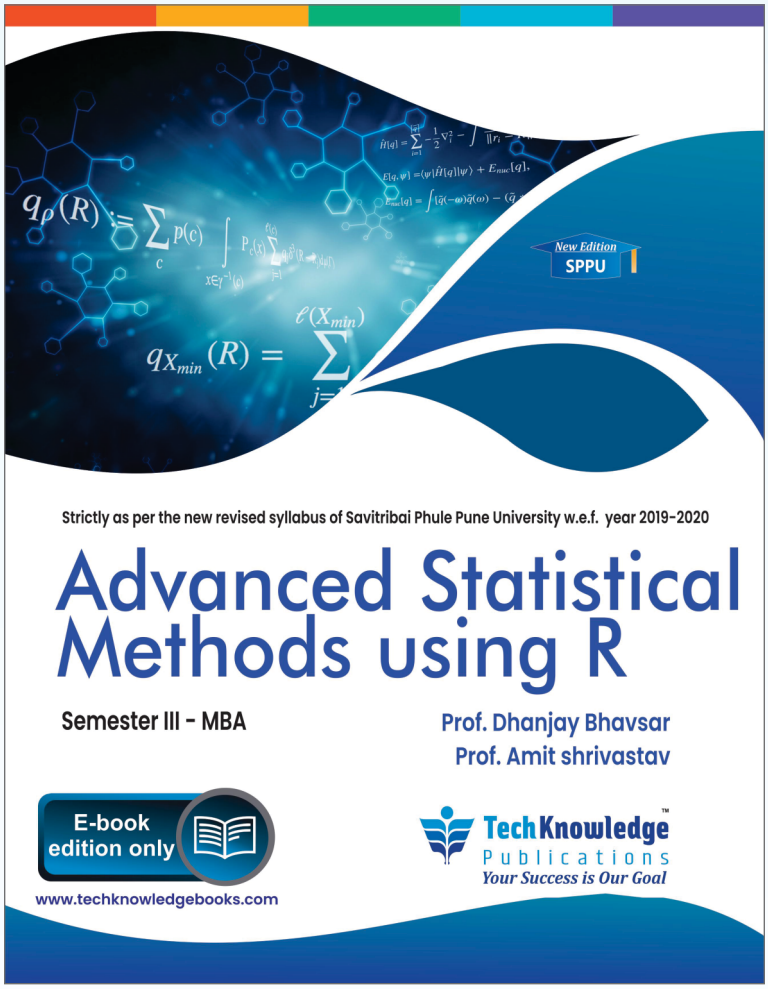
TechKnowledge™ Publications
Your Success is Our Goal

Advanced Statistical Methods using R

MBA | Sem 3 | SPPU



15 % DISCOUNT
Students Special
Buy NOW

~~Price : 200/-~~

After Discount Price : 170/-

ISBN : 978-81-947867-1-9

E-book Purchase link https://techknowledgebooks.com/product/advanced-statistical-methods-using-r/

MCQ link https://campusfunda.com/Book/mcq-view/NDk=

About TechKnowledge Publication

Our vision is to create an impact on education

Our mission is to bring quality content closer to the student community

We accomplish this by continuous innovation and research towards understanding how learning happens. Our authors are ambassadors of technical & management education and know how to enable the readers for success beyond exam results and into their careers.



THE WHOLE LIBRARY IS IN YOUR PHONE

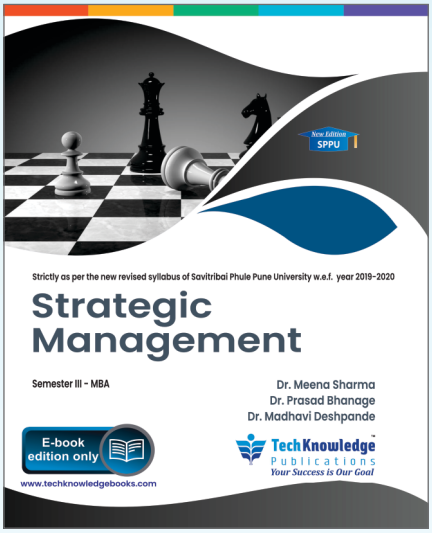
Step to Purchase E-Book

1. Visit the www.techknowledgebooks.com
2. From bookstore select appropriate University Branch and Subject.
3. Select 'Buy eBook' button.
4. You will be redirected to eBook portal -www.campusfunda.com
5. Signup / Login (mandatory) (Please confirm signup email in your inbox)
6. In ebooks section you will find the respective eBook for purchase.
7. You will have ebook in your account.

