

Применение кластерного анализа для идентификации масличного сырья

Клейменова Наталья Леонидовна

ФГБОУ ВО «Воронежский государственный университет
инженерных технологий»

Адрес: 394036, г. Воронеж, проспект Революции, д. 19

E-mail: klesha78@list.ru

Болгова Инэсса Николаевна

ФГБОУ ВО «Воронежский государственный
университет инженерных технологий»

Адрес: 394036, г. Воронеж, проспект Революции, д. 19

E-mail: bolgovainessa@yandex.ru

Копылов Максим Васильевич

ФГБОУ ВО «Воронежский государственный
университет инженерных технологий»

Адрес: 394036, г. Воронеж, проспект Революции, д. 19

E-mail: kopylov-maks@yandex.ru

Пегина Алла Николаевна

ФГБОУ ВО «Воронежский государственный университет
инженерных технологий»

Адрес: 394036, г. Воронеж, проспект Революции, д. 9

E-mail: toriss@yandex.ru

Кашолкина Дарья Андреевна

ФГБОУ ВО «Воронежский государственный университет
инженерных технологий»

Адрес: 394036, г. Воронеж, проспект Революции, д. 19

E-mail: toriss@yandex.ru

В данной статье осуществлен поиск математических методов для оценки уровня сбалансированности основных количественных признаков у масличных сортов. На основе экспериментальных данных применения кластерного анализа для оценки сортов различных масличных культур проанализированы основные количественные признаки и урожайность. С помощью кластерного анализа проведена идентификация различных сортов масличных культур, которые выращивают в разных регионах страны. Изучены основные посевные площади масличных культур за 2017-2019 годы. В результате проведенных исследований установлено сокращение площадей по культурам льна, горчицы и рыжика, что связано с урожайностью и малым экспортом. Проанализированы состояние роста и развития масличных культур, продолжительность их возделывания, зависящая от биологических особенностей сорта, а также технологии выращивания. Изучена специфика и особенности исследуемых масличных культур. С помощью метода выявлены единые меры, которые включают ряд исследуемых признаков. Данные признаки объединены с помощью метрики в один кластер сходства группирования исследуемых объектов. Сорта сгруппированы в три кластера по годам. Состав каждого кластера менялся в зависимости от природно-климатических условий. Осуществлен разведочный анализ с использованием точечной диаграммы. Построена дендограмма сравнительной характеристики сортов масличных культур по элементам структуры урожая. При помощи метода K-means clustering кластеризации определены евклидовы расстояния исследуемых параметров средних значений соответствующих 5 кластеров, которые имеют схожие сорта масличных культур. На основании проведенных исследований можно определять конкурентоспособность различных масличных культур, разрабатывать новые пищевые продукты.

Ключевые слова: кластер, кластерный анализ, масличные культуры, статистика, дендограмма, урожай

Введение

В настоящее время наблюдаются позитивные изменения развития агропромышленного комплекса, которые позволяют использовать передовые технологии для выращивания масличных культур в различных климатических условиях (El-haak, Atta, & Abd Rabo, 2015; Lobell & Gourdji, 2017).

Авторами (Тищенко, Панченко, & Чернышева, 2013) применялись выборки генотипов и их структурных элементов для пшеницы озимой и на основе Евклидовой метрики получена классификация сорта и определена степень сбалансированности показателей для урожая. Кластерный анализ применялся для оценки сортов мягкой яровой пшеницы по элементам структуры урожая в статье (Шаманин, Петуховский, & Краснова, 2016).

Получены результаты кластерного анализа, в которых разделена совокупность сортов яровой мягкой пшеницы на 5 групп, имеющих разное сочетание параметров длины корней с мощностью их развития как в контроле, так и в присутствии алюминия (Лисицын & Амунова 2014). Авторы предполагают, что повышение алюмоустойчивости сортов яровой мягкой пшеницы происходит косвенным путем, параллельно с целенаправленной селекцией на повышение уровня устойчивости растений к абиотическим факторам среды выращивания.

Авторами (Gebeyehu, Hammenhag, Ortiz, Tesfaye, & Geleta, 2020) проведено исследование по выявлению наличия значительных фенотипических и генотипических различий в генофонде эфиопской нуги, отраженных как на качественных, так и на количественных признаках. Кластерный анализ на основе евклидовых расстояний показал, что схема кластеризации образцов плохо коррелирует с географическим расстоянием между местами сбора проб. Отсутствие четкой модели кластеризации присоединений с учетом их географической близости предполагает сильный поток генов между популяциями в широком географическом регионе - Эфиопии.

Исследования авторов (Румянцев & Глуховцев, 2011; Бушнев, Подлесный, & Хатит, 2019; Котлярова, & Титовская, 2018), позволили предложить научную концепцию селекции зерновых культур по устойчивости к стрессовым факторам, что способствовало получению высокого урожая и высокопластичных сортов, которые имеют свойство противостоять экстремальным условиям в период вегетации.

Однако в природе имеются масличные культуры, которые имеют сбалансированный состав признаков, что обеспечивает им выживание в постоянно изменяющихся климатических условиях (Hussain et al., 2017; Hassan et al., 2019). Выращивание гибридов не способствует идеальному составу в генотипе масличного сырья (Тищенко, Ищенко, & Дубенец, 2016; Creissen, Jorgensen, & Brown, 2016; Bhattarai, Thapa, Ojha, Kharel, & Sapkota, 2017).

Известно, что лен и горчица выращиваются для экспорта, а из рыжика в перспективе можно получать масло и шрот (Patial, Paul, Sharma, Sood, & Kumar, 2019; Jankowski, Załuski, & Sokólski, 2020; Krzyżaniak et al., 2019).

Для успешного выращивания богатого урожая необходимо знать состояние роста и развития масличных культур, продолжительность возделывания, зависящее от биологических особенностей сорта, а также технологии выращивания (Abdorreza, Hossein, & Morteza, 2015; Embaye et al., 2018; Cuesta-Marcos et al., 2016; Agahi, Ahmadi, Oghan, Fotokian, & Orang, 2020; Majerova & Nevima, 2017; Mustafa, Farooq, Hasan, Bibim, & Mahmood, 2015; Khalid & Hameed, 2021; Venujayakanth, Dudhat, Swaminathan, & Anurag, 2017). В данном контексте важным элементом является изучение специфики и особенностей масличных культур.

В современных условиях развития сельского хозяйства для качества продукции актуальной проблемой является совершенствование модернизации оборудования на основе развития инновационных процессов, при этом знания в области оценки технологических приёмов выращивания сельскохозяйственных культур необходимы для повышения конкурентоспособности. Анализ агроморфологических признаков в зависимости от урожайности семян важен для повышения продуктивности масличной культуры. Поэтому важным аспектом является получение высокого урожая при наименьших затратах. С этой целью создаются новые перспективные удобрения, применяют высокопроизводительные агрегаты для обработки почвы и т.п.

Исследование направлено на поиск математических методов, с помощью которых можно оценить уровень сбалансированности основных количественных признаков у масличных сортов.

Цель исследования – провести анализ на основе экспериментальных данных применения кластерного анализа для оценки сортов различных

масличных культур по сбалансированности основных количественных признаков и урожайности.

Задача исследования – применить кластерный анализ для идентификации различных сортов масличных культур, произрастающих в разных регионах и установить сбалансированность количественных признаков, с помощью которых определяют адаптивные свойства, способствующие увеличению урожайности.

Материалы и методы

Материалы

Масличные культуры, используемые в кластерном анализе – рапс (яровой), подсолнечник, рыжик (яровой и озимый), рапсовый (пятнистая), лен (масличный), горчица (белая).

