

# Моделирование данных для анализа соответствия городов Поволжского региона формату цифрового государства

И.Н. Хаймович<sup>1,2</sup>, В.М. Рамзаев<sup>1</sup>, В.Г. Чумак<sup>1</sup>

<sup>1</sup>Самарский университет государственного управления «Международный институт рынка», ул. Г.С. Аксакова, 21, Самара, Россия, 443030

<sup>2</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

**Аннотация.** В статье предложена методика оценки готовности муниципальных образований Поволжского региона к внедрению технологий цифрового государства. Предложена модель статистических испытаний, основанная на многократном теоретико-вероятностном и статистическом моделировании параметрических величин элементов технологий цифровой экономики. Данная модель позволит определить возможность участия городов Поволжского региона в государственной программе цифровой экономики, подобрать города наиболее подходящие к внедрению современных технологий, выявить основные недостатки по возможности их интеграции в программу. Данное исследование может быть интересно специалистам в области цифровой экономики, управления данными Big Data.

## 1. Введение

В настоящее время в России осуществляется переход к цифровому развитию государства и цифровой экономики. В рамках жесткой конкуренции муниципальных образований вопрос внедрения информационных технологий становится особенно актуальным. Муниципальные образования (МО) – городские округа являются ключевыми элементами территориальной организации экономики страны, где формируются новые деловые, финансовые и культурные центры, стимулирующие изменения. Реализация в МО современных форм и инструментов цифровой экономики в настоящее время сдерживается следующими проблемами: ограничением доступа в муниципальных образованиях к цифровым системам, депрессивным состоянием экономик МО, отсутствием социальной вовлеченности местного самоуправления в процесс управления, отсутствием основ построения цифровой экономики целого ряда других.

Вместе с тем дальнейшее развитие МО не возможно без привлечения новых интеллектуальных услуг, одной из которых выступают технологии цифрового государства. Возникают вопросы:

- готовы ли МО к внедрению технологий цифрового государства, т.е. к устранению посредников при услугах, осуществлению прямых транзакций в экономике и т.д.;
- на основе какой методики осуществляется проверка готовности.

При управлении развитием региональной экономики РФ в современном цифровом формате требуется решить следующие задачи:

- рассмотреть основные рейтинги оценки соответствия МО принципам цифрового государства;
- разработать методику применения через анализ данных соответствия МО принципам «цифрового государства»;
- провести анализ городов Поволжья по этой методике.

**2. Анализ существующих рейтингов в концепции «цифрового государства»**

В концепцию «цифрового государства» входят следующие понятия:

- «умная экономика», которая включает высокий уровень показателей в областях инновации, занятость, торговля, производительность, физическая инфраструктура;
- «умная среда», которая связывает показатели качество воздуха, водоснабжение, уровень шума, качество окружающей среды, биоразнообразие, энергетику;
- «умное общество и культура», которое включает образование, здравоохранение, безопасность, жилье, культуру, социальную вовлеченность.

Все эти показатели должны осуществлять взаимодействие только через информационно-коммуникационные технологии (ИКТ).

Для анализа готовности и привлечения инвестиций в МО разработаны рейтинги «Smart City» Организацией Объединенных Наций [1] и рейтинг России по оценке устойчивого развития городов [2]. Анализ данных рейтингов приведен в таблице 1.

**Таблица 1.** Рейтинги «цифрового государства» в России и за рубежом.

<b>Рейтинги</b>	<b>Индикаторы</b>
<b>Рейтинг устойчивого развития городов РФ</b>	Экономика, городское хозяйство, социальная сфера, экологическая обстановка
<b>Система показателей умных городов</b>	Экономика (физическая инфраструктура, инновации, занятость, торговля, производительность, информационно-коммуникационные технологии). Окружающая среда (качество воздуха, водоснабжение, уровень шума, качество окружающей среды, биоразнообразие, энергетику). Общество и культура (образование, здравоохранение, безопасность, жилье, культуру, социальную вовлеченность)

Рассмотрим подробнее рейтинг устойчивого развития городов России, предложенный агентством ООО «Агентство ЭС ДЖИ ЭМ». В данный рейтинг включены свыше тридцати показателей, характеризующих МО: экономика, городское хозяйство, социальная сфера, экология. Респондентами выборки выступают административные центры России. Достоинством данного рейтинга является сбалансированность рассматриваемых показателей, ведь разбалансировка отрицательно влияет на устойчивое развитие МО. К недостаткам можно отнести тот факт, что высокие показатели не всегда определяют лидерские позиции МО в стране. Каждый год в данном рейтинге лидируют города – Москва, Санкт-Петербург и Уфа. К аутсайдерам ежегодно относят административные центры Поволжского региона. К недостаткам в разрезе рассматриваемой темы можно отнести обособленный характер показателей данного рейтинга к теме «цифровое государство». Например, показатель «демография», состоящий из критериев «коэффициент естественного прироста», «коэффициент миграционного прироста», «демографическая нагрузка» лишь опосредованно влияет на использование ИКТ во всех сферах деятельности города.

