

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Порохня Андрей Алексеевич

Должность: И.о. директора института перспективной инженерии

Дата подписания: 02.06.2026 12:52:49

Уникальный программный ключ:

990bea3aeec48c23bd8699e48288f8d1960c600

MINISTRY OF SCIENCE AND HIGHER EDUCATION

OF THE RUSSIAN FEDERATION

Federal State Autonomous Educational Institution of Higher Education

North Caucasus Federal University

APPROVE

Acting Director of the Institute,

Candidate of Technical Sciences,

Associate Professor A.A. Porohnia

FUND OF ASSESSMENT TOOLS FOR THE DISCIPLINE

Multidimensional data analysis

Direction of training

09.04.03 Applied Informatics

Focus (profile)

Knowledge Management

Year of starting studies

2026

Form of study

full-time

Implemented in the semester

1

Introduction

1. Purpose: The FOS is designed to objectively assess the level of competencies.

2. The FOS is an appendix to the program of the discipline "Multidimensional data analysis / Multidimensional data analysis"

3. Developer: Lapina M.A., Associate Professor of the Department of Information Security of Automated Systems.

4. The FOS has been reviewed.

Members of the expert group:

Chairman: I.V. Azarov, Head of the Department of Applied Informatics

Members of the Commission:

N.V. Kononova, Associate Professor of the Department of Applied Informatics

O.G. Orlinskaya, Associate Professor of the Department of Applied Informatics

Representative of the Employer Organization: S.A. Cherevko, General Director of Kraftkod LLC

Expert opinion: The FOS for the discipline allows you to assess the level of competencies development. Start the approbation.

« ____ » _____

5. The validity period of the FOS is determined by the duration of the educational program.

1. Description of the criteria for evaluating competencies at various stages of their development, and description of the evaluation scales

Levels of competency (ies) and indicator (s)	Descriptors			
	Minimum level not reached (Unsatisfactory) 2 points	Minimum level (Satisfactory) 3 points	Average level (Good) 4 points	High level (Excellent) 5 points
OPK-4				
ID-2 OPC-4. Apply new research methods in practice	Difficulty in applying new research methods in practice	Does not fully apply new research methods in practice	In general, it applies new research methods in practice	applies new research methods in practice
OPK -2				
ID -1 OPC-2. develop original algorithms for solving professional problems;	Difficulty in developing original algorithms for solving professional problems	Weakly develops original algorithms for solving professional tasks	He mainly develops original algorithms for solving professional problems.	develops original algorithms for solving professional tasks
ID-2 OPC-2. develop software tools for solving professional tasks	Difficulty in developing software for professional tasks	It has a weak development of software tools for solving professional tasks	He mainly develops software for solving professional tasks.	develops software for solving professional tasks
PC-3				
ID-1 PC-3. to develop fundamental and interdisciplinary knowledge	It cannot develop fundamental and interdisciplinary knowledge	Makes serious mistakes in the development of fundamental and interdisciplinary knowledge	It mainly develops fundamental and interdisciplinary knowledge.	develops fundamental and interdisciplinary knowledge
ID-2 PC-3. To develop scientific research methods	Hardly develops scientific research methods	Makes serious mistakes in the development of scientific research methods	Basically, he precisely develops the methods of scientific research	develops scientific research methods
OPK-6				
ID-1 OPC-6. to explore the current problems of applied informatics for the development of the information society	He is struggling to explore the current problems of applied informatics for the development of the information society.	It does not sufficiently explore the current problems of applied informatics for the development of the information society	In general, it explores the current problems of applied informatics for the development of the information society	explores the current problems of applied computer science for the development of the information society
PC-2				
ID-1 PC-2. Develop new competitive ideas in the field of applied informatics.	Difficult to develop new competitive ideas in the field of applied informatics	It does not develop new competitive ideas in the field of applied computer science	In general, it develops new competitive ideas in the field of applied computer science	develops new competitive ideas in the field of applied computer science
ID-2 PC-2. Implement new competitive ideas in the field of applied informatics	Difficult to implement new competitive ideas in the field of applied informatics	It does not implement new competitive ideas in the field of applied informatics	In general, it implements new competitive ideas in the field of applied informatics	implements new competitive ideas in the field of applied computer science

The level of competence in a discipline is assessed based on the "Regulations on the Conduct of Current Academic Performance Monitoring and Interim Assessment of Students in Higher Education Programs - Bachelor's Programs, Specialist Programs, and Master's Programs -

at the Federal State Autonomous Educational Institution of Higher Education "North Caucasus Federal University" in its current version.

