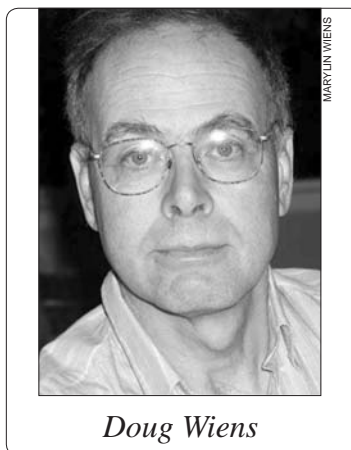## CJS: Coming Attractions

Douglas P. Wiens
Editor, CJS

The December 2006 issue of the ***Journal*** opens with 'Strictly monotone and smooth nonparametric regression for two or more variables', by Holger Dette and Regine Scheder (both Ruhr-Universität Bochum, Germany). Here, a monotone nonparametric estimate for a multiple regression function is proposed. The method starts with an unconstrained nonparametric regression estimate and makes successive refinements through one-dimensional isotonization procedures. Asymptotic properties are derived, and the methods are illustrated on two datasets.

Following this, Damião N. da Silva (Universidade Federal do Rio Grande do Norte, Brazil) and Jean Opsomer (Iowa State U.), in 'A kernel smoothing method of adjusting for unit nonresponse in sample surveys', propose a weighting adjustment for nonresponse based on estimating the response probabilities by kernel regression. A replication-based estimation method for the variance of the estimator is proposed, and illustrated by simulation.

Reinaldo B. Arellano-Valle (Pontificia Universidad Católica de Chile), Márcia D. Branco (Universidade São Paulo, Brazil) & Marc G. Genton (Texas A&M U.) propose a new definition of 'selection distribution', and as a consequence unify many existing definitions in the literature of multivariate skewed distributions. In 'A unified view on skewed distributions arising from selections', they describe motivating examples that involve various forms of selection distributions and which lead to skewed distributions, and provide a detailed description of the link between these classes of distributions.

There follow two papers on robustness of design. In 'Some robust design strategies for percentile estimation in binary response models', Stefanie Biedermann, Holger Dette (both Ruhr-Universität Bochum, Germany) and Andrey Pepelyshev (St. Petersburg State U., Russia) consider the problem of designing for efficient and

Doug Wiens

robust estimation of the percentiles of a dose response curve. They propose a maximin approach which yields locally optimal designs which are simultaneously efficient with respect to various parameter regions and link functions.

Zhide Fang (U. New Orleans), in 'Some robust designs for polynomial regression models', revisits the problem of designing for polynomial regression. The optimality criterion is a linear combination of the D-optimality criterion for the estimation of the low degree terms (included in the experimenter's model) and the Ds-optimality criterion for testing for higher degree (omitted) terms. Results on canonical moments and on continued fractions are utilized for this purpose.

Continuing the robustness theme, Debbie J. Dupuis (HEC Montréal) and Maria-Pia Victoria-Feser (U. Geneva) study estimation of the Pareto tail index using extreme order statistics. In 'A robust prediction error criterion for Pareto modelling of upper tails', they develop robust methods for the determination of the number of order statistics to be used, and for the estimation of the tail index.

The development of automated mapping and geographic information systems has resulted in an enormous amount of information stored in the form of spatial databases. Although there is a large literature

on modelling uncertainty associated with measurements at locations, the locations themselves are generally assumed to be correctly specified. In 'Modelling map positional error to infer true feature location', Jarrett J. Barber (U. Wyoming), Alan E. Gelfand (Duke U.) and John A. Silander Jr. (U. Connecticut) consider this problem of modelling the differences between locations as represented in a spatial database and the corresponding unobservable true locations.

In clinical and epidemiologic studies the event of interest is often recurrent. A common problem is that, even when the event times are known, their categories may be missing. In 'Multiple imputation methods for recurrent event data with missing event category', Douglas Schaubel (U. Michigan) and Jianwen Cai (U. North Carolina at Chapel Hill) propose methods for dealing with this. Asymptotic properties are derived, and the applicability of the methods in finite samples is examined through simulation.

