
Decision-making statistical methods

Finalità

Part one is a thorough and advanced revision of the topics presented in the basic Statistics course. We introduce new classes of tests to deal with more complex and realistic settings. Part two is a not-too-quick glance of today's analysis and management techniques that belongs to statistics and explorative data analysis.

Programma

PART ONE: BASIC MULTIVARIATE TOOLS.

Revision on random variables and statistical inference.

Classical Z, T and F tests for comparing parameters for two normal populations.

Adaptation and independence tests (Fisher-Irwin, chi-square, contingency tables).

Regression: coefficients determination (linear and multilinear models, linearization; coefficient of determination, analysis of residuals, weighted min-squares); inference on coefficients (T and F tests)

Analysis of variance (one-way, two-ways and with interactions).

SECOND HALF: EXPLORATIVE DATA ANALYSIS

Graphical representation of very large and/or high-dimensional data sets (multivariate gaussian distribution, correlation matrix, eigenvalues and eigenvectors)

Model adaptation (kernel functions, chi-squared test, Kolmogorov-Smirnov test)

Cluster analysis (distances; hierarchical tree clustering, linkage; k-means algorithms; EM algorithms, mixtures of measures, bayesian classification).

Factor analysis (principal component analysis, common factor analysis, variables reduction, factor interpretation, factors rotations).

Discriminant function analysis (Fisher linear methods, variables reduction).

Neural networks (multilayer perceptron).

Overfitting and overlearning: when the model does not fit the population but the sample.

Non-parametric tests (signs, ranked signs, Wilcoxon's, for the independence of the sample).

Bayesian parametric tests (overview).

Attività d'esercitazione

For part one only: practice is done on the PC and shows how to use a spreadsheet to solve selected exercises and to clarify some theoretical topics.

Modalità d'esame

For part one: examination on the PC. For part two: either a seminar on a selected topic (on a deeper level than the lesson) or an analysis on some data set. Depending on the complexity, such work could be taken by a small group of people.

Propedeuticità

Statistica, Analisi AB, Analisi C.

Testi consigliati

S. Ross - Probabilità e statistica per l'ingegneria e le scienze - Apogeo 2003

Hand, Mannila, Smyth - Principles of data mining.