

ОРИГИНАЛЬНАЯ СТАТЬЯ



DOI: 10.26794/2220-6469-2019-13-4-115-125

УДК 330.47(045)

JEL C38

Структурирование многомерных данных в исследовании развития информационного общества в регионах России

Н.М. Габдуллин

Казанский федеральный университет,
<https://orcid.org/0000-0003-3393-3707>

И.А. Киршин

Казанский федеральный университет,
<https://orcid.org/0000-0002-7407-7188>

АННОТАЦИЯ

В статье проведен анализ структуры многомерных данных по факторам использования информационных и коммуникационных технологий (ИКТ) для развития информационного общества в Российской Федерации. В качестве объектов наблюдений были выбраны субъекты РФ, а в качестве факторов каждого объекта наблюдений — среднегодовые величины соответствующих факторов мониторинга развития информационного общества в Российской Федерации за период с 2010 по 2017 г. Исследование выполнено в целях редукции множества факторов использования ИКТ в субъектах РФ и выявления их структуры. Анализ структуры факторов выполнен посредством применения метода главных компонент (Principal Component Analysis, PCA), реализованного авторами в интегрированной среде разработки Rstudio. В результате применения метода PCA выявлена структура исходных факторов использования ИКТ для развития информационного общества в Российской Федерации, произведено их редуцирование. Показано, что выявленные факторные нагрузки имеют определенные смысловые интерпретации, консолидирующие связи отдельных факторов использования ИКТ для развития информационного общества в Российской Федерации. Проведенный факторный анализ обосновывает эффективность применения метода PCA в исследованиях по развитию цифровизации регионов Российской Федерации для структурирования исходных данных и качественной интерпретации результатов. Доказано, что выявляемые посредством применения метода PCA факторные нагрузки имеют детерминированный экономический смысл.

Ключевые слова: Big Data; анализ структуры данных; цифровизация экономики; кластеризация; метод главных компонент; информационные и коммуникационные технологии

Для цитирования: Габдуллин Н.М., Киршин И.А. Структурирование многомерных данных в исследовании развития информационного общества в регионах России. *Мир новой экономики*. 2019;13(3):115-125. DOI: 10.26794/2220-6469-2019-13-4-115-125

ORIGINAL PAPER

Structuring Multidimensional Data in the Study of the Development of Information Society in Russian Regions

N.M. Gabdullin

Kazan Federal University,
<https://orcid.org/0000-0003-3393-3707>

I.A. Kirshin

Kazan Federal University,
<https://orcid.org/0000-0002-7407-7188>

ABSTRACT

In the article, the authors present the results of the analysis of the structure of multidimensional data on the factors of use of information and communication technologies (ICT) for the development of the information society in the

Russian Federation. We chose the subjects of the Russian Federation as the objects of observations, and the average annual values of the corresponding factors of monitoring the development of the information society in the Russian Federation for the period from 2010 to 2017 as the factors of each object of observations. The study was carried out to reduce the set of factors of ICT use in the subjects of the Russian Federation and identify their structure. The analysis of the structure of factors is carried out by applying the Principal Component Analysis (PCA) method, implemented by the authors in the integrated development environment Rstudio. Through the use of the PCA method, we identified the structure of the initial factors of ICT use for the development of the information society in the Russian Federation and made their reduction. Then, we showed that the identified factor loads have specific semantic interpretations that consolidate the links of individual factors of ICT use for the development of the information society in the Russian Federation. The factor analysis carried out by us proves the effectiveness of the DCA method in research on the development of digitalisation of the regions of the Russian Federation for structuring the initial data and qualitative interpretation of the results. We have proved that factor loadings revealed employing the application of a method of RSA have the deterministic economic sense.

Keywords: Big Data; data structure analysis; economy digitalisation; clustering; principal components method; information and communication technologies

For citation: Gabdullin N.M., Kirshin I.A. Structuring multidimensional data in the study of the development of information society in Russian regions. *Mir novoj ekonomiki = World of the New Economy*. 2019;13(4):115-125. DOI: 10.26794/2220-6469-2019-13-4-115-125

ВВЕДЕНИЕ

Цифровизация экономики обуславливает кардинальные трансформации всех сфер жизнедеятельности, определяя необходимость подготовки специалистов новой формации. Происходят изменения в структуре экономики, закономерностях ее развития, а поэтому активизируются исследования по взаимовлиянию этих процессов [1]. Безусловно, в составе ресурсного потенциала устойчивого роста национальных экономик в условиях применения цифровых технологий меняются приоритеты и значимость традиционных факторов производства [2].

Следуя мировой практике, Россия в последние годы реализует исследовательские программы, нацеленные на формирование методологии и методик по оценке уровня цифровизации отечественной экономики и определения актуальных характеристик цифровых технологий. Так, с 2017 г. московская школа управления «Сколково» составляет индекс «Цифровая Россия» на основе данных по всем субъектам РФ¹. Разработана оригинальная методология, сочетающая количественные показатели и качественную экспертную оценку, полученную в результате анализа метаданных, отражающих события и инициативы цифровизации российских регионов на основе публичных упоминаний в открытых источниках.