online or offline education learn better with Techknowledge eBooks

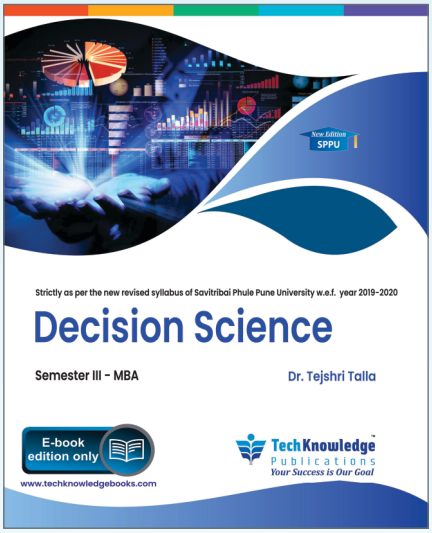
TechKnowledge™ Publications
Your Success is Our Goal

SPPU MBA SEM III e-books Available



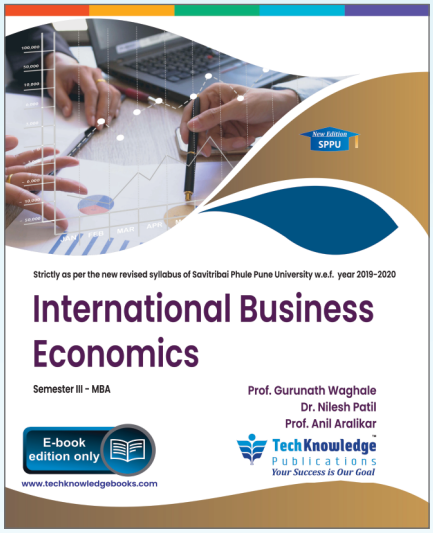
Price : ₹ 190/-

ISBN : 978-93-90428-34-2



Price : ₹ 190/-

ISBN : 978-81-947867-2-6



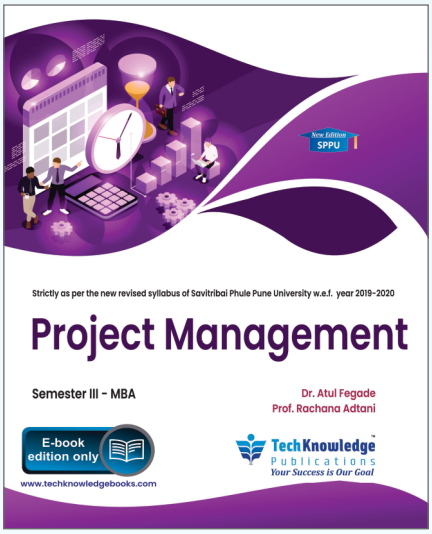
Price : ₹ 180/-

ISBN : 978-93-90428-26-7



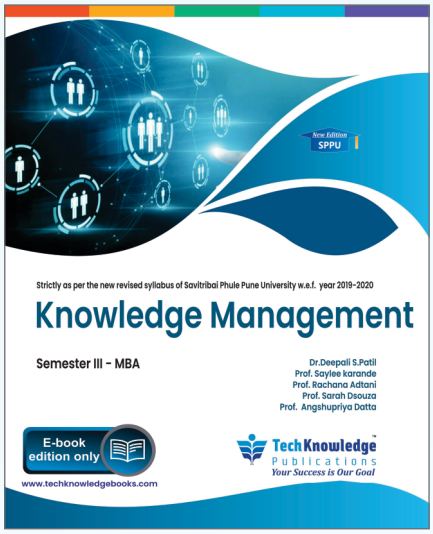
Price : ₹ 190/-

