

*Елена КЛОЧКОВА*

## МЕТОДЫ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ВЗАИМОСВЯЗЕЙ ПОКАЗАТЕЛЕЙ УРОВНЯ ИНФОРМАЦИОННОГО РАЗВИТИЯ РЕГИОНОВ

*Программно-целевые методы государственного управления в последние годы получили значимое развитие в России. Основная цель этих методов – обеспечение повышения эффективности и результативности государственного управления, создание условий для эффективного и своевременного достижения планируемых результатов. Инструментом реализации стратегических документов выступают государственные программы, обосновывающие оптимальный вариант достижения целей и приоритетов развития государственной политики. В области развития сферы информационных технологий реализуется Государственная программа «Информационное общество (2011–2020 годы)». Для проведения регулярной оценки результативности и эффективности ее реализации, а также оценки вклада в решение вопросов модернизации и инновационного развития экономики актуальным становится вопрос об адекватных методологических подходах. В статье предложена методика статистического анализа, позволившая определить факторы, оказывающие влияние на степень развития информационного общества в регионах. Предложена модель взаимосвязи основных показателей сектора ИКТ. В качестве информационной базы для анализа использованы показатели вышеназванной госпрограммы.*

**Ключевые слова:** информационно-коммуникационные технологии, информационное общество, корреляционно-регрессионный анализ, регион, федеральный округ

Российская Федерация достаточно сильно дифференцирована по уровню развития информационного общества. Можно также утверждать, что ей свойственны черты «цифрового неравенства». Об этом свидетельствует дифференциация субъектов Федерации по таким индикаторам сектора информационно-коммуникационных технологий (далее – ИКТ), как телефонная плотность фиксированной связи; проникновение подвижной радиотелефонной (сотовой) связи; число персональных компьютеров, используемых в учебных целях, на 100 обучающихся государственных и муниципальных общеобразовательных учреждений и др.<sup>1</sup>.

<sup>1</sup> Распоряжение Правительства Российской Федерации от 20 октября 2010 г. № 1815-р «О государственной программе Российской Федерации „Информационное общество (2011–2020 годы)“».

### ***Методика отбора факторов для оценки уровня развития сектора ИКТ***

Анализ показывает, что между развитостью сектора ИКТ и уровнем ИКТ-инфраструктуры региона существует зависимость.

Так, например, чем больше удельный вес отгруженной продукции сектора ИКТ в общем объеме отгруженной продукции региона, тем ниже уровень развития на его территории сферы ИКТ. Эта зависимость объясняется *высоким уровнем цен на услуги связи* в субъектах Федерации (таких, например, как регионы Северо-Кавказского федерального округа), где уровень проникновения Интернета и подвижной связи отстает от общероссийских значений<sup>2</sup>. Следовательно, величина стоимости отгруженных продуктов и оказанных услуг в этих регионах *зависит в большей степени от ценового фактора*, нежели от количества спроса на товары и услуги ИКТ. Данным фактом обуславливается и относительно высокая доля отгруженной продукции сектора ИКТ в этих регионах, что, в свою очередь, характеризует конъюнктуру рынка ИКТ товаров и услуг в регионе<sup>3</sup>.

Для подтверждения данной зависимости между показателем удельного веса отгруженных продуктов, выполненных работ, услуг сектора ИКТ в общем объеме отгруженных продуктов, выполненных работ, услуг и базовыми показателями развития ИКТ в субъектах Федерации можно применить методы множественного корреляционно-регрессионного анализа. Они, как мы полагаем, позволят с определенной долей вероятности *подтвердить наличие предполагаемой связи*, количественно определить силу и направление влияния рассматриваемых факторов, установить аналитическую форму их взаимозависимости<sup>4</sup>.

Для построения множественной регрессионной модели на основе наблюдений по регионам Российской Федерации в качестве результативного показателя (Y) выступает удельный вес объема отгруженных товаров (выполненных работ, услуг) собственного производства организаций сектора ИКТ в общем объеме отгруженных товаров (выполненных работ, услуг) собственного производства организаций региона.

В качестве независимых факторов выступают показатели, входящие в состав интегрального показателя государственной программы «Информационное общество (2011–2020 годы)» – «Степень дифференциации субъектов Российской Федерации по интегральным показателям информационного развития»<sup>5</sup>, где:

---

<sup>2</sup> Наиболее высокими тарифами обладают Чеченская Республика и Республика Ингушетия, где абонентская плата услуг доступа к Интернету составляет около 1300–1900 руб. в месяц (около 10% среднедушевого денежного дохода населения в данных регионах) при среднероссийском значении 540 руб. в месяц (2,1% от среднедушевого денежного дохода в целом по России).