Цифровая грамотность выступает атрибутом современного человеческого капитала и представляет набор знаний, умений и компетенций, необходимых

для безопасного и эффективного использования цифровых технологий и ресурсов интернета. Структурно индекс цифровой грамотности РФ формируется из 20 ключевых параметров². Большой датасет актуальных статистических данных по развитию информационного общества в Российской Федерации формирует Федеральная служба государственной статистики РФ в разделе «Мониторинг развития информационного общества в Российской Федерации»³ (далее — Мониторинг).

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

Число наблюдаемых факторов в вышеописанных датасетах и методиках формирования индексов достаточно большое, и взаимосвязи между ними сложны и труднотруктурируемы. Структурирование больших данных повышает эффективность работы с ними, выявляет интегрированные конфигурации факторов, характеризующих отдельные стороны исследуемых объектов. Структурирование исходных факторов, характеризующих уровень цифровизации российской экономики, возможно посредством редукции размерности их выборки. Выдвинем гипотезу, согласно которой существует малое число факторов, влияющих на измеряемые параметры цифровизации экономики, достаточно точно приближающих

¹ Индекс «Цифровая Россия». URL: <https://finance.skolkovo.ru/ru/sfice/research-reports/1779-2018-10-001-ru/> (дата обращения: 28.03.2019).

² Цифровая грамотность. URL: <http://цифроваяграмотность.рф/mindex/program/> (дата обращения: 25.03.2019).

³ Мониторинг развития информационного общества в Российской Федерации (по состоянию на 03.10.2018). URL: http://www.gks.ru/free_doc/new_site/figure/anketa1-4.html (дата обращения: 03.10.2018).

наблюдаемые параметры и описывающих связи между ними. Воспользуемся для этого методом главных компонент (Principal Component Analysis, PCA)⁴ [3]. Применение метода главных компонент позволяет произвести редукцию числа факторов (переменных модели) и выявить *структуру* взаимосвязей между наблюдаемыми объектами (переменными) [4, 5]. Метод главных компонент выявляет предварительную кластерную структуру данных по цифровизации регионов и характеристики построенных кластеров регионов [6].

Главные компоненты представляют собой линейные комбинации исходных факторов (предикторов) и могут быть использованы вместо исходных факторов [7]. При этом новые ортогональные (некоррелированные) факторы-компоненты строятся в ходе численного эксперимента в порядке убывания объясняемой ими доли суммарной дисперсии исходных величин, что позволяет зачастую ограничиться несколькими первыми компонентами [8, 9]. Остальные главные компоненты объясняют остаточную малую «случайную» дисперсию, которой можно пренебречь. Процедура выделения главных компонент аналогична вращению, максимизирующему дисперсию (варимакс) исходного пространства переменных. Этот вид вращения максимизирует дисперсию, так как критерий (цель) вращения заключается в максимизации дисперсии (изменчивости) «новой» переменной (фактора) и минимизации разброса вокруг нее [10, с. 34; 11].

Исходная матрица факторов, число которых подлежит редукции, составлена из части статистических данных Федеральной службы государственной статистики «Мониторинг развития информационного общества в Российской Федерации» за период с 2010 по 2017 г. В данном исследовании проведем последовательное редуцирование множества факторов использования информационных и коммуникационных технологий (ИКТ) для развития посредством применения метода PCA. Программная реализация метода PCA произведена нами в интегрированной среде разработки Rstudio.

В качестве объектов наблюдений были выбраны субъекты РФ, а в качестве факторов каждого объекта наблюдений — среднегодовые величины соответствующих факторов за рассмотренный в мониторинге период.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Начинаем структурирование с факторов подгруппы 2.1 «Электронное правительство». По результатам анализа долей дисперсии, объясняемых главными компонентами, выбираем первые три главные компоненты, в совокупности объясняющие 85,38% общей дисперсии. Собственные значения главных компонент — это их дисперсии и собственные векторы корреляционной матрицы — нагрузки на компоненты [11, р. 8–27]. Собственные векторы определяют связь между переменными и главными компонентами, т.е. факторные нагрузки на компоненты — это коэффициенты корреляции переменных и выделенных главных компонент [12].

Определим собственные векторы первых трех главных компонент. Главная компонента обозначений факторов представлена в *табл. 1*.

PC1 (первая главная компонента) относительно сильно отрицательно коррелирует с F211 и F212, и поэтому ее можно охарактеризовать как компоненту, определяющую готовность органов государственной власти и органов местного самоуправления к использованию ИКТ для осуществления управленческих функций и предоставления государственных услуг. PC2 (вторая главная компонента) сильно отрицательно коррелирует с F2141 и F2142, следовательно определяет возможность размещения заказов на поставку товаров, выполнение работ и оказание услуг для государственных и муниципальных нужд с использованием электронных торговых площадок. PC3 (третья главная компонента) сильно отрицательно коррелирует с F215 «Доля электронного документооборота между органами государственной власти, в общем объеме межведомственного документооборота».