ISBN : 978-93-90428-27-4



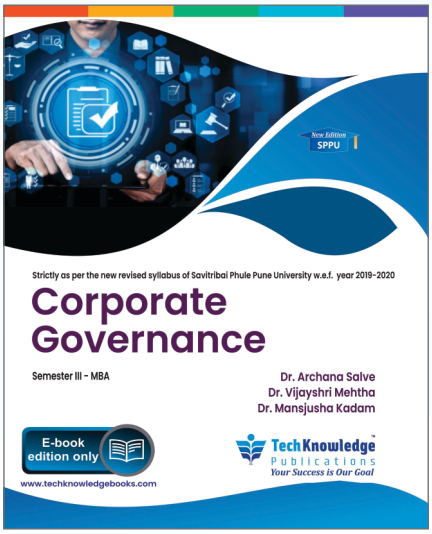
Price : ₹ 170/-

ISBN : 978-81-947867-8-8



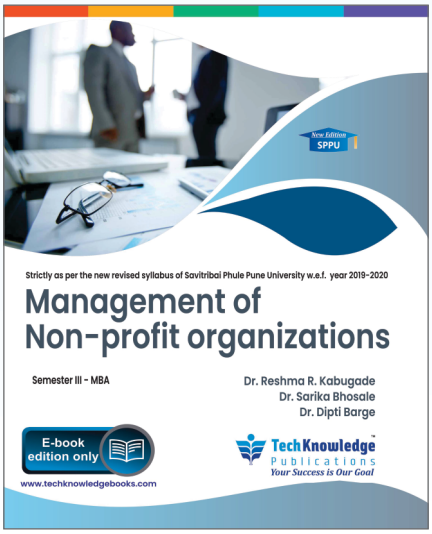
Price : ₹ 190/-

ISBN : 978-81-947867-5-7



Price : ₹ 180/-

ISBN : 978-93-90428-20-5



Price : ₹ 180/-

ISBN : 978-93-90428-28-1

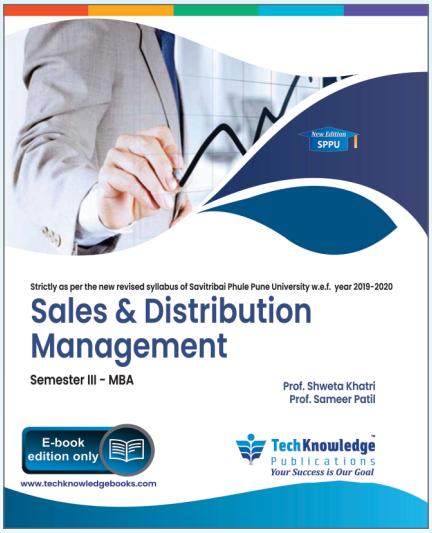


Price : ₹ 190/-

ISBN : 978-93-90428-19-9

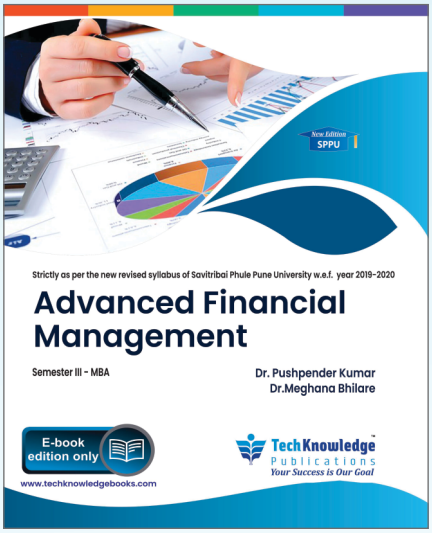
TechKnowledge™ Publications
Your Success is Our Goal

