Contents

1 Principal Compoment Analysis

1.1 Orthogonal versus Oblique Solutions

- This course will discuss only principal component analysis that result in **orthogonal solutions**. An orthogonal solution is one in which the components remain uncorrelated (orthogonal means uncorrelated).
- It is possible to perform a principal component analysis that results in correlated components. Such a solution is called an **oblique solution**. In some situations, oblique solutions are superior to orthogonal solutions because they produce cleaner, more easily-interpreted results.
- However, oblique solutions are also somewhat more complicated to interpret, compared to orthogonal solutions. For this reason, we will focus only on the interpretation of orthogonal solutions

1.2 Sampling adequacy (KMO Statistic)

- Measured by the Kaiser-Meyer-Olkin (KMO) statistics, sampling adequacy predicts if the analyses are likely to perform well, based on correlation and partial correlation. KMO can also be used to assess which variables to drop from the model because they are too multi-collinear.
- There is a KMO statistic for each individual variable, and their sum is the KMO overall statistic. KMO varies from 0 to 1.0 and KMO overall should be 0.60 or higher to proceed with factor analysis. Values below 0.5 imply that factor analysis or PCA may not be appropriate. (Approach to overcoming this: If it is not, drop the **indicator variables** with the lowest individual KMO statistic values, until KMO overall rises above 0.60.)
- Kaiser-Meyer-OlkinTo compute KMO overall, the numerator is the sum of squared correlations of all variables in the analysis (except the 1.0 self-correlations of variables with themselves, of course).
- The denominator is this same sum plus the sum of squared partial correlations of each variable i with each variable j, controlling for others in the analysis. The concept is that the partial correlations should not be very large if one is to expect distinct factors to emerge from factor analysis.

1.3 Bartlett's Test for Sphericity

- Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. For PCA and factor analysis to work we need some relationships between variables and if the correlation matrix were an identity matrix then all correlation coefficients would be zero. Therefore, we want this test to be signifcant (i.e. have a significance value less than 0.05).
- A significant test tells us that the correlation matrix is not an identity matrix; therefore, there are some relationships between the variables we hope to include in the analysis.
- For these data, Bartlett's test is highly significant ($p < 0.001$), and therefore factor analysis is appropriate.

2 Review of Important Definitions

- An observed variable can be measured directly, is sometimes called a measured variable or an indicator or a manifest variable.
- A principal component is a linear combination of weighted observed variables. Principal components are uncorrelated and orthogonal.
- A latent construct can be measured indirectly by determining its influence to responses on measured variables. A latent construct could is also referred to as a factor, underlying construct, or unobserved variable.
- Factor scores are estimates of underlying latent constructs.
- Unique factors refer to unreliability due to measurement error and variation in the data.
- Principal component analysis minimizes the sum of the squared perpendicular distances to the axis of the principal component while least squares regression minimizes the sum of the squared distances perpendicular to the x axis (not perpendicular to the fitted line).
- Principal component scores are actual scores.
- Eigenvectors are the weights in a linear transformation when computing principal component scores. Eigenvalues indicate the amount of variance explained by each principal component or each factor.
- Orthogonal means at a 90 degree angle, perpendicular. Obilque means other than a 90 degree angle.
- An observed variable ***loads*** on a factors if it is highly correlated with the factor, has an eigenvector of greater magnitude on that factor.
- Communality is the variance in observed variables accounted for by a common factors. Communality is more relevant to EFA than PCA.