3.17 E012 Statistical Models and Methods (Proposed elective for approval)

Credits. 1, 2, 3, 4, OR 5

Prerequisites. C206 (2.18) or equivalent, or as defined by the instructor/s.

Dependent Courses. None

Attributions. CBC; C, L, S

Rationale, Outlook, Purpose, Objectives, and Goals. This course is a flexible-credit choice-based/elective course with focus on advanced statistical methods and topics. A student can take this course for any integer number of credits between 1 and 5. Listed below in the syllabus section is a representative selection of topics rated at 1 credit each. With approval from the Centre, a qualified instructor may introduce additional 1-credit topics. Any selection of topics adding up to the requisite number of credits may be offered. The flexible-credit structure of the course is intentional; this will give enough freedom and flexibility for a student to (a) choose courses and topics of her/his interest, and (b) allow her/him to satisfy the requirement on the total number of choice-based credits.

Syllabus.

1. Linear and Logistic Regression. Simple Linear Regression. Least Squares and Maximum Likelihood. Properties of the Least Squares Estimators. Prediction. Multiple Regression. Model Selection. Logistic Regression.
2. Multivariate Models. Random vectors. Estimating the correlation. Multivariate normal. Multinomial.
3. Inference about Independence. Two binary variables. Two discrete variables. Two continuous variables. One continuous variable and one discrete variables.
4. Causal Inference. The counterfactual model. Beyond binary treatments. Observational studies and confounding. Simpson's paradox.
5. Directed Graphs and Conditional Independence. Conditional independence. Directed graphs. Probability and directed graphs. More independence relations. Estimation for directed graphs.
6. Undirected Graphs. Undirected graphs. Probability and graphs. Cliques and potentials. Fitting graphs to data.
7. Log-Linear Models. The log-linear model. Graphical log-linear models. Hierarchical log-linear models. Model generators. Fitting log-linear models to data.
8. Nonparametric Curve Estimation. The bias-variance trade-off. Histograms. Kernel density estimation. Nonparametric regression.
9. Smoothing Using Orthogonal Functions. Orthogonal functions and L_2 spaces. Density estimation. Regression. Wavelets.
10. Classification. Error rates and the Bayes classifier. Gaussian and linear classifiers. Linear regression and logistic regression. Relationship between logistic regression and linear discrimination analysis. Density estimation and naive Bayes. Trees. Assessing error rates and choosing a good classifier. Support vector machines. Kernelization. Other classifiers.
11. Time Series Analysis. Overview of the Box-Jenkins and Bayesian approaches. Principles of nonlinear and chaotic time series analysis.

Suggested Texts/References.

1. Larry Wasserman, *All of Statistics*. Springer-Verlag, 2004.
2. Larry Wasserman, *All of Nonparametric Statistics*. Springer-Verlag, 2006.

Notes on Pedagogy. The suggested syllabus is primarily the content of Part III of the first recommended reference book. Appropriate balance of theory and practice is recommended.

Contributor/s. Mihir Arjunwadkar (`http://cms.unipune.ac.in/~mihir`)