ASSESSMENT TOOLS FOR CHECKING THE LEVEL OF COMPETENCE DEVELOPMENT

Task number	Right answer	The content of the question	Competence
		Full-time study, semester 1	
1.	end-user tools that provide multidimensional visualization and data manipulation; the multidimensional view layer is abstracted from the physical data structure and treats data as multidimensional	Multidimensional data representation	OPK-4
2.	a means (language) for formulating multidimensional queries (the traditional relational SQL language is not suitable for this purpose) and a processor that can process and execute such queries	Multidimensional processing	OPK-6,
3.	tools for physically organizing data that can efficiently execute multidimensional queries	Multidimensional storage	PC-2
4.	this is a set of methods designed to divide a set of objects into homogeneous groups (clusters). In most clustering methods, it is not known in advance how many classes will be identified in a given set of objects. Unlike discriminant analysis, cluster analysis is called "unsupervised classification" because its methods do not use a training sample.	Cluster analysis	PC-3
5.	- classification of objects based on their characteristics; - verification of the data structure assumption; - construction of new classes.	Cluster analysis tasks	OPK-2
6.	It is mainly used for "data compression", i.e. reducing a set of elementary features to a small number of "generalized features", and identifying	Factor analysis	OPK-4

	latent (hidden, unmeasurable) factors. This task can be solved not only for features, but also for objects. Factor analysis can be used for classification purposes after identifying generalized indicators.		
7.	All statistical packages process information presented in the form of a matrix of one of three types: a data matrix, a proximity (distance) matrix, or a relationship (between variables) matrix. Each of these matrices can be an input matrix or an output matrix generated by a processing program. This division of data is primarily based on content, and the files used to store and input this data have the same logical organization, differing only in their contents.	Main types of multidimensional data representation	OPK-6,
8.	This matrix, which defines the "object-attribute" relationship, is one of the most common forms of data organization and is a two-dimensional table of $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix},$ size n*m.	Data matrix.	PC-2

9.	This matrix defines the feature-feature relationship and is a two-dimensional symmetric square matrix of size $n \times m$ $S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1m} \\ S_{21} & S_{22} & \cdots & S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mm} \end{bmatrix}$	Matrix of connections between features.	PC-3
10.	This matrix, which defines the object-object relationship, is a square symmetric matrix of size $n \times m$ with non-negative elements $D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$	Matrix of proximities (distances).	OPK-2
11.	Discriminant analysis is an alternative to multiple regression analysis and is used to decide which variables (dependent variables, usually numerical) distinguish (discriminate) two or more (independent variables, usually nominal) emerging sets (groups).	Discriminant analysis	OPK-4
12.	The main purpose of multidimensional scaling is to identify the structure of the studied set of objects, when the initial information is data on the difference or proximity (similarity) of objects.	Multidimensional scaling.	OPK-6,

13.	Cluster analysis is a procedure for arranging objects into relatively homogeneous classes based on a pairwise comparison of these objects using pre-defined and measured criteria. Cluster analysis options are a variety of simple computational procedures used for object classification. Object classification is the grouping of objects into classes so that the objects in each class are more similar to each other than to objects in other classes.	Cluster analysis	PC-2
14.	The combination or tree clustering method is used to form clusters of dissimilarity or distance between objects. These distances can be defined in one-dimensional or multidimensional space. Types of distance measures: - Euclidean distance. - Square of the Euclidean distance. - City block distance (Manhattan distance). - Chebyshev distance. - Power distance. - Percentage of disagreement.	Measures of difference (distance) in cluster analysis.	PC-3
15.	- dividing a set of subjects into groups based on measured characteristics in order to further test the causes of intergroup differences based on external criteria (for example, testing hypotheses about whether there are typological differences between subjects based on measured characteristics); - using cluster analysis as a much simpler and more visual alternative to factor analysis, where the goal is only to group characteristics based on their correlation; - classifying objects based on direct assessments of the differences between them (for example, studying	Tasks solved by cluster analysis	OPK-2

