Анализ канонических корреляций показателей эколого-социально-экономического развития региона

Сыровацкая Ирина Владимировна
Аспирант кафедры статистики и эконометрики
Оренбургский государственный университет
Irina_Sirovatskaya@mail.ru

Аннотация. В статье рассмотрен метод анализа канонических корреляций системы статистических показателей, характеризующих эколого-социально-экономическое развитие региона.

Abstract. In article the method of the analysis of initial correlations of system of the statistics characterizing ecological, social and economic development of region is considered.

Ключевые слова: эколого-социально-экономическое развитие региона, система статистических показателей, анализ канонических корреляций.

Keywords: ecological, social and economic development of region, system of statistics, the analysis of initial correlations.

Изучение закономерностей развития системы экономика-общество-природа является в настоящее время весьма актуальным: в центре внимания многих учёных находится важная задача соизмерения экономического развития с качеством жизни населения и возможностями окружающей природной среды воспринимать последствия такого развития.

Статистическое исследование эколого-социально-экономического развития региона возможно в рамках системного подхода, т.е. когда изучению подлежит влияние не отдельных факторов на процессы развития и их результаты, а целая система факторов и результатов [8]. Обоснование

того, что система показателей региональной статистики соответствует принципам системного подхода, представлено в работе [6, с.17- 18].

На основе системы статистических показателей, построенной в соответствии с принципами системного анализа и адекватно отражающей состояние и развитие экономики, общества и окружающей природной среды, может быть получена полная, своевременная и достоверная статистическая информация, которая будет в дальнейшем использована для принятия управленческих решений на уровне региона и его административно-территориальных образований.

В соответствии со Стратегией социально-экономического развития Оренбургской области [1] нами была предложена система статистических показателей, характеризующая экологическое, социальное и экономическое развитие региона в целом и в разрезе его административно-территориальных образований (городов и районов) [7].

Рис. 1 – Блочная архитектура системы статистических показателей, характеризующих эколого-социально-экономическое развитие региона

Докажем, что сформированная нами система статистических показателей, характеризующая эколого-социально-экономическое развитие региона является системой, отвечающей принципам системного анализа, а не произвольным набором, перечнем показателей, посредством изучения наличия, тесноты и направления связи между структурными элементами системы показателей. Ставится задача статистического исследования зависимостей между направлениями и блоками показателей сформированной системы методом анализа канонических корреляций. Практическая реализация метода осуществлялась в ППП Statistica 6.0 [2].

Анализ канонических корреляций проводился по 47 административно-территориальным образованиям (12 городам и 35 районам) Оренбургской области, относящейся к группе регионов со слабой устойчивостью тенденции роста экономики, которую формируют более одной трети регионов РФ, обеспечивающих экономический рост [5]. Период исследования определён шестилетием, начиная с 2004 г. Этот период характерен завершением перехода государственной статистики в РФ к новой системе классификации – по видам экономической деятельности, и учётом экономических показателей согласно ОКВЭД. Показатели, имеющие стоимостные единицы измерения, были пересчитаны в сопоставимые цены 2009 года с помощью индекса-дефлятора валового регионального продукта. Чтобы избежать так называемой «привязки» к территории, часть показателей была получена расчётным путём (на душу населения; на 1000 человек трудоспособного населения; на 1000 населения). Для анализа использовались официальные публикации территориального органа Федеральной службы государственной статистики по Оренбургской области.

В ходе проверки предпосылок применения метода канонических корреляций выявлено, что совокупность из 57 показателей имеет многомерное нормальное распределение. Анализ вариации показал, что практически по всем рассмотренным показателям наиболее выделяется г. Оренбург, т.е. является “выбросом”. При проведении анализа

канонических корреляций рекомендуется выявлять и исключать выбросы. Однако исключать выявленный “выброс” из совокупности городов и районов Оренбургской области не целесообразно, так как г. Оренбург представляет собой субъект экономической деятельности, имеющий хозяйственные связи с другими городами и районами региона.

