

Российская академия сельскохозяйственных наук

СЕВЕРО-КАВКАЗСКИЙ ЗОНАЛЬНЫЙ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
САДОВОДСТВА И ВИНОГРАДАРСТВА

КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

С.Н. ЩЕГЛОВ

ПРИМЕНЕНИЕ БИОМЕТРИЧЕСКИХ МЕТОДОВ
ДЛЯ УСКОРЕНИЯ СЕЛЕКЦИОННОГО ПРОЦЕССА
ПЛОДОВЫХ И ЯГОДНЫХ КУЛЬТУР

Краснодар 2005

УДК 57.087.1:631.527

ББК 41.366

Щ 334

Рецензент:

Доктор сельскохозяйственных наук, профессор,
заслуженный деятель науки Кубани и России,

И.А. Драгавцева

Щеглов С.Н.

Щ 334 Применение биометрических методов для ускорения селекционного процесса плодовых и ягодных культур. Краснодар: СКЗНИИСиВ; Кубанский гос. ун-т, 2005. 106 с.

ISBN 5-8209-0383-8

В монографии рассматриваются вопросы теории и практики использования математических методов и моделирования процессов, протекающих в живых системах. На многочисленных примерах показано использование современных методов анализа экспериментальных данных, существенно повышающих эффективность исследований в самых разных областях селекции. Автор описывает конкретные математические методы обработки результатов полевого опыта и экспериментальных данных.

Адресуется научным работникам специализированных плодовых предприятий, аспирантам, студентам высших образовательных учреждений.

Утверждено ученым советом ГНУ РАСХН СКЗНИИСиВ.

УДК 57.087.1:631.527

ББК 41.366

© Щеглов С.Н., 2005

© Кубанский государственный университет, 2005

ISBN 5-8209-0383-8

© Северо-Кавказский зональный НИИ садоводства и виноградарства, 2005

ВВЕДЕНИЕ

В настоящее время уже недостаточно использования в селекционной работе методов полевой, приближенной оценки гибридного материала по основным хозяйственно-ценным свойствам. Должны использоваться генетико-статистические методы реализации селекционного процесса, направленного на ускоренный отбор отдельных признаков и свойств. В связи с этим встает вопрос о необходимости поиска новых путей и разработки методологии для раскрытия и реализации биологического потенциала растений.

Характерной особенностью математизации биологии в наши дни является стремительный рост спроса на такие методы обработки эмпирического материала, которые обеспечивают комплексный подход к познанию живых организмов. Общеизвестная истина, что в природе все взаимосвязано, в исследовательской работе не всегда учитывается. Ярче всего это проявляется при организации наблюдений и экспериментов по принципу единственного фактора. Согласно этому принципу, изменениям подвергается один признак, а остальные признаки стараются поддерживать на постоянном уровне. Даже в лабораторных условиях трудно строго соблюдать принцип единственного фактора, не говоря об опытах в природных условиях. И все-таки, следуя этому принципу, вводили различные ограничения, оправдывали существенные упрощения, разрабатывались искусственные схемы исследований и т.д. В результате была допущена методологическая ошибка: сознательно игнорировался принцип единства живой природы. Нет таких методов, в том числе

математических, способных исправить ошибки, которые обусловлены упрощениями методологического характера. Конечно, здесь нельзя винить исследователей в желании намеренно упрощать реальную ситуацию. Принцип единственного фактора вошел в биологию по объективным причинам, среди которых первостепенно значение имеет факт, что человек не может непосредственно анализировать одновременное взаимодействие многих факторов. Однако такие математические методы, разработанные с всесторонним учетом принципа единства живой природы и возможности практической реализации их на ЭВМ, появились в последние десятилетия. И наконец, многочисленны случаи, где принцип единственного фактора оказывается целесообразным и даже незаменимым. Но в большинстве случаев (особенно это относится к экологическим исследованиям) полученные таким образом знания нельзя распространять на другие условия фона. Поэтому понятен усиленный интерес к методам многомерной математической статистики, которым не присущи методологические противоречия по отношению к принципу единства природы.

Методы многомерной статистики интенсивно развиваются. Углубляется понимание смысла возникающих задач, разрабатываются новые, более эффективные методы. С появлением персональных компьютеров существенно расширился круг исследователей и практических работников, применяющих аппарат многомерного статистического анализа. Этот процесс не может быть отнесен только к математической или прикладной статистике. Использование методов многомерной статистики предполагает обращение к системному анализу рассматри-

ваемого явления, основных его составляющих и связей, принятие решения о характере установленных закономерностей. Кроме того, программно-алгоритмическое обеспечение такого анализа имеет отношение к методам искусственного интеллекта (обобщение данных с помощью факторного и кластерного анализа, распознавание с помощью дискриминантного анализа).

Именно этот многогранный характер современного развития многомерной статистики, способствующий слиянию собственно статистики, системного анализа и принятия решений, информатики и искусственного интеллекта, обеспечивает постоянный рост числа специалистов, использующих эту методологию.

1. МЕТОДЫ МОДЕЛИРОВАНИЯ В ГЕНЕТИКО-СЕЛЕКЦИОННЫХ ИССЛЕДОВАНИЯХ

В настоящее время ввиду чрезвычайной сложности объектов генетико-селекционных исследований наиболее эффективными методами изучения признаны системный анализ и моделирование различной степени сложности.

Системный подход направлен на целостный охват изучаемых объектов и явлений и в то же время на углубленный анализ их структуры, на определение максимального числа взаимодействующих факторов и характера их взаимосвязей, а также на обобщенное агрегированное их представление.

Слияние методов системного анализа с методами моделирования приводит к интегральному методу системного моделирования – комплексу средств, методов и приемов изучения и конструирования сложных биологических структур и биологических систем.

Математическое моделирование может оказаться полезным при соблюдении нескольких условий.

1. Для составления математической модели биологического явления или процесса необходимо достичь определенного уровня знаний об этом процессе. В биологии имеется еще много явлений, которые мы понимаем пока поверхностно.

2. Неизбежное при математическом моделировании упрощение схемы моделируемого явления должно быть регламентировано определенными требованиями. Эти требования вытекают из знания существа процесса. Слишком отвлеченная модель может оказаться малополезной.

3. Модель должна базироваться на отражении определенных принципов, которым подчиняется моделируемая система или явление. Собственно модель с помощью адекватного математического языка отражает соблюдение определенного принципа или закона соответствующими параметрами или функциями, относящимися к моделируемой системе.

Общепринятой классификации моделей не существует. По разным критериям они могут разделяться на: вербальные, графические, иконографические, математические, физические (по исполнению); статические, экспертные, стохастические (по объему исходной информации); компаративные, экстраполяционные, прогнозные (по практическому назначению) и т.д.

Наиболее многообещающими для моделирования при проведении генетико-селекционных исследований признаны математические методы (с использованием ввиду сложности исследуемых объектов компьютерной техники) и математические модели.

Уже длительное время главной проблемой практического применения математического моделирования в биологии является наполнение разработанных моделей конкретной и качественной информацией. Точность и полнота первичной информации, реальные возможности ее сбора и обработки во многом определяют выбор типов прикладных моделей. Однако исследования по моделированию биологических процессов выдвигают новые требования к системе информации.

В зависимости от моделирования объектов и назначения моделей используемая в них исходная информация имеет существенно различный характер и происхождение.

ние. Она может быть разделена на две категории: информация о прошлом развитии и современном состоянии объектов (селекционные наблюдения и их обработка) и информацию о будущем развитии объектов, включая данные об ожидаемых изменениях их внутренних параметров и внешних условий (прогнозы). Вторая категория информации – результат самостоятельных исследований, которые также могут выполняться посредством моделирования.

По целевому назначению модели делятся на теоретико-аналитические, используемые при исследовании общих свойств и закономерностей, и прикладные, применяемые в решении конкретных задач (анализа, прогнозирования, управления) (Кольцов Ю.В., Коршиков А.Ф., Матвиенко В.Ю., 2001).

Простейшими являются модели масштабного типа. В таких моделях соблюдается геометрическое подобие оригинала и модели, но натурные значения всех размеров умножаются на постоянную величину – масштаб моделирования. Большие объекты представляются в уменьшенном виде, а малые – в увеличенном.

В аналоговых моделях исследуемые процессы изучаются не непосредственно, а по аналогичным явлениям, т.е. по процессам, которые имеют иную физическую природу, но описываются такими же математическими соотношениями. Для такого моделирования используются аналогии между механическими, тепловыми, гидравлическими, электрическими и другими явлениями.

Реальные (натурные, аналоговые) модели, если таковые удастся создать, отражают самые существенные черты оригинала. Например, селекционный участок мо-

жет служить моделью полевых посадок. Однако создание реальных моделей сопряжено с большими техническими трудностями, так как пока еще не удается достичь точного воспроизведения оригинала.

Знаковая модель представляет собой условное отображение оригинала с помощью математических выражений или подробного описания. Наибольшее распространение в современных генетико-статистических исследованиях получили концептуальные и математические модели и их многочисленные разновидности. Разновидности концептуальных моделей характеризуются подробным описанием системы (научный текст, схема системы, таблицы, графики и т.д.). Математические модели – более эффективный метод изучения биологических систем, особенно при определении количественных показателей.

Остановимся более подробно на характеристике таких классов математических моделей, которые в наибольшей степени отражают особенности методологии и техники моделирования (Ляпунов А.А., Багриновская Г.П., 1975).

В соответствии с общей классификацией математические модели подразделяются на функциональные, структурные и структурно-функциональные. В исследованиях на уровне биосферы чаще используются структурные модели, поскольку для планирования и управления большое значение имеют взаимосвязи подсистем. Функциональные модели широко применяются мутационной селекции, когда на поведение объекта («выход») воздействуют путем изменения «входа», например, изменяя дозу мутагена.

По характеру отражения причинно-следственных связей различают модели жесткодетерминированные и модели, учитывающие случайность и непосредственность. Необходимо различать неопределенность, описываемую вероятностными законами, и неопределенность, для описания которой законы теории вероятности неприменимы. Второй тип неопределенности гораздо более сложен для моделирования.

По способам отражения фактора времени математические модели делятся на статические и динамические. В статических моделях все зависимости относятся к одному моменту или периоду времени. Динамические модели характеризуют изменения биологических процессов во времени. По длительности рассматриваемого периода различаются модели краткосрочного (до года), среднесрочного (до 5 лет), долгосрочного (10–15 и более лет) прогнозирования и планирования. Само время в математических моделях может изменяться либо непрерывно, либо дискретно. Справедливость (объективность) модели во временном аспекте определяется организационным уровнем объекта моделирования.

Модели биологических процессов чрезвычайно разнообразны по форме математических зависимостей. Особенно важно выделить класс линейных моделей, наиболее удобных для анализа и вычислений и получивших вследствие этого большое распространение.

По соотношению экзогенных и эндогенных переменных, содержащихся в самой моделируемой системе и отражающих ее зависимость или связь с внешними факторами, модели могут разделяться на открытые и закрытые. Полностью открытых моделей не существует; мо-

дель должна содержать хотя бы одну эндогенную переменную. Полностью закрытые математические модели, т.е. не содержащие экзогенных переменных, исключительно редки; их построение требует полного абстрагирования от «среды», т.е. серьезного огрубления реальных биологических систем, всегда имеющих внешние связи. Подавляющее большинство математических моделей занимает промежуточное положение и различаются по степени открытости.

Для моделей важно деление на агрессивные и детализированные. В зависимости от того, включают ли модели пространственные факторы и условия или не включают, различают модели пространственные и точечные.

Выделяют также статические и динамические модели. Статические модели формализуют связь между показателями без учета переменной времени. Динамические модели используются для оценки явлений в развитии.

Важнейшей задачей моделирования является прогнозирование и управление объектом, однако находят применение и модели без управления, а также оптимизационные (для управления одним или несколькими параметрами).

Таким образом, общая классификация математических моделей достаточно неустойчива. С развитием генетико-селекционных исследований проблема классификации применяемых моделей усложняется. Наряду с появлением новых типов моделей (особенно смешанных типов) и новых признаков их классификации, осуществляется процесс интеграции моделей разных типов в более сложные модельные конструкции.

Важно, что использование компьютерных методов моделирования позволяет проанализировать гигантское количество вариантов, перебрать большое количество условий, проверить качественные и количественные вариации параметров модели, но только от исследователя зависит, каким образом будет производиться моделирование, какие свойства регионального объекта будут учтены. Это в конечном счете влияет на адекватность модели и реальность полученных результатов, поэтому правильный выбор метода и критериев адекватности особенно важен.

Преимущества математических моделей состоят в том, что они позволяют делать предсказания, которые можно сравнить с реальными данными, поставив эксперимент или проведя необходимые наблюдения, и установить необходимый уровень генерации, исходя из цели практического использования модели. Математические модели дают также возможность использовать всю совокупность накопленных знаний о поведении взаимосвязей, так что к логически связанным суждениям об изучаемой системе ученый приходит, не повторяя всех предыдущих исследований.

Недостатки математических моделей заключаются во внешней сложности символической логики. Эта сложность отчасти неизбежна – если изучаемая проблема сложна, вполне возможно (но все же не обязательно), что сложным окажется и математический аппарат, необходимый для ее описания. Недостаток математических моделей связан с тем искажением, которое можно привести в саму проблему, придавая слишком большое значение математическому аппарату.

