CSRPA DATA Bootcamp Series No. 1

Data Management with STATA for Social Science Research

The Center for Survey Research and Policy Analysis (CSRPA) prepared a data management workshop for undergraduate and graduate students and faculty for the upcoming winter period. This workshop will provide an opportunity to establish long-term data management skills to support quantitative social science research. The workshop will take place on campus over three days immediately before the beginning of the spring semester.

Meeting time: January 19th, 20th, and 21st. 3 sessions for two hours each (6:00-8:00 PM)
Requirement: Participants must bring a computer with the STATA program installed.
Place: ELABN TBD (classroom to be reserved)
Registration: please send your name and email address at csr@utrgv.edu.
Registration fee: $100 for undergraduates/$200 for graduates/$300 for faculty
If paid for by a department, the payments should be made through IDT to CSRPA before the first meeting session. Otherwise, registration fees can be paid either by ePay (the ePay link) or by check (payable to University of Texas Rio Grande Valley or UTRGV)
Instructor: Dr. Dongkyu Kim, Assistant Professor of Political Science

Topics: Foundations of Data Management

1. Basics of data structure – Jan 19th
   a. Working with STAT batch file
   b. Dealing with different data types (tab, sav, csv, xlsx, dta, etc.)
   c. Basic data structure for statistical analysis (cross-sectional, time-series, cross-sectional time-series)
   d. Increasing the efficiency of the data management process (data organization and using paths)
   e. Pulling what you need and making your work folder light

2. Basics of data manipulation – Jan 20th
   a. Basic data cleaning principles – Data labeling & value labeling
   b. Destring and encoding variables
   c. Recoding variables
   d. Generating variables
   e. Reshaping datasets

3. Data compilations – Jan 21st
   a. Data merging (Putting data side by side)
   b. Data appending (Vertically integrating data)

* For any inquiries, please email at csr@utrgv.edu.

CSRPA DATA Bootcamp Series Overview

The Center for Survey Research and Policy Analysis (CSRPA) will be offering a series of workshops for students and faculty to learn the necessary skills for quantitative research in the social sciences and humanities, but these skills may also be valuable to students in business, engineering, and the natural sciences. Starting with a workshop on data management skills, CSRPA will conduct a series of workshops on statistical estimations. The complete list of the workshop topics is listed below:

1. Data Management with STATA for Social Science Research - Feburary
   Participants will learn how to set up a long-term data management platform for individual research needs. This workshop is to learn about how to put data together in a reliable and consistent manner to prepare data for quantitative social science research.

2. Theories of Statistics and Test Statistics with STATA - March
   Participants will learn how to use STATA to learn basic statistical estimation skills, such as Z-test, T-test, Chi2-test, F-test, correlation, with STATA. If you want to strengthen your knowledge of theories of statistics and basic estimation skills, this workshop will be helpful.

3. Regression Analysis with STATA - April
   Participants will learn the bolts and nuts of regression analysis. The workshop covers a regression analysis of cross-sectional data and time-series cross-sectional data. Participants will learn how to build a regression model and make a publishable regression table with STATA.

4. Logistic Regression Analysis and Post-estimation with STATA - May
   Participants will learn basic estimation skills of logistic regression analysis. Using STATA, participants will run logistic regression analyses and learn how to calculate marginal effects and predicted probabilities.

5. Multilevel Regression Analysis with STATA - June
   Participants will learn basic concepts of multilevel regression analysis (MLM). With STATA, participants can expect to learn how to prepare data for the MLM analysis and run models.