ST697M/ST797M: Measurement Error: Models, Methods and Applications.
Fall 2012

Instructor: John Buonaccorsi
Meets: MWF, 10:10 - 11:00
Text: "Measurement Error - Models, Methods and Applications" by J. Buonaccorsi. Chapman & Hall/CRC.

Prerequisites: The ST697M version of the course is designed for students across any discipline with a background in probability and statistics (at the level of ST515-516 or equivalent), as well as familiarity with regression analysis, including being comfortable with linear regression models and methods expressed in matrix-vector form. Prior exposure to nonlinear and logistic regression is not essential; we will review the standard methods without measurement error when we get to these topics.

The ST797M version is designed for graduates students in Statistics or other areas with an added interest in some of the basic theory. It assumes the student has taken ST607-608 or equivalent and have taken ST705 (or an equivalent course) or is currently enrolled in it.

Course Description. Measurement error, which is ubiquitous, occurs when one or more of the variables of interest in a model cannot be observed exactly. This is typically due to sampling error, instrument error or a combination of the two. There has been an explosion of activity over the last twenty years on the impact of, and how to deal with, measurement error, especially in Epidemiology, Ecology and Economics as well as in the mainstream Statistics literature. Examples abound including measuring dietary intake, physical activity, pollution exposure, economic or financial variables (at the household, firm or some aggregate spatial level such as region), ecological variables over spatial regions (ground cover, acorn production, animal abundance), physiological variables (e.g., cholesterol or hormone levels) and chemical quantities (e.g., water pH or alkalinity), to name just some. It also includes misclassification of categorical variables; examples include misclassifying disease or some other status using a fallible diagnostic tool (e.g., imaging blood measures, self-report), misclassification of habitat type with the use of satellite imagery, and misclassifying whether a part meets specification using an error-prone inspection technique.

This course exams the impact of measurement errors on standard statistical analyses which ignore them (so-called "naive" analyses) and describes methods of correcting for measurement error using additional information or data about the measurement error process. The latter comes in the form or replication or validation data. Correcting is important given that measurement error can often induce serious biases into standard analyses. A variety of correction techniques and related methods of inference (including the bootstrap) will be explored.

We examine these questions for i) misclassification in estimating one or more proportions and in two-way contingency tables; ii) measurement error in predictors and/or the response in simple and multiple linear regression as well as error in the response in estimating and comparing one or many means; iii) measurement error in nonlinear regression, including binary regression (e.g., logistic or probit) and Poisson type models.

The focus of is on understanding models and methods and applying them to examples from a variety of disciplines (epidemiology, ecology, economics, etc.). ST797M will have more of a theory component (see above). Computing will primarily use SAS and STATA, the two packages for which measurement error programs have been developed. Some R packages are emerging which we will review. Some prior experience with statistical computing is obviously helpful but prior use of SAS or STATA is not essential.