Clinical Psychology: Making Decisions about Best Therapy for a Client: Using Data Mining to Explore the Structure of a Depression Instrument

David P. Armentrout, Ph.D. and Linda A. Miner, Ph.D.

The original data set for this example had 359 cases. The intent of the instrument was to measure various components of depression as an aid to practitioners as they organize the therapy of a client. It was important that the instrument did, in fact, measure depression and that it did so reliably. There were 164 questions in the original survey.

To determine the structure, we followed these procedures. A factor analysis first grouped the questions into meaningful groups. We thought that people suffering from depression would appreciate a smaller survey rather than a longer one. Feature selection indicated which of the individual questions were most important to each factor grouping. We reduced the number of questions in the survey by eliminating the questions that contributed the least to each factor. By finding the questions that most associated with their meaningful groups, we shortened the survey while retaining valuable information. The resulting data set produced the one that we used for this tutorial, which illustrates part of our procedures. We wanted to focus on the structure of the relationships of variables, not predicting the amount of depression. In fact, in predicting who was depressed and who was not, we found that simply asking the question "Are you depressed, yes