CONCEPTS OF STATISTICS II
ASSIGNMENT #3
SETH HOWELLS
07/26/2020

Factor Analysis Overview

Factor analysis has the primary purpose of identifying underlying structures among the variables. Factors, sets of variables that are highly intercorrelated, are latent variables that represent the common underlying dimensions. Since a dataset can contain a large number of variables under investigation, it is implausible for one to identify interrelated patterns by analyzing the relationship one by one. Instead, factor analysis provides data summarization and data reduction which allows the researcher to identify these patterns parsimoniously.

Seven Stages of Applying Factor Analysis

1. OBJECTIVES OF A FACTOR ANALYSIS

- The two types of Factor Analysis are Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA and CFA are very similar in their respective techniques but are very different in their application. Some similarities and differences are listed below:

Differences

- EFA does not have a priori fixed number of factors, unlike CFA.
- CFA assumes the researcher enters the factor analysis with a strong understanding about the factors and which variables are likely to be loaded onto each factor.
- EFA assumes the researcher enters the factor analysis with an aim to explore the interrelationships among the variables.

2. DESIGNING A FACTOR ANALYSIS
    a. R vs Q
    - **R-type** factor analysis: factors are calculated from the correlation matrix
    - **Q-type** factor analysis: factors are calculated from the individual respondent

    b. Variable Selection
    - Metric: easy to compute interrelations among variables under investigation
    - Nonmetric: need to use dummy variables (Boolean)
    - Reduce data, but keep at least 5 variables for each proposed factor in order to assess a proposed structure

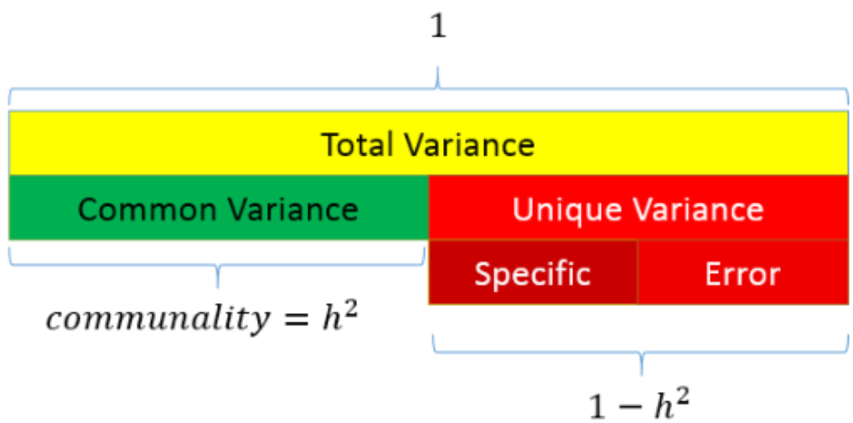
    c. Sample Size:
    - Minimum sample should be 100 or larger for proper analysis
    - General rule of thumb is to have at least 5 times as many cases as the number of variables under investigation
3. ASSUMPTIONS IN FACTOR ANALYSIS
    a. The design of Factor Analysis is more conceptual than statistical because the inferences drawn from the output are sensitive to the decision in the design.

- Basic assumption: no underlying variable exists. If correlation occurs, it does not guarantee to be relevant.
- Example: If the researcher is analyzing variables among males and females, but the structure is naturally geared toward one gender or the other, factor analysis would be inappropriate despite showing high correlation for the proposed, favored gender.

b. Rule of Thumb:

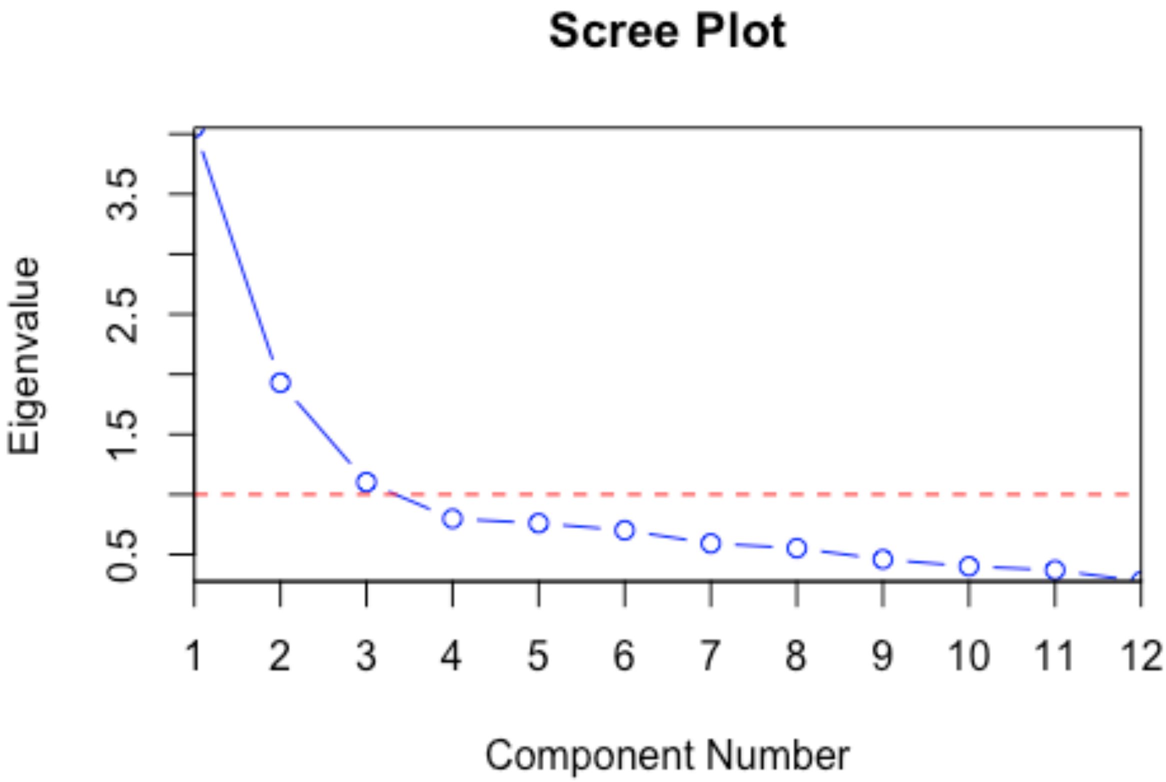
- Strong conceptual understanding to support the basic assumption.
- Bartlett's test of sphericity: significance level less than .05 indicates correlation exists.
- Measure of Sampling Adequacy (MSA): values greater than .50 for the overall test and each individual variable. Values less than .50 should be omitted