Метод PCA, относящийся к методам визуализации структуры данных [13, с. 143], позволяет идентифицировать кластеры как самих переменных, так и наблюдаемых объектов [14, 15] (в нашем случае — субъектов РФ). Визуализация вклада факторов в главные компоненты показана на графике biplot соответствующими факторными нагрузками (*рис. 1*). На полученном графике biplot представлена диаграмма рассеяния наблюдаемых объектов — регионов РФ в пространстве выделенных двух первых главных компонент. Порядковые номера субъектов РФ указаны в *табл. 2*.

В пакете RStudio предусмотрена возможность отображения наблюдений точками (номераами) на факторной плоскости [16]. По аналогии с факторными вкладами, вклады наблюдений показывают меру значимости наблюдения в качестве определителя факторной оси.

⁴ Factor Analysis: Statnotes from North Carolina State University. URL: https://www.encorewiki.org/download/attachments/25657/Factor+Analysis_+Statnotes+from+North+Carolina+State+University.pdf (дата обращения: 25.03.2019).

Таблица 1 / Table 1

Главная компонента обозначенных факторов / The main component of the indicated factors

№	Наименование исходного фактора	Обозначение
Готовность органов государственной власти (ОГВ) и местного самоуправления (ОМС) к использованию ИКТ для осуществления управленческих функций и предоставления государственных услуг		
2.1.1	Доля ОГВ и ОМС, использовавших интернет, в общем числе обследованных организаций ОГВ и ОМС – всего	F211
2.1.2	Доля ОГВ и ОМС, имевших скорость передачи данных через интернет не менее 2 Мбит/сек, в общем числе обследованных организаций ОГВ и ОМС	F212
Использование ИКТ в деятельности ОГВ и ОМС		
2.1.4.1	Доля размещенных заказов на поставки товаров, выполнение работ и оказание услуг для государственных и муниципальных нужд с использованием электронных торговых площадок в общем объеме размещенных заказов, количество торгов	F2141
2.1.4.2	Доля размещенных заказов на поставки товаров, выполнение работ и оказание услуг для государственных и муниципальных нужд с использованием электронных торговых площадок в общем объеме размещенных заказов, стоимость заключенных контрактов	F2142
2.1.5	Доля электронного документооборота между органами государственной власти в общем объеме межведомственного документооборота	F215
Использование организациями компьютеров и компьютерных сетей		
2.2.1	Доля организаций, использовавших персональные компьютеры, в общем числе обследованных организаций	F221
2.2.2	Число персональных компьютеров в расчете на 100 работников организаций	F222
2.2.3	Доля организаций, использовавших локальные вычислительные сети, в общем числе обследованных организаций	F223
2.2.4	Доля организаций, использовавших беспроводные локальные вычислительные сети, в общем числе обследованных организаций	F224
2.2.5	Доля организаций, использовавших интранет, в общем числе обследованных организаций	F225
2.2.6	Доля организаций, использовавших экстранет, в общем числе обследованных организаций	F226
2.2.7	Доля организаций, использовавших предоставляемые третьей стороной операционные системы с открытым исходным кодом (например, Linux), в общем числе обследованных организаций	F227
Использование интернета		
2.2.8	Доля организаций, использовавших интернет, в общем числе обследованных организаций	F228
2.2.9	Число персональных компьютеров, имевших доступ к интернету, на 100 работников организаций	F229
2.2.10	Доля организаций, использующих широкополосный доступ к сети интернет, в общем числе организаций	F2210
2.2.11	Доля организаций, использующих доступ к сети интернет со скоростью не менее 2 Мбит/с, в общем числе организаций	F2211
2.2.12	Доля организаций, имевших веб-сайт, в общем числе обследованных организаций	F2212
2.2.13	Доля организаций, имевших действующие информационно-справочные терминалы (инфоматы), в общем числе обследованных организаций	F2213
2.2.14	Доля организаций, использовавших электронную почту, в общем числе обследованных организаций	F2214

