

J.Špitálský, M. Vošvrda
Czechoslovak Academy of Sciences,
Institute of Information Theory and Automation,
Prague, Czechoslovakia

Abstract. STADIA is an opened interactive software system for solving scientific and engineering computation tasks. The system was worked out to advance effectiveness of the work of both experts for whom it makes mathematical data processing easy, and programmers for whom it makes it possible to connect (in an easy standard way) methods of application (without any additional connected software modification) to the general system that ensures very natural friendly communication, data handling and execution of methods.

The dialogue of the system (consisting commands, diagnostic and informative messages) is very natural and easily learnable. The possibility to create macrocommands as a choice from the system commands allows easy customization to a user's specific problem or application.

The system commands control a large software library of both well known and advanced techniques aimed at solving mathematical statistics, statistical analysis and modelling tasks. STADIA can handle large data sets without significant loss in through-put time. All data (up to 30.000 real numbers) are loaded into memory for top speed.

The software system STADIA-PC (mostly written in Fortran 77) can be run on PC under operational systems MS DOS 3.2 or greater.

Keywords. Parameter estimation; Time - series analysis; Robustness; Modelling.

INTRODUCTION

Statistical analysis - that is estimating and testing parameters of random subjects and their comparing and interpretation to reach qualitatively new conclusions. There are two extreme approaches to statistical data processing in statistical analysis:

1/ The data analytical approach, where a new reality and circumstances are drawn by an inductive process on the base of data analysis.

2/ The scientific recognition approach, where a hypothetic model of an initial reality is verified and specified by means of statistical files.

We usually find ourselves between these two positions when we process statistical data. In creating computational statistical apparatus, these two approaches are to be combined. The STADIA software system makes the choice and realization of methods which are now used for statistical analysis possible in both these approaches.

The possibility of an interaction between a man and a computer in solving problems of statistical analysis seems to be very important for choosing and combining suitable methods according to evaluated results of processing. The STADIA software system is the particular interactive system.

STADIA-PC's STATISTICAL ANALYSIS

The STADIA system includes working commands (those making the user's activity with the system very comfortable, e.g. the establishment of new methods into the system etc.) and functional commands for data input, data output, statistical estimation, modelling, testing and recognition.

Input data file commands /the key word READ of the initial menu/ assure data input from the keyboard, disc or different transformations of data from a powerful random number generator. The STADIA system enables to read from the disc integer or real, formatted or unformatted data files with direct or sequential access. In case of data modelling the system offers an input of data with uniform, normal, lognormal, exponential, Weibull's, gamma, Student's, Rayleigh's, Rice's, logistic, Cauchy's, Tukey's, Johnson's, chi-square, double - exponential, binomial, Poisson's and geometrical distributions and an autoregressive process.

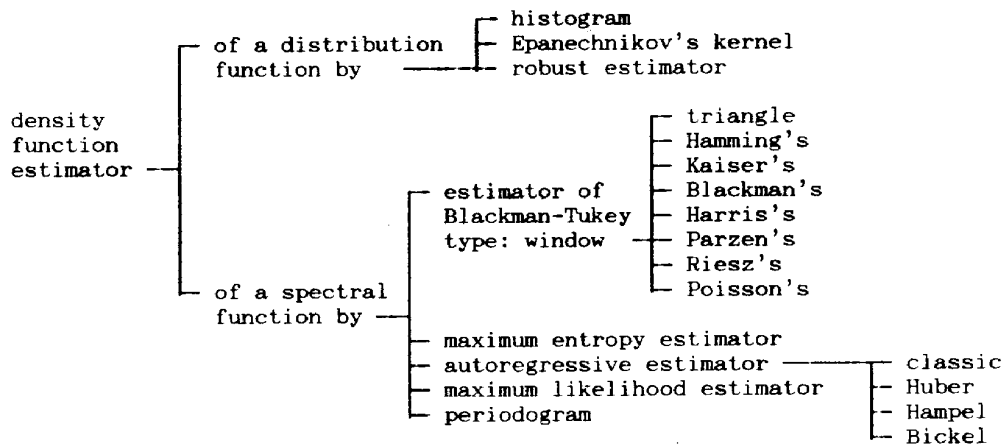
Statistical estimating commands /the key word **COMPUTE** of the initial menu/ assure estimates of location and scale parameters, covariance and correlation structure, density of both distribution and spectral functions, regression coefficients, model parameters, computation of critical and distribution function values of a given distribution and useful arithmetical computation (operations) over data files:

Estimate of a location parameter can be found by sample mean, Takeutchi's adaptive location estimator, Jaeckel's estimator, adaptive minimum divergence estimator or by sample median.

Estimate of a scale parameter may be executed by sample variance, adaptive minimum divergence estimator or Downton's estimator.

An estimating of a covariance (a correlation) structure may be executed by sample covariance (correlation) coefficients - mutual, multiple or partial.

An estimating of a density function has the following framework:



An estimating of regression coefficients may be executed by both methods of a linear regression and a method of regression quantile (Bassett-Koenker's estimator).

An estimating of model parameters may be executed by hidden periodicities method (Fisher's test).

A filtration of the given realization may be executed by Fisher's method or by robust methods (Huber, Hampel or Bickel).

A computing of critical values may be executed both for normal, chi-square and Student's distribution and for an empirical distribution (in this case by means of a sample quantile).

A computing of a distribution function values may be executed for normal, chi-square, Student's and Fisher's one.

Cross-spectral characteristics of time series - gain, phase and coherence may be estimated.

Parameters of autoregressive time series are estimated by Jones or Akaike methods, stability of these models can be asserted.

User can estimate likelihood function, error of prediction and change point of an autoregressive realization or an autoregressive class can be found from several autoregressive models.

The **STADIA** contains the following arithmetical operations over files: Product by scalar, shift, square root of vector elements, searching for minimum and maximum of vector elements, resampling, inversion, sum, inner product, collection or discollection of two data files.

Statistical testing commands /the key word **TEST** of the initial menu/ concern with hypothesis of symmetry (in contingency table), outliers (Dixon's test), coincidence and independence:

User can test coincidence hypothesis of mean with a given value (one or two sampled Student's test), variance (Bartlett's test), distribution - chi-square test of goodness of fit and Kolmogoroff - Smirnov's test for one (against a theoretical distribution) or two samples.

Independence hypothesis can be tested either between two realizations by nonparametrical Spearman's or Kendall's test or between two normally distributed realizations, or in contingency table by chi-square or Stuart's test, or by analysis of variance.

Output operations commands /the key word **WRITE** of the initial menu/ assure output of data files in either a digital form to display, printer and disc, or in a graphic form to display, printer or other graphic peripherals. Graphical output to display has an interactive and dynamical form.

The software system **STADIA** can be run on PC under operational system MS DOS.

CONCLUSION

STADIA represents very efficient purpose-performance ratio for statistics, modelling and data analysis software.

REFERENCE

Havlová, H., J. Špitálský, and M. Vošvrda (1989). STADIA'89 - User's Manual. UTIA ČSAV, Prague.

J. Špitálský, M. Vošvrda: Statistický dialogový systém STADIA, Robust'86, 1986, pp. 125-127.