OPINION
Chiang Mai J. Sci. 2007; 34(1) : first page - last page
www.science.cmu.ac.th/journal-science/josci.html

Musings of a Statistical Consultant

Timothy E. O'Brien

Those of us who work in applied statistics often have the good fortune to collaborate with researchers and decision-makers in fields as diverse as agriculture, biochemistry, commerce, education, medicine, sociology, and so on. Thus, in addition to our usual university teaching and research activities, we serve as statistical consultants, and indeed many universities (mirroring our commercial counterparts) have established statistical consulting centers to provide the necessary resources for these collaborations.

Statistical methods such as hypothesis testing are intrinsic to the scientific process in which conjectures are made and assessed juxtaposed with actual evidence (data) and so it is not surprising to witness the paramount role of the application of appropriate statistical methods to decision-making. Consider, for example, many of the important discoveries over the past several decades in the field of Genetics, in which learning and understanding has resulted via many iterations of conjectures and hypothesis testing, followed by refined conjectures and assessment, thereby resulting in the current "laws" or dogma.

Possibly no other field, however, has witnessed the mis-application of its techniques more than the field of applied statistics. This may well be the result of the plethora of easy-to-use (sometimes called "point and click") statistical and software packages, whereby researchers are often easily provided with all-important p-values but with no explanation of the underlying assumptions and requirements for their data. To understand this current state of affairs, let's first look to the recent past.

While I was completing my Ph.D. in Statistics at North Carolina State University (NCSU) in Raleigh, North Carolina, USA (1987-93), I witnessed first-hand my professors actively engaging in statistical consulting. Also during this time, we as graduate students completed two semesters of a course on consulting methods; we were assigned to sit in on consulting sessions between statistics faculty and researchers and to report back to the larger class on our experiences.

Timothy E. O'Brien, Ph.D., is an Associate Professor of Statistics in the Department of Mathematics and Statistics, Loyola University, Chicago (USA). During the period 18 October 2006 – 28 February 2007, he served as a Fulbright Scholar in the Department of Statistics, Faculty of Science, Chiang Mai University. Dr. O'Brien also has an affiliate appointment in the Department of Biology at Loyola, and has extensive work and consulting experience with US and international research institutes and government agencies and in the pharmaceutical industry, including Amylin Pharmaceutics (San Diego, USA), Ciba-Geigy (now Novartis, Basel, Switzerland), Glaxo (now GSK, North Carolina, USA), and Janssen Pharma (Beerse, Belgium).

The format of the consulting sessions was often the same. First, the researcher(s) explained the underlying theory and conjecture(s) to be tested. Next, a study or experiment was designed and implemented. Third, the data was obtained and checked. Finally, the data were analyzed and a report of the findings was prepared. Many of these experiments designed at NCSU were quite involved, and so specialized software such as SAS® was needed to perform the data analysis, and this usually required the programming skills of a statistician. Thus, it was typically the case that the statistics professor or graduate student needed to perform the data analysis and to discuss the results with the researchers.

These days, however, even spreadsheet packages such as Microsoft Excel® can perform basic statistical analyses, implementing methods such as multiple linear regression, ANOVA, binary logistic regression, and so on. In addition, many universities around the globe use "point-and-click" statistical packages such as Minitab® and SPSS® for teaching purposes, and researchers thereby have a wider array of statistical methods at their fingertips, including MANOVA, principal components and factor analysis, and time series analysis, to name but a few. The potential danger, of course, associated with the use of these methods is encountered when certain key conditions or assumptions are not met, thereby resulting in erroneous conclusions. For example, when faced with the table below of count data comparing two drugs, many researchers would find no significant difference between the two treatments using the (incorrect) chi-square test (p-value = 14.2%) or Fisher's Exact test (p-value = 16.2%), whereas the more powerful (and more appropriate) CMH test (p-value = 3.9%) or proportional odds model (p-value = 2.4%) do indicate a significant difference between these drugs.

	Outcome – Pain Relief				
	None	Some	Substantial	Complete	Total
Drug A	3	7	10	5	25
Drug B	7	11	5	2	25

It is thus incumbent upon us as applied statisticians to gently enlighten our colleagues to be mindful of underlying assumptions and conditions, and to exercise a great deal of humility in the process. This sense of modesty is an important part of an effective consultant's character, and is easily realized when we bear in mind the amazing pace with which new and better statistical techniques are being developed – retraining and retooling is very necessary and important for applied statisticians to remain current in the latest improvements and software.

For their part, researchers would also be well advised to include an applied statistician as a collaborator or co-author so as to ensure the quality of their findings; indeed, most statistical consultants would welcome the opportunity and be grateful to be an active part of the scientific process.

OPINION is a forum for the expression of the individual views of the author. Publication here does not imply that these views are endorsed by the *Chiang Mai J. Sci.*