

Благодарности

В работе использованы экспериментальные данные, полученные на уникальной научной установке «Сферический токамак Глобус-М», входящей в состав ФЦКП «Материаловедение и диагностика в передовых технологиях» (уникальный идентификатор проекта RFMEFI62119X0021). Коллектив авторов выражает благодарность научному сотруднику ФТИ им. А.Ф. Иоффе Курскиеву Г.С. за подробные консультации по анализу экспериментальных данных.

Список литературы

1. Шафранов В.Д., Бондаренко Б.Д., Гончаров Г.А., Лаврентьев О.А., Сахаров А.Д. К истории исследований по управляемому термоядерному синтезу // Успехи физических наук. 2001. № 8 (171). С. 877–886. URL: <https://ufn.ru/ru/articles/2001/8/o/> (дата обращения: 23.04.2020). DOI: 10.3367/UFNr.0171.200108o.0877.
2. Брюс П., Брюс Э. Практическая статистика для специалистов Data Science: 50 важнейших понятий / Пер. с англ. СПб.: БХВ-Петербург, 2018. 303 с.
3. Системный анализ и принятие решений: Словарь-справочник / Под ред. В.Н. Волковой, В.Н. Козлова. М.: Высшая школа, 2004. 616 с.
4. Vega J., Murari A., Gonzalez S. A universal support vector machines based method for automatic event location in waveforms and video-movies: Applications to massive nuclear fusion databases // Review of scientific instruments. 2010. №81.

УДК 330.1

doi:10.18720/SPBPU/2/id20-253

*Нгуен Тхи Тху Зунг*¹,

студент ИКНТ;

*Черенькая Людмила Васильевна*²,

д-р техн. наук, профессор ИКНТ

МОДЕЛЬ ДЛЯ АНАЛИЗА РАЗВИТИЯ ЭКОНОМИКИ РАЙОНА ТХАЙ-БИНГ (ВЬЕТНАМ) НА ОСНОВЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ МНОГОМЕРНОЙ СТАТИСТИКИ

^{1,2} Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия,

¹ thudung.mta.tb@gmail.com, ² ludmila@qmd.spbstu.ru

Аннотация. В социально-экономических системах широко используются методы многомерной статистики. Для определения факторов экономического развития был проведен сбор сведений о нескольких районах Вьетнама. На основе анализа методов многомерной статистики для решения поставленной задачи были выбраны два метода: метод факторного анализа и метод корреляционно-регрессионного анализа. Разработана модель, включающая методы расчета рейтинга и прогноза развития экономики нескольких районов Вьетнама. Модель реализована с

помощью пакета прикладных программ Matlab. Проведена апробация разработанной модели и анализ результатов. Выявлены факторы, влияющие на социально-экономические характеристики региона.

Ключевые слова: статистика, методы многомерной статистики, факторный анализ, корреляционно-регрессионный анализ, анализ развития экономики.

*Nguyen Thi Thu Dung*¹,
Master Student, Institute of Computer Science and Technology;
*Liudmila V. Chernenkaya*²,
Doctor of Technical Sciences, Professor,
Professor of Institute of Computer Science and Technology

MODEL FOR ANALYSIS OF THE ECONOMIC DEVELOPMENT OF THAI BINH (VIETNAM) ON THE BASIS OF MATHEMATICAL METHODS OF MULTIDIMENSIONAL STATISTICS

^{1,2} Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russia,
¹ thudung.mta.tb@gmail.com, ² ludmila@qmd.spbstu.ru

Abstract. In socio-economic problems, multidimensional statistical methods are widely used. For the determination of economic development factors, the accumulating of data about several regions of Vietnam has been carried out. Based on the analysis of multidimensional statistical methods, for solving of the stated task two methods were chosen: methods for factor analysis and correlation- regression analysis. The model, that includes methods for calculation of rating and the forecast for economic development of selected regions of Vietnam, has been developed. The simulation was realized with using of the application package Matlab. The developed model is tested, and the analysis of results has been carried out. Factors, that bring impacts on the social-economic characteristics of the region, are determined.

