

ОБЗОРЫ

УДК 338.242.2

**МЕТОДОЛОГИЯ КЛАСТЕРИЗАЦИИ ОБЪЕКТОВ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ПРИРОДЫ ПО УРОВНЮ РИСКА****Максимов Д.А., Митрофанов Д.П.***Российский экономический университет имени Г.В. Плеханова, Москва,**e-mail: maksimovdenis@mail.ru, dimsyb@gmail.com*

В работе представлены постановка задачи и описание численного алгоритма построения устойчивой классификации элементов иерархической социально-экономической системы по векторному критерию. Кластеризация элементов системы по однородным по уровню риска группам – актуальная задача, корректное решение которой обеспечит обоснованный выбор приоритетов развития системы и отдельных ее элементов. Авторами предложена методология кластеризации, основанная на использовании арсенала методов многомерного статистического анализа. В практической части работы представлен численный алгоритм и результаты классификации двадцати четырех предприятий холдинга АО «УГМК» по трём показателям – расширенному коэффициенту автономии, рентабельности рабочего капитала – ROA и рентабельности инвестиций в рабочий капитал – ROAAM – на четыре группы по уровню риска рыночной деятельности. Особое внимание уделено алгоритму получения устойчивой кластеризации методами иерархического и итеративного кластерного анализа, проверке её качества методами дискриминантного анализа и на основе результатов проверки многомерных статистических гипотез. Проведен всесторонний анализ полученной кластеризации, выделены предприятия холдинга, образующие отличные по уровню риска группы и сопоставлены уровни характеризующих их показателей. Была проанализирована возможность использования группировки при получении новых данных.

Ключевые слова: социально-экономические системы, интегрированные группы предприятий, риски рыночной деятельности, расширенный коэффициент автономии, коэффициент рентабельности рабочих активов, рентабельность инвестиций, многомерные статистические методы, стандартизация, итеративный и иерархический кластерный анализ, дискриминантный анализ, многомерные статистические гипотезы, критерий Хотеллинга

**METHODOLOGY OF CLUSTERING OF OBJECTS
OF SOCIO-ECONOMIC NATURE ACCORDING TO LEVEL OF RISK****Maksimov D.A., Mitrofanov D.P.***Plekhanov Russian University of Economics, Moscow, e-mail: maksimovdenis@mail.ru, dimsyb@gmail.com*

The article presents the formulation of the problem and the description of the numerical algorithm for constructing a stable classification of the elements of a hierarchical socio-economic system according to a vector criterion. Clustering elements of a system in groups that are homogeneous in terms of risk level is an urgent task, the correct solution of which will ensure a reasonable choice of priorities for the development of the system and its individual elements. The authors proposed a clustering methodology based on the use of an arsenal of methods for multivariate statistical analysis. In the practical part of the work, a numerical algorithm and classification results of twenty-four enterprises of JSC «UMMC» holding are presented according to three indicators – an extended autonomy ratio, working capital ROA profitability and working capital investment ROAAM – into four groups according to the risk level of market activity. Particular attention is paid to the algorithm for obtaining stable clustering using the methods of hierarchical and iterative cluster analysis, testing its quality using discriminant analysis methods and based on the results of testing multidimensional statistical hypotheses. A comprehensive analysis of the resulting clustering was carried out, the holding's enterprises were selected, forming groups of groups excellent in risk, and the levels characterizing their indicators were compared. The possibility of using grouping when receiving new data was analyzed.

Keywords: socio-economic systems, integrated group of enterprises, the risks of market activity, the equity / assets ratio, return on assets, return on investment, multivariate statistical methods, standardization, iterative and hierarchical cluster analysis, discriminant analysis, multivariate statistical hypothesis, the criterion Hotelling

Проблематика оценки и учета риска при выборе стратегии и тактики развития социально-экономических систем различного уровня – актуальная задача, рассмотрению которой посвящены работы цитируемых ниже Ф. Найта [1], Д. Власова [2], В. Гранатурова [3], И. Киселевой [4], А. Мищенко [5], В. Ступакова и Г. Токаренко [6], Н. Тихомирова [7], Л. Тэпмана [8], М. Халикова [9–11], А. Шеремета [12] и др. Актуальной является также проблематика

классификации объектов социально-экономической природы по уровню риска с целью выбора приоритетных стратегий развития этих систем в условиях высокой неопределенности макросреды их функционирования. Например, в работе [9] рассматривалась проблема кластеризации российских регионов по уровню социально-экономической безопасности. Для упрощения численного алгоритма решения этой и аналогичных задач классификации объ-

ектов социально-экономической природы авторы использовали интегральный показатель экономической безопасности региона в форме индекса, представляющего собой взвешенную сумму частных показателей безопасности жизненно важных сфер экономики и социальной сферы региона. Та же идея, но в приложении к крупным интегрированным структурам и, в частности, к вертикально-интегрированным холдингам использована и в работе М.А. Халикова [13]. В более поздней работе этого автора и его учеников [10] обобщены результаты кластеризации предприятий крупных промышленных компаний и вертикально-интегрированных холдингов по выбранному показателю риска.