Динамические модели предназначены для прогнозирования и оперативного управления производственным процессом с учетом складывающейся агрометеорологической обстановки. В основе динамического моделирования – описание системы с помощью обыкновенных дифференциальных уравнений в частных производных, параметры которых определяют по эмпирическим данным. Известны динамические модели формирования урожая, диагностики минерального питания растений, накопления и распада в агроэкосистемах пестицидов, нефтепродуктов, процессов в мелиорируемых почвах.

Модели теории исследования операций решают задачу оптимального управления в условиях, когда доступные ресурсы ограничены, т.е. регламентированы значения переменных. Оптимизационные модели – основа автоматизированного проектирования сельскохозяйственных технологий. Известны модельные разработки планирования хозяйственного использования земель в условиях радиоактивного загрязнения, оптимизации агротехнологий по принципу «игр с природой».

Важнейшим способом имитации сложных систем является создание многомерных моделей.

Многомерные модели часто представляются в виде композиции моделей меньшей размерности. Это позволяет разбить общую задачу моделирования исследуемого объекта на меньшие подзадачи, модули. Важно отметить, что общая адекватность модели зависит от адекватности решения каждой подзадачи и адекватности реакции модели на изменение каждого ее параметра. Иными словами, не только поведение каждого модуля, но и поведение модели в целом определяет ее адекватность. При этом

очевидно, что каждый модуль должен учитывать не только общие закономерности моделируемой части, но и некоторые, возможно, незначительные в отдельном модуле зависимости, существенно влияющие на поведение системы в целом (Дженифферс Дж., 1981).

Необходимость достижения компромисса между требованиями к сложности модели и ее адекватностью не позволяет решить проблему масштабируемости моделей в полной мере.

При создании масштабных моделей сложных объектов комплексное моделирование дает возможность использовать уже имеющиеся модели отдельных частей для последующего объединения их в общую комплексную модель. В отличие от многомерного моделирования, при комплексном моделировании можно объединять различные разнородные модели с целью получения общей модели объекта исследования.

Так как каждый метод моделирования имеет свои недостатки, отдаляющие модель от реальных объектов, то имеет смысл использовать комплекс моделей, при создании которых применялись различные методы и подходы. Это позволяет детально рассмотреть обобщенную модель, выявить особенности, которые проявляют себя не при всех подходах к моделированию, но имеют большое значение при реальном взаимодействии с объектом исследования. Таким образом, происходит объединение положительных качеств разнородных моделей в комплексной модели.

Итак, применение моделей дает возможность проводить контролируемые эксперименты в ситуациях, где экспериментирование на реальных объектах было бы

практически невыполнимо или экономически нецелесообразно. Не представленное экспериментирование с системой обычно состоит в варьировании ее некоторых параметров; при этом, поддерживая все остальные параметры неизменными, наблюдают результаты эксперимента. Для большинства систем, с которыми приходится иметь дело исследователю, это или практически недоступно, или слишком дорого, или и то и другое вместе. Когда ставить эксперимент на реальной системе слишком дорого и (или) невозможно, зачастую строят модель, на которой необходимые эксперименты могут быть проведены с относительной легкостью и недорого. При экспериментировании с моделью сложной системы имеется возможность получить больше знаний о ее внутренних взаимодействующих факторах, чем манипулируя реальной системой; это достигается благодаря описанности структурных элементов модели, контролируемости ее поведения, варибельности параметров и т.п.

К современным сортам растений предъявляется ряд требований, определяемых характером современного сельскохозяйственного производства. Быстрое повышение продуктивности культур в равной степени обусловлено как достижениями селекции, так и улучшением технологии возделывания растений.

Успехи практической селекции показывают большие возможности человека по перестройке генетических систем существующих сортов и синтезу новых генетических систем с наперед заданными свойствами. Роль материального образа, к которому стремится селекционер в своей работе по выведению новых форм, играет модель

сорта. Различают директивную и биологическую модели сорта.

В директивной модели на основе достигнутого селекцией уровня продуктивности, прошлой селекционной истории, состояния отрасли сельского хозяйства, экономических критериев устанавливаются количественные требования к уровню продуктивности, структуре элементов урожая, устойчивости к неблагоприятным условиям, вредителям и болезням, а также к качеству урожая и материальным затратам на удобрение. Эта модель объединяет усилия селекционеров, организует и направляет их работу. В ней обобщаются требования современного сельскохозяйственного производства и создается желаемый образ будущих сортов.

Биологическая модель сорта строится исходя из результатов глубокого изучения биологии растения. Она может быть фактической и перспективной. В фактической модели на основе изучения морфологии и физиологических особенностей существующих организмов фиксируется достигнутый уровень продуктивности, структура компонентов урожая, их соотношение. Перспективная биологическая модель призвана вскрыть генетический потенциал организма и предложить конкретные пути реализации директивной модели. В перспективной модели устанавливается желательный физиологический тип растений, задается соотношение частей организма (например, соотношение корней, стеблей, листьев), скорости прохождения фаз органогенеза, показатели его устойчивости и экологической пластичности, способность создавать высокопродуктивные агроценозы, отвечать требованиям технологии возделывания – иметь опреде-

ленную высоту, массу, сроки созревания, отзывчивость на внесение удобрений. Перспективную биологическую модель сорта невозможно составить, не имея данных о наличной генетической изменчивости по комплексу хозяйственно ценных признаков. Данные генетического анализа, наряду с результатами биохимического, физиологического и технологического изучения исходного материала, служат основой для разработки перспективной биологической модели сорта.

Перспективная биологическая модель сорта определяет направление селекционного процесса и фиксирует внимание селекционера на узловых проблемах перестройки наследственных свойств растений. Однако биологическая модель, векторизуя селекционный процесс и определяя его конечные результаты, вместе с тем не отвечает на вопрос о том, какими способами и путями следует достигать намеченных целей.

Современная экспериментальная генетика разработала комплекс способов эффективной перестройки наследственной основы организмов. В их числе различные способы скрещивания родительских форм, искусственного отбора, регулирования гетерозиса, искусственного вызывания мутаций, экспериментальной полиплоидии. Основу для успешной селекционной работы создает наличная генетическая изменчивость. Поэтому важной задачей селекционера является создание достаточного запаса наследственной изменчивости в селекционируемой популяции по интересующим его признакам. С этой целью на начальных этапах селекции привлекается генетический материал из мировой коллекции, проводится генетический анализ прерывистой и непрерывной изменчивости,

организуется программа скрещиваний. Выбор селекционной стратегии – ключевой момент осуществления любого биологического проекта сорта.

В основе современных методов планирования и прогнозирования результатов селекционного процесса лежат данные генетического анализа изменчивости. Оценку частоты и степени доминирования отдельных аллелей кладут в основу программы изоляции, скрещивания родителей и отбора в последующих поколениях нужных комбинаций дискретных аллелей, составляющих желательный генотип. Несколько иначе обстоит дело при составлении селекционных программ, направленных на повышение элементов продуктивности растений, а также на благоприятное изменение их физиологических и биохимических признаков. Непрерывная вариационная кривая по этим признакам возникает вследствие расщепления большого числа генов и модифицирующего влияния условий окружающей среды (Рокицкий П.Ф., Савченко В.К., Добина А.И., 1977).

Генетический анализ количественных признаков позволяет вскрыть структуру генетической изменчивости, оценить степень доминирования генов родителей. Эти данные служат основой для прогнозирования результатов селекционных мероприятий, выбора селекционной стратегии. Селекционную работу следует начинать только в том случае, если в исходном материале имеется достаточная генетическая изменчивость. Если же генетическая изменчивость в исходном материале мала или вообще отсутствует, то сначала необходимо позаботиться о том, чтобы увеличить генетическую вариацию за счет комбинативной или мутационной изменчивости. В про-

тивном случае любые селекционные мероприятия, как бы тщательно они ни проводились, не принесут успеха.

При наличии достаточной генетической изменчивости в исходной популяции важную роль для выбора селекционной стратегии играет соотношение аддитивного и неаддитивного компонентов дисперсии. Если аддитивный компонент достаточно велик или преобладает в генетической вариации, то можно рассчитывать на существенное изменение средней величины признака путем применения методов массивного, семейного или индивидуального отбора. Если же аддитивный компонент генетической изменчивости по изучаемому признаку мал, а преобладающими являются компоненты, связанные с доминированием и эпистазом, то нет оснований надеяться, на существенное изменение уровня количественного признака в процессе отбора. В этом случае селекционный процесс должен быть ориентирован на создание и использование в производственном масштабе гетерозисного эффекта (Меттлер Л., Грегг Т., 1972).

Необходимо подчеркнуть, что основные сельскохозяйственные растения в процессе селекционной проработки на протяжении многих поколений подвергались интенсивному искусственному отбору. В основном это приводило к тому, что внутри сортов уменьшилось генетическое разнообразие, поэтому отбор перестал приносить желаемые результаты.

Синтез сортов, отвечающих современным требованиям, – процесс длительный. Для этого иногда требуется получить несколько поколений. Синтез сорта включает несколько этапов: привлечение и изучение исходного материала; скрещивание родительских форм; отбор, завер-

шающийся формированием генетического ядра сорта; использование нового сорта в сельскохозяйственном производстве.

Правильный выбор исходного материала является решающим фактором успеха селекции. На это неоднократно указывал Н.И. Вавилов. Ему же принадлежит идея и труд по созданию обширной коллекции мировых растительных ресурсов. Опыт селекции свидетельствует также о большой ценности генофонда местных сортов и популяций как наиболее адаптированных к конкретным природно-экологическим условиям. Однако рассчитывать на успех, работая только с местными сортами и популяциями, нет оснований. Все более широкое применение в селекционном процессе находят дикие предки современных культурных растений. Они выступают в качестве доноров генов, обеспечивающих иммунитет к болезням и вредителям, устойчивость к неблагоприятным условиям среды (Савченко В.К., 1986).

Важное значение имеет всестороннее исследование исходного материала. Оно должно включать генетический анализ его прерывистой и непрерывной изменчивости, изучение физиологических, биохимических, онтогенетических, иммунных и иных особенностей и свойств. В селекционный процесс желательно включать самый обширный исходный материал из различных эколого-географических районов, так как это расширяет возможности комбинативной изменчивости.

Кооперация между генетикой и селекцией должна осуществляться также по линии разработки селекционного процесса и внедрения таких систем в существующих селекционных центрах. Эта система должна вклю-

чать способы максимально полной мобилизации исходного материала для селекции, использование разнообразных методов генетического анализа для его изучения и оценки, разные способы обоснования селекционной стратегии, методы направленного генетического синтеза. Большое значение имеет разработка и внедрение в селекционный процесс экспресс-методов генетического анализа, его автоматизация. Трудно переоценить роль и значение в оптимизации селекционного процесса специализированных вычислительных центров. Такие центры, оснащенные современными ЭВМ, позволяют создавать банки генетико-селекционных данных и характеристик, производить анализ генетико-статистических параметров, расчет ожидаемого сдвига средних величин при проведении искусственного отбора. Количественные оценки и критерии облегчают принятие тех или иных селекционных решений.

В последние годы ученые СКЗНИИСиВ довольно широко используют классические биометрические методы и их модификации в своих исследованиях.

Так, И.А. Драгавцева и А.Б. Дьяков (2001) показали возможности биометрических методов для оценки сортов плодовых культур, приспособленных к конкретным почвенно-климатическим условиям. В своей работе они использовали средние величины признаков, среднеквадратические отклонения, коэффициенты вариации, коэффициенты регрессии на индексы среды, величину эквалент и коэффициенты детерминации оценок урожайности индексами среды. Ими показано, что при одновременном использовании комплекса параметров, характеризующих разные аспекты приспособленности генотипов, можно

выявлять особенности адаптивности не только отдельных сортов, но и их групп, различающихся по биологическим особенностям.

Эти же исследователи предложили метод сравнительной оценки экологической адаптивности видов и сортов плодовых культур (2002). Важной особенностью разработанного метода является то, что строится график, на оси абсцисс которого откладываются оценки урожайности генотипа, а по оси ординат – величины того же признака сопоставляемого с ним генотипа. При этом на график наносится не только линия регрессии, но и линия, которая выражала бы характер связи между переменными, если бы нормы реакций изучаемых объектов были полностью тождественны.

И.А. Драгавцева и А.Б. Дьяков (2003), давая обзор способов оценки стабильности сортов, высказали мнение о том, что для подобных оценок непригодны комплексные параметры, в том числе и коэффициент вариации. Они предлагают использовать абсолютный показатель variability признака – дисперсию или стандартное отклонение.

И.А. Драгавцева, Н.М. Запорожец, Е.В. Луценко и Н.Е. Луценко (2003) разработали математическую модель, относящуюся к классу моделей распознавания образов и принятия решений. Она позволяет оценить степень адекватности условий в пункте выращивания конкретного набора плодовых культур и провести картографическую визуализацию результатов решения поставленной задачи с применением технологий существующих геоинформационных систем.

В.А. Драгавцев, И.А. Драгавцева, Л.М. Лопатина (2003) предложили организацию экологических испытаний дифференцировать согласно конкретным задачам, которые можно решать с помощью экологических испытаний. Такие задачи могут быть направлены на изучение реакции сортов в зависимости от особенностей: 1) почвенно-климатических условий; 2) метеоусловий в одной почвенно-климатической зоне при неполивном земледелии; 3) отдельных управляемых человеком факторов среды (доз удобрений, густоты посева, полива, химической обработки и пр.).