	the social structure of a group based on sociometric data and identifying interpersonal preferences).		
16.	1. Selection of objects for clustering. Objects can be, depending on the purpose of the study: a) subjects; b) objects that are evaluated by subjects; c) features measured on a sample of subjects. 2. Definition of a set of variables, 3. Definition of a measure of difference between the objects of clustering. 4. Selection and application of a classification method to create groups of similar objects. 5. Verification of the reliability of the division into classes.	General sequence of cluster analysis	OPK-4
17.	The algorithm starts by finding the two closest objects, which form the primary cluster. Each subsequent object is assigned to the cluster that is closest to it.	Single link method (nearest neighbor method)	OPK-6,
18.	The merging rule of this method implies that a new object is assigned to the cluster whose farthest element is closer to the new object than the farthest elements of other clusters.	Complete link method (far neighbor method)	PC-2
19.	At each step, the average arithmetic distance between each object in one cluster and each object in another cluster is calculated.	Medium link method (unweighted pairwise average)	PC-3
20.	In the calculations, the size of the corresponding clusters (i.e., the number of objects contained in them) is used as a weighting factor. The proposed method should be used when unequal cluster sizes are expected.	Weighted pairwise average.	OPK-2
21.	In this method, the distance between two clusters is defined as the distance between their centers of gravity.	Unweighted centroid method.	OPK-4

22.	The calculations use weights to account for the difference in cluster sizes (i.e., the number of objects in each cluster).	Weighted centroid method (median).	OPK-6,
23.	It uses variance analysis methods to estimate the distances between clusters. The method minimizes the sum of squares (SS) for any two (hypothetical) clusters that can be formed at each step.	Ward's method.	PC-2
24.	- Research of the structure of interrelations of variables. - Identification of factors as hidden (latent) variables - causes of interrelations of initial variables. - Calculation of factor values for subjects as new, integral variables. At the same time, the number of factors is significantly less than the number of initial variables.	Main tasks of factor analysis	PC-3
25.	Factor analysis methods are different ways of obtaining a factor structure with a given number of factors. These methods differ in solving the problem of commonality. There are: 1. Principal component analysis. 2. Factor analysis of images. 3. The method of principal axes. 4. The method of unweighted least squares. 5. The method of maximum likelihood	Methods of factor analysis	OPK-2
26.	2,3,4	Data Mining is the process of discovering patterns in raw data. (1) previously formulated hypotheses (2) non-obvious patterns (3) practical patterns (4) objective patterns (5) a large number of patterns	OPK-4
27.	1	In the support vector method, classification is based on ... (1) not the entire set of samples, but only a small portion of them that lies on the boundaries	OPK-6,

		(2) the entire set of samples (3) a portion of the samples that lies outside the boundaries	
28.	1	A group of neuron synapses is ... (1) unidirectional input connections that are connected to the outputs of other neurons (2) the output connection of a given neuron, from which the signal (excitation or inhibition) is transmitted to the synapses of the following neurons (3) one or more neurons that receive the same common signal	PC-2
29.	1	Feedforward networks are ... (1) Perceptron (2) Hopfield networks (associative memory tasks) (3) Kohonen networks (clustering tasks)	PC-3
30.	1	A controversial clustering object is an object that, based on its similarity, can be assigned to multiple clusters. (1) can be assigned to multiple clusters (2) cannot be assigned to any cluster (3) can be assigned to more than two clusters	OPK-2
31.	1	When using which method, you must specify the number of clusters? (1) k-means method (2) nearest neighbor method (3) the entire group of hierarchical methods (4) all answers are incorrect	OPK-4
32.	1	A set is called frequent if: (1) Its support is above the user-defined minimum value (2) Its support is below the user-defined maximum value (3) Its support is equal to a user-defined value	OPK-6,
33.	1,2,3,5	Traditional visualization techniques can be used for the following purposes: (1) Presenting information to the user in a visual format (2) Compactly describing patterns within the original data set (3) Reducing the dimensionality or compressing information (4) Simplifying calculations within the model	PC-2

		(5) Filling in gaps within the data set	
34.	1,2,3	The essential concepts of a decision support system are: (1) computer-based interactive (2) decision support (3) for semi-structured and unstructured problems (4) for structured and semi-structured problems	PC-3
35.	1	For what kind of data set is it important to determine the presence of a seasonal component: (1) for ordered data (2) for unordered data (3) for both	OPK-2
36.	1,2	Errors that occur during the use of cleaning tools (which are the two extremes of data cleaning) are: (1) solving a problem that does not actually exist with a data cleaning tool (2) errors that occur when cleaning tools completely miss an existing problem (3) errors that occur when cleaning tools fail to detect an existing problem	OPK-4
37.	4	Data is: (1) facts and graphs (2) text (3) pictures, sounds, analog or digital video segments (4) all of the above	OPK-6,
38.	1	Data Mining models are built for the following purposes: (1) to research or study the modeled object and obtain new knowledge necessary for decision-making (2) to select the fastest-performing model (3) to explore all possible properties and characteristics of the studied object	PC-2
39.	1	Data Mining is a ... that must be integrated into the business. (1) not only a tool, but also a process (2) a tool	PC-3

