StatNews #48

What is Factor Analysis?
December 2001

Factor analysis is a statistical model that assumes that a small number of unobserved (i.e., latent) constructs are responsible for the correlation among a large number of observed variables. The latent constructs, for example, academic ability, cannot be directly observed, but they affect observable variables, such as French, English, and Mathematics scores. Specifically, factor analysis assumes that the variance of each observed variable comes from two parts: a common part shared with other variables that cause correlation among them, and a unique part that is different from other variables. The common parts are called factors, and these factors represent the latent constructs.

We usually use factor analysis when we want to do one or more of the following:

1. determine whether a smaller set of factors exists that will explain the relationships among the observed variables,
2. determine a smaller set of uncorrected factors,
3. determine the number of underlying factors,
4. interpret these new factors,
5. evaluate individuals or experimental units in the data set on these new factors, and/or
6. use these new factors in other statistical analyses of the data.

For example, we have a data set about the test scores of a group of students on different subjects. These subjects are: French, English, Math, Physics, Music. Each student's performance in each of these subjects is due in part to overall academic ability that applies to all the subjects, and in part to the student's unique ability in each subject. The measured variables here could be modeled by following equations:

$$\text{French} = b_1 * F + u_1$$
$$\text{English} = b_2 * F + u_2$$
$$\text{Math} = b_3 * F + u_3$$
$$\text{Physics} = b_4 * F + u_4$$
$$\text{Music} = b_5 * F + u_5$$

Here F is the common part to all variables, and u_1, u_2, u_3, u_4, u_5 are the unique parts. The unique parts are independent of each other and independent of F. The coefficients b_1, b_2, b_3, b_4, b_5 represent the strength of the relations between F and the observed variables. We can interpret F as overall academic ability that influences students performances on all subjects, while the u's can be interpreted as the special ability relating to a specific subject. The model asserts that the common factor F among these performance variables is the reason that they are correlated with each other.

On the other hand, each variable could also be modeled in another way:

French = b1*F1 + u1
English = b2*F1 + u2
Math = b3*F2 + u3
Physics = b4*F2 + u4
Music = b5*F1 + u5

where we can interpret F1 as artistic ability and F2 as scientific ability. F1 and F2 are the common factors.

How to interpret the underlying factors and how many there are depend on different understandings and purposes of the research. Most of the time, a researcher who wants to use factor analysis has some beliefs about what the latent constructs conceptually may be. Then, factor analysis is used to test these beliefs. For this situation, factor analysis is used as a means to confirm theoretical and/or conceptual beliefs. This is called confirmatory factor analysis, and it is commonly used as a theory-testing procedure. In other situations, a researcher may have little or no idea about what are the latent constructs and how many there are. For this situation, factor analysis is used to explore the structure of the latent constructs by identifying factors that may represent them. This is called exploratory factor analysis, and it is commonly used as a theory-generating procedure.

To learn more about factor analysis, please read the following references:

1. Darlington, R., Factor Analysis: http://www.psych.cornell.edu/Darlington/factor.htm
2. Kim, J. & Mueller, C. W.: Introduction to Factor Analysis: What It Is and How To Do It. (Sage)
3. Kim, J. & Mueller, C. W.: Factor Analysis: Statistical Methods and Practical Issues. (Sage)

And if you need further help in conducting a factor analysis, please contact the Office of Statistical Consulting for appointments.

Author: Yun Wang

Back to StatNews Table of Contents