ISSN 2304-120X
КОНЦЕПТ
научно-методический электронный журнал

Деркаченко О. В. Кластеризация и дискриминантный анализ регионов Приволжского федерального округа по уровню отдельных социально-экономических показателей // Концепт. – 2016. – № 02 (февраль). – ART 16028. – 0,3 п. л. – URL: http://e-koncept.ru/2016/16028.htm. – ISSN 2304-120X.

ART 16028 **УДК 332.025.12**

Деркаченко Олег Валентинович,
кандидат экономических наук, начальник лаборатории ФГУ «3 Центральный научно-исследовательский институт» Министерства обороны Российской Федерации, г. Москва
derkachenko77@mail.ru

Кластеризация и дискриминантный анализ регионов Приволжского федерального округа по уровню отдельных социально-экономических показателей

Аннотация. *В статье приводятся результаты кластеризации регионов Приволжского федерального округа по социально-экономическим показателям и дана интерпретация кластеров. Построены дискриминантные функции для определения принадлежности «новых» регионов к кластерам. Выявлена принадлежность отдельного региона к кластеру со средним уровнем социально-экономических показателей. Получена эконометрическая модель объема инвестиций в основной капитал в зависимости от валового регионального продукта.*
Ключевые слова: *регионы, показатели, информация, кластерный анализ, интерпретация кластеров, дискриминантные функции, принадлежность, «новые» регионы, эконометрическая модель.*
Раздел: *(04) экономика.*

В условиях экономического кризиса и усиливающейся международной конкуренции устойчивость российской экономики во многом зависит от способности каждого региона успешно развиваться и конкурировать на рынках. Изменения в мире, вызванные политическими и экономическими причинами, требуют новых подходов к социально-экономическому развитию российских регионов. Социально-экономическое положение регионов неоднородно. По одним показателям регион может находиться в числе передовых, а по другим, наоборот, имеет средние или вообще низкие значения. Каждый регион обладает своими уникальными природно-климатическими условиями, географическим и геополитическим положением, в каждом – особая демографическая ситуация. В таких условиях для учета региональных различий по уровню социально-экономических показателей целесообразно воспользоваться статистическими методами многомерной классификации субъектов РФ, в частности методом кластерного анализа.

Главная особенность кластерного анализа состоит в том, что различия между объектами, входящими в выделенную группу (кластер), незначительны, а различия между группами существенны.

Методы кластерного анализа позволяют:

– объединять объекты в группы (кластеры) все более высокой общности на основе критерия минимума расстояния в пространстве изучаемых показателей, описывающих эти объекты;

– разбивать множество объектов на заданное число кластеров [1–3].

Дискриминантный анализ является разделом многомерного статистического анализа, который включает в себя методы классификации многомерных наблюдений по принципу максимального сходства при наличии обучающих признаков. В кластерном анализе рассматриваются методы многомерной классификации без обучения. В дискриминантном анализе новые кластеры не образуются, а формулируется правило,

по которому объекты подмножества, подлежащего классификации, относятся к одному из уже существующих (обучающих) подмножеств (классов) на основе сравнения величины дискриминантной функции классифицируемого объекта, рассчитанной по дискриминантным переменным, с некоторой константой дискриминации. Дискриминантная функция в общем виде запишется так:

$$У = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k. \quad (1)$$

В качестве зависимой переменной выступает номинальная переменная, идентифицирующая принадлежность объектов (регионов) к одной из нескольких групп (кластеров). Независимые переменные (x_1, x_2 … x_k) такие же, как и в регрессионном анализе (количественные и качественные).

В данной статье рассматривается применение кластерного и дискриминантного анализа для решения задачи классификации регионов Приволжского ФО по социально-экономическим показателям за 1-й квартал 2015 г. Для этого используется новая статистическая информация, приведенная на сайте Росстата (www.gks.ru), а именно ежеквартальное электронное издание «Статистическое обозрение» № 3 (94) 2015 г. [4] Статистические данные отдельных социально-экономических показателей регионов Приволжского ФО приведены в табл. 1. Для выделения однородных регионов Приволжского ФО по уровню отдельных социально-экономических показателей используется кластерный анализ [5]. Общая модель кластеризации имеет вид:

$$dab = \alpha ldlb + \alpha f\ dfb + \beta dlf + \gamma(dlb - dfb), \quad (2)$$

где α_l, α_f, β, γ – параметры, определяющие алгоритм образования кластеров;

d_{lb}, d_{fb}, d_{lf} – расстояние между соответствующими кластерами.

