figure 2: Correlations Between Rankings of the Nine Liquors

	Arm	Cal	Cog	Kir	Mar	Mir	Rum	Whi	Liq
Arm	1.00	0.21	0.37	-0.32	0.00	-0.31	-0.26	0.09	-0.38
Cal	0.21	1.00	0.09	-0.29	0.12	-0.30	-0.14	0.01	-0.39
Cog	0.37	0.09	1.00	-0.31	-0.04	-0.30	-0.11	0.12	-0.39
Kir	-0.32	-0.29	-0.31	1.00	-0.16	0.25	-0.13	-0.14	0.90
Mar	0.00	0.12	-0.04	-0.16	1.00	-0.20	-0.03	-0.08	-0.38
Mir	-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
Whi	0.09	0.01	0.12	-0.14	-0.08	-0.16	-0.20	1.00	-0.24
Liq	-0.38	-0.39	-0.39	0.90	-0.38	0.18	0.04	-0.24	1.00

Observations: Figure 1 shows a 9 by 9 correlation matrix based on the 1,442 observations, which shows the correlations between rankings of the nine liquors. Figure 2 displays the same correlation matrix, but represented as a data frame. This view is similar to what was seen in the paper, which shows the names of the French wines on the columns and on the rows. The data checks reveal that the object is a matrix object and that the matrix is symmetric. For example, the correlation matrix shows 1.00 along the diagonal and a lower/upper triangular pattern, which is an indicator that the matrix is symmetric.

Section 2: Three Factor Model with Varimax Rotation Using Maximum Likelihood Factor Analysis

Figure 3: Three Factor Model

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

Uniquenesses:
[1] 0.759 0.792 0.739 0.134 0.005 0.005 0.933 0.890 0.005

Loadings:
      Factor1 Factor2 Factor3   Nms
 [1,] -0.450           0.193   "Arm"
 [2,] -0.411   0.100   0.172   "Cal"
 [3,] -0.473           0.183   "Cog"
 [4,]  0.921  -0.121           "Kir"
 [5,]          0.996           "Mar"
 [6,]  0.293  -0.169  -0.938   "Mir"
 [7,]                  0.256   "Rum"
 [8,] -0.305                   "Whi"
 [9,]  0.923  -0.344   0.158   "Liq"

               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
```

2a & 2b Observations (see Appendix for graph): Figure 3 shows a three factor model with varimax rotation using maximum likelihood factor analysis. The results show that for Factor 1, there are two variables with very high positive loadings (Kir and Liq) and four medium loadings on the negative side (Arm, Cal, Cog and Whi). As a result, we can interpret factor 1 as a contrast between the two positive loadings and the four medium loadings on the negative side. We also see this in the graph in the appendix (strong vs. sweet from left to right). This is similar to what we saw in the Stoetzel paper in regards to sweet and strong liquors. Factor 2 shows a very high positive loading (0.996, Mar), while the other loadings are comprised of both low negative/positive loadings. We also see this in the graph in the appendix (inexpensive vs. expensive from top to bottom). This is similar to what we saw in the Stoetzel paper in regards to expensive vs. non-expensive. Factor 3 shows that the highest loading is -0.938 (Mir), while the other loadings are comprised of low/mediocre positive loadings. This is similar to what we saw in the Stoetzel paper in regards to local vs. national liquors. As a result, in comparison to these results to the Stoetzel paper, the factor loadings are different numerically. This is expected because the correlation matrix only had two-thirds of the matrix filled out and it used the Thurstone's centroid method. However, I expect that we would be able to reproduce a factor analysis and get the same loadings if the same method was used (e.g., varimax roation using maximum likelihood factor). On the other hand, the results produced the same qualitative results as Stoetzel. For example, the distinction between sweetness, price, and regional popularity were all common factors that were identified in the Stoetzel paper. As a result, I would expect to be able to reproduce a factor analysis and get the same factor interpretation.

2c Observations: The SS loadings for each factor is computed by squaring and adding the various loadings. Proportion Var is the Eigen values and explains the variance for factor 1, 2, and 3. The Cumulative Var is the cumulative variance for each factor. As a result, given this context, 52.7% of the variation is explained by the 3 factors. The null hypothesis for the chi-square test statistic is that 3 factors are sufficient and the alternative hypothesis is that more factors are needed. Since the p-value is 0 (small), we reject the null hypothesis that 3 factors are sufficient and conclude that more factors are needed. Therefore, the three factors that are used to describe this correlation matrix using statistical inference for the maximum likelihood factor analysis is incorrect.

Section 3: Five Factor Model with Varimax Rotation Using Maximum Likelihood Factor Analysis

Figure 4: Five Factor Model with Varimax Rotation

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

Uniquenesses:
[1] 0.660 0.664 0.603 0.005 0.005 0.005 0.768 0.005 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5 Nms
 [1,] -0.164   0.555                          "Arm"
 [2,] -0.141   0.552   0.101                  "Cal"
 [3,] -0.150   0.605                          "Cog"
 [4,]  0.937  -0.279                   0.193 "Kir"
 [5,] -0.142           0.986                  "Mar"
 [6,]         -0.497  -0.207  -0.228   0.806 "Mir"
 [7,]                         -0.171  -0.426 "Rum"
 [8,] -0.110                   0.985          "Whi"
 [9,]  0.834  -0.446  -0.273  -0.132          "Liq"

               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
```

