



## МЕТОДИЧЕСКИЕ ПОДХОДЫ К СТРАТЕГИЧЕСКОМУ ПЛАНИРОВАНИЮ РАЗМЕЩЕНИЯ МАСЛИЧНЫХ КУЛЬТУР МЕТОДОМ КЛАСТЕРНОГО АНАЛИЗА

**Н.Н. КРАВЧЕНКО**  
**Т.В. САВЧЕНКО**

*Алексеевский филиал  
Белгородского  
государственного  
университета*

*e-mail:  
Diana080904@Yandex.ru*

В статье освещены методические подходы к стратегическому планированию размещения масличных культур методом кластерного анализа. Рассматривается зонирование территории Белгородской области по уровню эффективности производства маслосемян подсолнечника. В статье используется агломеративный метод иерархической классификации или иерархический кластерный анализ. Полученный в результате классификации график предоставляет возможность выявить кластеры и интерпретировать их.

Ключевые слова: подсолнечник, масличные культуры, посевные площади, экономические показатели производства подсолнечника, кластерный анализ, кластер.

Развитие агропромышленного комплекса России обуславливает необходимость концентрации производства сельскохозяйственного сырья и его переработки в тех территориальных зонах, где имеются организационно-экономические и природно-климатические условия, обеспечивающие наибольшую его эффективность.

В Белгородской области выделяют три зоны: 1 – лесостепная западная зона; 2 – лесостепная центральная зона; 3 – степная юго-восточная зона. В основу данного зонирования положены различия в почвенно-климатических условиях и территориальном расположении административных районов области. Однако за последние двадцать лет в области сформировалась определенная региональная структура экономики, и, как показал анализ размещения производства и переработки масличных культур, фактическая специализация районов не вписывается в указанные выше зоны.

В настоящем исследовании нами было проведено зонирование территории области по уровню эффективности производства маслосемян подсолнечника методом кластерного анализа с использованием средних показателей за период 2006–2008 гг., отражающих потенциальные факторы размещения их производства и, следовательно, специализацию районов.

Кластерный анализ – это совокупность методов разделения множества элементов (объектов), заданных определенными параметрами своих признаков, на однородные группы.

Большое преимущество кластерного анализа состоит в том, что он позволяет производить группировку 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 [1].

В ходе исследования нами был выдвинут следующий ряд гипотез.

1. Типизация объектов (районов области) по концентрации производства семян подсолнечника (площадь, валовой сбор, доля посевов в общей площади пашни, уровень развития отрасли и др.) объединит объекты по степени специализации и концентрации производства.

2. Кластеры по показателям результативности производства подсолнечника (урожайность, себестоимость, рентабельность и др.) объединят объекты по уровню эффективности данной отрасли.



3. При выявлении нескольких кластеров появится возможность определить приоритетные стратегические направления развития для каждого из них с учетом расположения на территории предприятий по переработке масличных культур.

Сформулируем задачу кластерного анализа. В нашем распоряжении имеется 21 объект (по количеству анализируемых районов Белгородской области) в 17-мерном пространстве  $(x_1, \dots, x_{17})$ , где:

- $X_1$  – валовой сбор подсолнечника, ц;
- $X_2$  – площадь посева подсолнечника, га;
- $X_3$  – урожайность подсолнечника, ц/га;
- $X_4$  – затраты на 1 га, руб.;
- $X_5$  – себестоимость 1 ц, руб.;
- $X_6$  – себестоимость 1 ц реализованной продукции, руб.;
- $X_7$  – цена реализации 1 ц, руб.;
- $X_8$  – прибыль от реализации 1 ц, руб.;
- $X_9$  – уровень рентабельности реализованной продукции, %;
- $X_{10}$  – прибыль (убыток) с 1 га, руб.;
- $X_{11}$  – прибыль от реализации всей продукции, руб.;
- $X_{12}$  – площадь пашни (всего), га;
- $X_{13}$  – доля посевов подсолнечника в общей площади пашни;
- $X_{14}$  – внесено минеральных удобрений на 1 га посева, кг;
- $X_{15}$  – внесено органических удобрений на 1 га посева, т;
- $X_{16}$  – количество тракторов на 100 га пашни, ед.;
- $X_{17}$  – уровень развития отрасли по индексу объемов производства.

