Measurement, Scaling, & Dimensional Analysis
Prof. Adam Enders
ICPSR, 2019

Assignment 4:
Factor Analysis

I start my analysis by assessing the dimensionality of the values items. To reiterate, dimensionality is contingent on model assumptions about the data generating process, but most practitioners expect something like visual inspection of a scree plot or examination of the proportion of shared variance accounted for by each factor. A scree plot of eigenvalues against factors/dimensions is presented below.

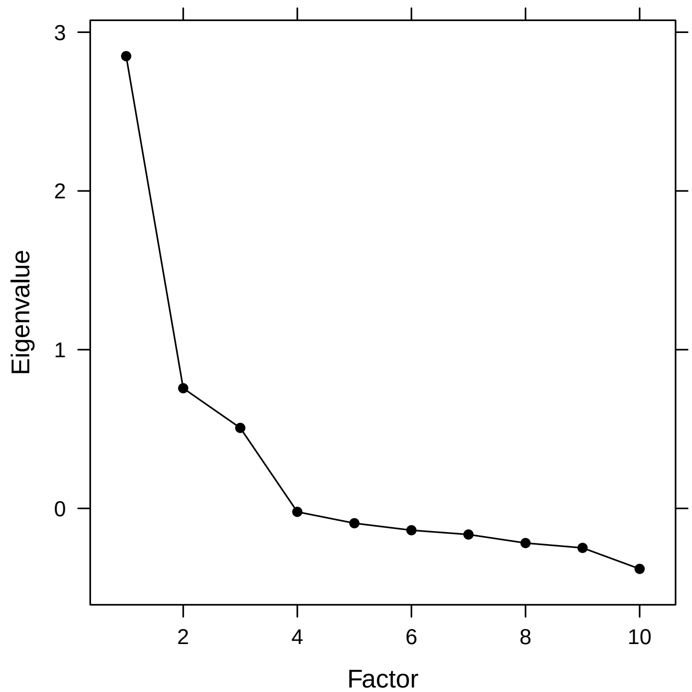
```
> check.scree <- fa(cor(values), fm = "pa", smc = TRUE, rotate = "none")
> xyplot(check.scree$values ~ 1:10,
+        aspect = 1,
+        type = "b",
+        col = "black",
+        xlab = "Factor",
+        ylab = "Eigenvalue",
+        pch = 16
+ )
```

Unfortunately, the scree plot does not contain a single, obvious "elbow" that I can use to make a determination about dimensionality. We observed elbows at 2 and 4 factors, suggesting either 1 or 3 factor solutions are most appropriate. A parallel analysis suggests that 3 factors is most appropriate. In the face of (partially) contradictory tests, and (more importantly) in light of theory, I am going to opt to retain 2 factors. Theoretically, equality and moral traditionalism are related, but distinct, value constructs, and I expect to observe latent factors that explain observed responses to the equality and moral traditionalism survey items.

Next, I re-estimated the common factor model retaining only 2 factors. Since unrotated factor loadings almost never conform to Thurstone's principle of "simple structure," I will forgo interpreting them substantively. Instead, I will move toward interpretation of the factor loadings (pattern coefficients) from an orthogonal (varimax) rotation, which are pictured below in both tabular and graphical form.

We can see from both the relative magnitude of the factor pattern coefficients ("loadings") and graphical depiction of the two factors and projected variable vectors ("shadow vectors") that the first factor is very highly related to the survey items dealing explicitly with issues of equality. Indeed, the first 6 values items, all of which are about equality, have loadings of 0.49 or higher on the first factor. The remaining items, which are about moral traditionalism, generally speaking, have loadings of 0.27 or less on the first factor. By a common rule of thumb, loadings of 0.30 or less should be treated with caution (not

Figure 1: Scree plot of eigenvalues against factors, in order.

