Principal Components and Factor Analysis

Bud Talbot

18 May 2017

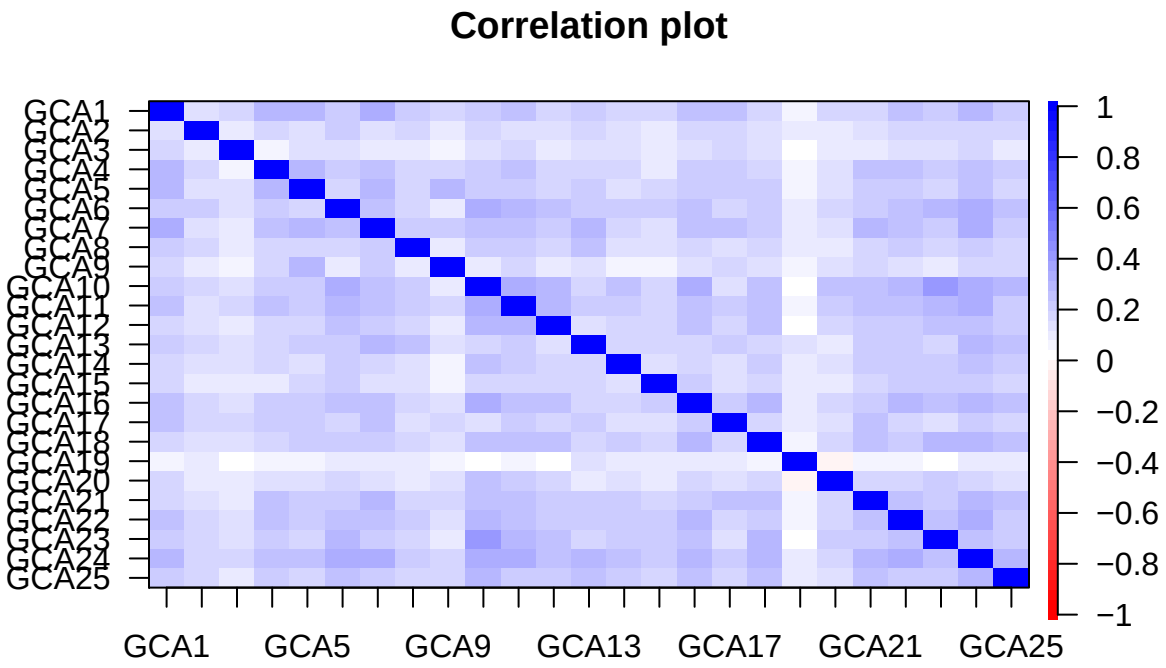
Initial Principal Components and Factor Analysis of GCA data.

The X1516GCA_FA data set is all Fall 15 and Spring 16 pre and post GCA scores (2346 case).

```
X1516GCA_FA <- read.csv(file = "1516GCA.csv", header = TRUE)
```