Продолжение табл. 1 / Continued Table 1

№	Наименование исходного фактора	Обозначение
Применение информационных и коммуникационных технологий (ИКТ) работниками организаций		
2.2.15	Доля работников организаций, использовавших персональные компьютеры не реже 1 раза в неделю, в общей численности работников списочного состава организаций	F2215
2.2.16	Доля работников организаций, использовавших интернет не реже 1 раза в неделю, в общей численности работников списочного состава организаций	F2216
2.2.17	Доля организаций, выделявших технические средства для мобильного доступа в интернет своим работникам, в общем числе обследованных организаций	F2217
2.2.18	Доля работников, использовавших предоставленные организацией средства мобильного доступа в интернет не реже 1 раза в неделю, в общей численности списочного состава организаций	F2218
Интеграция внутренних информационных систем (ИС) и совместный доступ к информации внутри организации		
2.2.19	Доля организаций, имевших специальные программные средства для управления закупками товаров (работ, услуг), в общем числе обследованных организаций	F2219
2.2.20	Доля организаций, имевших специальные программные средства для управления продажами товаров (работ, услуг), в общем числе обследованных организаций	F2220
2.2.21	Доля организаций, использовавших ERP-системы, в общем числе обследованных организаций	F2221
2.2.22	Доля организаций, использовавших CRM-системы, в общем числе обследованных организаций	F2222
2.2.23	Доля организаций, использовавших системы электронного документооборота, в общем числе обследованных организаций	F2223
Интеграция ИС организации с ИС контрагентов		
2.2.24	Доля организаций, использовавших электронный обмен данными между своими и внешними информационными системами по форматам обмена, в общем числе обследованных организаций	F2224
2.2.25	Доля организаций, использовавших SCM-системы, в общем числе обследованных организаций	F2225
Электронная коммерция		
2.2.26	Доля организаций, размещавших заказы на товары (работы услуги) в интернете, в общем числе обследованных организаций	F2226
2.2.27	Доля организаций, получавших заказы на выпускаемые товары (работы, услуги) по интернету, в общем числе обследованных организаций	F2227
Готовность учреждений образования к развитию на основе ИКТ		
2.3.1	Число персональных компьютеров, используемых в учебных целях, на 100 обучающихся государственных и муниципальных общеобразовательных учреждений	F231
2.3.2.2	Число персональных компьютеров, используемых в учебных целях, находящихся в составе локальных вычислительных сетей (ЛВС), в расчете на 100 студентов (обучающихся) по образовательным учреждениям среднего профессионального образования (СПО)	F2322
2.3.2.3	Число персональных компьютеров, используемых в учебных целях, находящихся в составе локальных вычислительных сетей (ЛВС), в расчете на 100 студентов (обучающихся) по образовательным учреждениям высшего профессионального образования (ВПО)	F2323
2.3.3.1	Доля образовательных учреждений ВПО, подключенных к интернету, в общем числе обследованных учреждений ВПО со скоростью 256 Кбит/сек и выше	F2331
2.3.4.2	Число персональных компьютеров, используемых в учебных целях, имеющих доступ к интернету, в расчете на 100 студентов (обучающихся) по образовательным учреждениям СПО	F2342

Окончание табл. 1 / End of Table 1

№	Наименование исходного фактора	Обозначение
2.3.4.3	Число персональных компьютеров, используемых в учебных целях, имеющих доступ к интернету, в расчете на 100 студентов (обучающихся) по образовательным учреждениям ВПО	F2343
2.3.5.2	Доля образовательных учреждений, имеющих веб-сайт в интернете, в общем числе самостоятельных образовательных учреждений СПО	F2352
Готовность учреждений здравоохранения к развитию на основе ИКТ		
2.4.1	Доля учреждений здравоохранения, использовавших персональные компьютеры, в общем числе обследованных учреждений здравоохранения	F241
2.4.2	Доля учреждений здравоохранения, имеющих локальные вычислительные сети, в общем числе обследованных учреждений здравоохранения	F242
2.4.3	Доля учреждений здравоохранения, использовавших интернет, в общем числе учреждений здравоохранения	F243
2.4.4	Число персональных компьютеров в расчете на 100 работников в учреждениях здравоохранения	F244
2.4.5	Число персональных компьютеров, имевших доступ к глобальным информационным сетям, на 100 работников в учреждениях здравоохранения	F245
2.4.6	Число персональных компьютеров, подключенных к интернету, на 100 работников учреждений здравоохранения	F246
Использование ИКТ в учреждениях здравоохранения		
2.4.7	Доля учреждений здравоохранения, имевших веб-сайт, в общем числе обследованных учреждений здравоохранения	F247
Готовность учреждений культуры к развитию на основе ИКТ		
2.5.1	Доля библиотек, имеющих персональные компьютеры, в общем числе библиотек	F251
2.5.3	Доля учреждений культуры, использовавших интернет, в общем числе обследованных учреждений культуры	F253
Информационные ресурсы и услуги учреждений культуры		
2.5.5	Доля учреждений культуры, имевших веб-сайт, в общем числе обследованных учреждений культуры	F255
2.5.6	Доля электронных документов на съемных носителях, в общем объеме библиотечного фонда	F256
2.5.7	Доля музейных предметов, внесенных в электронный каталог, в общем объеме общего музейного фонда	F257
2.5.8	Объем электронного каталога библиотеки, доступного в интернете	F258
2.5.9	Число документов библиотечного фонда, переведенных в электронную форму	F259
2.5.10	Число музейных предметов, внесенных в электронный каталог	F2510
2.5.11	Доля библиографических баз данных библиотеки, в общем объеме собственных баз данных библиотеки	F2511
2.5.12	Число доступных в интернете музейных предметов, внесенных в электронный каталог и имеющих цифровые изображения, на 10 000 предметов общего музейного фонда	F2512
2.5.13	Доля музейных предметов, внесенных в электронный каталог и имеющих цифровые изображения, в общем объеме музейного фонда	F2513
2.5.14	Доля библиотек, предоставляющих доступ через интернет к полнотекстовым электронным ресурсам библиотеки, в общем числе библиотек	F2514
2.5.15	Доля библиотек, имеющих персональные компьютеры для пользователей, в общем числе библиотек	F2515
2.5.16	Доля компьютеризированных посадочных мест с возможностью доступа к электронным ресурсам библиотеки в общем числе посадочных мест	F2516
2.5.17	Доля посадочных мест с возможностью выхода в интернет в общем числе посадочных мест библиотеки	F2517

