

УДК 004.9:372.851

DOI: 10.25559/SITITO.14.201804.960-965

О ПРИМЕНЕНИИ МАТЛАВ К РЕШЕНИЮ СТАТИСТИЧЕСКИХ ЗАДАЧ

Г.Д. Анисимова, С.И. Евсеева

Омский государственный технический университет, г. Омск, Россия

ON APPLICATION OF MATLAB TO THE SOLUTION OF STATISTICAL PROBLEMS

Galina D. Anisimova, Svetlana I. Evseeva

Omsk State Technical University, Omsk, Russia

© Анисимова Г.Д., Евсеева С.И., 2018

Ключевые слова

Информационные технологии;
математическая статистика;
Matlab; проверка статистиче-
ской гипотезы.

Аннотация

В настоящее время решение исследовательских задач невозможно представить без использования компьютерных технологий. При обработке результатов вычислительного эксперимента необходим статистический анализ большого объема данных, в связи с чем возникает потребность в применении информационных технологий для автоматизации этого процесса. Современные программные средства, такие как Matchcad, Matlab, Maple, Maxima, Statistica, Scilab, Statgraphics позволяют решать широкий круг задач, связанных с обработкой экспериментальных данных, используя встроенные функции или алгоритмы, разработанные пользователем. В Matlab существует огромное количество стандартных функций, необходимых для статистической обработки данных, графических средств для визуализации результатов исследования, а также возможность интеграции с MS Excel для обмена данными. Пакет прикладных программ Statistics Toolbox расширяет возможности системы Matlab в области статистической обработки данных, включает в себя функции для генерации и исследования случайных массивов с различными законами распределения. В пакете предусмотрено множество инструментов для графической визуализации данных. Учитывая вышеизложенное, была поставлена задача разработки алгоритма для написания программы, позволяющей проводить статистический анализ данных с определением числовых характеристик выборки, построением гистограммы и функции плотности распределения случайных величин, проверкой гипотезы о законе распределения по заданному критерию. В статье рассматривается использование пакета Matlab в курсе математической статистики технического университета, изучены возможности компьютерной системы Matlab для статистики, представлен алгоритм проверки статистической гипотезы о распределении генеральной совокупности и результаты его выполнения в Matlab.

Keywords

Information technology;
mathematical statistics; Matlab;
verification of the statistical
hypothesis.

Abstract

The solution of research problems without the use of computer technology can't be imagined at present. A statistical analysis of a large amount of data is necessary to processing the results of a computational experiment and the use of information technology contributes to automate this process. Modern software tools such as Matchcad, Matlab, Maple, Maxima, Statistica, Scilab, Statgraphics allow solving a lot of tasks related to the processing of experimental data using built-in functions or algorithms developed by the user. There are a lot of number of standard functions required for statistical processing of data, graphical tools for visualizing the results of the investigation, as well as the ability to integrate with MS Excel for data exchange in Matlab. The Statistics Toolbox extends the capabilities of the Matlab system in the field of statistical data processing, includes functions for generating and researching random arrays with different distribution laws. The package provides many tools for graphical visualization of data. Thus, the problem of development of the algorithm for writing a program that allows statistical analysis of data with the determination of numerical characteristics of the sample, the construction of a histogram and a distribution function of random variables, and testing the hypothesis of the distribution law according to a given criterion is set. The using of the Matlab in the course of mathematical statistics at the technical university is considered, the possibilities of the Matlab for statistics are studied, the algorithm for testing the statistical hypothesis on the distribution of the population and the results of its implementation in Matlab is presented.