Последний показатель был определен нами как отношение базового индекса физического объема производства подсолнечника к среднему индексу физического объема производства:

$$УРО = \frac{УФО(б)}{УФО(с)}, \quad (1)$$

где  $УРО$  – уровень развития отрасли;  $УФО(б)$  – базовый индекс физического объема производства, определенный как отношение валового сбора подсолнечника в 2008 году к 2005 году;  $УФО(с)$  – средний индекс физического объема производства, определенный как отношение валового сбора подсолнечника в среднем за 2006 – 2008 гг. к 2005 году.

Необходимо эффективно разделить эти объекты на ряд групп таким образом, чтобы:

- внутри группы объекты были максимально схожи между собой;
- группы максимально между собой различались.

Если такое разделение осуществимо, то, скорее всего, ввиду однородности внутри каждого кластера зависимость  $z = f(x_1, \dots, x_{17})$  будет представлять собой непрерывную функцию. Иначе говоря, в рассматриваемом пространстве  $(x_1, \dots, x_{17})$  требуется ввести количественную меру сходства между объектами. В нашем случае кажется удобным введение метрики «евклидово расстояние», которая вычисляется как корень из суммы квадратов по координатных разностей:

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

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

Однако введение метрики – это еще не все. Часто бывает так, что различные независимые переменные измеряются в разных шкалах с различными диапазонами, что влечет за собой необходимость проведения процедуры стандартизации переменных. Для этого необходимо вычислить среднее значение и дисперсию. Воспользуемся методом анализа данных «Описательная статистика» в редакторе электронных таблиц Microsoft Excel.



Результатом стандартизации является приведение всех переменных к единой шкале: данные изменяются в пределах от нуля до  $\pm 3$ , причем большая часть всех значений будет принадлежать интервалу  $(-1, 1)$ . Очень важно, что процедура стандартизации не изменяет структуру взаимодействий между переменными. Соответственно, стандартизация не влияет на структуру кластеров. Полученные в результате стандартизации переменные имеют нулевое среднее и единичную дисперсию.

В целом методы кластеризации делятся на агломеративные (от слова агломерат – скопление) и итеративные дивизивные (от слова division – деление, разделение).

Для определения числа кластеров будем использовать агломеративный метод иерархической классификации или иерархический кластерный анализ.

Сущность данного метода заключается в том, что на первом шаге каждый объект выборки рассматривается как отдельный кластер. Процесс объединения происходит последовательно: на основании матрицы расстояний объединяются наиболее близкие объекты. Если матрица сходства первоначально имеет размерность « $m \times m$ », то полностью процесс кластеризации завершается за  $m-1$  шагов, в итоге все объекты будут объединены в один кластер. Последовательность объединения легко поддается геометрической интерпретации и может быть представлена в виде графа – дерева (дендрограммы). На оси абсцисс дендрограммы откладываются имена наблюдений, а по оси ординат – расстояния объединения наблюдений в кластеры. Соответственно, чем выше расположена ветвь дерева на дендрограмме, тем позднее было проведено объединение объектов.

Полученный в результате иерархической кластеризации график (вертикальная дендрограмма) предоставляет возможность обнаружить кластеры (ветви) и интерпретировать их (рис. 1).