Методы

Цель кластерного анализа заключается в построении групп (или классов, или кластеров), обеспечивая при этом следующее свойство: внутри группы наблюдения должны быть как можно более похожими, в то время как наблюдения, принадлежащие к разным группам, должны быть как можно более разными.

Процедура исследования

В кластерном анализе в основу группировки было включено 5 признаков: урожайность зерна, продуктивная кустистость, число зерен в колосе, масса 1000 семян, масса зерна с колоса. На основе 5 признаков, которые характеризуют сорта исследуемых масличных культур, использовалась иерархическая кластеризация для поиска иерархии кластеров, которая представляет древовидную структуру – дендрограмма. На основании кластеризации k-средних необходимо вычислить расстояния, чтобы каждый кластер был связан с центроидом. Цель применения k-средних состоит в минимизации суммы расстояний между точками и их центроидом кластера.

Анализ данных

Анализ проводили, используя 54 сорта с помощью математического метода – кластерный анализ в программе Statistica.

Результаты и их обсуждение

Анализ литературных данных показал, что в России самой распространенной культурой является подсолнечник из всех исследуемых объектов. Урожайность масличных культур за 2017–2019 – года представлена на Рисунке 1.

Согласно Росстату¹, в России за 2019 год различные масличные культуры занимают 19 % основных посевных площадей. Из исследуемых масличных культур: пшеница – 15,3 %, подсолнечник – 10,7 %, ячмень – 10,2 %; рапс – 1,9 %, лен масличный – 1 %, горчица – 0,5%, рыжик – 0,1 %. Однако наблюдалось сокращение площадей по культурам льна, горчицы и рыжика, что связано с урожайностью и малым экспортом. Данные агрокультуры занимают малую долю для производства пищевых растительных масел.

При проведении исследования возникла проблема объединения по сходству объектов, которые характеризуются множеством признаков, выраженных в разных единицах измерения. Поэтому воспользовались математическим методом – кластерным анализом. С помощью метода выявлены единые меры, которые включают ряд исследуемых признаков. Данные признаки объединяются с помощью метрики в один кластер сходства группирования исследуемых объектов. Исследованию подверглись 54 сорта масличных культур, значения параметров взяты преимущественно с видов, высаженных в Центральной части России^{2,3}. Формирование кластеров основываются на исследовании сортов и выявлении значений элементов структуры урожая в среднем за 2017–2019 годы⁴. Для данного исследования сгруппировали сорта в три кластера по годам. Состав каждого кластера меняется в зависимости от природно-климатических факторов. В программе Statistica сформировали данные с анализируемыми параметрами (Таблица 1).

¹ Валовые сборы и урожайность сельскохозяйственных культур в Российской Федерации в 2019 году. URL: <https://rosstat.gov.ru> (дата обращения: 13.04.2021).

² Итоги сортоиспытания сельскохозяйственных культур. URL: <https://agro.amurobl.ru> (дата обращения: 20.05.2021).

³ Сорта растений, включенные в Государственный реестр селекционных достижений, допущенных к использованию. URL: <https://web.archive.org/web/20171101010903/http://reestr.gossort.com/reestr/1/15> (дата обращения: 10.04.2021).

⁴ Результаты сортоиспытания сельскохозяйственных культур на госсортоучастках Кировской области за 2017–2019 годы и сортовое районирование на 2020 год. Рекомендации составлены в помощь специалистам сельскохозяйственного производства. URL: <https://docviewer.yandex.ru> (дата обращения: 10.04.2021).

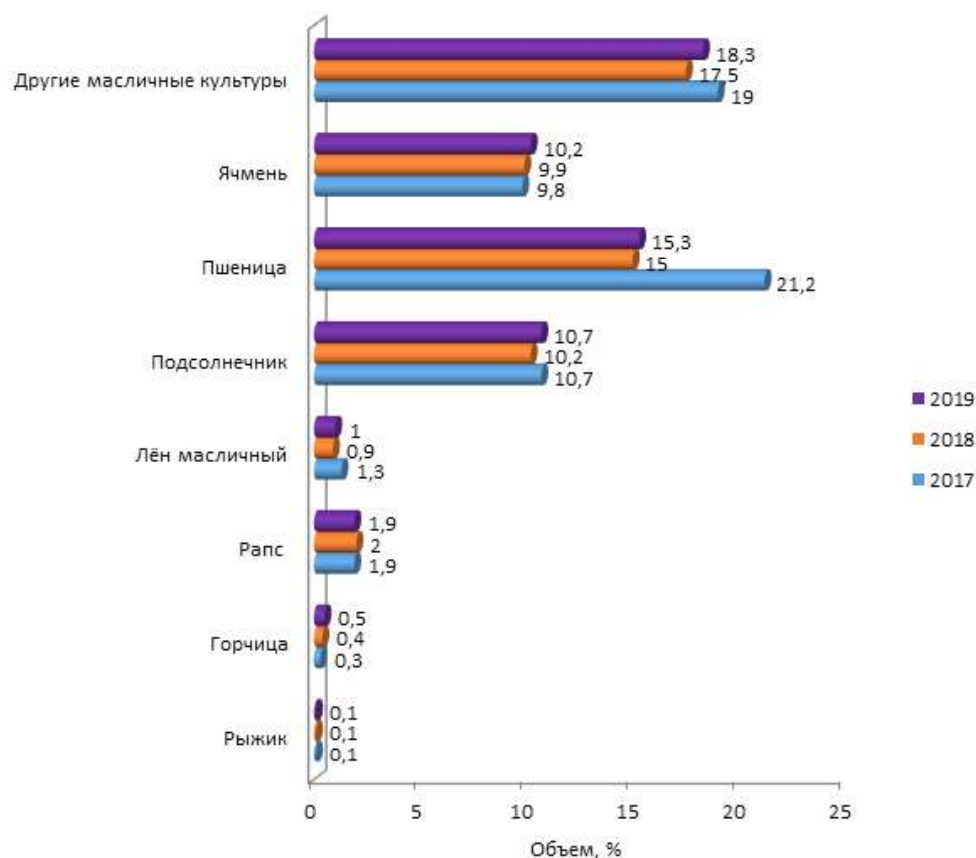


Рисунок 1. Посевные площади масличных культур за 2017-19 годы

Следующим этапом работы являлось проведение разведочного анализа с использованием точечной диаграммы (*Scatterplots*).

Анализ точечной диаграммы (Рисунок 2) показал, что по четырем исследуемым признакам анализируемые масличные культуры разбиваются на 5 кластеров, выделение кластеров необходимо для дальнейшего исследования анализа. При построении дендограммы использовалась Евклидова метрика и метод единичной связи (Мулянова & Косников, 2018). Сравнительная характеристика сортов масличных культур по элементам структуры урожая представлена на Рисунке 3.

Для того чтобы наглядно увидеть разбиение на кластеры используем изменение масштаба (Рисунок 4).

Анализ Рисунка 4 показал, что первый кластер представлен горчицей разных сортов, представителем второго кластера является подсолнечник, третьего – разные сорта льна и рыжика, четвертого – рапсороща, пятого – рыжик «Омич» и все сорта рапса. Известно, что различные сорта рапса

имеют невысокую зимостойкость и поэтому выращивают в регионах с мягким климатом. Рыжик «Омич» устойчив к климату Сибири, количество Омга 9 больше чем в других сортах рыжика.

Следующим этапом необходимо применить *K-means clustering* кластеризации (Venujayakanth et al., 2017). С помощью разведочного анализа построено 5 кластеров по результатам и графики средних значений для выбранных кластеров по элементам структуры урожая масличного сырья (Рисунок 5).

