

ПРИМЕНЕНИЕ МЕТОДА КЛАСТЕРНОГО АНАЛИЗА В ОЦЕНКЕ И ПРОГНОЗИРОВАНИИ УРОВНЯ БЕЗРАБОТИЦЫ В РЕГИОНЕ

В статье проведен анализ территориального размещения и оценка ситуации на рынке труда Оренбургской области. Для снижения размерности статистической совокупности выполнен корреляционный анализ, позволивший определить признаки, которые наилучшим образом характеризуют тенденции на рынке труда региона. Рассчитан прогнозный уровень безработицы для каждого кластера на основе регрессионных моделей.

Ключевые слова: уровень безработицы, корреляция, кластерный анализ, уравнение регрессии, прогнозирование, Оренбургская область.

Особенности ситуации на рынке труда в каждом регионе ориентируют на использование набора специфических факторов, изучение влияния которых на уровень безработицы позволяет использовать для ее прогнозирования методы корреляционно-регрессионного анализа, построение комплексной модели и использование факторов, статистически существенно влияющих на величину коэффициента безработицы.

Однако всегда трудно выбрать какой-то один признак в качестве основы группировки. Еще труднее проводить группировку по нескольким признакам. Комбинация двух признаков позволяет сохранить обзорность таблицы, но комбинация более двух признаков дает совершенно неудовлетворительный результат. Сохранить сложность описания групп и вместе с тем преодолеть недостатки комбинационной группировки позволяют структурные методы классификации объектов. Эти многомерные методы объединяются термином «кластерный анализ».

Английское слово «the cluster» имеет значение группа, пучок, то есть объединение каких-то однородных явлений. Кластерный анализ получил распространение благодаря использованию ПЭВМ и пакетов прикладных программ «STATISTICA».[1]

Основной целью кластерного анализа является выделение в исходных многомерных данных однородных групп, таких, чтобы объекты внутри групп были бы близки по выбранной метрике в многомерном признаковом пространстве, а объекты из разных групп были бы удалены друг от друга.

При изучении методов кластерного анализа возникают трудности, связанные с отсутствием формальной точной постановки задачи кластерного анализа. В первую очередь большие сложно-

сти вызывает определение кластера. Кроме того, в пространстве большей размерности возникают еще большие трудности с различными вариантами взаиморасположения объектов и кластеров. Поэтому для проведения классификации необходимо в первую очередь ввести понятие близости объектов в признаковом пространстве.

Большое достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. Кроме того, кластерный анализ в отличие от большинства математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов, и позволяет рассматривать множество исходных данных практически произвольной природы. Это имеет большое значение, например, для прогнозирования ситуации на региональном рынке труда, когда показатели имеют разнообразный вид, затрудняющий применение традиционных эконометрических подходов. [2]

Кластерный анализ позволяет рассматривать достаточно большой объем информации и резко сокращать, сжимать большие массивы социально-экономической информации, делать их компактными и наглядными. Кластерный анализ имеет применительно к совокупностям временных рядов, характеризующих экономическое развитие (например, товарной конъюнктуры). Здесь можно выделять периоды, когда значения соответствующих показателей были достаточно близкими, а также определять группы временных рядов, динамика которых наиболее схожа.

В задачах социально-экономического прогнозирования весьма перспективно сочетание кластерного анализа с другими количественными методами (например, с регрессионным анализом). [1]

Каждая единица совокупности в кластерном анализе рассматривается как точка в заданном признаковом пространстве. Значение каждого из признаков у данной единицы служит ее координатой в этом пространстве по аналогии с координатами точки в нашем реальном трехмерном пространстве. Таким образом, признаковое пространство – это область варьирования всех признаков совокупности изучаемых явлений. Если уподобить это пространство обычному пространству, имеющему евклидову метрику, то тем самым получается возможность измерять «расстояния» между точками признакового пространства. Эти расстояния называют евклидовыми, формула (1): [5]

$$\rho_E(x_i, x_j) = \sqrt{\sum_{l=1}^k (x_{il} - x_{jl})^2} \quad (1)$$

где x_{il}, x_{jl} – величина l -ой компоненты у i -го (j -го) объекта ($l = 1, 2, \dots, k; j = 1, 2, \dots, n$).