Таблица 2 / Table 2

Порядковые номера субъектов РФ / Numbers of subjects of the Russian Federation

№	Субъект РФ	№	Субъект РФ
1	Белгородская область	43	Ставропольский край
2	Брянская область	44	Республика Башкортостан
3	Владимирская область	45	Республика Марий Эл
4	Воронежская область	46	Республика Мордовия
5	Ивановская область	47	Республика Татарстан
6	Калужская область	48	Удмуртская Республика
7	Костромская область	49	Чувашская Республика
8	Курская область	50	Пермский край
9	Липецкая область	51	Кировская область
10	Московская область	52	Нижегородская область
11	Орловская область	53	Оренбургская область
12	Рязанская область	54	Пензенская область
13	Смоленская область	55	Самарская область
14	Тамбовская область	56	Саратовская область
15	Тверская область	57	Ульяновская область
16	Тульская область	58	Курганская область
17	Ярославская область	59	Свердловская область
18	г. Москва	60	Ханты-Мансийский автономный округ – Югра
19	Республика Карелия	61	Ямало-Ненецкий автономный округ
20	Республика Коми	62	Тюменская область без АО
21	Архангельская область	63	Челябинская область
22	Вологодская область	64	Республика Алтай
23	Калининградская область	65	Республика Бурятия
24	Ленинградская область	66	Республика Тыва
25	Мурманская область	67	Республика Хакасия
26	Новгородская область	68	Алтайский край
27	Псковская область	69	Забайкальский край
28	г. Санкт-Петербург	70	Красноярский край
29	Республика Адыгея	71	Иркутская область
30	Республика Калмыкия	72	Кемеровская область
31	Республика Крым	73	Новосибирская область
32	Краснодарский край	74	Омская область
33	Астраханская область	75	Томская область
34	Волгоградская область	76	Республика Саха (Якутия)
35	Ростовская область	77	Камчатский край
36	г. Севастополь	78	Приморский край
37	Республика Дагестан	79	Хабаровский край
38	Республика Ингушетия	80	Амурская область
39	Кабардино-Балкарская Республика	81	Магаданская область
40	Карачаево-Черкесская Республика	82	Сахалинская область
41	Республика Северная Осетия – Алания	83	Еврейская автономная область
42	Чеченская Республика	84	Чукотский автономный округ

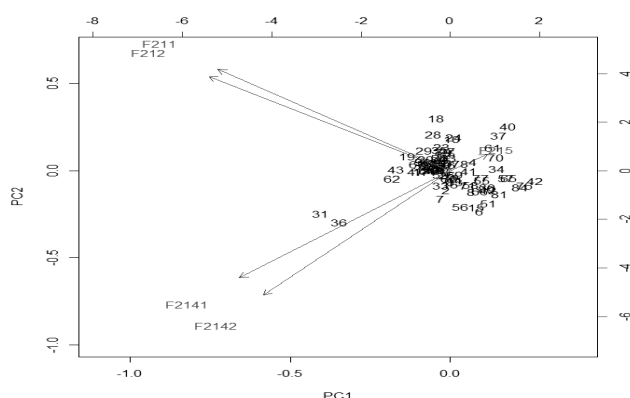


Рис. 1 / Fig 1. График Biplot по набору факторов подгруппы 2.1 / Biplot graph on a set of factors of subgroup 2.1

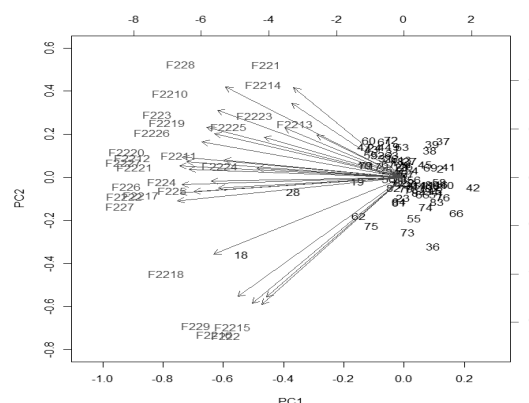


Рис. 2 / Fig 2. График Biplot по набору факторов подгруппы 2.2 / Biplot graph on a set of factors of subgroup 2.2

Продолжим редуцированием факторов подгрупп 2.2 «Электронный бизнес». Выбираем первые пять главных компонент, в совокупности объясняющие 83,5% общей дисперсии.

PC1 относительно сильно отрицательно коррелирует с F223, F226, F227, F2212, F2217–F2222, и поэтому ее можно охарактеризовать как компоненту, по совокупности определяющую интеграцию внутренних информационных систем (ИС) и совместный доступ к информации внутри организации. PC2 коррелирует с F222, F229, F2215, F2216, следовательно она интегрально определяет в основном применение ИКТ работниками организаций. PC3 сложно однозначно идентифицировать. PC4 (четвертая главная компонента) коррелирует с F2224 и F2225. Следовательно, она определяет интеграцию ИС организации с ИС контрагентов. И, наконец, PC5 сложно однозначно идентифицировать. На полученном графике biplot (рис. 2) представлена диаграмма рассеяния наблюдаемых объектов — регионов РФ.