Отметим, что такой подход является серьезным упрощением реальной картины, так как не раскрывает всей природы риска [7, 8] и концентрирует риск либо в одной сфере деятельности компании, либо заменяет этот риск каким-либо внешним или внутренним (например, рыночным) [4, 14].

Более корректным является подход, основанный на классификации объектов социально-экономической природы по однородным группам с использованием векторного критерия, компонентами которого являются показатели риска в отдельных сферах, по которым осуществляется оценка финансово-экономических результатов деятельности элементов социально-экономической системы. Цель исследования состоит в том, чтобы предложить подход к решению задачи кластеризации объектов социально-экономической природы, что позволит оценить различия между объектами.

Материалы и методы исследования

Рассмотрим эту задачу более подробно. Ниже представлена методология решения этой проблемы, включающая концепцию и численный алгоритм кластеризации сложной, многоуровневой социально-экономической системы по векторному критерию.

Для классификации элементов сложной социально-экономической системы по группам риска необходимо решить следующие задачи:

- выбрать и обосновать состав показателей риска;
- проверить возможность оценки выбранных показателей, а также возможность актуализации результатов при получении новых данных (например, на последующих временных интервалах);
- выбрать соответствующую типу исходных данных метрику признакового пространства;

– определить и обосновать количество групп риска;

– предложить метод кластеризации, основанный на инструментарии многомерного статистического анализа и позволяющий построить устойчивую кластеризацию. Устойчивость кластеризации необходима для обеспечения необходимого уровня качества модели.

Приведем возможный вариант и математическую модель решения поставленной проблемы. Введем следующие обозначения:

- n – число объектов;
- m – число критериев (признаков);
- i – номер объекта, $i = 1, n$;
- j – номер признака (показателя риска) $j = 1, m$;
- x_{ij} – значение j -го признака для i -го объекта;
- \bar{x}_j – среднее арифметическое j -го признака для всей совокупности объектов;
- σ_j – среднеквадратическое отклонение для j -го признака;

$$d(x_k, x_l) = \left(\sum_{i=1}^m |x_k - x_l|^p \right)^{1/p}$$

– расстояние между k -м и l -м объектами в общем виде рассчитываемое по формуле, приведенной в работе Н. Тихомирова, Т. Тихомировой и О. Ушмаева [15]. Параметр p устанавливается в зависимости от выбранной метрики.

- u – число кластеров;
- X_r, X_q – r -й и q -й кластеры;
- μ_r – центр r -го кластера.

Имея данные по показателям риска для каждого объекта необходимо провести стандартизацию данных с целью устранения негативный эффект влияния от масштаба. Стандартизация проводится по формуле [15–17]:

$$\dot{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \tag{1}$$

В результате стандартизации получим пригодные для исследования данные, по которым можно построить качественную модель кластеризации, независимую от разницы масштабов единиц измерения выбранных показателей риска.

Далее в зависимости от типа данных необходимо выбрать метрику признакового пространства. Метрика должна соответствовать типу данных для обеспечения точности модели кластеризации. Различают следующие виды расстояний [18]:

- евклидово расстояние – подходит для анализа количественных данных:

$$d(x_k, x_l) = \left(\sum_{i=1}^m |x_k - x_l|^2 \right)^{1/2} ; \tag{2}$$

– квадрат евклидова расстояния – также для оценки количественных данных, но эта метрика позволяет усилить различимость объектов в пространстве:

$$d(x_k, x_l) = \sum_{i=1}^m |x_k - x_l|^2; \quad (3)$$

– расстояние сити-блок – для анализа номинальных данных:

$$d(x_k, x_l) = \sum_{i=1}^m |x_k - x_l|. \quad (4)$$

На следующем этапе на основе визуального анализа проводится исключение «нетипичных» объектов с целью получения устойчивой кластеризации. К «нетипичным» относятся объекты, значения показателей риска которых значительно отличаются от основной группы объектов. Внесение данных объектов в модель кластеризации приводит к некорректной классификации.