Канонический анализ позволяет исследовать зависимость между двумя множествами переменных (в нашем случае – блоками показателей), при этом достаточно ограничиться рассмотрением небольшого числа наиболее коррелированных линейных комбинаций из каждого множества (блока показателей) [4].

Для рассмотрения берутся взвешенные суммы по множествам, чтобы веса, приписанные отдельным слагаемым, соответствовали реальной “структуре” переменных, т.е. их взаимной значимости:

$$a_1 y_1 + a_2 y_2 + \ldots + a_p y_p = b_1 x_1 + b_2 x_2 + \ldots + b_q x_q , \qquad (1)$$

При этом под знаком “ = ”, подразумевается наличие стохастической взаимосвязи между линейными комбинациями переменных обоих множеств.

Таким образом, если имеются два множества, содержащие p и q переменных соответственно, то будет исследоваться зависимость между взвешенными суммами переменных из каждого множества (т.е. между линейными комбинациями p и q переменных соответственно).

После того как сформулировано в общем виде “уравнение модели” для канонической корреляции, определяются веса для двух наборов переменных. При подборе весовых коэффициентов исходят из условия максимальной коррелированности двух множеств, т.е. рассматриваются все максимально коррелированные взвешенные суммы (максимизируется значение корреляции между каноническими переменными). При этом последовательно получаемые пары канонических переменных не коррелированны друг с другом и объясняют всё меньшую и меньшую долю изменчивости.

Взвешенные суммы определяют канонический корень, или каноническую переменную, число канонических корней равно числу переменных в меньшем множестве.

Проанализируем корреляции внутри и между множествами – показателями направлений “Развитие человеческого потенциала” и “Общеэкономические ориентиры”. В первом множестве (показатели направления “Развитие человеческого потенциала”) наибольшая зависимость наблюдается между переменными *Y3 - На 1000 человек трудоспособного возраста приходится лиц старше трудоспособного возраста* и *Y5 - Коэффициент смертности (на 1000 человек населения)*: коэффициент парной корреляции равен 0,87. Во втором множестве (показатели направления “Общеэкономические ориентиры”) сильная зависимость наблюдается между *X1 - Сальдированный финансовый результат (прибыль минус убыток) деятельности организаций на одно предприятие (тыс. руб.)* и *X4 - Инвестиции в основной капитал на душу населения (в сопост. ценах; руб.)*: коэффициент парной корреляции равен 0,92. Для переменных из разных множеств наибольшая корреляция выявлена между *Y28 - Оборот розничной торговли на душу населения (руб.)* и *X14 - Грузооборот автомобильного транспорта организаций всех видов деятельности (тыс. тонно-км)*: коэффициент парной корреляции равен 0,93. Предположительно, что переменные *Y3, Y5, Y28, X1, X4* и *X14* будут определять значение канонической корреляции между двумя множествами показателей.

Далее вычислено столько собственных значений матрицы $B = R^{-1}{}_{pp} \cdot R_{pq} \cdot R^{-1}{}_{qq} \cdot R_{qp}$, сколько имеется канонических корней (т.е. столько, сколько переменных содержит наименьшее множество). Собственные значения матрицы B, ранжированные по убыванию, равняются квадратам канонических корреляций (коэффициентам корреляций между множествами). Для каждого канонического корня значения канонической корреляции, статистики χ^2, p - уровень приведены на рис. 2.

Root Removed	Chi-Square Tests with Successive Roots Removed (Средн					
	Canonicl R	Canonicl R-sqr.	Chi-sqr.	df	p	**Lambda Prime**
0	0,999841	0,999682	755,5494	434	0,000000	0,000000
1	0,996830	0,993669	570,3406	390	0,000000	0,000000
2	0,992050	0,984163	453,9078	348	0,000122	0,000000
3	0,986153	0,972498	358,5627	308	0,025516	0,000000
4	0,963917	0,929136	275,9126	270	0,389812	0,000006
5	0,934259	0,872839	215,0317	234	0,807270	0,000087
6	0,908247	0,824912	167,5987	200	0,953385	0,000684
7	0,865402	0,748921	127,5220	168	0,991189	0,003909
8	0,823866	0,678756	95,7362	138	0,997577	0,015569
9	0,791569	0,626581	69,6185	110	0,999039	0,048466
10	0,743221	0,552378	46,9622	84	0,999637	0,129790
11	0,659626	0,435107	28,4747	60	0,999819	0,289954
12	0,642993	0,413440	15,3390	38	0,999602	0,513291
13	0,353432	0,124914	3,0690	18	0,999967	0,875086

