

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МНОГОМЕРНОЙ СТАТИСТИКИ В СЕЛЕКЦИИ ПШЕНИЦЫ ОЗИМОЙ

USING OF MULTIDIMENSIONAL STATISTICS METHODS FOR WINTER WHEAT SELECTION

Голева Г.Г. ¹ – кандидат с.-х. наук, доцент кафедры селекции и семеноводства

Вашченко Т.Г. ¹ – доктор с.-х. наук, профессор кафедры селекции и семеноводства

Голев А.Д. ² – кандидат техн. наук, доцент кафедры безопасности жизнедеятельности и правовых отношений

¹ Федеральное государственное бюджетное учреждение высшего образования

Воронежский государственный аграрный университет имени императора Петра I, кафедра агрохимии и почвоведения 394087, Россия, г. Воронеж, ул. Мичурина, д. 1
E-mail: biolog2011@rambler.ru

² Федеральное государственное бюджетное учреждение высшего образования Воронежский государственный лесотехнический университет им. Г.Ф. Морозова 394087, Россия, г. Воронеж, ул. Тимирязева, 8

Представлены результаты исследований по использованию методов многомерной статистики для оценки морфо-биологической разнокачественности сортообразцов пшеницы озимой, проведенных в 2005-2016 годах в условиях лесостепи Центрального Черноземья. Апробированы различные методы группировки образцов – кластерный анализ (метод последовательной дихотомии, *k*-средних, иерархической классификации), нейросетевая обработка данных (сеть «Кохена»). Методами дисперсионного и дискриминантного анализов установлено, что значимое влияние на кластеризацию сортообразцов оказывают только три признака: длина периода всходы-колошение, высота растения и масса 1000 зёрен. При этом принадлежность сортообразцов к определенной группе остаётся постоянной независимо от условий года. Для скрещиваний при гибридизации родительские компоненты подбирают с учетом их принадлежности к разным кластерам. Скрещивания проводят между сортообразцами с контрастными признаками. Использование метода *k*-средних для планирования скрещиваний позволяет получать ценные в селекционном отношении гибриды озимой пшеницы, которые характеризуются полиморфизмом, наличием благоприятных трансгрессий, высокой продуктивностью, зимостойкостью, устойчивостью к полеганию. Такой принцип подбора родительских компонентов был использован на практике при создании нового сорта озимой мягкой пшеницы Алексия, переданного в 2016 году в Государственное сортоиспытание. Апробированный метод *k*-средних в качестве способа оценки морфо-биологической разнокачественности сортообразцов упрощает подбор родительских компонентов для скрещиваний и позволяет создавать гибриды озимой пшеницы, которые в условиях ЦЧР характеризуются высокой продуктивностью, зимостойкостью, устойчивостью к неблагоприятным факторам среды.

Ключевые слова: озимая пшеница, селекционные образцы, морфо-биологические показатели, кластерный, дисперсионный и дискриминантный анализы, гибриды, подбор родительских компонентов для скрещиваний.

Goleva G.G. ¹ – Candidate of Agricultural Sciences, Associate Professor of the Dept. of Plant and Seed Selection Breeding

Vashchenko T.G. ¹ – Doctor of Agricultural Sciences, Professor of the Dept. of Plant and Seed Selection Breeding

Golev A.D. ² – Candidate of Engineering Sciences, Associate Professor of the Dept. of Life Safety and Legal Relations

¹ Federal State Budgetary Educational Institution of Higher Professional Education

«Voronezh State Agricultural University named after Emperor Peter I» 1, Michurina str., Voronezh 394087, Russia
E-mail: biolog2011@rambler.ru

² Federal State Budgetary Educational Institution of Higher Professional Education

«Voronezh State University of Forestry and Technologies named after G.F. Morozov» 8, Timiryazeva str., Voronezh 394087, Russia

The article presents the using of multidimensional statistics methods for assessing the morphological and biological quality of winter wheat varieties in experiments conducted under the conditions of the forest-steppe of the Central Chernozem region in 2005-2016. Agrotechnics is generally accepted for the region. Mathematical processing was carried out by the package Statistica 6.1. There were tested different methods of grouping the samples, such as cluster analysis (successive dichotomy, *k*-means, hierarchical classification), neural network data processing (Kohonen network). Only three characteristics had a significant effect on the clustering of varieties: the sprouting period, the height of the plant and the mass of 1000 grains; it was established by dispersion method and discriminant analysis. The parent components were selected on the basis of their belonging to different clusters in hybridization crosses. Crosses were between varieties with contrasting signs. The using of the *k*-averages method made it possible to obtain valuable winter wheat hybrids which possessed polymorphism, favorable transgressions, high productivity, winter hardiness, and lodging resistance. This selection principle was used in practice during the creating of a new variety of soft winter wheat Alexia, which was presented to the State Variety Test in 2016. The approved method of *k*-averages simplifies the selection of parent components for crosses and allows the creation of hybrids of winter wheat which is characterized by high productivity, winter hardiness and resistance to unfavorable environmental factors.

