Solution for the assignment of the tenth class

Kovacs Marton

9/29/2021

Importing data

```
processed <- read_tsv("data/boldog_processed.tsv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   neme = col_character(),
##   isk = col_character()
## )
## i Use `spec()` for the full column specifications.
```

Data exploration

```
skimr::skim(processed) %>%
  kable()
```

skim_type	skim_variable	n_missing	complete_rate	character.min	character.max	character.empty	character.n_unique	character.whitespace	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100	numeric.hist
character	neme	0	1.000	2	5	0	2	0	NA	NA	NA	NA	NA	NA	NA	NA
character	isk	0	1.000	7	11	0	4	0	NA	NA	NA	NA	NA	NA	NA	NA
numeric	index	0	1.000	NA	NA	NA	NA	NA	972.34	486.43	27	566	998.00	1377	1661.75	<U+2585><U+2585><U+2587><U+2587><U+2587>
numeric	eletkora	0	1.000	NA	NA	NA	NA	NA	52.52	20.07	25	38.00	45.00	62.50	81.00	<U+2581><U+2585><U+2587><U+2585><U+2582>
numeric	gyermekek	0	1.000	NA	NA	NA	NA	NA	1.75	1.02	0	1.00	2.00	2.00	9	<U+2585><U+2587><U+2581><U+2581><U+2581>
numeric	anyagi	0	1.000	NA	NA	NA	NA	NA	3.10	0.56	1	3.00	3.00	3.00	5	<U+2581><U+2581><U+2587><U+2581><U+2581>
numeric	pic_elmeny_percent	0	1.000	NA	NA	NA	NA	NA	58.46	20.04	18	40.00	60.00	80.00	90	<U+2582><U+2585><U+2587><U+2587><U+2587>
numeric	testi_fi	3	0.994	NA	NA	NA	NA	NA	4.16	0.95	2	4.00	4.00	5.00	6	<U+2581><U+2583><U+2587><U+2581><U+2581>
numeric	alt_lelki	4	0.992	NA	NA	NA	NA	NA	4.16	1.00	2	4.00	4.00	5.00	6	<U+2582><U+2583><U+2587><U+2581><U+2581>
numeric	alt_eg_all	7	0.986	NA	NA	NA	NA	NA	4.12	1.04	1	3.00	4.00	5.00	6	<U+2582><U+2583><U+2587><U+2581><U+2581>
numeric	fizero	6	0.988	NA	NA	NA	NA	NA	4.15	1.04	1	3.00	4.00	5.00	6	<U+2582><U+2585><U+2587><U+2581><U+2581>
numeric	aicocska	0	1.000	NA	NA	NA	NA	NA	2.57	0.94	1	2.00	2.00	3.00	7	<U+2587><U+2585><U+2581><U+2581><U+2581>
numeric	aigodalo	0	1.000	NA	NA	NA	NA	NA	2.77	1.04	1	2.00	2.50	4.00	6	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	ideges	0	1.000	NA	NA	NA	NA	NA	2.49	1.03	1	1.00	2.00	3.00	6	<U+2587><U+2582><U+2581><U+2581><U+2581>
numeric	feszult	0	1.000	NA	NA	NA	NA	NA	2.61	1.03	1	1.00	2.00	3.00	6	<U+2587><U+2583><U+2581><U+2581><U+2581>
numeric	nyugtala	0	1.000	NA	NA	NA	NA	NA	2.40	1.04	1	1.00	2.00	3.00	6	<U+2587><U+2582><U+2581><U+2581><U+2581>