При решении этих трех задач они предлагают использовать различный математический подход к оценке адаптивности сортов. Для этого сорта должны быть испытаны в 7–10 достоверно отличающихся друг от друга условиях. В первом и втором случае этого можно добиться, высевая сорта 2–3 года в 4–5 разных точках. Решая третью задачу при комплексном изучении реакции сортов на отдельные управляемые факторы, необходимо прибегать к теории планирования эксперимента с выходом на многофакторные регрессионные уравнения 1-го, 2-го, 3-го и последующего порядка, включая взаимодействия факторов (Лопатина Л.М., 1986; Лопатина Л.М., Ключков С.А., 1995; Лопатина Л.М., Кудряшов И.Н., 1996; Смиряев А.В., Гохман М.В., 1985).

Генотипы, адаптированные к условиям лимита факторов среды и слабоадаптированные к безлимитным условиям, показывают в разных экологических зонах примерно одинаковые уровни продуктивности, что характеризует их как низкопластичные. Генотипы, слабоадаптированные к условиям лимита факторов среды и высоко-

адаптированные к безлимитным условиям, ведут себя как отзывчивые на благоприятные факторы среды, имеют положительные коэффициенты регрессии на условия среды и относятся к пластичным. Пластичным является и генотип с отрицательной регрессией на условия среды, т.е. высокоадаптированный к среде с лимитами ее факторов и слабоадаптированный к безлимитным средам.

Если принять за генетическую пластичность сорта степень его реакции на изменяющиеся условия, то высокопластичным будет тот сорт, который быстро увеличивает изучаемый признак с улучшением условий, но и быстро снижает его в плохих условиях. Обычно это высокоинтенсивные сорта, пригодные для возделывания в хороших условиях при высокой культуре земледелия.

Низкопластичные сорта, которые в меньшей степени реагируют на изменение внешних условий, больше пригодны для жестких условий выращивания, где они не снижают урожай или снижают его в меньшей степени, чем сорта высокопластичные.

При использовании регрессионных моделей для оценки реакции сорта на изменение факторов внешней среды коэффициент регрессии выступает как показатель пластичности сорта.

Те же В.А. Драгавцев, И.А. Драгавцева, Л.М. Лопатина (2003) предложили новый метод экспрессной оценки генетической корреляции между генетическими системами аттракции и адаптивности (или любыми другими). Он основан на том факте, что экологическая линия регрессии координатных признаков отражает истинную (эталонную) реакцию двух признаков на улучшение условий роста данного генотипа (сорта).

И.А. Драгавцевой и Л.М. Лопатиной (2001) изучался отклик плодовых культур на изменения условий среды. Они предложили методику интегральной оценки экологической адаптивности сортов на примере косточковых культур, предусматривающую построение регрессионных моделей связей параметров продуктивности с лимитирующими факторами, что позволяет рассчитать урожай сорта по средним многолетним данным погодных условий зоны, а также величину урожая в предполагаемых экстремальных условиях (используя пороговые коэффициенты).

Л.М. Лопатиной (2001) разработана методика дифференцированной оценки эколого-генетической адаптивности плодовых культур и проведено моделирование причинно-следственных связей параметров генотипа с внешними факторами среды. В основе методики лежит многофакторный регрессионный анализ и оценка путевых коэффициентов на пути реализации генотипического потенциала в фенотипе.

Г.Н. Теренько, Н.К. Шафоростова, И.А. Мазурик (2002) на примере яблони дали оценку плодоношения деревьев по данным ежегодных урожаев в течение жизни дерева. Они использовали сглаживание ряда данных по скользящей средней, определяли средний прирост и некоторые производные характеристик прироста, что дает возможность наметить тенденцию к изменению ряда.

Н.Н. Сергеева и В.О. Улитин (2003) с помощью многофакторного корреляционно-регрессионного анализа изучили влияние на урожайность яблони климатических факторов и доз внесенных удобрений. В результате им удалось построить уравнение, описывающее влияние

цикличности плодоношения, возраста сада и фона удобрений на урожайность сада в результате действия как отдельных факторов, так и их взаимодействия.

Г.Н. Ключникова и В.О. Улитин (2003), исследуя влияние температуры и осадков на длительность фенологических периодов, урожайность и сахаристость винограда, установили связь между этими показателями у различных сортов.

Наряду с классическими методами статистической обработки данных начинается активное использование современных многомерных методов.

По данным многолетних учетов Ю.А. Волчков, А.П. Кузнецова и А.В. Алексеенко (2002) провели оценку коллекции сортоформ черешни на основе комплекса показателей уровня и стабильности полевой устойчивости к коккомикозу и монилиозу. Предлагаемый подход заключается в установлении такого критерия селекционной ценности образца, который включает в себя не только относительно низкий средний балл поражения болезнью, но и показатели его устойчивости в варьирующих условиях среды и характеристики популяции патогена. Ими предложен следующий алгоритм: 1) объединить в линейную комбинацию средний многолетний балл поражения, коэффициент линейной регрессии балла поражения на индекс среды (средний балл по коллекции в данном году) и дисперсию отклонений годовых баллов поражения данного сорта от линии регрессии; 2) определить координаты изучаемых форм в ортогональном пространстве главных компонент; 3) оценить в этом пространстве расстояние каждой из сортоформ до точки, отвечающей «идеальному» объекту (в качестве координат которого

используют наименьший в совокупности данных средний многолетний балл поражения, наименьшее значение коэффициента регрессии и наименьшее значение дисперсии отклонения от линии регрессии); 4) по наименьшему расстоянию до точки «идеального» объекта выбрать лучший по устойчивости сорт, гибрид (или их группу).

В.В. Яковенко, В.Н. Подорожный, С.Н. Щеглов и В.И. Лапшин (2002) изучали меж- и внутрисемейную изменчивость комплекса селекционно ценных признаков гибридных семей земляники. Использование метода главных компонент и последующая кластеризация методом Уорда позволили им выделить группы перспективных гибридных семей, провести отбор внутри каждой семьи и выявить лучшие генотипы по хозяйственно ценным признакам.

Аналогичный подход применил В.И. Лапшин (2002) при изучении изменчивости морфологических признаков листа в гибридных семьях земляники. Автор, продолжив исследования вегетативной изменчивости сортов земляники садовой и популяций земляники лесной (Горшкова В.В., Щеглов С.Н., 1997; Щеглов С.Н., 2000), приходит к выводу, что различная степень уклонений морфологических признаков в гибридном потомстве служит показателем уклонений по другим селекционно значимым признакам и при условии расширения набора морфологических признаков можно получить надежные характеристики для идентификации гибридов по хозяйственно ценным признакам.

2. МНОГОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ

Классические статистические методы хорошо разработаны и описаны во многих литературных источниках (Доспехов В.А., 1979; Лакин Г.Ф., 1990 и др.). В наших работах уже приводился краткий обзор наиболее часто применяемых методов статистической обработки данных (Щеглов Н.И., Щеглов С.Н., 1999; Щеглов С.Н., 2002).

В данной главе мы подробно остановимся на описании некоторых многомерных статистических методов, предполагающих описание объектов с помощью комплекса признаков. Более полный их обзор можно найти в ряде научных изданий (Айвазян С.А., Бухштабер В.М., Енюков И.С. и др., 1989; Андерсон Т., 1963; Андрукович П.Ф., 1973; Иберла К., 1980; Ким Дж., Мьюллер Ч., 1989; Кендалл М., Стьюарт А., 1976; Клекка У.Р., 1989; Олдендерфер М.С., Блэшфилд Р.К., 1989; Сажин Ю.В., Басова В.А., 2002); Тюрин В.В., Морев И.А., Волчков Ю.А., 2003).

2.1. Метод главных компонент и факторный анализ

При решении задач методом главных компонент главное заключается в выделении линейных комбинаций случайных величин, имеющих максимально возможную дисперсию.

Методы факторного анализа позволяют решать следующие четыре задачи. Первая задача заключается в «сжатии» информации до обозримых размеров, т. е. в извлечении из исходной информации наиболее существенной части за счет перехода от системы исходных пере-

менных к системе обобщенных факторов. При этом выявляются неявные, непосредственно не измеряемые, но объективно существующие закономерности, обусловленные действием как внутренних, так и внешних причин. Вторая задача сводится к описанию исследуемого явления значительно меньшим числом m обобщенных факторов (главных компонент) по сравнению с числом исходных признаков. Обобщенные факторы – это новые единицы измерения свойств явления, линейно зависящие от исходных, непосредственно измеряемых признаков. Третья задача связана с выявлением взаимосвязи наблюдаемых признаков с вновь полученными обобщенными факторами. Четвертая задача заключается в построении уравнения регрессии на главных компонентах с целью прогнозирования изучаемого явления.

Анализ главных компонент – это метод преобразования данной последовательности наблюдаемых переменных в другую последовательность переменных. Наиболее простой способ пояснить внутреннюю логику метода сводится к его изучению в двухмерном случае. Предположим, что есть две переменные X и Y с совместным нормальным распределением.

Совместное нормальное распределение величин, имеющих положительную корреляцию, представлено на рис. 1 с помощью кривых равных вероятностей. Эти кривые показывают, что благодаря положительной связи между X и Y данные представляют кластер, в котором большие величины X имеют тенденцию соответствовать большим величинам Y (и наоборот). Таким образом, в большинстве случаев точки попадают в первый и третий квадранты и реже – во второй и четвертый. Кривые рав-

ных вероятностей имеют форму эллипсоидов, две оси которых изображены пунктирными линиями. Главная ось P_1 проходит по линии, вдоль которой располагается основная часть данных; вторая ось P_2 – по линии, вдоль которой идет меньшая часть данных.

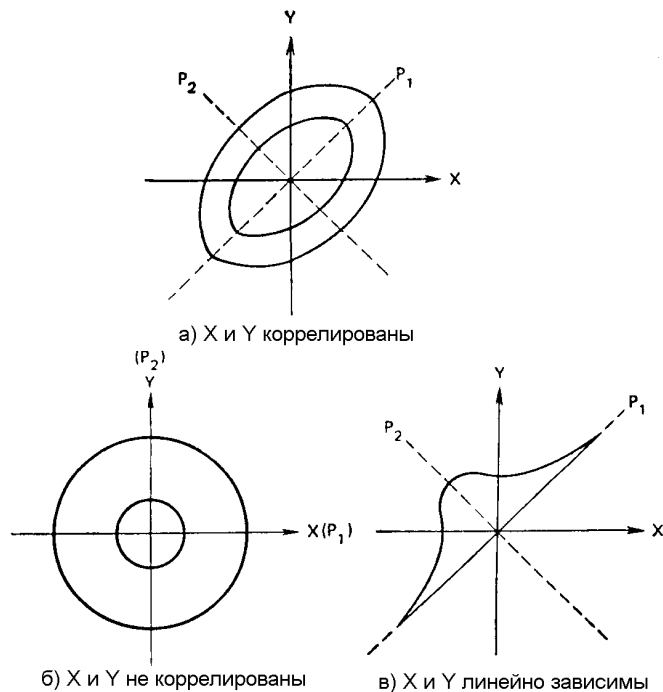


Рис. 1. Главные оси двухмерных распределений
(по Ким Дж., Мьюлеер Ч., 1989)

Теперь предположим, что нужно представить точки в терминах только одной размерности (оси). В этом случае естественно выбрать ось P_1 , потому что в целом она

ближе описывает данные наблюдений. Тогда первая главная компонента есть не что иное, как представление точек, расположенных вдоль выбранной главной оси. Например, точка с единичными значениями X и Y будет иметь координату больше 1 по оси P_1 и меньше 1 по оси P_2 . Если мы описываем каждую точку в терминах P_1 и P_2 (в новой системе координат), потери информации не произойдет. Тем не менее можно сказать, что первая ось (и первая компонента) более информативна в описании точек, так как связь между X и Y становится сильнее. В том случае, когда X и Y связаны линейной зависимостью, первая главная компонента будет содержать всю информацию, необходимую для описания каждой точки. Если X и Y независимы, то главная ось отсутствует и анализ главных компонент не способствует даже минимальному сокращению («сжатию») результатов наблюдений.

Поскольку первая компонента определена таким образом, что основная доля информации содержится именно в ней (дисперсия в направлении этой компоненты максимальна), вторая компонента определяется аналогичным образом при условии, что ее ось перпендикулярна первой. Следовательно, в двухмерном случае после фиксирования первой компоненты вторая становится известна автоматически. Если Y не является линейной функцией от X , то главных компонент будет две (для полного описания совместного распределения необходимы две оси).

При определении главных компонент не обязательно предполагать существование гипотетических факторов. Новые оси являются математическими (линейными)

функциями наблюдаемых переменных. Даже если с помощью анализа главных компонент достигается «сжатие» данных (выделение только нескольких первых компонент), задача состоит не в объяснении корреляции между переменными, а в объяснении максимальной доли дисперсии наблюдений. Однако для рассматриваемого двухмерного случая в факторном анализе потребуется лишь один фактор, и главной задачей будет объяснение корреляций между переменными. Итак, первая задача относится к объяснению дисперсий, а вторая – к объяснению корреляций.

При наличии более двух переменных принцип определения главных компонент тот же. Например, для трехмерного нормального распределения поверхность равной вероятности будет ограничивать овальное тело (эллипсоид), где первая ось – его наибольший диаметр, вторая – пройдет по наибольшему диаметру в плоскости, перпендикулярной первой оси; третья ось будет самой короткой, перпендикулярной первым двум осям.

Основной математический метод получения направлений главных осей основан на нахождении собственных чисел и векторов корреляционной (ковариационной) матрицы.

