

Homework 06
Psychology 312

1. (50 points). *Exploratory and Confirmatory Factor Analysis*. An article by Long and Perkins(2003) is online in the Statistics Handouts section. Professor Perkins graciously provided data from that study. Long and Perkins reviewed a previously published measure that presented 12 items purportedly representing 4 constructs. Long and Perkins performed two confirmatory factor analyses that they reported in their paper. One is a single factor model, and one is a 4 factor model with each factor loading on 3 variables. Here are descriptions of the 12 items:

I am going to read some things that people might say about their block. Each time I read one of these statements, please tell me if it is mostly true or mostly false about your block simply by saying "true" (2=MORE SOC) or "false" (1=LESS SOC).

- SCI1 I think my block is a good place for me to live.
- SCI2 People on this block do not share the same values. (reverse)
- SCI3 My neighbors and I want the same things from the block.
- SCI4 I can recognize most of the people who live on my block.
- SCI5 I feel at home on this block.
- SCI6 Very few of my neighbors know me. (reverse)
- SCI7 I care about what my neighbors think of my actions.
- SCI8 I have almost no influence over what this block is like. (reverse)
- SCI9 If there is a problem on this block people who live here can get it solved.
- SCI10 It is very important to me to live on this particular block.
- SCI11 People on this block generally don't get along with each other. (reverse)
- SCI12 I expect to live on this block for a long time.

NOTE: Items 2,6,8,11 are reverse coded! This means that the loading on this item may be negative on the designated factor.

Data for these items are available in the file *LongPerkinsTime1SCI.csv*.

Here is what I want you to do:

- (a) Using the `QuickCFA` function, replicate the two (one factor and 4 factor) confirmatory factor analyses of Long and Perkins, which they report in Table 2 of their article. You can load the data from the Internet and you should remove missing values, as below:

```
> # create new dataset without missing data
> data <- read.csv(
+   'http://www.statpower.net/R312/LongPerkinsTime1SCI.csv')
> ### remove missing values
> data <- na.omit(data)
```

- (b) Load the `AdvancedFactorFunctions` library. Compute the correlation matrix.

```
> source(  
+ 'http://www.statpower.net/R312/AdvancedFactorFunctionsV1.11.R')  
> R <- cor(data)
```

Use the `FA.Stats` and `Scree.Plot` functions to perform exploratory factor analyses on the correlation matrix, and report overall statistics for 1-5 factors. Examine the output and state what number of factors would be dictated by (1) the Scree test, (2) the Kaiser-Guttman rule, (3) the old-fashioned "sequential chi-square test," and (4) the RMSEA confidence interval plot. If there is ambiguity, don't be afraid to say so.

- (c) We're going to try to find a good model with only 3 factors. Using the exploratory-confirmatory approach of Joreskog (1978) via the `QuickJoreskog` function, produce a good-fitting model with 3 factors for the same 12 SCI variables, and provide your name and description for the constructs that you think each of these three factors are capturing. There is a problem with this solution, and you will find it - if you look carefully. Remember in class we talked about a "Heywood Case"? (Hint: A variance can never be negative.) Try a 4 factor orthogonal and oblique solution. Try a bifactor model with 4 factors. See what you can find! On substantive and statistical grounds, try to decide which model is your favorite.
2. (20 points) A manipulated sample of 20 voices were played to 30 judges, who rated them on 15 adjectives on a 14 point Likert scale. Averaging over the judges produced a sample with $n = 20$ and $p = 30$. The data are available online as *voices.csv*.

```
> voices.data <- read.csv("http://www.statpower.net/R312/voices.csv")
```

- (a) Examine the heuristic and statistical criteria, choose a number of factors.
- (b) Pick a rotation that reveals a structure you find meaningful.
- (c) Name the factors.
3.] (10 points) At the end of Chapter 13 of his book, Rencher attempts to characterize the distinction between principal components and common factors. Here is a quote from that discussion:

Additional differences are that (3) principal component analysis requires essentially no assumptions, whereas factor analysis makes several key assumptions; (4) the principal components are unique (assuming distinct eigenvalues of S), whereas the factors are subject to an arbitrary rotation; and (5) if we change the number of factors, the (estimated) factors change. This does

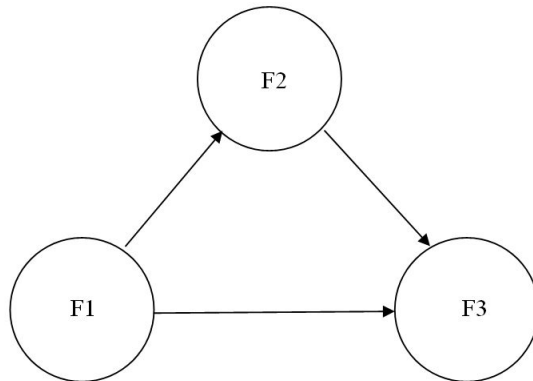
not happen in principal components. The ability to rotate to improve interpretability is one of the advantages of factor analysis over principal components. If finding and describing some underlying factors is the goal, factor analysis may prove more useful than principal components; we would prefer factor analysis if the factor model fits the data well and we like the interpretation of the rotated factors. On the other hand, if we wish to define a smaller number of variables for input into another analysis, we would ordinarily prefer principal components, although this can sometimes be accomplished with factor scores.

Rencher seems to believe that principal components cannot be rotated! In what (very limited) sense is he right, and in what (much more significant) sense is he completely wrong? (Hints: Suppose you have the first two principal components. If you rotated them, are the individual variables still principal components? Is the vectorspace spanned by the two components changed from before they were rotated? Does the amount of variability in the observed variables explained by the two components change after you rotate them?)

4. (20 points). Suppose you obtain 145 observations on 9 independent variables. Factor F_1 loads on Y_1, Y_2, Y_3 . Factor F_2 loads on $Y_4 - Y_6$, and factor F_3 loads on $Y_7 - Y_9$. Your data are online in a file *9vRawData.csv*.

```
> data <- read.csv("http://www.statpower.net/R312/9vRawData.csv")
```

You believe that the construct represented by factor F_1 affects F_3 , but that part of this effect is *mediated* through F_2 . In other words, the structural equations can be represented by the structural model shown in the figure below.



- (a) Use the `QuickSEM` function to fit the structural equation model described above. You can compute the covariance matrix of the raw data using the `cov()` command. Examining the output, I want you to address two primary issues.

- (b) Using conventional fit indices such as the RMSEA or the Chi-square test of perfect fit, how would you describe the fit of this model?
- (c) Examine the path coefficients for the structural part of the model. Is there significant mediation in this situation?