VIRTUAL TRAINING COURSE ON ADVANCED REGRESSION ANALYSIS

Purpose

This training course covers the methods and theories of essential topics in advanced regression. The goal of the course is for participants to acquire application skills in correctly implementing methodologies discussed in the training, and knowledge in advanced regression that will enable them to understand the empirical and theoretical literature as discussed in reports and academic journals. The focus of the course is on the quantitative analyses of cross-section, pooled, and panel data that are deemed useful for participants who are interested in doing research. This course is useful for those who want to work or are working in institutions doing programs or project evaluations.

Benefits to Participants

The participants will be able to handle it. More so, the participants will be able to execute the model-building process using statistical software.

Target Participants

Technical staff whose work requires examining data sets in finding meaningful structural relationships, building and using regression models for making evaluations, predictions, and policy analysis. Participants must have prior knowledge in computer operations, basic data management, basic statistics, and regression analysis. Equivalent recommended prerequisites are PSRTI training courses on Microsoft Excel for Data Management (SW 1), Basic Statistics 1, and Regression Analysis.

Course Outline

I. Logistic Regression
- Motivation: why logistic regression
- The logit function
- The odds
- Defining the logistic regression model
- Estimating a logistic regression model
- The Pseudo-R^2
- Model diagnostics
  - Hosmer-Lemeshow goodness-of-fit test
  - Cook's distance for presence of outliers
  - VIF for the presence of multicollinearity
- Interpreting effect of explanatory variables
  - Change in log of odds
  - Odds ratio interpretation
  - Marginal effects
- Prediction and classification
  - Moving forward
  - Probit regression
  - Multinomial and ordinal logistic regression

II. Poisson Regression
- Motivation: why Poisson regression
- The Poisson distribution
- Defining the Poisson regression model
- Estimating a Poisson regression model using R
- The Pseudo-R^2
- Interpreting effect of explanatory variables
  - Change in log
  - Factor change
  - Percentage change
  - Average marginal effects

- Model diagnostics
  - Likelihood ratio test for overall significance
  - Deviance and Pearson goodness-of-fit tests
  - VIF for the presence of multicollinearity
  - Over-dispersion test
- Moving forward: other count data models
  - Negative binomial regression
  - Zero truncated and zero-inflated count models

III. Multinomial Logistic Regression
- Motivation: why multinomial logistic regression
- The logit function in the multinomial context
- Defining the multinomial logistic regression model
- Estimating a multinomial logistic regression model
- Investigating the significance of estimated coefficients
- Interpreting the logistic regression model in terms of
  - Ratio of Relative Risks
  - Marginal effects
  - Change of baseline group
- Assessing model fit
  - Likelihood ratio test for overall model significance
  - The Pseudo-R^2

IV. Instrumental Variable Regression
- Endogeneity
- The endogeneity problem – an intuitive approach
- Endogeneity defined
- Causes of endogeneity
- Detecting and addressing endogeneity
- The instrumental variable
- Defining the instrumental variable
  - Instrument validity
  - Examples of instrumental variables
- Two-stage least squares
  - The general procedure of 2SLS
  - Instrumental variable regression in R

V. Heckman Selection Model
- Selection bias: an intuitive approach
- Selection bias and missing not at random
- Defining the Heckman selection model
- Estimating the Heckman selection model
- Interpretation of results

VI. Piecewise Regression
- Defining the piecewise regression model
- Performing piecewise regression in R
  - Data visualizations
  - Breakpoint tests and determining the number of breakpoints
  - Estimating the piecewise regression model
- Assessing model fit

VII. Analysis of Pooled Data
- What is pooled data
- Analyzing pooled data
- Difference-in-differences regression

VIII. Introduction to Panel Data Models
- The nature of panel data
- Panel data model with unobserved effects
- The first-differences model
- The fixed effects model
  - One-way fixed effects model
  - Two-way fixed effects model
- The random effects model

Duration: 5 days – 8:30am to 4:30pm, July 11-15, 2022