Computer Support of Creation Study Materials from Selected Statistical Methods

Dušan Knežo

{dusan.knezo@tuke.sk}

Technical University of Košice, Faculty of Manufacturing Technologies with a Seat in Prešov, Department of Natural Sciences and Humanities, Bayerova 1, 080 01 Prešov, Slovak Republic

Abstract. Interval estimations of parameters for normal distribution and testing hypothesis about normal distribution parameters are in practice often use and among into the basic statistical method and in school is on mentioned method donated special attention. Presented article describes methods for computer-generated training tasks, which are necessary for the creation of study materials from the above-mentioned statistical methods. Generating of tasks is made by computer, software apparats in language Free Pascal in environment Lazarus and are used pascal´s procedures for generating random values from normal distribution and evaluation of critical values of normal distribution, Students distribution and χ^2- distribution. Text and solutions is in format PDF and is created in typographical system LaTeX.

Keywords: parameters of normal distribution, interval estimation, hypothesis testing, PC creation of task and its solutions

1   Introduction

Interval estimations of normal distribution parameters and testing hypothesis about normal distribution parameters are very often used statistical methods and in learning process is devoted to method significant attention. The education process requires study materials with sufficient amount of training tasks for application of mentioned statistical methods. Generating the task using computer is represent by several notes:

- parameters of normal distribution are generated by computer or by user,
- software generate random selection of values from normal distribution,

- software generate text to tasks and its solutions in the format PDF.

2 Generating of random values and evaluation of critical values

Generating of study tasks is necessary to dispone with software tools:
- for generating random values from normal distribution
- for evaluation of critical values of the normal distribution,
- for evaluation of critical values of the Students distribution,
- for evaluation of critical values of the χ^2-distribution.

2.1 Generating of random values from uniform distribution

Generating random values from of the uniform distribution provides software tools of Free Pascal, specifically procedure Randomize and function Random. Generating random values using mentioned function and are represented by:
- range of generating random values n
- parameters μ and σ of normal distribution from which is generated random selection.

2.2 Generating of random values from normal distribution

Generating random values for normal distribution $N(0,1)$ can be used standard numerical method [1].
If distribution function $\phi(y)$ of normal distribution $N(0,1)$ is written in form

$$\phi(y) = \begin{cases} 0.5 - \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^{|y|} e^{-\frac{t^2}{2}} dt, & y < 0, \\[4mm] 0.5 + \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^{y} e^{-\frac{t^2}{2}} dt, & y \geq 0. \end{cases} \tag{1}$$

then evaluation values of distributing function is enough to find numerical solution of the integral

$$\int_0^y \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_0^y e^{-\frac{t^2}{2}}dt, \tag{2}$$

for $y \geq 0$. Because

$$\left|\varphi^{(4)}(t)\right| \leq \varphi^{(4)}(0) = \frac{3}{\sqrt{2\pi}}, \tag{3}$$

then integral (2) can be evaluated by Simpsons method with preselected selected accuracy $\varepsilon > 0$, with assumption, that step of numerical integration h is

$$h \leq \sqrt[4]{\frac{60 \cdot \varepsilon \cdot \sqrt{2\pi}}{y}}. \tag{4}$$

Generating of random values $x_1, x_2, \cdots, x_n$ from normal distribution $N(\mu, \sigma^2)$ is realized in three steps.

1. Using function Random are generated random values $p_1, p_2, \cdots, p_n$ z from even distribution $U(0,1)$.
2. For each value p_i is assigned random value y_i from normal distribution $N(0,1)$. Value y_i is the solution of the equation $\phi(y_i) = p_i$. Equation is solved by bisection method.
3. For each value y_i is assigned random value x_i from normal distribution $N(\mu, \sigma^2)$ by transformation $x_i = \mu + \sigma \cdot y_i$.

2.3 Evaluation of critical values of normal distribution

If function $F(x)$ is distributing function, the critical value x_α of the distribution on significance level α is values and is valid:

$$1 - F(x_\alpha) = \alpha. \tag{5}$$

Critical value y_α of normal distribution $N(0,1)$ is evaluated by numerical solution of the equation

$$1 - \phi(y_\alpha) = \alpha. \tag{6}$$

Equation is solved by using bisection method [2].