Рис. 2 – Канонические корреляции между показателями направлений “Развитие человеческого потенциала” и “Общеэкономические ориентиры”

Данный рисунок показывает, что статистически значимым являются первые десять канонических корней, которым соответствуют первые десять пар канонических переменных. Однако мы ограничимся рассмотрением первой пары канонических переменных, для которой наблюдается максимальное значение канонической корреляции $R = 0{,}9998$:

$$\begin{cases} V^1 = 0{,}14y_1 + 0{,}13y_2 + \ldots 1{,}04y_{19} - 0{,}93y_{20} + \ldots - 0{,}55y_{30} + 0{,}09y_{31} \\ U^1 = 0{,}36x_1 - 0{,}19x_2 + \ldots - 0{,}06x_{11} - 0{,}24x_{12} + 0{,}11x_{13} + 1{,}17x_{14} \end{cases} \quad (2)$$

Значение канонической корреляции $R = 0{,}9998$, свидетельствует о наличии сильной зависимости между показателями направлений “Развитие человеческого потенциала” и “Общеэкономические ориентиры”. Статистика $\chi^2 = 755{,}5$ и уровень значимости $p = 0{,}00$ показывают значимость данной канонической корреляции.

Так как канонический корень представляет собой две взвешенные суммы, по одной на каждое множество, то для толкования “смысла” канонического корня рассматриваются канонические веса, сопоставленных каждому множеству переменных. Чем больше приписанный вес (т.е.

абсолютное значение веса), тем больше вклад соответствующей переменной (показателя) в значение канонической переменной. Рассмотрение канонических весов позволяет увидеть, как конкретные переменные (показатели) в каждом множестве влияют на взвешенную сумму (т.е. каноническую переменную).

Из выражения (2) видно, что для левого множества (показатели направления "Развитие человеческого потенциала") наибольший вклад в значение первой канонической переменной вносят *Y19 - Численность врачей на 10000 человек населения*. Для правого множества (показатели направления "Общеэкономические ориентиры") наибольший вклад в значение первой канонической переменной вносит *X14 - Грузооборот автомобильного транспорта организаций всех видов деятельности (тыс. тонно-км)*.

Канонические веса соответствуют уникальному вкладу, вносимому соответствующей переменной (показателем) во взвешенную сумму или каноническую переменную; коэффициенты канонической корреляции соответствуют корреляции между взвешенными суммами по двум множествам переменных. Однако, они не говорят ничего о том, какую часть изменчивости (дисперсии) каждый канонический корень объясняет в переменных.

Тем не менее, можно сделать заключение о доле объясняемой дисперсии, рассматривая нагрузки канонических факторов. Нагрузки канонических факторов представляют собой обычные корреляции между каноническими переменными и показателями из каждого множества, и их можно интерпретировать так же, как и в факторном анализе (таблица 1).

Наиболее тесная корреляционная связь существует между канонической переменной V^1 и показателями первого множества *Y28 - Оборот розничной торговли на душу населения* (коэффициент корреляции равен 0,69) и *Y19 - Численность врачей на 10000 человек населения* (коэффициент корреляции равен 0,63).

