Principal Components and Factor Analysis

Bud Talbot

18 May 2017

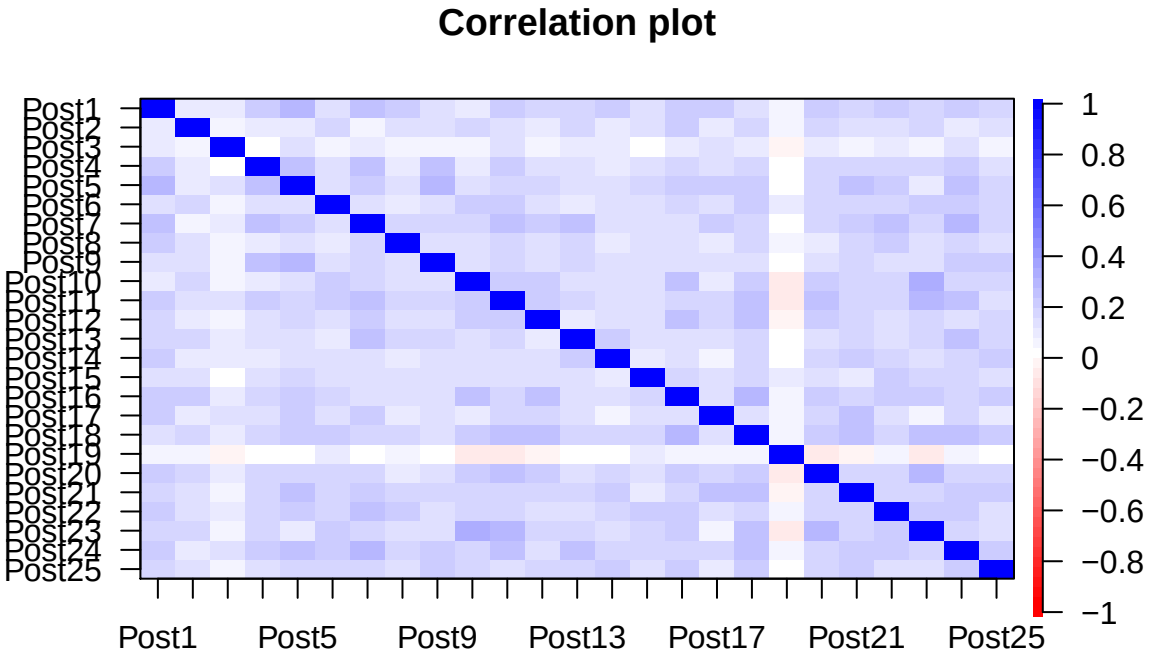
Initial Principal Components and Factor Analysis of GCA data.