Главные компоненты могут быть также использованы при классификации объектов наблюдений. В биологических исследованиях стремление полнее изучить исследуемое явление приводит к включению в модель все большего числа исходных переменных, которые зачастую отражают одни и те же свойства объекта наблюдения. Это приводит к высокой корреляции между переменными, т. е. к явлению мультиколлинеарности. При

этом классические методы регрессионного анализа оказываются малоэффективными. Преимущество уравнения, построенного на основе главных компонент, заключается в том, что последние не коррелированы между собой.

В основу факторного анализа и метода главных компонент положена линейная модель. Пусть получена информация, характеризующая n объектов по k признакам. Тогда модель факторного анализа можно представить в виде

$$z_{ij} = \sum_{v=1}^m a_{jv} f_{iv} + a_j w_{ij},$$

где z_{ij} – значение j -го нормированного признака, полученного из факторной модели для i -го объекта;

a_{jv} – весовой коэффициент v -го общего фактора на j -й переменной;

f_{iv} – значение v -го общего фактора, присутствующего более чем в одном признаке i -го объекта;

a_j – весовой коэффициент j -го индивидуального фактора на j -й переменной;

w_{ij} – значение j -го индивидуального фактора, присутствующего только j -му признаку i -го объекта.

Теперь рассмотрим модель метода главных компонент:

$$z_{ij} = \sum_{v=1}^m a_{jv} f_{iv},$$

По структуре она напоминает модель факторного анализа, но имеется существенная разница. Во-первых, отсутствуют специфические факторы. Во-вторых, число

главных компонент k здесь соответствует числу признаков k . Значит, в модели главных компонент исчерпывается вся дисперсия исследуемого процесса.

Главные компоненты являются характеристическими векторами ковариационной матрицы. Множество главных компонент представляет собой удобную систему координат, а их вклад в общую дисперсию характеризует статистические свойства главных компонент. Из общего числа главных компонент для исследования, как правило, оставляют m ($m < k$) наиболее весомых, т. е. вносящих максимальный вклад в объясняемую часть общей дисперсии.

Таким образом, несмотря на то, что в методе главных компонент для точного воспроизведения корреляции и дисперсии между переменными надо найти все компоненты, большая доля дисперсии между переменными объясняется небольшим числом главных компонент. Кроме того, можно по признакам описать факторы, а по факторам (главным компонентам) описать признаки.

Задача факторного анализа состоит в объяснении матрицы коэффициентов корреляции между реально наблюдаемыми показателями и набором новых гипотетических переменных – общих факторов. При заданном числе наблюдаемых показателей p максимально возможное число факторов m определяется соотношением $(p + m) < (p - m)^2$, которое должно выполняться.

После нахождения какого-либо решения, удовлетворяющего в статистическом смысле модели факторного анализа, система факторов подвергается такому «вращению», чтобы полученная в итоге система факторов (столь же хорошо описывающая выборочные данные, что

и исходная) была наиболее удобна с точки зрения содержательной интерпретации.

Сходство анализа главных компонент и факторного анализа заключается в том, что в обоих методах происходит сокращение данных. Зная величину собственных чисел, исследователь может принять, например, решение использовать только две первые компоненты, хотя эти компоненты не объясняют корреляции. Существует еще одно сходство двух методов – они применяются при исследовании взаимной зависимости переменных. Заметим, что в случае некоррелированных переменных главных компонент не существует, так как все они равноправны: каждой соответствует одинаковая доля дисперсии. Если же корреляция между переменными увеличивается, то доля, объясняемая несколькими первыми компонентами, возрастает.

Одним из отличий двух рассматриваемых методов является следующее. Факторный анализ представляет ковариационную структуру в терминах гипотетической модели, в то время как анализ главных компонент сокращает данные посредством использования нескольких линейных комбинаций наблюдаемых переменных. Выбор метода определяется целью исследования. Объяснение корреляций в терминах небольшого числа факторов возможно лишь при введении гипотетической модели. Если же иметь дело с линейными комбинациями переменных, то обращаться к какой-либо модели нет необходимости, при этом латентная факторная структура останется «вещью в себе».

2.2. Кластерный анализ

Кластерный анализ – это совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных X_1, X_2, \dots, X_n . Цель кластерного анализа – образование групп схожих между собой объектов, которые принято называть кластерами. Слово кластер (cluster) английского происхождения, переводится как сгусток, пучок, группа.

В отличие от комбинационных группировок, кластерный анализ приводит к разбиению на группы с учетом всех группировочных признаков одновременно. Например, если каждый наблюдаемый объект характеризуется двумя признаками X_1 и X_2 , то при выполнении комбинационной группировки вся совокупность объектов будет разбита на группы по X_1 , а затем внутри каждой выделенной группы – по X_2 . Такой подход получил название монотетического. Определить принадлежность каждого объекта к той или иной группе можно, последовательно сравнивая его значения X_1 и X_2 с границами выделенных групп. Образование группы всегда связано в этом случае с указанием ее границ по каждому группировочному признаку отдельно.

В кластерном анализе используется иной принцип образования групп, так называемый политетический подход. Все группировочные признаки одновременно участвуют в группировке, т. е. учитываются все сразу при отнесении наблюдения к той или иной группе. При этом, как правило, не указаны четкие границы каждой группы,

а также не известно заранее, сколько групп целесообразно выделить в исследуемой совокупности.

Методы кластерного анализа позволяют решать следующие задачи:

- проведение классификации объектов с учетом признаков, отражающих сущность, природу объектов (решение такой задачи, как правило, приводит к углублению знаний о совокупности классифицируемых объектов);

- проверка выдвигаемых предположений о наличии некоторой структуры в изучаемой совокупности объектов, т. е. поиск существующей структуры;

- построение новых классификаций для слабоизученных явлений, когда необходимо установить наличие связей внутри совокупности и попытаться привести в нее структуру.

Методы кластерного анализа можно разделить на две большие группы: агломеративные (объединяющие) и дивизимные (разделяющие). Агломеративные методы последовательно объединяют отдельные объекты в группы (кластеры), а дивизимные методы расчленяют группы на отдельные объекты. В свою очередь, каждый метод как объединяющего, так и разделяющего типа может быть реализован при помощи различных алгоритмов. В частности, наиболее доступным и поэтому наиболее популярным можно считать иерархический агломеративный кластерный анализ. Результаты реализации таких алгоритмов для большего удобства их визуального анализа принято изображать в виде дендрограммы (дерева).

Процедура иерархического кластерного анализа предусматривает группировку как объектов (строк мат-

рицы данных), так и переменных (столбцов). Можно считать, что в последнем случае роль объектов играют переменные, а роль переменных – столбцы.

Этот метод реализует иерархический агломеративный алгоритм. Его смысл заключается в следующем. Перед началом кластеризации все объекты считаются отдельными кластерами, которые в ходе алгоритма объединяются. Вначале выбирается пара ближайших кластеров, которые объединяются в один кластер. В результате количество кластеров становится равным $N-1$. Процедура повторяется, пока все классы не объединятся. На любом этапе объединение можно прервать, получив нужное число кластеров. Таким образом, результат работы алгоритма агрегирования зависит от способов вычисления расстояния между объектами и определения близости между кластерами.

Для определения расстояния между парой кластеров можно использовать различные подходы. Так, существуют методы, основанные на разных способах вычисления расстояний между объектами:

- среднего расстояния между кластерами (Between-groups linkage);

- среднего расстояния между всеми объектами пары кластеров с учетом расстояний внутри кластеров (Within-groups linkage);

- расстояния между ближайшими соседями – объектами кластеров (Nearest neighbor);

- расстояния между самыми далекими соседями (Furthest neighbor);

- расстояния между центрами кластеров (Centroid clustering);

– расстояния между центрами кластеров (Centroid clustering – центроидный метод); недостатком этого метода является то, что центр объединенного кластера вычисляется как среднее центров объединяемых кластеров, без учета их объема.

Кроме того, используется метод медиан – тот же центроидный метод, но центр объединенного кластера вычисляется как среднее всех объектов (Median clustering), а также метод Уорда (Ward's method). В качестве расстояния между кластерами берется прирост суммы квадратов расстояний объектов до центров кластеров, получаемый в результате их объединения.

Расстояния и меры близости между объектами. У нас нет возможности сделать полный обзор всех коэффициентов, поэтому остановимся лишь на характерных расстояниях и мерах близости для определенных видов данных.

Меры близости отличаются от расстояний тем, что они тем больше, чем более похожи объекты.

Пусть имеются два объекта $X = (X_1, \dots, X_m)$ и $Y = (Y_1, \dots, Y_m)$. Используя эту запись для объектов, можно определить основные виды расстояний, используемых в кластерном анализе:

– евклидово расстояние $d(X, Y) = \sqrt{\sum_{i=1}^m (X_i - Y_i)^2}$

(Euclidian distance);

– квадрат евклидова расстояния

$d(X, Y) = \sum_{i=1}^m (X_i - Y_i)^2$ (Squared Euclidian distance) (эвк-

лидово расстояние и его квадрат целесообразно использовать для анализа количественных данных);

– мера близости – коэффициент корреляции

$$S(X, Y) = \left(\sum_{i=1}^m Z_{X_i} Z_{Y_i} \right) / (m - 1), \text{ где } Z_{X_i} \text{ и } Z_{Y_i} - \text{компо-}$$

ненты стандартизованных векторов X и Y (эту меру целесообразно использовать для выявления кластеров переменных, а не объектов);

– расстояние χ -квадрат, получаемое на основе таблицы сопряженности, составленной из объектов X и Y , которые предположительно являются векторами частот. Здесь рассматриваются ожидаемые значения элементов, равные $E(X_i) = X \cdot (X_i + Y_i) / (X + Y)$ и $E(Y_i) = Y \cdot (X_i + Y_i) / (X + Y)$, а расстояние χ -квадрат имеет вид корня из соответствующего по-

$$\text{казателя } d(X, Y) = \sqrt{\sum_{i=1}^m \frac{(X_i - E(X_i))^2}{E(X_i)} + \sum_{i=1}^m \frac{(Y_i - E(Y_i))^2}{E(Y_i)}}.$$

В иерархическом кластерном анализе чаще всего применяется евклидово расстояние и его квадрат.

Непосредственное использование переменных в анализе может привести к тому, что классификацию будут определять переменные, имеющие наибольший разброс значений. Поэтому применяются следующие виды стандартизации:

1) Z-шкалы (Z-Scores) – из значений переменных вычитается их среднее и эти значения делятся на стандартное отклонение;

2) разброс от -1 до 1 – линейным преобразованием переменных добиваются разброса значений от -1 до 1 ;

3) разброс от 0 до 1 – линейным преобразованием переменных добиваются разброса значений от 0 до 1 ;

4) максимум 1 – значения переменных делятся на их максимум;

5) среднее 1 – значения переменных делятся на их среднее;

6) стандартное отклонение 1 – значения переменных делятся на стандартное отклонение;

Кроме того, возможны преобразования самих расстояний, в частности, можно расстояния заменить их абсолютными значениями (это актуально для коэффициентов корреляции) или все расстояния преобразовать так, чтобы они изменялись от 0 до 1.

Таким образом, работа с кластерным анализом может превратиться в увлекательную игру, связанную с подбором метода агрегирования, расстояния и стандартизации переменных с целью получения наиболее интерпретируемого результата. Желательно только, чтобы это не стало самоцелью и исследователь получил действительно необходимые содержательные сведения о структуре данных.

Поскольку кластерный анализ предназначен для создания однородных групп, естественно рассмотреть процедуры, позволяющие определить число полученных групп. Например, древовидная структура дендрограммы указывает на то, что в данных может находиться много различных групп и правомерен вопрос: где нужно «обрезать» дерево, чтобы получить оптимальное число групп. К сожалению, эта проблема до сих пор находится среди нерешенных задач кластерного анализа из-за отсутствия подходящей нулевой гипотезы и сложной природы многомерных выборочных распределений.

В кластерном анализе существуют также методы, которые трудно отнести к первой или второй группе. Это так называемые итеративные методы. Сущность их заключается в том, что процесс классификации начинается с задания некоторых начальных условий (количество образуемых кластеров, порог завершения процесса классификации и т.д.). Итеративные методы в большей степени, чем иерархические, требуют от пользователя интуиции при выборе типа классификационных процедур и задания начальных условий разбиения, так как большинство этих методов очень чувствительно к изменению задаваемых параметров. Один из таких методов – метод k-средних – принадлежит к группе итеративных методов эталонного типа. В отличие от иерархических процедур он не требует вычисления и хранения матриц расстояний между объектами. Алгоритм этого метода предполагает использование только исходных значений переменных. Для начала процедуры классификации должны быть заданы k случайно выбранных объектов, которые будут служить эталонами, т. е. центрами кластеров. Считается, что алгоритмы эталонного типа достаточно удобны и быстродейственны. В этом случае важную роль играет выбор начальных условий, которые влияют на длительность процесса классификации и его результаты.

2.3. Дискриминантный анализ

Дискриминантный анализ – метод классификации многомерных наблюдений в одну или несколько категорий или совокупностей в ситуации, когда исследователь обладает так называемыми обучающими выборками.

Как и многие другие многомерные методы, дискриминантный анализ основан на построении линейных комбинаций признаков – дискриминантных функций, в которые каждый из них входит со своим коэффициентом (вкладом).

$$DF = b_1 x_1 + \dots + b_i x_i + \dots + b_p x_p + C ,$$

где DF – значение дискриминантной функции;

x_i – численное значение i -го признака;

b_i – вклад i -го признака в значение функции;

p – число признаков;

C – константа.

Дискриминантный анализ обеспечивает объективное сравнение (разделение) групп за счет искусственной минимизации внутригруппового разнообразия (дисперсии).