Keywords: statistics, multidimensional statistical methods, factor analysis, correlation and regression analysis, analysis of economic development.

Введение

Каждый район Вьетнама представляет собой целостную социально-экономическую систему, которая характеризуется множеством параметров. Как для анализа любой социально-технической системы [1], так и для анализа развития экономики района целесообразно использовать многомерный статистический анализ – раздел математической статистики, посвященный математическим методам построения оптимальных планов сбора, систематизации и обработки многомерных статистических данных с целью выявления характера и структуры взаимосвязей между компонентами каждого исследуемого многомерного признака.

1. Постановка задачи

Цель исследования заключается в обосновании выбора математических методов многомерной статистики и построении модели анализа экономики района Тхай-бинг (Вьетнам). Статистический анализ многофакторной статистики позволяет перейти от исходной системы, которая характеризуется многими экономическими показателями, к новой модели, которая характеризуется неродственными компонентами или факторами [2, 3]. Метод факторного анализа позволяет уменьшить размерность исходных данных при минимальной потере исходной информации [4].

Для изучения взаимосвязей параметров системы использован корреляционно-регрессионный анализ, построено уравнение регрессии, характеризующее зависимость признака от определяющих его факторов [5].

Разработана модель для анализа развития экономики района Тхай-Бинга (Вьетнам), основанная на комбинации двух математических методов многомерной статистики: метода факторного анализа и метода корреляционно-регрессионного анализа.

2. Решение задачи

Для выполнения этих задач в этом исследовании были предприняты следующие шаги:

Шаг 1. Сбор статистических данных для анализа.

Шаг 2. Проведение анализа по методу факторного анализа и методу корреляционно-регрессионного анализа с помощью пакетов прикладных программ Matlab и Statistica.

Шаг 3. Анализ и подтверждение полученных результатов.

На подготовительном этапе практические данные района Тхай-Бинга, включающие 44 под-факторные переменные, принадлежащие 13 факторам, были закодированы. Факторы, которые были рассмотрены для проведения анализа данных, были следующими:

1. Рыночный спрос на продукцию
2. Доступ к материальным и нематериальным ресурсам
3. Доступ к земле и жилью
4. Деловой климат
5. Местные налоги
6. Местные школьные системы
7. Утилиты
8. Экологическая осведомленность
9. Спорт и культура
10. Качество рабочей силы
11. Транспортная инфраструктура
12. Политика правительства в области экономического развития
13. Другие

2.1. Анализ на основе методов факторного анализа

Замечание 1. Для получения при использовании факторного анализа надежных результатов рекомендуется, чтобы размер выборки был в 4–5 раз больше числа исходных переменных[6].

В данном случае размер выборки составлял 1000 опрошенных, поэтому первое требование выполняется.

