

*Рыкова Ю.С. Исследование количества пользователей единой системы идентификации и аутентификации в регионах российской федерации при помощи регрессионного и факторного анализов // Академия педагогических идей «Новация». Серия: Студенческий научный вестник. – 2019. – №2 (февраль). – АРТ 153-эл. – 0,3 п.л. - URL: <http://akademnova.ru/page/875550>*

**РУБРИКА: ЭКОНОМИЧЕСКИЕ НАУКИ**

**УДК 311:33**

**Рыкова Юлия Сергеевна**

студентка 1 курса магистратуры, экономический факультет

*Научный руководитель:* Бакуменко Л.П., д.э.н., профессор

ФГБОУ ВО "Поволжский государственный технологический университет"

г. Йошкар-Ола, Российская Федерация

e-mail: [yuliya-ryk96@yandex.ru](mailto:yuliya-ryk96@yandex.ru)

**ИССЛЕДОВАНИЕ КОЛИЧЕСТВА ПОЛЬЗОВАТЕЛЕЙ ЕДИНОЙ  
СИСТЕМЫ ИДЕНТИФИКАЦИИ И АУТЕНТИФИКАЦИИ В  
РЕГИОНАХ РОССИЙСКОЙ ФЕДЕРАЦИИ ПРИ ПОМОЩИ  
РЕГРЕССИОННОГО И ФАКТОРНОГО АНАЛИЗОВ**

*Аннотация:* В статье рассмотрено исследование количества пользователей единой системы идентификации и аутентификации и факторов, оказывающих на него влияние, при помощи методов статистического анализа.

*Ключевые слова:* эконометрическое моделирование, ИКТ, ЕСИА, регрессионный анализ, факторный анализ.

**Rykova Yuliya Sergeevna**

1st year undergraduate student, faculty of economics

Superisor: L. P. Bakumenko, DeS, Professor

Federal State Budgetary Educational Institution of Higher Education «Volga

State University of Technology»

Yoshkar-Ola, Russian Federation

e-mail: [yuliya-ryk96@yandex.ru](mailto:yuliya-ryk96@yandex.ru)

**STUDY OF THE NUMBER OF USERS OF A UNIFIED  
IDENTIFICATION AND AUTHENTICATION SYSTEM IN THE  
REGIONS OF THE RUSSIAN FEDERATION BY MEANS OF  
REGRESSIONAL AND FACTOR ANALYSIS**

*Abstract:* The article considers the study of the number of users of a unified system of identification and authentication and the factors influencing it using the methods of statistical analysis.

*Keywords:* econometric modeling, ICT, ESIA, regression analysis, factor analysis

Происходящие сегодня технологические, организационные, управленческие перемены, связанные с распространением информационных и коммуникационных технологий (ИКТ), охватывают все сферы деятельности – экономику, государственное управление, культуру, здравоохранение, образование, трансформируют повседневную жизнь человека и создают новые способы коммуникации. Фактический и потенциальный эффект развития ИКТ вызывает большой интерес на глобальном, национальном и региональном уровнях. Задачи анализа влияния развития ИКТ инфраструктуры на население привлекают многих.

Эконометрическое моделирование — мощный аналитический инструмент. Он позволяет на основе математического и статистического моделирования решить ряд задач. Эконометрические методы сейчас являются не только мощным инструментарием для получения новых знаний в экономике, но и широко применяемым аппаратом для принятия практических решений в прогнозировании.

Основные статистические показатели развития информационного общества в субъектах Российской Федерации и межрегиональные сравнения зависят от большого количества характеризующих их параметров, что обуславливает трудности, связанные с выявлением структуры их взаимосвязей. В подобных ситуациях, когда решение принимается на основе анализа стохастической, неполной информации необходимо применение методов многомерного статистического анализа.

Многомерный экономико-статистический анализ опирается на широкий спектр методов. Регрессионный и факторный анализы – это методы среди множества возможных вероятностно - статистических моделей позволяют обоснованно выбрать ту, которая наилучшим образом соответствует исходным статистическим данным, характеризующим реальное поведение исследуемой совокупности объектов, оценить надежность и точность выводов, сделанных на основании ограниченного статистического материала.

Для реализации рассмотренных методов многомерной классификации была выбрана система STATISTICA.

