



The University of Chicago
Department of Statistics

MASTER'S THESIS PRESENTATION

CHAD KEARNEY

Department of Statistics
The University of Chicago

Survey Response Coding as an Incomplete-Data Problem

MONDAY, December 6, 2010, at 1:00 PM

110 Eckhart Hall, 5734 S. University Avenue

ABSTRACT

We develop approaches to automated coding (that is, to the mechanical categorization of free-form survey responses) wherein the correct classifications of free-form responses are treated as missing data. We briefly discuss existing approaches to automated coding, and describe types of problems for which these approaches are generally ill-suited. We then discuss several models for incomplete data, and show how these models can be used to generate predicted classifications of linguistic responses in situations where other coding methods might be found wanting. Finally, we apply these models to data from the large-scale National Immunization Survey, considering whether a missing-data approach can be used to classify natural language descriptions of race.

Information about building access for persons with disabilities may be obtained in advance by calling Sandra Romero at 773.702-0541 or by email (sandra@galton.uchicago.edu).