Обратившись к дискриминантному анализу, можно не только оценить достоверность межгрупповых различий и оценить «расстояния» между группами, но и определить те признаки из числа учтенных, которые в первую очередь обуславливают межгрупповые различия. Более того, когда две (или большее число) группы уже разделены в дискриминантном анализе, можно определить принадлежность неизвестного объекта к одной из них.

Исходные данные для дискриминантного анализа должны представлять собой совокупность объектов, разделенных на несколько групп (классов). Число признаков, по которым описаны объекты, не может быть меньше двух и не должно превышать суммарного числа объектов.

Признаки должны быть количественными, а распределение их значений в каждом классе – нормальным.

Недопустимо включение в комплекс признаков с единичной (полной) корреляцией.

Описание объекта по комплексу из p признаков геометрически эквивалентно определению его координат в p -мерном пространстве. Единичному объекту в пространстве отвечает точка, группе объектов – облако точек. Два или более облака точек, отвечающих изучаемым группам объектов, нередко перекрываются, т.е. неоднозначно различимы. Практически всегда есть особи с такой ординацией в пространстве, которая не позволяет однозначно отнести их к той или иной группе. Задача дискриминантного анализа геометрически формулируется как построение нового пространства, в котором принадлежность объектов к группам определяется однозначно – перекрывание облаков точек становится минимальным.

Координаты объектов в новом ортогональном пространстве зависят от значений дискриминантных функций (DF) – специфических линейных комбинаций признаков, определяемых на основе двух корреляционных матриц: межгрупповой (корреляции групповых средних) и внутригрупповой (корреляции признаков особей в группах). Выбор таких осей производится по критерию максимума отношения межгрупповой дисперсии к внутригрупповой. Математически выбор основан на сравнении матрицы корреляции групповых средних с матрицей корреляции тех же признаков внутри групп.

Суть подхода иллюстрирует рис. 2. Пусть мы имеем две группы объектов, описанных по двум переменным. Эти переменные образуют плоскость, на которой каждая из групп представлена в виде облака точек. Центроиды

этих групп обозначены А и В. При рассмотрении проекций крайних объектов каждой группы на координатные оси признаков, например, ось X, видны области перекрытия групп. Иными словами, в каждой из групп имеются объекты, положение которых не определено. Они не могут быть однозначно отнесены к той или иной группе. На рисунке это объекты в интервале $x_{b_1} - x_{a_2}$. Но можно построить новую ось (на рис. 2 – DF), значения которой по-прежнему учтут исходную изменчивость. Внутригрупповая изменчивость будет минимизирована и за счет этого четко будут выражены межгрупповые различия – перекрытие облаков точек исключено. Новая ось (новый интегральный признак) является линейной комбинацией исходных признаков – дискриминантной функцией.

Из рис. 2 ясно, в результате чего достигается четкое разделение групп – минимизация внутригрупповых различий. Действительно, расстояние между проекциями объектов a_1, a_2 и b_1, b_2 на ось дискриминантной функции меньше, чем расстояния между их проекциями на ось X. Расстояние между проекциями центроидов групп А и В на ось DF за счет этого оказывается большим, чем расстояние между проекциями на ось X (и аналогично на ось Y). Данное расстояние, называемое расстоянием Махаланобиса, и оценивает действительные различия групп.

На рис. 2 иллюстрируется возможность дискриминантного анализа при разделении двух групп. Но их может быть и больше. Это потребует определения дополнительных дискриминантных функций, ортогональных к первой. В общем случае число необходимых дискриминантных функций на единицу меньше числа разделяемых групп. Однако информативная ценность этих функций разная, поскольку они

учитывают разную долю исходной изменчивости комплекса признаков. Первая дискриминантная функция учитывает максимум такой изменчивости, вторая – максимум остатка и т.д. И если первые две-три функции учли значительную, например, более 80%, долю исходной дисперсии, то оставшиеся функциями можно пренебречь при решении последующих задач дискриминантного анализа.

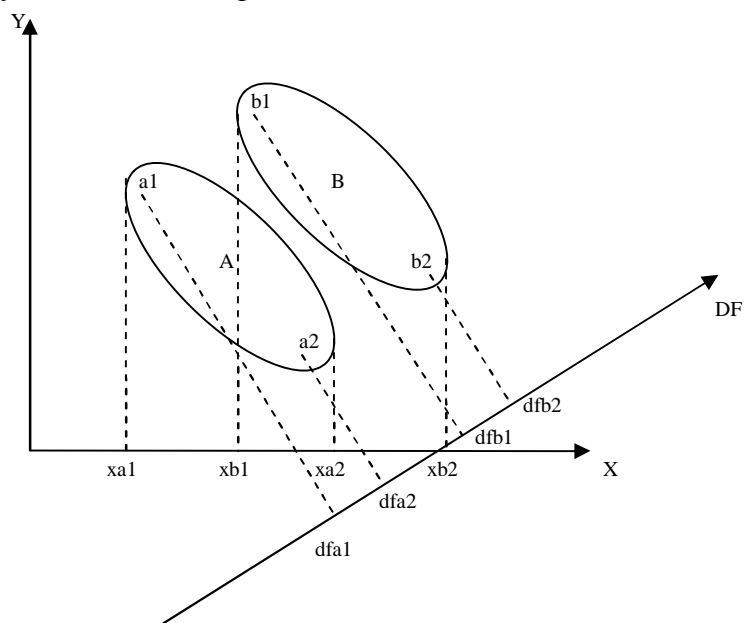


Рис. 2. Геометрическая интерпретация результатов дискриминантного анализа: А, В – центры разделяемых групп; x_{a1} , x_{b1} – минимальные значения признака X в группах А и В; x_{a2} , x_{b2} – максимальные значения признака в этих группах; d_{fa1} , d_{fa2} – проекции конкретных объектов группы А на ось дискриминантной функции; d_{fb1} , d_{fb2} – то же для конкретных объектов группы В (по Тюрину В.В., Мореву И.А., Волчкову Ю.А., 2003)

Процент учета дисперсии дискриминантными функциями не единственный и не самый главный критерий их значимости. Важной характеристикой, позволяющей оценить информативность дискриминантной функции, является коэффициент канонической корреляции (R). Каноническая корреляция оценивает меру связи между двумя множествами переменных. Чем больше значение R , тем выше разделительная способность дискриминантной функции.

Основной критерий оценки эффективности дискриминации для конкретной функции – величина λ -Уилкса. Этот критерий оценивает остаточную дискриминационную способность, под которой понимается способность различать группы, если исключить информацию, полученную с помощью ранее вычисленных функций. Когда остаточная дискриминация мала, вычисление очередной дискриминантной функции не имеет смысла. Статистика λ -Уилкса подчиняется χ^2 распределению, поэтому вычисление данного критерия позволяет оценить достоверность межгрупповых различий для каждой из полученных функций.

В результате дискриминантного анализа в зависимости от числа дискриминантных функций образуется ось, плоскость или пространство, дифференцирующее сравниваемые группы. В любом случае можно оценить степень сходства групп путем вычисления расстояния между их центроидами. В качестве меры сходства, как уже указано, выступает расстояние Махаланобиса (D^2), достоверность которого оценивается при помощи критерия Фишера.

В дискриминантном анализе различают два типа коэффициентов: нестандартизованные и стандартизованные. Посредством коэффициентов первого типа можно получить уравнение для каждого объекта, позволяющее вычислить собственно значение дискриминантной функции. В данное уравнение помимо значений признаков, умноженных на значения коэффициентов, также входит константа. Нормированные коэффициенты приводят к дискриминантным значениям, измеряемым в единицах стандартного квадратического отклонения, т.е. можно определить на сколько сигм конкретный объект отклоняется от главного центроида.

Стандартизованные коэффициенты позволяют судить об относительном вкладе конкретной переменной в дискриминантную функцию. То есть анализ абсолютных величин стандартизованных коэффициентов может дать ответ на вопрос, какова роль каждого признака в межгрупповых различиях. Если для всех значимых дискриминантных функций абсолютная величина коэффициента для данной переменной мала, то эту переменную можно исключить и тем самым сократить число переменных. Эта процедура называется определением информативного списка признаков.

К числу основных результатов дискриминантного анализа относят также вычисление структурных коэффициентов. Они представляют собой коэффициенты корреляции между конкретными переменными и дискриминантными функциями. Когда абсолютная величина такого коэффициента велика для конкретного признака, принято считать, что вся информация о дискриминантной функции заключена в этой переменной. Это позволяет

интерпретировать дискриминантные функции в соответствии с биологической ролью наиболее тесно скоррелированных с ней переменных.

При решении задач классификации – отнесения неизвестных объектов в одну из известных групп применяются классифицирующие функции. Они представляют собой уравнения, составленные для каждой группы. Неизвестный объект относится к классу, у которого в результате решения каждого из уравнений значение классифицирующей функции оказывается наибольшим.

Качество дискриминации групп также может быть проверено с использованием классификационной матрицы. Она вычисляется по классифицирующим функциям, определенным для каждого объекта совокупности. Критерием правильной дискриминации объектов служит процент их отнесения в «свою» группу.

3. ИСПОЛЬЗОВАНИЕ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МЕТОДОВ ДЛЯ ИЗУЧЕНИЯ ЭКОЛОГИЧЕСКОЙ ПЛАСТИЧНОСТИ СОРТОВ И ГИБРИДОВ ЯБЛОНИ

Сорт имеет огромное значение в плодоводстве. Однако не существует сортов универсальных, которые могли бы успешно выращиваться в любой местности, при любых почвенно-климатических условиях. Следует учитывать, что один и тот же сорт может очень сильно менять свойства, а следовательно, и производственную ценность, попадая в различные условия произрастания.

Генотип сорта определяет его реакцию на изменение среды, размах изменчивости каждого признака, а от среды зависит степень проявления данного признака в границах этого размаха. Сумма всех признаков и свойств каждого сорта, сложившаяся при данных условиях, называется его фенотипом и определяется взаимодействием генотипа и окружающей среды.

Материалом для нашего исследования послужили данные подеревного учета 90 сортов и гибридов яблони посадки 1986–1987 гг. общим числом 4200 в ОПХ «Центральное», предоставленные с.н.с. И.Л. Ефимовой. Таким образом, удалось получить данные о динамике урожайности каждого дерева с 1991 по 2000 г.

Для проведения предварительной оценки лучших генотипов было решено оставить сорта и гибридные комбинации, дававшие урожай до 2000 г. включительно. Таковых оказалось 50. Следует отметить, что выпадения деревьев наблюдались в течение всех 10 лет изучения. Данные с пропусками можно использовать в дальнейшем

для серьезных теоретических обобщений. В нашем случае оценивались 50 сортов и гибридов, что составляет от общего числа более 55% (1 412 деревьев).

Исследование было начато с двухфакторного дисперсионного анализа для количественной оценки влияния генотипа (сорта) и среды (года выращивания) на урожайность яблони. Следует отметить, что из анализа исключены данные за 1991–1992 гг., так как в них урожайность всех изученных деревьев равна нулю.

Влияние всех факторов признано статистически достоверным. В общей дисперсии доля фактора «сорт» составила 13%, фактора «год» – 30,4%, взаимодействия «сорт × год» – 20,7%. Как и ожидалось, средовая компонента оказалась весьма существенной.

В каждый год изучения, в зависимости от внешних условий, влияние сорта на урожай довольно разнообразно. Эффект сорта в каждом году сортоизучения статистически достоверен и вклад соответствующей дисперсии в общую варьирует от 17,2 (урожайность за 1993 г.) до 69,3% (урожайность за 1994 г.). Этот же факт наглядно иллюстрируют рис. 3 и 4.

Если проследить динамику каждого из 50 изученных сортов и гибридов яблони по самой простой схеме, т. е. возрастание или убывание урожайности в каждом последующем году, то обнаруживается 22 типа реакции, причем в большинство из них входят 1–3 формообразца. Такой метод довольно и трудоемок и его результат непредсказуем.

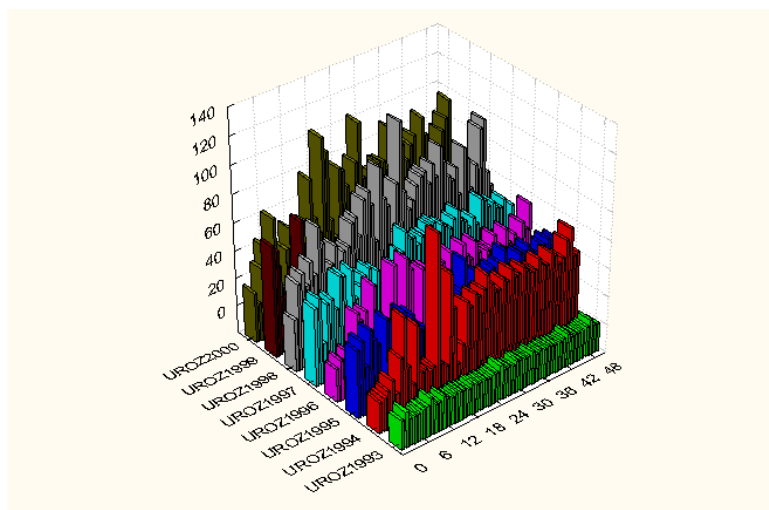


Рис. 3. Урожайность сортов и гибридов яблони

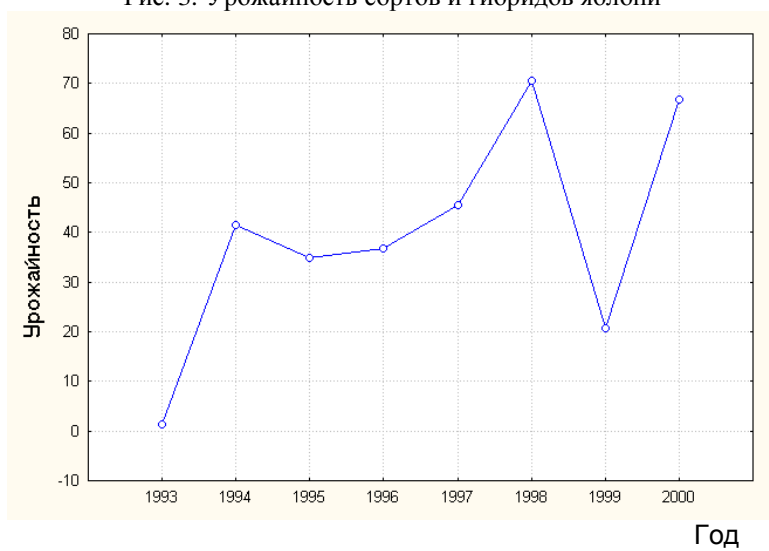


Рис. 4. Динамика средней урожайности сортов и гибридов яблони

Нами была предпринята попытка классификации типов реакции сортов с помощью многомерных статистических методов. Для этого данные по урожайности за последовательные годы плодоношения были преобразованы в линейные комбинации – главные компоненты, по которым был проведен кластерный анализ методом Уорда (рис. 5).