SPPU MBA SEM III e-books Available



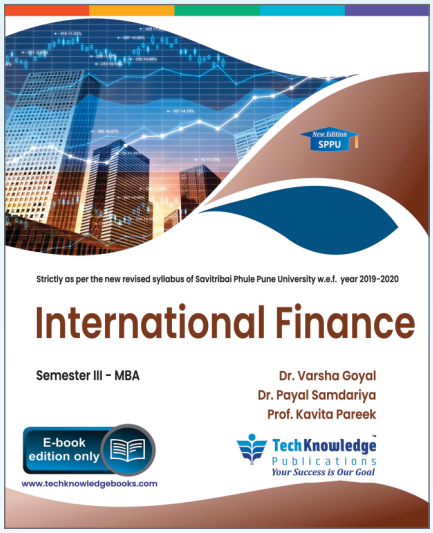
Price : ₹ 180/-

ISBN : 978-81-947867-9-5



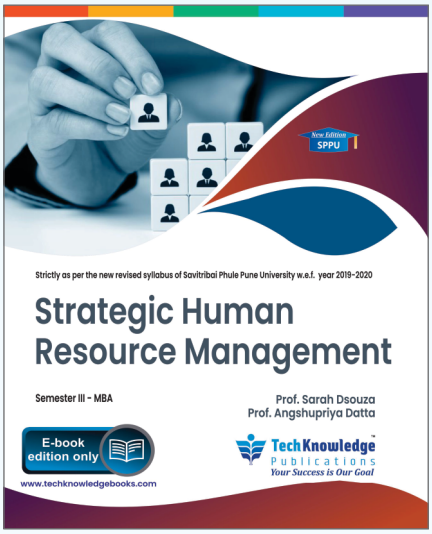
Price : ₹ 180/-

ISBN : 978-81-947867-0-2



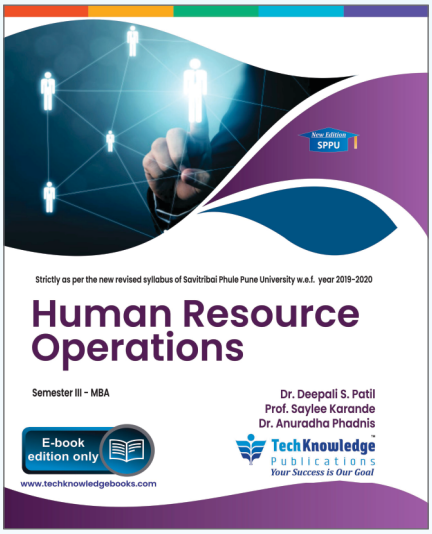
Price : ₹ 250/-

ISBN : 978-81-947867-3-3



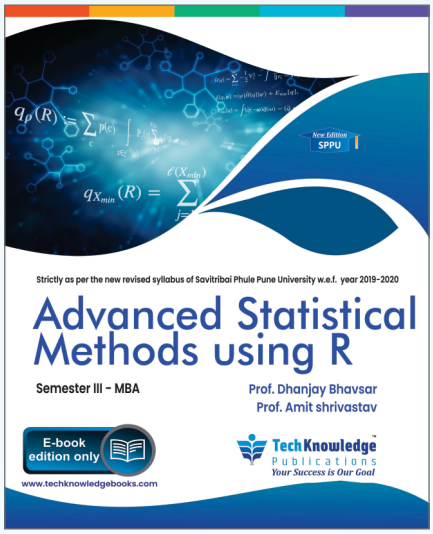
Price : ₹ 180/-

ISBN : 978-93-90428-35-9



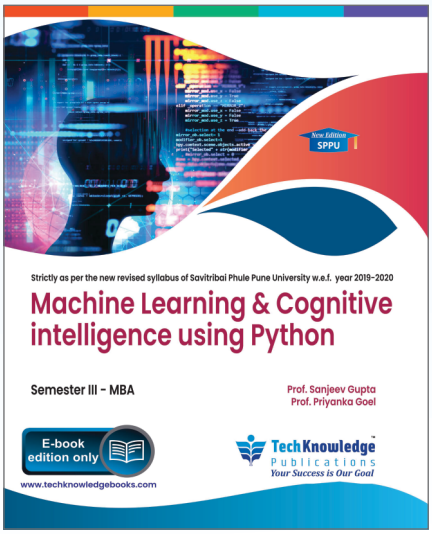
Price : ₹ 170/-

ISBN : 978-93-90428-25-0



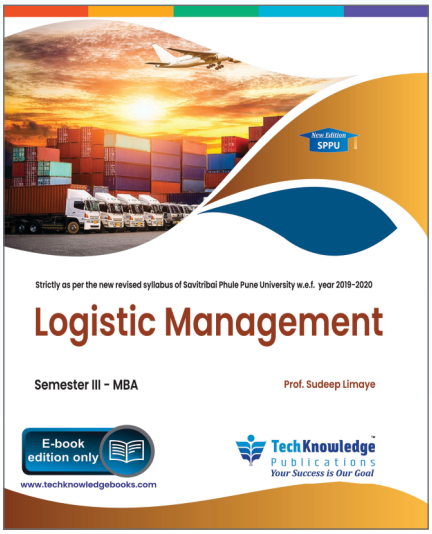
Price : ₹ 200/-

ISBN : 978-81-947867-1-9



Price : ₹ 220/-

ISBN : 978-81-947867-7-1



Price : ₹ 150/-

ISBN : 978-81-947867-6-4

Advanced Statistical Methods Using R

(Code : SC–BA-03)

Semester III – Master of Business Administration (MBA)
(Savitribai Phule Pune University)

Strictly as per the New Revised Syllabus (Rev. 2019) of Savitribai Phule Pune University w.e.f. academic year 2020-21 (As per Choice Based Credit System)

Mr. Dhananjay Narayan Bhavsar

BSc. (Physics), MCM , MCA, MBA (HR) ,
Ph D(pursuing)

Mr. Amit Shrivastava

M.Phil (computer Science)MCA ,Ph D(pursuing)

Assistant Professor

Dr. D.Y. Patil Institute of Management & Research
Pimpri- Pune

TechKnowledge™ Publications

ONPOMBA15A Price ₹ 200/-

(Book Code : ONPOMBA15A

Advanced Statistical Methods Using R

Mr. Dhananjay Narayan Bhavsar, Mr. Amit Shrivastava

Semester III – Master of Business Administration (MBA) (SPPU)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Edition : December 2020

This edition is designed for Savitribai Phule Pune University but can be globally purchased.

