

The University of Chicago
Department of Statistics
Seminar

Michael D. Larsen
Department of Statistics, University of Chicago

“Record Linkage and Analysis of Linked Files”

Monday, March 4, 2002 at 4:00 pm
133 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

Computerized record linkage can be used to join together two files that contain information on the same individuals, but lack unique personal identification codes. Mixture models, when fit to measurements of agreement between unlinked pairs of records on the two files, can be used to produce estimated probabilities that record pairs match. Given estimated probabilities, algorithms can be used to produce one-to-one assignments between individuals in the two files. Alternatively, models can be modified to reflect inherent one-to-one matching between files. In a study of the relationships between variables on the two matched files, errors in linkage produce bias in naive estimators. Estimators that adjust for matching error require estimated probabilities that pairs of records are matches. The interaction of methods of estimating probabilities, methods of estimating relationships, and methods of inference are explored. Applications mentioned include those from the U.S. Bureau of the Census, medical record linkage, and epidemiology.