<sup>3</sup> Индикаторы информационного общества: 2015: статистический сборник / Г.И. Абдрахманова, Л.М. Гохберг, М.А. Кевеш и др.; Национальный исследовательский университет «Высшая школа экономики». М.: НИУ ВШЭ, 2015.

<sup>4</sup> Статистика: учебник для бакалавров / Н.А. Садовникова [и др.] / под ред. В.Г. Минашкина. М.: Издательство «Юрайт», 2013. С. 160. Серия: Бакалавр. Базовый курс.

<sup>5</sup> Информационное общество: тенденции развития в субъектах Российской Федерации. Выпуск 2: статистический сборник / Г.И. Абдрахманова, Л.М. Гохберг, Г.Г. Ковалева; Национальный исследовательский университет «Высшая школа экономики». М.: НИУ ВШЭ, 2015.

$X_1$  – телефонная плотность фиксированной связи (включая таксофоны), штук на 100 чел. населения, единиц;

$X_2$  – проникновение подвижной радиотелефонной (сотовой) связи на 100 чел. населения, единиц;

$X_3$  – число пунктов коллективного пользования (доступа), имеющих выход в сеть Интернет, на 10 000 чел. населения, единиц;

$X_4$  – удельный вес домохозяйств, имеющих доступ к сети Интернет, в общем числе домашних хозяйств, в %;

$X_5$  – доля организаций, использующих широкополосный доступ к сети Интернет, в общем количестве организаций, в %;

$X_6$  – число персональных компьютеров, используемых в учебных целях, на 100 обучающихся государственных и муниципальных общеобразовательных учреждений, в %;

$X_7$  – доля учреждений здравоохранения, использовавших Интернет, в общем числе учреждений здравоохранения, в %.

Информационной базой анализа являются данные Федеральной службы государственной статистики (далее – Росстат)<sup>6</sup>.

На предварительном этапе анализа корреляционных зависимостей между перечисленными показателями и построением регрессионной модели оценивается экономическая сущность анализируемых показателей. Рассматриваемые показатели разносторонне описывают развитие информационного общества в регионах России.

Например, показатели  $X_1$  –  $X_3$  в целом характеризуют ИКТ-инфраструктуру и доступ ИКТ в регионе как фактор развития информационного общества, описывая соответственно уровень фиксированной телефонной связи, подвижной сотовой связи, использования компьютеров и доступа в Интернет. Группа показателей  $X_4$  –  $X_7$  характеризует уровень использования ИКТ для развития. При этом показатели  $X_4$  и  $X_5$  описывают уровень использования Интернета населением и бизнесом, а остальные два показателя – готовность учреждений образования и здравоохранения к развитию на основе ИКТ.

С учетом того, что в структуре объема отгруженной продукции организаций сектора ИКТ преобладают организации электросвязи, и, зная, что факторные показатели влияют обратно пропорционально на результивный показатель, следует предположить, что *большее влияние* на величину удельного веса отгруженной продукции сектора ИКТ окажут показатели ИКТ-инфраструктуры (уровень телефонной плотности фиксированной связи и проникновения подвижной связи). Также определенное влияние на нее будут оказывать два показателя доступа к Интернету домашних хозяйств и организаций:

- число персональных компьютеров, используемых в учебных целях, на 100 обучающихся государственных и муниципальных общеобразовательных учреждений;
- доля учреждений здравоохранения, использовавших Интернет, в общем числе учреждений здравоохранения.

<sup>6</sup> Мониторинг развития информационного общества в Российской Федерации / URL: [http://www.gks.ru/wps/wcm/connect/rosstat\\_main/rosstat/ru/statistics/science\\_and\\_innovations/it\\_technology/](http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/science_and_innovations/it_technology/).

Наши предположения о характере и силе взаимосвязи по каждому из представленных показателей отчасти подтверждаются рассчитанными автором парными коэффициентами корреляции (см. табл. 1).