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

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

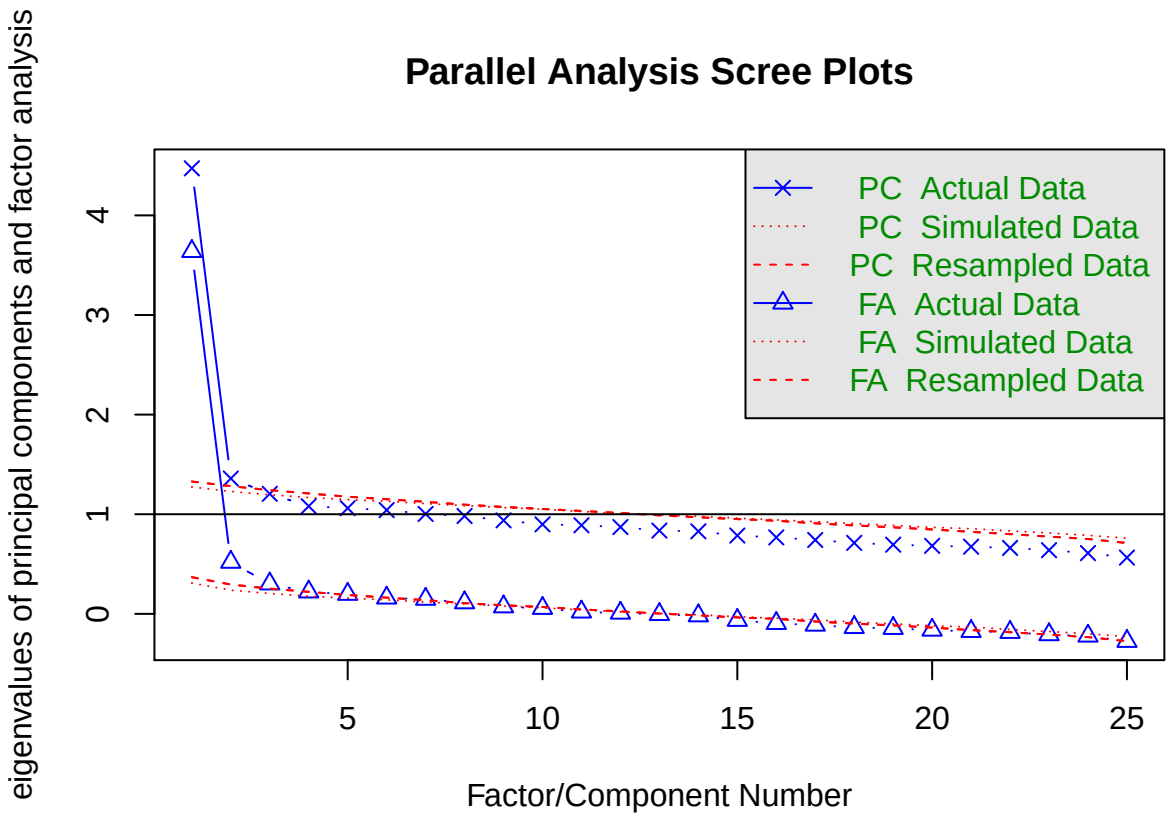
First look at the correlaton matrix for GCA items:

```
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 =  8  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
## Post1   0.51  0.07  0.20 0.30 0.70 1.3
## Post2  -0.03  0.23  0.43 0.24 0.76 1.5
## Post3   0.31  0.12 -0.16 0.14 0.86 1.8
## Post4   0.46  0.12  0.10 0.23 0.77 1.2
## Post5   0.59  0.07  0.10 0.36 0.64 1.1
## Post6   0.10  0.25  0.39 0.22 0.78 1.9
## Post7   0.56  0.11  0.10 0.34 0.66 1.1
## Post8   0.20  0.07  0.37 0.18 0.82 1.6
## Post9   0.44  0.07  0.14 0.22 0.78 1.3
## Post10  0.02  0.61  0.22 0.43 0.57 1.3
## Post11  0.36  0.46  0.04 0.34 0.66 1.9
## Post12  0.24  0.40  0.10 0.23 0.77 1.8
## Post13  0.32  0.09  0.28 0.19 0.81 2.1
## Post14  0.24  0.16  0.23 0.14 0.86 2.7
## Post15  0.08  0.05  0.50 0.26 0.74 1.1
## Post16  0.12  0.35  0.43 0.32 0.68 2.1
```

```
## Post17  0.51  0.03  0.02 0.26 0.74 1.0
## Post18  0.21  0.36  0.34 0.29 0.71 2.6
## Post19 -0.05 -0.48  0.55 0.53 0.47 2.0
## Post20  0.22  0.53  0.09 0.33 0.67 1.4
## Post21  0.44  0.19  0.16 0.25 0.75 1.6
## Post22  0.27  0.13  0.39 0.25 0.75 2.0
## Post23  0.00  0.65  0.23 0.48 0.52 1.3
## Post24  0.47  0.07  0.30 0.32 0.68 1.7
## Post25  0.25  0.15  0.34 0.20 0.80 2.3
##
##                        RC1  RC2  RC3
## SS loadings           2.74 2.22 2.08
## Proportion Var        0.11 0.09 0.08
## Cumulative Var        0.11 0.20 0.28
## Proportion Explained  0.39 0.32 0.30
## Cumulative Proportion 0.39 0.70 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  0.06
##  with the empirical chi square  2397.9  with prob <  0
##
## Fit based upon off diagonal values = 0.85
```