A new method of testing for multivariate uniformity is studied by José R. Berrendero, Antonio Cuevas (both Universidad Autónoma de Madrid) and Franciso Vázquez-Grande (U. Chicago). In 'Testing multivariate uniformity: the distance-to-boundary method', they derive a test statistic based on the distances of the observations from the boundary of the sample space. The method is shown to be distribution-free in many cases, and seems to be largely resistant to the "curse of dimensionality". Finally, Radu Herbei and Marten Wegkamp (both Florida State U.), in 'Classification with reject option', study a method of binary classification that allows for a reject option, in which case no decision is made. The reject option is to be made for those observations for which the conditional (on certain feature variables) class probabilities are close, and as such are hard to classify. Starting in the next issue of Liaison, this Coming Attractions column will be writ-

ten by the incoming *CJS/RCS* Editor Paul Gustafson (U. British Columbia), to whom I gratefully turn over the reins at the end of the 2006 calendar year. Paul will outline his vision for the next three years in his Editorial in the March 2007 Journal; my final Editor's Report will also appear in that issue.

CONSULTANTS' FORUM

A Statistician Embedded in SAS

John Amrhein
SAS Institute (Canada) Inc., Toronto

Cary, NC, late 1980's

An infielder approached a cluster of 2 or 3 players from the opposing team and, seeing "SAS" emblazoned across the fronts of their uniforms, asked what the acronym meant. "I don't know" was the unanimous answer. "It's a software company" said one. "Has something to do with statistics, I think" added another. I found this exchange curious, to say the least. How could a group of SAS employees not know what their company's software is all about?

I did not work for SAS in those days, only having joined SAS Canada in 1999. But the situation is no different today. Worldwide, SAS employs more than 10,000 individuals. Most of these employees are in sales, marketing, finance, operations, etc. – areas that do not require a comprehensive understanding of statistical methods implemented in the software. Further, most components of the SAS system do not involve statistics or statistical methods, so even within Research and Development or Customer Support, many employees have no need for an understanding of statistics.

So, what role(s) does a statistical instructor/consultant fill when employed by the largest producer of statistical analysis software, and the vast majority of his colleagues do not know, nor need to know, statistics, or even issues surrounding data collection, storage and retrieval? It's a bit of an ironic situation. More so, my employer, SAS Canada, is a subsidiary whose function is primarily software sales. Most of its employees work in either a sales or sales support capacity.

John Amrhein

Focus on Customers

Focusing on the customer is the best way of dealing with the feeling of exclusion that comes from being a statistician employed by a sales organization. As a member of the Education Department within SAS Canada, my role is one of customer service. The services I provide fall into one of three categories.

- Instruction: SAS public courses are scripted lectures. Each delivery of a given course, *Categorical Data Analysis Using Logistic Regression* for example, is more or less the same. This predictability presents me, as the instructor, with the challenge of keeping the class interesting. Not only for the customers, who are hearing the material for the first time, but for myself, who is presenting it for the umpteenth time. My level of interest, whether low or high, is easily detected by the participants. My primary method for preventing my own loss of interest is to encourage questions from my audience. I have been teaching some courses for 7 years, and I still hear questions that have never before been asked. It is really rather amazing. The more difficult questions present an opportunity for personal growth because they force me to investigate the answer, which, in turn, increases my statistical or SAS programming knowledge. If I approach each class as a learning opportunity for myself, then I remain enthusiastic, and this projects to the audience.

With public instruction, customers must translate learned methods to their own situations. Also, course data rarely mimic a customer's own. Therefore, from a customer's viewpoint, although public courses are the least expensive service, they are also furthest removed from production. This sometimes leads to the customer's decision to pay for a customized course. The customer provides the data and the course is developed to teach the statistical methods needed to meet their research or business objectives. This requires an extensive period of course development. However, if the customer anticipates a repeated need for training, say annually, for multiple employees, then investing in the development effort the first time is worthwhile.

- Mentoring: This service is one on one, or one on two, customized instruction. It takes place at the customer's workplace and involves the use of their data. No course development period is required because this service does not involve lecture. It is two (or three) people working together to complete an analysis. These engagements are unscripted and unpredictable, which makes them