Observations (See Appendix for Factor Models 1 to 6): Figure 4 shows a five factor model with varimax rotation using maximum likelihood factor analysis. 69.8% of the variation is explained by the 5 factors. The null hypothesis for the chi-square test statistic is that 5 factors are sufficient and the alternative hypothesis is that more factors are needed. Since the p-value is 9.34e-149 (large), we fail to reject the null hypothesis that 5 factors are sufficient and conclude that the correct number of factors to describe this correlation matrix is 5. As a result, the five factor model with varimax rotation using maximum likelihood factor analysis represents the correct number of factors. For example, Factor 1 shows a contrast between two positive loadings (Kir and Liq) and five loadings on the negative side, which represent sweet liqueurs. Factor 2 shows a contrast between three positive loadings (Arm, Cal, Cog) and 3 medium loadings on the negative side, which represent strong liqueurs. Factor 3 shows a very high positive loading (Mar), while the other loads are comprised of both low negative/positive loadings, which represents popular local liqueurs. Factor 4, shows a very high positive loading (Whi) and negative low loadings, which represent popular national liqueurs. Factor 5, shows a high positive loading (Mir) and very low negative loading (Rum), which represents price (e.g., inexpensive and expensive).

Section 4: Three Factor Model with Promax Rotation Using Maximum Likelihood Factor Analysis

Figure 5: Three Factor Model with Promax Rotation

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

Uniquenesses:
[1] 0.759 0.792 0.739 0.134 0.005 0.005 0.933 0.890 0.005

Loadings:
      Factor1 Factor2 Factor3 Nms
 [1,] -0.347   0.202  -0.102 "Arm"
 [2,] -0.331   0.193         "Cal"
 [3,] -0.371   0.186  -0.146 "Cog"
 [4,]  0.961                 "Kir"
 [5,] -0.139   0.123   0.977 "Mar"
 [6,] -0.186  -1.075  -0.101 "Mir"
 [7,]  0.123   0.287         "Rum"
 [8,] -0.248          -0.146 "Whi"
 [9,]  1.047   0.175  -0.186 "Liq"