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
##          RC2   RC1   RC3   RC4   h2   u2 com
## Post1   0.06  0.34  0.19  0.39 0.31 0.69 2.5
## Post2   0.27 -0.02  0.43  0.01 0.25 0.75 1.7
## Post3   0.09 -0.19 -0.05  0.64 0.46 0.54 1.2
## Post4   0.10  0.57  0.01  0.09 0.34 0.66 1.1
## Post5   0.04  0.54  0.05  0.29 0.38 0.62 1.6
## Post6   0.28  0.11  0.36  0.06 0.23 0.77 2.2
## Post7   0.09  0.37  0.09  0.44 0.34 0.66 2.1
## Post8   0.09  0.01  0.41  0.30 0.26 0.74 2.0
## Post9   0.06  0.61  0.03  0.01 0.38 0.62 1.0
## Post10  0.63  0.08  0.15  0.00 0.43 0.57 1.2
## Post11  0.44  0.13  0.04  0.41 0.38 0.62 2.2
## Post12  0.40  0.31  0.02  0.07 0.26 0.74 2.0
## Post13  0.10  0.19  0.28  0.28 0.20 0.80 3.0
## Post14  0.17  0.28  0.18  0.08 0.14 0.86 2.6
## Post15  0.09  0.14  0.49  0.00 0.26 0.74 1.2
## Post16  0.38  0.20  0.37  0.00 0.32 0.68 2.5
## Post17  0.00  0.20  0.06  0.52 0.31 0.69 1.3
## Post18  0.37  0.25  0.29  0.08 0.29 0.71 2.8
## Post19 -0.42 -0.03  0.60 -0.05 0.54 0.46 1.8
## Post20  0.52  0.16  0.04  0.19 0.34 0.66 1.5
## Post21  0.18  0.43  0.10  0.20 0.27 0.73 1.9
## Post22  0.15  0.15  0.40  0.26 0.27 0.73 2.4
```

```
## Post23  0.67  0.02  0.18  0.04 0.48 0.52 1.2
## Post24  0.07  0.32  0.29  0.37 0.33 0.67 3.0
## Post25  0.16  0.49  0.24 -0.11 0.33 0.67 1.8
##
##                        RC2  RC1  RC3  RC4
## SS loadings           2.26 2.26 1.83 1.77
## Proportion Var        0.09 0.09 0.07 0.07
## Cumulative Var        0.09 0.18 0.25 0.32
## Proportion Explained  0.28 0.28 0.23 0.22
## Cumulative Proportion 0.28 0.56 0.78 1.00
##
## Mean item complexity =  1.9
## 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  2831.87  with prob <  0
##
## Fit based upon off diagonal values = 0.82
```

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
## Post1  0.44 0.17  0.03 0.22 0.78 1.3
## Post2  0.09 0.33  0.08 0.12 0.88 1.3
## Post3  0.18 0.04 -0.07 0.04 0.96 1.4
## Post4  0.38 0.15 -0.03 0.16 0.84 1.3
## Post5  0.49 0.13  0.01 0.25 0.75 1.1
## Post6  0.17 0.34  0.05 0.15 0.85 1.5
## Post7  0.47 0.15 -0.04 0.25 0.75 1.2
## Post8  0.23 0.23  0.08 0.11 0.89 2.3
## Post9  0.35 0.15  0.01 0.15 0.85 1.3
## Post10 0.11 0.49 -0.18 0.28 0.72 1.4
## Post11 0.34 0.31 -0.21 0.25 0.75 2.7
## Post12 0.25 0.29 -0.11 0.16 0.84 2.3
## Post13 0.29 0.22  0.05 0.14 0.86 1.9
## Post14 0.23 0.22  0.03 0.11 0.89 2.0
## Post15 0.17 0.28  0.15 0.13 0.87 2.3
## Post16 0.20 0.44  0.06 0.23 0.77 1.4
## Post17 0.37 0.08  0.00 0.14 0.86 1.1
## Post18 0.26 0.40  0.01 0.23 0.77 1.7
## Post19 0.00 0.02  0.42 0.18 0.82 1.0
## Post20 0.24 0.36 -0.20 0.23 0.77 2.4
## Post21 0.38 0.22 -0.02 0.19 0.81 1.6
## Post22 0.29 0.30  0.09 0.18 0.82 2.2
## Post23 0.09 0.54 -0.24 0.35 0.65 1.4
## Post24 0.43 0.23  0.09 0.24 0.76 1.6
## Post25 0.27 0.27  0.08 0.15 0.85 2.2
```

```
##
##                        MR1  MR2  MR3
## SS loadings           2.20 2.02 0.44
## Proportion Var        0.09 0.08 0.02
## Cumulative Var        0.09 0.17 0.19
## Proportion Explained  0.47 0.43 0.09
## Cumulative Proportion 0.47 0.91 1.00
##
## Mean item complexity =  1.7
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  300  and the objective function was  2.85 with Chi Sq
## The degrees of freedom for the model are 228  and the objective function was  0.32
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.03
##
## The harmonic number of observations is  989 with the empirical chi square  418.44  with prob <  2.2e
## The total number of observations was  1173  with Likelihood Chi Square =  370.31  with prob <  7.4e-
##
## Tucker Lewis Index of factoring reliability =  0.938
## RMSEA index =  0.023  and the 90 % confidence intervals are  0.019 0.027
## BIC =  -1241.04
## Fit based upon off diagonal values = 0.97
## Measures of factor score adequacy
##                                                MR1  MR2   MR3
## Correlation of scores with factors             0.79 0.78  0.59
## Multiple R square of scores with factors       0.63 0.61  0.35
## Minimum correlation of possible factor scores  0.26 0.22 -0.30
```