Т а б л и ц а 1

**Матрица парных коэффициентов корреляции**

Показатели	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
Y	1	-0,446	-0,438	-0,191	-0,675	0,149	-0,324	-0,063
X <sub>1</sub>	-0,446	1	0,736	0,118	0,645	0,060	0,562	0,087
X <sub>2</sub>	-0,438	0,736	1	-0,008	0,511	0,221	0,520	-0,012
X <sub>3</sub>	-0,191	0,118	-0,008	1	0,136	-0,459	0,057	-0,325
X <sub>4</sub>	-0,675	0,645	0,511	0,136	1	0,109	0,430	0,071
X <sub>5</sub>	0,149	0,060	0,221	-0,459	0,109	1	0,138	0,433
X <sub>6</sub>	-0,324	0,562	0,520	0,057	0,430	0,138	1	0,065
X <sub>7</sub>	-0,063	0,087	-0,012	-0,325	0,071	0,433	0,065	1

Источник: Расчеты автора с использованием пакета IBM SPSS.

Следует отметить, что все рассчитанные парные коэффициенты корреляции по своей величине не достигают 0,8, что говорит об отсутствии мультиколлинеарности<sup>7</sup> факторов. Наибольшие значения парных коэффициентов характерны для показателя X<sub>1</sub> (телефонная плотность фиксированной связи, включая таксофоны, на 100 человек населения).

В частности, парный коэффициент корреляции ( $r_{X_1 X_2}$ ) между уровнем телефонной плотности фиксированной связи (X<sub>1</sub>) и уровнем проникновения подвижной радиотелефонной (сотовой) связи на 100 чел. населения (X<sub>2</sub>) составил 0,736, что говорит о сильной прямой взаимосвязи этих факторов. Данная взаимосвязь объясняется тем, что для регионов с достаточно высоким уровнем развития фиксированной связи уровень развития сотовой связи также велик.

Парный коэффициент корреляции между переменной X<sub>1</sub> и X<sub>4</sub> составляет 0,645, что говорит о сильной прямой зависимости между телефонной плотностью фиксированной связи и удельным весом домохозяйств, имеющих доступ к сети Интернет. Данная зависимость обусловлена тем, что доступ домохозяйств к Интернету осуществляется преимущественно через соединение фиксированной связи, поэтому в субъектах Федерации, где уровень предоставления технической возможности беспрепятственного доступа к Интернету высок, большее число домохозяйств имеет доступ в сеть.

В приведенной матрице парных коэффициентов корреляции достаточно много отрицательных значений коэффициентов корреляции при зависимом показателе, что подтверждает выдвинутое ранее предположение об обратном направлении связи. Согласно матрице парных

<sup>7</sup> Под мультиколлинеарностью понимается тесная корреляционная зависимость между факторными признаками.

коэффициентов корреляции наибольшее влияние на результативный показатель оказывают три фактора –  $X_1$ ,  $X_2$  и  $X_4$ . Наибольшее влияние на результативный показатель, согласно значению коэффициента корреляции, оказывает показатель доли домохозяйств, имеющих доступ в Интернет. Значение  $r_{YX_4}$  составляет  $-0,675$ , что говорит о сильной обратной по направлению связи между этими показателями.

Таким образом, можно утверждать, что, чем меньше уровень показателя доли домохозяйств, имеющих доступ к сети Интернет, чем меньше в них развито техническое оснащение электросвязи, тем выше уровень удельного веса объема отгруженных товаров, работ, услуг сектора ИКТ в общем объеме отгруженных товаров организаций региона. Значимой связью с  $Y$  обладает еще один показатель – Число персональных компьютеров, используемых в учебных целях, на 100 обучающихся государственных и муниципальных общеобразовательных учреждений ( $X_6$ ), что в целом вытекает из зависимости  $X_6$  и  $X_1$ .

Остальные парные коэффициенты корреляции результативного и факторных показателей –  $r_{YX_3}$ ,  $r_{YX_5}$ ,  $r_{YX_7}$  – статистически незначимы с вероятностью 99%.

Судя по матрице частных коэффициентов корреляции (см. табл. 2), умеренное влияние на результативный показатель оказывает признак  $X_4$ , частный коэффициент корреляции между ним и факторным показателем  $r_{YX_4 / X_1 X_2 X_3 X_5 X_6 X_7} = -0,599$ .