4. DERIVING FACTORS AND ASSESSING OVERALL FIT
    a. Component Analysis: Considers the total variance
    b. Common Factor Analysis: Considers the shared variance

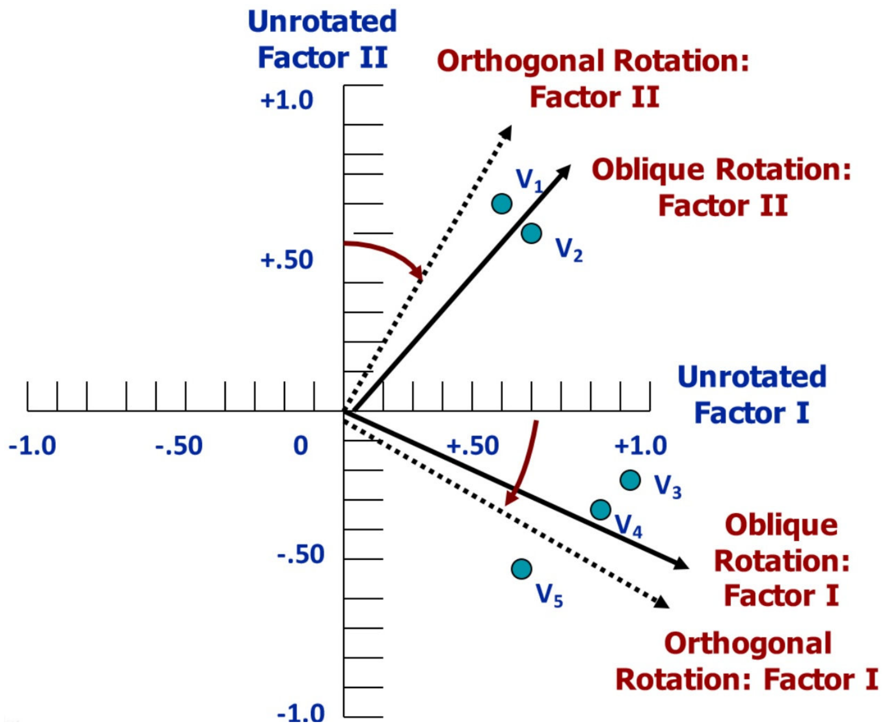


c. Scree Plot

- Displays eigenvalues (y-axis) to the components (x-axis) in a downward curve. The line at y = 1.0 suggests the number of factors to be generated in the analysis (3 factors per scree plot below)



5. INTERPRETING THE FACTORS
    a. Factor Rotation
    b. Orthogonal: axes maintained at a 90 degree angle. This is the most commonly used.
        i. **QUARTIMAX**: rows simplified in order for each variable to be loaded on a single factor
        ii. **VARIMAX**: columns simplified in order for the factors to be clearly associated and separated among variables
        iii. **EQUIMAX**: combination of QUARTIMAX and VARIMAX to simplify rows and columns simultaneously



c. Oblique: axes not maintained at a 90 degree angle. This is best suited for the goal of obtaining meaningful (theoretical) constructs since the "real world" tends to have few uncorrelated variables.

6. VALIDATION OF FACTOR ANALYSIS
   a. Factor Analysis at its core is primarily concerned with the best representation of all variables under investigation through the means of a simplified data reduction and summarization technique. In return, the researcher must analyze the generalizability of the results. The most direct method toward validating a factor analysis is move to a confirmatory perspective.
7. ADDITIONAL USES OF THE FACTOR ANALYSIS RESULTS
   a. Surrogate Variable
   - Objective: identify appropriate variables for application with other statistical techniques.
   - Direct and easy approach when one factor is significantly higher than the rest.
   b. Summated Scales
   - Comprised of several individual variables into one composite measure.
   - The average of variable loading highly on a factor is used as a replacement variable