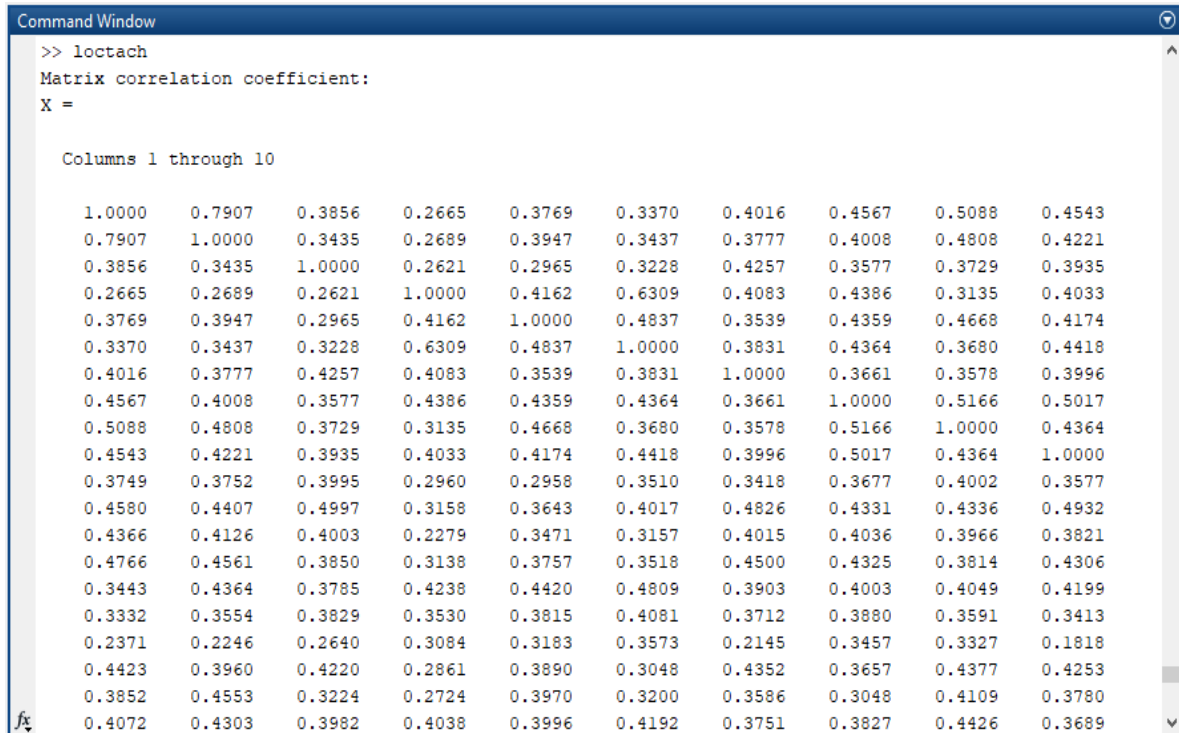


Рис. 1. Матрица корреляций

Проверка корреляционных коэффициентов между под-факторами. На рисунке 1 представлена корреляционная матрица, которая отражает попарную корреляцию между возможными парами переменных, включённых в анализ. Проанализировав полученную матрицу, можно сделать вывод, что большинство корреляционных коэффициентов между под-факторами больше 0,3, что хорошо для анализа.

Замечание 2. Факторный анализ эффективен, если между исходными переменными существует заметная корреляция.

Чтобы убедиться, что это так, осуществляется проверка по критерию выборочной адекватности Кайзера – Мейера – Олкина (КМО) [7]. Факторный анализ целесообразно применять, если значение КМО находится в пределах от 0,5 до 1. В нашем случае этот показатель равен 0.8976 (примерно 89,76 %), что является хорошим результатом.

На рисунке 2 представлен вывод графика каменистой осыпи (Scree Plot), т. е. графика последовательного убывания собственных значений матрицы коэффициентов корреляции между исходными переменными.

При этом собственное значение (eigenvalue) должно быть больше 1. Как следует из рисунка 2, имеется 10 компонент, для которых eigenvalues > 1, поэтому число компонент будет 10.