Далее определяется число классификационных групп по уровню риска. Для этого следует применить такой метод кластерного иерархического анализа, как метод дальнего соседа [19]. Согласно этому методу расстояние между кластерами рассчитывается на основе двух наиболее удаленных друг от друга объектов из разных кластеров по формуле

$$d(X_r, X_q) = \max d(x_k, x_l). \quad (5)$$

На следующем этапе проводится предварительная группировка предприятий одним из наиболее эффективных методов кластерного анализа – методом *k*-средних [15]. Суть метода состоит в расчете центров тяжести для каждого из кластеров. Далее измеряются расстояния от объектов до центров тяжести, на основе информации по которым объекты причисляются к ближайшему кластеру. Процедура повторяется до получения устойчивой кластеризации. При реализации этого метода минимизируется суммарное квадратичное отклонение точек кластеров от их центров:

$$\sum_{r=1}^u \sum_{x \in X_r} (x_i - \mu_r)^2 \rightarrow \min. \quad (6)$$

$$T^2(\alpha, m, n_1, n_2) = \frac{(n_1 + n_2 - 2)m}{n_1 + n_2 - m - 1} F(\alpha, m, n_1 + n_2 - m - 1). \quad (9)$$

Также на основе расчётных значений статистики Хотеллинга:

$$T^2 = n(\bar{x}_i - x_j)^T S_i^{-1}(\bar{x}_i - x_j), \quad (10)$$

где S_i – выборочная ковариационная матрица i -го кластера, \bar{x}_i – вектор математического ожидания i -й группы, x_j – вектор показателей j -го элемента выборки, осуществляется проверка гипотез о равенстве векторов, характеризующих уровень рисков нетипичных объектов, математическим ожиданиям центров групп.

Результаты расчетов по методу *k*-средних позволяют получить устойчивую кластеризацию усеченного набора объектов исследуемой системы.

На завершающем этапе проводится попытка включения нетипичных объектов в модель кластеризации с использованием двойной проверки качества модели методами дискриминантного анализа и путем проверки гипотезы многомерного критерия Хотеллинга [18]. Благодаря дискриминантному анализу становится возможной расклассификация нетипичных объектов системы по кластерам. Построение дискриминантных функций по предварительно полученному разбиению позволяет также оценить вероятности отнесения объектов к тому или иному кластеру. Формула для расчета дискриминантных функций представлена ниже:

$$D(x) = \sum_{j=1}^m \alpha_j x_j. \quad (7)$$

Дискриминантный анализ позволяет выявить ошибки в обучающей выборке.

Окончательная проверка качества полученной группировки реализуется с помощью критерия Хотеллинга – многомерного критерия, позволяющего проверить нулевую гипотезу о равенстве двух векторов математических ожиданий (центров групп). Формула для расчета статистики Хотеллинга выглядит следующим образом [15]:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^T S_*^{-1} (\bar{x}_1 - \bar{x}_2), \quad (8)$$

где S_*^{-1} – матрица, обратная к матрице, составленной из центрированных значений i -й выборки.

Таким образом, в случае, если нулевая гипотеза принимается, т.е. расчётное значение статистики Хотеллинга оказывается меньше, чем табличное значение критерия $T^2(\alpha, m, n_1, n_2)$, то делается вывод о неразличимости двух центров групп, что в итоге позволяет объединить две группы в одну:

Нулевая гипотеза принимается, если расчётное значение статистики Хотеллинга оказывается меньше, чем табличное значение критерия. На основе $T^2(\alpha, m, n - m)$ подтверждается или отклоняется гипотеза о равенстве вектора j -го объекта вектору матожидания i -го кластера, что ведет к отнесению этого объекта к i -й группе:

$$T^2(\alpha, m, n - m) = \frac{(n - 1)m}{n - m} F(\alpha, m, n - m). \quad (11)$$

В противном случае подтверждается специфичность этих объектов, что приводит к невозможности их отнесения к той или иной группе и необходимости исследовать их обособленно от полученных кластеров.

По завершении приведенного алгоритма получаем качественную классификацию элементов социально-экономической системы по уровню риска.

Результаты исследования и их обсуждение

По предложенному алгоритму авторами были проведены расчеты на примере предприятий вертикально-интегрированного горно-металлургического холдинга

АО «УГМК» (двадцать четыре предприятия) [20, 21].