Для приведения признаков к одинаковым единицам прибегают к нормировке каждого признака путем деления центрированной величины на среднее квадратическое отклонение (формула (2)):

$$x_{il}^H = \frac{x_{il} - \bar{x}_l}{s_l}, \quad (2)$$

где x_{il} – значение l -го признака у i -го объекта; \bar{x}_l – среднее арифметическое значение l -го признака; $s_l = \sqrt{\frac{1}{n} \sum_i (x_{il} - \bar{x}_l)^2}$ – среднее квадратическое отклонение l -го признака.

Результат процесса классификации – дендрограмма – дерево объединений кластеров с порядковыми номерами объектов на вертикальной оси и шкалой расстояний по горизонтальной оси. Поскольку полученные кластеры по признакам представляют собой однородную совокупность, то представляется возможным получить регрессионную модель. [5]

Для оценки степени различия административных территорий Оренбургской области по ситуации на рынке труда был проведен кластерный анализ по 35 районам Оренбургской области. Города Оренбургской области не включены в кластеры, так как по некоторым признаками отсутствует необходимая информация, что затрудняет процедуру кластеризации.

Прежде чем воспользоваться методом кластерного анализа, было выяснено, какие факторы целесообразно использовать, как основу

многомерной классификации. Более обоснованным методом для выбора факторов является корреляционный анализ, состоящий в отборе факторных признаков X для дальнейшего анализа, позволяющего количественно описать взаимосвязь между признаками.

Уровень безработицы Оренбургской области в 2011 г., в% к среднесписочной численности работников, был взят как основной показатель, характеризующий состояние безработицы региона (Y), факторными признаками (X) выступили следующие:

X_1 – доля безработных, которым назначено пособие по безработице, в общей численности зарегистрированных безработных, %;

X_2 – соотношение численности зарегистрированных безработных и численности ищущих работу граждан, не занятых трудовой деятельностью, %;

X_3 – коэффициент роста численности безработных, которым назначено пособие по безработице, в% к 2010 г.;

X_4 – коэффициент роста среднемесячной номинальной начисленной заработной платы работников в% к 2010 г.;

X_5 – коэффициент напряженности на рынке труда, чел. на одну вакансию;

X_6 – соотношение численности пенсионеров и численности населения в трудоспособном возрасте, на 1000 чел. населения («пенсия нагрузка»);

X_7 – общий коэффициент естественного прироста населения (на 1000 человек населения);

X_8 – общий коэффициент рождаемости (число родившихся на 1000 человек населения);

X_9 – коэффициент миграционного прироста (на 1000 человек населения);

X_{10} – доля убыточных организаций, % от общей численности организаций;

X_{11} – индекс физического объема инвестиций в основной капитал, в% к предыдущему году;

X_{12} – индекс физического объема оборота розничной торговли, в% к предыдущему году;

X_{13} – ввод в действие жилых домов на 1000 (квадратных метров общей площади). [3, 4, 6]

Влияние вышеназванных факторов на уровень безработицы определяется с помощью корреляционной матрицы (табл. 1).

По данным табл. 1, важно отметить, что все факторы ($X_1 - X_{13}$) в разной степени связаны с результативным показателем. Парные коэффи-

циенты корреляции определяют не только между зависимыми и факторными признаками. Большое значение имеют также коэффициенты корреляции, рассчитанные между факторными признаками.

Данные табл. 1 свидетельствуют, о том, что мультиколлинеарность между факторными признаками отсутствует.