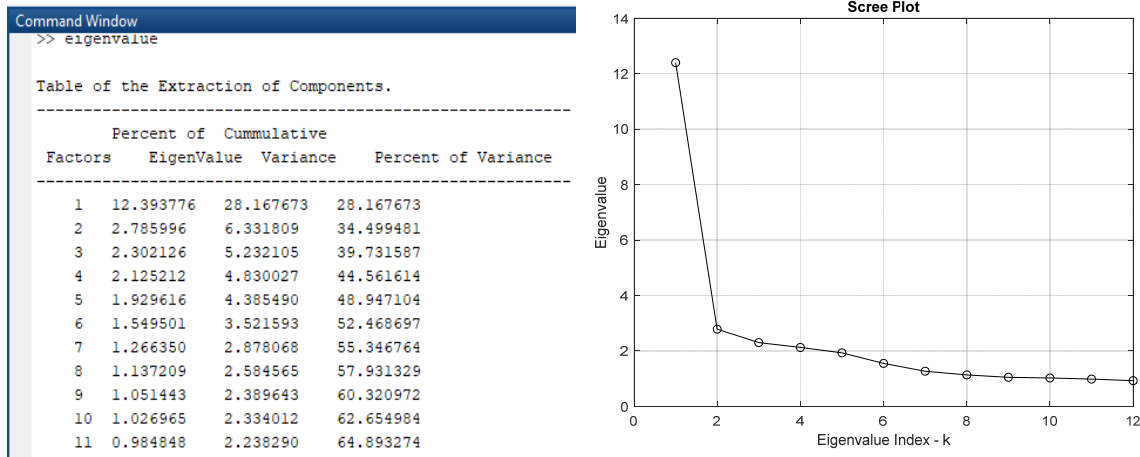


Рис. 2. График каменной осыпи

На рисунке 3 представлен результат применения метода вращения факторов.

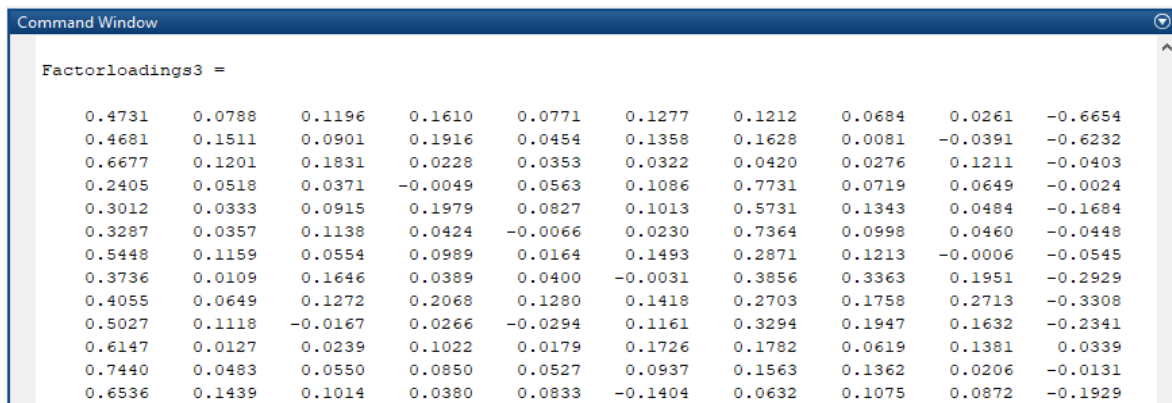


Рис. 3. Результат применения методов вращения факторов

Факторные нагрузки будут отсортированы следующим образом: сверху будут расположены исходные переменные, которые сильнее всего коррелируют с первым фактором, они будут отсортированы по убыванию модуля коэффициентов с этим фактором; затем аналогичным образом будут расположены исходные переменные, сильнее всего коррелирующие со вторым фактором и т. д. [5]. Кроме того, в данном случае факторные нагрузки, по модулю меньшие чем 0,5, не будут учитываться.

Сокращение количества переменных и определение структуры взаимосвязей между переменными было проведено с использованием факторного анализа. Получено сокращение 13 основных факторов и 44 под-фактора до 10 основных факторов и 33 под-факторов, что существенно упрощает дальнейшую обработку данных. Факторы, которые

оказывают наибольшее влияние на экономическое развитие региона, были выделены в первую группу факторов. К ним относятся «Деловой климат», «Местные налоги» и др. Вторая группа факторов включает, в частности, показатель «Доступ к материальным и нематериальным ресурсам». В третью группу факторов входят такие показатели, как «Среда жизни», «Возможность и квалификация сотрудников» и др.

Были установлены связи и степени влияния между переменными, определены расчетные значения зависимой переменной по методу корреляционно-регрессионного анализа [7]. В результате были получены модели зависимости показателей «Качество жизни», «Общий бизнесклимат», «Доступ к маркетинговым и рекламным услугам» и определено, что в соответствии с коэффициентами, рассчитанными в работе, наибольшее влияние оказывает показатель «Уровень преступности».