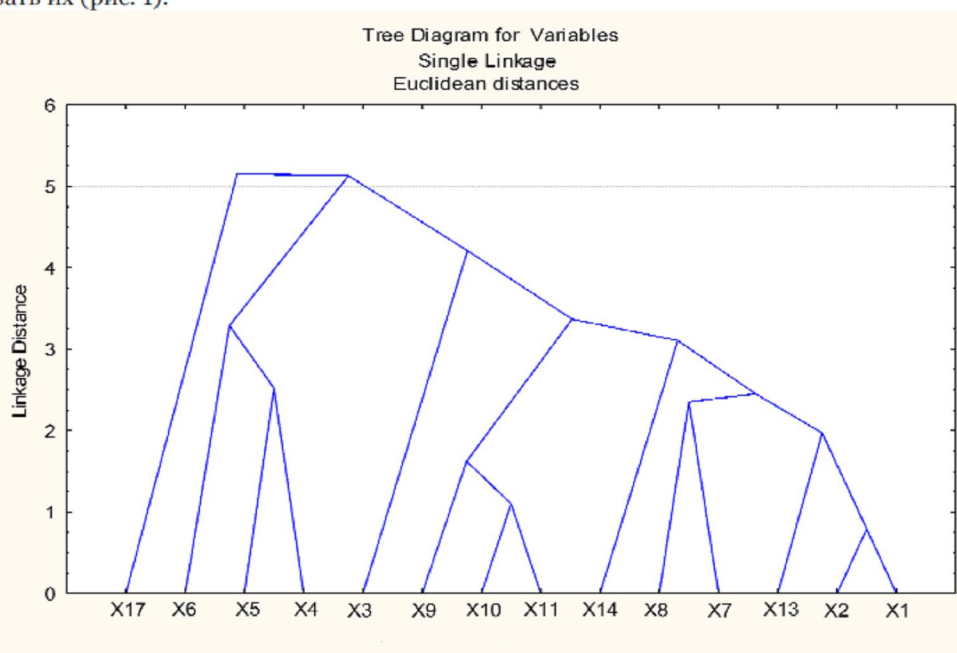


Рис. 1. Вертикальная дендрограмма древовидной классификации для исследуемых переменных

Древовидная диаграмма отображает историю объединения объектов в кластеры. Чем выше ветви дерева-графа, тем позднее объекты были объединены. На представленном выше рисунке, выделяются три ветви дерева, объединенные на разной высоте. Каж-



дая из этих ветвей имеет продолжение в виде скоплений ветвей гораздо меньшей высоты, еще называемых кучностями.

Объединение начинается с группы переменных: валовой сбор ( $X_1$ ); площадь ( $X_2$ ); цена реализации ( $X_7$ ); прибыль от реализации 1 ц ( $X_8$ ); всего площади пашни ( $X_{14}$ ) и доля посевов подсолнечника в общей площади пашни ( $X_{13}$ ).

Далее – вторая группа переменных: урожайность ( $X_3$ ); получено прибыли от реализации ( $X_{11}$ ); уровень рентабельности реализованной продукции ( $X_9$ ); прибыль (убыток) с 1 га ( $X_{10}$ ); внесено органических удобрений на 1 га ( $X_{15}$ ); количество тракторов на 100 га пашни ( $X_{16}$ ).

Завершает объединение третья группа: затраты на 1 га ( $X_4$ ), себестоимость 1 ц ( $X_5$ ), себестоимость 1 ц реализованной продукции ( $X_6$ ), внесено минеральных удобрений на 1 га ( $X_{14}$ ), уровень развития отрасли по индексу объемов производства ( $X_{17}$ ).

На вопрос о количестве кластеров мы ответили. Их три. Теперь необходимо выяснить, какой объект к какому кластеру принадлежит. Для этого воспользуемся итеративной процедурой, методом К-средних. Сущность метода заключается в том, что процесс классификации начинается с задания начальных условий. В нашем случае – это количество образуемых кластеров и центры этих кластеров. Далее, каждое многомерное наблюдение совокупности относится к тому кластеру, центр которого ближе всех к этому наблюдению. Затем выполняется проверка на устойчивость классификации. Если классификация устойчива, процесс останавливается. В противном случае происходит очередная процедура разбиения объектов по кластерам.

Процесс дальнейшей кластеризации будет проходить в два этапа:

- кластеризация переменных;
- кластеризация наблюдений.

На каждом этапе необходимо последовательно дополнительно выполнить три шага: произвести анализ дисперсии; определить среднее значение в кластерах и евклидово расстояние; произвести описательную статистику.

В результате проведенной кластеризации получим подтвержденное иерархической кластеризацией разделение переменных на три кластера.

В первый кластер вошли 6 переменных: валовой сбор ( $X_1$ ), площадь ( $X_2$ ), цена реализации ( $X_7$ ), прибыль от реализации ( $X_{11}$ ), всего площади пашни ( $X_{12}$ ) и доля посевов подсолнечника в общей площади пашни ( $X_{13}$ ).

Во второй кластер вошли тоже 6 переменных: урожайность ( $X_3$ ), прибыль от реализации 1 ц ( $X_8$ ), уровень рентабельности реализованной продукции ( $X_9$ ), прибыль (убыток) с 1 га ( $X_{10}$ ), количество органических удобрений на 1 га ( $X_{15}$ ), количество тракторов на 100 га пашни ( $X_{16}$ ).

В третий кластер вошли 5 переменных: затраты на 1 га ( $X_4$ ), себестоимость 1 ц ( $X_5$ ), себестоимость 1 ц реализованной продукции ( $X_6$ ), количество минеральных удобрений на 1 га ( $X_{14}$ ), уровень развития отрасли по индексу объемов производства.