Keywords: winter wheat, selection samples, morphological and biological indicators, cluster, dispersion and discriminant analyses, hybrids, selection of parent components for crosses.

Введение

Подбор родительских компонентов для скрещиваний – важный и ответственный этап селекционной работы, во многом определяющий её успех. Ежегодно в селекционных питомниках родительские формы для скрещиваний отбирают из сотен, а иногда и тысяч изучаемых сортообразцов. Для селекционной работы все изучаемые образцы целесообразно объединять в группы на основе их комплексной оценки. При этом родительские компоненты для скрещиваний следует подбирать с учётом их принадлежности к определённой группе, что позволит значительно сократить объём работы по гибридизации, удешевить и ускорить селекционный процесс.

Методика

Работа проведена в ФГБОУ ВО Воронежский ГАУ в 2005-2016 годах. Агротехника в полевых опытах – общепринятая для ЦЧР. Объект исследований – сортообразцы

«зимостойкость», «длина периода вегетации», «урожайность», «высота растений», «масса зерна с колоса», «содержание клейковины в зерне», «масса 1000 зёрен», «число зёрен с колоса».

Проводя кластеризацию сортообразцов на различное число групп и определяя функцию классификации с помощью дискриминантного анализа, мы добивались 100% распределения объектов по группам. Участие каждой переменной в дискриминации оценивали с помощью показателя лямбда Уилкса, и те переменные, для которых уровень значимости (p) был больше показателя 0,05, из модели исключали [4]. В результате было установлено, что значимое влияние на группировку оказывали только три признака из 16 первоначально взятых в расчёт: «длина периода всходы-колошение», «высота растения» и «масса 1000 зёрен», а значение лямбды Уилкса, равное 0,0074, свидетельствовало о хорошей дискриминации (табл. 1).

Таблица 1. Результаты анализа дискриминантных функций

Признак	Discriminant Function Analysis Summary (Spreadsheet126) No. of vars in model: 3; Grouping: CLUSTER (19 grps) Wilks' Lambda: 0,00074 approx. F (54,289)=54,881 $p < 0,0000$					
	Wilks' Lambda	Partial Lambda	F-remove (4,109)	p-level	Toler.	1-Toler. (R-Sqr.)
Длина периода всходы-колошение	0,007	0,110	43,7	0,000	0,999	0,001
Высота растений	0,017	0,044	115,7	0,000	0,992	0,008
Масса 1000 зёрен	0,004	0,179	24,6	0,000	0,992	0,008

озимой мягкой пшеницы различного эколого-географического происхождения, число которых в разные годы составляло от 200 до 350. Статистический анализ данных осуществляли с помощью пакета Statistica 6.1 [4].

Результаты

В ходе исследований необходимо было определить признаки и способ группировки сортообразцов. Вначале их группировали на основе оценки по 16 признакам (высота растения, длина колоса; число колосков в колосе (продуктивных и непродуктивных); число зёрен в колосе; масса растения, колоса, зерна с колоса, 10 зёрен колоса; коэффициент хозяйственного использования; длина периодов всходы-колошение, колошение-спелость, вегетации; густота продуктивного стеблестоя и урожайность. Затем, с помощью кластерного анализа были выявлены группы тесно связанных признаков, которые дублируют друг друга и из дальнейших расчётов они были исключены. Расчёт Эвклидова расстояния по оставшимся признакам показал, что расстояние между сортообразцами не изменилось, а, следовательно, исключённые из расчёта признаки не несут дополнительной информации. Проводя аналогичные вычисления, избыточные признаки исключались до тех пор, пока полученные результаты перестали различаться с первоначальным вариантом расчёта (по 16 признакам). При этом число показателей уменьшилось до следующих восьми:

Методом дисперсионного анализа было подтверждено, что все признаки существенно влияли на результаты классификации. Следует отметить, что число кластеров и их наполняемость ежегодно менялись, поскольку изменялся изучаемый сортимент, однако принадлежность сортообразцов к определённой группе была постоянной.