               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
```

4a Observations: Figure 5 shows a three factor model with promax rotation using maximum likelihood factor analysis. The results show that for Factor 1, there are two variables with very high positive loadings (Kir and Liq) and four medium loadings on the negative side (Arm, Cal, Cog and Whi). As a result, we can interpret factor 1 as a contrast between the two positive loadings and the four medium loadings on the negative side. This is similar to what we saw in the three and five factor varimax rotation and is consistent with the Stoetzel paper in regards to sweet and strong liquors. Factor 2 shows a very high negative loading (-1.075, Mir) and medium loading (0.287, Rum), while the rest are low/medium positive loadings. As a result, this shows the contrasts between inexpensive vs. expensive liqueurs (Rum being least expensive and Mir being the most expensive). It’s interesting to note that interpreting factor 2 using the promax rotation was easier to interpret than using the three factor model with varimax rotation and was on par with the five factor rotation with varimax rotation. These results were also more similar to the results we saw in the Stoetzel paper. Factor 3 shows that the highest loading is 0.977 (Mar), while the other loadings are comprised of low negative loadings. This shows the contrasts between local vs. national liquors that we saw in the five factor model with varimax rotation and is consistent with what we saw in the Stoetzel paper. As a result, interpreting Factor 3 using the promax rotation was easier to interpret than using the varimax rotation. Overall, this model has better interpretability than the three factor model with varimax rotation. Not only was it easier to interpret, but in a lot of ways it mimicked the results that we saw in the five factor varimax rotation and is more in line with the results we saw in the Stoetzel paper.

4b Observations: Figure 5 shows a five factor model with promax rotation using maximum likelihood factor analysis. 55.4% of the variation is explained by the 3 factors. The null hypothesis for the chi-square test statistic is that 3 factors are sufficient and the alternative hypothesis is that more factors are needed. Since the p-value is 0 (small), we reject the null hypothesis that 3 factors are sufficient and conclude that more factors are needed. Therefore, the three factors that are used to describe this correlation matrix using statistical inference for the maximum likelihood factor analysis is incorrect. *4b observations continued on next page….*

Figure 6: Five Factor Model with Promax Rotation

```
Call:
factanal(factors = 5, covmat = cor.matrix, n.obs = 1442, rotation = "promax")
Uniquenesses:
[1] 0.660 0.664 0.603 0.005 0.005 0.005 0.768 0.005 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5 Nms
 [1,]          0.570                  -0.105 "Arm"
 [2,]          0.579  -0.116                 "Cal"
 [3,]          0.633          -0.128         "Cog"
 [4,]  0.991                                 "Kir"
 [5,]                          0.998         "Mar"
 [6,]         -0.448  -0.108           0.930 "Mir"
 [7,] -0.216  -0.165  -0.183          -0.328 "Rum"
 [8,]                  1.011                 "Whi"
 [9,]  0.748  -0.284          -0.182  -0.146 "Liq"

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      1.603   1.377   1.091   1.085   1.036
Proportion Var   0.178   0.153   0.121   0.121   0.115
Cumulative Var   0.178   0.331   0.452   0.573   0.688

Factor Correlations:
        Factor1 Factor2 Factor3 Factor4 Factor5
Factor1  1.0000 -0.1060  0.0704 -0.2474  0.4426
Factor2 -0.1060  1.0000  0.1335  0.0266 -0.2564
Factor3  0.0704  0.1335  1.0000  0.0190 -0.0738
Factor4 -0.2474  0.0266  0.0190  1.0000 -0.1817
Factor5  0.4426 -0.2564 -0.0738 -0.1817  1.0000

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
```

4b Observations (continued): No, the factor rotation should not affect the statistical interference for the number of factors. For instance, I re-ran a five factor model with promax rotation and I obtained the same results as the five factor model with varimax rotation. For example, Factor 1 shows a contrast between two positive loadings (Kir and Liq) and one loading on the negative side, which represent sweet liqueurs. Factor 2 shows a contrast between three positive loadings (Arm, Cal, Cog) and 3 medium loadings on the negative side, which represent strong liqueurs. Factor 3 shows a very high positive loading (Whi), while the other loads are comprised of low negative loadings, which represents popular national liqueurs. Factor 4, shows a very high positive loading (Mar) and negative low loadings, which represent popular local liqueurs. Factor 5, shows a high positive loading (Mir) and very low negative loading (Rum), which represents price (e.g., inexpensive and expensive).

Section 5: Mean Absolute Error of the Residual Matrix Comparison

Three Factor Model with Varimax Rotation

