

## МАТЕМАТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ ЭКОНОМИКИ

УДК 330.101.54

А.Ю. Трусова\*

### СТАТИСТИКА ПОКАЗАТЕЛЕЙ ТРУДОВЫХ РЕСУРСОВ В ФОКУСЕ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МЕТОДОВ

В статье проведен анализ статистических данных структуры экономически занятого населения Самарского региона средствами многомерных статистических методов.

**Ключевые слова:** сжатие многомерных данных средствами факторного анализа, регрессионный анализ в пространстве латентных факторов, кластеризация, статистический пакет SPSS.

Вопросы статистики труда современного общества являются актуальными постоянно. Экономическая ситуация в регионах подвержена непрерывной изменчивости как в сторону улучшения показателей занятости населения, так и в сторону ухудшения. Статистические органы разного уровня ведут постоянный учет всех статистических показателей в сфере занятости и безработицы. Первичный статистический анализ представляется в виде многомерных данных, сгруппированных по фактору времени, пола, образования. Первичный анализ таких данных представлен в современной научной литературе достаточно полно. Использование аппарата многомерного статистического анализа в сфере экономической занятости населения в настоящее время представлено недостаточно. В этой связи анализ данных статистики экономической занятости населения средствами факторного анализа и кластерного анализа является актуальным и практически значимым.

В статье проводится визуализация и кластеризация данных среднегодовой численности и структуры занятых в экономике по видам экономической деятельности. Исходные данные взяты с официального сайта государственной статистики Самарской области за период с 2005 по 2015 год. Для исследования отобраны следующие статистические показатели:

$X_1$  – доля занятых сельским хозяйством, охотой и лесным хозяйством;  $X_2$  – доля занятых добычей полезных ископаемых;  $X_3$  – доля занятых обрабатывающим производством;  $X_4$  – доля занятых производством и распределением электроэнергии, газа и воды;  $X_5$  – доля занятых в строительстве;  $X_6$  – доля занятых в оптовой и розничной торговле, ремонтом автотранспортных средств, мотоциклов, бытовых изделий и предметов личного пользования;  $X_7$  – доля занятых в гостиницах и ресторанах;  $X_8$  – доля занятых в сфере транспорта и связи;  $X_9$  – доля занятых финансовой деятельностью;  $X_{10}$  – доля занятых операциями с недвижимым имуществом, арендой и предоставлением услуг;  $X_{11}$  – доля занятых в государственном управлении и обеспечении военной безопасности; социальном страховании;  $X_{12}$  – доля занятых в сфере образования;  $X_{13}$  – доля занятых в здравоохранении и предоставлении социальных услуг;  $X_{14}$  – доля занятых предоставлением прочих коммунальных, социальных и персональных услуг;  $X_{15}$  – доля занятых в государственной собственности;  $X_{16}$  – доля занятых в муниципальной собственности;  $X_{17}$  – доля занятых в собственности общественных и религиозных организа-

\* © Трусова А.Ю., 2017

Трусова Алла Юрьевна (a\_yu\_ssu@mail.ru), кафедра математики и бизнес-информатики, Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Российская Федерация, г. Самара, Московское шоссе, 34.

ций (объединений);  $X_{18}$  – доля занятых в частной собственности;  $X_{19}$  – доля занятых в смешанной российской собственности;  $X_{20}$  – доля занятых в иностранной, совместной российской и иностранной собственности;  $X_{21}$  – доля имеющих высшее образование;  $Y$  – израсходовано средств на мероприятия по охране труда, млн руб.

Обработка данных осуществлялась с использованием статистического пакета SPSS. На рис. 1 представлен фрагмент исходных данных.

	Y	X1	X2	X3	X4	X5	X6
1	1649,40	6,70	0,70	25,10	2,40	7,40	16
2	2088,60	6,50	0,80	24,60	2,30	7,30	16
3	2429,30	6,20	0,80	24,80	2,30	7,40	16
4	3424,20	6,20	0,80	24,00	2,40	7,40	16
5	2531,20	6,30	0,80	22,80	2,40	6,70	16
6	2952,00	6,30	0,70	22,20	2,50	6,70	16
7	3905,30	6,20	0,80	22,30	2,40	6,80	17
8	4221,60	6,10	0,90	21,90	2,80	6,90	17

Рис. 1. Фрагмент исходных данных

Исходный массив данных имеет размерность 11 x 22. Визуализация и комплексный анализ данных такой размерности затруднен. В этой связи средствами факторного анализа осуществлено сжатие массива данных до размерности 11 x 4. В последующем анализе рассматриваются 4 главных латентных фактора. В таблице 1 представлены собственные значения и дисперсии оставленных в анализе главных факторов.