Анализ проводился по следующим показателям:

у — количество пользователей ЕСИА в расчете на 100 чел. населения, человек;

$x_1$  — Число телефонных аппаратов (включая таксофоны) в расчете на 100 чел. населения, шт.;

$x_2$  — Число подключенных абонентских устройств подвижной радиотелефонной связи в расчете на 100 чел. населения, шт.;

$x_3$  — Удельный вес телефонизированных населенных пунктов в общем числе сельских населенных пунктов, проценты;

$x_4$  — Число абонентов фиксированного широкополосного доступа к Интернету в расчете на 100 чел. населения;

$x_5$  — Число абонентов мобильного доступа к Интернету в расчете на 100 чел. населения;

$x_6$  - Число пунктов коллективного пользования (доступа), имеющих выход в Интернет, в расчете на 10 000 чел. населения;

$x_7$  - Инвестиции в основной капитал организаций сектора ИКТ в расчете на 100 человек населения, тыс. рублей;

$x_8$  — Персональные компьютеры (в расчете на 100 домашних хозяйств), шт.;

$x_9$  - Мобильные телефоны (в расчете на 100 домашних хозяйств), шт.

Переменная  $y$  (количество пользователей ЕСИА) рассматривается, как зависимая (критериальная) переменная, а все остальные переменные являются независимыми (предикторы).

Проведем построение регрессионной модели с учетом всех факторов, как значимых, так и незначимых. Результаты представлены на рисунке 1 (стандартный метод).

Итоги регрессии для зависимой переменной: y (Таблица даны R= ,91999126 R2= ,84638392 Скорректир. R2= ,82744495 F(9,73)=44,690 p<0,0000 Станд. ошибка оценки: 43,786						
N=83	БЕТА	Стд. Ош. БЕТА	В	Стд. Ош. В	t(73)	p-уров.
<b>Св.член</b>			-6,87051	45,59751	-0,15068	0,880646
x1	-0,213812	0,085945	-2,69088	1,08164	-2,48778	0,015138
x2	-0,091238	0,078016	-0,27552	0,23559	-1,16948	0,246016
x3	0,050186	0,051370	0,25456	0,26057	0,97695	0,331822
x4	0,038006	0,079123	0,66330	1,38088	0,48035	0,632416
x5	0,058627	0,065530	0,41750	0,46666	0,89467	0,373905
x6	0,131483	0,052043	12,30035	4,86873	2,52640	0,013692
x7	0,880949	0,049472	4,08963	0,22967	17,80690	0,000000
x8	0,053550	0,072955	0,24307	0,33114	0,73402	0,465287
x9	-0,071265	0,059153	-0,20926	0,17370	-1,20475	0,232192

Рисунок 1 - Итоговая таблица регрессии (стандартный метод)

Регрессионная статистика показывает, что коэффициент корреляции равен 0,92, а коэффициент детерминации 0,846, что говорит о том, что модель является достаточно качественной и 84% факторов были учтены в модели.

С учетом всех факторов при проведении анализа данных получается следующая модель:

$$\bar{Y}_x = -6,87 - 2,69x_1 - 0,28x_2 + 0,25x_3 + 0,66x_4 + 0,42x_5 + 12,3x_6 + 4,09x_7 + 0,24x_8 - 0,21x_9$$

О данной модели можно сказать то, что каждый коэффициент регрессии при факторном признаке показывает, на сколько единиц изменится количество пользователей ЕСИА при аналогичном изменении какого-либо из исследуемых факторных признаков на 1 соответствующую ему единицу измерения.

Значение -6,87 отражает усредненное влияние всех факторных признаков, не входящих в модель.

Чтобы исключить факторы, которые не являются значимыми, можно воспользоваться стандартным методом, методом исключения и методом включения факторов.

При использовании стандартного метода получилось 3 значимых фактора ( $x_1, x_6, x_7$ ), представленных на рисунке 1.

Воспользуемся методом исключения (рисунок 2).

Данный метод позволил сделать вывод, что значимыми являются три фактора:  $x_1, x_6, x_7$ .

