Indian Institute of Management Indore

INTEGRATED PROGRAMME IN MANAGEMENT (IPM)
AY2017-18 BATCH: 2016-21 TERM: VI

TITLE OF THE COURSE: Multivariate Analysis
CREDITS: 4

Name of the Faculty	Faculty Blok/ Room No.	Email	Telephone Number
Sayantan Banerjee	Block A, Old Conference Room, Ground Floor	sayantanb@iimidr.ac.in	0731 2439 568

CONSULTATION TIME FOR STUDENTS

Name of the Faculty	Timing
Sayantan Banerjee	Th-Fri 11am-12pm

COURSE DESCRIPTION

Modern real-world datasets are becoming increasingly complex, with observations recorded on several variables, leading to the abundance of 'multivariate data'. Understanding the core of these kind of datasets and subsequent analyses has become extremely important in this present scenario. This course will introduce the concept of multivariate data and multiple random variables, and aims to serve as the foundation course for introducing several techniques for multivariate data analysis. The course will be a mix of theory and applications.

COURSE OBJECTIVES

1) Introduction to multivariate data analysis
2) Building theoretical foundations of properties of random vectors and their distributions
3) In-depth treatment of several important multivariate distributions
4) Identification and development of appropriate statistical tools to analyze real-world problems involving multivariate datasets

COURSE REQUIREMENTS

This course requires basic foundation in Calculus, Linear Algebra, Probability and Statistics. It is not mandatory, but it is strongly recommended that this course should be opted for by only those students who are comfortable with the above mentioned topics, and have obtained at least B in the courses related to the topics mentioned above.

PEDAGOGY/TEACHING METHOD:

Mix of lectures, discussions, practical applications using R.

TEXT BOOK FOR THE COURSE

Johnson, R. A., & Wichern, D. W. (2001). *Applied Multivariate Statistical Analysis* (Fifth Edition). Prentice-Hall.

EVALUATION

Individual Component	**Group Component**	**Weightage**
Midterm		30
	Group Project	30
Endterm		40
Total		**100%**

ACADEMIC DISHONESTY

IIM Indore believes in Academic honesty.
Academic dishonesty or misconduct is cheating that relates to an academic activity. It is a violation of trust between the Institute and its stakeholders. Plagiarism, fabrication, deception, cheating and sabotage are examples of unacceptable academic conduct. Please consult the Programme manual for the section on academic dishonesty.

SCHEDULE OF SESSIONS

Module I : Multivariate Data and Multiple Random Variables

Module Objective: Introduction to multivariate data, visualization, and concepts of random vectors

Session 1 **Objective:** **Reading:** **Case:**	Multivariate data and multiple random variables Introduce multiple random variables, discuss problems involving multivariate data, summary statistics, linear combinations, and geometric ideas Lecture notes; JW Chapter 1
Session 2 **Objective:** **Reading:** **Case:**	Random vectors Introduce random vectors: pmf/pdf, cdf, conditional and marginal distributions, population moments, transformations. Lecture notes; JW Chapter 2-3
Session 3 **Objective:** **Reading:** **Case:**	Visualization of multivariate data Graphical displays of multivariate data: panel displays, surface plots, 3D scatter plots, contour plots, other 2D representations Lecture notes; JW Chapter 1.4

Session 4 **Objective:** **Reading:** **Case:**	Multiple linear regression Discuss multiple linear regression, multiple correlation coefficient and partial correlations in the context of multivariate data analysis Lecture notes;
Session 5 **Objective:** **Reading:** **Case:**	Problem solving session Problem solving, with focus on applications with real datasets Lecture notes; problem set prepared by instructor

Module II Multivariate Normal Distribution

Module Objective: Introduction to the multivariate Normal distribution and its properties

Session 6 **Objective:** **Reading:** **Case:**	Multivariate Normal Distribution Introduction to Multivariate Normal distribution: density, geometry, moments, special case of Bivariate Normal Lecture notes; JW Chapter 4.1 – 4.3
Session 7, 8 **Objective:** **Reading:** **Case:**	Multivariate Normal Distribution: Properties Discuss the notion of equiprobability contours, role of correlation coefficient, marginal and conditional distributions, distribution of linear and quadratic forms, Cochran's theorem Lecture notes;
Session 9,10 **Objective:** **Reading:** **Case:**	Sampling from Multivariate Normal distribution Sampling from Multivariate Normal distribution, methods for generating random samples: spectral decomposition method, SVD method, Cholesky factorization method; comparison of performances of different methods Lecture notes;

Session 11 **Objective:** **Reading:** **Case:**	Distributions emerging from multivariate Normal Introduction to Wishart distribution and its properties, Hotelling's T^2 distribution, Mahalanobis distance Lecture notes; JW Chapter 5.3

Module III Statistical Inference for Multivariate Normal Distribution

Module Objective: To carry out statistical inference in the context of multivariate normal populations

Session 12 **Objective:** **Reading:** **Case:**	Estimation and large sample behevious Maximum likelihood estimation – unconstrained case; large sample behavior of sample mean and sample covariance Lecture notes; JW Chapter 4.4 – 4.5
Session 13 **Objective:** **Reading:** **Case:**	Confidence intervals and Hypothesis testing Discuss confidence intervals of mean vector; discuss hypothesis testing: likelihood ratio test for mean and covariance Lecture notes; JW Chapter 5.2 – 5.5
Session 14, 15 **Objective:** **Reading:** **Case:**	Multiple Linear regression Inference in multiple linear regression with Gaussian errors, brief introduction to variable selection in regression Lecture notes; JW Chapter 7
Session 16 **Objective:** **Reading:** **Case:**	Correlations among normal random variables To study multiple correlation coefficient and partial correlations with respect to multivariate Normal distribution, with brief introduction to graphical models Lecture notes;

Module IV Multinomial, Categorical and Dirichlet distributions

Module Objective: Introduction to other important multivariate distributions and their properties

Session 17 **Objective:** **Reading:** **Case:**	Multinomial distribution Introduction to multinomial distribution, marginal and conditional distribution, properties Lecture notes
Session 18 **Objective:** **Reading:** **Case:**	Sampling from multinomial distribution Simulating multinomial random variables, applications, Hardy-Weinberg principle Lecture notes
Session 19 **Objective:** **Reading:** **Case:**	Categorical and Dirichlet distributions Introduce categorical distribution and Dirichlet distribution: properties and applications Lecture notes
Session 20 **Objective:** **Reading:** **Case:**	Problem solving session Problem solving session, brief idea of multivariate data analysis techniques: principal components and factor analysis Lecture notes; problem set prepared by instructor

Please indicate the changes made in the course outline based on the measurement of assurance of learning (closing the loop)/student feedback:

1. Not applicable – Elective course being offered for the first time.

Please give the details of the book if students need to buy the book

Author	**Title**	**Publisher**	**Edition**	**Remarks, if any**

Additional Readings

The following books are recommended for supplementary reading:

1. Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis* (Third Edition). Wiley.

******