9. Photo: UT Austin Available at Mercury - ROBOTS: Your Guide to the World of Robotics (ieee.org)

10. California Institute of Technology. "LEONARDO, the bipedal robot, can ride a skateboard and walk a slackline." ScienceDaily. ScienceDaily, 6 October 2021.Available at LEONARDO, the bipedal robot, can ride a skateboard and walk a slackline -- ScienceDaily

11. Robert Perkins (2021) LEONARDO, the Bipedal Robot, Can Ride a Skateboard and Walk a Slackline Available at LEONARDO, the Bipedal Robot, Can Ride a Skateboard and Walk a Slackline | www.caltech.edu

12. Seong Chiun Lim, Gik Hong Yeap (2012) The Locomotion of Bipedal Walking Robot with Six Degree of Freedom International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012) Procedia Engineering 41 (2012) 8 – 14 Available at The Locomotion of Bipedal Walking Robot with Six Degree of Freedom - ScienceDirect

UDC 519.23

O. Aziukovskyi[1], I. Udovyk[1], A. Kozhevnykov[1], T. Powroźnik[2]
[1]Dnipro University of Technology, Dnepr, Ukraine
[2]Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Kraków, Polska

CREATING USING THE MATHCAD SYSTEM OF LABORATORY EXPERIMENTATION ON THE SUBJECT «INTELLIGENT DATA ANALYSIS»

Annotation. Methodical recommendation for laboratory experimentation on the subject of "Intelligent Data Analysis", based on the MathCAD system, will be briefly described. The experimentation for each laboratory research corresponds to open-source code, transparently related to mathematical models on the topic of the work. The developed experimentation consists of 6 research variations. Input data for each research has 32 options.

Keywords: *intelligent data analysis, mathematical statistics, random values, probabilistic distributions, point and interval estimates, correlation, regression, cluster.*

Introduction. Currently, in the training of IT specialists, one of the main educational components of the professional constituent is the subject "Intelligent Data Analysis". Known approaches to the organization of laboratory experimentation, in this subject, are based on the use of analytical platforms such as WEKA [1, 2] and Deductor Academic [3].

These products are powerful tools for solving practical Data Mining tasks, but they mostly use tools of intelligent data analysis as "black boxes". In addition, laboratory experimentation based on analytical platforms usually don’t contain individual task options.

At the same time, while introducing students to the tools of intelligent data analysis, it is desirable for software implementations of these tools to have open-source transparently related to their mathematical models. Solving such problems, even with the use of specialized languages that are adapted to mathematical calculations, such as MATLAB or Python, not to mention the popular C and Java, is laborious and doesn’t provide transparency of the generated code. The best software in this case, in our opinion, is the system of computer algebra MathCAD [4], or its further development MathCAD Prime [5]. The main difference, between these systems and similar ones, is the availability of graphic templates for entering program sentences. The templates reproduce the symbols used in mathematics, making the program code compact and transparent. MathCAD uses its own scripting language. To organize the simplest cycles, ranked variables are used, more complex algorithms are implemented within program blocks.

Purpose of the research is to create, using MathCAD system methodical recommendation and software for laboratory experimentation on the subject of "Intelligent Data Analysis", in which each laboratory research corresponds to open-source code, transparently related to mathematical models on the topic of the work**.**

The main content of the work. The developed laboratory experimentation consists of 6 research variations. Input data for each research has 32 options. A brief overview, of these research variations, is given below.

Research № 1, 2. *Name:* Probabilistic distributions of discrete and continuous random variables and point estimates of their parameters.

Input data:

- names and formulas for differential theoretical distributions of discrete and continuous random variables;
- formulas: calculation of distribution parameters depending on the option number, mathematical expectations and variances of distributions;
- names of the MathCAD functions used.

Brief content of research:

- determination of the effective range of variation of a random value;
- generation of a random values sample with a given probability distribution;
- plotting of theoretical and empirical differential probability distributions of a random value;
- plotting of theoretical and empirical cumulative probability distributions of a random value;
- determination of random value parameter estimates: mean, variance, asymmetry, excess mode and median of a random variable using built-in MathCAD, according to sample data and empirical distribution;
- plotting of the dependence mean and statistical variance of a random value on the volume of the sample.