Далее рассмотрим систему показателей умных городов, разработанную Европейской экономической комиссией Организации Объединенных Наций. Степень готовности к

внедрению новых ИКТ оценивается через показатели инновационности города для повышения уровня жизни населения. В данной системе идет оценка эффективности деятельности и услуг для удовлетворения будущих поколений в различных аспектах деятельности.

Данная система состоит из следующих трех блоков:

Блок 1. Экономика: инфраструктура ИКТ, инновации, занятость, торговля – электронная торговля, торговля – экспорт/импорт, производительность, физическая инфраструктура – водоснабжение, электроснабжение, здравоохранение, транспорт, здания.

Блок 2. Окружающая среда: качество воздуха, водоснабжение, шум, качество окружающей среды, биоразнообразие, энергетика.

Блок 3. Общество и культура; образование, здравоохранение, безопасность – ликвидация последствий бедствий, безопасность – чрезвычайные ситуации, безопасность – ИКТ, жилье, культура, социальная вовлеченность. В последний показатель входит расчет уровней общественного участия, гендерного равенства доходов, возможности людей с особыми потребностями, привлекательности для квалифицированных кадров, коэффициента Джинни.

Достоинством данной системы является подробный расчет всех показателей. Недостатком является применимость данной оценки только для европейских стран.

После проведенного подробного анализа европейских и российских рейтингов и систем оценки была разработана методика оценки городов к внедрению технологий «цифрового государства» с учетом всех особенностей Поволжского региона.

### **3. Модель вовлеченности МО Поволжского региона в технологии «цифровой экономики»**

Несмотря на большое количество исследований в данной области [3-6,8] отсутствует единая методика оценки вовлеченности МО в технологии «цифрового государства». Международные рейтинги и системы только апробируются и отвечают на конкретные вопросы: устойчива ли экономика города, есть ли в ней элементы «умного» города. В них нет оценки показателей интеллектуализации городской среды.

Далее будет рассмотрена модель статистических испытаний, основанная на многократном теоретико-вероятностном и статистическом моделировании параметрических величин концепции «цифрового государства». Данная модель будет связана с анализом большого количества данных, что потребует автоматизации показателей оценки, может быть на основе технологии Big Data [9,10].

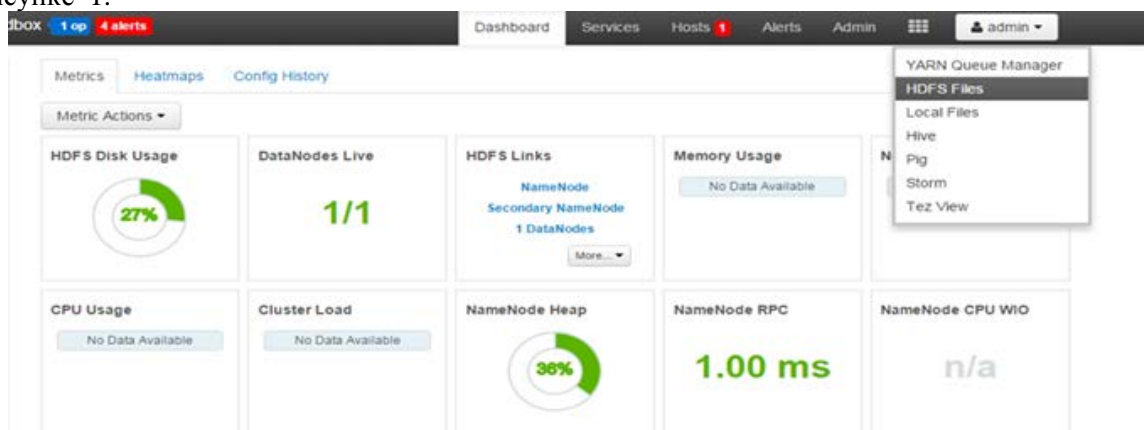
Метод использования интеллектуального анализа данных для решения поставленной задачи состоит в следующем:

1. формирование набора больших данных в hadoop из twitter по фильтру «Самарская область», выявляющему количество обращений;
2. разделение сформированного набора по различным фильтрам, связанным с показателями оценки вовлеченности МО в технологии «цифрового государства»;
3. проведение мониторинга потокового анализа неструктурированной информации по фильтрам;
4. разработка программы на языке Scala для работы с фильтрацией в области Больших Данных;
5. отладка и тестирование программы с набором практических данных;
6. анализ результатов вычисления.

Для получения данных используется социальная сеть «twitter», так как это «открытый» продукт, его применение не требует дополнительных инвестиций, а 50% пользователей Интернет имеют профили в данной программе. Twitter является второй по популярности сетью среди пользователей во всем мире, уступая лишь Facebook. Однако в отличие от Facebook, который не предоставляет открытый доступ к своим данным, Twitter такой доступ предоставляет, отсутствуют ограничения на доступ к наборам данных сервера. Пользователи данной социальной сети обмениваются в основном текстовой информацией, что является несомненным плюсом при обработке. Twitter не является предметной сетью и наиболее широко отражает общественное мнение по многим интересующим вопросам, поэтому для формирования зон малого бизнеса в регионе обработка данных из этой социальной сети была оптимальной.