Ответ на вопрос – случайны или нет полученные результаты расчетов – можно получить, проверив значимость парных коэффициентов корреляции, с помощью t-критерия Стьюдента (табл. 2). [5]

По данным табл. 2 видно, что наблюдаемые значения t-критерия по модулю больше критического, следовательно, парные коэффициенты корреляции статистически значимы и полученные значения сформировались не случайно. Следовательно, в основу классификации, целе-

сообразно включить следующие признаки:

X_5 – коэффициент напряженности на рынке труда, чел. на одну вакансию;

X_9 – коэффициент миграционного прироста (на 1000 человек населения);

X_{10} – доля убыточных организаций, % от общей численности организаций;

X_{11} – индекс физического объема инвестиций в основной капитал, % к предыдущему году;

X_{12} – индекс физического объема оборота розничной торговли, % к предыдущему году;

X_{13} – ввод в действие жилых домов на 1000 человек населения (квадратных метров общей площади).

Для объединения городов и районов Оренбургской области в кластеры по признакам, указанным выше воспользовались методом Уорда и Евклидовым расстоянием (ППП «STATISTICA 7.0»). [2]

Таблица 1. Матрица парных коэффициентов корреляции для совокупности муниципальных образований Оренбургской области*

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃
Y	1,00													
X ₁	0,14	1,00												
X ₂	0,04	-0,12	1,00											
X ₃	0,11	0,18	0,01	1,00										
X ₄	0,21	-0,20	0,09	0,01	1,00									
X ₅	0,74	0,09	0,09	-0,24	-0,07	1,00								
X ₆	-0,05	0,10	0,09	-0,24	-0,12	1,00	1,00							
X ₇	-0,06	-0,08	0,07	0,11	0,09	0,12	0,12	1,00						
X ₈	0,25	-0,20	0,10	0,09	0,36	-0,15	-0,16	0,24	1,00					
X ₉	-0,68	0,10	0,06	-0,12	-0,56	0,14	0,15	0,14	-0,16	1,00				
X ₁₀	0,58	-0,08	0,05	-0,38	0,06	0,23	0,21	-0,31	-0,23	0,06	1,00			
X ₁₁	-0,68	0,01	0,18	-0,01	0,10	-0,05	-0,06	0,08	0,16	0,11	0,35	1,00		
X ₁₂	-0,66	0,02	0,07	-0,10	-0,20	-0,01	-0,01	0,24	0,06	0,42	0,04	0,06	1,00	
X ₁₃	-0,63	0,24	0,32	0,21	-0,42	-0,23	-0,23	0,07	-0,18	0,59	0,22	0,18	0,26	1,00

*Рассчитано автором по данным [3, 4, 6]

Таблица 2. Проверка значимости парных коэффициентов корреляции по t-критерию

Парные коэффициенты корреляции	Значение r_{yx}	$t_{кр} (\alpha = 0,05; \nu = 35 - 2)$	Модуль $t_{набл}$	Оценка значимости r_{yx}
r_{yx5}	0,741	2,040	2,312	значим
r_{yx9}	-0,683	2,040	2,556	значим
r_{yx10}	0,581	2,040	2,299	значим
r_{yx11}	-0,680	2,040	2,914	значим
r_{yx12}	-0,662	2,040	2,309	значим
r_{yx13}	-0,634	2,040	2,209	значим

Результат объединения – дендрограмма (рис. 1.), по оси ординат, которой отражены города и районы Оренбургской области, а по оси абсцисс показано значение интегрального показателя, представленного величиной, сформированной на основе исследуемых показателей. Данный показатель не имеет единицы измерения, а является многомерной статистической оценкой. В нашем случае – оценка ситуации на рынке труда Оренбургской области.

По результатам многомерной группировки получено 3 кластера (табл. 3), определяющие специфику экономической ситуации на рынке труда в регионе.

Анализируя полученные описательные характеристики (табл. 4) можно отметить, что по средним значениям, районы, попавшие в третий кластер (20,0% от общего числа районов), можно отнести к районам, имеющим благополучную ситуацию на рынке труда. Об этом свидетельствуют низкие значения коэф-

фициента напряженности и доли убыточных организаций.