Матрица факторного отображения без варимаксного вращения представлена в таблице 2.

Таблица 1

#### Дисперсии латентных факторов без вращения

Латентные факторы	Начальные собственные значения	% дисперсии	Кумулятивный %
$F_1$	13,833	65,871	65,871
$F_2$	3,388	16,136	82,007
$F_3$	1,744	8,305	90,312
$F_4$	0,882	4,199	94,511

Таблица 2

## Матрица факторного отображения

Наименование	Латентные факторы			
	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>
X <sub>1</sub>	<b>-0,837</b>	-0,168	-0,147	0,457
X <sub>2</sub>	<b>0,704</b>	0,574	0,14	0,031
X <sub>3</sub>	<b>-0,963</b>	-0,043	-0,034	-0,116
X <sub>4</sub>	<b>0,822</b>	0,204	-0,231	0,099
X <sub>5</sub>	-0,27	<b>0,9</b>	-0,066	-0,096
X <sub>6</sub>	<b>0,943</b>	0,127	-0,248	0
X <sub>7</sub>	-0,243	0,38	<b>0,814</b>	0,289
X <sub>8</sub>	<b>-0,897</b>	-0,208	0,296	-0,128
X <sub>9</sub>	<b>0,835</b>	0,076	0,258	-0,453
X <sub>10</sub>	<b>0,934</b>	0,085	-0,126	0,261
X <sub>11</sub>	0,359	-0,898	-0,182	-0,038
X <sub>12</sub>	<b>0,865</b>	-0,417	-0,14	0,153
X <sub>13</sub>	<b>0,761</b>	-0,458	0,412	-0,018
X <sub>14</sub>	<b>0,786</b>	-0,014	0,521	0,185
X <sub>15</sub>	<b>0,878</b>	0,414	-0,022	0,111
X <sub>16</sub>	<b>-0,883</b>	-0,446	0,001	0
X <sub>17</sub>	<b>-0,804</b>	0,284	-0,28	0,373
X <sub>18</sub>	<b>0,956</b>	0,213	-0,146	-0,074
X <sub>19</sub>	<b>-0,93</b>	0,311	-0,05	-0,15
X <sub>20</sub>	<b>0,858</b>	-0,411	0,206	0,163
X <sub>21</sub>	<b>0,923</b>	0,09	-0,334	-0,027

Как видно из матрицы факторного отображения, максимальную корреляцию имеет большинство показателей с первым латентным фактором. Следовательно, для визуализации и графического представления достаточно первых двух латентных факторов. На рис. 2 представлены изучаемые временные периоды в двумерном пространстве латентных факторов. В частности, по оси абсцисс представлен первый латентный фактор, по оси ординат – второй. Из графика видно, что наблюдается улучшение показателей экономической занятости населения Самарского региона по изучаемой структуре. Второй латентный фактор свидетельствует о периодических колебаниях изучаемых показателей интегрально. Такие временные периоды, как 2009, 2010 и 2011 годы, имеют отрицательные показатели по изучаемой структуре.

Проверка гипотезы о достаточном количестве оставленных в анализе факторов подтверждается на 5 %-ном уровне значимости. Таким образом, средствами факторного анализа исходный массив данных размерности 11 x 21 сжат до размерности 11 x 2. В таблице 3 представлены изучаемые временные периоды в координатах латентных факторов.

