

1 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-Olkin To 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.

1.1 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 R- matrix were an identity matrix then all correlation coefficients would be zero. Therefore, we want this test to be significant (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.