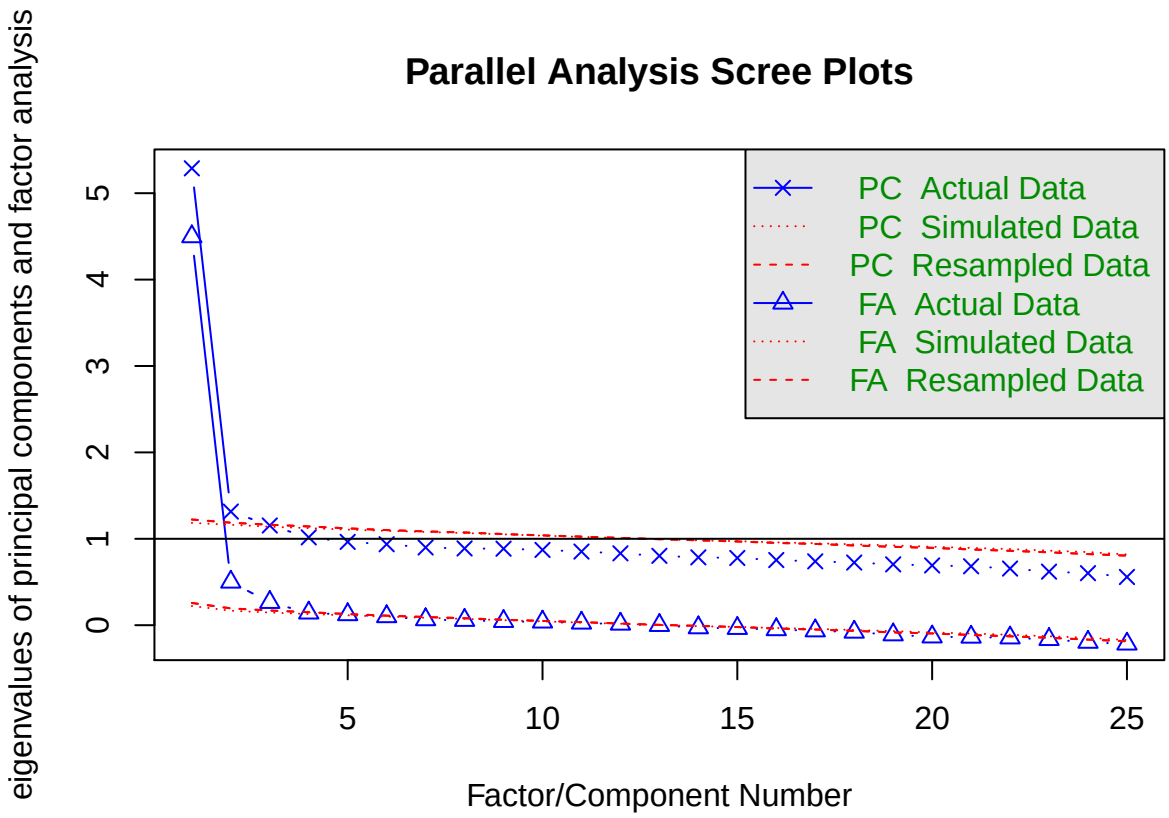
First look at the correlaton matrix for GCA items:

```
library(psych)
library(GPArotation)
corPlot(X1516GCA_FA)
```



One way to determine the number of factors is to compare the solution to a set of simulated random data with properties similar to the GCA data set (a parallel analysis). Running this parallel analysis also produces the scree plot:

```
fa.parallel((X1516GCA_FA))
```




```
## Parallel analysis suggests that the number of factors =  6  and the number of components =  3
```