2.2. Анализ на основе корреляционно-регрессионного анализа

Регрессионный анализ проводился на основе построения и последующего анализа регрессионных моделей. Для моделей были определены важные переменные, зависимости и было построено уравнение регрессии.

Пример регрессионного результата для переменной показан на рисунке 4.

	b*	Std. Err. of b*	b	Std. Err. of b	t(1989)	p-value
Intercept			-0.440021	0.171720	-2.56243	0.010467
Q1b	0.646492	0.016762	0.602196	0.015614	38.56859	0.000000
Q1c	-0.004747	0.015671	-0.007238	0.023897	-0.30289	0.762007
Q2a	-0.004433	0.017267	-0.006372	0.024820	-0.25673	0.797410
Q2b	-0.014540	0.015978	-0.012626	0.013874	-0.90999	0.362939
Q2c	0.003449	0.017518	0.003503	0.017789	0.19690	0.843926
Q3a	0.019110	0.015628	0.024942	0.020398	1.22280	0.221552
Q3b	0.071158	0.016600	0.088468	0.020639	4.28653	0.000019
Q4a	0.084726	0.016874	0.087319	0.017390	5.02111	0.000001
Q4b	0.054184	0.016496	0.063248	0.019256	3.28465	0.001039
Q4c	0.016696	0.015583	0.024741	0.023093	1.07138	0.284129
Q4d	0.014882	0.017475	0.021524	0.025275	0.85160	0.394541
Q5a	0.050620	0.016180	0.044188	0.014124	3.12849	0.001783
Q5b	0.008675	0.016398	0.008247	0.015589	0.52904	0.596839
Q6a	-0.140735	0.020659	-0.115631	0.016974	-6.81236	0.000000
Q6b	0.044938	0.019335	0.040226	0.017307	2.32420	0.020215

Рис. 4. Пример результатов расчета регрессии

Коэффициент детерминации $R^2 = 0.7251$ показывает, что изменчивость зависимой переменной на 72,51 % объясняется изменчивостью других переменных. Значение F -критерия, который используется для проверки гипотезы, равно 121,01. Значение уровня значимости гипотезы о равенстве нулю коэффициентов регрессии p мало [8].