skim_type	skim_variable	n_missing	complete_rate	character.min	character.max	character.empty	character.n_unique	character.whitespace	numeric.mean	numeric.sd	numeric.p0	numeric.p25	numeric.p50	numeric.p75	numeric.p100	numeric.hist
numeric	diener1	0	1.000	NA	NA	NA	NA	NA	5.514000	1.328659	1.00	5.0000	6.0000	6.000000	7	<U+2581><U+2581><U
numeric	diener2	0	1.000	NA	NA	NA	NA	NA	5.228000	1.321934	1.00	4.0000	6.0000	6.000000	7	<U+2581><U+2581><U
numeric	diener3	0	1.000	NA	NA	NA	NA	NA	5.574000	1.231078	1.00	5.0000	6.0000	6.000000	7	<U+2581><U+2581><U
numeric	diener4	0	1.000	NA	NA	NA	NA	NA	5.418000	1.261132	1.00	5.0000	6.0000	6.000000	7	<U+2581><U+2581><U
numeric	diener5	0	1.000	NA	NA	NA	NA	NA	5.756000	1.072734	1.00	5.0000	6.0000	7.000000	7	<U+2581><U+2581><U
numeric	diener6	0	1.000	NA	NA	NA	NA	NA	5.768000	1.124592	1.00	5.0000	6.0000	7.000000	7	<U+2581><U+2581><U
numeric	diener7	0	1.000	NA	NA	NA	NA	NA	5.620000	1.373816	1.00	5.0000	6.0000	7.000000	7	<U+2581><U+2581><U
numeric	diener8	0	1.000	NA	NA	NA	NA	NA	5.532000	1.214877	1.00	5.0000	6.0000	6.000000	7	<U+2581><U+2581><U
numeric	jollet	0	1.000	NA	NA	NA	NA	NA	4.488667	1.061084	1.00	4.0000	4.6667	5.333300	6	<U+2581><U+2582><U
numeric	savor	0	1.000	NA	NA	NA	NA	NA	4.514667	1.013386	1.00	4.0000	4.6667	5.000000	6	<U+2581><U+2582><U
numeric	a_vhat	0	1.000	NA	NA	NA	NA	NA	4.595600	0.939396	1.04	4.0000	4.8000	5.200000	6	<U+2581><U+2581><U
numeric	onreg	0	1.000	NA	NA	NA	NA	NA	4.217333	1.221736	1.08	3.3333	4.3333	5.083325	6	<U+2582><U+2583><U
numeric	rezil	0	1.000	NA	NA	NA	NA	NA	4.044667	1.021045	1.03	3.3333	4.0000	4.666700	6	<U+2581><U+2583><U
numeric	m_flow	0	1.000	NA	NA	NA	NA	NA	4.773999	0.789097	1.98	4.3333	5.0000	5.333300	6	<U+2581><U+2581><U
numeric	g_jerz	0	1.000	NA	NA	NA	NA	NA	4.152800	1.031377	1.05	3.4000	4.2000	5.000000	6	<U+2581><U+2583><U
numeric	g_jpszi	0	1.000	NA	NA	NA	NA	NA	4.282000	0.989284	1.03	3.7500	4.5000	5.000000	6	<U+2581><U+2582><U
numeric	g_jszoc	0	1.000	NA	NA	NA	NA	NA	3.949500	1.078920	1.09	3.0000	4.0000	5.000000	6	<U+2582><U+2585><U
numeric	g_jspir	0	1.000	NA	NA	NA	NA	NA	4.245500	1.200316	1.07	3.5000	4.2500	5.250000	6	<U+2582><U+2583><U
numeric	g_jerzpsz	0	1.000	NA	NA	NA	NA	NA	4.217400	0.972023	1.14	3.6187	4.3375	4.925000	6	<U+2581><U+2582><U
numeric	g_jspszoc	0	1.000	NA	NA	NA	NA	NA	4.097500	1.013821	1.01	3.3750	4.1250	5.000000	6	<U+2581><U+2583><U
numeric	pik_mm	0	1.000	NA	NA	NA	NA	NA	3.027333	0.462674	1.87	2.6667	3.1667	3.500000	4	<U+2581><U+2582><U
numeric	pik_av	0	1.000	NA	NA	NA	NA	NA	3.175000	0.623218	1.89	2.7500	3.2500	3.750000	4	<U+2581><U+2581><U
numeric	pik_onr	0	1.000	NA	NA	NA	NA	NA	3.064000	0.791213	1.31	2.6667	3.3333	3.666700	4	<U+2581><U+2582><U
numeric	pik_rez	0	1.000	NA	NA	NA	NA	NA	3.150667	0.742419	1.24	2.6667	3.3333	3.666700	4	<U+2581><U+2582><U
numeric	p_poz_erz	0	1.000	NA	NA	NA	NA	NA	21.7425	5.027513	[illegible]	20.0000	23.0000	26.000000	30	<U+2581><U+2581><U
numeric	p_elmely	0	1.000	NA	NA	NA	NA	NA	22.3604	4.558580	[illegible]	20.0000	23.0000	26.000000	30	<U+2581><U+2582><U
numeric	p_poz_kapc	0	1.000	NA	NA	NA	NA	NA	22.4425	5.567100	[illegible]	19.0000	24.0000	26.000000	30	<U+2581><U+2581><U
numeric	p_ert_cel	0	1.000	NA	NA	NA	NA	NA	23.3744	4.679093	[illegible]	21.0000	24.0000	27.000000	30	<U+2581><U+2581><U
numeric	p_telj	0	1.000	NA	NA	NA	NA	NA	22.7164	4.342009	[illegible]	21.0000	24.0000	26.000000	30	<U+2581><U+2581><U
numeric	p_boldog	0	1.000	NA	NA	NA	NA	NA	7.4040	2.012170	[illegible]	7.0000	8.0000	9.000000	10	<U+2581><U+2581><U
numeric	p_egeszs	0	1.000	NA	NA	NA	NA	NA	22.2605	5.224804	[illegible]	19.0000	23.0000	26.000000	30	<U+2581><U+2581><U
numeric	p_neg_erz	0	1.000	NA	NA	NA	NA	NA	9.7660	5.708977	[illegible]	5.0000	9.0000	13.000000	30	<U+2587><U+2587><U
numeric	p_magany	0	1.000	NA	NA	NA	NA	NA	3.1620	3.056540	[illegible]	0.0000	2.0000	5.000000	10	<U+2587><U+2582><U
numeric	perma	0	1.000	NA	NA	NA	NA	NA	169.373	31.026722	[illegible]	155.000	170.500	191.000	[illegible]	<U+2581><U+2581><U

