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

COURSE HANDOUT

Part A: Content Design

Course Title	ADVANCED STATISTICAL TECHNIQUES FOR ANALYTICS
Course No(s)	SSZG 536
Credit Units	
Course Author	
Version No	
Date	

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

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

Text Books

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

Reference Books

No	Author(s), Title, Edition, Publishing House
R1	Miller and Freund's Probability and statistics for Engineers, 8th Edition, PHI
R2	Statistics for Business and Economics by Anderson, Sweeney and Wiliams, 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
   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

7. Applied Multivariate Analytics

6.1 Introduction

6.2 Joint distributions – Discrete & Continuous

6.3 Multivariate Normal Distribution

6.4 Principal Component Analysis

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

Content Structure
Module Summary:

No   Title of the Module

M1   Descriptive Statistics

M2   Inferential Statistics

M3   Predictive Analytics(Part 1)

M4   Predictive Analytics(Part 2)

M5   Predictive Analytics(Part 3)

M6   Applied Multivariate Analysis

Part B: Contact Session Plan

Academic Term	Second Semester 2020-2021

Course Title	ADVANCED STATISTICAL TECHNIQUES FOR ANALYTICS
Course No	SS ZG536
Lead Instructor	AKANKSHA BHARADWAJ

Detailed Structure

Glossary of Terms

Contact Hour (CH) stands for a hour long live session with students conducted either in a physical classroom or enabled through technology. In this model of instruction, instructor led sessions will be for 22 CH.

        Pre CH = Self Learning done prior to a given contact hour

        During CH = Content to be discussed during the contact hour by the course instructor

        Post CH = Self Learning done post the contact hour

RL stands for Recorded Lecture or Recorded Lesson. It is presented to the student through an online portal. A given RL unfolds as a sequences of video segments interleaved with exercises

SS stands for Self-Study to be done as a study of relevant sections from textbooks and reference books. It could also include study of external resources.

LE stands for Lab Exercises

HW stands for Home Work.

M stands for module. Module is a standalone quantum of designed content. A typical course is delivered using a string of modules. M2 means module 2.

Course Contents

Contact hour-1, Module 1:

Time	Type	Description	References
Pre- CH-1	RL	RL – 1.1.1 (Data representation)	
During CH-1	CH-1	Discussion on data representation	T1:Chapter 1
Post-CH-1	HW	T1: Chapter 1	T1:Chapter 1
Lab			

Contact hour-2, Module 1

Time	Type	Description	References
Pre-CH-2	RL	RL – 1.1.2 (Data visualization)	
During CH-2	CH-2	Measures of Central Tendency, Measures of Variability	T1:Chapter 1
Post-CH-2	HW	T1: Chapter 1	T1:Chapter 1
Lab			

Contact hour-3, Module 1:

Time	Type	Description	References
Pre-CH-3	RL	RL – 1.1.3(Introduction to Probability)	
During CH-3	CH-3	Review of set theory basics, Probability - Introduction and Basics	T1:Chapter 2
Post-CH-3	HW	T1: Chapter 2(Exercise Problems on probability)	T1:Chapter 2
Lab			

Contact hour-4, Module 1:

Time	Type	Description	Reference
Pre-CH-4	RL	RL – 1.2.1(Conditional Probability)& RL -1.2.2& RL – 1.2.3) (Baye's theorem & Applications)	
During CH-4	CH-4	Problems on conditional probability and Baye's theorem	T1:Chapter 2
Post-CH-4	HW	T1: Chapter 2(Exercise Problems on Conditional probability & Baye's theorem)	T1:Chapter 2
Lab			

Contact hour-5, Module 1:

Time	Type	Description	Reference
Pre-CH-5	RL	RL – 1.2.4(Random variables)	
During CH-5	CH-5	Problems on Random variables – Discrete RV	T1:Chapter 3
Post-CH-5	HW	T1: Chapter 3(Exercise problems on discrete random variables)	T1:Chapter 3
Lab			

Contact hour-6, Module 1:

Time	Type	Description	Reference
Pre-CH-6	RL	RL – 1.2.4(Random variables)	
During CH-6	CH-6	Problems on Random variables – Continuous RV	T1:Chapter 4
Post-CH-6	SS	T1: Chapter 4(Exercise problems on Continuous RV)	T1:Chapter 4
Lab			

Contact hour-7, Module 1

Time	Type	Description	References

Pre-CH-7	RL	RL – 1.3.1 & 1.3.2(Binomial & Poisson distributions)	
During CH-7	CH-7	Review & Problems on discrete probability distributions	T1:Chapter 3 & Chapter 4
Post-CH-7	HW	T1: Chapter 3 & 4(Exercise problems on Binomial & Poisson Distributions)	T1:Chapter 3 & Chapter 4
Lab			

Contact hour-8, Module 1

Time	Type	Description	References
Pre-CH-8	RL	RL – 1.3.4(Normal distributions)	
During CH-8	CH-8	Review of Normal Distribution	T1:Chapter 4
Post-CH-8	HW	T1: Chapter 4(Exercise problems on Normal Distributions)	T1:Chapter 4
Lab			