```
> dim(lamda.f1)
[1] 9 3
> dim(approx.f1)
[1] 9 9
> mae.f1
[1] 0.04857863
```

Three Factor Model with Promax Rotation

```
> dim(lamda.g1)
[1] 9 3
> dim(approx.g1)
[1] 9 9
> mae.g1
[1] 0.1102698
```

Five Factor Model with Varimax Rotation

```
> mae.f1
[1] 0.02004459
```

Five Factor Model with Promax Rotation

```
> mae.g1
[1] 0.1238722
```

Observations: The MAE identified for the three factor model with varimax rotation is 0.05. The MAE is calculated by taking the estimated correlation matrix vs. the actual correlation matrix and then comparing them by taking the absolute deviations, then adding them up, and then taking the mean (mean absolute error). In comparison, the MAE identified for the three factor model with promax rotation is 0.11, which is higher than the varimax rotation. Therefore, the three factor model with varimax rotation better approximates the correlation matrix than the three factor model with promax rotation (although if you add-in the five factor model with varimax rotation, this model approximates the correlation matrix the best out of all the models). However, as seen in the previous sections, the three factor model with promax rotation has better interpretability than the three factor model with varimax rotation. Similar results are also seen when the MAE is computed for a five factor model with varimax rotation versus promax rotation (except that the MAE for the varimax rotation is lower than all the other factor models). In conclusion, we are making a tradeoff between accuracy and interpretability. For instance, if we choose the three factor model with promax rotation, we have less accuracy, but better interpretability.

Conclusion/Summary

In section 1, we prepped the data by loading a 9 by 9 correlation matrix based on the 1,442 observations, which shows the correlations between rankings of the nine liquors.

In section 2, we estimated a three factor model with a VARIMAX rotation using maximum likelihood factor analysis and compared it to Stoetzel's estimated three factor model. The results showed that in comparison to these results to the Stoetzel paper, the factor loadings are different numerically. This is expected because the correlation matrix only had two-thirds of the matrix filled out and it used the Thurstone's centroid method. However, I expect that we would be able to reproduce a factor analysis and get the same loadings if the same method was used (e.g., varimax rotation using maximum likelihood factor). On the other hand, the results produced the same qualitative results as Stoetzel. For example, the distinction between sweetness, price, and regional popularity were all common factors that were identified in the Stoetzel paper. As a result, I would expect to be able to reproduce a factor analysis and get the same factor interpretation. Additionally, we saw in the three factor model that 52.7% of the variation is explained by the 3 factors. Additionally, since the p-value was 0 (small), we rejected the null hypothesis that 3 factors are sufficient and concluded that more factors were needed. Therefore, the three factors that were used to describe this correlation matrix using statistical inference for the maximum likelihood factor analysis is incorrect.