Для работы с BIG DATA в социальных сетях используют методы сбора, обработки и анализа данных. Сбор данных осуществляется в режиме реального времени, в пределах определенной геолокации, либо в пределах всей сети, по определенным шаблонам. Информация, представляющая интерес для анализа это: локация, дата и время, контент, «автор» контента (пользователь), связи между пользователями. Сбор данных в социальных сетях можно осуществлять с помощью следующих инструментов: Apache Hadoop, BigInsights (IBM), Cloudera, Hortonworks, Storm. Для выполнения исследования в области вовлеченности МО в «цифровую экономику» был выбран Hortonworks. Для работы применялся Twitter Application (apps.twitter.com), в котором определялись и уточнялись ключевые параметры: API key, API secret, Access token, Access token secret.

Для сбора данных с использованием Hortonworks, Twitter App использовался конфигурационный файл сервиса flume в виртуальной машине Hortonworks Sandbox. После установки виртуальной машины Hortonworks\_Sandbox версии 2.3 и настройки сервиса flume система готова к загрузке данных из twitter. Для просмотра и загрузки скаченных файлов переходим в папку HDFS, где осуществляем обработку данных. Вид файловой структуры HDFS в виртуальной машине Hortonworks при решении задачи в сфере вовлеченности МО показан на рисунке 1.



**Рисунок 1.** Визуализация HDFS в Hortonworks при загрузке файлов для решения задачи в области вовлеченности МО в «цифровую экономику».

Собранные данные необходимо структурировать (т.е. обработать) в соответствии с парадигмой MapReduce. MapReduce — это фреймворк для выполнения распределенных задач с использованием большого количества компьютеров, образующих кластер.

Использование MapReduce позволило структурировать поток данных из социальных сетей по критериям: шрифты, размер текста, цвет, ссылка на профиль пользователя, локация, время и прочее.

Для определения данных для анализа показателей МО в нашем исследовании нужны данные следующих типов: размещение, текст, язык и время. Для того, чтобы извлечь лишь эту информацию можно использовать технологию MapReduce, встроенную в инструменте Hortonworks Sandbox. Для обработки данных используем СУБД Hive в среде Hadoop, позволяющую осуществлять операции над данными и их анализ путем SQL – подобных запросов. Для этого создаем файл обработки и создания нужных таблиц hiveddl.sql.

Запускаем данный файл командой: Hive\_f hiveddl.sql. Структурированные данные будут размещены в таблице 2.

**Таблица 2.** Вид заголовков для анализа структурированных данных в задачах для оценки показателей МО.

A	B	C	D	E	F
Data/Time	Time/Zona	language	Text	location	Sentiments

Эта таблица была получена из данных социальной сети «twitter», показанных на рисунке 2.



Рисунок 2. Собранные данные из социальной сети «twitter».

Благодаря технологии BIG DATA можно хранить и обновлять данные в файловой системе «hadoop» по фильтру «Самарская область» (filter1= {Самарская область}). Затем необходимо данную область отфильтровать по базовым параметрам оценки МО, установив, например, следующие фильтры: Filter2 (экономика) = { дороги, товары}; Filter3 (окружающая среда)= {лес, воздух}; Filter4 (общество и культура)= {ночной клуб, концерт, сейшн, тусовка}.

Можно получить графики зависимости количества обращений пользователей по фильтрам (т.е. величину показателя интернетизации МО) от времени сбора данных.

Время сбора данных из Интернета в технологии BIG DATA неограниченно.

В итоге получаем динамическое изменение информации в режиме реального времени из системы Интернет, что позволяет с минимальными инвестициями проводить мониторинг потокового анализа неструктурированной информации (технология In-Memory Data Processing and Stream) по фильтрам. Для реализации данного метода была написана программа на языке Scala:

```
val file = spark.textFile("hdfs://... ")
val errors=file.filter(line=>line.contains("Самарская область"))
//count all the data
errors.count()
//count data mentioning Filter
errors.filter(line=>line.contains("концерт")).count()
//Fetch the filter as an array of string
errors.filter(line=>line.contains("прием врача")).collect()
```

