

# BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

## WORK INTEGRATED LEARNING PROGRAMMES

### COURSE HANDOUT

#### Part A: Content Design

<b>Course Title</b>	Advanced Statistical Techniques for Analytics
<b>Course No(s)</b>	SS ZG536
<b>Credit Units</b>	4
<b>Course Author</b>	
<b>Version No</b>	
<b>Date</b>	

#### Course Description

This course will cover the statistical techniques which are very important in Data analytics. It covers the models related to descriptive statistics, inferential statistics, predictive analytics and applied multivariate analytics. And also this course introduces the important tools used in Data Analytics in Particular R language.

#### Course Objectives

<b>CO1</b>	Understanding the data representation and analysis which is very important in Data Analytics
<b>CO2</b>	Understanding the predictive & inferential statistical models used in Data Analytics
<b>CO3</b>	Understanding the tools used in data analytics

#### Text Books

T1	Mathematical Statistics and Data Analysis, John A Rice, 3 <sup>rd</sup> Edition, Cengage Learning
T2	Applied Logistic Regression, Hosmer and Lemeshow, 3 <sup>rd</sup> Edition, Wiley
T3	Regression Analysis by Example, Samprit Chatterjee, Ali S. Hadi 5th Edition, Wiley

#### Reference Books

R1	<i>Business Analytics: Methods, Models, and Decisions</i> , 1 <sup>st</sup> edition, James R. Evans, Pearson/PH
R2	Statistics for Business and Economics by Anderson, Sweeney and Williams, CENAGE learning
R2	R programming for Data Science, Roger D Peng,

## **Modular Content Structure**

1. Descriptive Statistics
  - 1.1 Data representation & Visualisation
  - 1.2 Measures of Central Tendency
    - 1.2.1. Mean, Median, Mode
    - 1.2.2. Skewness and Kurtosis
  - 1.3 Measures of Variability
  - 1.4 Probability
  - 1.5 Conditional probability
    - 1.5.1. Bayes theorem
  - 1.6 Probability Distributions
    - 1.6.1. Binomial distribution
    - 1.6.2. Poisson distribution
    - 1.6.3. Normal distribution
2. Inferential Statistics
  - 2.1 Sampling
  - 2.2 Central Limit Theorem
  - 2.3 Estimation
  - 2.4 Type I, Type II errors
  - 2.5 Testing of Hypothesis – Mean & Proportions
3. Predictive Analytics – 1 (Linear Regression)
  - 3.1 Covariance
  - 3.2 Correlation
  - 3.3 Sum of Least Squares of Errors of parameters
  - 3.4 Simple linear regression
  - 3.5 Maximum likelihood
  - 3.6 Ridge Models & Lasso Model
  - 3.7 Assumptions of linear regression and implications
    - 3.7.1 Heteroscedasticity
    - 3.7.2 Multicollinearity
    - 3.7.3 Serial correlation
  - 3.8 Model validation
  - 3.9 Multiple linear regression
4. Predictive Analytics – 2 (Logistic Regression)
  - 4.1 Introduction
  - 4.2 Logistic regression function

- 4.3 Training the model
- 4.4 Polynomial Logistic Regression
- 4.5 Fitness of the model
- 4.6 Dependent variable prediction
- 5. Predictive Analytics – 3 (Forecasting Models)
  - 5.1 Principles of Forecasting
  - 5.2 Time series Analysis
    - 5.2.1. Smoothing & decomposition methods
    - 5.2.2. ARIMA Model
    - 5.2.3 GARCH models
    - 5.2.4 Holt – Winter method
    - 5.2.4 Multivariate time series models
  - 5.3 Casual methods
  - 5.4 Moving Averages
  - 5.5 Exponential smoothing
  - 5.6 Forecast Data patterns – level, seasonality ,cyclical
- 6. Applied Multivariate Analytics
  - 6.1 Introduction
  - 6.2 Multivariate Normal distribution
  - 6.3 Principal Component Analysis
  - 6.4 Factor Analysis
  - 6.5 Discriminant Analysis
  - 6.6 MANOVA

**Learning Outcomes:**

No	Learning Outcomes
LO1	Clear understanding of the difference between Business Intelligence and data analytics
LO2	Various statistical models & tools used in Data Analytics
LO3	Understanding the use of R language in Data Analytics