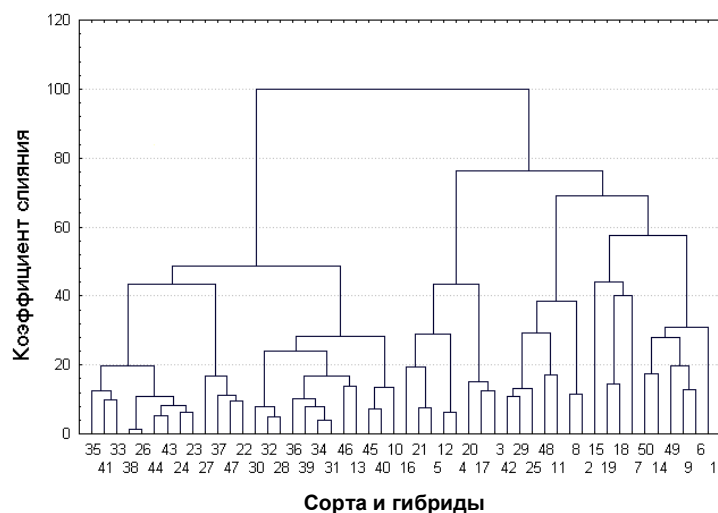


Рис. 5. Результаты кластерного анализа 50 сортов и гибридов яблони

Как видно из рис. 5, на полученном иерархическом древе можно выделить шесть кластеров. Для проверки правомерности кластерного решения был сделан дисперсионный анализ с фактором «кластер», подтвердивший наличие статистически достоверных различий между кластерами. Вклад соответствующей дисперсии довольно велик и варьирует от 37,0 (урожайность за

1998 г.) до 79,7% (урожайность за 1993 г.). Следует заметить, что с помощью кластерного анализа удалось выявить шесть групп сортов и гибридов, которые стабильно выделяются в каждом году сортоизучения и, что очень важно, достоверно различаются. Теперь дальнейшая работа будет проводиться только по шести выделенным группам. Обсуждение динамики их урожайности (рис. 6) составляет следующую часть нашего исследования.

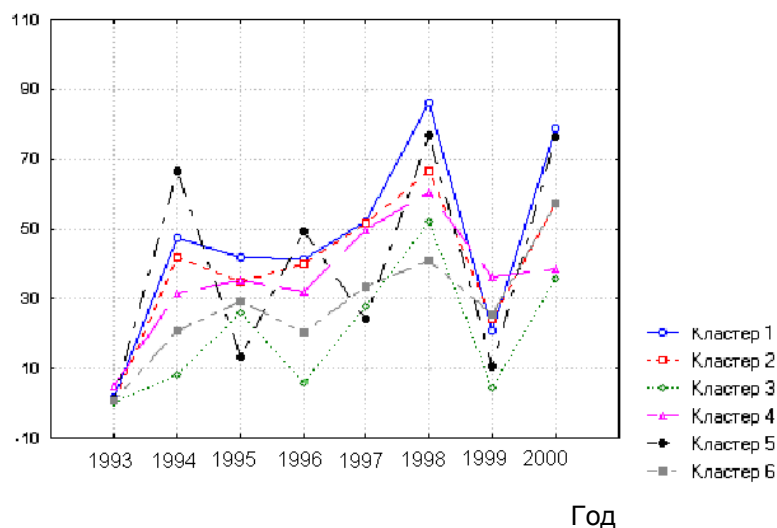


Рис. 6. Динамика урожайности шести групп сортов и гибридов яблони

На следующем этапе следовало отобрать группу сортообразцов, перспективных для дальнейшей селекционной работы. Дисперсионный анализ показал различия между группами сортообразцов, но не ответил на вопрос о их сходстве внутри каждого года исследований, т. е. требовалось провести их попарное сравнение с помощью

множественного рангового теста. Установлено, что в каждом году исследований 1–3 группы сортообразцов имели максимальную урожайность. Они расположились в следующем порядке: 1993 г. – 4-я группа; 1994 г. – 5-я группа; 1995 г. – 1-я и 4-я группы; 1996 г. – 5-я, 1-я и 2-я группы; 1997 г. – 1-я, 2-я и 4-я группы; 1998 г. – 1-я и 5-я группы; 1999 г. – 4-я и 6-я группы; 2000 г. – 1-я и 5-я группы. Для удобства отбора можно отобразить эту информацию в виде таблицы (табл. 1).

Таблица 1

Группы сортов и гибридов яблони
с максимальной урожайностью

Год	Группа					
	1	2	3	4	5	6
1993				*		
1994					*	
1995	*			*		
1996	*	*			*	
1997	*	*		*		
1998	*				*	
1999				*		*
2000	*				*	

Приведенные данные табл. 1 позволяют легко оценить изученные сорта и гибриды яблони. Видно, что группы 2, 3 и 6 имеют неудовлетворительные показатели урожайности и ее динамики. Наибольший интерес представляют группы 1, 4 и 5. Известно, что яблоне свойственна периодичность плодоношения, что хорошо видно в группах 4 и 5. Следует отметить, что если в саду будут

сорта одновременно из групп 4 и 5, урожай всегда будет хорошим, так как эти группы как бы дополняют друг друга (отсутствие урожая в одной из групп компенсируется другой группой). Более ровная динамика плодоношения наблюдается в группе 1, которая также обладает максимальной суммарной урожайностью из всех изученных. Это позволяет выделить вошедшие в нее сортообразцы как наиболее перспективные для дальнейшей селекции.

4. КЛАССИФИКАЦИЯ КОЛЛЕКЦИИ СОРТОВ ЯБЛОНИ ПО КОМПЛЕКСУ ПРИЗНАКОВ

В увеличении производства и повышении качества плодов решающая роль принадлежит сорту. В связи с интенсификацией садоводства к сорту как основному средству производства предъявляются повышенные требования. Сорты для интенсивных садов должны обладать целым комплексом положительных качеств. В настоящее время большинство промышленных сортов не отвечает повышенным требованиям производства. Необходимо планомерное обновление сортимента. Основным путем улучшения сортимента является селекция. Новые сорта должны быть интенсивного типа, с высокой зимостойкостью и адаптационной способностью, устойчивые к болезням и вредителям, скороплодные и урожайные, с плодами высоких товарных и потребительских качеств, с повышенным содержанием биологически активных веществ, разных сроков созревания.

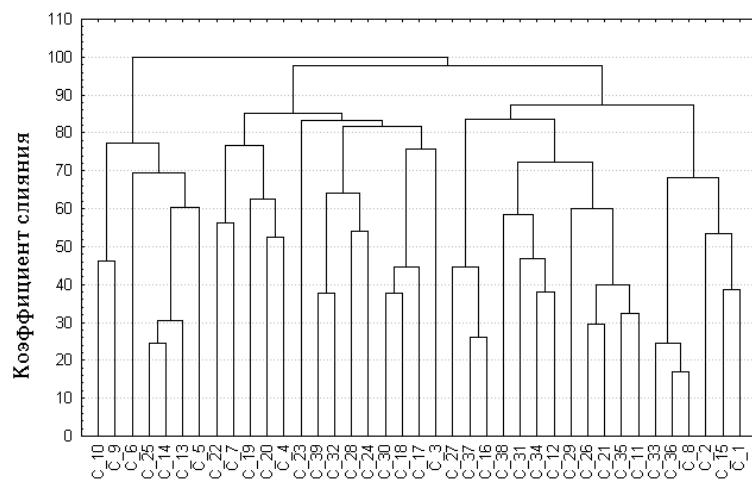
Материалом для исследования послужили 39 сортов яблони коллекции селекционного центра СКЗНИИСиб (кураторы коллекции – С.Н. Артюх, И.Л. Ефимова, Е.В. Ульяновская.), описанные по следующим признакам:

- 1) URDERSL – урожай с дерева на слаборослом подвое, кг/дер;
- 2) URGASL – урожай с гектара на слаборослом подвое, т/га;
- 3) URSERSR – урожай с дерева на среднерослом подвое, кг/дер;

- 4) URGASR – урожай с гектара на среднерослом подвое, т/га;
- 5) MPMIN – масса плода минимальная, г;
- 6) MPMAX – масса плода максимальная, г;
- 7) VKUS – дегустационная оценка, балл;
- 8) SVMIN – содержание сухих веществ минимальное, мг/%;
- 9) SVMAX – содержание сухих веществ максимальное, мг/%;
- 10) SAXMIN – содержание сахаров минимальное, мг/%;
- 11) SAXMAX – содержание сахаров максимальное, мг/%;
- 12) KISMIN – содержание кислот минимальное, мг/%;
- 13) KISMAX – содержание кислот максимальное, мг/%;
- 14) VCMIN – содержание витамина С минимальное, мг/%;
- 15) VCMAX – содержание витамина С максимальное, мг/%;
- 16) PAMIN – содержание Р-активных веществ минимальное, мг/%;
- 17) PAMAX – содержание Р-активных веществ максимальное, мг/%;
- 18) PARSHA – устойчивость к парше, балл;
- 19) MUCHROS – устойчивость к мучнистой росе, балл.

В связи с ориентацией исследования на системный анализ для обработки данных использовались многомерные статистические методы. Вся информация была пре-

образована в линейные комбинации с помощью метода главных компонент и подвергнута кластерному анализу методом Уорда для выявления сортов, сходных по комплексу из 19 учетных признаков. Результаты кластерного анализа представлены на рис. 7.



Сорта

Рис. 7. Результаты кластерного анализа сортов яблони

В результате выявлено шесть групп сортов. Они распределились по группам следующим образом:

1-я группа – Вадимовка, Восход Кубани, Делишес спур, Джонаред, Кальвиль снежный, Киддс оранж ред, Память Сергееву;

2-я группа – Белоснежка, Голден делишес, Кубань, Кубань спур, Мелба;

3-я группа – Арго, Кубанка, Кубанское багряное, Палитра, Память есаулу, Прикубанское, Ред делишес, Ренет шампанский, Вагнер;

4-я группа – Красный мак, Пепин лондонский, Фортуна;

5-я группа – Джонатан, Золотое летнее, Ламбурне, Пармен зимний золотой, Прима, Ренет кубанский, Слава переможцам, Старк, Бойкен;

6-я группа – Айдаред, Алёнушкино, Делишес, Корей, Рояль ред делишес, Старкримсон.

Для проверки правильности кластерного решения был проведен дисперсионный анализ, оценивающий различия по выделенным группам. В результате анализа статистически достоверные различия показали 13 признаков из 19. Соответствующая дисперсия варьировала от 22,5 (урожай с дерева на среднерослом подвое) до 60,4% (содержание сахаров максимальное). Такой результат на выборке из 39 сортов можно рассматривать как положительный.

По признакам, показавшим достоверные различия по группам, был построен график Sun Rays (солнечных лучей), на осях которого откладывались нормированные значения признаков (рис. 8). На этом рисунке отчетливо видно, что группы различаются по своим значениям признаков.

Далее нами был использован дискриминантный анализ – метод классификации многомерных наблюдений в одну или несколько категорий или совокупностей в ситуации, когда исследователь обладает так называемыми обучающими выборками (табл. 2).

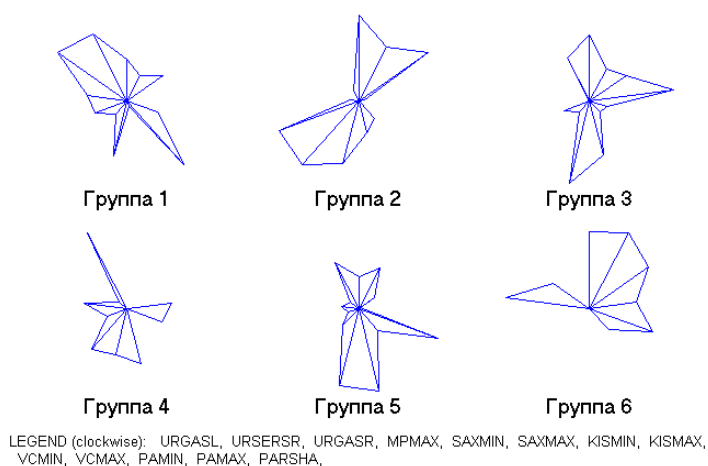


Рис. 8. Нормированные значения признаков в выделенных группах

Таблица 2

Результаты дискриминантного анализа групп сортов яблони

Дискриминантная функция	Собственное значение	Лямбда Уилкса	χ -квадрат	df	Уровень значимости
1	14,89	0,0001	230,02	95	0,000
2	5,96	0,0019	159,49	72	0,000
3	5,08	0,0133	110,01	51	0,000
4	3,09	0,0081	63,95	32	0,000
5	2,00	0,3332	28,01	15	0,021

Приведенные в табл. 2 низкие значения лямбды Уилкса свидетельствуют о хорошем качестве и мощности разделения групп, но самое важное – уровень значимости показы-

ваает на работу всех дискриминантных функций в этом разделе.

