Analysis of Survey Data
94-830, Mini Spring 2017
The Heinz School, Carnegie Mellon University

Lectures: Tuesday and Thursday, 9:00-10:20, HBH 1007
Lab & Review Sessions Friday, 1:30-2:50, HBH 2009
Mini 4 March 20th – May 5th 2017
Course Web Page http://www.cmu.edu/blackboard
Instructor Claude Messan Setodji
Office TBD
e-mail: csetodji@andrew.cmu.edu or setodji@rand.org
Office Hours By appointment
Teaching assistant TBD (tbd@andrew.cmu.edu), Room TBD
T.A Office Hours TBD
Required Textbook None.
Iain Pardoe. Applied Regression Modeling, Second Edition. ebook that can be found at CMU library

Software SAS: The SAS software is also available in the Heinz computing clusters.
The TA will help on the basic fundamentals of SAS for non-experienced SAS users.
Grading Policy
The course grade will be determined by performance on homework assignments (40%), the class participation and attendance (20%) and a team project (40%). All projects need to be approved by me by April 6th, 2017.

Policy on Team Projects
Student will be required to form teams of three to conduct a team project where they will pick a policy question, find appropriate data and conduct the suitable analysis of the survey data for inference. The group of students will work together and write a report of their analysis results including the SAS codes used for analysis. The project grade will contribute to 40% of the final student grade.

Policy on Collaboration
You may collaborate on homework assignments. Of course, collaboration does not mean simply copying what someone else has done. Each individual must turn in a separate assignment and each individual is responsible for the content of the assignment that she/he turns in. To receive full credit, you must show your work and attach copies of SAS programs and output.

Anyone found to have cheated on any class activity will be dealt with according to the Heinz School and CMU policy. Furthermore, you will fail the class.

Engagement in the Course
An analyst benefits from working in cooperation with others as opposed to working alone. We will use class sessions as an opportunity to improve our analysis skills by engaging in small group activities on the topics of the day. Therefore, I expect you to attend each class and be ready to participate in the discussions and activities. This course does not have excused absences. You will earn points during each class session for both attendance and participation. Choosing not to attend, to not arrive on time or to stay for the entire session, or to not participate in the group discussion will result in the loss of some or all points for that class session.

Policy on Lateness and Grading of Homework and Tests
The due date for homework assignments will be printed on the assignments. You will generally have 7 days to do assignments. Assignments will be distributed in lecture and collected (and subsequently returned to you) in review session.

Assignments will be due at the beginning of review session or lectures. Late assignments will not be accepted except under extreme circumstances and with a documented excuse.

You are responsible for collecting your returned homework assignments. If your assignment is not returned to you, you must notify the instructor or the TA within one week of the return date in order to receive credit.
Course Prerequisites
- An empirical methods course that includes linear regression
- Basic experience using SAS is not required but is preferable (experience with other statistical program is not a substitute). Basics in SAS will be given at the beginning for non-experienced SAS users.

Course Description
This course is designed to teach students appropriate analytic methods for answering research questions using existing survey datasets. We will begin by discussing the creation of analysis plans, which will include basic descriptive and inferential analytic methods along with more advanced regression methods. Students will learn how to apply regression methods to survey data, so must have a solid foundation in linear regression and a basic understanding of logistic regression. We will focus on learning the types of data that can be used with particular analytic methods, as well as on learning which types of analytic methods are appropriate for which types of research questions. As such, we will learn when and how to manipulate existing data through the creation of dummy variables or composite variables, through data transformations, and through coding procedures for changing text into analyzable numeric data. We will also learn how to assess the degree of missing data and how to handle problematic levels of missing data, how to interpret statistical results, and how to present findings in text and in graphic format. This course will focus on research question-driven analysis. We will use SAS to conduct our analyses, and students will learn SAS procedures appropriate to those analyses. Students who have limited experience with SAS should attend optional Friday SAS workshops, in which they will receive basic instruction on navigating the SAS environment, importing datasets, and writing SAS code.

Course Objectives and Topics

Upon completion of this course, the student will be able to:

1. Assess the degree of missing data and determine whether missing data are associated with a particular set of respondents
2. Use appropriate coding techniques to transform text responses into analyzable numeric data
3. Manipulate data by creating dummy and composite variables
4. Create an analysis plan for answering the guiding research questions
5. Perform basic descriptive and relational analyses on collected data
6. Determine effects of explanatory variables on outcomes of interest using linear and logistic regression
7. Report survey findings in a clear, understandable manner that focuses on overall meaning rather than individual results