На Рисунке 5 получены средние переменные для пяти кластеров: урожайность, продуктивность куста, число зерен в колосе, масса 1000 зерен и масса зерна с колоса. Урожайность, масса 1000 зерен и с колоса незначительно влияют на классы, следовательно, можно игнорировать (исключить). Число зерен в колосе зависит от фаз развития, внесения питательных веществ, от погодных условий.

С помощью раздела описательной статистики для каждого класса определяется принадлежность исследуемых объектов к каждому классу (Рисунок 6).

Таблица 1

Формирование кластеров при исследовании различных масличных сортов

Номер кластера	Сорт	Урожайность, т/га	Продуктивность куста	Число зёрен в колосе, шт.	Масса 1000 зёрен, г	Масса зерна с колоса, г.
1 (2017)	Рапс Яровой "Амулет"	2,3	1,2	22	3,7	8,03
	Рапс Озимый "Лорис"	3,5	1,5	24	3,9	8,85
	Рапс "Галант"	2,1	1,1	26	3,1	8,1
	Подсолнечник "Бузулук"	3,1	1,3	417,6	55	33,3
	Подсолнечник "Джинн"	3,2	1,2	440	56	35,4
	Подсолнечник "Умник"	3,3	1,1	412,2	58	30,9
	Рыжик "Карат"	2,1	0,99	98,2	0,9	0,46
	Рыжик "Омич"	2	1,1	101,3	1,3	0,48
	Рыжик "Пензяк"	2,3	1,3	105,4	0,9	0,49
	Рапсоропша "Пятнистая"	1,6	0,93	35,2	0,9	6,5
	Рапсоропша "Амулет"	1	0,99	40,3	1,1	4,2
	Рапсоропша "Панацея"	0,5	1,01	39,4	1	7,1
	Лён «Нилин»	1,9	0,94	87	5,5	0,86
	Лён "Бирюза"	2,5	0,97	96	7,4	0,9
	Лён "Даник"	2,5	0,98	101	7,9	1,05
	Горчица Белая «Колла»	1,9	0,97	530,2	5,1	1,44
	Горчица "Люкс"	2,3	0,95	510,6	3,2	1,87
	Горчица "Ника"	2,8	0,94	537,2	3,4	1,6
	Рапс Яровой "Амулет"	2,6	1,3	25	3,9	7,9
	Рапс Озимый "Лорис"	3,3	1,4	24	4,1	8,02
	Рапс "Галант"	2,3	1,3	25	2,9	8,7
	Подсолнечник "Бузулук"	2,8	1,2	410,1	58	35,7
	Подсолнечник "Джинн"	3	1,1	432,3	69	38,3
	Подсолнечник "Умник"	3,1	1	405,9	60	34,9
	Рыжик "Карат"	2	1,1	100,3	0,93	0,47
	Рыжик "Омич"	1,9	1	104,2	1,2	0,49
	Рыжик "Пензяк"	2,2	1,3	99,9	1,2	0,46
	Рапсоропша "Пятнистая"	1,6	1,09	38,4	0,8	7,3
	Рапсоропша "Амулет"	1,2	0,98	38,7	1	4,2
2 (2018)	Рыжик "Омич"	1,9	1	104,2	1,2	0,49
	Рыжик "Пензяк"	2,2	1,3	99,9	1,2	0,46
	Рапсоропша "Пятнистая"	1,6	1,09	38,4	0,8	7,3
	Рапсоропша "Амулет"	1,2	0,98	38,7	1	4,2
	Рапсоропша "Панацея"	0,6	0,99	40,1	0,9	5,6
	Лён «Нилин»	2,1	1,01	90	5,3	1,02
	Лён "Бирюза"	2,4	0,96	99	7,5	0,99
	Лён "Даник"	2,6	0,98	96	7,8	0,97
	Горчица Белая «Колла»	2	0,96	540,3	5,2	1,93
	Горчица "Люкс"	2	0,98	533,4	3,5	1,77
	Горчица "Ника"	3,2	0,99	524,2	3,3	1,6
	Рапс Яровой "Амулет"	2,4	1,1	22	4,1	8,1
3 (2019)	Рапс Озимый "Лорис"	4,0	1,5	23	4,2	8
	Рапс "Галант"	2,5	1,2	25	3	8,3

Номер кластера	Сорт	Урожайность, т/га	Продуктивность куста	Число зёрен в колосе, шт.	Масса 1000 зёрен, г	Масса зерна с колоса, г.
3 (2019)	Подсолнечник “Бузулук”	3	1,3	407,2	62	36,3
	Подсолнечник “Джинн”	3,1	1,1	421,3	71	34,2
	Подсолнечник “Умник”	3,5	1,1	405,1	62	37,5
	Рыжик “Карат”	2,3	1,2	97,6	0,92	0,44
	Рыжик “Омич”	2,1	1	101,2	1,3	0,46
	Рыжик «Пензяк»	2	1,3	102,1	1,3	0,49
	Расторопша “Пятнистая”	1,5	0,97	41,2	1	6,7
	Расторопша “Амулет”	1,1	0,99	39,5	1,1	7,2
	Расторопша “Панацея”	0,6	1,01	40,3	1,2	7,9
	Лён «Нилин»	2	0,98	97	5,4	0,99
	Лён “Бирюза”	2,6	0,95	101	7,9	1,09
	Лён «Даник»	2,7	1,03	104	8	1,02
	Горчица Белая «Колла»	2,1	0,98	534,2	5,2	1,74
	Горчица “Люкс”	2,2	1,01	540,6	3,3	1,95
	Горчица “Ника”	3,1	0,97	538,9	3,7	1,87