Для районов третьего кластера данные показатели достаточно высоки, экономическую ситуацию следует считать благополучной.

Значения показателей (табл.4), рассчитанные для районов второго кластера, по величине больше значений показателей третьего кластера и меньше значений показателей первого кластера. Следовательно, ситуацию на рынке труда районов второго кластера следует считать хорошей относительно ситуации на рынке труда районов первого кластера.

Высокие средние значения показателей для первого кластера, позволяют характеризовать ситуацию, сложившуюся на рынке труда районов, попавших в первый кластер, как неблагоприятную. Так, доля убыточных организаций для районов первого кластера достаточно велика, что следует рассматривать как негативную тенденцию для рынка труда этих районов.

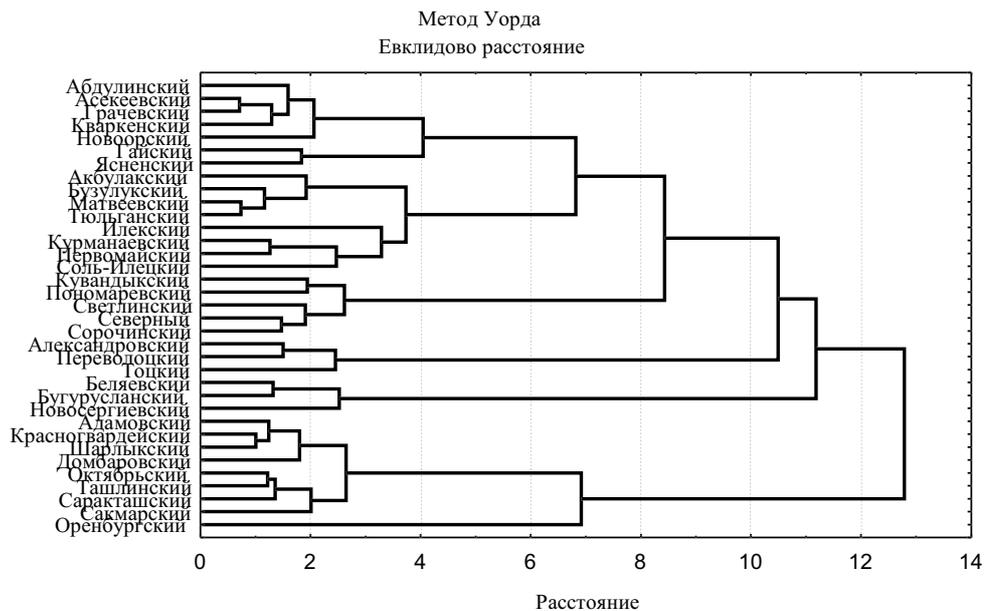


Рисунок 1. Дендрограмма кластеризации методом Уорда

Таблица 3. Результаты кластеризации районов Оренбургской области

№ кластера	Количество районов	Наименование районов
1	20	Абдулинский, Асекеевский, Бузулукский, Гайский, Грачевский, Домбаровский, Кваркенский, Кувандыкский, Курманаевский, Матвеевский, Новоорский, Первомайский, Пономаревский, Светлинский, Северный, Соль-Илецкий, Сорочинский, Тюльганский, Шарлыкский, Ясенский
2	8	Акулацкий, Александровский, Беляевский, Бугурусланский, Илекский, Новосергиевский, Перволоцкий, Тоцкий
3	7	Адамовский, Красногвардейский, Октябрьский, Оренбургский, Сакмарский, Саркатынский, Ташлинский

На основе графического представления (рис. 2) средних значений признаков можно провести аналогичную экономическую интерпретацию для каждого кластера по ситуации на региональном рынке труда.

С помощью метода регрессионного анализа, позволяющего определить аналитическое выражение связи между результативным и факторными признаками, рассмотрено влияние факторов на уровень безработицы в Оренбургской области. Воспользовавшись модулем Multiple Regression (Множественная регрессия) пакета прикладных программ Statistica, получили следующие результаты, представленные в таблице 5.