Итоги регрессии для зависимой переменной: y (Таблица дан R= ,91650295 R2= ,83997765 Скорректир. R2= ,83390085 F(3,79)=138,23 p<0,0000 Станд. ошибка оценки: 42,960						
N=83	БЕТА	Стд. Ош. БЕТА	B	Стд. Ош. B	t(79)	p-уров.
Св.член			-34,2581	17,69216	-1,93634	0,056402
x1	-0,192983	0,045394	-2,4287	0,57129	-4,25130	0,000058
x6	0,148701	0,046616	13,9111	4,36097	3,18992	0,002042
x7	0,871051	0,046746	4,0437	0,21701	18,63353	0,000000

Рисунок 2 – Итоговая таблица регрессии (метод с исключением)

По регрессионной статистике можно сделать вывод, о том, что модель является качественной, так как коэффициент детерминации  $R_2=0,84$ . Из этого следует, что 84% фактора были учтены в модели. Наблюдаемое значение F велико ( $F_{набл.} = 138,23$ ) и превышает F критическое, следовательно, можно сделать вывод, о том, что гипотеза о статистической значимости и надежности уравнения принимается при уровне значимости 0,05. Показатель множественной корреляции, равный 0,917 превышает отметку 0,7, т.е. связь количества пользователей ЕСИА с совокупностью факторов сильная.

Уравнение регрессии имеет следующий вид:

$$\bar{Y}_x = -34,2581 - 2,42,87x_1 + 13,9111x_6 + 4,0437x_7$$

По данной модели можно сделать вывод, что наибольшее влияние на количество пользователей ЕСИА оказывает число пунктов коллективного пользования (доступа), имеющих выход в Интернет.

Интерпретируем данную модель:

При уменьшении числа телефонных аппаратов (включая таксофоны) в расчете на 100 чел. населения на 1 штуку количество пользователей ЕСИА (в расчете на 100 чел. населения) уменьшится на 2 человека.

При увеличении числа пунктов коллективного пользования (доступа), имеющих выход в Интернет, в расчете на 10 000 чел. населения на 1 штуку количество пользователей ЕСИА (в расчете на 100 чел. населения) увеличится на 13 человек.

При увеличении инвестиций в основной капитал организаций сектора ИКТ в расчете на 100 человек населения на 1 тысячу рублей количество пользователей ЕСИА (в расчете на 100 чел. населения) увеличится на 4 человека.

Проведем факторный анализ при помощи метода главных компонент. Количество выделенных факторов равно трем, а их собственные значения соответственно составляют 3,25942; 1,53036 и 1,23161.

На следующем шаге сформируем таблицу собственных значений, назначением и полезностью которых является принятие решений о том, сколько факторов нужно оставить. Таблица собственных значений и процентов общей дисперсии представлена на рисунке 3.

Значен.	Собственные значения (Таблица данных1) Выделение: Главные компоненты			
	Соб. зн.	% общей дисперс.	Кумулятивн. собст. знач.	Кумулятивн. %
1	3,259422	36,21580	3,259422	36,21580
2	1,530356	17,00396	4,789778	53,21976
3	1,231610	13,68455	6,021388	66,90431

Рисунок 3 – Таблица собственных значений и процентов общей дисперсии

Как видно из таблицы, собственное значение для первого фактора равно 3,259422; а доля дисперсии, объясненная первым фактором равна приблизительно 36,22%. Второй фактор включает в себя около 17% дисперсии, собственное значение равно 1,530356. Третий фактор включает в себя около 13,68% дисперсии, собственное значение равно 1,231610. Можно сделать вывод, что первый фактор наиболее значимый, т.к. он объясняет более трети вариаций переменных.

Используя критерий Кайзера, оставляем факторы с собственными значениями большими 1. Из приведенной выше таблицы следует, что критерий приводит к выбору полученных трех факторов.

Факторные нагрузки можно интерпретировать как корреляции между факторами и переменными. Поэтому они представляют наиболее важную информацию, на которой основывается интерпретация факторов. Чем теснее связь с рассматриваемым фактором, тем выше значение факторной нагрузки. Если коэффициент факторной нагрузки больше 0,7, то он выделяется красным цветом и соответствующая переменная включается в состав фактора.

Сначала посмотрим на повернутые (без вращения) факторные нагрузки для трех факторов (рисунок 4).

Перемен.	Фактор. нагрузки (без вращ.) (Табл. Выделение: Главные компоненты (Отмечены нагрузки >,700000))		
	Фактор 1	Фактор 2	Фактор 3
x1	0,814095	0,036471	-0,190270
x2	0,839570	-0,239915	0,002084
x3	-0,011671	0,760797	0,214712
x4	0,789329	-0,010499	-0,105318
x5	0,690742	-0,268354	0,106749
x6	-0,048504	0,619340	-0,599564
x7	0,207424	0,113306	-0,720307
x8	0,795346	0,218943	0,158588
x9	0,336959	0,613330	0,472592
Общ. дис.	3,259422	1,530356	1,231610
Доля общ	0,362158	0,170040	0,136846

Рисунок 4 – Факторные нагрузки (без вращения)

Чем теснее связь данного признака с рассматриваемым фактором, тем выше значение факторной нагрузки. Положительный знак факторной нагрузки указывает на прямую (а отрицательный знак - на обратную) связь данного признака с фактором.