Данные по уровням рисков рыночной деятельности, включаемым в векторный критерий кластеризации, рассчитаны по методу, приведенному в работе [11], и приведены в табл. 1. В цитируемой работе в качестве показателей риска рассматриваются: коэффициент расширенной автономии (Ka), коэффициент эффективности операционной сферы (ROA) и показатель эффективности инвестиционной сферы (ROAAM).

Была проведена стандартизация данных. Полученные значения показателей находятся в пределах от -2,9 до 3,5. Так как данные являются количественными и значения показателей лежат в достаточно узком интервале, было принято решение использовать метрику квадрата евклидова расстояния. В этом случае будет получена достаточная различимость объектов в признаковом пространстве.

На следующем этапе после очищения выборки от «нетипичных» предприятий – структурных бизнес-единиц (СБЕ) холдинга было установлено, что метод дальнего соседа подтверждает предположение о необходимости разбиения выборки на пять групп по уровню риска (рис. 1).

Таблица 1

Показатели риска предприятий холдинга АО «УГМК»

№ п/п	Предприятие	Коэффициенты		
		Расширенной автономии	Эффективности операционной сферы	Эффективности инвестиционной сферы
1	ООО «БАШМЕДЬ»	0,241	0,085	0,195
2	АО «БУРИБАЕВСКИЙ ГОК»	0,786	0,122	0,145
3	ПАО «ГАЙСКИЙ ГОК»	0,558	0,066	0,083
4	ПАО «КЗОЦМ»	0,525	0,062	0,110
5	ООО «ММСК»	0,490	0,100	0,198
6	ПАО «РЗ ОЦМ»	0,171	0,057	0,255
7	ПАО «РКЗ»	0,689	0,073	0,079
8	АО «САФЬЯНОВСКАЯ МЕДЬ»	0,705	0,242	0,375
9	АО «СВЕРДЛОВСКАВТОДОР»	0,538	0,040	0,075
10	ПАО «СВЯТОГОР»	0,663	0,083	0,117
11	АО «СИБИРЬ-ПОЛИМЕТАЛЛЫ»	0,264	0,216	0,124
12	АО «СИБКАБЕЛЬ»	0,665	0,087	0,117
13	ПАО «СУМЗ»	0,125	0,146	0,291
14	АО «СУХОЛОЖСКОЕ ЛИТЬЕ»	0,303	0,207	0,279
15	АО «УГМК – ВЦМ»	0,126	0,027	0,219
16	АО «УК «КУЗБАССРАЗРЕЗУГОЛЬ»	0,392	0,134	0,218
17	АО «УРАЛКАБЕЛЬ»	0,493	0,089	0,181
18	АО «УРАЛЭЛЕКТРОМЕДЬ»	0,738	0,153	0,182
19	АО «УРУПСКИЙ ГОК»	0,859	0,183	0,207
20	АО «УЧАЛИНСКИЙ ГОК»	0,793	0,086	0,101
21	ПАО «ЧЦЗ»	0,816	0,156	0,194
22	АО «ШААЗ»	0,596	0,084	0,134
23	АО «ЭКЗ»	0,556	0,104	0,189
24	ПАО «ЭЛЕКТРОЦИНК»	0,269	0,060	0,237

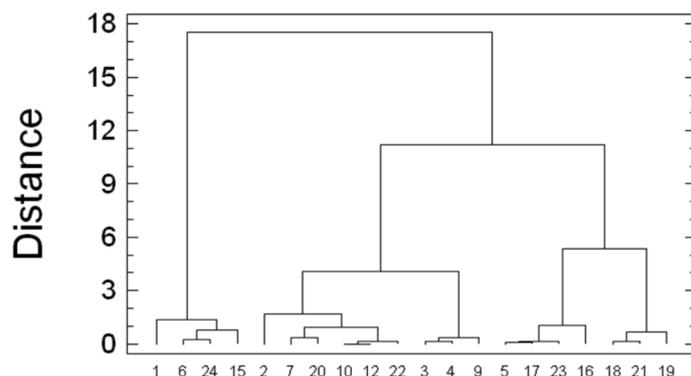


Рис. 1. Дендрограмма кластерного анализа предприятий холдинга АО «УГМК» методом «дальнего соседа»

Таблица 2

Распределение СБЕ АО «УГМК» по группам в соответствии с уровнем рисков рыночной деятельности методом k -средних