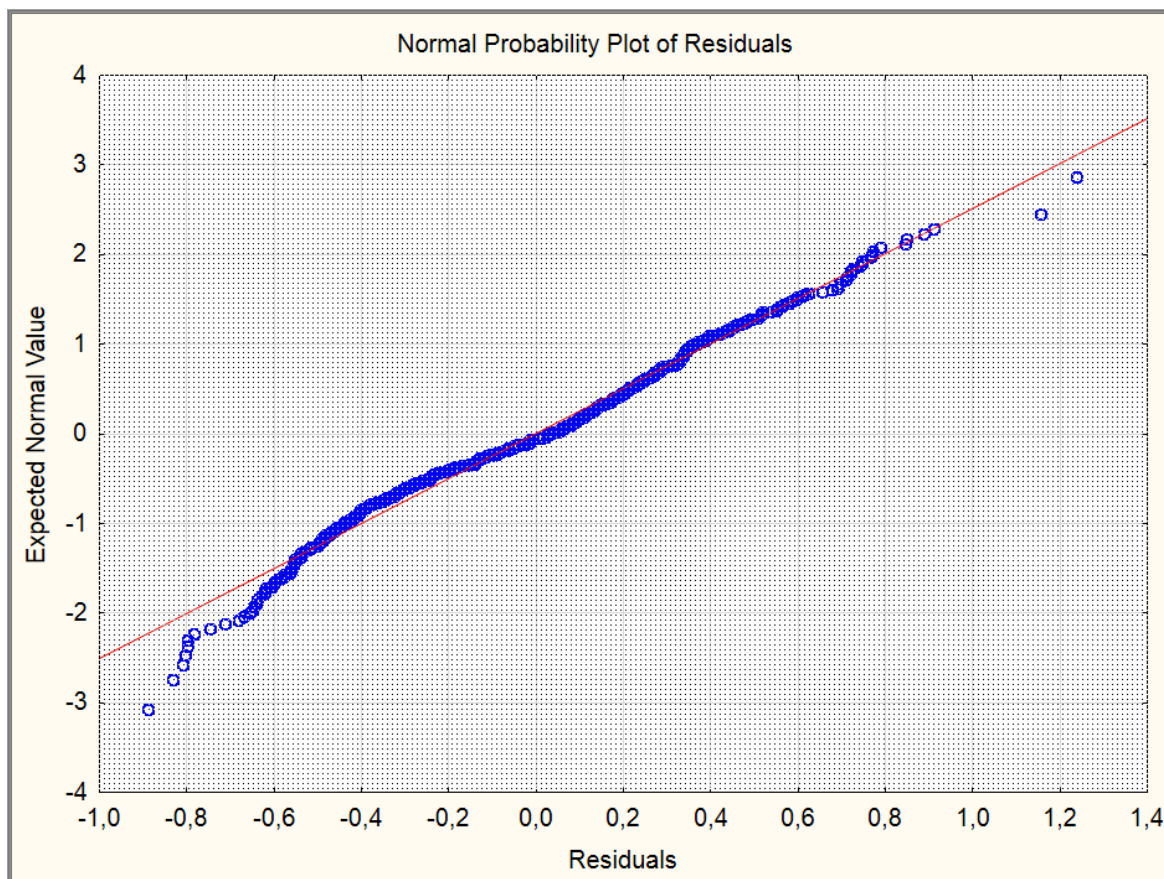


Рис. 5. График остатков для переменной x_1

Остатки находятся на линии, соответствующей нормальному правилу, при условии, что выполняется нормальная ошибка распределения. Поэтому запрошенная модель имеет вид:

В результате модель регрессии для переменной x_1 можно представить следующим образом:

$$\begin{aligned}
 x_1 = & -0.44 + 0.602x_2 - 0.007x_3 - 0.006x_4 + 0.013x_5 + 0.004x_6 + 0.025x_7 + \\
 & + 0.088x_8 + 0.088x_9 + 0.063x_{10} + 0.025x_{11} + 0.022x_{12} + 0.044x_{13} + \\
 & + 0.008x_{14} - 0.116x_{15} + 0.04x_{16} - 0.053x_{17} + 0.097x_{18} - 0.073x_{19} - \\
 & - 0.025x_{20} + 0.064x_{21} + 0.113x_{22} + 0.135x_{23} + 0.018x_{24} - 0.107x_{25} + \\
 & + 0.172x_{26} + 0.075x_{27} + 0.031x_{28} + 0.023x_{29} + 0.011x_{30} + 0.063x_{31} - \\
 & - 0.006x_{32} + 0.012x_{33} - 0.06x_{34} - 0.08x_{35} + 0.011x_{36} - 0.021x_{37} - \\
 & - 0.077x_{38} - 0.084x_{39} - 0.039x_{40} + 0.051x_{41} - 0.112x_{42} + 0.096x_{43} - \\
 & - 0.106x_{44}.
 \end{aligned}$$

Из управления видно, какие переменные влияют на значение x_1 в большей или меньшей степени в зависимости от величины стоящего перед переменной множителя. Положительный коэффициент свидетельст-

вует о прямой зависимости между переменными, отрицательный – об обратной зависимости.

Другими словами, чем больше значения факторов x_2, x_3, x_4, \dots , тем больше значение x_1 , и, наоборот, чем больше значение факторов x_2, x_3, x_4, \dots , тем меньше значение x_1 . Таким образом, из построенной модели мы получили, что, чем больше значения «Близость к рынку для вашей продукции», «Доступ к сырью/материалам производства», «Доступ к производственным/ производственным услугам», ..., тем больше значение «Доступ к рынку для продукции» и наоборот, чем больше значение «Доступ к инженерным, исследовательским и производственным объектам», ..., тем меньше значения «Доступ к рынку для вашей продукции».