Об авторах:

Анисимова Галина Дмитриевна, аспирант, кафедра высшей математики, Омский государственный технический университет (644050, Россия, г. Омск, пр. Мира, д. 11), ORCID: <http://orcid.org/0000-0002-4621-6726>, gdanisimova@gmail.com

Евсеева Светлана Ивановна, аспирант, кафедра химической технология и биотехнологии, преподаватель кафедры высшей математики, Омский государственный технический университет (644050, Россия, г. Омск, пр. Мира, д. 11), ORCID: <http://orcid.org/0000-0003-4109-8131>, evseevasve@rambler.ru



Введение

В современном мире информационные технологии масштабно применяются как в образовании, так и в научных исследованиях. Проведение практических занятий в аудиториях, оснащенных компьютерным оборудованием, стало неотъемлемым элементом системы обучения в техническом вузе. Существует множество программных средств и специализированных пакетов для решения математических задач [1, 2], использование которых способствует улучшению образовательного процесса [3-6].

Различные программные средства, такие, как Statistica, Matlab, Mathcad, Maple, применяются при решении задач из курса теории вероятностей и математической статистики. Благодаря статистике становится возможным обработка больших объемов данных.

Из многообразия вышеперечисленных пакетов наше внимание привлёк Matlab [7-9], поскольку с его помощью можно не только производить сложные вычисления, но и оформлять теоретические выкладки. Система MATLAB разработана К. Молером в конце 70-х гг. и с тех пор повсеместно используется при решении задач математической статистики. [8]

Некоторые особенности MATLAB:

- сбор и обработка экспериментальных данных,
- генерация наборов случайных чисел,
- визуализация результатов эксперимента,
- оформление отчёта,
- возможность диалога с другими системами (Maple, Mathcad, MS Word, MS Excel),
- наличие большого числа встроенных функций,
- высокая скорость вычислений.

В данной работе подробно рассмотрена задача проверки статистической гипотезы о законе распределения генеральной совокупности по выборке. Цель исследования – построение гистограммы, выдвижение и принятие или опровержение гипотезы о распределении генеральной совокупности средствами пакета Matlab.

Задачи исследования:

- изучить возможности пакета Matlab для статистики,

- реализовать алгоритм решения поставленной задачи в Matlab,
- найти числовые характеристики выборки,
- визуализировать промежуточные и конечные результаты исследования,
- проверить согласованность теоретического и статистического распределений.

Построение эмпирической функции распределения и гистограммы

Дана выборка R объёма n (файл *.xlsx), содержащая сто случайных элементов (табл. 1). Упорядочим выборку R , используя функцию *sort*. Найдём наименьшее R_{min} и наибольшее R_{max} числа выборки и разность между ними *inter*. Интервал *inter* разделим на m равных частей, где $m = \text{round}(n^{0.5})$.

Результат выполнения команд:

```
R_min = 1
R_max = 9
inter = 8
m = 10
```

Таблица 1. Выборка из 100 случайных элементов

Table 1. A sample of 100 random elements

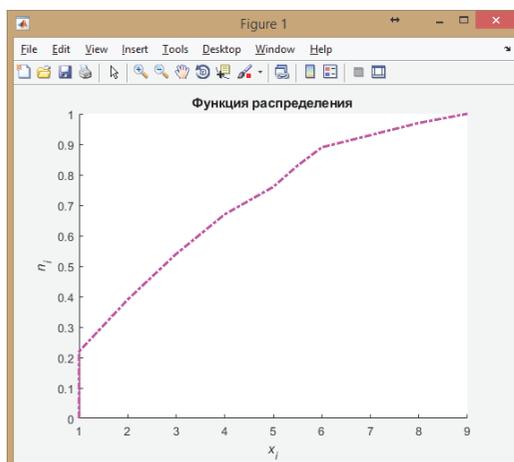
3.00	2.00	4.00	5.50	5.00	4.00	2.00	1.00	2.00	5.00
4.00	6.00	5.50	4.00	4.00	3.00	4.00	6.00	2.00	5.00
5.50	9.00	5.00	8.00	3.00	1.00	5.00	6.00	2.00	2.00
8.00	4.00	2.00	2.00	4.00	1.00	8.00	3.00	5.50	1.00
1.00	5.50	3.00	1.00	9.00	3.00	1.00	3.00	3.00	4.00
4.00	8.00	5.00	3.00	1.00	1.00	2.00	5.50	2.00	5.50
6.00	1.00	6.00	9.00	7.00	1.00	7.00	3.00	1.00	1.00
2.00	3.00	1.00	3.00	5.00	1.00	7.00	4.00	1.00	1.00
2.00	2.00	1.00	5.00	2.00	4.00	5.00	1.00	6.00	1.00
3.00	1.00	3.00	7.00	3.00	2.00	2.00	2.00	4.00	1.00