		(3) a process	
40.	1,2,3	There are the following options for implementing Data Mining tools: (1) purchasing ready-made Data Mining software (2) purchasing Data Mining software tailored to a specific business (3) a combination of these options, including the use of various libraries, components, and toolkits for developers to create embedded Data Mining applications	OPK-2
41.	1	The SAS Enterprise Miner package is particularly useful for data analysis in ... (1) large organizations (2) medium-sized organizations (3) medium-sized and small organizations	OPK-4
42.	3	The PolyAnalyst system architecture ... (1) is a single-user option (2) is a multi-server enterprise solution (3) provides the ability to scale the system from a single-user option to a multi-server enterprise solution	OPK-6,
43.	1	The Cognos 4Thought software product is based on the technology of ... (1) neural networks (2) multiple regression (3) decision trees	PC-2
44.	1	Oracle Data Mining is a ... (1) feature in Oracle Enterprise Edition (2) standalone application (3) depending on the edition	PC-3
45.	2	Setting the task, building the optimal model, understanding the model, applying the results. The steps listed above are stages: (1) The traditional Data Mining process (2) The KXEN approach (3) both (4) neither one nor the other	OPK-2
46.	1,2,3	Data Mining services can be provided ...	OPK-4

		(1) in certain territories (2) in certain subject areas (3) using certain methods	
47.	1,2,3	Which of the following are the stages of Data Mining? (1) Free Search (2) Predictive Modeling (3) Exception Analysis (4) Rule Induction	OPK-6,
	1	Which of the following tasks establishes patterns between related events in a data set? (1) Association rule mining tasks (2) Sequential association rule mining tasks (3) Outlier analysis tasks	PC-2
48.	1	Classification is ... (1) assigning objects to one of the known classes (2) assigning objects to one of the unknown groups (3) the process of forming groups and assigning objects to one of them.	PC-3
49.	1	Which of the following is the main unit of time for which a forecast is made? (1) forecast period (2) forecast horizon (3) forecast interval	OPK-2
50.	1,2	Highlight two main areas of Web Mining: (1) Web Content Mining (2) Web Usage Mining (3) Web Text Mining	OPK-4
51.	1,2	Central tendency measurement characteristics: (1) mean (2) median (3) minimum (4) variance	OPK-6,

52.	1	If the dependent variable takes discrete values, the decision tree method solves the following problems: (1) classification (2) numerical forecasting (3) classification and numerical forecasting	PC-2
53.	1	Data Mining is the process of discovering knowledge in raw data that is necessary for: (1) making decisions in various fields of human activity (2) replacing an analyst in the decision-making process (3) increasing the cost of data analysis	PC-3
54.	1	Classification by the support vector method is considered good if the area between the boundaries is ... (1) empty (2) minimally filled (3) maximally filled	OPK-2
55.	1	A neuron has an axon, which is a ... (1) the output connection of a given neuron, from which the signal (excitation or inhibition) is transmitted to the synapses of the following neurons (2) unidirectional input connections that are connected to the outputs of other neurons (3) one or more neurons that receive the same common signal	OPK-4
56.	2,3	Feedforward networks are... (1) Perceptron (2) Hopfield networks (associative memory tasks) (3) Kohonen networks (cluster analysis tasks)	OPK-6,
57.	1	Which method requires a priori information about the number of clusters? (1) k-means method (2) nearest neighbor method (3) the entire group of hierarchical methods (4) all answers are incorrect	PC-2
58.	1	A transaction is a set of events that occurred ...	PC-3