The parallel analysis suggest 3 components and 5 factors and the scree plot shows 3-4 components with eigenvalue > 1, so run PCA (descriptive model) with 3 factors, varimax rotation

```
principal(X1516GCA_FA, nfactors=3, rotate = "varimax")
```

```
## Principal Components Analysis
## Call: principal(r = X1516GCA_FA, nfactors = 3, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##         RC1   RC2   RC3    h2   u2 com
## GCA1   0.21  0.52  0.13 0.332 0.67 1.5
## GCA2   0.21  0.12  0.34 0.175 0.83 1.9
## GCA3   0.18  0.22  0.03 0.081 0.92 1.9
## GCA4   0.19  0.51  0.08 0.301 0.70 1.3
## GCA5   0.13  0.63  0.02 0.410 0.59 1.1
## GCA6   0.49  0.14  0.27 0.334 0.67 1.7
## GCA7   0.19  0.53  0.23 0.370 0.63 1.6
## GCA8   0.24  0.17  0.37 0.225 0.78 2.2
## GCA9  -0.05  0.61 -0.05 0.378 0.62 1.0
## GCA10  0.68  0.12  0.00 0.477 0.52 1.1
## GCA11  0.48  0.32  0.01 0.335 0.66 1.7
## GCA12  0.51  0.13  0.04 0.278 0.72 1.1
## GCA13  0.16  0.33  0.43 0.316 0.68 2.2
## GCA14  0.42  0.06  0.22 0.225 0.77 1.5
## GCA15  0.38  0.03  0.24 0.199 0.80 1.7
## GCA16  0.46  0.22  0.20 0.296 0.70 1.9
```

```
## GCA17  0.09  0.51  0.17 0.292 0.71 1.3
## GCA18  0.47  0.17  0.16 0.272 0.73 1.5
## GCA19 -0.12 -0.04  0.73 0.552 0.45 1.1
## GCA20  0.43  0.20 -0.26 0.299 0.70 2.1
## GCA21  0.29  0.41  0.10 0.262 0.74 1.9
## GCA22  0.42  0.25  0.20 0.280 0.72 2.1
## GCA23  0.64  0.08 -0.01 0.422 0.58 1.0
## GCA24  0.43  0.34  0.27 0.372 0.63 2.6
## GCA25  0.34  0.21  0.33 0.274 0.73 2.7
##
##                        RC1  RC2  RC3
## SS loadings           3.43 2.70 1.63
## Proportion Var        0.14 0.11 0.07
## Cumulative Var        0.14 0.25 0.31
## Proportion Explained  0.44 0.35 0.21
## Cumulative Proportion 0.44 0.79 1.00
##
## Mean item complexity =  1.7
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  4018.23  with prob <  0
##
## Fit based upon off diagonal values = 0.91
```

Now compare to PCA with 4 factors:

```
principal(X1516GCA_FA, nfactors=4, rotate = "varimax")
```

```
## Principal Components Analysis
## Call: principal(r = X1516GCA_FA, nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##         RC1   RC2   RC3   RC4   h2   u2 com
## GCA1   0.20  0.46  0.11  0.29 0.35 0.65 2.3
## GCA2   0.23  0.11  0.33  0.05 0.18 0.82 2.1
## GCA3   0.10 -0.01 -0.04  0.86 0.75 0.25 1.0
## GCA4   0.22  0.54  0.08 -0.05 0.35 0.65 1.4
## GCA5   0.14  0.61  0.02  0.13 0.41 0.59 1.2
## GCA6   0.51  0.12  0.25  0.06 0.34 0.66 1.6
## GCA7   0.20  0.52  0.22  0.12 0.37 0.63 1.8
## GCA8   0.25  0.13  0.35  0.17 0.23 0.77 2.6
## GCA9  -0.02  0.66 -0.03 -0.11 0.45 0.55 1.1
## GCA10  0.68  0.10 -0.03  0.06 0.48 0.52 1.1
## GCA11  0.47  0.27 -0.02  0.22 0.34 0.66 2.0
## GCA12  0.52  0.13  0.02  0.00 0.29 0.71 1.1
## GCA13  0.16  0.25  0.40  0.32 0.35 0.65 3.0
## GCA14  0.41  0.01  0.19  0.18 0.24 0.76 1.8
## GCA15  0.39  0.01  0.22  0.04 0.20 0.80 1.6
## GCA16  0.47  0.21  0.18  0.07 0.30 0.70 1.8
## GCA17  0.07  0.42  0.14  0.39 0.35 0.65 2.3
## GCA18  0.48  0.16  0.13  0.04 0.28 0.72 1.4
## GCA19 -0.08 -0.01  0.75 -0.10 0.57 0.43 1.1
## GCA20  0.43  0.20 -0.28  0.02 0.30 0.70 2.2
## GCA21  0.30  0.42  0.09  0.04 0.27 0.73 2.0
## GCA22  0.43  0.23  0.18  0.11 0.28 0.72 2.0
```

```
## GCA23  0.65  0.07 -0.04  0.02 0.43 0.57 1.0
## GCA24  0.44  0.30  0.24  0.16 0.37 0.63 2.7
## GCA25  0.37  0.23  0.32 -0.04 0.30 0.70 2.7
##
##                        RC1  RC2  RC3  RC4
## SS loadings           3.51 2.46 1.52 1.29
## Proportion Var        0.14 0.10 0.06 0.05
## Cumulative Var        0.14 0.24 0.30 0.35
## Proportion Explained  0.40 0.28 0.17 0.15
## Cumulative Proportion 0.40 0.68 0.85 1.00
##
## Mean item complexity =  1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  4496.83  with prob <  0
##
## Fit based upon off diagonal values = 0.9
```

4 factors only accounts for 4% more variance than the 3 factor PCA. But look at communalities for GCA3 in each solution (0.081 in 3 factor, 0.75 in 4 factor)?