При определенных значениях параметров α_l, α_f, β, γ общее уравнение (2) преобразуется в конкретный алгоритм. В данном исследовании использовался алгоритм «средняя связь»:

$$d_{ab} = N_l/(N_l + N_f)\ d_{lb} + N_f/(N_l + N_f)\ d_{fb}, \quad (3)$$

где – N_l, N_f соответственно кластеры «l» и «f» с числом объектов N.

Зависимость (3) получена из общего уравнения (2) при следующих значениях параметров:

$$\alpha_l = N_l/(N_l + N_f);\ \alpha_f = N_f/(N_l + N_f) \text{ и } \beta = \gamma = 0.$$

Для классификации регионов на однородные группы используется пакет прикладных программ “Statistica” (модуль – кластерный анализ). Схема работы следующая: Ввод исходных данных в таблицу – Анализ – Многомерный анализ – Кластерный анализ – Иерархическая классификация – ОК – Файл данных (исходные данные). Переменные строки. Правило объединения: метод полной связи. Мера близости: евклидово расстояние – ОК – Выбор переменных (Var1-Var5) – ОК – Вертикальная дендрограмма – Результат: дендрограмма классификации 14 регионов Приволжского ФО по пяти переменным (x_1-x_5). На рисунке: по оси абсцисс – номера регионов, по оси ординат – евклидово расстояние. Для выделения кластеров используется граничное значение евклидова расстояния, которое задается в пределах $d_{гр} = (0{,}5 – 0{,}9)d_{max}$. Из рисунка видно, что $d_{max} = 18{,}0$. Задаемся коэффициентом, равным 0,6. Следовательно, граничное значение равно 10,8. Для этого значения проведем на рисунке линию, параллельную оси абсцисс, которая позволит выделить три кластера (см. табл. 2).

Таблица 1

Социально-экономические показатели регионов Приволжского ФО за 1-й квартал 2015 г.