Т а б л и ц а 2

Матрица частных коэффициентов корреляции

Показатели	Y	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
Y	1	0,173	-0,262	-0,002	-0,599	0,325	-0,045	-0,047
$X_1$	0,173	1	0,606	0,073	0,423	-0,255	0,240	0,223
$X_2$	-0,262	0,606	1	-0,081	-0,125	0,335	0,143	-0,268
$X_3$	-0,002	0,073	-0,081	1	0,108	-0,336	0,048	-0,184
$X_4$	-0,599	0,423	-0,125	0,108	1	0,277	0,039	-0,016
$X_5$	0,325	-0,255	0,335	-0,336	0,277	1	0,086	0,368
$X_6$	-0,045	0,240	0,143	0,048	0,039	0,086	1	0,011
$X_7$	-0,047	0,223	-0,268	-0,184	-0,016	0,368	0,011	1

При фиксации влияния всех остальных факторов связь между  $Y$  и  $X_1$  поменяла свое направление, по сравнению с парным коэффициентом, и стала незначимой, а между  $Y$  и  $X_2$  – уменьшилась сила связи при сохранении ее направления. Чистое влияние фактора  $X_5$  на  $Y$  показывает наличие статистически значимой слабой прямой связи между ними. Исходя из значений матрицы частных коэффициентов, можно заключить – фактором, наиболее сильно влияющим на результативный показатель, является показатель  $X_4$ .

### Моделирование взаимосвязей показателей уровня развития ИКТ

Для оценки силы взаимосвязи набора факторных и результативного показателя используются также множественные коэффициенты корреляции (см. табл. 3).

Влияние исследуемых факторов достаточно велико, что подтверждает возможность использования показателя удельного веса объема отгруженных товаров, работ, услуг сектора ИКТ в общем объеме отгруженных товаров организаций региона в качестве результативного для построения регрессионной модели.

Множественные коэффициенты корреляции  $r_{X_1/X_2 X_3 X_4 X_5 X_6 X_7 Y} = 0,834$ ,  $r_{X_2/X_3 X_4 X_5 X_6 X_7 Y X_1} = 0,800$  и  $r_{X_4/X_5 X_6 X_7 Y X_1 X_2 X_3} = 0,800$  говорят о *тесной взаимозависимости* набора факторных признаков. Таким образом, в дальнейшем исследовании в качестве факторов могут быть использованы показатели уровня телефонной плотности фиксированной связи регионов ( $X_1$ ), проникновение подвижной радиотелефонной (сотовой) связи ( $X_2$ ) и удельный вес домохозяйств, имеющих доступ к сети Интернет ( $X_4$ ).

Т а б л и ц а 3

#### Множественные коэффициенты корреляции

Коэффициент	Значение коэффициента
$r_{Y/X_1 X_2 X_3 X_4 X_5 X_6 X_7}$	0,738
$r_{X_1/X_2 X_3 X_4 X_5 X_6 X_7 Y}$	0,834
$r_{X_2/X_3 X_4 X_5 X_6 X_7 Y X_1}$	0,800
$r_{X_3/X_4 X_5 X_6 X_7 Y X_1 X_2}$	0,524
$r_{X_4/X_5 X_6 X_7 Y X_1 X_2 X_3}$	0,800
$r_{X_5/X_6 X_7 Y X_1 X_2 X_3 X_4}$	0,668
$r_{X_6/X_7 Y X_1 X_2 X_3 X_4 X_5}$	0,594
$r_{X_7/Y X_1 X_2 X_3 X_4 X_5 X_6}$	0,522

Из результатов предшествующего анализа зависимости  $Y$  от  $X_i$  следует, что взаимосвязь результативной переменной и признаков имеет линейный характер. Таким образом, регрессионное уравнение (1) по рассматриваемой группе показателей, построенное на данных в региональном разрезе по совокупности субъектов Федерации за 2013 г., с включением всех исходных показателей имеет вид:

$$\hat{y} = 22,069 + 0,149x_1 - 0,067x_2 - 0,007x_3 - 0,293x_4 + 0,204x_5 - 0,060x_6 - 0,040x_7 \quad (1)$$

$$t_{\text{набл}} \quad (2,065) \quad (1,468) \quad (-2,268) \quad (-0,015) \quad (-6,265) \quad (2,878) \quad (-0,379) \quad (-0,391)$$

Коэффициент детерминации для данной модели  $R^2 = 0,544$ . Это означает, что *только 54,4%* вариации зависимой переменной объясняется факторными показателями. Уравнение регрессии значимо

$F_{\text{набл}} = 11,93 > F_{\text{кр}} = 2,25$  (0,05; 7; 74). Величина стандартной ошибки оценки регрессии  $\hat{S} = 3,99$ . При этом коэффициенты уравнения при показателях  $X_1, X_3, X_6, X_7$  не значимы, поскольку наблюдаемые значения  $t_i$  критерия Стьюдента для этих коэффициентов не превосходят критическое  $t_{\text{кр}} (\alpha = 0,1; \nu = 78) = 1,671$ .