На втором этапе проведем кластеризацию наблюдений (районов Белгородской области) для переменных: валовой сбор ( $X_1$ ), площадь ( $X_2$ ), цена реализации ( $X_7$ ), прибыль от реализации 1 ц ( $X_8$ ), площадь пашни ( $X_{12}$ ) и доля посевов подсолнечника в общей площади пашни ( $X_{13}$ ). Построенная на втором этапе вертикальная дендрограмма для районов Белгородской области будет выглядеть следующим образом.

Выполним кластеризацию наблюдений методом К-средних на стандартизованных данных. В результате получим разделение наблюдений на три кластера.

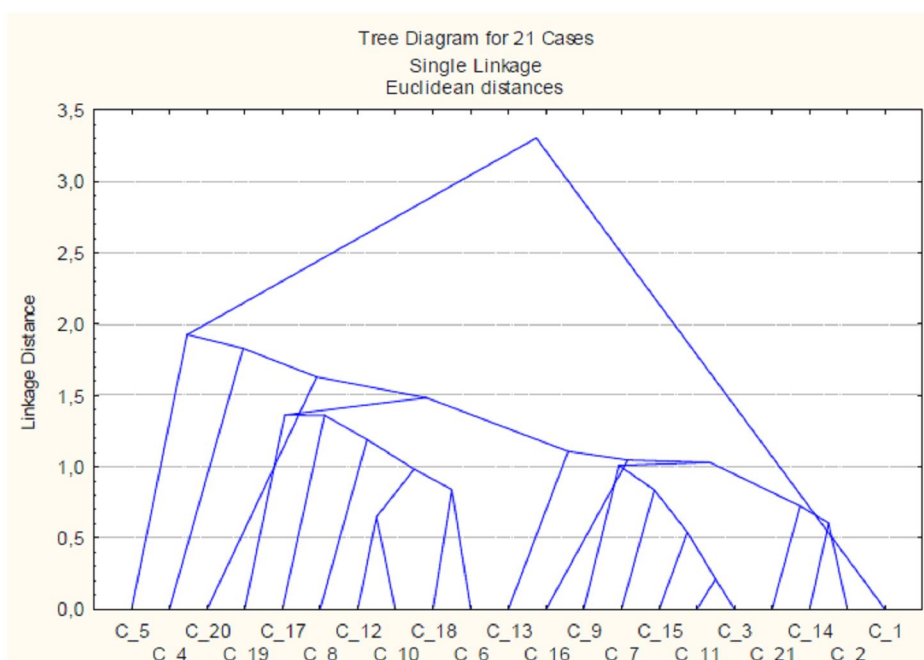


Рис. 2. Вертикальная дендрограмма древовидной классификации для исследуемых наблюдений

В первый кластер вошли 10 наблюдений (объектов): Алексеевский (C<sub>1</sub>), Валуйский (C<sub>4</sub>), Вейделевский (C<sub>5</sub>), Волоконовский (C<sub>6</sub>), Губкинский (C<sub>8</sub>), Корочанский (C<sub>10</sub>), Красногвардейский (C<sub>12</sub>), Ровенской (C<sub>17</sub>), Чернянский (C<sub>19</sub>), Шебекинский (C<sub>20</sub>) районы.

Во второй кластер вошли 6 наблюдений (объектов): Борисовский (C<sub>3</sub>), Красненский (C<sub>11</sub>), Прохоровский (C<sub>15</sub>), Ракитянский (C<sub>16</sub>) районы.

В третий кластер вошли 5 наблюдений (объектов): Белгородский (C<sub>2</sub>), Грайворонский (C<sub>7</sub>), Ивнянский (C<sub>9</sub>), Краснояружский (C<sub>13</sub>), Новооскольский (C<sub>14</sub>) районы.

В таблице и на рис. 3 приведено окончательное распределение районов области по соответствующим кластерам и выведены средние кластерообразующие значения.

Таблица

**Распределение районов по кластерам и среднее значение кластерообразующих**

№ кластера	Ранг	Районы области	Валовой сбор, ц	Площадь, га	Всего площадь пашни, га	Доля посевов подсолнечника в общей площади пашни, %	Цена реализации 1 ц, руб.	Прибыль от реализации 1 ц, руб.
1	2	3	4	5	6	7	8	9
1	1	Алексеевский	168126	8390	79999	10,5	1390,4	574,7
1	4	Валуйский	139658	8178	63100	13,0	810,5	237,8
1	5	Вейделевский	118133	6762	74820	9,0	1041,6	460,8
1	6	Волоконовский	87043	4196	69943	6,0	857,9	349,7
1	8	Губкинский	74472	4332	79110	5,5	744,1	202,9



Окончание табл.