In section 3, we fitted factor models for k=1 through 6 and determined that the five factor model with varimax rotation using maximum likelihood factor analysis represents the correct number of factors. For instance, 69.8% of the variation is explained by the 5 factors and since the p-value is 9.34e-149 (large), we fail to reject the null hypothesis that 5 factors are sufficient and conclude that the correct number of factors to describe this correlation matrix is 5. Additionally, the results for Factor 1 showed a contrast between two positive loadings (Kir and Liq) and five loadings on the negative side, which represent sweet liqueurs. Factor 2 showed a contrast between three positive loadings (Arm, Cal, Cog) and 3 medium loadings on the negative side, which represent strong liqueurs. Factor 3 showed a very high positive loading (Mar), while the other loads are comprised of both low negative/positive loadings, which represents popular local liqueurs. Factor 4, showed a very high positive loading (Whi) and negative low loadings, which represent popular national liqueurs. Factor 5, showed a high positive loading (Mir) and very low negative loading (Rum), which represents price (e.g., inexpensive and expensive).

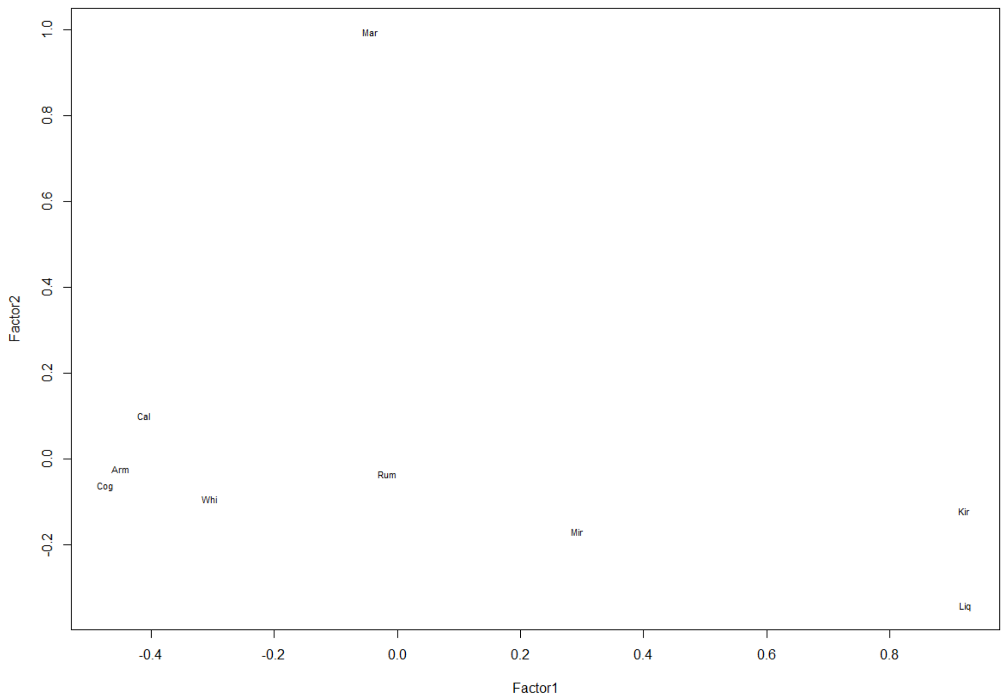
In section 4, we fit a three factor model with a PROMAX rotation using maximum likelihood factor analysis. After analyzing the results, we concluded that this model has better interpretability than the three factor model with varimax rotation. Additionally, we were also able to determine that since p-value was 0 (small), we rejected the null hypothesis that 3 factors are sufficient and concluded that more factors were needed. Therefore, the three factors that are used to describe this correlation matrix using statistical inference for the maximum likelihood factor analysis is incorrect. Furthermore, we also concluded that the factor rotation should not affect the statistical interference for the number of factors.

For instance, I re-ran a five factor model with promax rotation and obtained the same results as the five factor model with varimax rotation.

Lastly, we used the factor loadings and the specific (or unique) variances to approximate the correlation matrix. We then measured the fit of these approximations using the Mean Absolute Error of the residual matrix so that we could see which factor model better approximates the correlation matrix. The results showed that the three factor model with varimax rotation better approximates the correlation matrix than the three factor model with promax rotation (although if you add-in the five factor model with varimax rotation, this model approximates the correlation matrix the best out of all the models). However, as seen in the previous sections, the three factor model with promax rotation has better interpretability than the three factor model with varimax rotation. Similar results are also seen when the MAE is computed for a five factor model with varimax rotation versus promax rotation (except that the MAE for the varimax rotation is lower than all the other factor models). In conclusion, we are making a tradeoff between accuracy and interpretability. For instance, if we choose the three factor model with promax rotation, we have less accuracy, but better interpretability.

Appendix

Section 2a & 2b



Section 3

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

Uniquenesses:
[1] 0.855 0.848 0.848 0.187 0.859 0.966 0.999 0.943 0.005

Loadings:
      Factor1
 [1,] -0.381
 [2,] -0.390
 [3,] -0.390
 [4,]  0.902
 [5,] -0.376
 [6,]  0.184
 [7,]
 [8,] -0.238
 [9,]  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
```

