

## BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI WORK INTEGRATED LEARNING PROGRAMMES

M .Tech (Data Science & Engineering)

**II Semester, 2021-22**

Course Handout(Revised on Jan 2020)

<b>Course Title</b>	<b>Introduction to Statistical Methods</b>
<b>Course No(s)</b>	DSECL ZC413

### Course Description

This course will cover the statistical techniques which are very important in Data Science. It covers the models related to descriptive statistics, inferential statistics, predictive analytics and applied multivariate analytics.

### Course Objectives

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

### Text Books

<b>No</b>	<b>Author(s), Title, Edition, Publishing House</b>
T1	Probability and Statistics for Engineering and Sciences, 8 <sup>th</sup> Edition, Jay L Devore, Cengage Learning
T2	Applied Logistic Regression, Hosmer and Lemeshow, 3 <sup>rd</sup> Edition, Wiley
T3	Introduction to Time Series and Forecasting, Second Edition, Peter J Brockwell, Richard A Davis, Springer.

### Reference Books

<b>No</b>	<b>Author(s), Title, Edition, Publishing House</b>
R1	Miller and Freund's Probability and statistics for Engineers, 8 <sup>th</sup> Edition, PHI
R2	Statistics for Business and Economics by Anderson, Sweeney and Williams, CENAGE learning

### Modular Content Structure



1. Descriptive Statistics
  - 1.1. Data Visualisation
  - 1.2. Measures of Central Tendency
  - 1.3. Measures of Variability
2. Probability
  - 2.1 Probability – Introduction and Basics
  - 2.2 Conditional probability
  - 2.3 Bayes' theorem
3. Probability Distributions
  - 3.1. Random variables – Discrete & Continuous(single and multiple)
  - 3.2. Probability Distributions
    - 3.2.1. Binomial Distribution
    - 3.2.2. Poisson Distribution
    - 3.2.3. Normal Distribution
4. Testing of Hypothesis
  - 4.1. Sampling & Estimation
  - 4.2. Type I, Type II errors
  - 4.3. Testing of Hypothesis – Mean – one and two mean
  - 4.4. Testing of hypothesis – Proportions – one and several Proportions
  - 4.5. ANOVA
5. Regression
  - 5.1. Covariance
  - 5.2. Correlation
  - 5.3. Sum of Least Squares
  - 5.4. Simple linear regression
  - 5.5. Ridge Models & Lasso Model
  - 5.6. Assumptions of linear regression
  - 5.7. Model validation
  - 5.8. Multiple linear regression
  - 5.9. Nonlinear regression
  - 5.10. Logistic regression
6. Forecasting Model
  - 6.1. Principles of Forecasting
  - 6.2. Time series Analysis
    - 6.2.1. Smoothing & decomposition methods
    - 6.2.2. ARIMA Model
    - 6.2.3 Moving Averages

## 6.2.4 Exponential smoothing

### Learning Outcomes:

No	Learning Outcomes
LO1	Clear understanding of the various statistical models to model the data
LO2	Drawing conclusions from the models selected to understand the data

### Part B: Course Handout

<b>Academic Term</b>	II semester ,2020 – 21
<b>Course Title</b>	<b>Introduction to Statistical Methods</b>
<b>Course No</b>	DSECLZ ZC413

### Course Contents

#### Contact Session 1: Module 1(Descriptive Statistics)

Contact Session	List of Topic Title	Reference
CS - 1	Descriptive Statistics: Data Visualisation, Measures of Central Tendency, Measures of Variability	T1:Chapter 1
HW	Problems on Descriptive Statistics	T1:Chapter 1
Lab		

#### Contact Session 2: Module 2 - Probability

Contact Session	List of Topic Title	Reference
CS - 2	Probability - Introduction and Basics, Conditional probability, Bayes' theorem	T1:Chapter 2
HW	Problems on probability	T1:Chapter 2

Lab		
-----	--	--

### Contact Session 3: Module 3 – Probability Distributions

Contact Session	List of Topic Title	Reference
CS - 3	Random Variables – Discrete & Continuous (single variable)	T1:Chapter 3 & 4
HW	Problems on Random Variables	T1:Chapter 3 & 4
Lab		

### Contact Session 4: Module 3 – Probability Distributions

Contact Session	List of Topic Title	Reference
CS - 4	Random Variables – Discrete & Continuous (Multi variates)	T1:Chapter 3 & 4
HW	Problems on Joint RVs	T1:Chapter 3 & 4
Lab		

### Contact Session 5: Module 3 – Probability Distributions

Contact Session	List of Topic Title	Reference
CS - 5	Probability Distributions – Binomial, Poisson and Normal Distributions	T1:Chapter 3 & 4
HW	Problems on probability distributions	T1:Chapter 3 & 4
Lab		

### Contact Session 6: Module 4 – Testing of Hypothesis

Contact Session	List of Topic Title	Reference
CS - 6	Sampling & Estimation	R1
HW	Problems on Interval Estimation	R1
Lab		

**Contact Session 7: Module 4 – Testing of Hypothesis**

Contact Session	List of Topic Title	Reference
CS - 7	Testing of Hypothesis - Type I & II errors, Critical region, t – test, Chi – Square test and F – test(Introduce and discuss these tests)	T1:Chapter 7 ,8,9 & 10
HW	Problems on Testing of Hypothesis	T1:Chapters 7 to 10
Lab		

**Contact Session 8:**

Contact Session	List of Topic Title	Reference
CS - 8	REVISION OF THE TOPICS COVERED	
HW		
Lab		

**MID SEMESTER EXAMINATION**
**Contact Session 9 & 10: Module 4 – Testing of Hypothesis**

Contact Session	List of Topic Title	Reference
CS – 9 & 10	Testing of Hypothesis - mean and proportions related models (one mean, two mean, one proportion and Several proportions with small and big samples wherever applicable)	T1:Chapter 7 ,8,9 & 10
HW	Problems on Testing of Hypothesis	T1:Chapters 7 to 10
Lab		

**Contact Session 11: Module 5 – Regression**

Contact Session	List of Topic Title	Reference
CS - 11	Covariance, Correlation, Rank Correlation	T1:Chapter 12 & 13

HW	Problems on correlation and co variance	T1:Chapter 12 & 13
Lab		

### Contact Session 12: Module 5 – Regression

Contact Session	List of Topic Title	Reference
CS - 12	Simple Linear regression model, Assumption of the model, interpretation of the model	T1:Chapter 12 & 13
HW	Problems on Linear regression	T1:Chapter 12 & 13
Lab		

### Contact Session 13: Module 5 – Regression

Contact Session	List of Topic Title	Reference
CS - 13	Multiple linear regression model, non – linear regression & Logistic regression( Introducing the model as a continuation of regression models)	T1:Chapter 12 & 13 and T2
HW	Problems on Linear regression	T1:Chapter 12 & 13
Lab		

### Contact Session 14: Module 6 – Forecasting Models

Contact Session	List of Topic Title	Reference
CS - 14	Principles of Forecasting, Time series models _ smoothing and decomposition methods, AR,MA,ARIMA Models(Introducing the models only)	T3
HW	Case studies on Time series models	
Lab		

### Contact Session 15: Module 6 – Forecasting Models

Contact Session	List of Topic Title	Reference
CS - 13	Moving Averages and Exponential smoothing models	T3
HW	Case studies on Time series models	
Lab		

**Contact Session 16:**

Contact Session	List of Topic Title	Reference
CS - 16	<b>REVISION OF THE SYLLABUS</b>	
HW		
Lab		