Значимость признаков, участвующих в формировании главной компоненты, можно установить коэффициентом информативности. Проверим информативность данных компонентов при помощи данного коэффициента, рассчитываемого по формуле:

$$K_{и} = \frac{\sum a_i^2(>0,7)}{\sum a_i^2(\text{все})} * 100\%$$

Получены следующие  $K_{и}$  для данных факторов:

$$K_{и_1} = 80,48; K_{и_2} = 37,82; K_{и_3} = 42,13.$$

На первый фактор приходится наибольшее количество значимых факторных нагрузки, на второй и третий факторы приходится приблизительно одинаковое количество значимых факторных нагрузки.

В таблице на рисунке 4 приведены существенные нагрузки на первый фактор (коэффициент информативности составляет 80,48%) для переменных  $x_1$  (число телефонных аппаратов (включая таксофоны) в расчете на 100 чел. Населения),  $x_{21}$  (число подключенных абонентских устройств подвижной радиотелефонной связи в расчете на 100 чел. населения),  $x_4$  (число абонентов фиксированного широкополосного доступа к Интернету в расчете на 100 чел. населения) и  $x_8$  (персональные компьютеры в расчете на 100 домашних хозяйств). Второй фактор имеет наименьшую нагрузку (коэффициент информативности составляет 37,82%) для переменной  $x_3$  (удельный вес телефонизированных населенных пунктов в общем числе сельских населенных пунктов). Третий фактор имеет чуть большую чем у второго фактора нагрузку (коэффициент информативности составляет 42,13%) для переменной  $x_7$  (инвестиции в основной капитал организаций сектора ИКТ в расчете на 100 человек населения).

Произведем регрессионный анализ, используя переменную  $y$  (число пользователей ЕСИА) и три полученных фактора. Уравнение множественной регрессии в обычной форме будет выглядеть следующим образом:

$$Y = 1,49431 * f_1 + 19,5434 * f_2 - 72,2102 * f_3 + 14,9440$$

Таким образом, при увеличении первого фактора на 1 единицу, количество пользователей ЕСИА увеличится на 1 человека (в расчете на 100 человек населения) или при росте второго фактора на 1 единицу, количество пользователей ЕСИА увеличится на 19 человек (в расчете на 100 человек населения), или при росте 3 фактора на 1 единицу, количество

пользователей ЕСИА уменьшится на 72 человека (в расчете на 100 человек населения).

Регрессионный и факторный анализы позволили установить, что на количество пользователей ЕСИА влияют различные показатели развития ИКТ-инфраструктуры. Расчеты показали, что роль каждого показателя имела более или менее выраженный статистический характер. Таким образом, количественную характеристику показателей, оказывающих влияние на количество пользователей ЕСИА по регионам РФ возможно построить только на основе системы статистических показателей. В комплексе перечисленные группы факторов позволяют получить достаточно полное и всестороннее представление.

#### **Список использованной литературы:**

1. Артамонов, Н. В. Введение в эконометрику / Н.В. Артамонов. - М.: МЦНМО, 2016. - 224 с.;
2. Заварина, Е. С. Основы региональной статистики / Е. С. Заварина, К. Г. Чобану. - М. : Финансы и статистика, 2016. - 416 с.;
3. Колпаков, В. Ф. Экономико-математическое и эконометрическое моделирование. Компьютерный практикум [Текст] : учеб. пособие / В. Ф. Колпаков. - М. : Инфра-М, 2017. - 396 с.;
4. Кочетыгов, А. А. Основы эконометрики / А.А. Кочетыгов, Л.А. Толоконников. - М.: Издательский центр "МарТ", 2015. - 352 с.;
5. Статистические сборники ВШЭ - НИУ ВШЭ. - Режим доступа: <https://www.hse.ru/primarydata/>

*Дата поступления в редакцию: 05.02.2019 г.*

*Опубликовано: 12.02.2019 г.*

*© Академия педагогических идей «Новация». Серия «Студенческий научный вестник», электронный журнал, 2019*

*© Рыкова Ю.С., 2019*