STAT 8801 – Case Study #2

February 1st, 2012

Group #3: Xiaoyi, Chris, Wen, Jenny, Felipe

In this case, the client wants to combine six scores from a questionnaire into one score, and use it as an overall indicator of motivation for volunteering.

Concerns and questions before doing combination:

First of all, we will ask the client why she needs to combine six scores into one? Is this really necessary? Since no matter what method we use to combine, we will lose some information in this process; and on the other hand, six scores is not too many for us to handle, so if there is not a good reason, we will suggest her not to do this combination.

If we've decided this combination is necessary, we have several additional questions for her:

1. What are the contents of her questionnaire, and how did she sample the persons in her questionnaire? This is very important since if the questions or the sampling is not proper, we cannot get a good result.
2. What are the correlations between these six scores? It is not reasonable to assume these scores are independent with each other, and correlation is very important information to determine how to combine these six scores.
3. Can we categorize these six scores? If there are some scores which are highly correlated with each other, we can put these scores into one category, and combine the scores in each category; then combine the category scores into one. Using this method we can decrease the within group variance.

Methods we suggest to use:

There are two methods we can use: principal component analysis, and factor analysis. Both of these techniques are used to analyze groups of correlated variables representing one or more common domains, which can be applied in this case. But there is difference between these two techniques: principal components analysis is used to find best ways of combining variables into a small number of subsets, while factor analysis may be used to identify the structure underlying such variables and to estimate scores to measure latent factors. Another difference between the two approaches is about the variance: in PCA, all of the observed variance is analyzed, while in factor analysis it is only the shared variances that are analyzed.

In this case, we will suggest the client use factor analysis. There are two reasons: first, we want to get the underlying models of the latent variables which cause the observed variables; second,

the factors in factor analysis can be conceptualized as "real world" phenomena, while the PCA is simply reducing the dimension of the data, which cannot easily be explained using real world phenomena.