## Part B: Contact Session Plan

<b>Academic Term</b>	First Semester 2018-2019
<b>Course Title</b>	Advanced Statistical Techniques for Analytics
<b>Course No</b>	SS ZG536
<b>Lead Instructor</b>	YVK RAVI KUMAR

### Course Contents

<b>Contact Session (2Hrs)</b>	<b>List of Topic Title (from content structure in Part A)</b>	<b>Topic # (from content structure in Part A)</b>	<b>Text/Ref Book/external resource</b>
1	Descriptive Statistics <ul style="list-style-type: none"> <li>✓ Data representation &amp; Visualisation</li> <li>✓ Measures of Central Tendency</li> <li>✓ Mean, Median, Mode</li> <li>✓ Skewness and Kurtosis</li> <li>✓ Measures of Variability</li> <li>✓ Probability</li> </ul>		
2	<ul style="list-style-type: none"> <li>✓ Conditional probability</li> <li>✓ Bayes theorem</li> </ul>		
3	<ul style="list-style-type: none"> <li>✓ Probability Distributions</li> <li>✓ Binomial distribution</li> <li>✓ Poisson distribution</li> <li>✓ Normal distribution</li> </ul>		
4	Inferential Statistics <ul style="list-style-type: none"> <li>✓ Sampling</li> <li>✓ Central Limit Theorem</li> <li>✓ Estimation</li> <li>✓ Type I, Type II errors</li> </ul>		
5	<ul style="list-style-type: none"> <li>✓ Testing of Hypothesis – Mean &amp; Proportions</li> </ul>		
6	Predictive Analytics – 1(Linear Regression) <ul style="list-style-type: none"> <li>✓ Covariance</li> <li>✓ Correlation</li> <li>✓ Sum of Least Squares of Errors of parameters</li> <li>✓ Simple linear regression</li> </ul>		
7	<ul style="list-style-type: none"> <li>✓ Maximum likelihood</li> </ul>		

	✓ Ridge Models & Lasso Model		
8	<b>Revision</b>		
9	<ul style="list-style-type: none"> <li>✓ Assumptions of linear regression and implications</li> <li>✓ Heteroscedasticity</li> <li>✓ Multicollinearity</li> <li>✓ Serial correlation</li> <li>✓ Model validation</li> </ul> Multiple linear regression		
10	Predictive Analytics–2(Logistic Regression) <ul style="list-style-type: none"> <li>✓ Introduction</li> <li>✓ Logistic regression function</li> <li>✓ Training the model</li> </ul>		
11	<ul style="list-style-type: none"> <li>✓ Polynomial Logistic Regression</li> <li>✓ Fitness of the model</li> <li>✓ Dependent variable prediction</li> </ul>		
12	Predictive Analytics–3(Forecasting Models) <ul style="list-style-type: none"> <li>✓ Principles of Forecasting</li> <li>✓ Time series Analysis</li> <li>✓ Smoothing &amp; decomposition methods</li> <li>✓ ARIMA Model</li> <li>✓ GARCH models</li> <li>✓ Holt – Winter method</li> </ul>		
13	<ul style="list-style-type: none"> <li>✓ Multivariate time series models</li> <li>✓ Casual methods</li> <li>✓ Moving Averages</li> <li>✓ Exponential smoothing</li> <li>✓ Forecast Data patterns – level, seasonality, cyclical</li> </ul>		
14	Applied Multivariate Analytics <ul style="list-style-type: none"> <li>✓ Introduction</li> <li>✓ Multivariate Normal distribution</li> <li>✓ Principal Component Analysis</li> </ul>		
15	<ul style="list-style-type: none"> <li>○ Factor Analysis</li> <li>○ Discriminant Analysis</li> <li>○ MANOVA</li> </ul>		
16	<b>REVISION</b>		

## **Evaluation Scheme**

Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Quiz-1/Assignment-I	Online		5%	September 10 to 20, 2018
	Quiz-2	Online		5%	October 20 to 30, 2018
	Quiz – 3 / Assignment-II	Online		10%	November 10 to 20, 2018
EC-2	Mid-Semester Test	Closed Book	2 hours	30%	29/09/2018 (AN) 2 PM – 4 PM
EC-3	Comprehensive Exam	Open Book	3 hours	50%	24/11/2018 (AN) 2 PM – 5 PM

**Note** - Evaluation components can be tailored depending on the proposed model.

## **Important Information**

Syllabus for Mid-Semester Test (Closed Book): Topics in Weeks 1-7

Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study

Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Announcements regarding the same will be made in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.