Statistical Analysis & Basic Econometrics for decision making with STATA/R

Health/Business/Economics

Statistical Analysis and Basic Econometrics for decision making with STATA/R is an instructor-led course. By the end of this course, you would have learned:

- **Basic STATA/R commands**
- Descriptive Statistic (tabulation and graphical methods for qualitative data)
- Numerical and Graphical Methods for quantitative data
- Correlation Analysis
- Inferential Statistics
- Two variable Regression Analysis
- Multiple Regression
- Dummy Variable model
- Health Statistics - I (Survival Analysis)
- Health Statistics – II (Observational Studies)

Approximate Training Hours: 65-70 hours spanning over 9 weeks.

Prerequisites: Basic understanding of probability, Basic Calculus and descriptive statistics like location and dispersion.

Detailed course outline of the topics covered under Statistical Analysis for Decision Making with STATA/R:

Topic	Details of the topic
Introduction	**Introduction to the STATA software and** essential STATA commands, such as gen, rename, label, egen, recode, label, tag, destring (with ignore and force), count, list, describe, substring, split, sort, browse.
	Learning about types of data (cross-sectional and time series), Nature of variable (nominal/ordinal/interval/ratio scale), and tools for descriptive statistics and correlation.

Descriptive Statistic	Descriptive Statistic (qualitative and quantitative data):
	Learning about Tabulation and Graphical Methods for categorical data-
	Two-way graphs: Bar graph and Pie chart (one/two/three/four variable using by and over option), horizontal and vertical bar graph, segmented bar graph with stack format.
	Tabulation with one variable, two variables, three variables, and four variables. Option for groups, over options, by option, if option, controlling the categorical axis, controlling legends, changing the look, and graphing by groups.
Numerical and Graphical Methods for quantitative data	Numerical and Graphical Methods for quantitative data:
	Learning about Location and Dispersion Methods-
	Arithmetic, Geometric and Harmonic mean, Median, Mode, Range, Standard Deviation, Variance, Inter Quartile Range, Coefficient of Variation, Identify outliers, Skewness, and Kurtosis. Criteria for selecting mean or median as the centre and standard deviation or Inter Quartile Range as dispersion. Comparison across groups and categories.
	Learning about Graphical methods-
	Range plot, Distribution plot, Box plot, Dot plot, and Histogram.
Correlation Analysis	Correlation Analysis:
	For Nominal, Ordinal and Ratio scale variables-
	Direction of association, and Degree of association. Nominal scale with nominal scale, Likert scale with nominal scale, ordinal scale with ordinal scale, ratio scale with ratio scale. Graphical analysis with scatter plot. Numeric Parametric and Non-Parametric methods (Karl Pearson, Spearman's rank correlation, Kendall's rank correlation, Pearson's chi-squared with Cramer's V). Discussing common mistakes in correlation analysis and relevance of correlation analysis for regression analysis.
Inferential Statistics	Inferential Statistics:
	Learning about distributions, properties, and probabilities for usage in hypothesis testing.
	Process of hypothesis testing with confidence interval/p-value/test of significance methods. Formulation of the *null and alternative* hypothesis,

	Selection of the Level of significance (with Type I and Type II error), Selection of the test statistic, Critical Value, Decision Rule, One tail/two-tail, and role of Central limit theorem.
	Applications with test of hypothesis (t-test, z-test, proportion test, and variance comparison test), One sample, two sample using group, two sample using variables and paired, One way ANOVA, Two-way ANOVA and Normality test (Skewness and Kurtosis test and Shapiro Wilk normality test). Non-parametric methods and Test of Hypothesis-Wilcoxon Matched Pairs signed-rank test, Wilcoxon rank-sum test, Kruskal Wallis rank test.
Two variable Regression Analysis	Two variable Regression Analysis: Stochastic specification of PRF, SRF, method of *Ordinary Least Squares* (OLS), Assumptions underlying the Method of Least Squares, Properties of Least Square Estimates (Linearity, Unbiasedness, and Minimum Variance-Gauss Markov Theorem), Goodness of Fit and Normality test of Residual. Evaluating Regression Results- Individual significance of intercept and slope coefficient, Overall significance of the model, and Goodness of fit (r^2).
Multiple Regression	Multiple Regression: Estimation of multiple regression, Interpretation of multiple regression equation, Multiple coefficients of determination, Adjusted R square ($\overline{R^2}$) and evaluation of results.
Dummy Variable model	Dummy Variable model: For qualitative independent variable- OLS method-ANOVA models, ANOVA model with two qualitative variables, Mixture of quantitative and qualitative regressors (ANCOVA), Interaction Effects, Use in seasonal Analysis and Piece-wise Linear Regression.
Multicollinearity, Heteroscedasticity and Autocorrelation	Multicollinearity, Heteroscedasticity and Autocorrelation- Nature, Causes, Practical Consequences, Detection method (Graphical and Numerical) and Remedial Measures.

Econometric Modelling	Econometric Modelling- Model specification and Diagnostic Testing, Types of specification errors, consequences of model specification errors (underfitting a model and overfitting a model), detection tests, incorrect specification of the stochastic error term, nested versus non-nested models, Model selection criteria: The R^2 criterion, Adjusted R square ($\overline{R^2}$), Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), Mallows's Cp Criterion.
Health Statistics	Survival Analysis- Time to event data, Kaplan–Meier survival curve, The logrank test, The hazard ratio and Modelling time to event data Observational studies – Risk and rates,Taking a random sample, Questionnaire and form design, Cross-sectional surveys, Non-randomised studies, Cohort studies, Case–control studies and Association and causality

NOTE: The entire course is completed with STATA first to understand theory, method, results and interpretation and then we move on to R to learn the same (R is optional).