Вариация результирующего фактора (Y – уровень безработицы, в% к среднесписочной численности) описывается вариацией включенных в модели признаков-факторов на 47,2%, 70,9% и 80,8% соответственно. Степень обусловленности результата факторами – от слабой до сильной.

Оценку надежности уравнения регрессии в целом и коэффициента множественной регрессии R дает F-критерий Фишера, табличные значения F-критерия меньше фактических (табл. 5, столбец 4), отсюда подтверждается значимость полученных для каждого кластера уравнений регрессии и множественного коэффициента детерминации.

Таблица 4. Характеристика районов Оренбургской области по состоянию рынка труда

№ кластера	X ₅	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃
1	9,77	-5,22	43,55	62,91	91,50	190,79
2	8,23	-0,30	39,19	64,45	98,53	210,39
3	7,37	5,74	25,57	67,77	105,11	444,13

*Рассчитано автором по данным [3, 4, 6]

Анализ уравнения регрессии для первого кластера (табл. 5) показал, что с ростом коэффициента напряженности и доли убыточных организаций на одну единицу своего измерения, уровень безработицы в Оренбургской области будет иметь тенденцию к росту.

Таблица 5. Характеристика регрессионных моделей

№ кластера	Модель	Множественный коэффициент детерминации	F-критерий (табличное значение при α=0,05)	Относительная ошибка аппроксимации, %
1	$y_x = 18,5 + 0,3X_5 + 0,2X_{10}$	0,472	10,231 (3,49)	3,06
2	$y_x = 15,6 - 0,7X_9 - 0,2X_{11}$	0,709	8,911 (4,46)	2,75
3	$y_x = 12,5 - 0,5X_{12} - 0,4X_{13}$	0,808	9,123 (4,74)	3,88

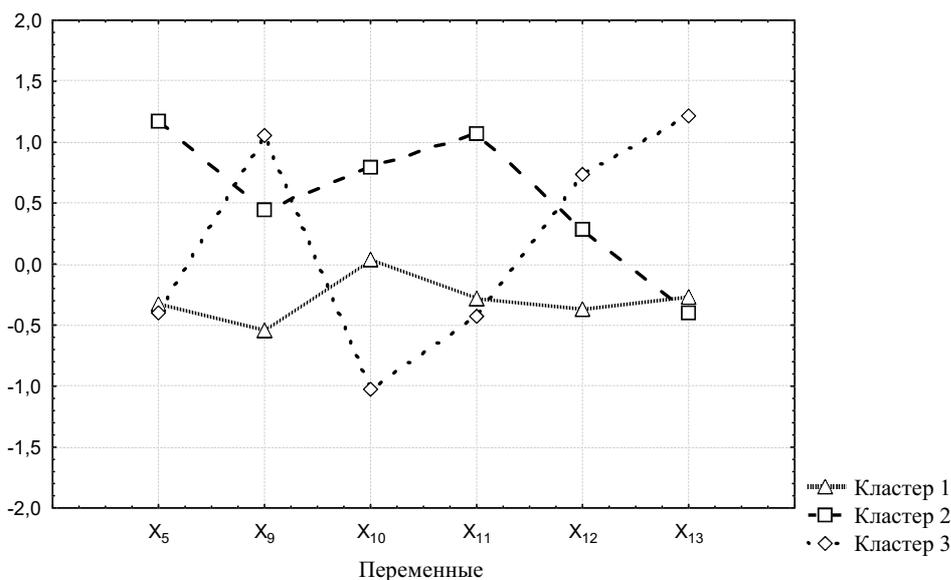


Рисунок 2. График средних значений по кластерам

Рост коэффициентов миграционного прироста и индексов физического объема инвестиций в основной капитал на одну единицу своего измерения для районов второго кластера приведет к снижению уровня безработицы. Вывод вполне закономерен, поскольку инвестиции в основной капитал оказывают значительное влияние на экономику региона, определяют спрос на рабочую силу, поскольку создаются новые рабочие места. Рост инвестиций в основной капитал во многом способствует вовлечению людей в производство, и как следствие, сокращению безработицы. Рост коэффициента миграционного прироста на 1 мигранта на 1000 человек населения приведет к снижению уровня безработицы на 0,2%. (табл. 5)