ISBN : 978-81-947867-1-9

Published by
TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex,
Taware Colony,Aranyeshwar Corner,
Pune - 411 009. Maharashtra State,
India. Ph : 91-20-24221234, 91-20-24225678.

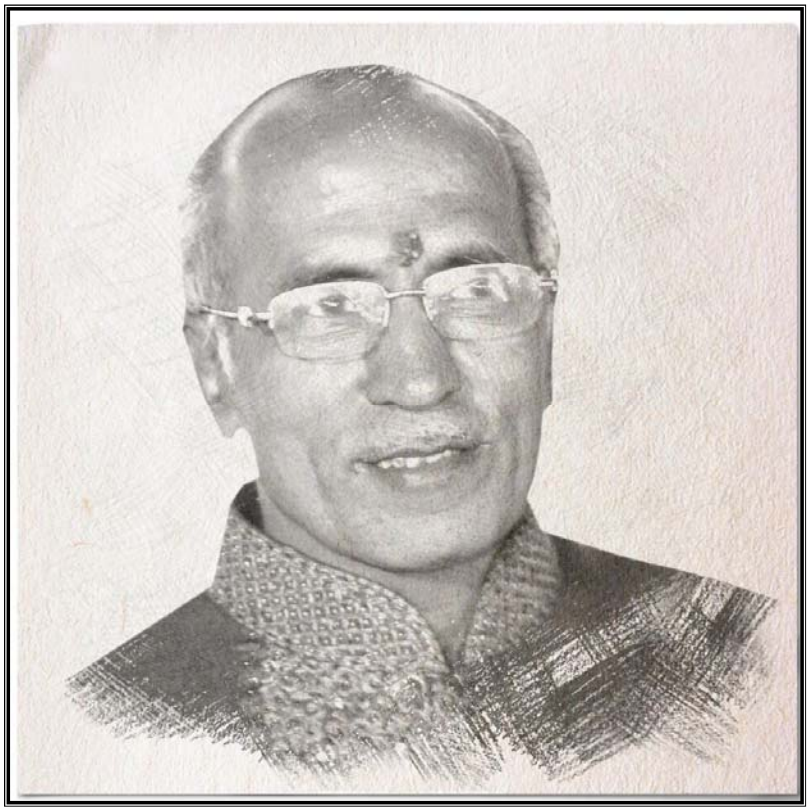
[Code : SC–BA-03] (FID : ONPOMBA15) (Book Code : ONPOMBA15A)

We dedicate this Publication soulfully and wholeheartedly,

in loving memory of our beloved founder director,

Late Shri. Pradeepji Lalchandji Lunawat,

who will always be an inspiration, a positive force and strong support behind us.



"My work is my prayer to God"

– Lt. Shri. Pradeepji L. Lunawat

Soulful Tribute and Gratitude for all Your Sacrifices, Hardwork and 40 years of Strong Vision…

Preface

Dear Students,

We are extremely happy to present the book of **"Advanced Statistical Methods Using R"** for you. We have divided the subject into small chapters so that the topics can be arranged and understood properly. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late. Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of **"TechKnowledge Publications".** He will always be remembered in our hearts and motivate us to achieve our new milestone.

We are thankful to Seema P. Lunawat, Vaishali Nisargand and Dr. Deepali Patil for the encouragement and support that they have extended.We are also thankful to the staff members of TechKnowledge Publications and others for their efforts to make this book as good as it is. We have made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve the book quality further.

We are thankful to my family members and friends for their patience and encouragement.

- Authors

❑❑❑

(Book Code : ONPOMBA15A

About the Authors

Mr. Dhananjay Narayan Bhavsar

Assistant Professor

BSc. (Physics), MCM, MCA, MBA (HR), Ph D(pursuing)

An Assistant Professor having 12 years of teaching experience in various Management Institutes affiliated to S.P. Pune University. His areas of interest include Web technology & Data Analysis. He has published 7 research papers in International and National Conferences & got Best Paper Presentation award in 2009 MAMI conference. He feels that, IT and Modern Management education have a Problem of mass obsolescence as today's knowledge becomes obsolete the next day. We need to gear up to face those challenges and overcome them with effective global solution knowledge.



Mr. Dhananjay Narayan Bhavsar

Mr. Amit Shrivastava

Assistant Professor

M.Phil (Computer Science) MCA, Ph D(pursuing)

An Assistant Professor, in Dr. D.Y. Patil Institute of Management & Research Pimpri- Pune, from last 11 years. He has completed MCA from RGTU, Bhopal and also completed M.Phil in computer Science from Alagappa University, Karaikudi. He has participated in Smart India Hackathon 2019 as well as in Aviskar SPPU Pune as a mentor. currently he is pursuing PhD from Bharti Vidyapeeth in Computer Application.