1	2	3	4	5	6	7	8	9
1	10	Корочанский	83137	5022	61855	8,1	810,4	110,3
1	12	Красногвардейский	77770	4648	66446	7,0	873,6	138,3
1	17	Ровеньский	101366	6305	71718	8,8	798,4	333,1
1	19	Чернянский	97962	5198	48880	10,6	847,6	253,9
1	20	Шебекинский	150899	7530	91251	8,3	772,7	241,1
		Среднее по кластеру	109857	6056	70712	8,56	924,8	313,85
2	3	Борисовский	47006	2535	37597	6,7	803,9	260,3
2	11	Красненский	40905	2666	37011	7,2	807,5	286,7
2	15	Прохоровский	39816	2769	45078	6,1	795,7	189,2
2	16	Ракитянский	17523	688	50405	1,4	870,9	533,1
2	18	Старооскольский	81661	3880	57760	6,7	808,7	312
2	21	Яковлевский	36472	2011	54562	3,7	824,3	352
		Среднее по кластеру	43897	2425	47069	5,15	812,4	300,53
3	2	Белгородский	38443	2024	65497	3,1	755,9	127,5
3	7	Грайворонский	30444	1790	40927	4,4	760,7	69
3	9	Ивнянский	15569	836	27368	3,1	756,9	240,2
3	13	Краснояржский	11191	509	23310	2,2	605,4	-33,8
3	14	Новооскольский	25883	1792	61086	2,9	821,2	43,5
		В среднем по кластеру	24306	1390	43638	3,18	757,3	94,6
		В среднем по области	70642	3905	57511	6,8	891,1	293,5



Рис. 3. Распределение районов Белгородской области по уровню производства подсолнечника

Таким образом, наибольший интерес, с точки зрения развития масложирового подкомплекса Белгородской области, представляет первый кластер, куда входит десять административных районов. По данному кластеру средняя площадь посевов подсолнечника в расчете на один район больше в 2,5 раза, чем во втором, и в 4,4 раза – чем в третьем кластере. Доля посевов подсолнечника в общей площади пашни здесь составляет



8,6%, а валовой сбор подсолнечника составляет 74% от всего сбора данной культуры по области. При этом именно в этой группе районов сосредоточены три крупных завода по переработке масличных культур (Чернянский, Алексеевский и Валуйский), на долю которых приходится 98,8% производимого в области подсолнечного масла.

Районы, вошедшие в данный кластер, должны сосредоточить внимание на интенсификации производства масличных культур без расширения посевных площадей под ними путем внедрения инноваций в области производства масличных культур, минимизации издержек при выращивании и доработке маслосемян, повышения качества сырья. При этом рост сырьевого обеспечения должен быть сбалансирован с учетом имеющихся в регионе перерабатывающих мощностей.

#### Литература

1. Дубов А.М. Многомерные статистические методы / А.М. Дубов, В.С. Мхитарян, Л.И. Трошин – М.: Финансы и статистика, 2003. – 352 с.
2. Мхитарян В.С. Эконометрика: учеб.-метод. комплекс / В.С. Мхитарян, М.Ю. Архипова, В.П. Сиротин – М.: Изд. центр ЕАОИ, 2008. – 144 с.

### **METHODICAL APPROACHES TO STRATEGIC PLANNING OF OIL-YIELDING CROPS PLACING BY MEANS OF CLUSTER ANALYSIS**

**H.H. KRAVCHENKO  
T.V. SAVCHENKO**

*Alexeyevka Branch  
of Belgorod State University*

*e-mail:  
Diana080904@Yandex.ru*

The article deals with methodical approaches to strategic planning of oil-yielding crops placing by means of cluster analysis. The author examines Belgorod region area zoning according to sunflower seeds production efficiency level. The article contains agglomerative hierarchical algorithms or cluster analysis. The resulting diagram gives an opportunity to discover and interpret clusters.

Key words: sunflower, oil-yielding crops, sown areas, economic indices of sunflower production, cluster analysis, cluster.