Аналогичные второму кластеру тенденции наблюдаются и в третьем кластере: для районов, попавших в этот кластер, рост индексов физического объема оборота розничной торговли и ввода в действие жилых домов на 1000 человек населения приведет к снижению уровня безработицы на 0,5% и 0,4% соответственно. (табл. 5)

В таблице 6 представлены три прогнозных варианта для каждого кластера.

При условии, что параметры регрессионных моделей сохраняются, уровень безработицы, % к среднесписочной численности работников, в Оренбургской области может составить:

– в 1-ом кластере при минимальном значении факторов – 23,3%; при среднем значении факторов – 28,3%; при максимальном значении факторов – 35,7%;

Таблица 6. Прогноз уровня безработицы Оренбургской области, % к среднесписочной численности работников

Прогноз	1 кластер	2 кластер	3 кластер
Пессимистичный	23,3	5,6	4,4
Реалистичный	28,3	12,3	11,8
Оптимистичный	35,7	16,1	13,2

– во 2-ом кластере: при минимальном значении факторов – 5,6%; при среднем значении факторов – 12,3%; при максимальном значении факторов – 16,1%;

– в 3-ем кластере при минимальном значении факторов – 4,4%; при среднем значении факторов – 11,8%; при максимальном значении факторов – 13,2%.

Таким образом, деятельность органов управления Оренбургской области должна быть сконцентрирована на улучшении ситуации на регистрируемом рынке труда для районов с неблагоприятной обстановкой на рынке труда (1 кластер), социально-трудового климата для районов с относительно благополучной ситуацией (2 кластер), социально-экономических показателей для районов с благоприятной обстановкой на рынке труда (3 кластер), оказывающих основное влияние на величину уровня безработицы конкретного района области. Положительные тенденции на региональном рынке труда, сформировавшиеся вследствие проведения ряда мероприятий по снижению уровня безработицы, обеспечат повышение уровня благосостояния населения Оренбургской области.

30.01.2012

Список литературы:

1. Башина О.Э., Иванова Н.Ю. Многомерные статистические группировки [Текст] / О.Э. Башина, Н.Ю. Иванова. – М.: Изд-во МГУК, 2001. – с. 24.;
2. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере [Текст] / В.П. Боровиков, Г.И. Ивченко. – М.: Финансы и статистика, 2003;
3. Областной статистический ежегодник [Текст] / Стат.сб. / Территориальный орган Федеральной службы государственной статистики по Оренбургской области. – Оренбург, 2011;
4. Труд и занятость в Оренбургской области. [Текст] / 2005-2011: Стат. сб. / Территориальный орган Федеральной службы государственной статистики по Оренбургской области. – Оренбург. – 2005-2011;
5. Эконометрика [Текст] / Под ред. И.И. Елисейевой – М.: Проспект, 2011;
6. www.orenburg-gov.ru – Портал Правительства Оренбургской области.

Сведения об авторе: **Портнова Лидия Владимировна**, аспирантка кафедры финансов и статистики Российского государственного торгово-экономического университета
E-mail: Lidiyaoren@rambler.ru

UDC 332

Portnova L.V.

THE PRACTICE OF APPLICATION OF THE METHOD OF CLUSTER ANALYSIS IN THE ASSESSMENT AND PREDICTION OF THE LEVEL OF REGIONAL UNEMPLOYMENT

The article deals with the analysis of unemployment, its evaluating and territory locating. The component analysis has been made to decrease statistical aggregate weighting. It helps much defining the particular combinations of clusters presenting the regional labour market situation in full. Special attention is paid to forecasting the unemployment level for each cluster based on regression models.

Key words: unemployment level, regression equation, cluster analysis, impact analysis, main components analysis, forecasting.