Mr. Amit Shrivastava

❑❑❑

Syllabus

Semester III		304 – Advanced Statistical Methods using R
3 Credits	LTP: 2:1:1	Subject Core (SC) Course – Business Analytics

Course Outcomes: On successful completion of the course the learner will be able to

CO#	COGNITIVE ABILITIES	COURSE OUTCOMES
CO 304BA .1	Remembering	RECALL all basic statistical concepts and associated values, formulae.
CO 304BA .2	Understanding	EXPLAIN the statistical tools and DESCRIBE their applications in multiple business domains and scenarios
CO 304BA .3	Applying	APPLY time series analysis in prediction of various trends.
CO 304BA .4	Analysing	DISCRIMINATE between various types of probability and probability distributions.
CO 304BA .5	Evaluating	FORMULATE and TEST hypothesis using tools of R.
CO 304BA .6	Creating	COMBINE various tools and functions of R programming language and use them in live analytical projects in multiple business domains and scenarios

1. **Statistics with R :** Computing basic statistics, Business Hypothesis Testing concepts, Basics of statistical modeling, Logistic Regression, Comparing means of two samples, Testing a correlation for significance, Testing a proportion, t test, z Test, F test, Basics of Analysis of variance (ANOVA), One way ANOVA, ANOVA with interaction effects, Two way ANOVA, Summarizing Data, Data Mining Basics, Cross tabulation. Case studies in different domains- using R. **(7+2)**

2. **Linear Regression :** Concept of Linear regression, Dependency of variables, Ordinary Least Sum of Squares Model, Multiple Linear Regression, Obtaining the Best fit line, Assumptions and Evaluation, Outliers and Influential Observations, Multi-collinearity, Case studies in different domains- using R. Dimension Reduction Techniques – Concept of latent dimensions, need for dimension reduction, Principal Components Analysis, Factor Analysis. Case studies in different domains- using R. **(7+2)**

3. **Probability :** Definition, Types of Probability, Mutually Exclusive events, Independent Events, Marginal Probability, Conditional Probability, Bayes Theorem. Probability Distributions – Continuous, Normal, Central Limit theorem, Discrete distribution, Poison distribution, Binomial distribution. **(7+2)**

4. **Predictive Modeling :** (a) Multiple Linear Regression: Concept of Multiple Linear regression, Step wise Regression, Dummy Regression, Case studies in different domains- using R (b) Logistic regression: Concept of Logistic Regression, odds and probabilities, Log likelihood ratio test, Pseudo R square, ROC plot, Classification table, Logistic regression & classification problems, Case studies in different domains using R (c) Linear Discriminant Analysis: Discriminant Function, Linear Discriminant Analysis, Case studies in different domains- using R **(7+2)**

5. **Time Series :** Time Series objects in R, Trends and Seasonality Variation, Decomposition of Time Series, autocorrelation function (ACF) and partial autocorrelation (PACF) plots, Exponential Smoothing, holts Winter Method, Autoregressive Moving Average Models (ARMA), Autoregressive Integrated Moving Average Models (ARIMA), Case studies in different domains- using R. **(7+2)**