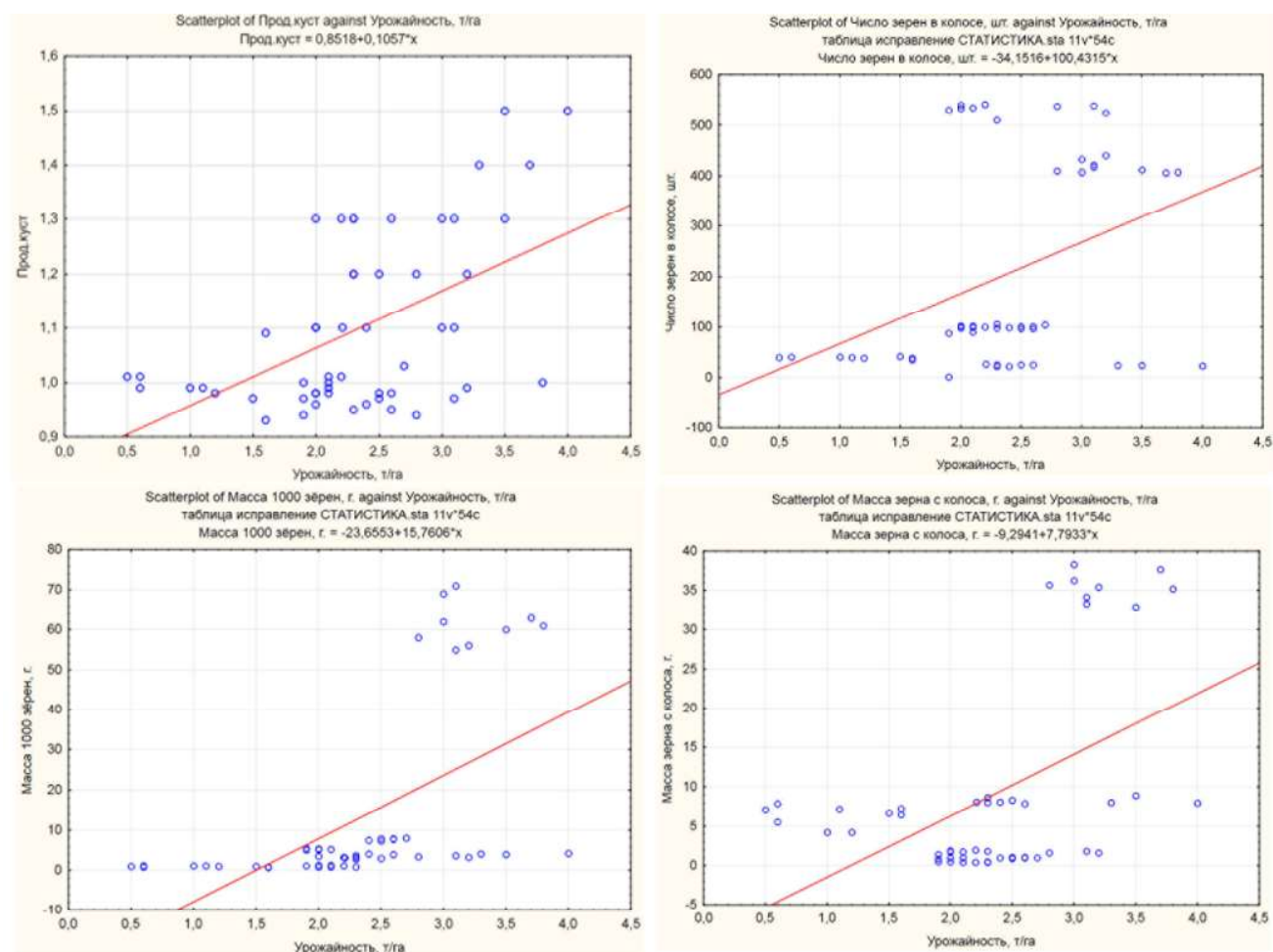


Рисунок 2. Точечная диаграмма из программы Statistica

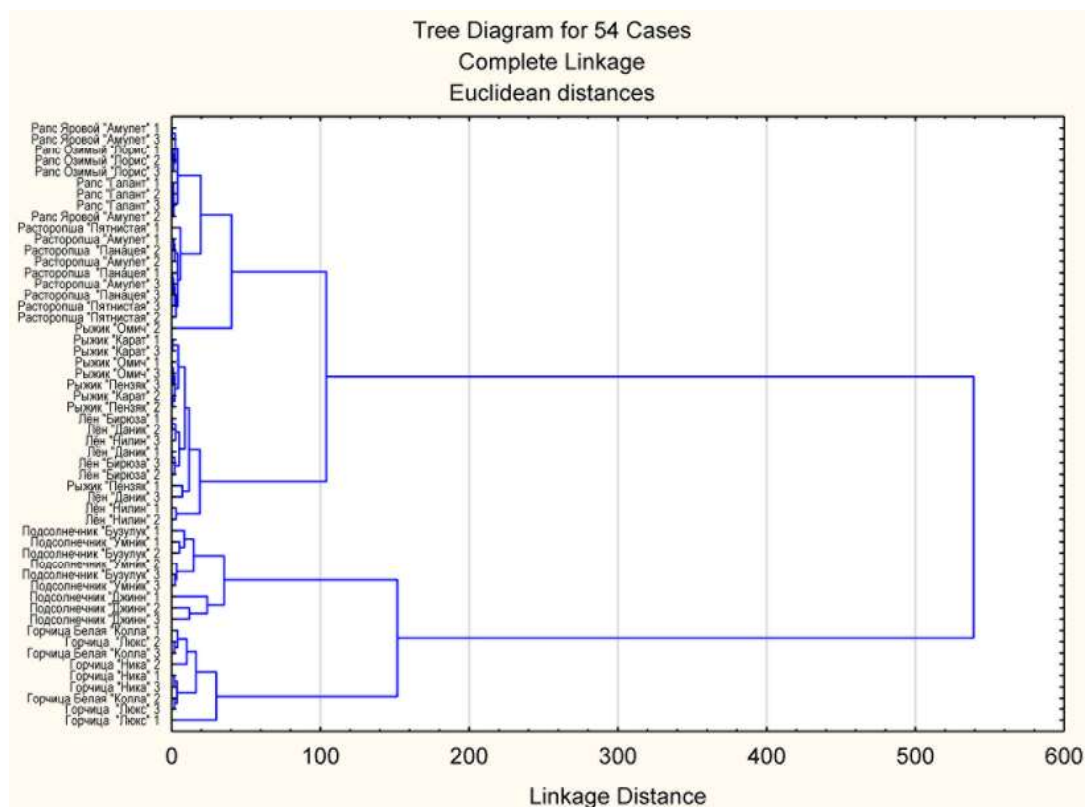


Рисунок 3. Сравнительная характеристика сортов масличных культур по элементам структуры урожая в среднем за 2017-2019 гг.