Приступим к редуцированию факторов подгрупп 2.3 «Электронное образование». Выбираем первые три главные компоненты, в совокупности объясняющие 81,57% общей дисперсии.

PC1 относительно сильно положительно коррелирует с F231, F2323, F2342, F2343 и поэтому ее можно охарактеризовать как компоненту, определяющую число персональных компьютеров, используемых в учебных целях, имеющих доступ к интернету и находящихся в составе локальных вычислительных сетей (ЛВС), в расчете на 100 студентов (обучающихся) СПО и ВПО. PC2 коррелирует с F231 и F2352. Следовательно, она определяет в основном долю образовательных учреждений, имеющих веб-сайт

в интернете, в общем числе самостоятельных образовательных учреждений СПО. PC3 выделяется отрицательной корреляцией с F2331 и характеризует долю образовательных учреждений ВПО, подключенных к интернету, в общем числе обследованных учреждений ВПО со скоростью 256 Кбит/сек и выше.

Визуализация вклада факторов в главные компоненты представлена на графике biplot соответствующими факторными нагрузками (рис. 3).

Приступим к редуцированию факторов подгрупп 2.4 «Электронное здравоохранение». Выбираем первые две главные компоненты, в совокупности объясняющие 81,09% общей дисперсии.

PC1 относительно сильно отрицательно коррелирует с F241, F245–F247 и поэтому ее можно охарактеризовать как компоненту, определяющую использование ИКТ в учреждениях здравоохранения. PC2 отрицательно коррелирует с F243 и F243, следовательно она определяет в основном готовность учреждений здравоохранения к развитию на основе ИКТ.

Визуализация вклада факторов в главные компоненты представлена на графике biplot соответствующими факторными нагрузками (рис. 4).

Проведем редуцирование факторов подгрупп 2.5 «Электронная культура». Выбираем первые шесть главных компонент, в совокупности объясняющие 80,4% общей дисперсии.

PC1 относительно сильно отрицательно коррелирует с F2514–F2517 и поэтому ее можно охарактеризовать как компоненту, определяющую электронные ресурсы библиотек. PC2 отрицательно коррелирует с F258–F2510 и положительно — с F2512 и F2513. Следовательно, она определяет в основном доступность электронных каталогов, музеев и библиотек. PC3

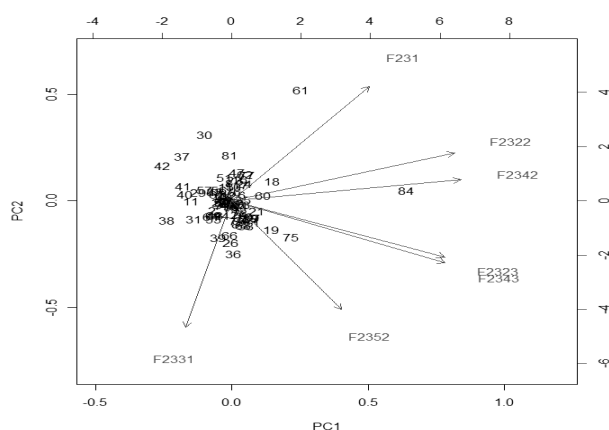


Рис. 3 / Fig. 3. График Biplot по набору факторов подгруппы 2.3 / Biplot graph on a set of factors of subgroup 2.3

относительно сильно коррелирует с F253 и F2511 и отражает долю учреждений культуры, использовавших интернет, в общем числе обследованных учреждений культуры. PC4 характеризует долю музейных предметов, внесенных в электронный каталог, в общем объеме общего музейного фонда. PC5 — долю учреждений культуры, имевших веб-сайт, в общем числе обследованных учреждений культуры. PC6 — долю электронных документов на съемных носителях в общем объеме библиотечного фонда. Визуализация вклада факторов в главные компоненты представлена на графике biplot соответствующими факторными нагрузками (рис. 5).

ЗАКЛЮЧЕНИЕ

Таким образом, в результате применения метода РСА произведено редуцирование 61 фактора использования ИКТ для развития информационного общества в Российской Федерации до 16 (табл. 3).

В рамках данного комплексного анализа многомерных данных были использованы следующие подходы:

- сравнительный анализ многомерных данных;
- многофакторный анализ данных (исследования результатов с целью идентификации главных компонент, интегрирующих характеристики групп исходных факторов).