❑❑❑

Index

Unit 1

Chapter 1 : Statistics with R **1-1 to 1-28**

1.1 Statistics ... 1-2

1.1.1 Sorts of Statistics ... 1-2

1.1.2 Variables ... 1-2

1.1.3 R - Mean, Median and Mode ... 1-3

1.2 Business Hypothesis Testing concepts ... 1-5

1.2.1 Sorts of Hypothesis ... 1-6

1.2.2 Sorts of Statistical Hypotheses ... 1-6

1.3 Basic Concepts Concerning Testing of Hypotheses ... 1-7

1.4 Procedure of Hypothesis Testing ... 1-8

1.5 Important Parametric Tests ... 1-9

1.5.1 z-test ... 1-9

1.5.2 t- test ... 1-10

1.5.3 F-test ... 1-10

1.6 Basics of Analysis of variance (ANOVA), ... 1-10

1.7 ONE-WAY ANOVA ... 1-11

1.8 TWO-WAY ANOVA ... 1-14

Unit 2

Chapter 2 : Linear Regression **2-1 to 2-43**

2.1 Linear Regression ... 2-2

2.1.1 Coefficients. ... 2-4

2.1.2 Residuals ... 2-5

2.1.3 Detect Influential Points. ... 2-10

2.2 Multiple (Linear) Regression ... 2-12

2.2.1 Relative Importance ... 2-14

2.2.2 Regression Diagnostics ... 2-15

2.2.3 ANOVA ... 2-17

2.2.4 MANOVA ... 2-19

2.3 Dimensionality reduction Techniques: ... 2-28

2.3.1 Principal Component Analysis ... 2-28

2.3.2 Factor Analysis ... 2-33

Unit 3

Chapter 3 : Probability **3-1 to 3-32**

3.1 Definition of Probability ... 3-2
3.1.1 EVENT COMPOSITION ... 3-3
3.1.2 PROBABILITY AXIOMS AND RULES ... 3-3
3.1.3 INDEPENDENT AND DEPENDENT EVENTS ... 3-5
3.2 Conditional Probability ... 3-6
3.3 Bayes Theroem ... 3-7
3.4 Probability Distribution ... 3-16
3.4.1 Poisson Distribution ... 3-17
3.4.2 Binomial Distribution in R ... 3-20
3.4.3 Normal Distribution ... 3-23

Unit 4

Chapter 4 : Predictive Modeling **4-1 to 4-31**

4.1 Predictive Modelling ... 4-2
4.1.1 Steps to apply the multiple linear regression in R ... 4-2
4.1.2 Assumptions for Regression Model: ... 4-2
4.2 Multiple Linear Regression ... 4-8
4.2.1 Dummy Variables ... 4-8
4.2.2 Model accuracy assessment ... 4-11
4.3 Logistic Regression ... 4-12
4.3.1 Likelihood Ratio Tests ... 4-14
4.3.2 ROC Curves ... 4-15
4.3.3 Classification_table: Binary Logistic Regression: Classification Table ... 4-17
4.3.4 Case studies using R : ... 4-18
4.4 Linear Discriminant Analysis ... 4-20
4.4.1 Quick start R code ... 4-21
4.4.2 Case Study : Wine Classification ... 4-24

Unit 5

Chapter 5 : Time Series **5-1 to 5-52**

5.1 Time Series and Forecasting ... 5-2
5.2 Trends and Seasonality Variation ... 5-4
5.3 Introduction to time series ... 5-6
5.3.1 Electrical equipment manufacturing ... 5-7
5.3.2 Seasonally adjusted data ... 5-8
5.4 Moving averages ... 5-9

5.4.1 Moving average smoothing 5-10
5.4.2 Moving averages of moving averages 5-13
5.4.3 Estimating the trend-cycle with seasonal data 5-14
5.4.4 Weighted moving averages 5-15
5.5 Classical Decomposition 5-16
5.5.1 Comments on classical decomposition 5-18
5.6 X11 decomposition 5-18
5.7 Autocorrelation 5-20
5.7.1 ACF Function 5-21
5.7.2 Partial Autocorrelation Function (PACF) 5-22
5.8 Exponential Smoothing 5-23
5.9 Holt's Method 5-26
5.10 Auto-Regressive Integrated Moving Average 5-29
5.10.1 ARIMA modeling 5-37
5.10.2 Choosing the best model 5-43
5.11 Case Study - 5-45
5.11.1 Case Study on Business Analytics for Healthcare 5-46
5.11.2 Case Study on Insurance customer retention and growth 5-46

❑❑❑

Unit 1
Statistics with R



Syllabus

Computing basic statistics, Business Hypothesis Testing concepts, Basics of statistical modeling, Logistic Regression, Comparing means of two samples, Testing a correlation for significance, Testing a proportion, t test, z Test, F test, Basics of Analysis of variance (ANOVA), One way ANOVA, ANOVA with interaction effects, Two way ANOVA, Summarizing Data, Data Mining Basics, Cross tabulation. Case studies in different domains- using R.

TechKnowledge™ Publications

Unit 2
Linear Regression



Syllabus

Concept of Linear regression, Dependency of variables, Ordinary Least Sum of Squares Model, Multiple Linear Regression, Obtaining the Best fit line, Assumptions and Evaluation, Outliers and Influential Observations, Multi-collinearity, Case studies in different domains- using R. Dimension Reduction Techniques – Concept of latent dimensions, need for dimension reduction, Principal Components Analysis, Factor Analysis. Case studies in different domains- using R.

TechKnowledge™ Publications

Unit 3
Probability



Syllabus

Definition, Types of Probability, Mutually Exclusive events, Independent Events, Marginal Probability, Conditional Probability, Bayes Theorem. Probability Distributions – Continuous, Normal, Central Limit theorem, Discrete distribution, Poison distribution, Binomial distribution.