Группы предприятий (СБЕ) холдинга	Номера СБЕ
Первая: с критическим уровнем риска (4 СБЕ)	1, 6, 15, 24
Вторая: с высоким уровнем риска (3 СБЕ)	3, 4, 9
Третья: со средним уровнем риска (5 СБЕ)	7, 10, 12, 20, 22
Четвёртая: с приемлемым уровнем риска (4 СБЕ)	5, 16, 17, 23
Пятая: с низким уровнем риска (4 СБЕ)	2, 18, 19, 21

Таблица 3

Итоговая группировка подразделений АО «УГМК» по уровню риска

Группы предприятий (СБЕ) холдинга	Номера СБЕ
Первая: с критическим уровнем риска (4 СБЕ)	1, 6, 15, 24
Вторая: с высоким уровнем риска (3 СБЕ)	3, 4, 9
Третья: со средним уровнем риска (5 СБЕ)	7, 10, 12, 20, 22
Четвёртая: с приемлемым уровнем риска (4 СБЕ)	5, 16, 17, 23
Пятая: с низким уровнем риска (4 СБЕ)	2, 18, 19, 21

В табл. 2 приводятся результаты расчетов методом k -средних.

Применение дискриминантного анализа не позволило внести нетипичные объекты в кластеризацию.

Согласно критерию Хотеллинга была выявлена существенная различимость полученных кластеров.

После уточнения состава кластеров и подтверждения качества модели была получена итоговая группировка (табл. 3).

Первая группа характеризуется наименее худшими значениями показателей устойчивости в трёх сферах деятельности

предприятий холдинга, в то время как предприятия пятой группы являются наиболее устойчивыми (уровень риска минимален) (рис. 2).

Заключение

По результатам работы можно сделать вывод о том, что показатели ROA и ROAMM, как правило, растут при переходе от групп предприятий с наибольшим уровнем риска к группам с наименьшим. Так, ROA в первой группе составляет 0,06, а в четвертой – 0,15. Показатель Ка также коррелирует с уровнем риска.

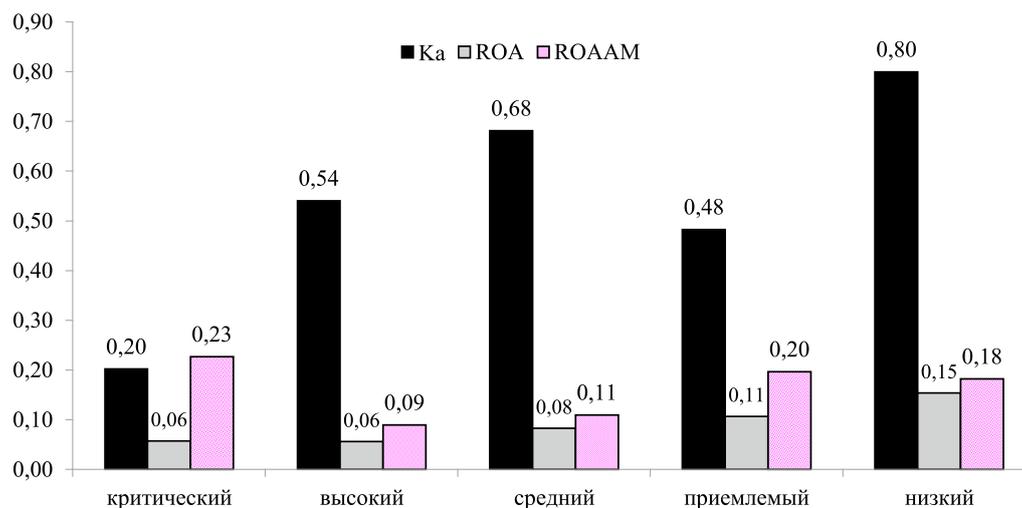


Рис. 2. График центров групп по уровням рисков рыночной деятельности

Таким образом, для формирования однородных по уровню риска рыночной деятельности групп предприятий холдинга целесообразно использовать метод k-средних при условии выделения пяти однородных по уровню риска групп предприятий – (критический, высокий, средний, приемлемый, низкий). Это разбиение является устойчивым по следующим причинам:

– прежде всего, применение в качестве первичных данных взвешенных значений используемых показателей рисков рыночной деятельности влечет к снижению случайных колебаний;

– реализация разнородных методов кластерного анализа позволила временно исключить подразделения холдинга, имеющие значительные отличия по показателям риска от основной массы предприятий, определить обоснованное число кластеров и получить наилучшее по выбранному критерию распределение подразделений по кластерам;

– возможность распределения ранее исключённых из выборки предприятий по уже выделенным кластерам позволяет существенно повысить качество классификации с учетом изменчивого финансово-экономического состояния подразделений холдинга.