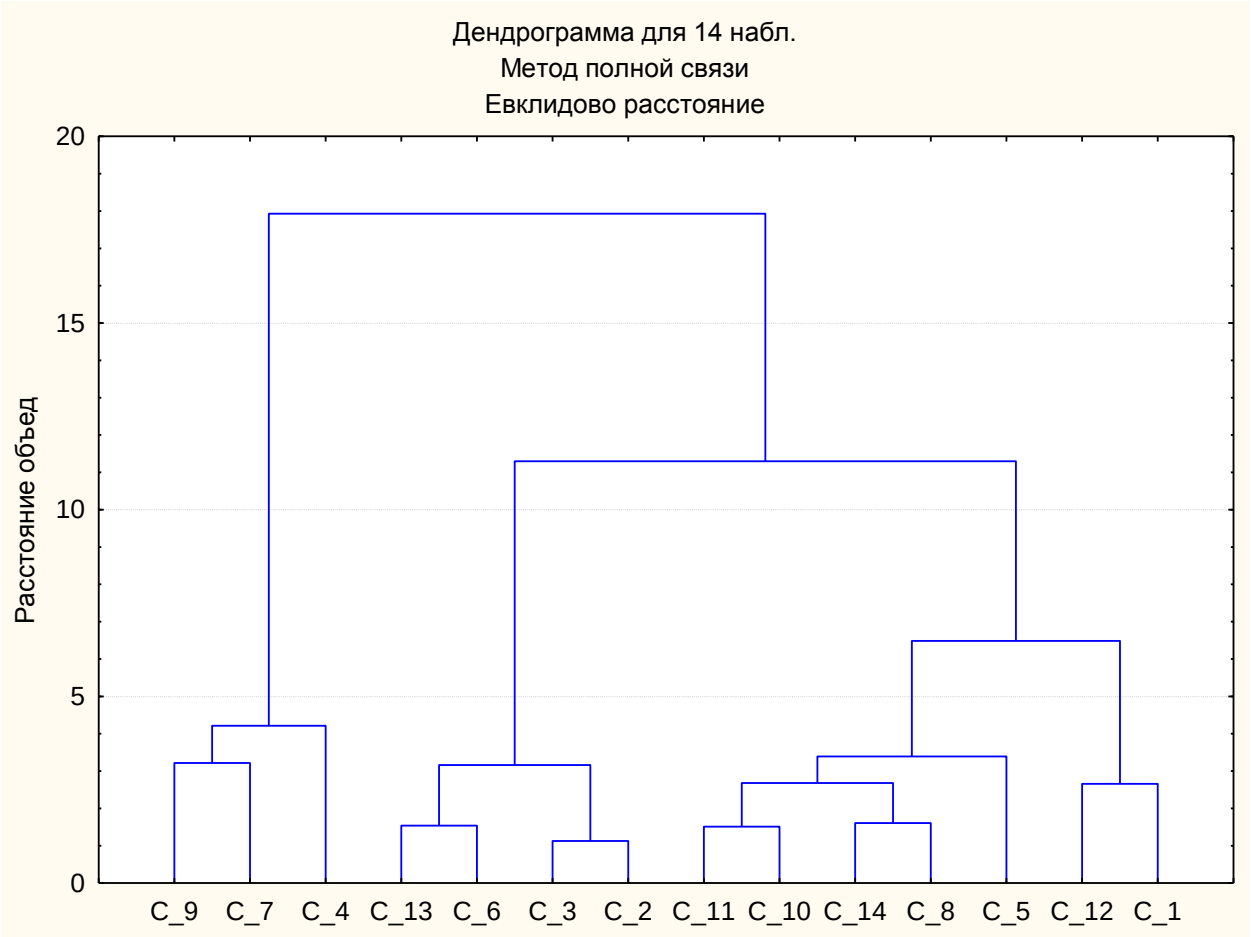
Регионы Приволжского ФО	***Уровень безработицы, в процентах от экономически активного населения, x_1***	***Среднедушевые денежные доходы населения (в месяц), тыс. рублей, x_2***	***Среднедушевые денежные расходы населения на покупку товаров и оплату услуг (в месяц) , тыс. рублей, x_3***	***Среднемесяч- ная номинальная начисленная заработная плата, тыс. рублей, x_4***	***Средний размер назначенных пенсий (на 1 апреля 2015 г.), тыс. рублей, x_5***
С1. Республика Башкортостан	6,5	21,225	18,721	24,166	11,432
С2. Республика Марий Эл	5,2	15,628	11,285	19,880	10,732
С3. Республика Мордовия	4,6	14,954	10,725	20,249	10.823
С4. Республика Татарстан	4,8	25,840	21,080	27,208	11,601
С5. Удмуртская Республика	4,9	20,957	14,186	23,524	11,807
С6. Чувашская Республика	5,7	16,559	12,681	19,901	10,911
С7. Пермский край	6,0	29,351	19,363	26,271	11,983
С8. Кировская область	5,2	19,208	14,319	20,637	11,779
С9. Нижегородская область	4,2	26,974	18,975	25,137	11,884
С10. Оренбургская область	5,8	18,459	14,502	22,794	11,000
С11. Пензенская область	4,7	18,420	14,601	21,767	11.039
С12. Самарская область	4,3	22,072	17,981	25,067	11,816
С13. Саратовская область	5,1	16,787	13,044	21,250	10,958
С14. Ульяновская область	5,5	20,452	13,889	21,139	11,069

Таблица 2

Номера регионов в кластерах

Номер кластера	***Номер региона***
1	9, 7, 4
2	13, 6, 3, 2
3	11, 10, 14, 8, 5, 12, 1

Первый кластер характеризуется низким уровнем безработицы и высокими показателями x_2–x_5 по сравнению с другими кластерами. Второй кластер – средний уровень безработицы, но в регионах низкие доходы, заработная плата и пенсия. Третий кластер имеет более высокую безработицу, но средние социально-экономические характеристики. К этому кластеру относится Пензенская область. Поэтому по уровню отдельных социально-экономических показателей первый кластер характеризуется как кластер с хорошими показателями, третий – со средними и второй – с удовлетворительными.



