



THE UNIVERSITY OF
CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

PETER BLANKENSHIP

Department of Statistics
The University of Chicago

A Comparison of Techniques for Handling Missing Data

FRIDAY, November 7, 2014, at 10:00 AM
Eckhart 117, 5734 S. University Avenue

ABSTRACT

Missing data are one of the most pervasive problems in all of statistics. This paper examines the impact that missing values can have on statistical estimates and several methods that can be used to address the missing value problem. In particular, the goal was to see how regression coefficients and their standard errors are affected when the data include missing or imputed values. To do this, data from the 2007 Oregon Health Experiment is used to estimate a linear model when no missing values are present. Then, missing values are artificially generated in different ways to simulate an assortment of missing data scenarios. Following this, the missing values are “filled in” using ad-hoc methods such as mean imputation, regression imputation, and stochastic imputation as well as more principled methods like maximum likelihood and multiple imputation. Finally, the linear model is then re-estimated using each imputed data set to determine whether the regression coefficients and their accompanying standard errors have changed as compared to the model under full data.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.