TechKnowledge™ Publications

Unit 4
Predictive Modeling



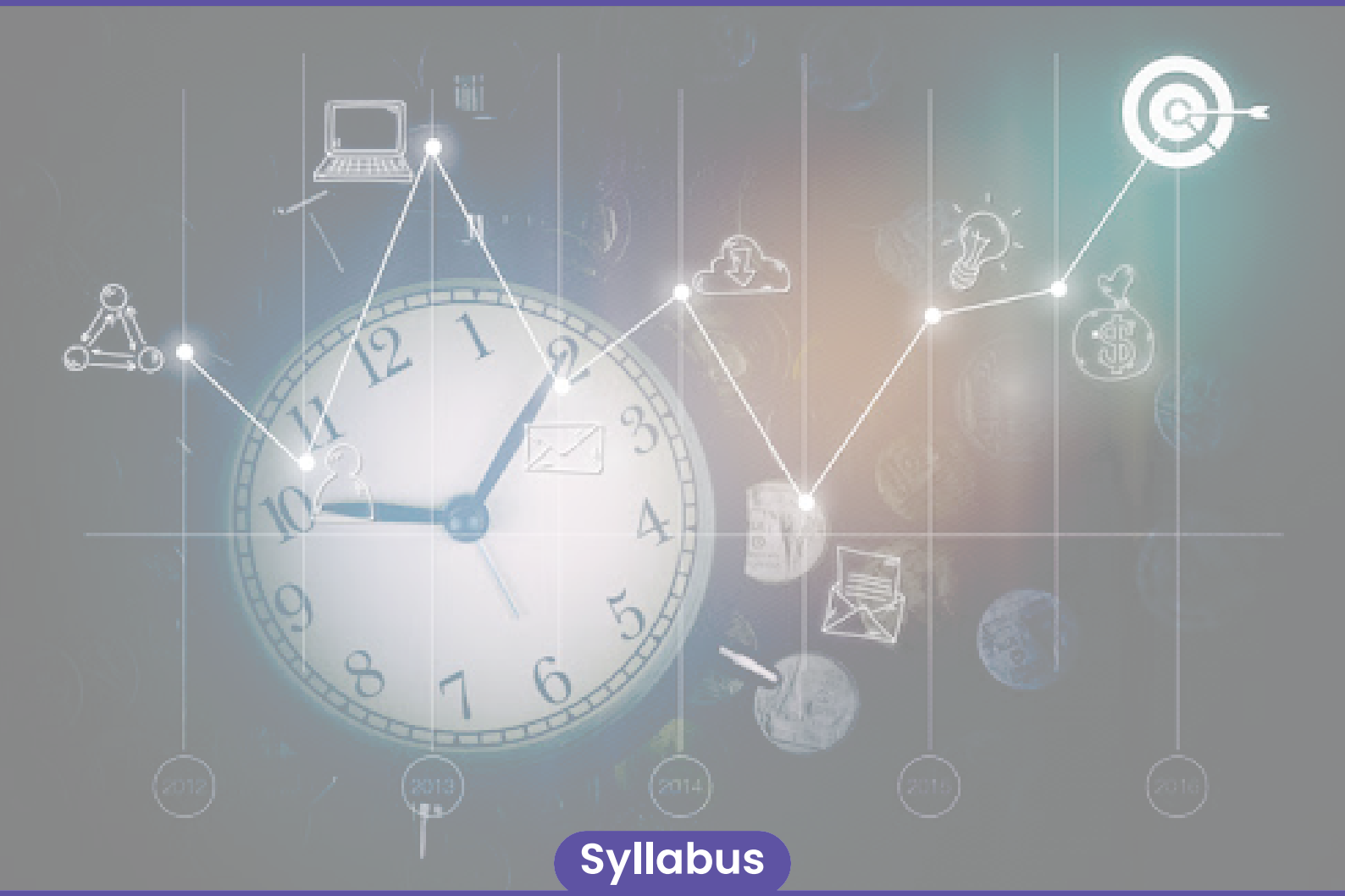
Syllabus

Multiple Linear Regression: Concept of Multiple Linear regression, Step wise Regression, Dummy Regression, Case studies in different domains- using R Logistic regression: Concept of Logistic Regression, odds and probabilities, Log likelihood ratio test, Pseudo R square, ROC plot, Classification table, Logistic regression & classification problems, Case studies in different domainsusing R Linear Discriminant Analysis: Discriminant Function, Linear Discriminant Analysis, Case studies in different domains- using R

TechKnowledge™ Publications

Unit 5

Time Series



Syllabus

Time Series objects in R, Trends and Seasonality Variation, Decomposition of Time Series, autocorrelation function (ACF) and partial autocorrelation (PACF) plots, Exponential Smoothing, holts Winter Method, Autoregressive Moving Average Models (ARMA), Autoregressive Integrated Moving Average Models (ARIMA), Case studies in different domains- using R.

TechKnowledge™ Publications