Simplify Your Statistics

One of the most common and costly mistakes made by graduate students (or their committees) is the development of statistical plans that are more complex than necessary to address the questions of interest. This mistake is often made because of the erroneous notion that more complicated techniques offer better results, and because committees often use the thesis or dissertation research to test their students' statistical abilities. Unfortunately, the employment of statistical analyses that are more complex than necessary can (a) needlessly extend the length of time necessary for the completion of the project, (b) confuse both students and the readers, and (c) even stop some students from ever completing the project. Many of those who are ABD (all-but-dissertation) are really ABS (all-but-statistics). Additionally, any attempt by advisors and committees to use the dissertation or thesis as a test of students' statistical abilities should be guided by the *appropriate* use of statistics, not the complexity of the methods selected.

The basic principle that should be followed is to *implement the minimally sufficient analysis*. Begin with your research questions and try to find the simplest analysis that will answer those questions. More complex analyses that provide answers to questions beyond the scope of your study only serve to cloud your results and increase the likelihood of getting bogged down. For example, it is true that a structural equation modeling framework can be used to perform an independent samples t-test. The conclusion from both approaches would be exactly the same. However, the added time required for data preparation and programming necessary to implement the structural equation model test is clearly not justified if the only goal is to determine if two groups differ on an interval or ratio level variable. This is not to say that some research questions do not require complicated statistical techniques, but it is still desirous to employ simpler techniques when they are adequate. An interesting side note is that nonparametric analysis are often simpler to perform and interpret that parametric analyses, and, depending on the type of data one is dealing with, have the further advantage that they can be more appropriate.

A related consideration in simplifying your analyses is to *use appropriate statistical software*. Matching the proposed analysis to a particular software package can be time consuming for those without experience, but the reward can be a drastic reduction in time spent on programming and analysis. Even some spreadsheet programs, such as Microsoft EXCEL (using the “Data Analysis” option), provide a surprising variety of statistical tests, and many students are already familiar with spreadsheet programs. New versions of popular statistical packages often allow analyses to be conducted without learning a new computer language using pull-down menus and dialog boxes (SPSS is probably the easiest statistical package to use). New versions of structural equation modeling programs provide a graphical interface that also greatly reduces programming time. For structural equation modeling, AMOS provides the simplest method for performing most analyses.

A final point relates to the quality of the data. Often, the quality and/or quantity of data necessitates that a simpler analytic approach be taken. The development in recent years of more sophisticated analytic techniques such as hierarchical linear modeling, mixture modeling, and structural equation modeling are seductive in their elegance but often require data of an extremely high quality. Although we often ignore warnings that certain statistical tests require “large samples,” or that we are being presented with an “asymptotic significance level,” they can imply the need for thousands of subjects and data adhering closely to multivariate normality. Some theses and dissertations meet these data requirements, but most do not (in this sense they are no different than other research projects), and the application of the advanced techniques in such circumstances is counterproductive. In general, simpler statistical techniques make fewer assumptions about the data and are more robust to violations of their assumptions.

The advantages of relying on the principal of the minimally sufficient analysis are that (a) it requires a shorter, more straightforward Results section, (b) it keeps the analytic focus on your research questions rather than the large number of other questions that can be addressed with more complicated techniques, and (c) it increases the probability that the conclusions from the study will directly match the questions posed in the introduction. Adhering to the principle is fairly simple once the decision has been made. The first step is the explicit statement of the research questions of interest. Second, each of the research questions is converted into a statistical hypothesis (see the chapter on this topic). Third, the simplest techniques that can be employed to test those statistical hypotheses are selected. Fourth, the tests are conducted and the results as they relate to each of the research questions are stated. It seems simple enough, but many students go astray, or are led there, by not following this principle.

In conclusion, there is no correspondence between the complexity of one's statistical analysis and the quality of the thesis or dissertation. *The important issue is the appropriateness of the analytic strategy with respect to the nature of the data, the variables to be employed, and the questions one is asking*. The chapter on "Choosing an Appropriate Statistical Test" will be of help in this regard also. Adhering to the principle of the minimally sufficient analysis will assist you in answering your research questions in the most straightforward manner possible.

John C. Caruso

Director of Services, Dissertation Statistical Services

www.diss-stat.com