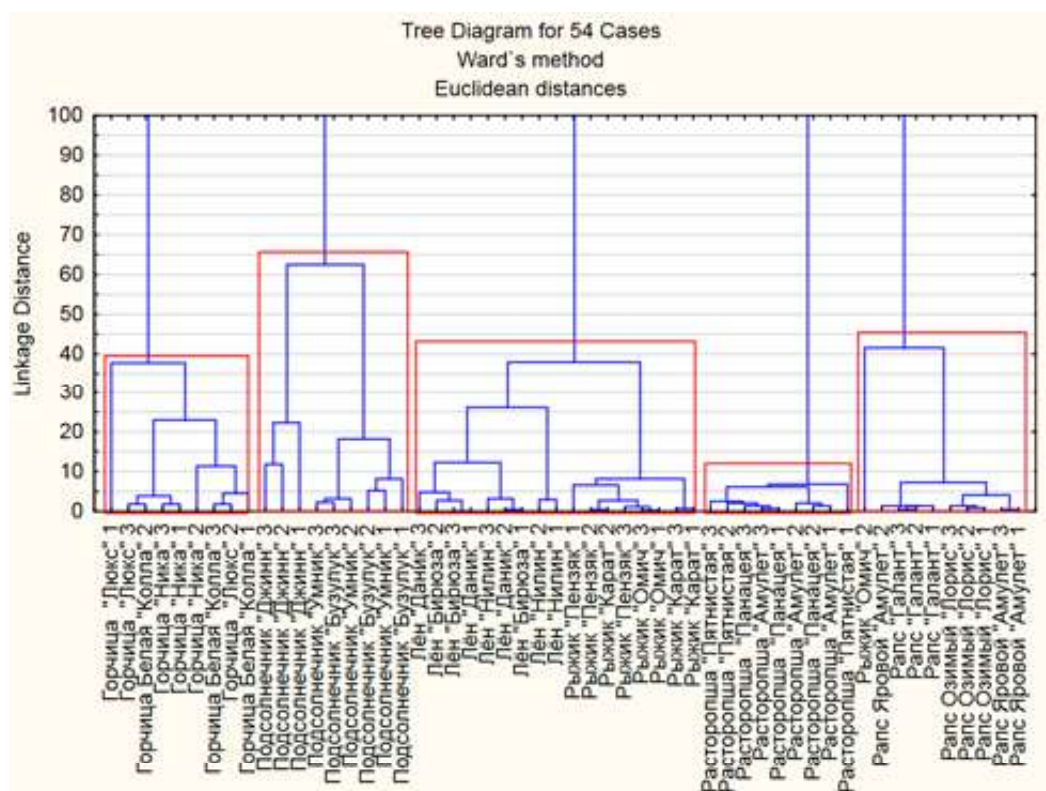


Рисунок 4. Объединенные кластеры на дендограмме

Средние значения в разрезе кластеров

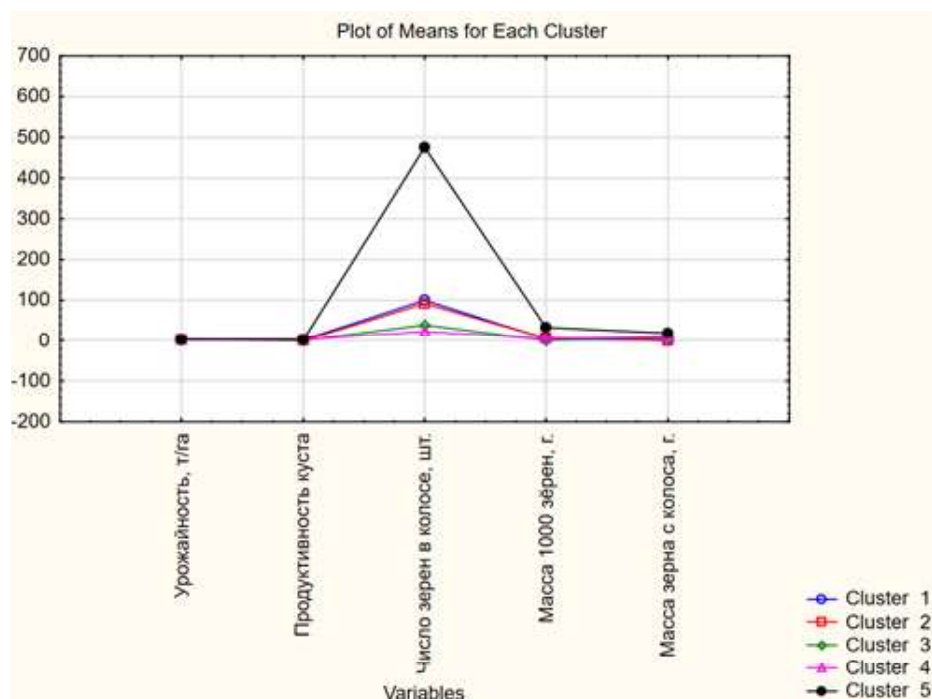


Рисунок 5. График средних расстояний для элементов структуры урожая масличного сырья 2017-2019 гг.

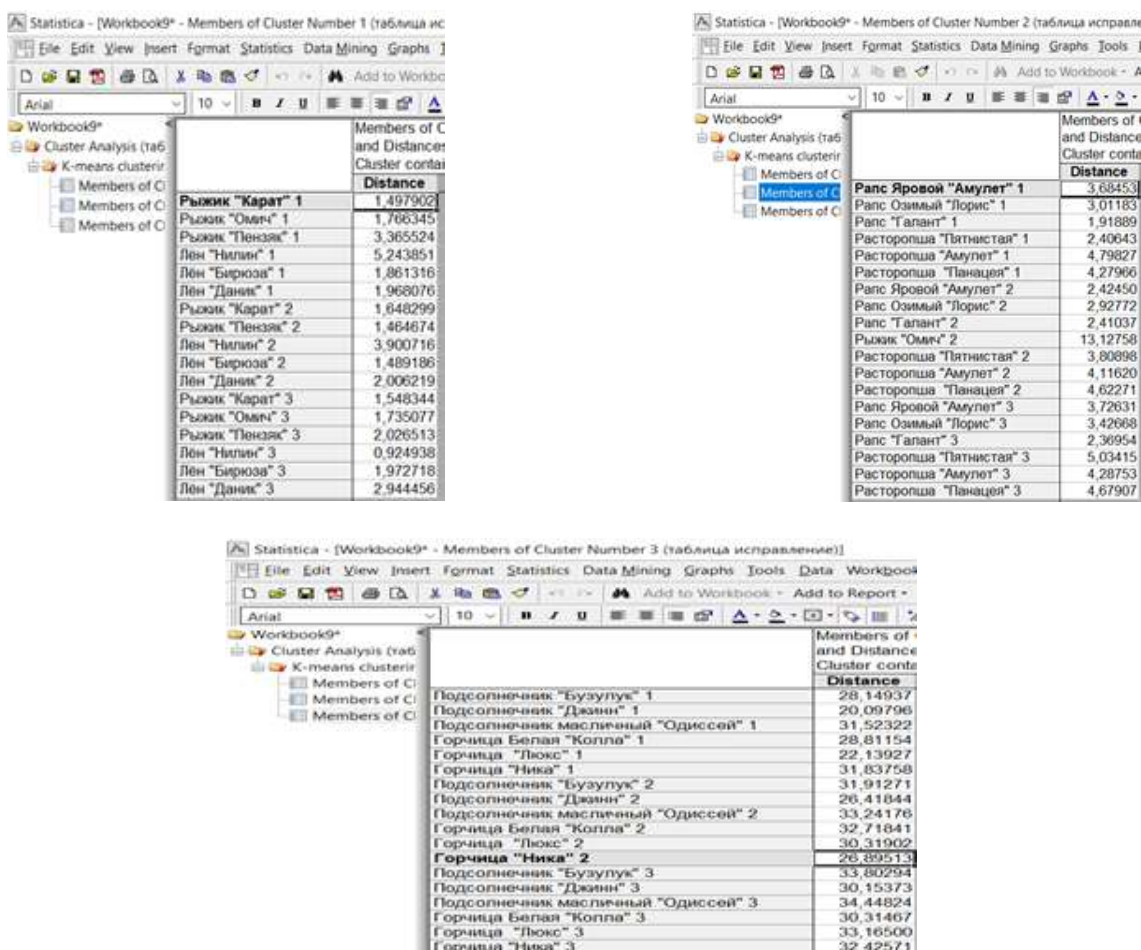


Рисунок 6. Результативные таблицы разбиения на кластеры

Анализ Рисунка 6 показал, что во втором столбце представлено евклидово расстояние от центра кластера до исследуемой культуры, т.е. средние величины.

Полученные данные по средним значениям исследуемых показателей 5 кластеров необходимы для определения как они различаются.

С помощью дисперсионного анализа для кластеров получены значения $p < 0,05$, что свидетельствует о значимом различии (Рисунок 7). Определены

евклидовы расстояния исследуемых параметров средних значений соответствующих 5 кластеров (Рисунок 8).

Рассчитаны элементы кластеров масличного сырья, получены евклидовы расстояния объектов от центров (средних значений) соответствующих им кластеров (Рисунок 9)

Следует отметить, что полученные 5 кластеров имеют схожие сорта масличных культур.

Variable	Analysis of Variance					
	Between SS	df	Within SS	df	F	signif. p
Урожайность, т/га	20	4	12,76	49	19,1612	0,000000
Продуктивность куста	0	4	0,90	49	5,7155	0,000741
Число зерен в колосе, шт.	2082955	4	62111,29	49	410,8141	0,000000
Масса 1000 зерен, г.	10635	4	15350,46	49	8,4871	0,000028
Масса зерна с колоса, г.	2906	4	5205,42	49	6,8388	0,000186

Рисунок 7. Дисперсионный анализ

Variable	Cluster Means				
	Cluster No. 1	Cluster No. 2	Cluster No. 3	Cluster No. 4	Cluster No. 5
Урожайность, т/га	2,2462	2,27500	1,07778	2,70100	2,8222
Продуктивность куста	1,0915	0,97500	0,99556	1,26000	1,0917
Число зерен в колосе, шт.	100,6154	92,25000	39,23333	21,74020	474,6444
Масса 1000 зерен, г.	3,4962	6,50000	1,00000	3,41000	32,8278
Масса зерна с колоса, г.	0,6838	0,93750	6,30000	7,44900	18,5983