Таблица 1 - Факторная нагрузка канонических переменных V^1 и U^1

Каноническая переменная V^1				Каноническая переменная U^1	
Y1	0,17	*Y17*	0,08	*X1*	0,11
Y2	-0,14	*Y18*	0,42	*X2*	-0,37
Y3	0,02	*Y19*	0,63	*X3*	-0,03
Y4	-0,17	*Y20*	0,27	*X4*	0,12
Y5	0,00	*Y21*	0,29	*X5*	-0,06
Y6	-0,07	*Y22*	0,17	*X6*	0,09
Y7	-0,23	*Y23*	-0,04	*X7*	0,09
Y8	-0,17	*Y24*	0,01	*X8*	0,21
Y9	-0,18	*Y25*	0,09	*X9*	-0,07
Y10	-0,12	*Y26*	-0,05	*X10*	0,08
Y11	0,08	*Y27*	0,04	*X11*	-0,06
Y12	0,11	*Y28*	0,69	*X12*	-0,02
Y13	-0,03	*Y29*	0,49	*X13*	0,09
Y14	-0,12	*Y30*	0,17	*X14*	0,84
Y15	0,14	*Y31*	-0,03		
Y16	0,21				

Каноническая переменная U^1 тесно связана с *X14 - Грузооборот автомобильного транспорта организаций всех видов деятельности* (коэффициент корреляции равен 0,84).

На основе этих корреляций между каноническими переменными и показателями из каждого множества получают меру избыточности множества переменных, которая показывает насколько реальная изменчивость в одном множестве переменных объясняется другим множеством. Общая избыточность для первого множества составляет 56,1 %, т.е. показатели направления “Развитие человеческого потенциала” объясняют 56,1 % вариации показателей направления “Общеэкономические ориентиры”; которые в свою очередь объясняют 79,7 % вариации показателей первого множества.

Аналогичным образом были исследованы взаимосвязи показателей направлений “Развитие человеческого потенциала” и “Создание комфортной

среды обитания"; "Общеэкономические ориентиры" и "Создание комфортной среды обитания" (таблицы 2 и 3 соответственно).

Таблица 2 – Общие результаты анализа канонических корреляций между направлениями "Развитие человеческого потенциала" и "Создание комфортной среды обитания"

<table>
<tr><th>Направление "Развитие человеческого потенциала"</th><th>Направление "Создание комфортной среды обитания"</th></tr>
<tr><td colspan="2">Каноническая корреляция $R = 0{,}99740$. Зависимость сильная
$\chi^2 = 589.46$ $p = 0{,}0000$ Каноническая корреляция значима.
Максимальная корреляция между показателями:
Y21 - Число больничных коек на 10000 человек населения и Z9 - Уровень преступности (коэффициент парной корреляции r_{xy}=0,71).</td></tr>
<tr><td>Левое множество</td><td>Правое множество</td></tr>
<tr><td>Общая избыточность: 49,3%
Наибольший вклад в зависимость вносит: Y19 - Численность врачей на 10000 человек населения</td><td>Общая избыточность: 78,1%
Наибольший вклад в зависимость вносит: Z9 - Уровень преступности</td></tr>
</table>

Существующую зависимость направлений "Развитие человеческого потенциала" и "Создание комфортной среды обитания" во многом определяют показатели *Y19 – Численность врачей на 10000 человек населения* и *Z9 – Уровень преступности* (таблица 2). Общая избыточность правого множества составляет 78,1%, левого – 49,3%.

Взаимосвязь направлений "Общеэкономические ориентиры" и "Создание комфортной среды обитания" наиболее сильно проявлена через вклад показателей *X4 – Инвестиции в основной капитал на душу населения* и *Z8 – Наличие квартирных телефонных аппаратов сети общего пользования на 1000 человек населения* (коэффициент канонической корреляции равен 0,95). Рассматриваемые множества в равной степени объясняют вариацию друг друга (таблица 3).