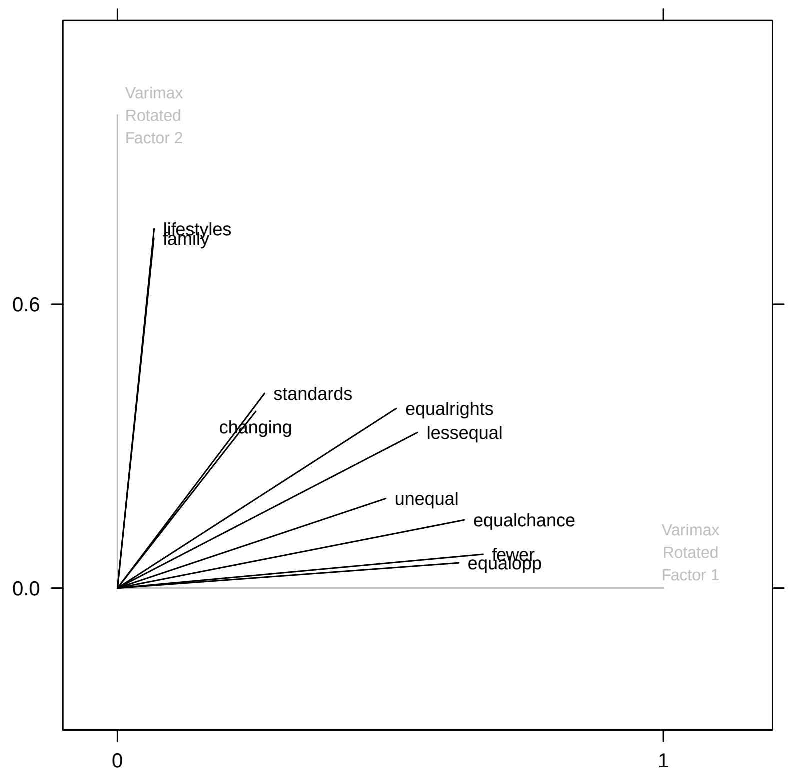


necessarily negligible, probably quite small *relative to other loadings*). Conversely, the moral traditionalism items have loadings between 0.37 and 0.76 on the second orthogonally-rotated factor, and the quality items have loadings between 0.05 and 0.38. Thus, we might interpret the first factor as "equality," and the second factor as "moral traditionalism."

```
Standardized loadings (pattern matrix) based upon correlation matrix
             PA1  PA2   h2   u2 com
equalopp    0.63 0.05 0.39 0.61 1.0
equalrights 0.51 0.38 0.41 0.59 1.8
equalchance 0.64 0.14 0.42 0.58 1.1
lessequal   0.55 0.33 0.41 0.59 1.6
unequal     0.49 0.19 0.28 0.72 1.3
fewer       0.67 0.07 0.45 0.55 1.0
changing    0.25 0.37 0.20 0.80 1.8
lifestyles  0.07 0.76 0.58 0.42 1.0
standards   0.27 0.41 0.24 0.76 1.7
family      0.07 0.74 0.55 0.45 1.0

                        PA1  PA2
Proportion Var         0.22 0.18
Cumulative Var         0.22 0.39
```

Figure 2: Varimax-rotated (orthogonal) factors from two-factor solution.



Although the varimax rotation provided factor loadings that aided me in lending some substantive interpretation to the factors (and, one that is pretty congruent with theory, in this case), I might want to rotate the factors in an oblique fashion as well. The reasons for doing so are twofold: 1) on a theoretical level, equality and moral traditionalism are probably not uncorrelated, and 2) oblique rotations have the potential to provided loading patterns that conform even better to Thurstone's "simple structure" principle.

The factor pattern coefficients (loadings) from the promax (oblique) rotation are depicted below. The substantive interpretation I lended to the factors via an examination of the loadings from a varimax rotation still hold. The only real difference between loadings is that some already large loadings increased, and some small loadings decreased. This is exactly what we'd want in order to achieve simple structure. We want as many loadings as close to 1 or 0 possible to make substantive interpretation "cleaner." It also turns out that, when not restricted to orthogonality, the two factors have a correlation of 0.55. Again, this makes good theoretical sense, especially since these variables have been recoded such that larger numerical values denote more conservative responses to the original survey items. At this point, I could write up my results and terminate my analysis with a discussion of value structure, or estimate individuals' positions along the latent factors for a subsequent statistical analysis.

```
Standardized loadings (pattern matrix) based upon correlation matrix
              PA1   PA2   h2   u2 com
equalopp     0.70 -0.15 0.39 0.61 1.1
equalrights  0.45  0.26 0.41 0.59 1.6
equalchance  0.68 -0.05 0.42 0.58 1.0
lessequal    0.52  0.19 0.41 0.59 1.3
unequal      0.50  0.05 0.28 0.72 1.0
fewer        0.74 -0.15 0.45 0.55 1.1
changing     0.16  0.34 0.20 0.80 1.4
lifestyles  -0.18  0.85 0.58 0.42 1.1
standards    0.17  0.38 0.24 0.76 1.4
family      -0.18  0.82 0.55 0.45 1.1

                       PA1  PA2
Proportion Var        0.22 0.17
Cumulative Var        0.22 0.39

 With factor correlations of
     PA1  PA2
PA1 1.00 0.55
PA2 0.55 1.00
```