1. Run PCA with varimax rotation

Select the variables.

```
vars <- c("p_elmeny_percent", "testi_fi", "alt_lelki", "alt_eg_all", "fizero", "arcocska", "aggodalo",

task_data <-
  processed %>%
  select(all_of(vars))
```

Create a correlation matrix for the scree plot.

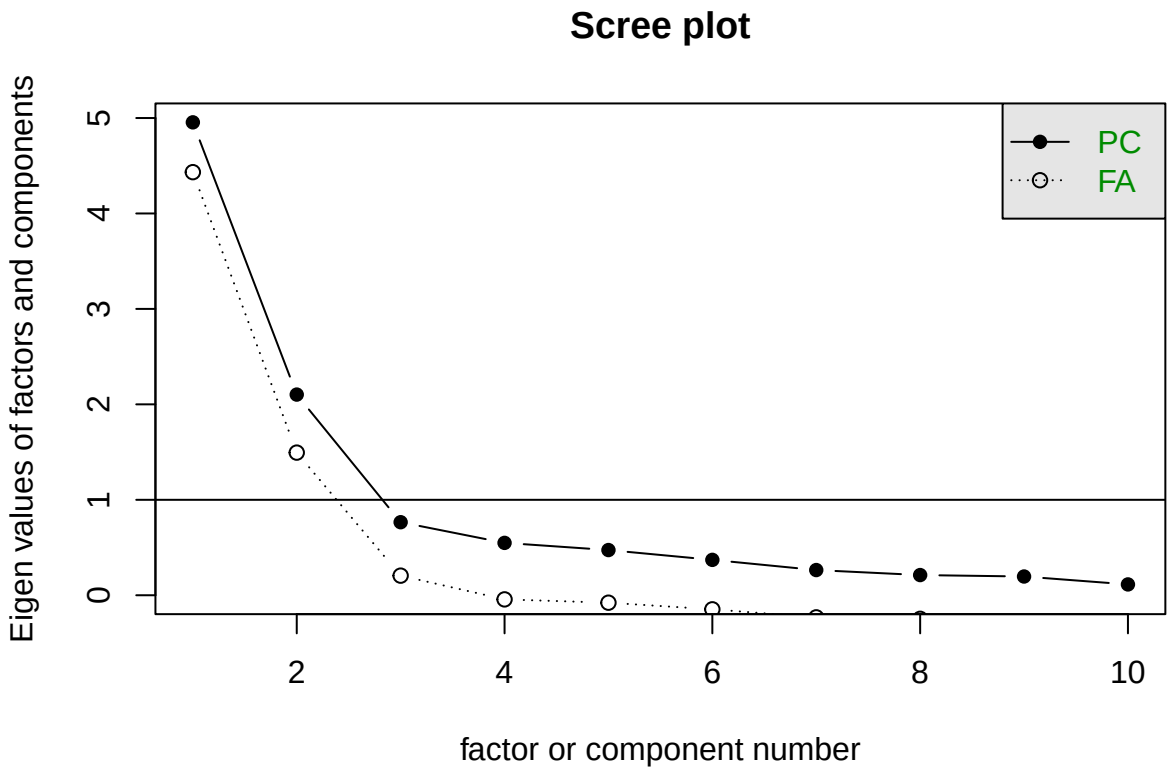
```
cor_matrix <- cor(task_data, use = "pairwise.complete.obs")

cor_matrix
```

```
##                  p_elmeny_percent   testi_fi  alt_lelki alt_eg_all     fizero
## p_elmeny_percent        1.0000000  0.2412126  0.4139176  0.2649702  0.2604031
## testi_fi                0.2412126  1.0000000  0.4718061  0.7848280  0.7589951
## alt_lelki               0.4139176  0.4718061  1.0000000  0.5174615  0.4637163
## alt_eg_all              0.2649702  0.7848280  0.5174615  1.0000000  0.7344209
## fizero                  0.2604031  0.7589951  0.4637163  0.7344209  1.0000000
## arcocska               -0.5278257 -0.3947918 -0.6073819 -0.3944909 -0.3986080
## aggodalo               -0.4301349 -0.1528917 -0.3950597 -0.2013013 -0.1606268
## ideges                 -0.4902466 -0.2066668 -0.4111001 -0.2436228 -0.2153273
## feszult                -0.4854919 -0.2088797 -0.4516150 -0.2478686 -0.2163127
## nyugtala               -0.3861882 -0.2093938 -0.3545616 -0.2728469 -0.2426588
##                    arcocska   aggodalo     ideges    feszult   nyugtala
## p_elmeny_percent -0.5278257 -0.4301349 -0.4902466 -0.4854919 -0.3861882
## testi_fi         -0.3947918 -0.1528917 -0.2066668 -0.2088797 -0.2093938
## alt_lelki        -0.6073819 -0.3950597 -0.4111001 -0.4516150 -0.3545616
## alt_eg_all       -0.3944909 -0.2013013 -0.2436228 -0.2478686 -0.2728469
## fizero           -0.3986080 -0.1606268 -0.2153273 -0.2163127 -0.2426588
## arcocska          1.0000000  0.4880817  0.5256719  0.5324749  0.4390953
## aggodalo          0.4880817  1.0000000  0.8027936  0.8253730  0.5731352
## ideges            0.5256719  0.8027936  1.0000000  0.8830490  0.6425057
## feszult           0.5324749  0.8253730  0.8830490  1.0000000  0.6318551
## nyugtala          0.4390953  0.5731352  0.6425057  0.6318551  1.0000000
```

Make the scree plot.

```
scree(cor_matrix)
```



We have two PCs with bigger than 1 Eigenvalues.

Run the PCA.

```
pca <- psych::principal(task_data, rotate = "varimax", nfactors = 2, scores = TRUE)

pca
```

```
## Principal Components Analysis
## Call: psych::principal(r = task_data, nfactors = 2, rotate = "varimax",
##     scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                    RC1   RC2   h2   u2 com
## p_elmeny_percent -0.60  0.26 0.43 0.57 1.4
## testi_fi         -0.09  0.90 0.82 0.18 1.0
## alt_lelki        -0.46  0.59 0.57 0.43 1.9
## alt_eg_all       -0.14  0.89 0.81 0.19 1.1
## fizero           -0.11  0.88 0.79 0.21 1.0
## arcocska          0.62 -0.45 0.58 0.42 1.8
## aggodalo          0.88 -0.04 0.78 0.22 1.0
## ideges            0.91 -0.10 0.85 0.15 1.0
## feszult           0.92 -0.10 0.86 0.14 1.0
## nyugtala          0.74 -0.15 0.57 0.43 1.1
##
##                        RC1  RC2
## SS loadings           4.01 3.05
## Proportion Var        0.40 0.30
## Cumulative Var        0.40 0.71
## Proportion Explained  0.57 0.43
## Cumulative Proportion 0.57 1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  169.04  with prob <  6.3e-23
##
## Fit based upon off diagonal values = 0.98
```

Based on the results the two PCs explain 71% of the variance. Not bad!