Числовые характеристики в MATLAB вычисляются по формулам (табл. 2).

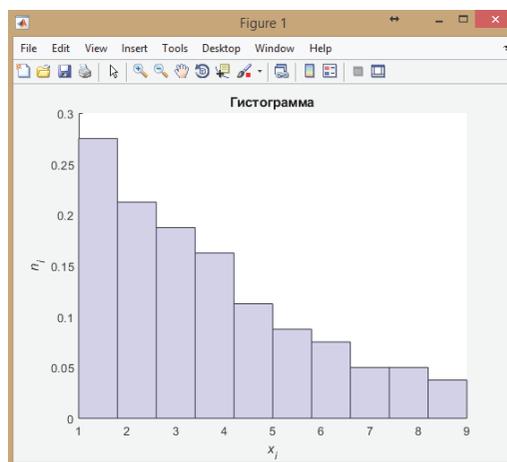
Таблица 2. Числовые характеристики

Table 2. Number characteristics

Числовая характеристика	Команда в MATLAB
Выборочное среднее (\bar{x})	mean (R)
Выборочная дисперсия (D)	var (R)
Выборочное среднее квадратическое отклонение (σ)	std (R)



a



b

Рис. 1. а – график эмпирической функции распределения; б – гистограмма

Fig. 1. a - Empirical distribution function; b - histogram



Значения числовых характеристик, рассчитанных в MATLAB для выборки (табл.1):

```

Выборочное среднее :
M = 3.5950
Выборочная дисперсия :
D = 4.9328
Выборочное среднее квадратическое отклонение :
S = 2.2210
  
```

Построим график функции распределения, используя команду $[Fi,xi]=ecdf(R)$, где Fi – вектор значений эмпирической функции распределения, вычисленный для упорядоченных данных x_i выборки R . При построении нормированной гистограммы также используем встроенную команду $ecdfhist(Fi,xi,m)$. Результаты выполнения команд представлены на рисунке 1.

Выдвижение гипотезы о распределении генеральной совокупности

По виду построенной гистограммы сделаем предположение о законе распределения генеральной совокупности. Для выбора подходящего закона распределения эмпирическую функцию или гистограмму сравнивают с известными законами распределения. В математической статистике чаще всего встречаются следующие распределения: нормальное, показательное (экспоненциальное), равномерное, Релея и гамма. Каждое распределение характеризуется определенным набором параметров. Для вычисления параметров распределения в MATLAB используется функция $mle(R,'distribution','Name')$, где $Name$ – название закона распределения из списка $distribution$.

Таблица 3. Вычисление параметров и функции плотности распределения в MATLAB
Table 3. Calculation of parameters and density distribution functions in MATLAB

Параметры распределения	Вычисление параметров в MATLAB	Вычисление функции плотности распределения в MATLAB
Нормальное распределение		
a, σ	$a_sigma = mle(R,'distribution','Normal')$	$f_norm = normpdf(x,a_sigma(1),a_sigma(2))$
Показательное распределение		
$\lambda = \frac{1}{a}$	$lambda = mle(R,'distribution','Exponential')$	$f_exp = expdf(x,lambda)$
Равномерное распределение		
a, σ	$a_sigma_unif = mle(R,'distribution','Uniform')$	$f_unif = unifpdf(x,a_sigma_unif(1),a_sigma_unif(2))$
Закон Релея		
σ	$sigma = mle(R,'distribution','Rayleigh')$	$f_rayl = raylpdf(x,sigma)$
Гамма-распределение		
$a = k\theta, \sigma = \theta\sqrt{k}$, где k и θ – коэффициенты масштаба	$a_sigma = mle(R,'distribution','Gamma')$	$f_gamma = gampdf(x,a_sigma(1),a_sigma(2))$

Используя функции, приведенные в таблице 2, покажем на рисунке 2, как графики функций плотности распределений накладываются на гистограмму, изображенную на рисунке 16.

