Cornell University
Cornell Statistical Consulting Unit

StatNews #49

Principal Components: Not Just Another Factor Analysis
February 2002

When a large number of variables are correlated, some information in the set of variables is redundant. Furthermore, including many correlated variables in a regression model can make interpreting results difficult, and, in extreme cases, lead to multicollinearity. Principal components analysis is a linear transformation technique that provides a smaller set of uncorrelated variables (called components) from a set of correlated variables while maintaining most of the information in the original data set. These components can then be used in place of the original variables in the regression model.

A principle component obtained from four observed variables, for example, can be presented mathematically as:

C = W1*X1 + W2*X2 + W3*X3 + W4*X4

where C is the value of the principal component, X1 to X4 are the observed variables, and W1 to W4 are weights.

The principal components are formed in decreasing order of importance, which means: 1) the first principal component accounts for as much of the variation in the observed variables as possible, 2) each succeeding component accounts for as much of the remaining variation as is possible, and 3) all components are uncorrelated with each other.

Principal components analysis and factor analysis are similar to each other in two ways. First, both are data reduction techniques that create a new smaller and manageable set of variables (principal components for one and factors for the other). Second, one of the extraction methods (i.e., principal axis method) used for factors in factor analysis is actually the same as the one used for principal components.

But there are also important differences between principal components and factor analysis. Conceptually, factor analysis requires a meaningful model that assumes that the correlation among variables is due to the fact that they are a manifestation of one or more common underlying factors (see StatNews #48: What is factor analysis?). In contrast, principal components analysis makes no assumption about a model; it is simply a mathematical linear transformation of the original variables. Furthermore, because factor analysis is a model similar to regression, we expect that, for each observed variable, some of the variability will be explained by the model and some will not. In contrast, in principal components analysis, all variability in the original variables will be explained by the components.

In summary, when you just want to reduce the number of observed variables to a smaller set of uncorrelated variables, then you would use principal components analysis. But, if you want to model some meaningful underlying constructs in your variables, you would use factor analysis.

Author: Yun Wang

Back to StatNews Table of Contents

(This newsletter was distributed by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at cscu@cornell.edu. Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at http://www.cscu.cornell.edu).