2. Run EFA on the same variables

```
efa <- fa(task_data, nfactors = 2, rotate = "varimax", fm = "pf")
```

```
## factor method not specified correctly, minimum residual (unweighted least squares  used
```

```
efa
```

```
## Factor Analysis using method =  minres
```

```
## Call: fa(r = task_data, nfactors = 2, rotate = "varimax", fm = "pf")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                     MR1   MR2   h2   u2 com
## p_elmeny_percent -0.53  0.25 0.34 0.66 1.4
## testi_fi         -0.10  0.88 0.78 0.22 1.0
## alt_lelki        -0.44  0.53 0.47 0.53 1.9
## alt_eg_all       -0.15  0.85 0.75 0.25 1.1
## fizero           -0.12  0.83 0.70 0.30 1.0
## arcocska          0.57 -0.42 0.50 0.50 1.8
## aggodalo          0.85 -0.06 0.73 0.27 1.0
## ideges            0.91 -0.11 0.85 0.15 1.0
## feszult           0.93 -0.11 0.87 0.13 1.0
## nyugtala          0.66 -0.18 0.47 0.53 1.1
##
##                        MR1  MR2
## SS loadings           3.71 2.76
## Proportion Var        0.37 0.28
## Cumulative Var        0.37 0.65
## Proportion Explained  0.57 0.43
## Cumulative Proportion 0.57 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  45  and the objective function was  6.96 with Chi Squa
## The degrees of freedom for the model are 26  and the objective function was  0.28
##
## The root mean square of the residuals (RMSR) is  0.04
## The df corrected root mean square of the residuals is  0.05
##
## The harmonic number of observations is  496 with the empirical chi square  58.94  with prob <  0.0003
## The total number of observations was  500  with Likelihood Chi Square =  139.67  with prob <  1.6e-17
##
## Tucker Lewis Index of factoring reliability =  0.942
## RMSEA index =  0.093  and the 90 % confidence intervals are  0.079 0.109
## BIC =  -21.91
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  MR1  MR2
## Correlation of (regression) scores with factors   0.97 0.95
## Multiple R square of scores with factors          0.94 0.90
## Minimum correlation of possible factor scores     0.88 0.80
```

With varimax rotation looking for 2 factors, and with PAF estimation (due to the non normal distributions of the variables), I do not find any variables with lower communality than 0.25 (see h2).

3. Check out KMO and Bartletts test results!

```
KMO(cor_matrix)
```

```
## Kaiser-Meyer-Olkin factor adequacy
```

```
## Call: KMO(r = cor_matrix)
## Overall MSA =  0.88
## MSA for each item =
## p_elmeny_percent         testi_fi        alt_lelki       alt_eg_all
##             0.94             0.81             0.91             0.83
##           fizero         arcocska         aggodalo           ideges
##             0.86             0.91             0.91             0.85
##          feszult         nyugtala
##             0.84             0.96
```

The overall KMO score and the individual items scores are all good.

```
cortest.bartlett(cor_matrix, n = 500, diag = FALSE)
```

```
## $chisq
## [1] 3442.281
##
## $p.value
## [1] 0
##
## $df
## [1] 45
```

The p value is so small that the test rounds it to 0. We can run an EFA the variables are good for factor analysis.