Now compare the PCAs to a factor analysis (structural model) specifying 3 factors, varimax rotation, do not impute values for missing, use minimum residual factoring method (default) and view loading matrix

```
fa(X1516GCA_FA, nfactors = 3, rotate = "varimax")
```

```
## Factor Analysis using method =  minres
## Call: fa(r = X1516GCA_FA, nfactors = 3, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##         MR1  MR2   MR3    h2   u2 com
## GCA1   0.23 0.43  0.14 0.257 0.74 1.8
## GCA2   0.21 0.15  0.19 0.102 0.90 2.8
## GCA3   0.17 0.16  0.06 0.059 0.94 2.3
## GCA4   0.23 0.40  0.08 0.221 0.78 1.7
## GCA5   0.18 0.51  0.04 0.292 0.71 1.2
## GCA6   0.44 0.18  0.21 0.266 0.73 1.8
## GCA7   0.23 0.44  0.21 0.291 0.71 2.0
## GCA8   0.24 0.19  0.23 0.148 0.85 2.9
## GCA9   0.06 0.40  0.02 0.165 0.83 1.0
## GCA10  0.62 0.14  0.00 0.405 0.60 1.1
## GCA11  0.43 0.29  0.05 0.272 0.73 1.8
## GCA12  0.41 0.17  0.06 0.200 0.80 1.4
## GCA13  0.21 0.28  0.31 0.222 0.78 2.7
## GCA14  0.34 0.12  0.16 0.156 0.84 1.7
## GCA15  0.30 0.11  0.15 0.124 0.88 1.8
## GCA16  0.41 0.23  0.16 0.247 0.75 1.9
## GCA17  0.15 0.38  0.16 0.191 0.81 1.7
## GCA18  0.40 0.19  0.14 0.216 0.78 1.7
## GCA19 -0.02 0.03  0.37 0.142 0.86 1.0
## GCA20  0.32 0.18 -0.10 0.145 0.86 1.7
## GCA21  0.29 0.34  0.12 0.209 0.79 2.2
## GCA22  0.38 0.25  0.17 0.235 0.76 2.2
## GCA23  0.55 0.12  0.00 0.322 0.68 1.1
## GCA24  0.42 0.31  0.24 0.333 0.67 2.5
## GCA25  0.33 0.22  0.25 0.219 0.78 2.7
```

```
##
##                        MR1  MR2  MR3
## SS loadings           2.77 1.92 0.74
## Proportion Var        0.11 0.08 0.03
## Cumulative Var        0.11 0.19 0.22
## Proportion Explained  0.51 0.35 0.14
## Cumulative Proportion 0.51 0.86 1.00
##
## Mean item complexity =  1.9
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  300  and the objective function was  3.54 with Chi Sq
## The degrees of freedom for the model are 228  and the objective function was  0.17
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.02
##
## The harmonic number of observations is  2056 with the empirical chi square  422.44  with prob <  8.8
## The total number of observations was  2346  with Likelihood Chi Square =  401.54  with prob <  1e-11
##
## Tucker Lewis Index of factoring reliability =  0.971
## RMSEA index =  0.018  and the 90 % confidence intervals are  0.015 0.021
## BIC =  -1367.85
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                MR1  MR2   MR3
## Correlation of scores with factors            0.82 0.75  0.60
## Multiple R square of scores with factors      0.68 0.56  0.36
## Minimum correlation of possible factor scores  0.36 0.12 -0.29
```