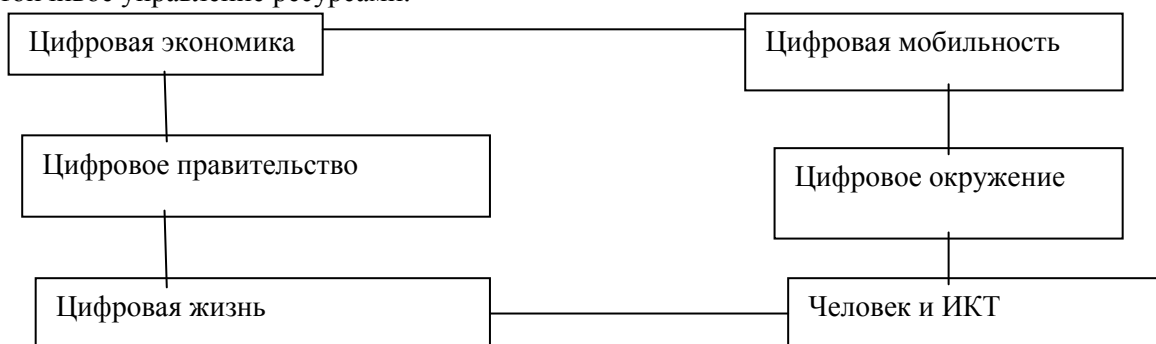
После работы программы получаем динамическое изменение параметров в среде BIG DATA, которые позволяют определять некоторые из показателей вовлеченности МО в «цифровую экономику» с учетом неструктурированной информации. Данную методу сбора информации для оценки параметров можно использовать и для других социальных систем и сайтов, также для сбора информации можно использовать и статистические данные официальных источников, выложенные в Интернет.

Таким образом, предложен инструмент для сбора данных в системе индикаторов МО в «цифровую экономику».

Система индикаторов вовлеченности МО к внедрению технологий «цифрового государства» представлена на рисунке 3.

К показателю в указанной выше системе индикаторов можно отнести следующие:

- цифровая экономика: показатель инновационности, предпринимательство, конкурентоспособность города, показатель технологичности производства, рынок труда, показатель финансовой независимости;
- цифровая мобильность: местная транспортная система, (интер-) национальная доступность, ИКТ-инфраструктура, устойчивость транспортной системы;
- человек и ИКТ: показатель интеллектуальности, обучение на протяжении всей жизни, этническое множество;
- цифровая жизнь: культурные и развлекательные заведения, состояние здоровья, индивидуальная безопасность, качество жилья, учебные заведения, туристическая привлекательность, социальная сплоченность;
- цифровое правительство: политическая осведомленность, общественные и социальные услуги, эффективное и прозрачное администрирование;
- цифровое окружение: качество воздуха (без загрязнения); экологическая осведомленность, устойчивое управление ресурсами.



**Рисунок 3.** Система индикаторов МО в технологиях «цифрового государства».

Данные показатели имеют экспертные значения. Для сравнения различных показателей необходимо стандартизировать значения по выборкам нескольких городов. В исследовании используется метод стандартизации z-преобразование по следующей формуле:

$$z_i = \frac{x_i - \bar{x}}{S}$$

где  $\bar{x}$  - среднее значение в выборке,  $S$  – стандартное отклонение по выборке. Этот метод преобразует все значения индикаторов в стандартизованные значения со средним 0 и стандартным отклонением 1. Метод имеет преимущества для рассмотрения неоднородности внутри групп и поддержания его метрической информации. Кроме того, достигается высокая чувствительность к изменениям.

Чтобы получать результаты по уровню показателей, индикаторам и окончательному результату для каждого города, необходимо суммировать значения на уровне индикатора. Для агрегирования соответствующей группы показателей по доменам также учитываем коэффициент охвата каждого показателя. Определенный результат от индикатора, охватывающего все города, весит больше, чем от показателя, охватывающего только, например, 6 городов. Помимо этой небольшой коррекции результаты были агрегированы на всех уровнях без какого-либо взвешивания. Агрегация была добавлена, но разделена на количество добавленных значений. Это позволяет включать также города, которые не охватывают все показатели. Их результаты рассчитываются по имеющимся значениям. Тем не менее, необходимо обеспечить хороший охват всех городов, чтобы получить разумные результаты.

Некоторые показатели могут быть не только экспертные, но и расчетные, к ним можно отнести показатель технологичности производства, показатель инновационности, показатель интернетизации, показатель интеллектуализации, показатель финансовой независимости, показатель энергоэффективности, показатель внедрения креативных технологий. Данные группировки могут позволить получить быстрые управленческие решения в зависимости от

средних значений индикаторов. Если среднее значение находится в границах 3,7 и выше, то МО готово к внедрению технологий цифровой экономики. Если среднее значение находится в границах от 2,5 до 3,7, то МО имеет среднюю готовность. Если среднее значение находится в границах от 1,95 до 2,5, то МО имеет удовлетворительную готовность. Если среднее значение находится ниже значения 1,95, то МО не готово к внедрению технологий цифровой экономики [7].

#### 4. Результаты и обсуждение

По разработанной выше методике были проведены расчеты показателей и индикаторов готовности городов Поволжского региона к внедрению технологий «цифрового государства».