В целях совершенствования модели применяется метод пошагового исключения переменных для отбора тех факторов, которые оказывают наибольшее влияние на зависимую переменную  $Y$ . На начальном этапе реализации метода находится и исключается переменная, которая является незначимой на заданном уровне значимости. Затем оценки параметров уравнения пересчитываются без исключенной переменной, и процедура проводится снова. Последовательное исключение приведет к тому, что в модели регрессии останутся только значимые переменные<sup>8</sup>.

Таким образом, конечное уравнение регрессии строится на основе трех объясняющих переменных –  $X_2, X_4, X_5$ . Оставшиеся переменные не включаются в модель, т.к. коэффициенты регрессии при этих переменных не значимы при  $\alpha = 0,05$ .

$$\hat{y} = 17,178 - 0,041x_2 - 0,265x_4 + 0,177x_5 \quad (2)$$

$$t_{\text{набл}} \quad (3,453) \quad (-1,929) \quad (-6,576) \quad (3,127)$$

Уравнение (2) на 53% обосновывает вариацию удельного веса объема отгруженных товаров, работ, услуг собственного производства организаций сектора ИКТ – коэффициент детерминации  $R^2 = 0,530$ . Стандартная ошибка регрессии  $\hat{S} = 3,94$ . Чем выше значение стандартной ошибки оценки параметров регрессии, тем в большей степени величины оценок регрессии отличаются от наблюдаемых значений зависимой переменной и тем менее они надежны, поэтому полученное при помощи шагового отбора уравнение регрессии точнее по этому параметру, чем уравнение, включающее все показатели.

Уравнение регрессии значимо, т.к.  $F_{\text{набл}} = 27,787 > F_{\text{кр}} = 2,25$  (0,05; 6; 73). Значимы все коэффициенты модели –  $\beta_2, \beta_4, \beta_5$  – с вероятностью ошибки 0,1.

Коэффициент при переменной  $X_5$  свидетельствует о том, что при увеличении доли организаций, использующих широкополосный доступ к сети Интернет, на 1% удельный вес отгруженной продукции сектора ИКТ в общем объеме отгруженной продукции организаций региона уменьшается на 0,177 процентного пункта. Коэффициенты при переменных  $X_2$  и  $X_4$  означают, что увеличение количества подключенных к сетям абонентских станций (абонентских устройств) подвижной радиотелефонной связи в сети общего пользования в расчете на 100 чел. населения и удельного веса домохозяйств, имеющих доступ к сети Интернет, на 1 единицу своего измерения влечет за собой уменьшение результирующего показателя на 0,041% и 0,265%.

<sup>8</sup> Статистика: учебник для бакалавров... С. 160–161.

\* \* \*

Осуществленный автором корреляционно-регрессионный анализ показал *наличие статистически значимой взаимосвязи* между удельным весом объема отгруженных товаров (выполненных работ, услуг) собственного производства организаций сектора ИКТ в общем объеме отгруженных товаров (выполненных работ, услуг) собственного производства и базовыми показателями развития ИКТ. В число факторов, оказывающих наибольшее влияние, входят «Проникновение подвижной радиотелефонной (сотовой) связи на 100 человек населения» ( $X_2$ ), «Удельный вес домохозяйств, имеющих доступ к сети Интернет, в общем числе домашних хозяйств» ( $X_4$ ) и «Доля организаций, использующих широкополосный доступ к сети Интернет, в общем количестве организаций» ( $X_5$ ).

Это доказывает, что развитие информационного общества невозможно без перманентного инфраструктурного преобразования существующих телекоммуникационных сетей и информационных технологий, предоставления свободного доступа к таким технологиям, формирования навыков их использования. Сфера информационно-коммуникационных технологий характеризуется чрезвычайно высокой степенью трансформации, что влечет за собой появление в краткосрочном горизонте абсолютно новых сегментов позиционирования ИТ-продукции, открытие новых технологических направлений и связанные с этим изменения в направлениях фундаментальных исследований отрасли.