Psychology 405: Psychometric Theory Homework on Factor analysis and structural equation modeling

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

May, 2020

Outline

The problem

1. Given the data set at http://personality-project.org/r/datasets/psychometrics.prob2.txt
   - Do basic descriptive statistics
   - Find the basic correlation matrix
2. Exploratory Factor analysis
   - How many factors?
   - What are they?
3. Factor Extension
   - Factor the first 5 variables
   - Extend to the last 3
4. Do this as a confirmatory model
   - With sem
   - With lavaan

Read and describe

```
#Give the file name (location/path)
> fn <-"http://personality-project.org/r/datasets/psychometrics.prob2.txt"

 #Read in the data
> dataset <- read.table(fn,header=TRUE)
 # Do basic descriptive statistics
> describe(dataset)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ID	1	1000	500.50	288.82	500.50	500.50	370.65	1.0	1000.00	999.00	0.00	-1.20	9.13
GREV	2	1000	499.77	106.11	497.50	498.75	106.01	138.0	873.00	735.00	0.09	-0.07	3.36
GREQ	3	1000	500.53	103.85	498.00	498.51	105.26	191.0	914.00	723.00	0.22	0.08	3.28
GREA	4	1000	498.13	100.45	495.00	498.67	99.33	207.0	848.00	641.00	-0.02	-0.06	3.18
Ach	5	1000	49.93	9.84	50.00	49.88	10.38	16.0	79.00	63.00	0.00	0.02	0.31
Anx	6	1000	50.32	9.91	50.00	50.43	10.38	14.0	78.00	64.00	-0.14	0.14	0.31
Prelim	7	1000	10.03	1.06	10.00	10.02	1.48	7.0	13.00	6.00	-0.02	-0.01	0.03
GPA	8	1000	4.00	0.50	4.02	4.01	0.53	2.5	5.38	2.88	-0.07	-0.29	0.02
MA	9	1000	3.00	0.49	3.00	3.00	0.44	1.4	4.50	3.10	-0.07	-0.09	0.02

```
>
```

The correlation matrix

```
> R <- lowerCor(dataset)
```

```
       ID    GREV  GREQ  GREA  Ach   Anx   Prelm GPA   MA
ID      1.00
GREV   -0.01  1.00
GREQ    0.00  0.73  1.00
GREA   -0.01  0.64  0.60  1.00
Ach     0.00  0.01  0.01  0.45  1.00
Anx    -0.01  0.01  0.01 -0.39 -0.56  1.00
Prelim  0.02  0.43  0.38  0.57  0.30 -0.23  1.00
GPA     0.00  0.42  0.37  0.52  0.28 -0.22  0.42  1.00
MA     -0.01  0.32  0.29  0.45  0.26 -0.22  0.36  0.31  1.00
```

Drop the ID field and redo the analysis

```
> my.data <- dataset[-1]
> R <-lowerCor(my.data)
       GREV  GREQ  GREA  Ach   Anx   Prelm GPA   MA
GREV    1.00
GREQ    0.73  1.00
GREA    0.64  0.60  1.00
Ach     0.01  0.01  0.45  1.00
Anx     0.01  0.01 -0.39 -0.56  1.00
Prelim  0.43  0.38  0.57  0.30 -0.23  1.00
GPA     0.42  0.37  0.52  0.28 -0.22  0.42  1.00
MA      0.32  0.29  0.45  0.26 -0.22  0.36  0.31  1.00
```

How many factors?

```
> nfactors(my.data)

Number of factors
Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,
    n.obs = n.obs, plot = FALSE, title = title)
VSS complexity 1 achieves a maximimum of 0.74  with  2  factors
VSS complexity 2 achieves a maximimum of 0.88  with  2  factors
The Velicer MAP achieves a minimum of 0.06  with  2  factors
Empirical BIC achieves a minimum of  -71.29  with  2  factors
Sample Size adjusted BIC achieves a minimum of  -23.74  with  3  factors

