Exploratory Factor Analysis

The data for this EFA exercise consist of correlations among 9 'ability' variables collected by Holzinger & Swineford in one Chicago junior high school in 1939. The 9 variables consist of 3 perceptual tests, 3 verbal tests and 3 counting/math tests, so one might expect a 3-factor solution would be sufficient to account for the correlations, with 3 non-overlapping, orthogonal factors.

The data set is unusual, in that it is entered in correlation form, as the lower half of the correlation matrix, supplemented by other observations that give the means, std. deviations and other information. This form is typical when data is re-analyzed from a published study, rather than from raw data.

1. Read the data, **`psych9.sas`** into SAS. (It is in **N: \psy6140\data**).

   ```
   %include data(psych9);
   proc print data=psych9; run;
   ```

   - What was the sample size in this study?
   - What were the mean, standard deviation and reliability[1] for variable X1?

2. Carry out a simple **principal factor analysis** of these data. The option `priors=SMC` says to use the squared multiple correlations as communality estimates. This just does a principal components analysis of the correlation matrix, but with diagonal entries replaced by $R^2_{i|others}$.

   ```
   proc factor data=Psych9 method=PRINCIPAL priors=SMC
        Scree;
   run;
   ```

   - How many factors were extracted? By what criterion?
   - What does the scree plot suggest?
   - You can set another criterion, or ask for a given # of factors, but we'll do that in the next step.

3. **Iterated principal factor** analysis is a better method for exploratory factor analysis. Ask for `NFact=3` and a varimax rotation. The `round` option makes the loadings easier to read, but don't be mislead to think that '`*`' means significant.

   ```
   proc factor data=Psych9 method=PRINIT
        Nfact=3 rotate=varimax round;
   run;
   ```

[1] The reliabilities are not ordinarily used in PROC FACTOR. How they can be used is discussed in the lecture.

4. *Time to quit fooling around*. We'll test the hypothesis that k=2 factors are sufficient with a **maximum likelihood** analysis. Here, we just look at the test of: `H0: 2 Factors are sufficient`

```
proc factor data=Psych9 method=ML Nfact=2;
run;
```

5. Try this again, this time with k=3 factors.

```
proc factor data=Psych9 method=ML Nfact=3
      round rotate=varimax;
run;
```

   - What does the test of k=3 factors are sufficient imply here? You can also compare the k=2 model with the k=3 model using the AIC (smaller=better) and the Tucker-Lewis reliability (larger=better).
   - How do the rotated loadings relate to the idea of 3 independent (uncorrelated) factors for these variables?

6. In EFA, there are many different rotation methods: varimax, promax, oblimax, etc, each of which optimizes some criterion of 'simple structure'. When you have a specific factor structure in mind to test, a better idea is often a '**Procustes rotation**' --- rotating the loadings to conform as closely as possible to the hypothesized structure. To do this, you first create a 'target' data set, containing the (transposed) target loadings, as a matrix of 1s (target loadings) and 0s (non-target).

```
data mytarget;
      input _name_ $ X1 X2 X4 X6 X7 X9 X10 X12 X13;
datalines;
FACTOR1  1  1  1  0  0  0  0  0  0
FACTOR2  0  0  0  1  1  1  0  0  0
FACTOR3  0  0  0  0  0  0  1  1  1
;
proc factor data=Psych9 method=ml NFact=3 round
      rotate=procrustes target=mytarget
      plot;
      run;
```

   - Try to interpret the results from the Rotated Factor Pattern (factor loadings, regression coefficients) and the Factor Structure (correlations of variables with factors).
   - This analysis allows the factors to be correlated, to achieve the best fit to the unrotated loadings. Try to interpret the Inter-Factor Correlations also in this context

If we were serious about testing this "theory of abilities," we would more likely do **confirmatory factor analysis** (CFA), using PROC CALIS (or Amos, Lisrel, EQS, etc.). More on this later.

EFA in R

The same correlation matrix is available in R in the `psych` package as Holzinger.9. Factor analysis using the MLE is provided by the factanal() function. Note that with a correlation matrix as input, it is necessary to supply the number of observations as the `n.obs` argument.

The following small script is available in **N:\psy6140\tutorials\psych9.R**.

```
# Holzinger-Swineford 9 ability variables
library(psych)
data(bifactor)
Holzinger.9

# ML factor analysis, with varimax rotation (by default)
factanal(covmat=Holzinger.9, factors=2, n.obs=145)
factanal(covmat=Holzinger.9, factors=3, n.obs=145)
factanal(covmat=Holzinger.9, factors=3, n.obs=145, rotation="promax")

# plot rotated loadings
loadings <- factanal(covmat=Holzinger.9, factors=3, n.obs=145)$loadings
plot(loadings, pch=16, main="Holzinger-Swineford, k=3 factor solution")
text(loadings[,1], loadings[,2], rownames(loadings), pos=1)
abline(h=0, v=0, col="gray")

plot(loadings[,c(1,3)], pch=16, main="Holzinger-Swineford, k=3 factor
solution")
text(loadings[,1], loadings[,3], rownames(loadings), pos=1)
abline(h=0, v=0, col="gray")
```

It is sometimes helpful to view a table of factor loadings as a table plot, which shows the size of the loadings as circles, together with the numerical values:

```
if(!require(tableplot)) install.packages(tableplot); library(tableplot)

tableplot(t(round(100*loadings)),
  cell.specs = list(list(cell.fill="yellow", back.fill="gray90",
         scale.max=100, label=1)), side.rot=90)
```