

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB
(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

4th SEMESTER

BSCDA-401: DATA VISUALISATION AND DATA DRIVEN DECISION MAKING

Objectives

To enable the students to:

- get an introduction to Data Analytics and its role in business decisions process based on Analytics and how to represent it in right format

Units	Course Content
1	What is Data Analytics? Solving business problems using data analytics Making business-defining decisions using data analytics Why do you need a data and analytics framework? The 4 aspects of the data and analytics framework Data and analytics framework: tools and techniques Make better and faster decisions with data and analytics Identifying, organizing and processing data Data based decision process
2	<u>Types of data analysis techniques</u> The role of Excel The role of SAS The role of R The role of Python The Power of Visualization The role of Tableau/ Power BI

References

1. Now You See It: Simple Visualization Techniques for Quantitative Analysis, By Stephen Few
2. Data Smart: Using Data Science to Transform Information Into Insight, By John W Foreman
3. Data Science & Analytics (with Python, R and SPSS Programming), By V.K. Jain
4. Jeeva Jose – Beginner's Guide for Data Analysis using R Programming

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB
(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA-402 STATISTICAL ANALYSIS USING SPSS

Objectives

To enable the students to:

- General objective of the course is to introduce students in knowledge of basic concepts of statistics and the ability to analyse quantitative survey data using special statistical software SPSS

Units	Course Content
1	<ol style="list-style-type: none">1. Basic strategies of quantitative research: research questions, operationalisation, variables;2. How to prepare data for the analysis using SPSS -(module files; edit, view, utilities)3. Distribution of categorical data and univariate analysis (module descriptive statistics - frequencies, explore);
2	<ol style="list-style-type: none">1. Distribution of interval data and their analysis;2. Transformation of data (module transform, recode, compute, count, rank cases);3. Normal Distribution and hypothesis testing - statistical inference;
3	<ol style="list-style-type: none">1. Comparison of means: t-test, one-sample t-test; independent-samples t-test); analysis of variance;2. Bivariate analysis – cross tabulation;3. Strength of association - coefficients of association and correlation;
4	<ol style="list-style-type: none">1. Spurious correlations, elaboration, partial correlation;2. Linear regression;3. Factor analysis.

References

1. NORUŠIS, M. J. SPSS introductory statistics : student guide2.
2. SPSS Base 14.0 : user's guide. Chicago: SPSS Inc., 2005.
3. Data Science & Analytics (with Python, R and SPSS Programming), By V.K. Jain

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB
(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA-403: INTRODUCTION TO REGRESSION ANALYSIS

Objectives

To enable the students to:

- get an introduction to Data Analytics and its role in business decisions process based on Analytics and how to represent it in right format

Units	Course Content
1	Regression and correlation coefficients of a single bivariate normal distribution, Combination of Probabilities in tests of significance
2	Analysis of Variance (ANOVA) Introduction: Heterogeneity and Analysis of Variance and Covariance, Linear Hypothesis, Orthogonal splitting of total variation, Selection of Valid Error. Applications of the ANOVA technique to: one-way classified data, two-way classified data with equal number of observations per cell, testing simple regression correlation ratio, linearity of simple regression, multiple correlation and partial correlation coefficients.

References

1. Scheffe H. (1959): The Analysis of Variance, John Wiley
2. A Linear Model Approach to Regression Analysis and its Applications, Suddhendu Biswas, New Central book agency

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB
(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA-404: GENERALISED LINEAR MODELS

Objectives

To enable the students to:

- Gain the knowledge about construction methodology and procedure and also the ability to apply them in the enterprise reality

Units	Course Content
1	<ol style="list-style-type: none"> 1. GLM basics: components, exponential family, model fitting, frequent inference: analysis of deviance, stepwise selection, goodness of fit 2. Bayesian inference in GLMs (basics): priors, posterior, comparison with frequentist approach, posterior computation, MCMC strategies (Gibbs, MetropolisHastings) 3. Binary & categorical response data: (a) Basics: link functions, form of posterior, approximations, Gibbs sampling via adaptive rejection (b) Latent variable models: Threshold formulations, probit models, discrete choice models, logistic regression & generalizations, data augmentation algorithms (Albert & Chib + other forms) 4. Count data: Poisson & over-dispersed Poisson log-linear models, prior distributions, applications 5. Bayesian variable selection: problem formulation, mixture priors, stochastic search algorithms, examples, approximations 6. Bayesian hypothesis testing in GLMs: one- and two-sided alternatives, basic decision theoretic approaches, mixture priors, computation, order restricted inference 7. Survival analysis: censoring definitions, form of likelihood, parametric models, discrete-time & continuous time formulations, proportional hazards, priors for hazard functions, computation 8. Missing data: problem formulation, selection & pattern mixture models, shared variable approaches, examples 8. Multistate & stochastic modeling: motivating examples (epidemiologic studies with periodic observations of a disease process), discrete time approaches, joint models, computation 9. Correlated data (basics): mixed models for longitudinal, frequentist alternatives (marginal models, GEEs, etc) 10. Generalized linear mixed models (GLMM): definition, examples, normal linear case - induced correlation structure, priors, computation, multi-level models, covariance selection 11. Generalized additive models: definition, frequentist approaches for inference & computation (Hastie & Tibshirani), Bayesian approaches using basis functions, priors, computation 12. Factor analytic models: Underlying normal formulations, mixed discrete & continuous outcomes, generalized factor models, joint models for longitudinal and event time data, covariance selection, model identifiability issues, computation 13. Mixtures of GLMs: finite mixtures, latent class models, computational approaches

References

- 1) Generalized Linear Models -- Ulff Olsson, Overseas Press.

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB
(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA-405: PROJECT I

Objectives

To enable the students to:

- To give the students exposure to dummy projects on Analytics using specific softwares/ tools.

Software/ tools: SAS/MS Excel, Python