Typical conclusions:

- polygons (plots) of differential theoretical probability distributions of a random value, constructed using the distribution formulas and the built-in MathCAD function matches;
- polygons (plots) of differential and cumulative theoretical probability distributions of random value matches with the corresponding polygons (histograms) of empirical distributions within the statistical error;
- the values of random value statistical parameters (mean, variance, asymmetry, excess, and median), calculated using the built-in MathCAD functions and calculated based on sample data matches, and calculated based on empirical distribution differ slightly from them;
- as the sample size increases, the values of the statistical parameters of the random value (mean and variance) are directed, respectively, to the values of the mathematical expectation and variance of the general population.

Research № 3. *Name:* Special functions of mathematical statistics, interval estimates of random values parameters and tests to verification statistical hypotheses

Input data are the data sets for research variations № 1, 2.

Brief content of research:

- plotting of the dependencies of the confidence interval of the mathematical expectation on the confidence probability using the inverse cumulative functions of the normal distribution and the Student's distribution;
- generation array of random value samples with given probability distributions;
- determining the estimations of the confidence probabilities of a random value mathematical expectation and the variance confidence intervals using the random trials method. Comparison of the obtained estimations with the confidence probability for which confidence intervals were calculated;
- calculating mean and variance for the sum large number of identically distributed random values;
- plotting of theoretical and empirical differential probability distributions for the mean large number of identically distributed random values;
- compliance check between theoretical and empirical differential probability distributions for the mean large number of identically distributed random values using by χ^2 criterion.

Typical conclusions:

- confidence interval increases with increasing confidence probability and decreases with increasing sample size;
- with a known variance of the general population and other identical conditions, the values of the confidence interval for the mathematical expectation are less than if the variance is calculated from the sample;
- the probabilities estimation for falling in the confidence intervals mathematical expectation and variance determined by the method of statistical trials differ slightly from the confidence probabilities;

- the mean and statistical variance of the empirical probability distributions for the mean large number of identically distributed random values coincide, respectively, with the mathematical expectation and variance of the theoretical distribution;
- the empirical distribution for the mean large number of identically distributed random values visually corresponds to the normal distribution. The χ^2 test also shows that the empirical distribution for the mean large number of identically distributed random values corresponds to the normal distribution.

Research № 4. *Name:* Correlation analysis.

Initial data is a subsample of an appendicitis diagnostics results sample. The sample is a data of observation matrices which presented in the ordinal scale. Matrices are given in two forms: with ordinary ordinal data and with rank data. Each row of the matrix corresponds to the patient. The columns of the matrices contain the codes of the resultant sign – the clinically confirmed absence of appendicitis or its morphological form and 8 independent signs – the values of the diagnostic attributes.

Brief content of research:

- formation of a data subsample from a sample in accordance with the option;
- calculating matrices of pairwise Pearson's correlation coefficients and vectors of multiple correlation coefficients for ordinary ordinal data and for rank data;
- checking the significance of the minimum in modulus paired and minimum multiple correlation coefficients;
- calculating matrices of Spearman's rank correlation coefficients and Kendall's rank correlation coefficients for ordinary ordinal data and for rank data with and without tied ranks correction.

Typical conclusions:

- obtained correlation indices vary within the relevant ranges of values: Pearson's coefficient and Spearman's and Kendall's coefficients with and without tied ranks corrections from -1 to 1. The multiple correlation coefficient varies from 0 to 1;
- all correlation matrices are symmetric, on their main diagonals are ones;
- the resultant sign has high correlation coefficients with independent signs, which indicates the correct diagnostic (independent) attributes choice;
- Pearson's coefficients have similar values for the input data presented in ordinary ordinal and rank units. The Kendel's coefficients give the same values in these cases. To calculate the Spearman's coefficients, only rank values can be used;
- Spearman's coefficients, with/without corrections for tied ranks, differ significantly due to the large number of tied ranks. The values of the Pearson's and the Spearman's coefficients with corrections for tied ranks coincide. The calculation of Kendall's coefficients is possible only with the corrections for tied ranks;
- the minimum in modulus paired and minimum multiple correlation coefficients are not significant.

Research № 5. *Name:* Regression analysis.

Initial data is a nonlinear regression model and a subsample of the quantitative data sample. Each item, of the sample, contain resultant and one independent sign.

