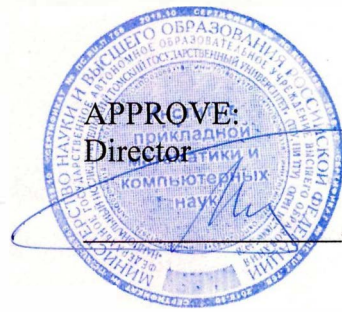
Appendix 1

Ministry of Science and Higher Education of the Russian Federation
NATIONAL RESEARCH
TOMSK STATE UNIVERSITY (NR TSU)

Institute of Applied Mathematics and Computer Science



APPROVE:
Director
_____A. V. Zamyatin

Evaluation materials of the current control and intermediate certification in the discipline

(Evaluation tools by discipline)

Mathematics and Statistics for Data Science – I

in the major of training

01.04.02 Applied mathematics and informatics

Orientation (profile) of training:

Big Data and Data Science

Tomsk – 2024

ET was implemented:
PhD,
Associate Professor of the Department
of Probability Theory and Mathematical Statistics — T.V. Kabanova

Reviewer:
PhD,
Associate Professor of the Department
of System Analysis and Mathematical Modeling — Zh.N. Zenkova

Evaluation tools were approved at a meeting of the educational and methodological commission of the Institute of Applied Mathematics and Computer Science (EMC IAMCS).

Protocol dated 20.05.2024 № 2

Chairman of the EMC IAMCS,
Dr. tech. Sciences, Professor — S.P. Sushchenko

Evaluation tools (ET) are an element of the system for assessing the formation of competencies among students in general or at a certain stage of its formation.

The ET is developed in accordance with the work program (WP) of the discipline.

1. Competencies and training outcomes, obtained upon the discipline mastery

Competencies	**Competence indicator**	**Code and name of planned training outcomes that characterize the stages of competency formation**	**Criteria for evaluating training outcomes**			
			Excellent	Good	Satisfactory	Unsatisfactory
GPC-1 – the ability to solve actual problems of fundamental and applied mathematics	IOPC-1.3 Demonstrates the skills of using the basic concepts, facts, principles of mathematics, computer science and natural sciences to solve practical problems related to applied mathematics and computer science.	TO-1.1.1. the student will be able to: - choose an adequate method for solving the problem; - implement the selected method in the data analysis program; - draw conclusions and interpret the results.	Demonstration of a high level of knowledge of the mathematical foundations and basic concepts that are necessary to understand the statistical methods of data analysis.	In general, successful, but containing some gaps, knowledge of the mathematical foundations and basic concepts that are necessary to understand the statistical methods of data analysis.	Fragmentary, incomplete knowledge without gross errors of the mathematical foundations and basic concepts that are necessary to understand the statistical methods of data analysis.	Does not know the mathematical foundations and basic concepts that are necessary to understand the statistical methods of data analysis.
GPC-2 – the ability to improve and implement new mathematical methods for solving	IOPC-2.1 Uses the results of applied mathematics to adapt new methods for solving problems in the field of his professional interests.	TO-2.1.1. the student will be able to: adapt models to describe the processes of a real subject area	Demonstration of a high level of knowledge of the mathematical foundations and	In general, successful, but containing some gaps, knowledge of	Fragmentary, incomplete knowledge without gross errors of the	Does not know the mathematical foundations and basic concepts that are necessary to

applied problems.	IOPC-2.1 Uses the results of applied mathematics to adapt new methods for solving problems in the field of his professional interests. IOPC-2.2 Implements and improves new methods, solving applied problems in the field of professional activity.	TO-2.2.1. the student will be able to: implement and interpret the constructed models to describe the processes of a real subject area. TO-2.3.1. The student will be able to implement a qualitative and quantitative analysis of the constructed models and the forecasts obtained on their basis and choose the most optimal one in accordance with the selected metric	basic concepts that are necessary to understand the statistical methods of data analysis.	the mathematical foundations and basic concepts that are necessary to understand the statistical methods of data analysis.	mathematical foundations and basic concepts that are necessary to understand the statistical methods of data analysis.	understand the statistical methods of data analysis.