Рисунок 8. Кластерное расстояние

Members of Cluster and Distance: Cluster contains		Members of Cluster and Distance: Cluster contains		Members of Cluster and Distance: Cluster contains	
Distance		Distance		Distance	
Рыжик "Карат" 1	1,590967	Расторопша "Пятнистая" 1	1,821801	Подсолнечник "Бузулук" 1	28,14937
Рыжик "Омич" 1	1,038654	Расторопша "Амулет" 1	1,054880	Подсолнечник "Джинн" 1	20,09796
Рыжик "Пензяк" 1	2,437888	Расторопша "Панацея" 1	0,447619	Подсолнечник масличный "Одиссей" 1	31,52322
Лен "Даник" 1	1,987598	Расторопша "Пятнистая" 2	0,634993	Горчица Белая "Колла" 1	28,81154
Рыжик "Карат" 2	1,165419	Расторопша "Амулет" 2	0,970528	Горчица "Люкс" 1	22,13927
Рыжик "Пензяк" 2	1,084416	Расторопша "Панацея" 2	0,543951	Горчица "Ника" 1	31,83758
Лен "Бирюза" 2	1,937780	Расторопша "Пятнистая" 3	0,917246	Подсолнечник "Бузулук" 2	31,91271
Рыжик "Карат" 3	1,777821	Расторопша "Амулет" 3	0,422288	Подсолнечник "Джинн" 2	26,41844
Рыжик "Омич" 3	1,024179	Расторопша "Панацея" 3	0,890647	Подсолнечник масличный "Одиссей" 2	33,24176
Рыжик "Пензяк" 3	1,197397			Горчица Белая "Колла" 2	32,71841
Лен "Нилин" 3	1,836431			Горчица "Люкс" 2	30,31902
Лен "Бирюза" 3	1,992586			Горчица "Ника" 2	26,89513
Лен "Даник" 3	2,532312			Подсолнечник "Бузулук" 3	33,80294
				Подсолнечник "Джинн" 3	30,15373
				Подсолнечник масличный "Одиссей" 3	34,44824
				Горчица Белая "Колла" 3	30,31467
				Горчица "Люкс" 3	33,16500
				Горчица "Ника" 3	32,42571

Рисунок 9. Элементы кластеров масличного сырья

Выводы

В результате идентификации масличного сырья установлено, что сорта, принадлежащие одному роду как правила входят в один кластер. Однако есть исключение – рыжик «Омич» 2 входит в кластер 5, в котором присутствуют только сорта рапса.

К одному кластеру принадлежат родственные сорта: кластер 1 – горчица, кластер 2 – подсолнечник, кластер 3 – лен и рыжик, кластер 4 – рапшопш. В результате кластерного анализа было выяснено, что третий кластер представлен родственными сортами льна и рыжика, что возможно объясняется приспособлением к климатическим условиям – засухе.

Полученные результаты исследования позволили сгруппировать разные сорта масличных культур и выделить близкие. Таким образом, применяя Евклидову метрику, объединили сорта масличных культур и сгруппировали их.

Следует отметить, используя огромные выборки генотипов и их структурных элементов, а также анализируя соотношение их уровня с помощью Евклидовой метрики, классифицированы сорта масличных культур и выявлена степень сбалансированности основных составляющих урожая. Данные можно использовать для определения конкурентоспособности различных масличных культур при разработке новых пищевых продуктов, а также определения ниши, в которой лучше позиционировать выводимый на рынок продукт.

Литература

- Бушнев, А. С., Подлесный, С. П., & Хатит, А. Б. (2019). Влияние нормы высевы семян на некоторые элементы структуры урожая сортов и гибридов подсолнечника. Масличные культуры, 2, 69-74. <https://doi.org/10.25230/2412-608X-2019-2-178-69-74>.
- Котлярова, Е. Г., & Титовская, Л. С. (2018). Изменчивость биометрических параметров гибридов подсолнечника в зависимости от способов основной обработки почвы и листовых подкормок. Вестник Мичуринского государственного аграрного университета, 2, 17-22.
- Лисицын, Е. М., & Амунова, О. С. (2014). Генетическое разнообразие сортов яровой мягкой пшеницы по алюмоустойчивости. Вавиловский журнал генетики и селекции, 18(3), 497-505.
- Мулянова, Ю. Н., & Косников, С. Н. (2018). Кластерный анализ в сельском хозяйстве. Economics, 5, 43-57.
- Румянцев, А. В., & Глуховцев, В. В. (2011). Роль селекции зерновых и кормовых культур в повышении урожайности и экономической стабильности сельскохозяйственного производства в условиях Среднего Поволжья. Инновация и модернизация сельскохозяйственного производства в условиях меняющегося климата: Материалы международной научно-практической конференции (с. 19-23). Оренбург.
- Тищенко, В. Н., Ищенко, А. Г., & Дубенец, Н. В. (2016). Идентификация сортов и селекционных линий озимой пшеницы в кластерном анализе по сбалансированности количественных признаков в адаптивной селекции. Вестник Курганской ГСХА, 1, 41-44.
- Тищенко, В. Н., Панченко, П. М., Чернышева, О. П. (2013). Идентификация сортов и селекционных линий пшеницы озимой по сбалансированности количественных признаков с использованием кластерного анализа. Вестник Полтавской государственной аграрной академии, 3, 28-35.
- Шаманин, В. П., Петуховский, С. Л., & Краснова, Ю. С. (2016). Кластерный анализ сортов мягкой яровой пшеницы по элементам структуры урожая в Южной лесостепи Западной Сибири. Вестник КрасГАУ, 4, 147-152.
- Abdorreza, J., Hossein, N., & Morteza, N. (2015). Relationship between agronomic and morphological traits in barley varieties under drought stress condition. International Journal of Basic and Applied Sciences, 9(9). 1507-1511.
- Agahi, K., Ahmadi, J., Oghan, H. A., Fotokian, M. H., & Orang, S. F. (2020). Analysis of genotype× environment interaction for seed yield in spring oilseed rape using the AMMI model. Crop Breeding and Applied Biotechnology, 20(1). <https://doi.org/10.1590/1984-70332020v20n1a2>
- Bhattarai, R. P., Thapa, D. B., Ojha, B. R., Kharel, R., & Sapkota, M. (2017). Cluster analysis of Elite spring wheat (Triticum aestivum L.) genotypes based on yield and yield attributing traits under irrigated condition. International Journal of Experimental Research and Review, 10, 9-14.
- Creissen, H. E., Jorgensen, T. H., & Brown, J. K. M. (2016). Increased yield stability of field-grown winter barley (Hordeum vulgare L.) varietal mixtures through ecological processes. Crop Protection, 85, 1-8. <https://doi.org/10.1016/j.cropro.2016.03.001>.
- Cuesta-Marcos, A., Kling, J. G., Belcher, A. R., Filichkin, T., Fisk, S. P., Graebner, R., Helgersson, L., Herb, D., Meints, B., Ross, A. S., Hayes, P. M., & Ulrich, S. E. (2016). Barley: Genetics and Breeding. In: Wrigley, C., Corke, H., and Seetharaman, K., Faubion, J. (eds.) Encyclopedia of Food Grains (2nd ed., pp. 287-295). Oxford: Academic Press.