Brief content of research:

- formation of a data subsample from a sample in accordance with the option;
- composing of normal equations systems with respect to unknown coefficients estimates for linear and nonlinear regression;
- calculation of coefficients estimates for linear and nonlinear pairs regression by solving of normal equations systems;
- calculation of values for total, regression and residual variances and coefficients of determination for linear and nonlinear regression;
- plotting in a common field of the subsample initial data, graphs of linear and nonlinear regression;
- checking the significance of the linear regression equation and its coefficients.

Typical conclusions:

- the result of plotting in a common field of the subsample initial data, graphs of linear and nonlinear regression shows good match between them. Also graphics shows that nonlinear regression is better than linear to approximation of initial data;
- better quality of nonlinear regression is confirmed by the results of calculations which show that nonlinear regression has a lower residual variance and a higher coefficient of determination;
- the linear regression equation is statistically significant, as are both of its coefficients.

Research № 6. *Name:* Cluster analysis.

Initial data is the data set for research variation № 4.

Brief content of research:

- formation of a data subsample from a sample in accordance with the option;
- clustering of the initial data by the method K-means;
- calculation of coefficients of pair correlation of diagnostic (independent) attributes with the result – the true diagnosis;
- determination of the cluster number corresponding to an unconfirmed diagnosis of appendicitis. For this cluster, the diagnostic attribute, most correlated with the resultant one, has the minimum value;
- calculation of probabilities estimates for 1: (a healthy patient belongs to one of the clusters with diagnosed appendicitis) and 2: (a patient who has appendicitis belongs to a cluster with an unconfirmed diagnosis) kind errors;
- clustering of the initial data by the method fuzzy C-means;
- determination of the cluster number corresponding to an unconfirmed diagnosis of appendicitis;
- calculation of probabilities estimates for 1: (a healthy patient belongs to one of the clusters with diagnosed appendicitis) and 2: (a patient who has appendicitis belongs to a cluster with an unconfirmed diagnosis) kind errors;
- plotting the dependence of the average variance of the cluster's number on the exponential weighting coefficient.

Typical conclusions:

- clustering of the patients diagnosis results with suspected appendicitis, which were performed by the methods of K-means and Fuzzy C-means (for exponential weighting coefficient w=1,5) show that the chosen diagnostic attributes system allows to clearly distinguish clinically unconfirmed cases and cases of appendicitis;
- the probabilities estimates for the 1-st kind errors is equal zero and the 2-nd kind consist of no more than 0,1;
- the dependence of the average variance of the cluster's number on the exponential weighting coefficient in the range of argument values (1.1. – 2.0) has monotonically increasing character.

Conclusions. This paper has briefly described methodical recommendations for laboratory experimentation on the subject of "Intelligent Data Analysis" based on the MathCAD system. The experimentation for each laboratory research corresponds to open-source code, transparently related to mathematical models on the topic of the work. The developed experimentation consists of 6 research variations. Input data for each research has 32 options. Methodical recommendations and appropriate software are posted on the distant education platform of Dnipro University of Technology, Department of Computer System's Software [6].

REFERENCES

1. Eibe Frank, Mark A. Hall, and Ian H. Witten. The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques” Morgan Kaufmann, 4-th ed. [Электрон. ресурс]. Режим доступа: https://www.cs.waikato.ac.nz/ml/weka/ Witten_et_al_2016_appendix.pdf (дата звернення: 09.12.2021).
2. Нестеров С.Н. Основы интеллектуального анализа данных. Лабораторный практикум: учебное пособие. [Текст] / СПб.: Лань, 2020, 40 с.
3. Мороз Б.І. Лабораторний практикум з курсу: “Аналіз даних та процесів”. [Электрон. ресурс]. Режим доступа: https://do.nmu.org.ua/course/view.php?id=3214 (дата звернення: 09.12.2021).
4. User’s Guide Mathcad 14.0. Parametric Technology Corporation. [Электрон. ресурс]. Режим доступа: http://www2.peq.coppe.ufrj.br/Pessoal/Professores/Arge/Nivelamento/ Mathcad/ 2-Apostilas/Mathcad%20Users%20Guide.pdf (дата звернення: 09.12.2021).
5. Brent Maxfield. Essential PTC Mathcad Prime 3.0. A Guide for New and Current Users. [Текст] / Elsevier Inc., 2014, 563 с.
6. Кожевников А.В. Лабораторний практикум та індивідуальні завдання з дисципліни “Інтелектуальний аналіз даних”. [Электрон. ресурс]. Режим доступа: https://do.nmu.org.ua/course/view.php?id=1817 (дата звернення: 09.12.2021).