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

Uniquenesses:
[1] 0.831 0.848 0.807 0.148 0.005 0.946 0.998 0.910 0.005

Loadings:
      Factor1 Factor2
 [1,] -0.404
 [2,] -0.343   0.185
 [3,] -0.437
 [4,]  0.866  -0.321
 [5,]  0.178   0.981
 [6,]         -0.217
 [7,]
 [8,] -0.298
 [9,]  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

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

Uniquenesses:
[1] 0.759 0.792 0.739 0.134 0.005 0.005 0.933 0.890 0.005

Loadings:
      Factor1 Factor2 Factor3
 [1,] -0.450           0.193
 [2,] -0.411   0.100   0.172
 [3,] -0.473           0.183
 [4,]  0.921  -0.121
 [5,]          0.996
 [6,]  0.293  -0.169  -0.938
 [7,]                  0.256
 [8,] -0.305
 [9,]  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


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

Uniquenesses:
[1] 0.696 0.727 0.639 0.005 0.005 0.005 0.781 0.765 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4
 [1,] -0.189   0.515                 
 [2,] -0.172   0.477   0.121         
 [3,] -0.176   0.573                 
 [4,]  0.943  -0.252           0.206 
 [5,] -0.151           0.985         
 [6,]         -0.536  -0.227   0.808 
 [7,]         -0.161          -0.426 
 [8,]          0.469                 
 [9,]  0.848  -0.438  -0.280         

               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

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

Uniquenesses:
[1] 0.660 0.664 0.603 0.005 0.005 0.005 0.768 0.005 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5
 [1,] -0.164   0.555                         
 [2,] -0.141   0.552   0.101                 
 [3,] -0.150   0.605                         
 [4,]  0.937  -0.279                   0.193 
 [5,] -0.142           0.986                 
 [6,]         -0.497  -0.207  -0.228   0.806 
 [7,]                         -0.171  -0.426 
 [8,] -0.110                   0.985         
 [9,]  0.834  -0.446  -0.273  -0.132         

               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
```

```
> factanal(factors = 6, covmat = cor.matrix, n.obs = 1442, rotation = "varimax")
Error in factanal(factors = 6, covmat = cor.matrix, n.obs = 1442, rotation = "vari
max") :
  6 factors are too many for 9 variables
```

Section 5

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

Uniquenesses:
[1] 0.660 0.664 0.603 0.005 0.005 0.005 0.768 0.005 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5
 [1,] -0.164   0.555
 [2,] -0.141   0.552   0.101
 [3,] -0.150   0.605
 [4,]  0.937  -0.279                   0.193
 [5,] -0.142           0.986
 [6,]         -0.497  -0.207  -0.228   0.806
 [7,]                         -0.171  -0.426
 [8,] -0.110                   0.985
 [9,]  0.834  -0.446  -0.273  -0.132

               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
> mae.f1
[1] 0.02004459


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

Uniquenesses:
[1] 0.660 0.664 0.603 0.005 0.005 0.005 0.768 0.005 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5
 [1,]          0.570                  -0.105
 [2,]          0.579  -0.116
 [3,]          0.633          -0.128
 [4,]  0.991
 [5,]                          0.998
 [6,]         -0.448  -0.108           0.930
 [7,] -0.216  -0.165  -0.183          -0.328
 [8,]                  1.011
 [9,]  0.748  -0.284          -0.182  -0.146

               Factor1 Factor2 Factor3 Factor4 Factor5
```

```
SS loadings      1.603   1.377   1.091   1.085   1.036
Proportion Var   0.178   0.153   0.121   0.121   0.115
Cumulative Var   0.178   0.331   0.452   0.573   0.688

Factor Correlations:
        Factor1 Factor2 Factor3 Factor4 Factor5
Factor1  1.0000 -0.1060  0.0704 -0.2474  0.4426
Factor2 -0.1060  1.0000  0.1335  0.0266 -0.2564
Factor3  0.0704  0.1335  1.0000  0.0190 -0.0738
Factor4 -0.2474  0.0266  0.0190  1.0000 -0.1817
Factor5  0.4426 -0.2564 -0.0738 -0.1817  1.0000

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

> lamda.g1 <- g.1$loadings;
> mae.g1
[1] 0.1238722
```