Statistics by number of factors
  vss1 vss2   map dof   chisq     prob sqresid  fit RMSEA BIC SABIC complex  eC
1 0.73 0.00 0.071  20 9.7e+02 3.7e-192     4.5 0.73 0.218 829 892.6     1.0 1.1
2 0.74 0.88 0.056  13 2.6e+01  1.9e-02     2.0 0.88 0.031 -64 -23.0     1.4 1.9
3 0.54 0.82 0.101   7 2.4e+00  9.4e-01     1.7 0.90 0.000 -46 -23.7     1.8 7.4
4 0.53 0.81 0.179   2 4.6e-02  9.8e-01     1.7 0.90 0.000 -14  -7.4     1.9 2.3
5 0.46 0.68 0.344  -2 1.3e-02       NA     1.6 0.91    NA  NA    NA     2.1 8.2
6 0.44 0.68 0.516  -5 1.7e-09       NA     1.4 0.92    NA  NA    NA     2.1 1.8
7 0.44 0.68 1.000  -7 0.0e+00       NA     1.4 0.92    NA  NA    NA     2.1 3.0
8 0.44 0.68    NA  -8 0.0e+00       NA     1.4 0.92    NA  NA    NA     2.1 3.0
```

What are the two factors?

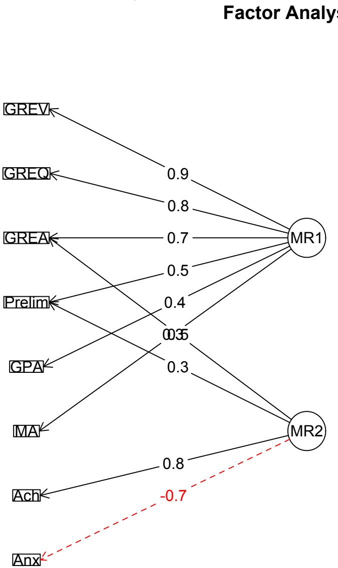
```
> f2 <- fa(my.data,2)
> f2

Factor Analysis using method =  minres
Call: fa(r = my.data, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
         MR1   MR2   h2   u2 com
GREV    0.91 -0.14 0.79 0.21 1.0
GREQ    0.84 -0.13 0.67 0.33 1.0
GREA    0.70  0.46 0.84 0.16 1.7
Ach    -0.06  0.81 0.63 0.37 1.0
Anx     0.07 -0.71 0.48 0.52 1.0
Prelim  0.47  0.31 0.39 0.61 1.7
GPA     0.45  0.27 0.33 0.67 1.6
MA      0.35  0.29 0.25 0.75 1.9

                       MR1  MR2
SS loadings           2.65 1.73
Proportion Var        0.33 0.22
Cumulative Var        0.33 0.55
Proportion Explained  0.60 0.40
Cumulative Proportion 0.60 1.00
```

two factors (continued)

```
 With factor correlations of 
     MR1  MR2
MR1 1.00 0.23
MR2 0.23 1.00
Mean item complexity =  1.4
Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are  28  and the objective function was  3.32
               with Chi Square of  3304.93
The degrees of freedom for the model are 13  and the objective function was  0.03 

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  1000 with the empirical chi square  18.51  with prob <  0.14 
The total number of observations was  1000  with MLE Chi Square =  25.56  with prob <  0.019 

Tucker Lewis Index of factoring reliability =  0.992
RMSEA index =  0.031  and the 90 % confidence intervals are  0.012 0.049
BIC =  -64.24
Fit based upon off diagonal values = 1
Measures of factor score adequacy             
                                                MR1  MR2
Correlation of scores with factors             0.95 0.90
Multiple R square of scores with factors       0.91 0.82
Minimum correlation of possible factor scores  0.81 0.64
>
```

Show the two factor solution

```
> fa.diagram(f2,simple=FALSE)
```



What about a three factor solution?

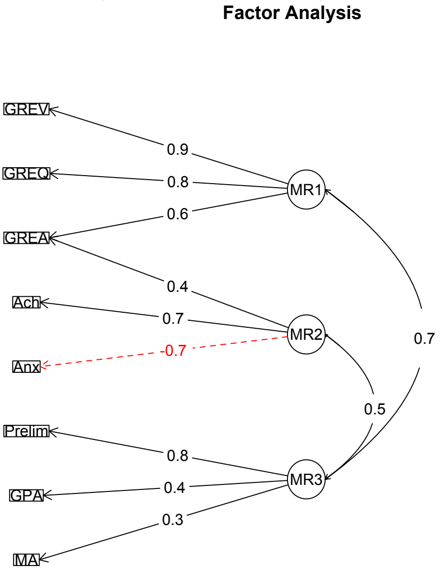
```
> f3 <- fa(my.data,3)
> f3