Результаты кластеризации регионов по социально-экономическим показателям

Следующим этапом исследования является решение задачи отнесения «нового» региона к определенному кластеру. Новый регион – это регион, не участвующий в кластеризации, например Тамбовская область. Такую задачу можно решить, используя дискриминантный анализ [6]. Построив дискриминантные функции и подставив в них социально-экономические характеристики Тамбовской области, можно определить, к какому кластеру она относится. Показатели Тамбовской области приведены в табл. 3.

Таблица 3

Показатели Тамбовской области

X_1	X_2	X_3	X_4	X_5
4,6	19,853	16,135	20,016	10,751

По данным табл. 1 и дискриминантного анализа в ППП “Statistica” построены три модели для каждого кластера:

$$Y_1 = – 1168,94 + 42,14X_1 – 23,34X_2 – 10,59X_3 + 23,34X_4 + 199,75X_5.$$
$$Y_2 = – 1241,78 + 49,80X_1 – 34,92X_2 – 12,59X_3 + 25,26X_4 + 219,93X_5.$$
$$Y_3 = – 1256,87 + 48,24X_1 – 30,55X_2 – 11,61X_3 + 24,65X_4 + 218,21X_5.$$

В каждую модель подставим характеристики Тамбовской области (см. табл. 3) и определим дискриминантные функции. Имеем: Y_1 = 1004,821; Y_2 = 1000,67; Y_3 = 1010,568. По максимальному значению функции выбирается кластер, к которому принадлежит «новый» объект. В данном случае Тамбовская область входит в третий кластер. Принадлежность региона к кластеру можно определить на основе апостериорной вероятности (табл. 4). Из таблицы видно, что Тамбовская область (15) относится к третьему кластеру с вероятностью 0,995.

Таблица 4

Вероятности отнесения регионов к кластеру

Апостериорные вероятности (Таблица.sta) Неправильные классификации отмечены *

Номер	***Наблюдения***	***G_1:1***	***G_2:2***	***G_3:3***
1	G_3:3	0,000000	0,000141	0,999859
2	G_2:2	0,000000	0,997006	0,002994
3	G_2:2	0,000000	0,999385	0,000615
4	G_1:1	1,000000	0,000000	0,000000
5	G_3:3	0,000000	0,002394	0,997606
6	G_2:2	0,000000	0,965815	0,034185
7	G_1:1	1,000000	0,000000	0,000000
8	G_3:3	0,000000	0,033395	0,966605
9	G_1:1	1,000000	0,000000	0,000000
10	G_3:3	0,000000	0,296256	0,703744
11	G_3:3	0,000000	0,040532	0,959468
12	G_3:3	0,000007	0,000004	0,999989
13	G_2:2	0,000000	0,917579	0,082421
14	G_3:3	0,000000	0,001775	0,998225
15	---	0,004909	0,000057	0,995034

Таким образом, применение кластерного анализа позволило выявить однородные регионы Приволжского ФО по социально-экономическим показателям, установить кластеры с хорошими, средними и удовлетворительными характеристиками. Для кластера с удовлетворительными показателями целесообразно разработать сценарий развития этих субъектов с целью перехода в третий или первый кластер. На основе дискриминантного анализа построены дискриминантные модели и выявлена принадлежность Тамбовской области к кластеру со средними показателями.

Рассмотренные отдельные социально-экономические показатели регионов Приволжского ФО в значительной степени зависят от инвестиций в основной капитал. Эти инвестиции сильно связаны с валовым региональным продуктом (ВРП). Поэтому возникает необходимость в разработке статистической модели связи инвестиций в основной капитал с ВРП. Статистические данные приведены в табл. 5.

Таблица 5

Инвестиции в основной капитал и ВРП регионов Приволжского ФО за 1-й квартал 2015 г.