Теперь мы подходим к основному – построению модели сортов, соответствующих каждой из выделенных групп. Построить такие модели позволяют функции классификации, полученные в дискриминантном анализе. В том, что они работают правильным образом, можно убедиться с помощью классификационной таблицы (табл. 3).

Таблица 3

Классификационная таблица дискриминантного анализа

Группа	Процент правильных отнесений	Группа					
		1	2	3	4	5	6
1	100,0	7	0	0	0	0	0
2	100,0	0	5	0	0	0	0
3	100,0	0	0	9	0	0	0
4	100,0	0	0	0	3	0	0
5	100,0	0	0	0	0	9	0
6	100,0	0	0	0	0	0	6

Как видно из табл. 3, в результате анализа удалось найти правило отнесения сорта по комплексу признаков к одной из шести групп со 100% точностью. Вычисленные для этого функции классификации приведены в табл. 4.

Теперь, если исследователю потребуется отнести любой сорт яблони к определенной группе, значения комплекса признаков этого сорта следует подставить в уравнение. Например, для группы 1 оно следующее:

Группа 1 = $3,22 \cdot \text{URDERSL} - 4,11 \cdot \text{URGASL} + 0,07 \cdot \text{URSERSR} + 0,95 \cdot \text{URGASR} - 0,88 \cdot \text{MPMIN} + 0,423 \cdot$

$\cdot \text{MPMAX} + 271,46 \cdot \text{VKUS} + 12,48 \cdot \text{SVMIN} - 13,93 \cdot$
 $\cdot \text{SVMAX} - 7,02 \cdot \text{SAXMIN} + 8,29 \cdot \text{SAXMAX} - 191,31 \cdot$
 $\cdot \text{KISMIN} + 194,28 \cdot \text{KISMAX} - 2,52 \cdot \text{VCMIN} + 1,67 \cdot$
 $\cdot \text{VCMAX} + 2,39 \cdot \text{PAMIN} - 0,380 \cdot \text{PAMAX} + 3,58 \cdot \text{PAR-}$
 $\text{SHA} + 11,44 \cdot \text{MUCHROS} - 773,62.$

Таблица 4

Функции классификации для групп сортов яблони

Признак	Группа					
	1	2	3	4	5	6
URDERSL	3,22	3,13	2,12	2,88	2,83	3,24
URGASL	-4,11	-3,86	-3,39	-4,55	-3,95	-4,29
URSERSR	0,07	0,07	0,25	0,29	0,28	0,21
URGASR	0,95	1,42	1,17	0,93	1,13	1,53
MPMIN	-0,88	-0,91	-1,08	-1,03	-0,97	-1,01
MPMAX	0,423	0,40	0,68	0,46	0,43	0,51
VKUS	271,4	258,83	258,69	291,07	261,01	277,61
SVMIN	12,48	9,63	11,62	14,19	13,40	13,48
SVMAX	-13,93	1,03	-3,25	-3,82	-4,02	-6,67
SAXMIN	-7,02	-2,80	0,68	1,19	3,96	0,39
SAXMAX	8,29	-9,18	-12,05	-14,68	-11,53	-9,50
KISMIN	-191,31	-153,69	-187,47	-193,10	-166,68	-205,64
KISMAX	194,28	178,43	214,58	199,81	194,09	213,96
VCMIN	-2,52	0,78	-3,11	-1,39	-0,46	-2,57
VCMAX	1,67	1,21	-0,14	-0,08	-1,03	-0,12
PAMIN	2,39	2,12	1,90	2,25	1,88	2,29
PAMAX	-0,38	-0,40	-0,30	-0,41	-0,35	-0,45
PARSHA	3,58	0,49	1,5916	3,21	2,56	1,67
MUCHROS	11,44	3,56	6,68	11,27	6,58	8,12
Константа	-773,62	-713,45	-654,22	-791,01	-684,38	-749,97

Для остальных групп уравнение строится подобным образом. В результате новый сорт можно отнести к той группе, для которой классификационное значение будет максимальным. Мы сейчас не обсуждаем достоинства сортов в выделенных группах. Нами предложен и доказан сам принцип их разделения и отнесения к ним новых сортов. Следует отметить, что группы выделены не только по их средним значениям, но и с учетом системы корреляций признаков, учтенных главными компонентами. Это позволяет сделать вывод, что в группах сортов могут находиться сходные генотипы, реализованные фенотипически в различной степени. Сорты из различных групп можно использовать при подборе родительских пар при скрещивании.

**5. МИНИМИЗАЦИЯ ЭФФЕКТОВ УСЛОВИЙ ГОДА
ПРИ ОЦЕНКЕ ГЕНЕТИЧЕСКИ ОБУСЛОВЛЕННЫХ РАЗЛИЧИЙ
В УСТОЙЧИВОСТИ ЧЕРЕШНИ К МОНИЛИОЗУ
МЕТОДОМ ДИСКРИМИНАНТНОГО АНАЛИЗА**

В настоящее время в связи с массовым выпуском отечественной промышленностью джемов, йогуртов, мороженого возникает серьезная потребность в плодах черешни и вишни. На сегодняшний день эта потребность удовлетворяется не более чем на 50%. Это связано с тем, что многие широко используемые в сельском хозяйстве сорта черешни подвержены грибковым заболеваниям, что приводит не только к значительным потерям урожая, но и к повреждениям осевых ветвей дерева и, через несколько лет к гибели растения. В условиях Краснодарского края наибольшую опасность представляет монилиоз, поскольку грибы рода *Monilia* хорошо приспособлены к экологическим условиям Кавказского региона. Благодаря теплым и влажным зимам спороношение гриба наблюдается на протяжении почти всего зимнего периода, что способствует быстрому распространению заболевания. В настоящее время особое беспокойство вызывает расширение специализации гриба.

Наиболее перспективный и экологически безопасный путь борьбы с монилиозом черешни – выращивание сортов, устойчивых к грибу. В связи с постоянной эволюцией возбудителя необходимо постоянное обновление садов устойчивыми высокопродуктивными сортами черешни.

Работа осуществлялась совместно с А.П. Кузнецовой на базе ОПХ «Центральное» СКЗНИИСиВ в 1998–

2000 гг. Материалом для исследования послужили образцы коллекции из 135 сортов черешни.

Проведена визуальная оценка степени поражения сортов монилиозом. Балл поражения учитывался согласно методике селекции плодово-ягодных и орехоплодных культур по следующей шкале: 0 – поражение отсутствует; 1 – поражено 10% соцветий; 2 – 10–20% соцветий; 3 – 20–30% соцветий; 4 – более 40% соцветий.

В число необходимых характеристик современного сорта включаются не только продуктивность, качество, устойчивость к болезням и вредителям, но и экологическая пластичность. Под ней понимают способность сохранять хозяйственно ценные признаки в изменяющихся условиях среды. В связи с длительным периодом хозяйственного использования растений это свойство оказывается для плодовых одной из наиболее важных характеристик селекционной и хозяйственной ценности сорта.

В генетическом плане понятие пластичности ассоциируется прежде всего с уровнем модификационной изменчивости. Она возникает как реакция на изменяющиеся условия среды, и пределы её (норма реакции) заданы генотипом. Именно это обстоятельство оправдывает сравнение сортов по уровню пластичности как генетически детерминированного свойства и целенаправленную селекцию на её повышение.

В основе большинства методов оценки пластичности лежат работы Р. Фишера по дисперсионному анализу. В рамках двухфакторного дисперсионного комплекса, где в качестве факторов выступают сорт (генотип) и условия выращивания (условия года, района, агротехнический фон), реально можно количественно оценить дис-

персию взаимодействия «генотип × среда». Дальнейшее развитие методов было направлено на разложение общей дисперсии взаимодействия на вклады, вносимые каждым из изученных сортов. Величина такого вклада и является оценкой пластичности сорта.

Предложен и апробирован целый ряд методов такого разложения, которые различаются лишь некоторыми статистическими особенностями. Центральной осталась идея разложения, предложенная С. Эберхартом и Н. Расселом (Eberhart S., Russel W., 1966), которая основана на оценке пластичности параметрами регрессии урожая в данном году (или данном регионе) на так называемый индекс среды. Последний представляет собой урожай по совокупности изученных сортов в данном году (регионе). Иными словами, индекс среды – это интегральная оценка условий самими изучаемыми объектами.

В нашем случае изучался не урожай с дерева, а балл поражения монилиозом. Мы сочли возможным внести соответствующие коррективы в метод, предложенный С. Эберхартом и В. Расселом. Индекс среды вычислялся как балл поражения монилиозом по совокупности сортов в течение нескольких лет изучения. Для каждого из сортов были вычислены три статистические характеристики. Первая из них – коэффициент регрессии балла поражения монилиозом на индекс среды. Чем он меньше, тем менее сорт реагирует на изменение условий года. Вторая характеристика – дисперсия отклонений от линии регрессий, оценивающая регулярность ответных реакций сорта на изменение условий. Как и в случае коэффициента регрессии, низкое значение этой дисперсии – опти-

мальное свойство сорта. Третья характеристика – средний многолетний балл поражения монилиозом.

Выбранные характеристики имеют между собой очень слабую корреляционную зависимость ($r = 0,2$) и не дублируют друг друга. Также они подчиняются нормальному распределению (рис. 9).

«Идеальным» с этих позиций представляется сорт с наименьшими значениями коэффициента регрессии, дисперсии отклонений и многолетнего балла поражения монилиозом.

Объединение характеристик в комплекс требует построения их линейной комбинации. Наилучший метод – главные компоненты.

Уже первые две из трёх возможных учли в совокупности 78% полной дисперсии комплекса характеристик (первая – 44%). Это позволило перевести задачу оценки соответствия сортов «идеальному» в оценку расстояний между соответствующими точками в плоскости I и II главных компонент.

В результате кластерного анализа сортов по методу Уорда (рис. 10) было выделено семь групп.

Метод Уорда относится к категории агломеративных (объединяющих), его идея состоит в том, чтобы проводить объединение, дающее минимальное приращение внутри групповой суммы квадратов отклонений. При известном числе выделяемых кластеров мы решили воспользоваться методом группировки k -средних, относящемся к итеративным разъединяющим методам и работающим непосредственно с объектами, а не с матрицей сходства. В методе k -средних объект относится к тому классу, расстояние до которого минимально. Расстояние

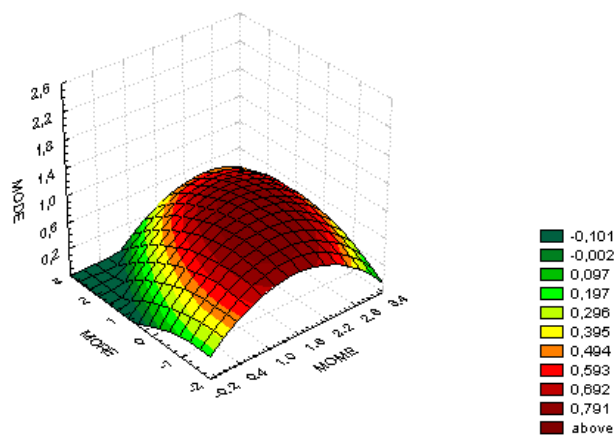


Рис. 9. Распределение характеристик поражения монилиозом коллекции сортов черешни:
 MODE – дисперсия отклонений от линии регрессии;
 MORE – коэффициент регрессии балла поражения монилиозом;
 MOME – средний многолетний балл поражения монилиозом

понимается как евклидово расстояние, т. е. объекты рассматриваются как точки евклидова пространства. При заданных изначально семи кластерах метод позволил успешно провести их формирование.

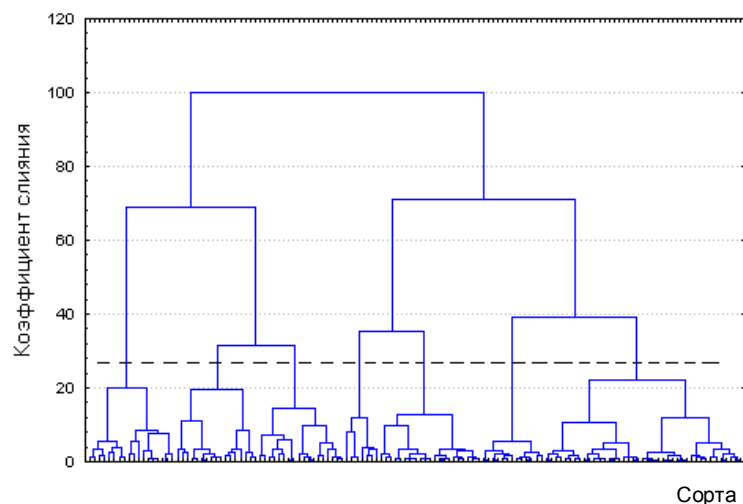


Рис. 10. Кластерный анализ сортов черешни по характеристикам поражения монилиозом

Статистически достоверные различия выделенных групп подтверждаются соответствующим дисперсионным анализом, выполненным по фактору «кластер». Установлены статистически достоверные межкластерные различия. Вклад соответствующей дисперсии в общую довольно велик: от 68,3 (евклидово расстояние до «идеального» сорта) до 77,3% (коэффициент регрессии).