Factor Analysis using method =  minres
Call: fa(r = my.data, nfactors = 3)
Standardized loadings (pattern matrix) based upon correlation matrix
         MR1   MR2   MR3   h2   u2 com
GREV    0.85 -0.10  0.07 0.79 0.21 1.0
GREQ    0.85 -0.05 -0.03 0.68 0.32 1.0
GREA    0.61  0.44  0.14 0.84 0.16 1.9
Ach    -0.10  0.75  0.10 0.63 0.37 1.1
Anx     0.01 -0.74  0.05 0.50 0.50 1.0
Prelim -0.02 -0.04  0.76 0.53 0.47 1.0
GPA     0.20  0.11  0.39 0.36 0.64 1.7
MA      0.14  0.15  0.32 0.26 0.74 1.8

                       MR1  MR2  MR3
SS loadings           2.05 1.44 1.10
Proportion Var        0.26 0.18 0.14
Cumulative Var        0.26 0.44 0.57
Proportion Explained  0.45 0.31 0.24
Cumulative Proportion 0.45 0.76 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.13 0.68
MR2 0.13 1.00 0.54
MR3 0.68 0.54 1.00
```

More 3 factor output

```
 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.13 0.68
MR2 0.13 1.00 0.54
MR3 0.68 0.54 1.00

Mean item complexity =  1.3
Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are  28  and the objective function was  3.32 with Chi Square of
The degrees of freedom for the model are 7  and the objective function was  0

The root mean square of the residuals (RMSR) is  0
The df corrected root mean square of the residuals is  0.01

The harmonic number of observations is  1000 with the empirical chi square  0.74  with prob <  1
The total number of observations was  1000  with MLE Chi Square =  2.38  with prob <  0.94

Tucker Lewis Index of factoring reliability =  1.006
RMSEA index =  0  and the 90 % confidence intervals are  NA 0.01
BIC =  -45.98
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1  MR2  MR3
Correlation of scores with factors             0.95 0.90 0.89
Multiple R square of scores with factors       0.90 0.81 0.79
Minimum correlation of possible factor scores  0.79 0.62 0.57
```

Show the three factor model

fa.diagram(f3,simple=FALSE)



Factor extension

1. Originally developed for the problem of new variables being added (Dwyer, 1937; Mosier, 1938; Horn, 1973)
    - Find the correlations of the new variables with the old factors
    - Do this by finding the factor score weights of the old variables on the factors
    - Then find the correlations of the new variables with those "scores"
    - Don't actually have to calculate the scores to do this
2. Can be used if we want to keep the original factor structure and "extend" it to new variables
    - `fa.extend` and `fa.extension` will do this
    - `fa.extend` will do the factor analysis on the old, and then merge output with the new
    - `fa.extension` takes the original factor analysis and the correlation of original with new and finds just the loadings on the new.

fa.extend

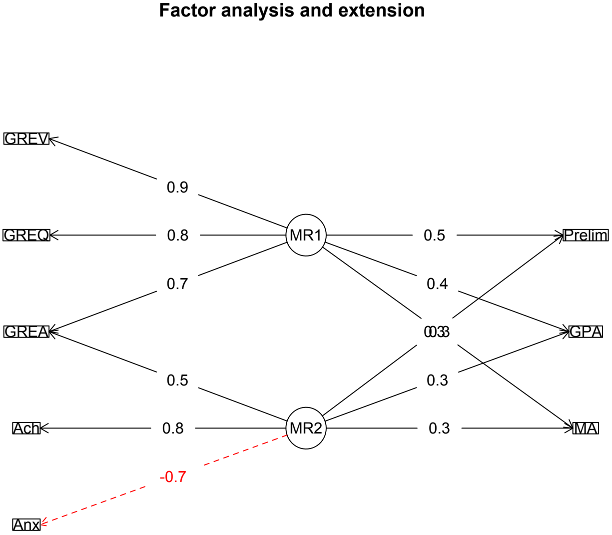
```
> f2e <- fa.extend(my.data,2,ov=1:5,ev=6:8)
> f2e