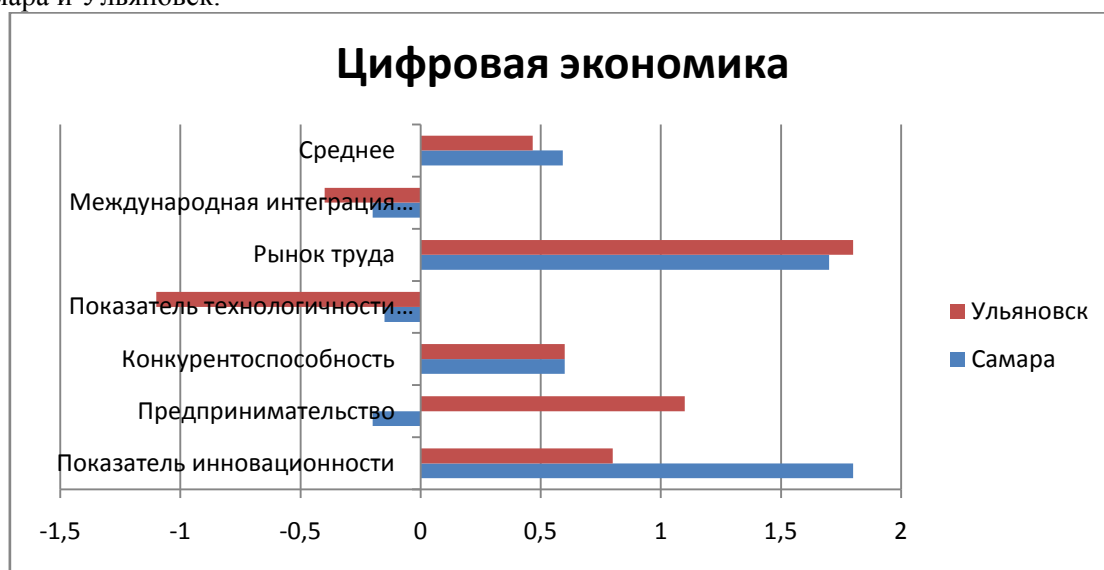
В таблице 3 рассчитываются относительные и абсолютные показатели МО Поволжского региона с учетом  $z$  – преобразования.

**Таблица 3.** Показатели муниципальных образований Поволжского региона с учетом  $z$  – преобразования.

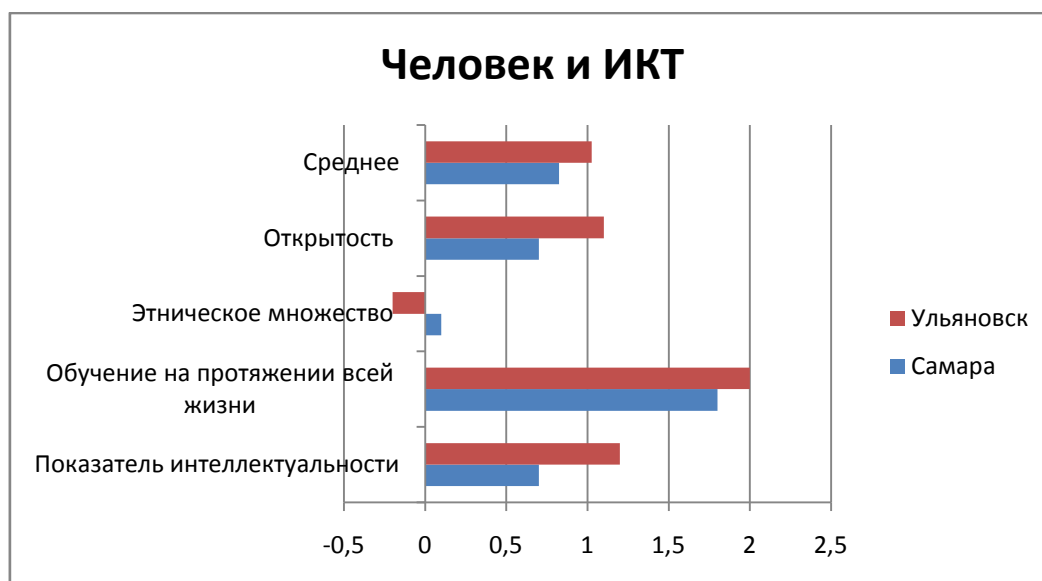
Цифровая экономика	МО (абс)	откл	МО (отн)
Показатель инновационности	3	0,547722558	0,912871
Предпринимательство	3		0,912871
Конкурентоспособность	2		-0,91287
Показатель технологичности производства	3		0,912871
Рынок труда	2		-0,91287
Межд.интеграция(показательфиннезависимости)	2		-0,91287
<b>Всего</b>	<b>15</b>		
<b>Человек и ИКТ</b>			
Показатель интеллектуальности	1	1,707825128	-1,0247
Обучение на протяжении всей жизни	2		-0,43916
Этническое множество	3		0,146385
Открытость	5		1,317465
<b>Всего</b>	<b>11</b>		
<b>Цифровая мобильность</b>			
Местная транспортная система	2	2,217355783	-0,56373
(Интер-) национальная доступность	1		-1,01472
ИКТ-инфраструктура	4		0,338241
Устойчивость транспортной системы	6		1,240216
<b>Всего</b>	<b>13</b>		
<b>Цифровая жизнь</b>			
Культ. и развлекательные заведения	6	0,975900073	1,610235
Состояние здоровья	5		0,58554
Индивидуальная безопасность	3		-1,46385
Показатель энергоэффективности	4		-0,43916
Учебные заведения	4		-0,43916
Туристическая привлекательность	5		0,58554
Социальная сплоченность	4		-0,43916
<b>Всего</b>	<b>31</b>		
<b>Цифровое правительство</b>			