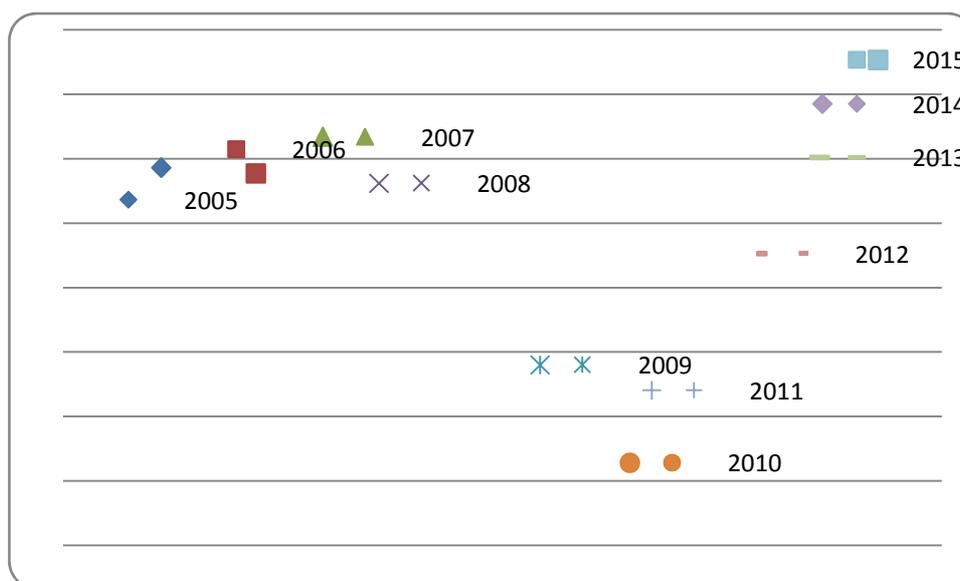


Рис. 2. Временные периоды изучаемой структуры экономической занятости населения Самарского региона в двумерном пространстве латентных факторов

Таблица 3

### Временные периоды в пространстве латентных факторов

t	Y	F <sub>1</sub>	F <sub>2</sub>
2005	1649,4	-1,60965	0,429175
2006	2088,6	-1,2327	0,386093
2007	2429,3	-0,96552	0,669615
2008	3424,2	-0,74188	0,308868
2009	2531,2	-0,09997	-1,09982
2010	2952	0,257555	-1,85999
2011	3905,3	0,344421	-1,2968
2012	4221,6	0,76449	-0,23636
2013	4483,3	1,013239	0,508138
2014	5408,1	1,024352	0,924984
2015	5519,9	1,245669	1,266097

Используя пространство новых факторов, мы построили регрессионную модель зависимости инвестиций в улучшение условий труда с целью снижения травматизма от латентных факторов структуры занятого в экономике Самарского региона населения. Статистика регрессионного анализа представлена в таблицах 4–6 (средствами MSExcel).

Таблица 4

### Вывод итогов. Регрессионная статистика

Множественный R	0,962669144
R-квадрат	0,926731881
Нормированный R-квадрат	0,908414851
Стандартная ошибка	396,6574514
Наблюдения	11

Таблица 5

**Вывод итогов. Дисперсионный анализ**

Наименование	df	SS	MS	F	Значимость F
Регрессия	2	15920631,2	7960315,61	50,594004	2,88178E-05
Остаток	8	1258697,07	157337,134		
Итого	10	17179328,3			

Таблица 6

**Вывод итогов. Статистика оценок параметров**

Наименование	Коэффициенты	Стандартная ошибка	t-статистика	P-значение
Y-пересеч.	3510,263636	119,5967215	29,35083497	1,96699E-09
F <sub>1</sub>	1192,216159	125,4329458	9,504808737	1,23856E-05
F <sub>2</sub>	413,0182192	125,4305015	3,292805293	0,010976

Уравнение линейной регрессии имеет вид

$$Y = 3510,26 + 1192,216F_1 + 413,018F_2.$$

Как видно, с увеличением доли занятого населения в сферах, связанных с производством, любой формы собственности необходимо увеличивать расходы на улучшение условий труда в среднем на 1192,2 млн рублей.

На рис. 3 представлен график частной регрессии переменной Y от латентного фактора F<sub>1</sub>.

**График частной регрессии**

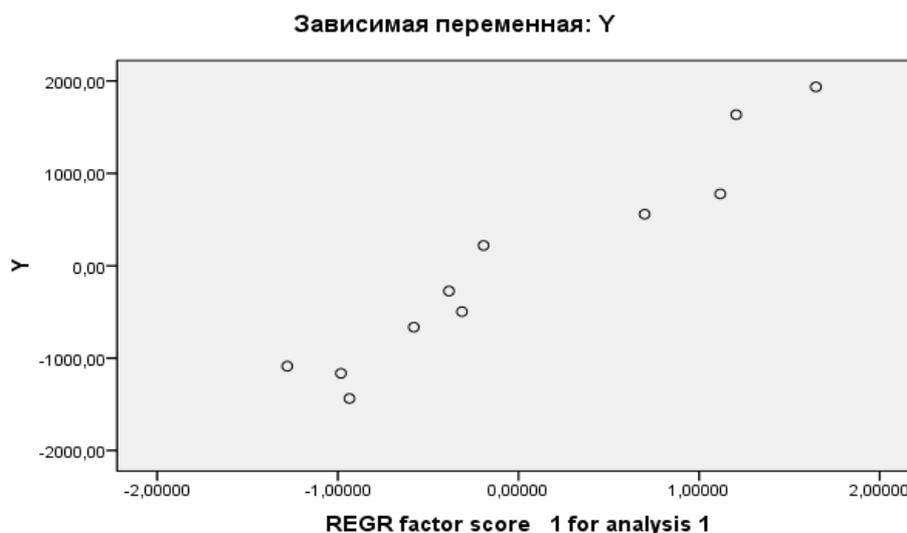


Рис. 3. Графическая зависимость частной регрессии переменной Y от латентного фактора F<sub>1</sub>