Factor Analysis using method =  minres
Call: fa.extend(r = my.data, nfactors = 2, ov = 1:5, ev = 6:8)
Standardized loadings (pattern matrix) based upon correlation matrix
         MR1   MR2   h2   u2 com
GREV    0.90 -0.11 0.79 0.21 1.0
GREQ    0.84 -0.10 0.68 0.32 1.0
GREA    0.69  0.49 0.84 0.16 1.8
Ach    -0.06  0.80 0.63 0.37 1.0
Anx     0.06 -0.71 0.49 0.51 1.0
Prelim  0.46  0.31 0.37 0.63 1.8
GPA     0.44  0.28 0.32 0.68 1.7
MA      0.34  0.29 0.24 0.76 1.9

                       MR1  MR2
SS loadings           2.60 1.75
Proportion Var        0.33 0.22
Cumulative Var        0.33 0.54
Proportion Explained  0.60 0.40
Cumulative Proportion 0.60 1.00

 With factor correlations of
     MR1  MR2
MR1 1.00 0.19
MR2 0.19 1.00
```

fa.extension (not showing the fa of the first 5 variables)

```
> f2o <- fa(my.data[1:5],2)
> f2e <- fa.extension(cor(my.data[1:5],my.data[6:8]), f2o)
> f2e


Call: fa.extension(Roe = cor(my.data[1:5], my.data[6:8]), fo = f2o)
Standardized loadings (pattern matrix) based upon correlation matrix
        MR1  MR2   h2   u2
Prelim 0.46 0.31 0.37 0.63
GPA    0.44 0.28 0.32 0.68
MA     0.34 0.29 0.24 0.76

                       MR1  MR2
SS loadings           0.59 0.33
Proportion Var        0.20 0.11
Cumulative Var        0.20 0.31
Proportion Explained  0.64 0.36
Cumulative Proportion 0.64 1.00
     MR1  MR2
MR1 1.00 0.19
MR2 0.19 1.00
```

Factor analysis and factor extension

fa.diagram(f2o,fe.result=f2e,simple=FALSE,cut=.2)



Three very good ways to do confirmatory analysis

1. *sem* by Fox, Nie & Byrnes (2013)
    - Uses RAM path notation
    - Can get some help from *psych* (see the psych-for-sem vignette)
2. *lavaan* by Rosseel (2012)
    - Somewhat easier to use
3. *OpenMx* by Neale, Hunter, Pritikin, Zahery, Brick, Kickpatrick, Estabrook, Bates, Maes & Boker (2016)
    - Most powerful package

Test a model with lavaan

```
z.data <- data.frame(scale(my.data) )#standardize
m1.model <-  'ability =~ GREV + GREQ + GREA
                     motive =~ GREA + Ach + Anx
                     perform =~ Prelim + GPA + MA'
fit <- sem(m1.model,data=z.data, auto.var=TRUE, auto.fix.first=TRUE,
              auto.cov.lv.x=TRUE)
summary(fit)
lavaan (0.5-16) converged normally after  22 iterations

  Number of observations                          1000

  Estimator                                         ML
  Minimum Function Test Statistic              306.074
  Degrees of freedom                                19
  P-value (Chi-square)                           0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
```

More lavaan

```
Latent variables:
  ability =~
    GREV              1.000
    GREQ              0.881    0.024   36.266    0.000
    GREA              0.784    0.023   33.433    0.000
  motive =~
    GREA              1.000
    Ach               1.048    0.027   38.579    0.000
    Anx              -0.931    0.028  -33.036    0.000
  perform =~
    Prelim            1.000
    GPA               0.781    0.029   26.653    0.000
    MA                0.671    0.030   22.076    0.000

Covariances:
  ability ~~
    motive            0.025    0.034    0.742    0.458
    perform           0.621    0.024   26.268    0.000
  motive ~~
    perform           0.716    0.022   31.955    0.000

Variances:
    GREV              0.185    0.019
    GREQ              0.339    0.021
    GREA              0.123    0.017
    Ach               0.421    0.027
    Anx               0.543    0.029
    Prelim            0.516    0.032
    GPA               0.620    0.032
    MA                0.719    0.035
    ability           1.000
    motive            1.000
    perform           1.000
```

Dwyer, P. S. (1937). The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, *2*(3), 173–178.

Fox, J., Nie, Z., & Byrnes, J. (2013). *sem: Structural Equation Models*. R package version 3.1-3.

Horn, J. L. (1973). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research*, *8*(4), 477 – 489.

Mosier, C. (1938). A note on Dwyer: The determination of the factor loadings of a given test. *Psychometrika*, *3*(4), 297–299.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kickpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.