Finally, I conducted a principal components analysis on the standardized values data using the "princomp" function.

```
> pcafit <- princomp(scale(values))
> pcafit
```

The proportion of variance explained by each principal component is different than the proportion of variance explained by each factor in the common factor analysis. In the PCA, the first two components account for about 51% of *total* variance in the dataset. In the factor analysis, the first two factors account for approximately 39% of *shared* or *common* variance among the set of variables. We should expect to see some differences because factor analysis and PCA are designed to explain different "types" of variance.

```
> summary(pcafit, loadings = TRUE, cutoff = 0)
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5    Comp.6
Standard deviation     1.8820111 1.2312952 1.1011701 0.84290850 0.76138710 0.7592693
Proportion of Variance 0.3542618 0.1516367 0.1212799 0.07106256 0.05798171 0.0576596
Cumulative Proportion  0.3542618 0.5058985 0.6271784 0.69824096 0.75622266 0.8138823
                           Comp.7     Comp.8     Comp.9    Comp.10
Standard deviation     0.75026080 0.69144973 0.64725302 0.63317000
Proportion of Variance 0.05629949 0.04781908 0.04190136 0.04009781
Cumulative Proportion  0.87018176 0.91800083 0.95990219 1.00000000
```

```
Loadings:
            Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
equalopp    -0.308 -0.346  0.214  0.031  0.827 -0.163 -0.118  0.083  0.066 -0.047
equalrights -0.370 -0.035 -0.336 -0.179 -0.105 -0.552  0.106 -0.232 -0.507 -0.281
equalchance -0.341 -0.265  0.227  0.452 -0.326 -0.027  0.147  0.580 -0.263  0.150
lessequal   -0.369 -0.108 -0.373 -0.237 -0.123 -0.187  0.248  0.126  0.641  0.343
unequal     -0.305 -0.218 -0.406 -0.247  0.046  0.735 -0.143  0.062 -0.255 -0.035
fewer       -0.328 -0.335  0.253  0.286 -0.307  0.116 -0.200 -0.636  0.254 -0.130
changing    -0.269  0.249  0.483 -0.301  0.057  0.237  0.667 -0.168 -0.093  0.008
lifestyles  -0.285  0.507 -0.106  0.317  0.171  0.014 -0.199 -0.242 -0.194  0.620
standards   -0.288  0.262  0.382 -0.488 -0.213 -0.099 -0.588  0.243  0.054 -0.047
family      -0.280  0.499 -0.177  0.366  0.089  0.112  0.038  0.179  0.282 -0.611
```

The calculated PCA row scores and estimated factor analysis scores are also somewhat different. Though the first factor/component scores are fairly highly correlated (0.95), the second factor/component scores are much more weakly correlated (0.53). With an orthogonal rotation, this correlation would be much higher. However, we have theoretical reason to expect correlated factors. In this case, rotation could make a substantive difference when it comes to inferences using estimated factor scores.

```
> facscores <- factor.scores(values, promax.factors, method = "Thurstone")$scores
> pcascores <- pcafit$scores[,1:2]
>
> cor(facscores, pcascores)
        Comp.1     Comp.2
PA1 -0.9477263 -0.3134376
PA2 -0.8313804  0.5290413
```

IMPORTANT NOTE: A three factor solution probably fits the data best, and that's likely what I'd do if I had no particularly strong theoretical prior. The root mean squared residual drops from 0.07 in the two factor solution to 0.02 in the three factor solution. And, I could just as easily interpret the "elbow" in the scree plot at the fourth factor (meaning I should retain 3 factors) as I could at the second third factor (meaning I should retain 2 factors). Even a unidimensional, single factor solution seems reasonable, to some degree. I've uses these exact same items in a summated rating scale. Admittedly, a SRS is not a good test of dimensionality, but the items to perform flawlessly in an item analysis (i.e., nearly linear, and always monotonic, estimated item response functions; high Cronbach's alpha estimate of reliability). DIMENSIONALITY IS INHERENTLY SUBJECTIVE AND THEORY IS KING.