2. Stages of competency formation and types of evaluation tools

№	Stages of competency formation (discipline sections)	Code and name of training outcomes	Type of evaluation tool (tests, assignments, cases, questions, etc.)
1.	Mathematical foundations	TO-1.1.1, TO-2.1.1, TO-2.2.1, OP-2.3.1.	Solution of practical problems; answers to questions in an exam or test
2.	Introduction to statistical analysis.	TO-1.1.1, TO-2.1.1, TO-2.2.1, TO-2.3.1.	Practical works; answers to questions in an exam or test
3.	Tests for comparing groups.	TO-1.1.1, TO-2.1.1, TO-2.2.1, TO-2.3.1.	Practical works; answers to questions in an exam or test
4.	Correlation analysis.	TO-1.1.1, TO-2.1.1, TO-2.2.1, TO-2.3.1.	Practical works; answers to questions in an exam or test
5.	Paired regression.	TO-1.1.1, TO-2.1.1, TO-2.2.1, TO-2.3.1.	Practical works; answers to questions in an exam or test

3. Typical control tasks or other materials necessary for the assessment of educational training outcomes

3.1. Typical tasks for conducting ongoing monitoring of progress in the discipline: tests, questions for colloquia, assignments for laboratory work.

Examples of practical tasks on mathematical foundations

Exercise. Solving systems of linear equations.

For a given system of equations, find a solution using the inverse matrix method.

Exercise. Eigenvalues and eigenvectors.

For a given matrix, find eigenvalues and eigenvectors.

Exercise. Functions of many variables.

For a given function of many variables, calculate the gradient, find the extrema of the function.

Examples of tasks for practical work on statistical analysis

Practical work. Data preprocessing

Exercise.

1. Import the given data set.
2. Check for gaps and outliers.

3. For quantitative indicators, build histograms.
4. Find estimates of numerical characteristics.
5. Test the hypothesis of normality.
6. Construct range diagrams by groups based on the breakdown of quantitative indicators by levels of categorical features.

Practical work. Exploratory Analysis

Runs in R.

Exercise.

Import a table with data into R.

1. Build graphs to visualize data and its relationships.
2. Check the relationships of factors with each other and their influence on the dependent target variable, choosing the appropriate criterion, depending on the types of data.
3. Test hypotheses about the significance of the relationship.

Practical work. Paired regression. Generation.

Runs in R.

Exercise.

1. Set the sample size n (from 50 to 150).
2. Generate a vector of predictor variable values.
3. Set the noise vector corresponding to the Gauss-Markov assumptions.
4. Set regression parameters.
5. Form a vector of values of the dependent variable according to the linear regression model.
6. Build a scatterplot and, if necessary, change the parameters.
7. Build OLS estimates of parameters, check their significance, compare with initial values
8. Find the SD of the residuals.
9. Check the quality of the model.

Practical work. Paired regression for real data. Linear and non-linear models.

Runs in R.

Exercise.

Import a table with data into R.

1. Build graphs to visualize data and their relationships.
2. Check the relationship of the factor with the dependent target variable.
3. Build and analyze a paired regression model of the target variable from the factor.
4. Build linear, exponential, exponential, logarithmic and inverse relationships.
5. Assess the quality of each model.
6. Choose the most adequate model.

3.2. Typical tasks for conducting intermediate certification in the discipline.

An approximate list of theoretical questions and topics for preparing for the exam:

1. Solution of systems of linear equations.
2. Eigenvectors and eigenvalues of a matrix.
3. Functions of several variables. The concept of a gradient.
4. Method of gradient descent.
5. Total probability formula and Bayes formula.
6. Types of data and ways to represent them.
7. Parametric tests for comparing groups.
8. Nonparametric tests for comparing groups.
9. Correlation analysis of quantitative data.
10. Rank correlation.
11. Correlation analysis of categorical data.
12. Paired regression. Model. OLS-estimates of parameters.
13. Descriptive statistics of estimates of parameters of paired regression.
14. Gauss-Markov theorem for the case of paired regression.
15. Checking the quality of the paired regression equation.

4. Methodological materials that determine the procedures for evaluating training outcomes

4.1. Methodological materials for assessing the of progress in the discipline.

For the ongoing certification, it is necessary to have attendance of at least 75% of all classes held at the time of certification and pass all practical work given at the time of certification. The practical work is graded for pass/fail.

4.2. Methodological materials for conducting intermediate certification in the discipline.

The results of the examination are "excellent", "good", "satisfactory", "unsatisfactory".

For a 15 question test. For each question, depending on its complexity, you can get from 1 to 3 points. Max 30.

excellent	from 26 to 30
good	from 21 to 25
satisfactory	from 16 to 20
unsatisfactory	from 0 to 15