Contact hour-9, Module 2

Time	Type	Description	References
Pre-CH-9	RL	RL – 2.1.1(Sampling & Estimation), RL – 2.1.2(Sampling methods), RL – 2.1.3(CLT),RL – 2.2.1(Testing of Hypothesis), RL – 2.2.2 – Types of errors	
During CH-9	CH-9	Discussion on Sampling and estimation with examples	R1
Post-CH-9	HW	R1: Problems on interval estimation	R1
Lab			

Contact hour-10, Module 2

Time	Type	Description	References
Pre-CH-10	RL	RL – 2.2.3 & RL – 2.2.4 (Testing of hypothesis with mean) RL – 2.3.1 & RL – 2.3.2 (Testing of hypothesis with proportions)	
During CH-10	CH-10	Discussion on testing of hypothesis & problem solving	T1:Chapter 7&8&9 and 10
Post-CH-	HW	T1:Chapter 7&8(Problems on testing of hypothesis)	T1:Chapter

			7&8&9 and 10
10			
Lab			

Contact hour-11

Time	Type	Description	References
Pre-CH-11		REVIEW	
During CH-11	CH-11	REVIEW	
Post-CH-11	HW	REVIEW	
Lab			
Syllabus for Mid-Semester Test: Topics covered in the first 10 contact session			

Contact hour-12, Module 2

Time	Type	Description	Content Reference
Pre-CH-12	RL	RL – 2.2.3 to RL – 2.3.2	
During CH-12	CH-12	Problems on Testing of Hypothesis – discussion of all the tests used	T1:Chapter 7&8&9 and 10
Post-CH-12	HW	T1:Chapter 7 to 10 (Problems on testing of hypothesis)	T1:Chapter 7&8&9 and 10
Lab			

Contact hour-13, Module 3

Time	Type	Description	Reference
Pre-CH-13	RL	RL – 3.1.1 & RL – 3.1.2 & RL – 3.1.3	
During CH-13	CH-13	Discussion and problem solving on Regression & Correlation	T1:Chapter 12 &13
Post-CH-13	HW	Problems on regression and correlation	T1:Chapter 12 &13
Lab			

Contact hour-14, Module 3

Time	Type	Description	References

Pre-CH-14	RL	RL – 3.2.1 & RL – 3.2.2	
During CH-14	CH-14	Problems on multinomial regression and polynomial regression	T1:Chapter 12 &13
Post-CH-14	HW	T1:Exercise problem from Chapter 12 &13	T1:Chapter 12 &13
Lab			

Contact hour-15, Module 4

Time	Type	Description	References
Pre-CH-15	RL	RL –4.1	
During CH-15	CH-15	Discussion on Logistic regression	T1:Chapter 13
Post-CH-15	HW	Problems on Logistic regression	TBD
Lab			

Contact hour-16 Module 5

Time	Type	Description	References
Pre-CH-16	RL	RL –5.1 & RL – 5.2	
During CH-16	CH-16	Discussion on Time Series Analysis	T3:Chapter 1 & 2
Post-CH-16	HW	Problems on Time Series	T3:Chapter 1 & 2
Lab			

Contact hour-17, Module 5

Time	Type	Description	References
Pre-CH-17	RL	RL –5.3	
During CH-17	CH-17	Problems on Smoothing model, AR, MA,ARMA & ARIMA Models	T3:Chapter 3
Post-CH-17	HW	Problems on Time Series	T3:Chapter 3
Lab			

Contact hour-18, , Module 5

Time	Type	Description	References
Pre-CH-18	RL	Module 5	
During CH-18	CH-18	Discussion on Time series with Examples	T3
Post-CH-18	HW		
Lab			

Contact hour-19, , Module 6

Time	Type	Description	Content Reference
Pre-CH-19	RL	RL –6.1.1	
During CH-19	CH-19	Bivariate Distribution	T1:Chapter 5
Post-CH-19	HW	T1:Exercise problems from Chapter 5	T1:Chapter 5
Lab			

Contact hour-20, Module 6

Time	Type	Description	References
Pre-CH-20	RL	RL –6.1.2 & 6.2	
During CH-20	CH-20	Bivariate Distribution – conditional & Normal	T1:Chapter 5
Post-CH-20	HW	Problems on conditional distributions	T1:Chapter 5
Lab			

Contact hour-21, Module 6

Time	Type	Description	References
Pre-CH-21	RL	RL –6.3	
During CH-21	CH- 6.3	PCA	Class notes
Post-CH-21	HW	Problem on PCA	Class notes
Lab			

Contact hour-22

Time	Type	Description	References
Pre-CH-22	RL	REVIEW	
During CH-22	CH-22	REVIEW	
Post-CH-22	HW	REVIEW	
Lab			
Syllabus for Comprehensive Exam (Open Book) All topics given in Plan			

Evaluation Scheme

Legend: EC = Evaluation Component

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Quiz-1		*	5%	February 1-15, 2021
	Quiz-2		*	5%	March 1-15, 2021
	Assignment		*	10%	April 1-15, 2021
EC-2	Mid-Semester Test	Closed Book	2 hours	30%	Friday, 05/03/2021 (AN) 2 PM – 4 PM
EC-3	Comprehensive Exam	Open Book	3 hours	50%	Friday, 30/04/2021 (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 Ch 1-10

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

Evaluation Guidelines:

1. 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.

2. 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.

3. 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 centers 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.