Политическая осведомленность	3	0,577350269	-0,57735
Общественные и социальные услуги	3		-0,57735
Эффективное и прозрачное администрирование	4		1,154701
<b>Всего</b>	<b>10</b>		
<b>Цифровое окружение</b>			
Качество воздуха (без загрязнения)	4	1,154700538	0,57735
Экологическая осведомленность	4		
Устойчивое управление ресурсами	2		
<b>Всего</b>	<b>10</b>		

Далее на рисунках 4-9 показаны гистограммы индикаторов цифрового города для городов Самара и Ульяновск.



**Рисунок 4.** Показатели «цифровой экономики» для городов Самара и Ульяновск.



**Рисунок 5.** Показатели для индикатора «Человек и ИКТ» для городов Самара и Ульяновск.



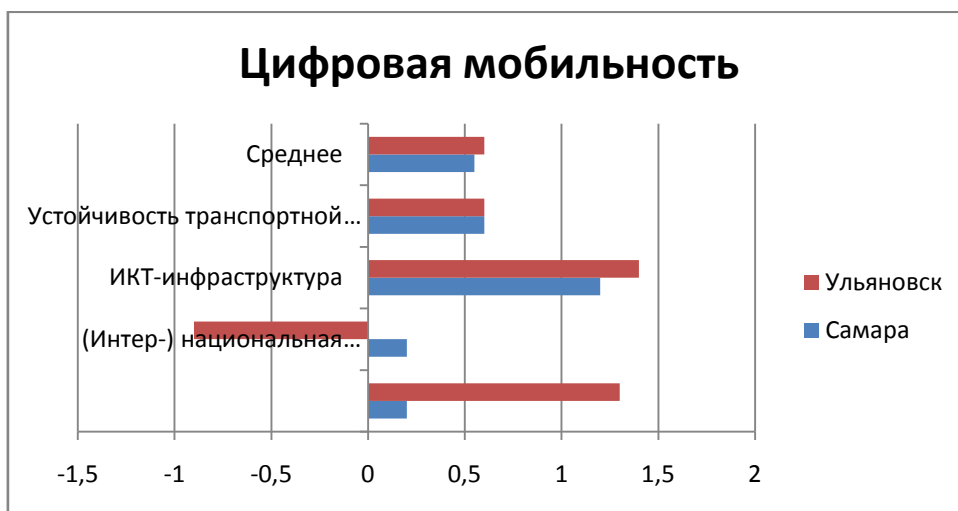


Рисунок 6. Показатели «цифровой мобильности» для городов Самара и Ульяновск.

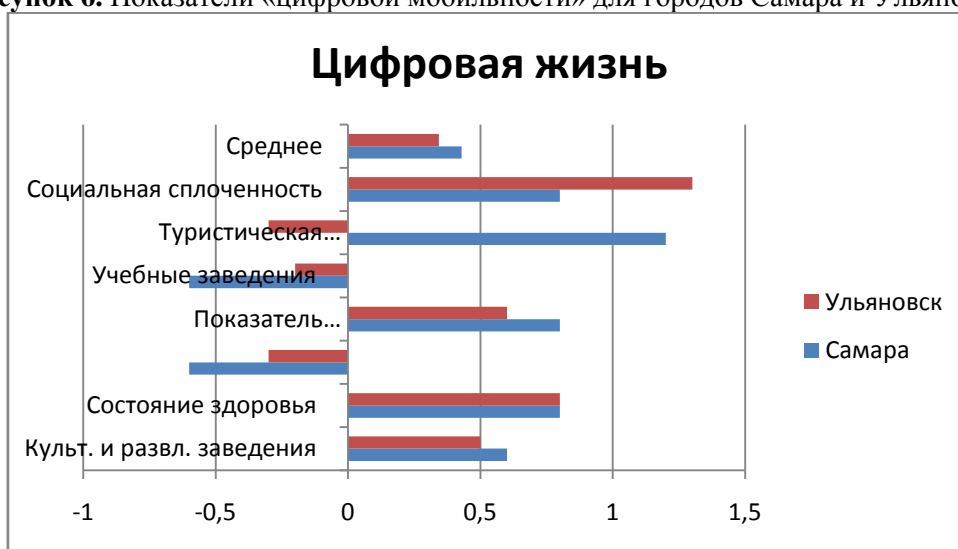


Рисунок 7. Показатели для индикатора «цифровая жизнь» для городов Самара и Ульяновск.

