STAT 530/J530
October 6th, 2005

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Homework 5

A classic data set collected by Holzinger and Swineford in 1939 is often referenced as being found in Harman's 1976 text. The original data set consisted of 24 psychological tests given to 145 seventh and eighth grade students.

a) Determine the number of factors suggested by Kaiser's criterion and the scree plot, and also the maximum number according to the degrees of freedom formula.

b) Using the iterated principal factor method, choose a number of factors that you think does a good job of balancing a good fitting model and a model that should be simple enough to work with. Justify your choice

c) Also fit the principal components factor analysis model for the number of factors you chose in b. Compare the fit of the two methods (iterated and pca).

d) For your first factor only, indicate which loadings you think are definitely practically significant and definitely not practically significant. Also indicate which loadings are likely to be statistically significant.

Drug-Use Items Continued

In order to have output worth looking at the resulting residual plots should look ok.

In order to be useful, the model should have factors with enough strong loadings to be interpretable.

This can often be aided by using rotations.

Estimating Them

In addition to using the loadings to getting a feel for the underlying factor structure, it is often desired to assign each observation predicted factor scores.

Getting the F Values

Given the model

$X_1=\lambda_{11}F_1+\lambda_{12}F_2+\cdots+\lambda_{1k}F_k+u_1$

$X_2=\lambda_{21}F_1+\lambda_{22}F_2+\cdots+\lambda_{2k}F_k+u_2$

$\vdots$

$X_q=\lambda_{q1}F_1+\lambda_{q2}F_2+\cdots+\lambda_{qk}F_k+u_q$

We can use either regression or maximum likelihood methods to estimate the factor values for each subject.

Problems

The formulas good properties are based on assuming a multivariate normal model holds.

The values depend a great deal on the particular sample being used.

Alternatives

Pick a representative item

Use the sum score

Last Things…

Normal Representation

Path Diagrams

Confirmatory Analysis