2.4 Evaluation of critical values of Students distribution

Distributing function of Students distribution with n degrees of freedom can be written in the form

$$F_n(x) = \begin{cases} 0.5 - c_n \cdot \displaystyle\int_0^{|x|} g_n(t)\,dt, & x < 0, \\[2em] 0.5 + c_n \cdot \displaystyle\int_0^{x} g_n(t)\,dt, & x \geq 0, \end{cases} \tag{7}$$

where

$$c_n = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\dfrac{n}{2}\right)} \tag{8}$$

and

$$g_n(t) = \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \tag{9}$$

then for evaluation values of distributing function is enough to numerical evaluate the integral

$$\int_0^x g_n(t)\,dt = \int_0^x \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt \tag{10}$$

for $x \geq 0$. Because

$$|g_n''(t)| \leq |g_n''(0)| = \frac{n+1}{n}, \tag{11}$$

then integral (10) can be solved by trapezoidal rule with preselected precision $\varepsilon > 0$, with integration step h as follows

$$h \leq \sqrt{\frac{12n\varepsilon}{x \cdot (n + 1)}} \cdot \qquad (12)$$

Critical value x_α of the Students distribution is calculated as numerical solving the equation

$$1 - F_n(x_\alpha) = \alpha. \qquad (13)$$

Equation is solved using bisection method [2].

2.3 Evaluation of critical values for χ^2-distribution

Values of the distribution function χ^2–distribution with n degrees of freedom is defined by prescription

$$F_n(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\dfrac{n}{2}\right)} \cdot \displaystyle\int_0^x x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}} \cdot dt, & x > 0, \\[2em] 0, & x \leq 0, \end{cases} \qquad (14)$$

can be calculated using Richardson extrapolation and the Simpson's rule.

Critical value x_α χ^2–distribution with n degrees of freedom can be calculated as numerical solving of the equation (13).

3   Task generating

Task generating in the area of interval estimation parameters and hypothesis testing of normal distribution was created the program in programing language Free Pascal in integrated development environment Lazarus.

Basic inputs for program contains input file, which consist of data groups, number of generating task set by user. Each group consist of following information:

- parameters μ and σ^2 of normal distribution represent base of generating task of the random $x_1, x_2, \cdots, x_n$,
- number of decimal places to generating values.

3.1 Interval estimation

Subsequently after loading of each input data groups is in software processes following steps:

1. Generating of natural number n from even distribution $U(n_1, n_2)$, which represent range of random selection.
2. Generating random selection $x_1, x_2, \cdots, x_n$ from normal distribution $N(\mu, \sigma^2)$.
3. Randomly selection of interval estimation of the mean or variance. If the estimate is around the mean value, it randomly select whether variance is known or unknown. Known value is imported automatically from input data file.
4. Randomly choose interval selection and its form if is two-sided, left-hand or right-hand.
5. Calculates the necessary numerical characteristics of the random selection, the necessary critical values and determines the selected interval estimate.
6. In a single output file writes the task text and its result, writes the task text and the complete solution to the second output file. Output file text is in LaTeX typographic format.

Processing all input data groups software call the external software pdfLaTeX which is included in LaTeX pack and is created final pdf file.

3.2 Hypothesis testing

Generating hypothesis testing tasks, the program proceeds analogously, but randomly chooses a single-parameter or two-parameter test at the start of each input group. In the case of two-parameter tests are generated two random (steps 1 and 2).

4   Conclusions

Presented article describes methods of creation study material supported with computer technics. Methods can be used for creation study materials, such as ex-

ercises and homework and also in the process of creating the study literature. The advantage of the method is the possibility of creation a large number of different tasks in an automated approach in PDF format. Another advantage is the possibility of working with the LaTeX source text created in program.

Acknowledgement

References

[1] Knežo, D.: Inverse transformation method for normal distribution and the standard numerical methods. International journal of interdisciplinarity in theory and practice, No. 5 (2014). ISSN 2344-2409.

[2] Knežo, D., Zajac, J., Michalik, P.: Calculation of critical values of several probability distributions using standard numerical methods. In: SmartCity360 2016. - Gent : EAI, 2017 S. 1-9. ISBN 978-1-63190-149-2

[3] Knežo, D.: Application of numerical methods for the teaching of statistics. International journal of interdisciplinarity in theory and practice, No. 5 (2014). ISSN 2344-2409.

[4] Walck, Ch.: Hand-book on statistical distributions for experimentalists. Universitet Stokholms 2007. [Online] Stokholms universitet [10.5.2017] http://www.fysik.su.se/~walck/suf9601.pdf.