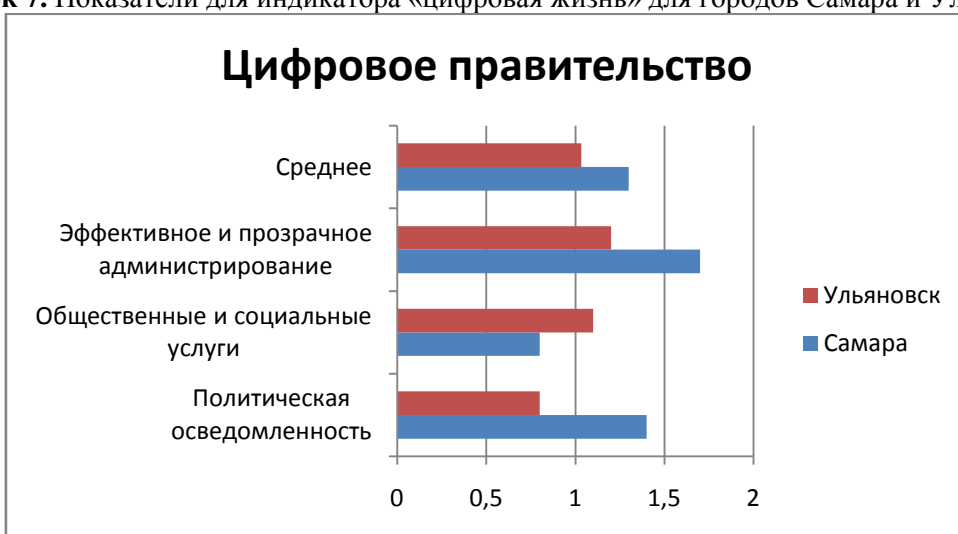


Рисунок 8. Показатели для индикатора «цифровое правительство» для городов Самара и Ульяновск.

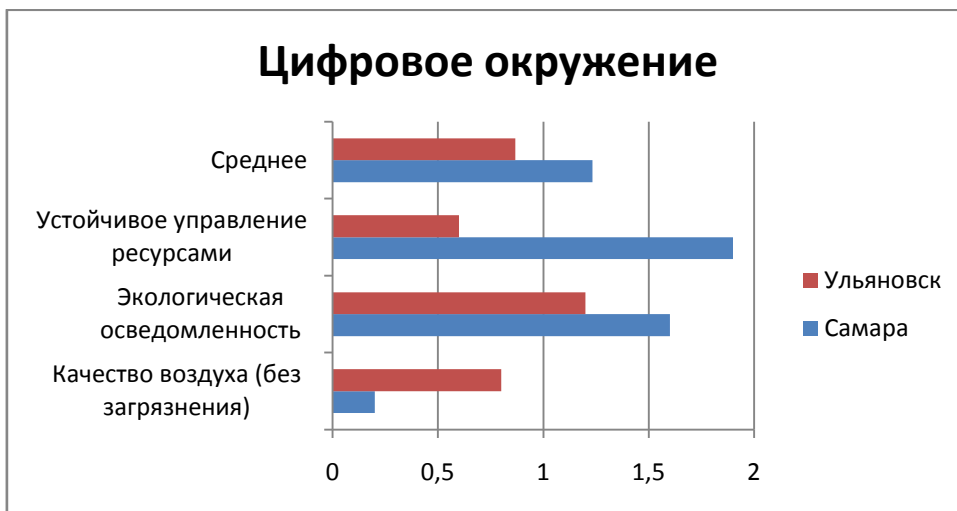


Рисунок 9. Показатели «цифрового окружения» для городов Самара и Ульяновск

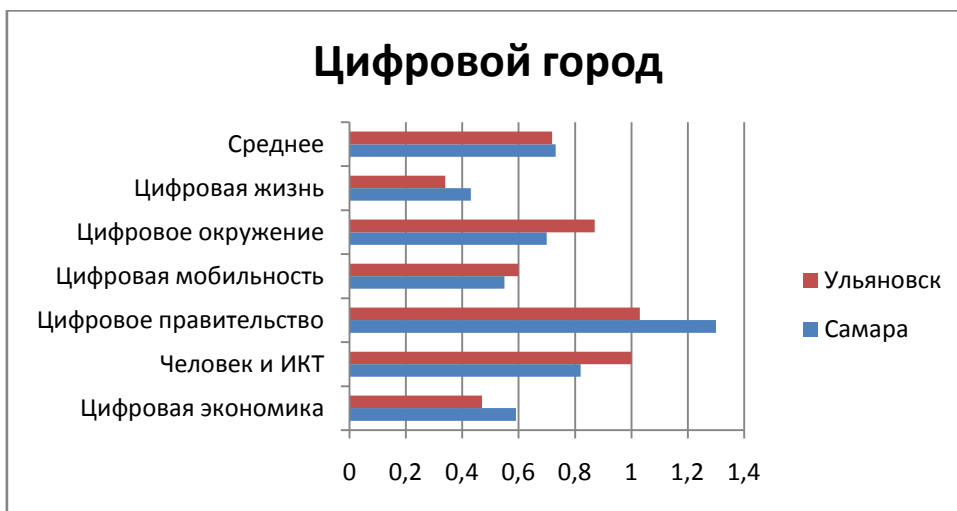


Рисунок 10. Индикаторы «цифрового города» для городов Самара и Ульяновск



Рисунок 11. График уровней индикаторов «цифрового города» для городов Поволжского региона.