Регион	***Инвестиции в основной капитал, млн руб. (y)***	***Валовой региональный продукт, млн руб. (z)***
Республика Башкортостан	152842,1	1154 056
Республика Марий Эл	30603,6	117 598
Республика Мордовия	35320,4	132 474
Республика Татарстан	322327,6	1 436 933
Удмуртская Республика	60236,7	371 498
Чувашская Республика	34853,9	217 034
Пермский край	183029,7	897 598
Кировская область	39683,3	212 370
Нижегородская область	175091,0	838 599
Оренбургская область	118522,1	629 370
Пензенская область	53535,0	240 335
Самарская область	226694,2	941 611
Саратовская область	89561,6	477 352
Ульяновская область	42080,2	244 230

По статистическим данным (табл. 5) с использованием MS Excel (этап – регрессия) построена однофакторная модель (табл. 6).

Таблица 6

Результаты построения однофакторной модели

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,953491729
R-квадрат	0,909146477
Нормированный R-квадрат	0,90157535
Стандартная ошибка	27966,55242
Наблюдения	14

Дисперсионный анализ

	df	*SS*	*MS*	*F*
Регрессия	1	93918510877	93918510877	120,080734
				Значим. F
Остаток	12	9385536650	782128054,1	1,32E-07
Итого	13	1,03304E+11		

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	-1834,883389	12778,52472	-0,143591176	0,88820642
				1,32127E-
Переменная X 1 (Z)	0,200993314	0,018341926	10,95813552	07

Эконометрическая модель имеет вид:

$$У = – 1834,883 + 0,201\ Z.$$

Из табл. 6 видно, что коэффициент детерминации (R^2) равен 0,909 и по критерию Фишера модель является значимой. Полученная эконометрическая модель может

быть использована для прогнозирования объема инвестиций в зависимости от валового регионального продукта.

Без вложения инвестиций в основной капитал не может быть эффективного развития региональной экономики.

Ссылки на источники

1. Деркаченко В. Н., Зубков А. Ф. Многомерные статистические методы: учеб. – Пенза: Изд-во Пензен. гос. технол. акад., 2011. – 179 с.
2. Статистика: учеб. / под ред. В. С. Мхитаряна. – М.: Экономистъ, 2006. – 671 с.
3. Деркаченко В. Н. Эконометрика: учеб. пособие. – Пенза: Изд-во Пензен. гос. технол. акад., 2013. – 140 с.
4. Статистическое обозрение. – М., 2015. – № 3(94).
5. Деркаченко В. Н., Зубков А. Ф. Указ. соч.
6. Там же.

Oleg Derkachenko,
Candidate of Economic Sciences, the chief of the laboratory of federal office of state "3 the central scientific research institute of the RF Ministry of Defense", Moscow
derkachenko77@mail.ru

Clustering and the discriminant analysis of the regions of the Volga federal region on the level of the separate social and economic indices

Abstract. The paper deals with the results of clustering the regions of the Volga federal region according to the social and economic indices interpretation to clusters are given. Discriminators for determining the belonging "new" regions with the clusters are constructed. The belonging of separate region with the cluster with the average level of social and economic indices is revealed. The econometric model of the volume of fixed asset formation depending on gross regional product is obtained.

Key words: regions, indices, information, cluster analysis, interpretation of clusters,.

References

1. Derkachenko, V. N. & Zubkov, A. F. (2011). *Mnogomernye statisticheskie metody: ucheb.*, Izd-vo Penz. gos. tehnol. akad., Penza, 179 p. (in Russian).
2. Mhitarjan, V. S. (ed.) (2006). *Statistika: ucheb.,* Jekonomist#, Moscow, 671 p. (in Russian).
3. Derkachenko, V. N. (2013). *Jekonometrika: ucheb. posobie*, Izd-vo Penz. gos. tehnol. akad., Penza, 140 p. (in Russian).
4. (2015). *Statisticheskoe obozrenie*, Moscow, № 3(94) (in Russian).
5. Derkachenko, V. N. & Zubkov, A. F. (2011). Op. cit.
6. Ibid.

Рекомендовано к публикации:
Горевым П. М., кандидатом педагогических наук,
главным редактором журнала «Концепт»

Поступила в редакцию *Received*	28.12.15	Получена положительная рецензия *Received a positive review*	30.12.15
Принята к публикации *Accepted for publication*	30.12.15	Опубликована *Published*	29.02.16

ISSN 2304-120X 02
9 772304 120166
www.e-koncept.ru

© Концепт, научно-методический электронный журнал, 2016
© Деркаченко О. В., 2016