30 Hours Certificate Course on Data Analysis Using STATA, R & PYTHON

Offered by the Department of Economics

DINABANDHU ANDREWS COLLEGE
GARIA, KOLKATA

Title

Data Analysis Using STATA, R & PYTHON

Course Overview: This certificate course aims to discuss data analysis techniques based on statistical and econometric models using STATA, R, and Python computer programs. This course will help those who wish to pursue a career as a data analyst. Given the importance of data analysis today, every student should be familiar with the techniques for advancing their careers in this field. This course is primarily aimed at Economics Honours students. This course, however, is useful for students from other disciplines also to develop skills in the field of data analysis.

Course Duration: 30 Hours (6 hours per week x 5 weeks)

The course consists of theory and practical classes

Final Assessment on the Last Day

Course Fee: Free of Cost

Pedagogy: Lecture on Theory and Practical Classes on Computer Software

Faculty Requirements: Internal- Department of Economics
Course Coordinator: Dr. Jhumur Sengupta
Joint Coordinator: Dr. Sajal Jana
Instructors: Prof. Jhumur Sengupta, Prof. Sajal Jana, Prof. Sumana Dutta, Prof. Ranjini Lahiri

Syllabus Structure of the Course

Module	Unit Title	Hours
Module-1	Exploratory Data Analysis	3
Module-2	Uncertainty in Data Analysis	4
Module-3	Predictive Analysis	4
Module-4	Time Series Analysis	5
Module-5	Panel data Models	5
Module-6	Non-linear Regression Model	4
Module-7	Introduction to Python Machine Learning	5
Module-8	Total	30

Module 1: Exploratory Data Analysis

1.1 Origin of Data
1.2 Preparing Data for Analysis
1.3 Data Types
1.4 Graphical Representation & Analysis
1.5 Summary of Data

Module 2: Uncertainty in Data Analysis

2.1 Merit of Normal Distribution in Data Analysis
2.2 Central Limit Theorem
2.3 Resampling & Bootstrap

2.4 Estimation & Confidence Intervals
2.5 Hypothesis Testing & False Discovery Rate Control
2.6 Bayesian Inference
2.7 Illustrations in Computer Program - R

Module 3: Predictive Analysis

3.1 Regression Modeling
3.2 Linear Regression Model
3.3 Dummy Variable Regression Model
3.4 Deviance & Likelihood
3.5 Regression Uncertainty

Module 4: Time Series Analysis

4.1 Stationary and Non-Stationary Time Series
4.2 Unit Root Tests in Time Series Models
4.3 Data Generation Process in Time Series Models-Random Walk, Moving Average, Autoregressive Process, ARMA Process, ARIMA Process
4.4 Autocorrelation Functions & Partial Autocorrelation Functions (ACF & PACF)
4.5 Box Jenkins Methodology & Forecasting
4.6 Vector Autoregressive Models (VAR)
4.7 Volatility in Time Series – ARCH & GARCH Models
4.8 Seasonality in Time Series
4.9 Analysis with Real Data Sets in Computer Programs-STATA & R

Module 5: Panel Data Model

5.1 The Fixed Effect & Least Squares Dummy Variable (LSDV) Models
5.2 The Random Effects Models
5.3 Pooled Regression Models
5.4 Fixed Versus Random Effects Models-Hausman Tests
5.5 Introduction to Dynamic Panel Data Models
5.6 Analysis with Real Data Sets in Computer Program – STATA

Module 6: Non-Linear Regression Model

6.1 The Linear Probability Model
6.2 Functional Forms: Logit & Probit Model
6.3 The Censored Regression
6.4 Binary Response Model
6.5 Goodness of Fit Measures
6.6 Analysis of Non-Linear Models in Computer Program STATA

Module 7: Introduction to Python Machine Learning

7.1 Machine Learning Models
7.2 Econometrics Versus Machine Learning
7.3 Why Machine Learning in Python?

7.4 Getting Started with Anaconda Platform & Jupyter Notebook
7.5 Descriptive Analysis in Python
7.6 Supervised Machine Learning Models Using Python
7.6.1 Splitting the Data Sets into Training & Test Data
7.6.2 Fitting Regression Models-Codes for Building Models
7.6.3 Model Diagnostics
7.6.4
7.6.5
7.6.6 Fitting Multiple Linear Regression Models

COURSE OUTCOME

Participants will gather knowledge of data analysis using real data and they will be able to interpret the results obtained from the analysis. They will be equipped with software skills based on STATA, R, and Python computer packages. This will help them to get jobs in the field of data analysis in the future.

LEARNING RESOURCES

SL. NO.	Title of the Book	Authors	Publisher
1.	Basic Statistics for Business & Economics	Douglas A. Lind, Willam G. Marchal, Samuel, and A. Wathen	McGraw Hill (2006)
2.	Marketing Research: An Applied Orientation	Naresh K. Malhotra and Stayabhushan Dash	6th Edition, Pearson.
3.	Basic Econometrics	Damodar N. Gujarati, and Dawn C. Porter	5th Edition, McGraw Hill Irwin.
4.	Introduction To Econometrics	Sengupta, Jhumur	1st Edition, Sultan Chand & sons
5.	Applied Econometric Time Series	Walter Enders	4th edition, Wiley.
6.	Econometric Analysis of Panel Data	Badi H. Baltagi	6th Edition, Springer.
7.	Data Management Using Stata: A Practical Handbook	Michael N. Mitchell	Stata Press
8.	R Programming for Beginners	Arora, Sandhya and Latesh, Malik	Universities Press, 2020.
9.	Machine Learning Using Python	Pradhan Manaranjan, and Kumar, U. Dinesh	Wiley, 2019

EVALUATION POLICY OF THE ADD-ON COURSE

The basic philosophy behind the Evaluation policy for this 30-hour add-on course is to objectively evaluate whether the participants have understood the basic concepts and to assess their ability to solve the data analysis problems by using computer packages STATA, R, and Python

The Evaluation would be done through 2 components –
i) C1 Course-end Assessments (Written Test) [Total Marks: 30]
ii) C2 Practical /LAB [Total Marks: 20]
Total Marks of the Evaluation process would be – 50 Marks

TABLE FOR QUALIFICATION

Total Score (Full Marks-50)	Grade
45-50	O-Outstanding
40-44	E-Excellent
35-39	A – Very Good
30-34	B-Good
25-29	C-Fair
Below 25	F-Fail

GENERAL RULES

1. Students must attend and appear for all the Module-End Assessments. If any student fails to submit any of the Module-End Assignments or fail to attend any of the Module-End Assessment examinations, the Student would NOT BE ELIGIBLE FOR CERTIFICATE.
2. Students must attend and appear for the Course-End Assessment Examination. If any student fails to submit the Course-End Assessment or fails to attend the Course-End Assessment Examination, the Student would NOT BE ELIGIBLE FOR CERTIFICATE.
3. Students must attend and appear for the Course-End Viva. If any student fails to fail to attend the Course-End Viva, the Student would NOT BE ELIGIBLE FOR CERTIFICATE.
4. Full marks for the evaluation of the course is 50.
5. Minimum 50% Marks must be scored to receive any Certificate. There will be only ONE Attempt allowed for each of the Module-End Assessments and the Course-End Assessment.
6. There will be NO PROVISION for Backlog Clearance.
7. General Rules and Regulations of the College must be followed without any exception.
8. A minimum of 75% attendance is required to receive the certificate of the course.