4. Is the model bad based on significance?

```
efa
```

```
## Factor Analysis using method =  minres
## Call: fa(r = task_data, nfactors = 2, rotate = "varimax", fm = "pf")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                     MR1   MR2   h2   u2 com
## p_elmeny_percent -0.53  0.25 0.34 0.66 1.4
## testi_fi         -0.10  0.88 0.78 0.22 1.0
## alt_lelki        -0.44  0.53 0.47 0.53 1.9
## alt_eg_all       -0.15  0.85 0.75 0.25 1.1
## fizero           -0.12  0.83 0.70 0.30 1.0
## arcocska          0.57 -0.42 0.50 0.50 1.8
## aggodalo          0.85 -0.06 0.73 0.27 1.0
## ideges            0.91 -0.11 0.85 0.15 1.0
## feszult           0.93 -0.11 0.87 0.13 1.0
## nyugtala          0.66 -0.18 0.47 0.53 1.1
##
##                        MR1  MR2
## SS loadings           3.71 2.76
## Proportion Var        0.37 0.28
## Cumulative Var        0.37 0.65
## Proportion Explained  0.57 0.43
```

```
## Cumulative Proportion 0.57 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  45  and the objective function was  6.96 with Chi Squa
## The degrees of freedom for the model are 26  and the objective function was  0.28
##
## The root mean square of the residuals (RMSR) is  0.04
## The df corrected root mean square of the residuals is  0.05
##
## The harmonic number of observations is  496 with the empirical chi square  58.94  with prob <  0.0003
## The total number of observations was  500  with Likelihood Chi Square =  139.67  with prob <  1.6e-17
##
## Tucker Lewis Index of factoring reliability =  0.942
## RMSEA index =  0.093  and the 90 % confidence intervals are  0.079 0.109
## BIC =  -21.91
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  MR1  MR2
## Correlation of (regression) scores with factors   0.97 0.95
## Multiple R square of scores with factors          0.94 0.90
## Minimum correlation of possible factor scores     0.88 0.80
```

The RMSEA is 0.093 CI90[0.079, 0109].

The total number of observations was 500 with Likelihood Chi Square = 139.67 with prob < 1.6e-17

5. Use a promax rotation

```
efa_promax <- fa(task_data, nfactors = 2, rotate = "promax", fm = "pf")
```

```
## factor method not specified correctly, minimum residual (unweighted least squares  used

## Loading required namespace: GPArotation
```

```
efa_promax
```

```
## Factor Analysis using method =  minres
## Call: fa(r = task_data, nfactors = 2, rotate = "promax", fm = "pf")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                     MR1   MR2   h2   u2 com
## p_elmeny_percent -0.51  0.13 0.34 0.66 1.1
## testi_fi          0.13  0.94 0.78 0.22 1.0
## alt_lelki        -0.34  0.46 0.47 0.53 1.9
## alt_eg_all        0.07  0.90 0.75 0.25 1.0
## fizero            0.09  0.88 0.70 0.30 1.0
## arcocska          0.51 -0.30 0.50 0.50 1.6
## aggodalo          0.92  0.17 0.73 0.27 1.1
## ideges            0.98  0.14 0.85 0.15 1.0
```

```
## feszult           0.99  0.14 0.87 0.13 1.0
## nyugtala          0.68 -0.01 0.47 0.53 1.0
##
##                        MR1  MR2
## SS loadings           3.78 2.69
## Proportion Var        0.38 0.27
## Cumulative Var        0.38 0.65
## Proportion Explained  0.58 0.42
## Cumulative Proportion 0.58 1.00
##
##  With factor correlations of
##       MR1   MR2
## MR1  1.00 -0.47
## MR2 -0.47  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  45  and the objective function was  6.96 with Chi Squa
## The degrees of freedom for the model are 26  and the objective function was  0.28
##
## The root mean square of the residuals (RMSR) is  0.04
## The df corrected root mean square of the residuals is  0.05
##
## The harmonic number of observations is  496 with the empirical chi square  58.94  with prob <  0.0003
## The total number of observations was  500  with Likelihood Chi Square =  139.67  with prob <  1.6e-17
##
## Tucker Lewis Index of factoring reliability =  0.942
## RMSEA index =  0.093  and the 90 % confidence intervals are  0.079 0.109
## BIC =  -21.91
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  MR1  MR2
## Correlation of (regression) scores with factors   0.97 0.95
## Multiple R square of scores with factors          0.95 0.91
## Minimum correlation of possible factor scores     0.89 0.82
```

p_elmeny_percent, arcocska, aggodalo, ideges, feszult, nyugtala loads for the 1st factor, but with p_elmeny_percent loading negatively. alt_lelki and arcocska has a high complexity.

Here it is possible for the items to correlate with each other.