Indian Institute of Management Ahmedabad

Statistical Methods and Data Analysis.

PGP –II, Slots 9-10.

Instructor: Prof. Karthik Sriram

Contact email karthiks@iimahd.ernet.in , phone extn: 4875

Course Summary and Objective:

Companies and regulators alike are placing more and more emphasis on objective decision making based on data. Strategic decisions of companies depend on findings from data analysis tasks such as estimating market share of a product, deciphering the latent reasons underlying a consumer's basket of product preferences, prediction of defaults on a bank's loan portfolio etc. Therefore, design, collection and analysis of data have been playing an increasingly crucial role in decision making.

This course introduces participants to some modern statistical techniques in data analysis. While the course is primarily application focused, it will provide adequate theoretical background so that participants can confidently formulate and apply these techniques in real life situations.

Session wise list of topics is provided below.

Pre-requisites

a) Good knowledge of concepts of probability and statistics introduced on PS-1, PS-2 and PS-3 courses in PGP 1st year.
b) Knowledge of mathematics at graduation level is desirable.
c) Familiarity with computer programming in any language is desirable.

Software

a) Participants need to learn soft wares R, Excel and mini-tab on their own.
b) The R software is freely downloadable from http://cran.rproject.org
c) It is expected that participants will use R, Excel or Minitab for solving most of the assignments.
d) It is expected that participants bring their laptop to class for working on data analysis problems.

Evaluation: Quizzes: 30% , Projects: 40%, Assignments: 30% .

References

1. (SDA) Lohr,S. L. (1999), Sampling Design and Analysis, Duxbury Press
2. (AMD) James Lattin, J. Douglas Carroll, Paul E. Green (2003), Analysing Multivariate Data.
3. (CDA) Alan Agresti (2007), An Introduction to Categorical Data Analysis.
4. (OCDA) Alan Agresti (2007), An Introduction to Ordinal Categorical Data Analysis.
5. (GLM) Piet De Jong and Gillian Z. Heller (2008), Generalized Linear Models for Insurance Data.
6. (CART) Brieman, Friedman, Olshen, Stone (1984), Classification and Regression Trees.

Note: Relevant readings for the sessions will be provided to the students.

Session 1: Introduction, Simple Random Sampling

Ref: Chapters 1 and 2 of SDA. (Part 1 material: pages 1 to 32)

Session 2: Ratio and Regression Estimation

Ref: Chapter 3 of SDA (Part 1 material: pages 33 to 50)

Session 3: Stratified Sampling

Ref: Chapter 4 of SDA (Part 1 material: pages 51 to 69)

Session 4: Cluster Sampling with Equal Probabilities

Ref: Chapter 5 of SDA (Part 1 material: pages 71 to 95)

Session 5: Sampling with Unequal Probabilities

Ref: Chapter 6 of SDA (Part 1 material: pages 97 to 117)

Session 6: Sampling with Unequal Probabilities

Probability Proportional to Size Sampling

Ref: Chapter 6 of SDA (Part 1 material: pages 97 to 117)

Session 7 Complex Surveys

Ref: Chapter 7 of SDA (Part 1 material: pages 119 to 135, 155 to 170)

Session 8 Descriptive Analysis of Multivariate Data

(Part 1 material: pages 171 to 184)

Session 9 Multivariate Normal Distribution

(Part 1 material: pages 171 to 184)

Session 10 Principal Component Analysis

Ref: Chapter 4 of AMD (see corresponding chapter in Part 2 of material)

Session 11 Factor Analysis

Ref: Chapter 5 of AMD (see corresponding chapter in Part 2 of material)

Session 12 Cluster Analysis and Multidimensional scaling

Ref: Chapter 7, 8 of AMD (see corresponding chapter in Part 2 of material)

Session 13	Cluster Analysis and Multidimensional scaling
	Ref: Chapter 7, 8 of AMD (see corresponding chapter in Part 2 of material)
Session 14	Quiz 1
Session 15	Presentation of Project on Sampling
Session 16	Discriminant Analysis: 2 groups
	Ref: Chapter 12 of AMD (see corresponding chapter in Part 2 of material)
Session 17	Discriminant Analysis: Multiple groups
	Ref: Chapter 12 of AMD (see corresponding chapter in Part 2 of material)
Session 18	Logistic Regression
	Ref: Chapter 4 of CDA (Part 1 material: pages 185 to 204)
Session 19	Building and Applying Logistic Regression
	Ref: Chapter 5 of CDA (Part 1 material: pages 205 to 224)
Session 20	Multi-category Logit Models
	Ref: Chapter 6 of CDA (Part 1 material: pages 225 to 240)
Session 21	Generalized Linear Models – Responses as Counts.
	Ref: Chapter 6 of GLM (Part 1 material: pages 241 to 250)
Session 22	Generalized Linear Models – Continuous Responses.
	Ref: Chapter 8 of GLM (Part 1 material: pages 251 to 255)
Session 23	Classification and Regression Trees
	Ref: chapters 1 and 2 of CART (Part 1 material: pages 257 to 287)
Session 24	Quiz 2
Session 25	Presentation on Data Analysis