Аналогично проводится анализ для остальных переменных.

3. Рекомендации по развитию экономики района Тхай-Бинг, Вьетнам

Анализ результатов проведенного исследования показал, что для развития экономики района Тхай-Бинг следует сосредоточиться на факторах, относящихся к первой группе. Среди них можно отметить минимизацию уровня преступности, в частности, уровня экономических преступлений. Для этого, как показал корреляционно-регрессионный анализ, необходимо поддерживать благоприятный деловой климат, предоставлять налоговые льготы, вводить государственную систему регулирования. Эти факторы имеют первостепенное значение для компаний, организующих свой бизнес в данном районе Вьетнама. В то же время власти района Тхай-Бинг должны пропагандировать принимаемые законодательные меры, рекламировать изменения в законодательной сфере, а также предотвращать возможные нарушения экономического характера со стороны компаний.

Заключение

В работе были изучены статистические данные, с использованием факторного анализа классифицированы группы новых факторов с уменьшением общего количества факторов. Для решения задач исследования был использован пакет прикладных программ Matlab.

Полученным новым группам факторов присвоены более подходящие по смыслу названия, данные новые элементы использованы как база для практического анализа развития каждого района.

Применение корреляционно-регрессионного анализа позволило получить взаимосвязь между факторами и определить факторы, оказывающие наиболее сильное влияние на основные факторы, которые требуется проанализировать.

На основании комплексной методики, объединяющей различные методы многомерной статистики, и проведения практического анализа определены направления экономического развития в каждом районе Вьетнама.

Предложенные модели для анализа развития экономики районов Вьетнама на основе математических методов многомерной статистики прошли апробацию для нескольких районов Вьетнама и могут найти практическое применение при проведении экономических исследований в стране.

Список литературы

1. Volkova V.N., Loginova A.V., Chernenkaja L.V., Romanova E.V., Chernyy Y.Y., Lankin V.E. Problems of sustainable development of socio-economic systems in the implementation of innovations // Proceedings of the 3rd International Conference ERGO-2018: Human Factors in Complex Technical Systems and Environments (ERGO-2018). 4–7 July 2018, Saint Petersburg Electrotechnical University “LETI”, St. Petersburg, Russia. Publisher: IEEE, 2018. P. 53–56.

2. Leonova T.I., Mager V.E., Mikeladze B.D., Chernenkaya L.V., Chernenkii A.V. Support of decision-making in organizations' quality management // Proceedings of the 2017 XX IEEE International conference on soft computing and measurement (SCM 2017). 1 May 2017, Saint Petersburg Electrotechnical University “LETI”, St. Petersburg, Russia. Publisher: IEEE, 2017. P. 843–845.

3. Chernenkaya L.V., Desyatirikova E.N., Belousov V.E., Chepelev S.A., Sergeeva S.I., Slinkova N.V. Optimal planning of distributed control systems with active elements. // Proceedings of the 2017 IEEE II International Conference on Control in Technical Systems (CTS 2017). 3–4 October 2017, Saint Petersburg Electrotechnical University “LETI”, St. Petersburg, Russia. Publisher: IEEE, 2017. P. 37–40. DOI: 10.1109/CTSYS.2017.8109482.

4. Чураков Е. П. Введение в многомерные статистические методы: Учебное пособие. СПб.: Лань, 2016. 146 с.

5. Кулаичев А.П. Методы и средства комплексного статистического анализа данных: Учебное пособие. 5-е изд., перераб. и доп. М.: Форум : ИНФРА-М, 2017. 484 с.

6. SPSS survival manual: A step by step guide to data analysis using SPSS. Sydney: Allen & Unwin, 2013. 334 p.

7. Кисляк Н.В., Шорохова И.С., Мариев О.С. Статистические методы анализа: Учебное пособие. М.: Флинта, 2017. 300 с.

8. Алабин М.А. Корреляционно-регрессионный анализ статистических данных в двигателестроении. М.: Машиностроение, 1974. 122 с.