В качестве способа группировки сортообразцов было апробировано несколько методов статистического анализа. Метод последовательной дихотомии (метод k -средних) [3], основан на разделении образцов последовательно на два кластера, хотя фактически их может быть больше. Нейросетевая обработка данных («Сеть Кохена» [2]) также не вполне отвечала нашим требованиям, поскольку нейронные сети работают как «чёрный ящик», и не всегда понятно, почему они приходят к определённому решению. Недостаток иерархической классификации на основе Эвклидова расстояния – это отсутствие возможности проверки качества кластеризации и значения признака для качества кластеризации.

Всех указанных выше недостатков лишён, по нашему мнению, метод k -средних (без ограничения числа групп), который позволяет итеративно подбирать число кластеров, оценивать адекватность разбиения на кластеры с помощью дисперсионного анализа и эффективность проведённой классификации – с помощью дискриминантного анализа [1].

Таблица 2. Характеристика гибридов озимой пшеницы

Гибридная комбинация	Число зёрен в колосе, шт.	Масса зерна, г	
		колоса	10 шт.
Росинка х Нана	34,5	1,54	0,46
Дон 93 х 494J6.11	38,0	1,73	0,45
Одесская 200 х Черноземка 99	37,4	1,87	0,53
Жатва Алтая х Зихтус	45,5	1,97	0,42
Безенчукская 380 х CADET	45,8	1,91	0,42
Московская 56 х Нана	49,0	2,26	0,46

Родительские компоненты для гибридизации мы подбираем с учётом их принадлежности к кластерам, стараемся проводить скрещивания между контрастными сортообразцами. Благодаря такому подходу созданы гибриды, характеризующиеся высокой продуктивностью и наличием благоприятных трансгрессий (табл. 2).

Данный способ подбора пар для скрещиваний был апробирован при выведении сорта пшеницы озимой мягкой Алексия, который с 2016 года проходит государственное сортоиспытание и характеризуется высокой зимостойкостью, продуктивностью (1,8-2 г) и крупным зерном (масса 1000 зерен 45-55 г).

Выводы

Методами кластерного и дискриминантного анализов определено минимальное число признаков группировки сортообразцов: «высота растения», «длина периода всходы-колошение», «масса 1000 зёрен», которые являются критериями оценки сорто-, экотипа и продуктивности генотипов пшеницы озимой. Апробированный метод к-средних (без ограничения числа групп) в качестве способа оценки морфо-биологической отдаленности сортообразцов позволяет упростить процесс планирования схем скрещиваний для более эффективной работы по гибридизации.

● ЛИТЕРАТУРА

1. Кацко И. А. Практикум по анализу данных на компьютере [Текст] / И. А. Кацко, Н. Б. Паклин. – Москва: КолосС, 2009. – 278 с.
2. Нейронная сеть как способ классификации исходного материала озимой пшеницы [Текст] // И. А. Русанов, Т. Г. Ващенко, Н. Т. Павлюк, Г.Г. Голева // Вестник Воронежского государственного аграрного университета. – 2010. – № 3 (26). – С. 27-31.
3. Проблема классификации селекционного материала [Текст] / А.Г. Буховец, И. А. Русанов, Г. Г. Голева, Н. Т. Павлюк // Биологические основы и методы селекции и семеноводства сельскохозяйственных культур: сборник научных трудов / ВГАУ. – Воронеж: ВГАУ, 2006. – С. 35-47.
4. Халафян А. А. Statistica 6. Математическая статистика с элементами теории вероятностей [Текст]: учебник / А. А. Халафян. – Москва: Бином, 2010. – 496 с.

● REFERENCES

1. Kacko I. A. Praktikum po analizu dannyh na komp'yutere [Tekst] / I. A. Kacko, N. B. Paklin. – Moskva: KolosS, 2009. – 278 s.
2. Nejronnaya set' kak sposob klassifikacii iskhodnogo materiala ozimoy pshenicy [Tekst] // I. A. Rusanov, T. G. Vashchenko, N. T. Pavlyuk, G. G. Goleva // Vestnik Voronezhskogo gosudarstvennogo agrarnogo universiteta. – 2010. – № 3 (26). – S. 27-31.
3. Problema klassifikacii selekcionnogo materiala [Tekst] / A.G. Buhovec, I. A. Rusanov, G. G. Goleva, N. T. Pavlyuk // Biologicheskie osnovy i metody selekcii i semenovodstva sel'skhozajstvennyh kul'tur: sbornik nauchnyh trudov / VGAU. – Voronezh: VGAU, 2006. – S. 35-47.
4. Halafyan A. A. Statistica 6. Matematicheskaya statistika s ehlementami teorii veroyatnostej [Tekst]: uchebnik / A. A. Halafyan. – Moskva: Binom, 2010. – 496 s.