Средствами SPSS была проведена кластеризация временных периодов. При кластеризации использовался агломеративный метод «ближнего соседа». Временные периоды можно охарактеризовать как однородные по показателям структуры занятого в экономике населения Самарского региона

следующим образом: 2005, 2006, 2007 и 2009 годы образуют однородный кластер, 2008, 2010, 2011, 2012, 2013 годы – второй однородный кластер; третий кластер – периоды 2014 и 2015 годы. Данные результаты хорошо согласуются с результатами факторного анализа. Каждый кластер характеризуется средним показателем по структурным факторам. Самый высокий средний показатель наблюдается у третьего кластера.

Таким образом, в работе средствами факторного анализа проведено сжатие многомерного массива данных. Представлена графическая визуализация данных, построена модель линейной регрессии на латентных факторах, проведена кластеризация временных периодов по структуре экономически занятого населения Самарского региона.

### Библиографический список

1. URL: [http://samarastat.gks.ru/wps/wcm/connect/rosstat\\_ts/samarastat/ru/statistics/employment](http://samarastat.gks.ru/wps/wcm/connect/rosstat_ts/samarastat/ru/statistics/employment).
2. Кремер Н.Ш., Путко Б.А. Эконометрика. М.: Юнити-Дана, 2002. 311 с.
3. Многомерный статистический анализ в экономике / Л.А. Сошникова [и др.]. М.: Юнити-Дана, 1999. 598 с.
4. Наследов А.Д. SPSS 15: профессиональный статистический анализ данных. СПб.: Питер, 2008.
5. Трусова А.Ю., Ильина А.И. Моделирование и анализ динамических данных // Вестник Самарского государственного университета. 2013. № 7 (108) С. 127–133.
6. Трусова А.Ю., Черепанова М.В. Кластерный анализ в социальных исследованиях // Труды X международной ФАМЭТ, 2010 конференции / под ред. О.Ю. Воробьева. Красноярск: КГТЭИ, СФУ, 2011. С. 261–264.

### References

1. Retrieved from: [http://samarastat.gks.ru/wps/wcm/connect/rosstat\\_ts/samarastat/ru/statistics/employment](http://samarastat.gks.ru/wps/wcm/connect/rosstat_ts/samarastat/ru/statistics/employment) [in Russian].
2. Kremer N.Sh., Putko B.A. *Ekonometrika* [Econometrics]. M.: Yuniti-Dana, 2002, 311 p. [in Russian].
3. *Mnogomernyy statisticheskiy analiz v ekonomike*, L.A. Soshnikova [i dr.] [Multidimensional statistical analysis in economics, L.A. Soshnikova [et al.]]. M.: Yuniti-Dana, 1999, 598 p. [in Russian].
4. Nasledov A.D. *SPSS 15: professional'nyy statisticheskiy analiz dannykh* [SPSS 15: professional statistical analysis of data]. SPb.: Piter, 2008 [in Russian].
5. Trusova A.Yu., Il'ina A.I. *Modelirovaniye i analiz dinamicheskikh dannykh* [Modeling and analysis of dynamic data]. In: *Vestnik Samarskogo gosudarstvennogo universiteta* [Vestnik of Samara State University], 2013, no. 7 (108), pp. 127–133 [in Russian].
6. Trusova A.Yu., Cherepanova M.V. *Klasternyy analiz v sotsial'nykh issledovaniyakh* [Cluster analysis in social research]. In: *Trudy X mezhdunarodnoy FАMЕТ, 2010 konferentsii, pod red. O.Yu. Vorob'yeva* [Proceedings of the X International FАMЕТ, 2010 conference, ed. O.Yu. Vorobuyev]. Krasnoyarsk: KGTEI, SFU, 2011, pp. 261–264 [in Russian].

*A.Yu. Trusova\**

### STATISTICS INDICATORS OF LABOR RESOURCES THE FOCUS OF MULTIVARIATE STATISTICAL METHODS

In work the analysis of statistical data structure of the economically employed population of the Samara region by means of multivariate statistical methods.

**Key words:** compression of multidimensional data by means of factor analysis, regression analysis in the space of latent factors, clustering, statistical package SPSS.

Статья поступила в редакцию 12/II/2017.

The article received 12/II/2017.

---

\* *Trusova Alla Yuriyevna* (a\_yu\_ssu@mail.ru), Department of Mathematics and Business Informatics, Samara National Research University, 34, Moskovskoye shosse, Samara, 443086, Russian Federation.