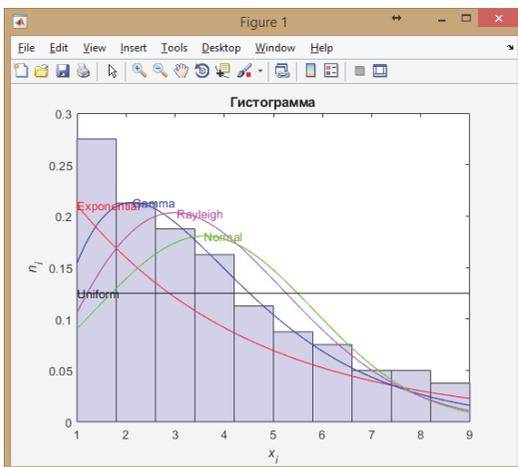


Рис. 2. Графики функции плотности распределений
Fig. 2. Density function graph

Подберем закон распределения исходя из вида гистограммы (сравниваем гистограмму с графиками плотностей распределения). В данном случае можем предположить, что выборка подчиняется показательному закону распределения. Это предположение назовем статистической гипотезой.

Проверка согласия теоретического и статистического распределений

Процесс проверки статистической гипотезы состоит из следующих этапов:

I. Выдвижение основной (H_0) и альтернативных гипотез (H_k).

II. Выбор критерия для проверки гипотезы H_0 .

III. Определение уровня значимости α и критической области M .

IV. Вычисление критерия по данным опыта.

V. Если найденное значение критерия принадлежит критической области M , то гипотеза H_0 отвергается, иначе гипотеза H_0 принимается на уровне значимости α [10].

Проверим согласие гипотетического закона распределения с данными выборки, используя критерий Пирсона. В качестве меры расхождения теоретического и наблюдаемого распределе-



ний выбирается величина χ^2 , определяемая формулой:

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}, \quad (1)$$

где n – объем выборки; m – число интервалов разбиения; n_i – число элементов выборки в i -ом интервале; p_i – теоретическая вероятность попадания случайной величины в i -й интервал, определяемая в соответствии с выбранным законом распределения.

При достаточно большом объеме выборки число степеней свободы распределения χ^2 вычисляется следующим образом:

$$k = m - l - 1, \quad (2)$$

где m – число интервалов, l – число параметров закона распределения.

По таблице значений критерия согласия Пирсона (χ^2) в со-

ответствии с заданным уровнем значимости определяется критическое значение χ^2 . Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, то выдвинутая гипотеза о законе распределения принимается, иначе – отвергается на уровне значимости α . Для вычисления критического значения в Matlab используется функция $\text{chi2inv}(1-\alpha, k)$.

Проверим выдвинутую гипотезу (H_0) о показательном законе распределения по критерию Пирсона χ^2 . Заметим, что альтернативными (H_k) являются гипотезы о нормальном, равномерном, Релеевском и Гамма-распределении; их также проверим по критерию Пирсона. Зададим уровни значимости $\alpha=[0.01; 0.05]$. Вычислим значения теоретических вероятностей по формуле, соответствующей показательному закону распределения:

$$p_i = e^{-\lambda a_i} - e^{-\lambda b_i},$$

где $[a_i; b_i]$ – i -й интервал. Вычислим в Matlab значения теоретических вероятностей для каждого закона из основной и альтернативных гипотез по формуле Симпсона [11]. Рассчитанные вероятности появятся на экране:

```
pi_norm =
0.0882    0.1176    0.1379    0.1423    0.1292    0.1031    0.0724    0.0447
0.0243    0.0116
pi_exp =
0.1511    0.1209    0.0968    0.0775    0.0620    0.0497    0.0397    0.0318
0.0255    0.0204
pi_unif =
0.1000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000
0.1000    0.1000
pi_rel =
0.1198    0.1582    0.1678    0.1530    0.1235    0.0894    0.0586    0.0350
0.0191    0.0095
pi_gamma =
0.1471    0.1721    0.1596    0.1311    0.1000    0.0725    0.0507    0.0345
0.0229    0.0150
```

Подставим вычисленные значения вероятностей в (1), найдем $\chi^2_{\text{набл}}$:

```
Наблюдаемые значения критерия хи-квадрат:
chi_norm = 28.7143
chi_exp = 16.2405
chi_unif = 37.4000
chi_rel = 17.1173
chi_gamma = 6.8144
```

Определим число степеней свободы по формуле (2): $k = 10 - 1 - 1 = 8$:

Число степеней свободы:

```
r_norm = 7
r_exp = 8
r_unif = 7
r_rel = 8
r_gamma = 7
```

На уровнях значимости $\alpha=[0.01; 0.05]$ и $k = 8$ найдем $\chi^2_{\text{кр}}$:

$\alpha=0.01$				$\alpha=0.05$			
Табличные значения критерия хи-квадрат:				Табличные значения критерия хи-квадрат:			
chi_tabl_norm =	18.4753			chi_tabl_norm =	14.0671		
chi_tabl_exp =	20.0902			chi_tabl_exp =	15.5073		
chi_tabl_unif =	18.4753			chi_tabl_unif =	14.0671		
chi_tabl_rel =	20.0902			chi_tabl_rel =	15.5073		
chi_tabl_gamma =	18.4753			chi_tabl_gamma =	14.0671		



Сравнивая значения $\chi^2_{0.01}$ и $\chi^2_{0.05}$ с $\chi^2_{набл}$, программа выдаст результат:

$\alpha=0.01$				$\alpha=0.05$			
Распределение генеральной совокупности подчиняется	случайных величин в совокупности не подчиняется	случайных величин в совокупности не подчиняется	закону нормального закона	Распределение массива не подчиняется	случайных величин в совокупности не подчиняется	случайных величин в совокупности не подчиняется	закону нормального закона распределения
Распределение экспоненциальному закону распределения	случайных величин в совокупности подчиняется	случайных величин в совокупности подчиняется	закону экспоненциальному закону распределения	Распределение массива не подчиняется	случайных величин в совокупности подчиняется	случайных величин в совокупности подчиняется	закону экспоненциальному закону распределения
Распределение равномерному закону распределения	случайных величин в совокупности не подчиняется	случайных величин в совокупности не подчиняется	закону равномерному закону	Распределение массива не подчиняется	случайных величин в совокупности не подчиняется	случайных величин в совокупности не подчиняется	закону равномерному закону распределения
Распределение закону Рэлея	случайных величин в совокупности подчиняется	случайных величин в совокупности подчиняется	закону Рэлея	Распределение массива не подчиняется	случайных величин в совокупности не подчиняется	случайных величин в совокупности не подчиняется	закону Рэлея
Распределение гамма-распределению	случайных величин в совокупности подчиняется	случайных величин в совокупности подчиняется	закону гамма-распределению	Распределение массива подчиняется	случайных величин в совокупности подчиняется	случайных величин в совокупности подчиняется	закону гамма-распределению

Обратим внимание, что на уровне значимости $\alpha=0.01$ подтверждается не только основная гипотеза о показательном законе распределения, но и альтернативные гипотезы о Релеевском и гамма-распределениях. Однако, на уровне значимости $\alpha=0.05$ основная гипотеза отвергается, подтверждается лишь одна из альтернативных – гипотеза о гамма-распределении.

