

STATISTICS COLLOQUIUM

MONDAY, SEPTEMBER 17, 2007

TALK: 4:00 PM — SCIENCE CENTER RM. 309

RECEPTION: 5:15 PM — SCIENCE CENTER, Rm. 302

“A Semi-Theoretician’s Mid-Day Confession:
The True Meaning of i.i.d. in (Applied) Statistics”

Xiao-Li Meng
Statistics Department
Harvard University

Doing good (applied) statistics is inherently – and increasingly – difficult. The size of the data and the complexity of their structure are increasing, as are the depth and specificity of the investigation goal. Yet the available time for conducting the study is decreasing due to intensive competition, especially for funding. Statistical consultants are therefore increasingly asked to perform magic, such as providing scientifically valid causal conclusions for a deeply stratified subgroup based on a handful of weighted samples with wildly varying weights. And by the way, the analysis must be done in one week and the methods must be implementable by (and explainable to) analysts whose statistical experience comes largely from reading output from SAS/SPSS/Stata.

This is not a cynical observation, but rather a real challenge that we as statisticians must face in order for our profession to remain at the core of quantitative scientific investigation. But of course with every challenge comes an opportunity! In this talk, I will report my own experiences both crying and smiling as a member of a team of statistical consultants for the National Latino and Asian American Study (NLAAS), a very recent survey of psychiatric epidemiology, which measured over 5000 variables and embedded experiments on different survey instruments.

I will report on the success and failure of using Bayesian modeling and multiple imputation to deal with the respondents’ lying behavior, as detected by the embedded experiments. I will also show how the ambiguity in the Institute of Medicine’s definition of “healthcare disparity” led to the concepts of conditional disparity and marginal disparity, which correspond to two extreme causal models. Yet the theory needed for the more realistic concept of joint disparity is still far from reach. The true meaning of i.i.d. will become clear only by the end of my talk; unless, of course, you have already deciphered it from this abstract.