- El-haak, M. A., Atta, B. M., & Abd Rabo, F. F. (2015). Seed yield and important seed constituents for naturally and cultivated milk thistle (*silybum marianum*) plants. *The Egyptian Journal of Experimental Biology*, 11(2), 141–146.
- Embaye, W. T., Bergtold, J. S., Archer, D., Flora, C., Andrango, G. C., Odening, M., & Buysse J. (2018). Examining farmers' willingness to grow and allocate land for oilseed crops for biofuel production. *Energy Economics*, 71, 311–320.
- Gebeyehu, A., Hammenhag, C., Ortiz, R., Tesfaye, K., & Geleta, M. (2021). Characterization of Oilseed Crop Noug (*Guizotia abyssinica*) Using Agro-Morphological Traits. *Agronomy*, 11(8), 1479. <https://doi.org/10.3390/agronomy11081479>.
- Amiri Oghan, H., N. Sabaghnia, V. Rameeh, H.R. Fanaee and E. Hezarjeribi. Univariate stability
- Hassan, A. O., Hossein, Z. T., Reza, F. H., Khattoon, K. N., Gholamreza, G., Amirkhosro, D., & Bagher, V. M. (2019). Stability Study of Seed Yield in Promising Lines of Spring Oilseed Rape in Southern-Worm Regions of Iran. *Journal of crop breeding*, 11(31), 42–54.
- Hussain, F., Rafiq, M., Ghias, M., Qamar, R., Razzaq, M. K., Hameed, A., Habib, S., Saad, H., & Mustafa, H. S. B. (2017). Genetic Diversity for Seed Yield and its Components Using Principal Component and Cluster Analysis in Sunflower. *Life Science Journal*, 14(5), 71–78. <https://doi.org/10.7537/marslsj140517.10>
- Jankowski, K. J., Załuski, D., & Sokółski, M. (2020). Canola-quality white mustard: Agronomic management and seed yield. *Industrial Crops and Products*, 145, 112138. <https://doi.org/10.1016/j.indcrop.2020.112138>
- Khalid, A. & Hameed, A. (2021). Genetic divergence in wheat genotypes based on seed biochemical profiles through agglomerative hierarchical clustering and association analysis among traits. *Pakistan Journal of Botany*, 53(4). [https://doi.org/10.30848/PJB2021-4\(7\)](https://doi.org/10.30848/PJB2021-4(7))
- Krzyżaniak, M., Stolarski, M. J., Tworowski, J., Puttick, D., Eynck, C., Załuski, D., & Kwiatkowski, J. (2019). Yield and seed composition of 10 spring camelina genotypes cultivated in the temperate climate of Central Europe. *Industrial Crops and Products*, 138, 111443. <https://doi.org/10.1016/j.indcrop.2019.06.006>
- Lobell, D. B., & Gourdji, S. M. (2012). The influence of climate change on global crop productivity. *Plant Physiology*, 160(4), 1686–1697. <https://doi.org/10.1104/pp.112.208298>
- Majerova, I., & Nevima, J. (2017). The measurement of human development using the Ward method of cluster analysis. *Journal of International Studies*, 10(2), 239–257. <https://doi.org/10.14254/2071-8330.2017/10-2/17>
- Mustafa, H. S. B., Farooq, J., Hasan, E. U., Bibim T., & Mahmood, T. (2015). Cluster and principle component analyses of maize accessions under normal and water stress conditions. *Journal of Agricultural Sciences Belgrade*, 60(1), 33–48. <https://doi.org/10.2298/JAS1501033M>
- Patial, R., Paul, S., Sharma, D., Sood, V. K. & Kumar, N. (2019). Morphological characterization and genetic diversity of linseed (*Linum usitatissimum* L.). *Journal of Oilseeds Research*, 36(1), 8–16.
- Venujayakanth, B., Dudhat, A. S., Swaminathan, B. & Anurag, M. L. (2017). Assessing Crop Genetic Diversity using Principle Component Analysis: A Review. *Trends in Biosciences*, 10(2), 523–528.

Application of Cluster Analysis for Oilseeds Identification

Natalia L. Kleymenova

*Voronezh State University of Engineering Technologies
19, Revolutsii Prospect, Voronezh, 394036, Russian Federation
E-mail: klesha78@list.ru*

Inessa N. Bolgova

*Voronezh State University of Engineering Technologies
19, Revolutsii Prospect, Voronezh, 394036, Russian Federation
E-mail: bolgovainessa@yandex.ru*

Maxim V. Kopylov

*Voronezh State University of Engineering Technologies
19, Revolutsii Prospect, Voronezh, 394036, Russian Federation
E-mail: kopylov-maks@yandex.ru*

Alla N. Pegina

*Voronezh State University of Engineering Technologies
19, Revolutsii Prospect, Voronezh, 394036, Russian Federation
E-mail: toriss@yandex.ru*

Daria A. Koshelkina

*Voronezh State University of Engineering Technologies
19, Revolutsii Prospect, Voronezh, 394036, Russian Federation
E-mail: toriss@yandex.ru*

A search for mathematical methods to assess the level of balance of the main quantitative traits in oilseed varieties is described in the article. The main quantitative traits and crop yield are analyzed on the basis of experimental data of cluster analysis application for assessment varieties of oilseeds. The identification of varieties of oilseeds, grown in different regions of the country was carried out with the help of cluster analysis. The main cultivation areas of oilseeds in 2017-2019 have been also studied. As a result of the research, it was established that the cultivation area of flax, mustard and camelina crops has reduced, which is connected with the crop yield and low exports. The state of growth and the development of oilseeds, the duration of their cultivation, which depends on the biological characteristics of the breed, as well as the cultivation technology, have been analyzed. The specific features of the oilseeds have been studied. Unified measures which include a number of investigated features are identified with the help of the method. These features are unified using the metric into one resemblance cluster of the grouping of the studied objects. The breeds are grouped into three clusters according to years. The composition of each cluster changes depending on the natural and climatic conditions. Exploratory analysis using a point diagram is performed. A dendrogram of the comparative characteristics of oilseed varieties by the elements of the yield structure is given. Using the K-means clustering method, the Euclidean distances of the studied parameters of the average values of the corresponding 5 clusters that have similar varieties of oilseeds are determined. The research shows that it is possible to determine the competitiveness of various oilseeds and to develop new food products.

Keywords: cluster, cluster analysis, oilseeds, statistics, dendrogram, harvest

References

- Bushiev, A. S., Podlesnyi, S. P., & Khatit, A. B. (2019). Vliyanie normy vyseva semyan na nekotorye elementy struktury urozhaya sortov i gibridov podsolnechnika [Influence of the seed-ing rate on some elements of the structure of the yield of varieties and hybrids of sunflower]. *Maslichnye kul'tury* [Oilseeds], 2, 69-74. <https://doi.org/10.25230/2412-608Kh-2019-2-178-69-74>.