По результатам проведенной работы устойчивость группировки была подтверждена. В случае потребности в актуализации используемых в расчетах данных необходимо пересчитывать показатели рисков рыночной деятельности с учетом новой информации. Однако повторная реализация использованного в работе инструментария не требуется. Тем не менее

необходима проверка качества модели кластеризации.

Практические расчеты по предложенной модели кластеризации с векторным критерием, проведенные на примере вертикально-интегрированного холдинга АО «УГМК», подтвердили приведенные выводы.

Список литературы

1. Найт Ф. Риск, неопределенность и прибыль. М.: Дело, 2003. 352 с.
2. Власов Д.А. Возможности Wolfram-технологий в исследовании простейших рискованных ситуаций // Фундаментальные и прикладные разработки в области технических и физико-математических наук: сборник научных статей по итогам работы Шестого международного круглого стола. М.: «КОНВЕРТ», 2018. С. 124–125.
3. Гранатуров В.М. Экономический риск: сущность, методы измерения, пути снижения. М.: Дело и Сервис, 2010. 208 с.
4. Kiseleva I.A., Tsetsgee B., Simonovich N.E. Risk management in the conditions of the economic crisis. Ulaanbaatar: Linograph. 2017, 153 p.
5. Мищенко А.В., Иванова А.В., Нестерович Л.Г., Кошелев П.С. Оптимизационные модели управления финансовыми ресурсами в логистике в условиях неопределенности и риска // Логистика сегодня. 2017. № 2. С. 82–100.
6. Ступаков В.С., Токаренко Г.С. Риск-менеджмент. М.: Финансы и статистика, 2005. 288 с.
7. Тихомиров Н.П. Риск-анализ в экономике. М.: Экономика, 2010. 318 с.
8. Тэпман Л.Н. Риски в экономике. М.: ЮНИТИ-ДАНА, 2002. 380 с.
9. Халиков М.А., Мынин Д.В. Методика комплексной оценки уровня социально-экономического развития регионов российской федерации и направления ее совершенствования // Региональная экономика: теория и практика. 2007. № 7. С. 59–63.
10. Халиков М.А., Хечумова Э.А., Щепилов М.В. Модели и методы выбора и оценки эффективности рыночной и внутрифирменной стратегий предприятия. М.: Российский экономический университет им. Г.В. Плеханова, 2015. 595 с.

11. Khalikov M.A., Maximov D.A., Shabalina U.M. Risk indicators and risk management models for an integrated group of enterprises. *Journal of Applied Economic Sciences*. 2018. vol. 13. no. 1 (55). P. 52–64.
12. Шеремет А.Д. Анализ и диагностика финансово-хозяйственной деятельности предприятия. М.: Инфра-М, 2011. 367 с.
13. Халиков М.А. Методы анализа и оценки риска рыночной деятельности подразделений иерархической производственной структуры // *Менеджмент в России и за рубежом*. 2009. № 1. С. 108–120.
14. Mishchenko A.V., Khalikov M.A. Distribution of organic resources in the problem of optimizing the production of an enterprise. *Journal of Computer and Systems Sciences International*. 1993. vol. 31. no. 6. P. 113–118.
15. Тихомиров Н.П., Тихомирова Т.М., Урмаев О.С. Методы эконометрики и многомерного статистического анализа. М.: Экономика, 2011. 647 с.
16. Власов Д.А. Инструментальное средство @RISK в системе прикладной математической подготовки // *Ярославский педагогический вестник*. 2018. № 3. С. 101–108.
17. Власов Д.А., Синчуков А.В. Потенциал Wolfram-технологий в построении и исследовании эконометрических моделей // *Современные информационные технологии и ИТ-образование*. 2017. Т. 13. № 4. С. 289–295.
18. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 1998. 352 с.
19. Дюран Б., Одел П. Кластерный анализ. М.: Статистика, 1977. 128 с.
20. Бухгалтерская (финансовая) отчетность предприятий [Электронный ресурс]. URL: <http://e-colog.ru> (дата обращения: 20.05.2019).
21. Официальный сайт АО «УГМК» [Электронный ресурс]. URL: <http://ugmk.com> (дата обращения: 20.05.2019).