Заключение

Применение Matlab при решении поставленной задачи позволило оценить значимость математических пакетов при обучении; быстро обработать большой массив данных; потратить меньше времени на громоздкую вычислительную работу; использовать графические возможности для визуализации результатов исследования.

Отметим, что при выполнении курсовых проектов часто приходится проводить статистические исследования (например, группировка или анализ данных), требующие больших затрат времени. Использование Matlab при выполнении индивидуальных заданий, а также выпускных квалификационных работ повысит их сложность и объем научных исследований.

Список использованных источников

- [1] Чернова О.В., Ковалева Л.А., Тарасова О.А. Анализ современных компьютерных пакетов для преподавания математических дисциплин // Научный альманах. 2016. № 4-2. С. 317–320. DOI: 10.17117/na.2016.04.02.317
- [2] Эсетов Ф.Э., Гаджиева З.Д., Азизова Л.Н. Проблемы использования математических пакетов в преподавании математических дисциплин в вузе // Известия Дагестанского государственного педагогического университета. Психолого-педагогические науки. 2016. Т. 10. №. 2. С. 103–106. URL: <https://elibrary.ru/item.asp?id=27200156> (дата обращения: 27.09.2018).
- [3] Ахметова Ф.Х., Ласковая Т.А., Чигирёва О.Ю. Методика обработки результатов эксперимента с помощью системы

Matlab в курсе «Математическая статистика» // Инженерный вестник. 2016. № 4. С. 1001–1011. URL: <https://elibrary.ru/item.asp?id=26165429> (дата обращения: 27.09.2018).

- [4] Власова Е.А., Меженная Н.М., Попов В.С., Пугачев О.В. Методические аспекты обеспечения дисциплины «теория вероятностей» в техническом университете // Современные наукоемкие технологии. 2017. № 11. С. 96–103. URL: <https://elibrary.ru/item.asp?id=30725892> (дата обращения: 27.09.2018).
- [5] Иглин С.П. Теория вероятностей и математическая статистика на базе MATLAB. Харьков: НТУ ХПИ, 2006. 612 с.
- [6] Плохотников К.Э., Николенко В.Н. Теория вероятностей в пакете MATLAB. М.: Горячая линия-Телеком, 2014. 612 с. URL: <https://elibrary.ru/item.asp?id=24057299> (дата обращения: 27.09.2018).
- [7] Ануфриев И.Е., Смирнов А.В., Смирнова Е.В. MATLAB7. СПб.: БХВ, 2005. 1104 с.
- [8] Дьяконов В.П. MatLAB 6/6.1/6.5 + Simulink 4/5 в математике и моделировании. М.: Солон-Пресс, 2010. 582 с. URL: <https://elibrary.ru/item.asp?id=30657586> (дата обращения: 27.09.2018).
- [9] Потемкин В.Г. Система MATLAB. М.: Диалог-МИФИ, 1997. 350 с.
- [10] Соколовский М.Н., Цветкова В.Д. Математическая статистика. Омск: ОмГТУ, 2005. 68 с.
- [11] Вержбицкий В.М. Численные методы (математический анализ и обыкновенные дифференциальные уравнения). М.: Издательский дом «ОНИКС 21 век», 2005. 400 с. URL: <https://elibrary.ru/item.asp?id=20305699> (дата обращения: 27.09.2018).
- [12] MathWorks [Электронный ресурс]. URL: <http://www.mathworks.com> (дата обращения: 27.09.2018).
- [13] Exponenta.ru [Электронный ресурс]. URL: <http://old.exponenta.ru/> (дата обращения: 27.09.2018).

Поступила 27.09.2018; принята в печать 10.11.2018; опубликована онлайн 10.12.2018.



