APPLICATION OF MULTIVARIATE ANALYSIS IN THE PROCESSING OF MEDICAL DATA

M.V. Shin
(Tomsk, National Research Tomsk Polytechnic University)
marishapicke@gmail.com

Abstracts. Medical data frequently represent multidimensional datasets as investigated factors and clinical and laboratory parameters coverage is huge. This research area is very important in terms of practical applications. We were given monthly lipid metabolism and hormonal status data of children (including children suffering from obesity) of Siberian region during a year. In this article some research results appear.

Keywords: multidimensional data, data analysis, statistical methods, obesity, research.

Multivariate methods of data processing. Accumulated in archives medical data is essential for studying of obesity mechanism among children. This data contains plenty of information about various diseases, process of treatment and its complexity and also a value of different clinical and laboratory indexes.

Such numeric arrays of indexes are made as big numeric tables, because the coverage of investigated factors and clinical and laboratory parameters is huge. Specific number of rows (or objects) in these tables can grow up to several tens of thousands and specific number of columns (or features) – up to several hundred [1].

Making statistic information as multidimensional arrays is proven and indexes that are contained in these arrays represent as multidimensional data.

Multivariate data analysis is a field of knowledge in that processes and systems are investigated. These processes and systems are described with a set of values. This research area is very important in terms of practical application, because a lot of huge systems are described with a set of values. Therefore, the problem is to define a set of technical, program and organizational support and staff that could provide medical staff with essential information about patient state in time. In addition, they must have an opportunity to make a research that shows the influence of different factors of patients on analysis parameters state and to define correlations between health activities and clinical and laboratory parameters.

We were given monthly lipid metabolism and hormonal status data of children (including children suffering from obesity) of Siberian region during a year.

Multivariate methods of data processing, tool selection. Methods of multivariate statistical analysis are the math statistical methods that are used for forming the best collection plans, systematization and multivariate statistical data processing. They aimed at definition of nature and structure of relationships between components of investigated multivariate feature. These components are used for getting scientific and practical inference.

There are plenty of programs for statistical data processing. We chose program STATISTICA (StatSoft company) from the most popular programs for statistical data processing. This program implements functions of data analysis, data management, data mining, data visualizing involving statistical methods [2].

Results. Initial data is a selection from 250 patients that are described with more than 15 attributes (there are clinical and laboratory parameters). Initial parameters that were given us as a research result are: arithmetical mean, median, mode, frequency of mode, standard deviation.

In Fig. 1 we can see a graph showing relationships between average value of weight before and after treatment and obesity group (50 patients per group, 5 group total).

Research of parameters relationship. We considered parameters using Pearson's correlation coefficient. It shows whether two parameters are in linear relationship. The correlation analysis in Graph 2 shows reverse relationship between patient's weight before treatment (root1) and cholesterol level after treatment (var6). That means that after treatment cholesterol level decreases, therefore child's weight decreases too.
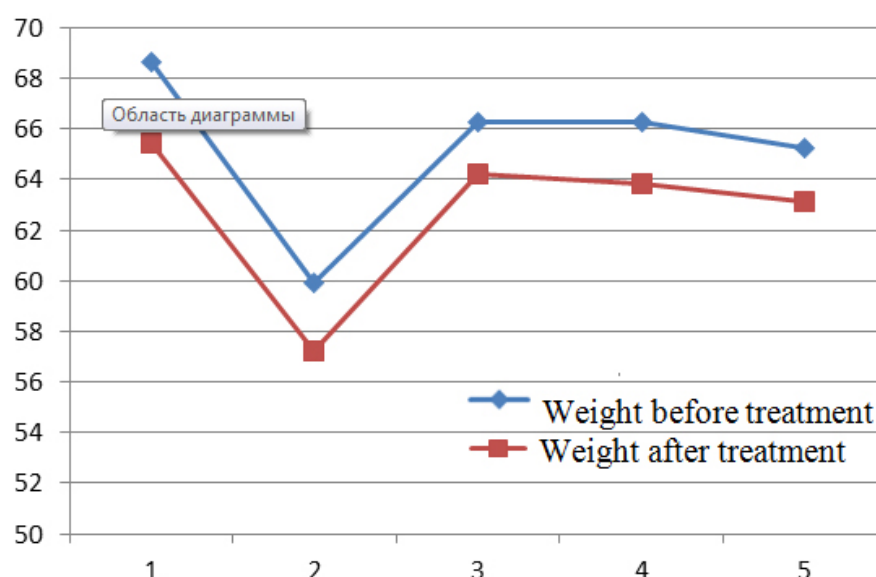
Fig. 1. Relationship between average value of weight before and after treatment and obesity group

Fig. 2. Correlation matrix

Conclusion. Application of multivariate analysis in the processing of medical data is logical for solving problems of object analysis that are described with a plenty of features. In this article it is showed that the solution of obesity problem may be an analysis of clinical and laboratory parameters which are the basis of children's clinical state and may help to find the regularity that can help in treatment.

References

1. Dyuk, Century Data processing on the personal computer in examples [Text] / V. Dyuk. – CPb.: Peter, 1997. – 240 pp.

2. Borovikov V. STATISTICA: Art of the analysis of the data on the computer. For professionals. SPb.: Peter, 2001, 656 p.