С помощью метода главных компонент удалось построить сравнительно небольшой набор главных компонент, характеризующих уровень развития цифровизации регионов РФ. Полученные характеристики главных компонент позволяют более эффективно проводить многофакторный анализ региональных данных, фокусируясь лишь на выделенных главных

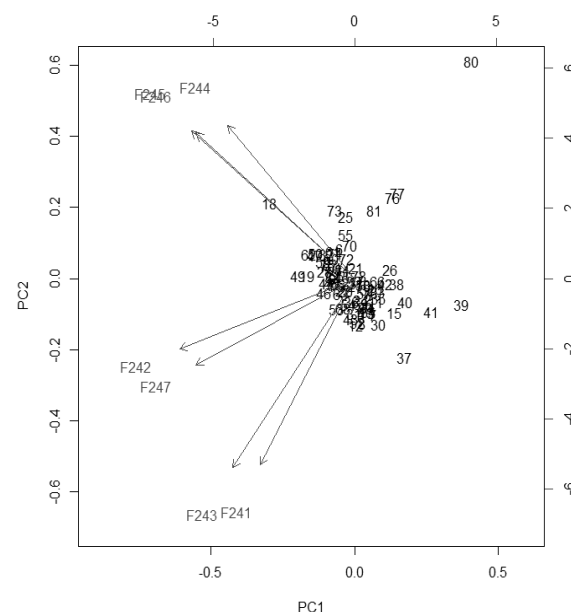


Рис. 4 / Fig. 4. График Biplot по набору факторов подгруппы 2.4 / Biplot graph on a set of factors of subgroup 2.4

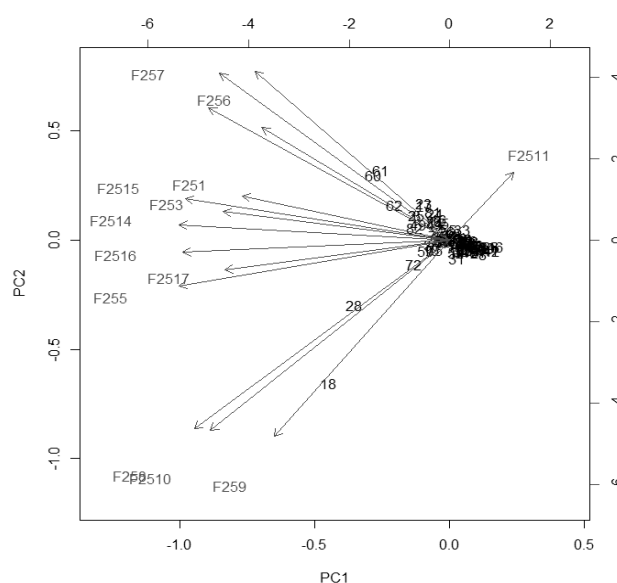


Рис. 5 / Fig. 5. График Biplot по набору факторов подгруппы 2.5 / Biplot graph on a set of factors of subgroup 2.5

компонентах. Визуальное представление субъектов РФ на диаграммах рассеяния существенно облегчает сравнительный анализ текущего состояния цифровизации в субъектах РФ по отдельным составляющим.

На основе анализа данных доказательно обосновано наличие устойчивой совокупности субъектов РФ,

Таблица 3 / Table 3

Результаты редуцирования числа факторов использования ИКТ для развития информационного общества в Российской Федерации / The results of reducing the number of factors to the use of ICT for the development of the information society in the Russian Federation

№	Подгруппа факторов	Главные компоненты	Характеристика
1	2.1. «Электронное правительство»	PC1	Готовность ОГВ и ОМС к использованию ИКТ для осуществления управленческих функций и предоставления государственных услуг
2		PC2	Готовность размещать заказы на поставки товаров, выполнение работ и оказание услуг для государственных и муниципальных нужд с использованием электронных торговых площадок
3		PC3	Готовность к ведению электронного документооборота в органах государственной власти
4	2.2. «Электронный бизнес»	PC1	Интеграция внутренних информационных систем (ИС) и совместный доступ к информации внутри организации
5		PC2	Применение ИКТ работниками организаций
6		PC4	Интеграция ИС организации с ИС контрагентов
7	2.3. «Электронное образование»	PC1	Оснащенность персональными компьютерами, используемыми в учебных целях, имеющими доступ к интернету
8		PC2	Наличие у образовательных учреждений СПО веб-сайта в интернете
9		PC3	Оснащенность образовательных учреждений ВПО интернетом со скоростью 256 Кбит/сек и выше
№	Подгруппа факторов	Главные компоненты	Характеристика
10	2.4. «Электронное здравоохранение»	PC1	Использование ИКТ в учреждениях здравоохранения
11		PC2	Готовность учреждений здравоохранения к развитию на основе ИКТ
12	2.5. «Электронная культура»	PC1	Электронные ресурсы библиотек
13		PC2	Электронные каталоги музеев и библиотек
		PC3	Оснащенность учреждений культуры интернетом
14		PC4	Полнота музейных электронных каталогов
15		PC5	Наличие у учреждений культуры веб-сайта
16		PC6	Ведение электронного документооборота в библиотечном фонде

отличающихся от большинства субъектов РФ особыми характеристиками использования ИКТ для развития.