- Kotlyarova, E. G., Titovskaya, L. S. (2018). Izmenchivost' biometricheskikh parametrov gibridov podsolnechnika v zavisimosti ot sposobov osnovnoj obrabotki pochvy i listovykh podkormok [Variability of biometric parameters of sunflower hybrids depending on the methods of basic tillage and leaf fertilizing]. *Vestnik Michurinskogo gosudarstvennogo agrarnogo universiteta* [Bulletin of Michurinsky State Agrarian University], 2, 17–22.
- Lisicyn, E. M., Amunova, O. S. (2014). Geneticheskoe raznoobrazie sortov yarovoj myagkoj pshenicy po alyumoustojchivosti [Genetic diversity of spring soft wheat varieties by aluminum resistance]. *Vavilovskij zhurnal genetiki i selekcii* [Vavilovsky Journal of Genetics and Breeding], 18(3), 497–505.
- Mulyanova, Yu. N., & Kosnikov, S. N. (2018). Klasternyi analiz v sel'skom khozyaistve [Cluster analysis in agriculture]. *Economics*, 5, 43–57.
- Rumyantsev, A. V., & Glukhovtsev, V. V. (2011). Rol' selekcii zernovykh i kormovykh kul'tur v povyshenii urozhainosti i ekonomicheskoi stabil'nosti sel'skokhozyaistvennogo proizvodstva v usloviyakh Srednego Povolzh'ya [The role of selection of grain and fodder crops in increasing the yield and economic stability of agricultural production in the conditions of the Middle Volga region]. In *Innovatsiya i modernizatsiya sel'skokhozyaistvennogo proizvodstva v usloviyakh menyayushchegosya klimata: Materialy mezhdunarodnoi nauchno-prakticheskoi konferentsii* [Innovation and modernization of agricultural production in a changing climate: Proceedings of the international scientific and practical conference] (pp. 19–23). Orenburg.
- Tishchenko, V. N., Ishchenko, A. G., & Dubenets, N. V. (2016). Identifikatsiya sortov i selekcionnykh linii ozimoi pshenitsy v klasternom analize po sbalansirovannosti kolichestvennykh priznakov v adaptivnoi selekcii [Identification of varieties and breeding lines of winter wheat in cluster analysis for the balance of quantitative traits in adaptive breeding]. *Vestnik Kurganskoi GSKhA* [Bulletin of the Kurgan State Agricultural Academy], 1, 41–44.
- Tishchenko, V. N., Panchenko, P. M., & Cherny'sheva, O. P. (2013). Identifikatsiya sortov i selekcionnykh linii pshenicy ozimoi po sbalansirovannosti kolichestvennykh priznakov s ispol'zovaniem klasternogo analiza [Identification of varieties and breeding lines of winter wheat by the balance of quantitative traits using cluster analysis]. *Vestnik Poltavskoy gosudarstvennoy agrarnoy akademii* [Bulletin of Poltava State Agrarian Academy], 3, 28–35.
- Shamanin, V. P., Petukhovskii, S. L., & Krasnova, Yu. S. (2016). Klasternyi analiz sortov myagkoj yarovoi pshenitsy po elementam struktury urozhaya v Yuzhnoi lesostepi Zapadnoi Sibiri [Cluster analysis of soft spring wheat varieties by elements of the yield structure in the southern forest-steppe of Western Siberia]. *Vestnik KrasGAU* [Bulletin of the Krasnoyarsk State Agrarian University], 4, 147–152.
- Abdorreza, J., Hossein, N., & Morteza, N. (2015). Relationship between agronomic and morphological traits in barley varieties under drought stress condition. *International Journal of Basic and Applied Sciences*, 9(9), 1507–1511.
- Agahi, K., Ahmadi, J., Oghan, H. A., Fotokian, M. H., & Orang, S. F. (2020). Analysis of genotype × environment interaction for seed yield in spring oilseed rape using the AMMI model. *Crop Breeding and Applied Biotechnology*, 20(1). <https://doi.org/10.1590/1984-70332020v20n1a2>.
- Bhattarai, R. P., Thapa, D. B., Ojha, B. R., Kharel, R., & Sapkota, M. (2017). Cluster analysis of Elite spring wheat (*Triticum aestivum* L.) genotypes based on yield and yield attributing traits under irrigated condition. *International Journal of Experimental Research and Review*, 10, 9–14.
- Creissen, H. E., Jorgensen, T. H., & Brown, J. K. M. (2016). Increased yield stability of field-grown winter barley (*Hordeum vulgare* L.) varietal mixtures through ecological processes. *Crop Protection*, 85, 1–8. <https://doi.org/10.1016/j.cropro.2016.03.001>.
- Cuesta-Marcos, A., Kling, J. G., Belcher, A. R., Filichkin, T., Fisk, S. P., Graebner, R., Helgersson, L., Herb, D., Meints, B., Ross, A. S., Hayes, P. M., & Ulrich, S. E. (2016). Barley: Genetics and Breeding. In: Wrigley, C., Corke, H., and Seetharaman, K., Faubion, J. (eds.) *Encyclopedia of Food Grains* (2nd ed., pp. 287–295). Oxford: Academic Press.
- El-haak, M. A., Atta, B. M., & Abd Rabo, F. F. (2015). Seed yield and important seed constituents for naturally and cultivated milk thistle (*silybum marianum*) plants. *The Egyptian Journal of Experimental Biology*, 11(2), 141–146.
- Embaye, W. T., Bergtold, J. S., Archer, D., Flora, C., Andrango, G. C., Odening, M., & Buysse J. (2018). Examining farmers' willingness to grow and allocate land for oilseed crops for biofuel production. *Energy Economics*, 71, 311–320.
- Gebeyehu, A., Hammenhag, C., Ortiz, R., Tesfaye, K., & Geleta, M. (2021). Characterization of Oilseed Crop Noug (*Guizotia abyssinica*) Using Agro-Morphological Traits. *Agronomy*, 11(8), 1479. <https://doi.org/10.3390/agronomy11081479>.
- Hassan, A. O., Hossein, Z. T., Reza, F. H., Khattoon, K. N., Gholamreza, G., Amirkhosro, D., & Bagher, V. M. (2019). Stability Study of Seed Yield in Promising Lines of Spring Oilseed Rape in Southern-Worm Regions of Iran. *Journal of crop breeding*, 11(31), 42–54.
- Hussain, F., Rafiq, M., Ghias, M., Qamar, R., Razzaq, M. K., Hameed, A., Habib, S., Saad, H., &

- Mustafa, H. S. B. (2017). Genetic Diversity for Seed Yield and its Components Using Principal Component and Cluster Analysis in Sunflower. *Life Science Journal*, 14(5), 71-78. <https://doi.org/10.7537/marslsj140517.10>.
- Jankowski, K. J., Załuski, D., & Sokólski, M. (2020). Canola-quality white mustard: Agronomic management and seed yield. *Industrial Crops and Products*, 145, 112138. <https://doi.org/10.1016/j.indcrop.2020.112138>.
- Khalid, A. & Hameed, A. (2021). Genetic divergence in wheat genotypes based on seed biochemical profiles through agglomerative hierarchical clustering and association analysis among traits. *Pakistan Journal of Botany*, 53(4). [https://doi.org/10.30848/PJB2021-4\(7\)](https://doi.org/10.30848/PJB2021-4(7)).
- Krzyżaniak, M., Stolarski, M. J., Tworowski, J., Puttick, D., Eynck, C., Załuski, D., & Kwiatkowski, J. (2019). Yield and seed composition of 10 spring camelina genotypes cultivated in the temperate climate of Central Europe. *Industrial Crops and Products*, 138, 111443. <https://doi.org/10.1016/j.indcrop.2019.06.006>.
- Lobell, D. B., & Gourdji, S. M. (2012). The influence of climate change on global crop productivity. *Plant Physiology*, 160(4), 1686-1697. <https://doi.org/10.1104/pp.112.208298>
- Majerova, I., & Nevima, J. (2017). The measurement of human development using the Ward method of cluster analysis. *Journal of International Studies*, 10(2), 239-257. <https://doi.org/10.14254/2071-8330.2017/10-2/17>.
- Mustafa, H. S. B., Farooq, J., Hasan, E. U., Bibim T., & Mahmood, T. (2015). Cluster and principle component analyses of maize accessions under normal and water stress conditions. *Journal of Agricultural Sciences Belgrade*, 60(1), 33-48. <https://doi.org/10.2298/JAS1501033M>.
- Patial, R., Paul, S., Sharma, D., Sood, V. K. & Kumar, N. (2019). Morphological characterization and genetic diversity of linseed (*Linum usitatissimum* L.). *Journal of Oilseeds Research*, 36(1), 8-16.
- Venujayakanth, B., Dudhat, A. S., Swaminathan, B. & Anurag, M. L. (2017). Assessing Crop Genetic Diversity using Principle Component Analysis: A Review. *Trends in Biosciences*, 10(2), 523-528.