		(1) simultaneously (2) one after another (3) both answers are incorrect	
59.	2,3,4	Traditional visualization techniques can be used for the following purposes: (1) as a supporting tool for data analysis (2) to reduce the dimensionality or compress information (3) to fill in gaps in a data set (4) to identify noise and outliers in a data set (5) all of the above	OPK-2
60.	1,4	Describe unstructured problems (1) have only a qualitative description based on the decision-maker's judgment, and the quantitative relationships between the main characteristics of the problem are unknown (2) are characterized by significant relationships that can be expressed quantitatively (3) combine quantitative and qualitative relationships, with the less well-known and uncertain aspects of the problem tending to dominate (4) these are the types of problems faced by managers	OPK-4
61.	1	If the data set is ordered and has a seasonal or cyclical component, what is the minimum amount of data required for analysis? (1) data for one season/cycle (2) data for half a season/cycle (3) data for two seasons/cycles	OPK-6,
62.	1	A Type 1 error occurs when ... (1) the data cleaning tool tries to solve a problem that doesn't actually exist (2) the data cleaning tool completely misses an existing problem (3) the data cleaning tool fails to detect an existing problem	PC-2
63.	4	Data can be obtained as a result of: (1) measurements (2) experiments	PC-3

		(3) arithmetic and logical operations (4) all of the above	
64.	1,2,3	The characteristics of a model are ... (1) the simplicity of the model in comparison with the object under study (2) the selection of the most significant factors in the object (3) the abstractness of the model	OPK-2
65.	1	Data Mining is not only a tool, but also a process that... (1) should be integrated into the business (2) can exist separately from the business (3) should precede the business	OPK-4
66.	1,2,3	There are the following options for implementing Data Mining tools: (1) development of a Data Mining product by a third-party company (2) development of a Data Mining product in-house (3) a combination of these options, including the use of various libraries, components, and toolkits for developers to create embedded Data Mining applications	OPK-6,
67.	3	Data Mining projects in SAS Enterprise Miner can be developed: (1) locally (2) in a client-server architecture (3) both options are possible	PC-2
68.	4	The unit of Data Mining research in PolyAnalyst is... (1) project tree (2) graph (3) rule (4) project	PC-3
69.	3	Cognos 4Thought is designed for ... (1) modeling (2) forecasting (3) both	OPK-2
70.	4	The Oracle Data Mining module is available from the following editions: (1) Personal Edition	OPK-4

		(2) Standard Edition (3) OneStandard Edition (4) Enterprise Edition (5) from all of the above editions	
71.	These are data that characterize the state of a parameter of an object under study. Most often, such data is represented by real numbers. Examples of numerical data include wages, population, blood pressure, and air temperature.	Numerical Data – ...	OPK-6,
72.	These are data that form a characteristic of belonging to a group. Examples of categorical data are an exam grade, the color of a car, and a person's level of education.	Categorical Data – ...	PC-2
73.	1) cluster analysis; 2) factor analysis; 3) discriminant function analysis; 4) multidimensional scaling; 5) log-linear analysis; 6) canonical correlation; 7) stepwise linear and nonlinear regression; 8) correspondence analysis; 9) analysis of time series and classification trees.	Methods of multidimensional analysis	PC-3
74.	• variable distribution analysis procedure (for example, to detect variables with non-symmetrical or non-Gaussian distributions, including bimodal distributions); • correlation matrix analysis to search for coefficients that exceed certain threshold values; • analysis of multi-input frequency tables (for example, "layer-by-layer" sequential analysis of combinations of control variable levels).	The main means of statistical analysis include:	OPK-2
75.	This is a method of classification analysis; its main purpose is to divide a set of objects and features under study into groups that are homogeneous in	Cluster analysis	OPK-4

	some sense, or clusters. This is a multivariate statistical method, so it is assumed that the initial data can be of significant volume, i.e., both the number of objects under study (observations) and the number of features that characterize these objects can be significantly large.		
76.	This is a procedure that reduces a large number of variables related to existing observations to a smaller number of independent variables called factors: variables that are highly correlated with each other are combined into a single factor.	Factor analysis	OPK-6,
77.	In discriminant analysis, a predictive model is created for group membership. This model constructs a discriminant function (or, when there are more than two groups, a set of discriminant functions) as a linear combination of predictor variables that best separates the groups.	Discriminant function analysis	PC-2
78.	A method of data analysis and visualization using the arrangement of points corresponding to the objects being studied (scaled) in a space of lower dimensionality than the feature space of the objects. The points are arranged so that the pairwise distances between them in the new space differ as little as possible from the empirically measured distances in the feature space of the objects being studied. If the elements of the distance matrix are obtained using interval scales, the method of multidimensional scaling is called metric. When the scales are ordinal, the method of multidimensional scaling is called non-metric. The measure of the differences in distances in the original and new spaces is called the stress function.	Multidimensional scaling	PC-3