К ним относятся: Московская область, г. Москва, г. Санкт-Петербург, Республика Крым, Чеченская Республика, Республика Татарстан, Ханты-Мансийский автономный округ — Югра, Ямало-Ненецкий автономный округ, Республика Тыва, Магаданская область, Чукотский автономный округ. Это вполне закономерно, так как именно эти субъекты РФ отличаются экстремумами значений факторов использования ИКТ для развития. Полученные результаты согласуются с данными рейтинга Совета по региональной информатизации Правительственной комиссии по использованию информационных технологий для улучшения качества жизни и условий ведения

предпринимательской деятельности. Так, из выявленного перечня обособленных объектов в первую пятерку данного рейтинга попали г. Москва, Республика Татарстан, Ханты-Мансийский автономный округ — Югра. В числе замыкающих рейтинг отметим выявленные в ходе исследования Республику Крым, Чеченскую Республику, Республику Тыва, Магаданскую область и Чукотский автономный округ.

Результаты проведенного факторного анализа обосновывают эффективность применения метода РСА для структурирования многомерных данных по использованию ИКТ для развития информационного общества в РФ. Выявленные факторные нагрузки не являются случайными, а имеют детерминированный экономический смысл.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Dossani R., Kenney M. Lift and Shift; Moving the back office to India. *Information Technologies and International Development*. Massachusetts Institute of Technology, Cambridge, MA. 2003;1(2):21–37.
2. Womack J., Jones D., Roos D. The machine that changed the world. New York: Rawson Associates; 1990.
3. Shlens J.A. Tutorial on principal components analysis. URL: <https://datajobs.com/data-science-repo/PCA-Tutorial-Shlens.pdf> (дата обращения: 25.03.2019).
4. Cornillon P.A., Imam W., Matzner-Løber E. Forecasting time series using principal component analysis with respect to instrumental variables. *Computational Statistics & Data Analysis*. 2008;3(52):1269–1280.
5. Tabachnick B.G., Fidell L.S. Using multivariate statistics. Fourth Edition. Allyn and Bacon, Boston; 2001.
6. Díaz A. Jareño, F., Navarro E. Term structure of volatilities and yield curve estimation methodology. *Quantitative Finance*. 2010. DOI: 10.1080/14697680903473286
7. Shappell Scott A., Wiegmann Douglas A. The human factors analysis and classification system—HFACS. 2000. Office of Aviation Medicine, Washington, DC 20591 DOT/FAA/AM-00/7.
8. Lekkos I. A Critique of factor analysis of interest rates. *Journal of Derivatives*. 2000;1(8):72–83.
9. Litterman R., Scheinkman J. Common factors affecting bond returns. *Journal of Fixed Income*. 1991;1(1):54–61.
10. Халафян А.А. Современные статистические методы медицинских исследований. Изд. 3-е. М.: ЛЕНАНД; 2014. 320 с. Khalafyan A.A. Modern statistical methods of medical research. 3 ed. Moscow: LENAND; 2014. 320 p.
11. Novales A., Benito S. A factor analysis of volatility across the term structure: the Spanish case. *Revista de Economía Financiera*. 2007;13:8–27.
12. Everitt B.S., Landau S., Leese M., Cluster analysis. 4th ed. London: Arnold; 2001.
13. Шипунов А.Б., Балдин В.М., Волкова П.А., Коробейников А.И., Назарова С.А., Петров С.В., Суфиянов В.Г. Наглядная статистика. Используем R! М.: ДМК Пресс; 2017. Shipunov A.B., Baldin V.M., Volkova P.A., Korobeynikov A.I., Nazarova S.A., Petrov S.V., Sufiyanov V.G. Visual statistics. Use R! Moscow: DMK Press; 2017.
14. Priestley M., Subba Rao T., Tong H. Applications of principal component analysis and factor analysis in the identification of multivariate systems. IEEE: *Transactions in Automatic Control*. 1974; AC-19:730–734.
15. Scholz M., Fraunholz M., Selbig J. Nonlinear principal component analysis: Neural network models and applications. In: Gorban A.N. et al., eds. LNCSE 58, Springer; 2007. URL: <https://pdfs.semanticscholar.org/9d31/23542031a227d2f4c4602066cf8ebceab7a.pdf>.
16. Урнышев Р. О мониторинге развития информационного общества в субъектах Российской Федерации. URL: http://tomedu.ru/wp-content/uploads/2015/02/Vopros_8_Rejting.pdf. Urnyshev R. On monitoring the development of information society in the subjects of the Russian Federation. URL: http://tomedu.ru/wp-content/uploads/2015/02/Vopros_8_Rejting.pdf.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Нашль Маратович Габдуллин — кандидат экономических наук, доцент кафедры управления корпоративными финансами, Казанский федеральный университет, Казань, Россия
nail56@yandex.ru

Игорь Александрович Киришин — доктор экономических наук, профессор Высшей школы бизнеса, Казанский федеральный университет, Казань, Россия
kia1125@mail.ru

ABOUT THE AUTHORS

Nail M. Gabdullin — Candidate of Economic Sciences, Associate professor, Department of Corporate Finance Management, Kazan Federal University, Kazan, Russia
nail56@yandex.ru

Igor A. Kirshin — Doctor of Economics, Professor, Higher School of Business, Kazan Federal University, Kazan, Russia
kia1125@mail.ru

Статья поступила 02.06.2019; принята к публикации 05.07.2019.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article received on 02.06.2019; accepted for publication on 05.07.2019.

The authors read and approved the final version of the manuscript.