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
##         MR2   MR1  MR4   MR3    h2   u2 com
## Post1  0.16  0.34 0.28 -0.01 0.218 0.78 2.4
## Post2  0.33  0.07 0.08 -0.05 0.126 0.87 1.3
## Post3  0.03  0.14 0.11  0.08 0.039 0.96 2.6
## Post4  0.15  0.33 0.19  0.05 0.166 0.83 2.1
## Post5  0.14  0.59 0.07  0.01 0.368 0.63 1.2
## Post6  0.32  0.11 0.18 -0.01 0.146 0.85 1.8
## Post7  0.08  0.29 0.45  0.08 0.297 0.70 1.9
## Post8  0.20  0.10 0.28 -0.06 0.131 0.87 2.2
## Post9  0.15  0.35 0.13  0.01 0.163 0.84 1.7
## Post10 0.46  0.06 0.12  0.22 0.278 0.72 1.7
## Post11 0.25  0.20 0.31  0.26 0.269 0.73 3.7
## Post12 0.28  0.23 0.11  0.14 0.166 0.83 2.8
## Post13 0.17  0.12 0.37 -0.02 0.179 0.82 1.7
## Post14 0.21  0.16 0.19 -0.01 0.106 0.89 2.9
## Post15 0.28  0.13 0.13 -0.12 0.126 0.87 2.4
## Post16 0.48  0.23 0.03 -0.04 0.289 0.71 1.5
## Post17 0.08  0.33 0.17  0.02 0.145 0.86 1.7
```

```
## Post18 0.39  0.20 0.18  0.03 0.229 0.77 2.0
## Post19 0.03 -0.02 0.06 -0.40 0.166 0.83 1.1
## Post20 0.34  0.21 0.13  0.23 0.232 0.77 2.8
## Post21 0.21  0.31 0.23  0.04 0.192 0.81 2.7
## Post22 0.28  0.21 0.23 -0.06 0.181 0.82 2.9
## Post23 0.49 -0.02 0.20  0.29 0.364 0.64 2.0
## Post24 0.18  0.25 0.42 -0.05 0.279 0.72 2.1
## Post25 0.28  0.22 0.17 -0.06 0.157 0.84 2.7
##
##                        MR2  MR1  MR4  MR3
## SS loadings           1.84 1.47 1.21 0.49
## Proportion Var        0.07 0.06 0.05 0.02
## Cumulative Var        0.07 0.13 0.18 0.20
## Proportion Explained  0.37 0.29 0.24 0.10
## Cumulative Proportion 0.37 0.66 0.90 1.00
##
## Mean item complexity =  2.1
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  300  and the objective function was  2.85 with Chi Sq
## The degrees of freedom for the model are 206  and the objective function was  0.27
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.03
##
## The harmonic number of observations is  989 with the empirical chi square  343.3  with prob <  6.1e-
## The total number of observations was  1173  with Likelihood Chi Square =  308.87  with prob <  4.5e-
##
## Tucker Lewis Index of factoring reliability =  0.95
## RMSEA index =  0.021  and the 90 % confidence intervals are  0.016 0.025
## BIC =  -1147
## Fit based upon off diagonal values = 0.97
## Measures of factor score adequacy
##                                                MR2  MR1   MR4   MR3
## Correlation of scores with factors            0.76 0.73  0.66  0.60
## Multiple R square of scores with factors      0.58 0.53  0.44  0.36
## Minimum correlation of possible factor scores  0.17 0.05 -0.12 -0.27
```

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.