Проведем сравнительный анализ по индикаторам «цифрового города» для муниципальных образований с более низким уровнем готовности, т.е. находящихся ниже нулевого уровня, т.е. требующих значительных инвестиций во внедрение ИКТ. Сравнительный график результатов исследования приведен на рисунке 11.

В итоге можно определить основные тренды для проведения инвестиционных вложений в МО Поволжского региона на основе сравнительного анализа индикаторов «цифрового города». Из рисунка 9 видно, что вложение в область с показателями ниже нулевого уровня не рентабельно. В этой зоне находятся многие города Поволжья, например, Жигулевск. Лучше осуществлять вложения в МО с уровнем готовности выше нулевого уровня, например, Ульяновск. Эти города больше подходят под концепцию «цифрового города», они практически готовы к внедрению технологий «цифрового государства».

Таким образом, модель оценки позволит определить уровень развития муниципальных образований, готовых к внедрению цифрового государства, выявить недостатки, входящих в группу «неготовых к внедрению», позволит улучшить показатели МО на основании подробного анализа данных по всем крупным городам Поволжского региона.

## 5. Литература

- [1] Показатели «умных» устойчивых городов, разработанные ЕЭК ООН-МСЭ. [Электронный ресурс] – Режим доступа: [http://www.unece.org/fileadmin/DAM/hlm/documents/2015/ECE\\_HBP\\_2015\\_4.ru.pdf](http://www.unece.org/fileadmin/DAM/hlm/documents/2015/ECE_HBP_2015_4.ru.pdf) (9.11.2017).
- [2] Рейтинг устойчивого развития городов Российской Федерации. Sustainable growth management agency (ООО «Агентство ЭС ДЖИ ЭМ») [Электронный ресурс] – Режим доступа: <http://agencysgm.com/projects/Рейтинг%20устойчивого%20развития-2015.pdf> (9.11.2017).
- [3] Akaslan, D. An Analogy Between Womb and Home for Supporting the Aspects of Smart Cities / D. Akaslan, S. Taskin // 4th Int. Istanbul Smart Grid Congress and Fair. – New York: IEEE Press, 2016.
- [4] De Domenico, M. Personalized Routing for Multitudes in Smart Cities / M. De Domenico, A. Arenas, A. Lima, M.C. Gonzalez // EPJ Data Science. – 2015. – Vol.1. – P. 1-11.
- [5] Glebova, I.S. Possibilities of «Smart City» Concept Implementing: Russia's Cities Practice / I.S. Glebova, Y.S. Yasnitskaya, N.V. Maklakova // Mediterranean J. of Social Sciences. – 2014. – Vol.12. – P. 129-133.
- [6] Ishkineeva, G. Major Approaches Towards Understanding Smart Cities Concept / G. Ishkineeva, F. Ishkineeva, S. Akhmetova // Asian Social Science. – 2015. – Vol. 5. – P. 70-73.
- [7] Komarevtseva, O.O. Smart city technologies: new barriers to investment or a method for solving the economic problems of municipalities? / O.O. Komarevtseva // R-Economy. – 2017. – Vol. 3. – P. 32-39.
- [8] Khatoun, R. Smart Cities: Concepts, Architectures, Research Opportunities, Association for Computing Machinery / R. Khatoun, S. Zeadally // Communications of the ACM. – 2016. – Vol.8. – P. 46-57.
- [9] Khaimovich, I.N. Use of big data technology in public and municipal management / I.N. Khaimovich, V.M. Ramzaev, V.G. Chumak // CEUR Workshop Proceedings. – 2016. – Vol. 1638. – P. 864-872.
- [10] Khaimovich, I.N. Challenges of data access in economic research based on Big Data technology / I.N. Khaimovich, V.M. Ramzaev, V.G. Chumak // CEUR Workshop Proceedings. – 2015. – Vol. 1490. – P. 327-337.

## Modeling data for the analysis of the correspondence of cities in the Volga region to the digital state format

I.N. Khaymovich<sup>1,2</sup>, V.M. Ramzaev<sup>1</sup>, V.G. Chumak<sup>1</sup>

<sup>1</sup>Samara University of Public Administration “International Market Institute”, 21, G.S. Aksakova Street, Samara, Russia 443030

<sup>2</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

**Abstract.** The article suggests the methodology for assessing the readiness of Volga region municipal entities to introduce digital state. The authors worked out a model of statistic tests based on multiple probability theoretic and statistical modelling of parameter values of technological elements of digital economy. This model will allow the Volga region cities to define their possibility to participate in the State Program of Digital Economy, to choose cities most suitable for introduction of modern technologies, to identify the main shortcomings hindering their integration into the program. This research may be of interest to experts in the field of digital economy and Big Data management.

**Keywords:** digital state, data modeling, probability-theoretic modeling, introduction of digital technologies.