And finally run the FA with 4 factors:

```
fa(X1516GCA_FA, nfactors = 4, rotate = "varimax")
```

```
## Factor Analysis using method =  minres
## Call: fa(r = X1516GCA_FA, nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##         MR1  MR2   MR3   MR4    h2   u2 com
## GCA1   0.24 0.43  0.17 -0.07 0.278 0.72 2.0
## GCA2   0.21 0.14  0.19  0.00 0.103 0.90 2.8
## GCA3   0.17 0.17  0.08 -0.05 0.066 0.93 2.6
## GCA4   0.23 0.40  0.10  0.03 0.220 0.78 1.7
## GCA5   0.18 0.50  0.04  0.06 0.291 0.71 1.3
## GCA6   0.44 0.17  0.23  0.01 0.269 0.73 1.8
## GCA7   0.23 0.44  0.22  0.03 0.292 0.71 2.0
## GCA8   0.25 0.19  0.25 -0.05 0.161 0.84 2.9
## GCA9   0.04 0.41 -0.01  0.18 0.202 0.80 1.4
## GCA10  0.61 0.13  0.01  0.11 0.407 0.59 1.2
## GCA11  0.43 0.29  0.07  0.03 0.273 0.73 1.8
## GCA12  0.41 0.16  0.06  0.09 0.202 0.80 1.5
## GCA13  0.20 0.28  0.32  0.03 0.224 0.78 2.7
## GCA14  0.33 0.11  0.16  0.10 0.157 0.84 1.9
## GCA15  0.30 0.10  0.16  0.01 0.125 0.87 1.8
## GCA16  0.40 0.22  0.16  0.11 0.249 0.75 2.1
## GCA17  0.16 0.38  0.17  0.00 0.194 0.81 1.8
```

```
## GCA18  0.39 0.18  0.13  0.14 0.223 0.78 2.0
## GCA19 -0.03 0.03  0.36  0.05 0.131 0.87 1.1
## GCA20  0.33 0.18 -0.08 -0.01 0.146 0.85 1.7
## GCA21  0.27 0.33  0.11  0.19 0.228 0.77 2.8
## GCA22  0.38 0.25  0.19 -0.03 0.245 0.75 2.2
## GCA23  0.56 0.12  0.01 -0.02 0.332 0.67 1.1
## GCA24  0.41 0.30  0.25  0.12 0.333 0.67 2.8
## GCA25  0.30 0.18  0.26  0.35 0.313 0.69 3.4
##
##                        MR1  MR2  MR3  MR4
## SS loadings           2.72 1.86 0.79 0.29
## Proportion Var        0.11 0.07 0.03 0.01
## Cumulative Var        0.11 0.18 0.21 0.23
## Proportion Explained  0.48 0.33 0.14 0.05
## Cumulative Proportion 0.48 0.81 0.95 1.00
##
## Mean item complexity =  2
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  300  and the objective function was  3.54 with Chi Sq
## The degrees of freedom for the model are 206  and the objective function was  0.15
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.02
##
## The harmonic number of observations is  2056 with the empirical chi square  355.87  with prob <  4.3
## The total number of observations was  2346  with Likelihood Chi Square =  345.33  with prob <  4e-09
##
## Tucker Lewis Index of factoring reliability =  0.975
## RMSEA index =  0.017  and the 90 % confidence intervals are  0.014 0.02
## BIC =  -1253.32
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                 MR1  MR2   MR3   MR4
## Correlation of scores with factors             0.82 0.75  0.60  0.48
## Multiple R square of scores with factors       0.67 0.56  0.36  0.23
## Minimum correlation of possible factor scores  0.35 0.11 -0.27 -0.53
```

The 4 factor FA only accounts for 1% more variance than the 3 factor FA, and is likely harder to interpret. Also the communalities (variance accounted for in each item by all factors in the solution) does not increase much for 4 factors versus 3.