References

- [1] Chernova O.V., Kovaleva L.A., Tarasova O.A. The analysis of modern software package for teaching mathematics. *Science Almanac*. 2016; 4-2:317–320. (In Russian) DOI: 10.17117/na.2016.04.02.317
- [2] Esetov F.E., Gadzhieva Z.D., Azizova L.N. Problems of Using the Mathematical Software in Teaching the Mathematical Disciplines at the University. *Dagestan State Pedagogical University. Journal. Psychological and Pedagogical Sciences*. 2016; 10(2):103-106. Available at: <https://elibrary.ru/item.asp?id=27200156> (accessed 27.09.2018). (In Russian)
- [3] Ahmetova F.H., Laskovaya T.A., Chigireva O.J. Metodika obrabotki rezul'tatov eksperimenta s po-moshch'yu sistemy Matlab v kurse "Matematicheskaya statistika". *Engineering Bulletin*. 2016; 4:1001–1011. Available at: <https://elibrary.ru/item.asp?id=26165429> (accessed 27.09.2018). (In Russian)
- [4] Vlasova E.A., Mezhenaya N.M., Popov V.S., Pugachev O.V. Methodological aspects of the discipline "Probability theory" in a technical university. *Modern high technologies*. 2017; 11:96-103. Available at: <https://elibrary.ru/item.asp?id=30725892> (accessed 27.09.2018). (In Russian)
- [5] Iglin S.P. Teoriya veroyatnostej i matematicheskaya statistika na baze MATLAB [Probability theory and mathematical statistics based on MATLAB]. Har'kov: NTU KPI, 2006. 612 p. (In Russian)
- [6] Plohotnikov K.E., Nikolenko V.N. Teoriya veroyatnostej v pakete MATLAB [Probability theory in MATLAB]. M.: Goryachaya liniya-Telekom, 2014. 612 p. Available at: <https://elibrary.ru/item.asp?id=24057299> (accessed 27.09.2018). (In Russian)
- [7] Anufriev I.E., Smirnov A.B., Smirnova E.V. MATLAB7. SPb.: BHV, 2005. 1104 p. (In Russian)
- [8] D'yakov V.P. MATLAB 6/6.1/6.5 + Simulink 4/5 v matematike i modelirovanii [MATLAB 6/6.1/6.5 + Simulink 4/5 in Mathematics and Modeling]. Moscow: SOLON-Press, 2010. 582 p. Available at: <https://elibrary.ru/item.asp?id=30657586> (accessed 27.09.2018). (In Russian)
- [9] Potemkin V.G. Sistema MATLAB. M.: Dialog-MEPi, 1997. 350 p. (In Russian)
- [10] Sokolovskij M.N., Cvetkova V.D. Matematicheskaya statistika [Mathematical Statistics]. Omsk: OmGTU, 2005. 68 p. (In Russian)
- [11] Verzhbitskii V.M. Numerical Methods (Mathematical Analysis and Ordinary Differential Equations). M.: ONIX 21 century, 2005. 400 p. Available at: <https://elibrary.ru/item.asp?id=20305699> (accessed 27.09.2018). (In Russian)
- [12] MathWorks. Available at: <http://www.mathworks.com> (accessed 27.09.2018).
- [13] Exponenta.ru. Available at: <http://old.exponenta.ru/> (accessed 27.09.2018). (In Russian)

Submitted 27.09.2018; revised 10.11.2018;
published online 10.12.2018.

About the authors:

Galina D. Anisimova, postgraduate student of the Department of High Mathematics, Omsk State Technical University (11 Mira Av, Omsk 644050, Russia), ORCID: <http://orcid.org/0000-0002-4621-6726>, gdanisimova@gmail.com

Svetlana I. Evseeva, postgraduate student; lecturer of the Department of High Mathematics, Omsk State Technical University (11 Mira Av, Omsk 644050, Russia), ORCID: <http://orcid.org/0000-0003-4109-8131>, evseevasve@rambler.ru



This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted reuse, distribution, and reproduction in any medium provided the original work is properly cited.