Таблица 3 – Общие результаты анализа канонических корреляций между направлениями “Общеэкономические ориентиры” и “Создание комфортной среды обитания”

<table>
<tr><th>Направление “Общеэкономические ориентиры”</th><th>Направление “Создание комфортной среды обитания”</th></tr>
<tr><td colspan="2">Каноническая корреляция $R = 0{,}95287$ Зависимость сильная
$\chi^2 = 250.49$ $p = 0{,}0000$ Каноническая корреляция значима
Максимальная корреляция между показателями:
X8 - Число организаций, имевших веб-сайт и Z8 - Наличие квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (коэффициент парной корреляции r_{xy}=0,74).</td></tr>
<tr><td>Левое множество</td><td>Правое множество</td></tr>
<tr><td>Общая избыточность: 43,4%
Наибольший вклад в зависимость вносит: X4 - Инвестиции в основной капитал на душу населения</td><td>Общая избыточность: 44,1%
Наибольший вклад в зависимость вносит: Z8 - Наличие квартирных телефонных аппаратов сети общего пользования на 1000 человек населения</td></tr>
</table>

Анализ канонических корреляций между блоками показателей внутри направления “Развитие человеческого потенциала” выявил, что наиболее сильная зависимость существует между блоками показателей “Демографическая структура и воспроизводство населения” и “Сфера образования, культуры и искусства” (коэффициент канонической корреляции равен 0,942). Наибольший вклад в зависимость двух множеств вносят показатели: *Y3 – На 1000 человек трудоспособного возраста приходится лиц старше трудоспособного возраста* и *Y24 – Число мест в учреждениях культурно-досугового типа на 1000 человек населения.*

Внутри направления “Общеэкономические ориентиры” показатели *X4 – Инвестиции в основной капитал на душу населения* и *X12 – Ввод в действие жилых домов на 1000 человек населения* вносят наибольший вклад в сильную зависимость между блоками “Финансы и инвестиции” и “Производство товаров и услуг (по основным видам экономической деятельности” (коэффициент канонической корреляции составил 0,853).

Внутри направления “Создание комфортной среды обитания” между блоками “Экологическая сфера” и “Жилищные условия населения” проявлена умеренная зависимость (коэффициент канонической корреляции равен 0,725), отметим, что показатели блока “Жилищные условия населения” способствуют существенным изменениям в социальной сфере – каноническая корреляция с блоком “Безопасность социальной сферы” равна 0,571.

Аналогичным образом проводился анализ канонических корреляций между блоками показателей различных направлений.

Таким образом, результаты канонического анализа показателей сформированной системы (рис. 1) показал наличие статистической зависимости между структурными элементами. Данное обстоятельство говорит о том, что предложенная нами система статистических показателей эколого-социально-экономического развития региона не является произвольным перечнем показателей, отвечает основным принципам системного анализа и является системой как таковой.

Библиографический список:

1. Стратегия развития Оренбургской области до 2020 года и на период до 2030 года [Электронный ресурс] : утверждена постановлением Правительства Оренбургской области от 20 августа 2010 года № 551-пп. – Оренбург, 2010. - Режим доступа : http://www.orenburg-gov.ru/magnoliaPublic/regportal/strateg/2030.html. - Дата обращения 15.10.2010.

2. Боровиков, В. П. STATISTICA. Искусство анализа данных на компьютере : для профессионалов. – СПб. : Питер, 2003. – 688 с.

3. Дрогобыцкий, И. Н. Системный анализ в экономике : учеб. пособие / И. Н. Дрогобыцкий. – М. : Финансы и статистика, 2007. – 512 с.

4. Дубров, А. М. Многомерный статистические методы : учеб. / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – М. : Финансы и статистика, 2000. – 352 с.

5. Зарова, Е. В. Качество экономического роста региона : методологические аспекты статистического исследования / Е. В. Зарова, М. А. Котякова // Вопросы статистики, 2006. - № 5. – С. 51-61.

6. Региональная статистика : учеб. / под ред. Е. В. Заровой. – М. : Финансы и статистика, 2006. – 624 с.

7. Сыровацкая, И. В. Формирование системы статистических показателей, характеризующих социально-экономическое развитие региона: методологические аспекты / И. В. Сыровацкая // Вестник СГЭУ, 2011. - №5 (79). – С.86-90.

8. Юзбашев, М. М. Системный эффект в экономике и жизни / М. М. Юзбашев, Т. М. Михайлова // Вопросы статистики, 2007. - № 11.- С.71-73.