Распределение сортов, включая «идеальный» (обозначен как F), в пространстве I и II главных компонент представлено на рис. 11.

Хотя различия выделенных семи кластеров очевидны, была предпринята попытка уменьшить число групп по баллу поражения монилиозом путем множественного рангового теста. Было доказано, что отсутствуют достоверные различия между первым и четвёртым, а также третьим и вторым кластерами, что даёт основание для их объединения. Проведённый повторно множественный ранговый тест подтвердил справедливость внесённой корректировки.

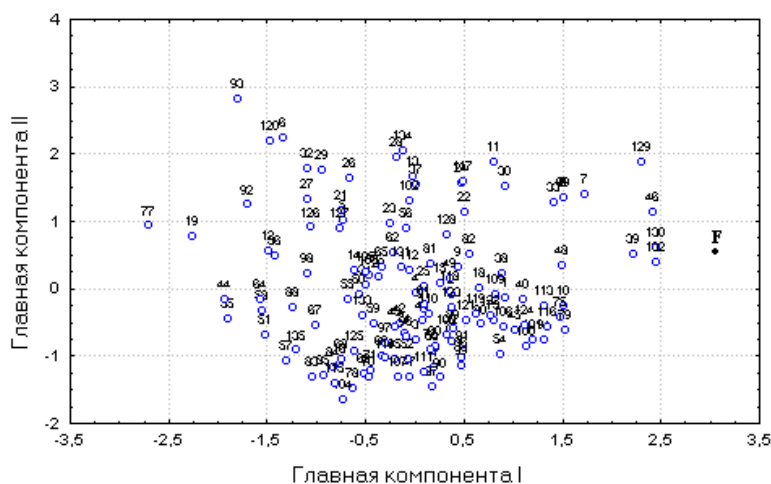


Рис. 11. Распределение сортов черешни в пространстве главных компонент

Решение об объединении кластеров подтвердилось также повторным дисперсионным анализом.

Межкластерные различия были статистически подтверждены. Доля соответствующей дисперсии в общей

составляла от 12,1 (дисперсия отклонения от линии регрессии) до 77,0 (коэффициент регрессии балла поражения монилиозом).

Итак, в результате предыдущих анализов было выявлено пять групп сортов черешни по устойчивости к монилиозу. В многомерном статистическом анализе существует раздел, содержащий статистические методы классификации многомерных наблюдений по одной из нескольких категорий, или совокупностей, называемый дискриминантным анализом. Его цель состоит в том, чтобы на основе измерения различных характеристик объекта классифицировать его, т. е. отнести к одной из нескольких групп (кластеров) некоторым оптимальным способом. Под оптимальным способом понимается минимум вероятности ложной классификации. В нашем случае следовало найти классификационное правило для отнесения сортов черешни по устойчивости к монилиозу в одну из выделяемых пяти групп.

Проведенный нами дискриминантный анализ показал, что распознавание группы принципиально возможно. Данные табл. 5 показывают, что дифференциация групп сортов осуществляется вполне успешно.

Таблица 5

Основные характеристики дискриминантного анализа, распознающего группы сортов черешни по устойчивости к монилиозу

Дискриминантная функция	Процент учета дисперсии	Критерий Уилкса	χ -квадрат	Уровень значимости
1	61,0	0,064	356,3	0,00
2	99,0	0,293	159,3	0,00
3	100,0	0,940	7,9	0,02

Из табл. 5 видно, что различия между группами установлены по значениям всех трех дискриминантных функций. Значение статистики Уилкса, лежащее около нуля, свидетельствует о хорошей дискриминации, что подтверждается распределением сортов в пространстве дискриминантных функций (рис. 12)

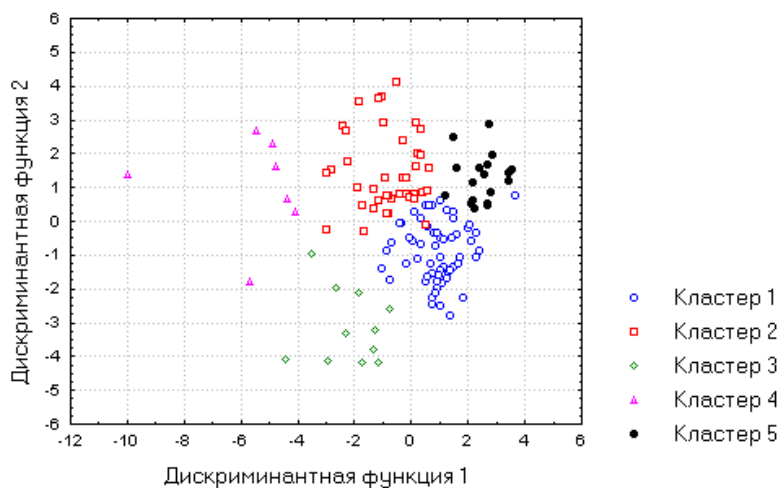


Рис. 12. Распределение сортов черешни в пространстве дискриминантных функций

В результате дискриминантного анализа были также получены функции классификации, позволяющие относить новые сорта черешни к одной из групп (табл. 6). Нами предложены следующие градации устойчивости к монилиозу: высокая, выше средней, средняя, ниже средней, низкая.

Таблица 6

Функции классификации отнесения сортов черешни в группы по устойчивости к монилиозу

Признак	Устойчивость к монилиозу				
	Высокая (3)	Выше средней (1)	Средняя (5)	Ниже средней (2)	Низкая (4)
Средний многолетний балл поражения (MOME)	6,38	5,52	7,27	14,21	23,45
Коэффициент регрессии (MORE)	-4,94	4,88	10,80	2,32	-10,36
Дисперсия отклонений от линии регрессии (MODE)	1,13	3,06	4,81	4,07	-0,73
Константа	-5,73	-6,64	-19,67	-17,79	-32,85

С помощью функций представленных в табл. 7, можно вычислить классификационные значения (метки) для вновь наблюдаемых сортов черешни по формулам:

1) устойчивость высокая:

$$6,38 \cdot \text{MOME} - 4,94 \cdot \text{MORE} + 1,13 \cdot \text{MODE} - 5,73;$$

2) устойчивость выше средней:

$$5,52 \cdot \text{MOME} + 4,88 \cdot \text{MORE} + 3,06 \cdot \text{MODE} - 6,64;$$

3) устойчивость средняя:

$$7,27 \cdot \text{MOME} + 10,80 \cdot \text{MORE} + 4,81 \cdot \text{MODE} - 19,67;$$

4) устойчивость ниже средней:

$$14,21 \cdot \text{MOME} + 2,32 \cdot \text{MORE} + 4,08 \cdot \text{MODE} - 17,79;$$

5) устойчивость низкая:

$$23,45 \cdot \text{MOME} - 10,36 \cdot \text{MORE} - 0,73 \cdot \text{MODE} - 32,85.$$

Новый сорт будет относиться к той группе устойчивости, для которой классификационное значение максимально.

6. ИДЕНТИФИКАЦИЯ ПРЕДСТАВИТЕЛЕЙ РОДА PRUNUS L. ПО МОРФОЛОГИЧЕСКИМ ПРИЗНАКАМ ЛИСТА И КОСТОЧКИ ПЛОДА

В селекции сливы акцент сделан на сочетании в гибридном потомстве комплекса ценных признаков. В основу нашего исследования положено изучение морфологических признаков листа и косточки сливы, сложное строение которых открывает перспективы детального описания отдельных растений и оценки различий между ними.

Использованы архивные материалы, содержащие описание 33 сортов сливы домашней, 3 сортов алычи и 3 сортов терна по следующим признакам: высота косточки (VK); ширина косточки (SK); толщина косточки (TK); длина листа (ODL); ширина левой части листа (SL); ширина правой части листа (SP); длина черешка (DCH); угол левой доли основания листа (U1); угол правой доли основания листа (U2); высота листа от основания до самой широкой части (PD); количество жилок слева (ZL); количество жилок справа (ZP).

Известно, что слива получена как гибрид алычи и тёрна, поэтому представляет интерес обнаружение различий этих видов по учтённым признакам. С этой целью нами использован однофакторный дисперсионный анализ. В результате были обнаружены статистически достоверные различия по двум из трёх признаков, характеризующих косточку плода. Доля соответствующей дисперсии варьировала от 31,3 (ширина косточки) до 35,0% (высота косточки). Исключение составил признак (толщина косточки), по которому различия не обнаружены.

По признакам морфологии листа различия обнаружены по 5 признакам из 9. Доля соответствующей дисперсии колебалась от 9,0 (длина черешка) до 21,0% (ширина левой части).

Далее необходимо было получить количественную характеристику изменчивости морфологических признаков листа отдельно по каждой породе с помощью дисперсионного анализа. Оказалось, что сорта сливы домашней статистически достоверно различаются по всем признакам морфологии листа, доля межсортовой дисперсии составляет от 21,7 (количество жилок справа) до 72,9% (угол левой доли основания листа); у алычи не обнаружено статистически достоверных различий по признакам – по длине черешка, углу левой доли основания листа, количеству жилок слева и справа, по остальным признакам доля межсортовой дисперсии составила от 23,7 (угол правой доли основания листа) до 70,6% (высота листа от основания до самой широкой части); у тёрна, как и у сливы домашней, сорта различаются по всем признакам, причём доля соответствующей дисперсии колеблется от 33,7 (длина листа) до 73,3% (угол правой доли основания листа). Такой результат был вполне ожидаем, так как различия сортовых генотипов обусловлены предшествующей селекцией.

Затем предстояло выявить, насколько велики различия по всем категориям учтенных признаков между сортами алычи, тёрна и сливы. Ввиду гибридного происхождения сливы домашней от алычи и тёрна должны быть формы, сходные с ними по морфологическому строению листа и плода. Для обнаружения таких форм было реше-

но воспользоваться методами многомерного статистического анализа.

Интегральная характеристика обеих категорий признаков получена с помощью метода главных компонент, который показал, что в дальнейших исследованиях достаточно использовать только первые пять главных компонент, которые в совокупности учли 94,2% общей дисперсии.

Сократив размерность исходного пространства 12 признаков до 5 главных компонент, мы провели кластерный анализ сортов слив, алычи и тёрна методом Уорда (рис. 13).

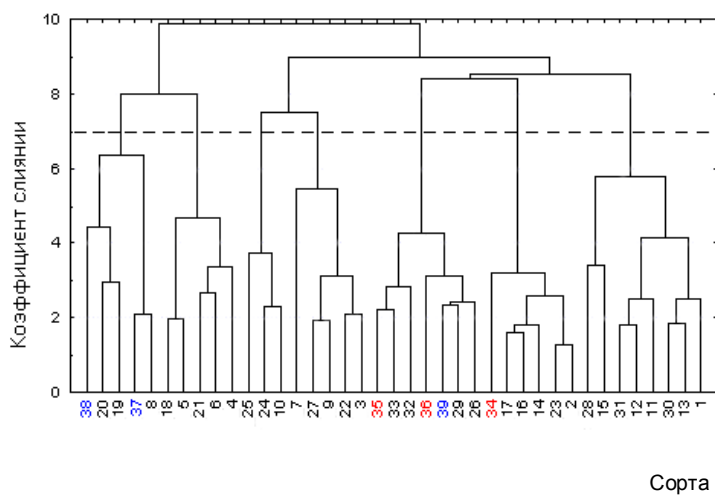


Рис.13. Кластерный анализ сортов сливы, алычи и тёрна

Из рис. 13 видно, что выделяется 7 групп сортов. Справедливость кластерного решения проверена нами в

дисперсионном анализе с фактором «кластер». Установлено, что статистически достоверные различия обнаружены по всем признакам, кроме признака «толщина косточки», но поскольку этот признак не показывал различий и между видами рода *Prunus* L., такой результат можно считать удовлетворительным.

Как видно из рис. 13, родительские виды (алыча и тёрн) вошли в первый, пятый и шестой кластеры. Причём в первый вошёл только тёрн, в пятый – алыча и тёрн, а в шестой – только алыча. Следовательно, сорта сливы, попавшие в эти кластеры, несут черты, сходные с родительскими видами. Таким образом, можно сделать вывод, что сорта сливы Ренклад ранний, Ренклад Карбышева, Пердригон полосатый сходны по морфологическим признакам листа и плода с терном; сорта сливы *Prunus Marianna*, Климас, Ломбард, Царь сходны по морфологии с алычой и терном, а сорта сливы Терновка Ронгоша, Тернослив 44–21, Ренклад советский, *Prunus Ussuriensis*, Терновка ажанская сходны по морфологии с алычой. Остальные сорта можно отнести к так называемым новообразованиям, т. е. растениям, имеющим отличия от родительских форм.

Проблему идентификации видов рода *Prunus* L. мы решили с помощью дискриминантного анализа – метода классификации многомерных наблюдений по одной из нескольких категорий, или совокупностей. Его цель состоит в том, чтобы на основе измерения различных характеристик объекта классифицировать его, т. е. отнести к одной из нескольких групп некоторым оптимальным способом (табл. 7).

Таблица 7

Результаты дискриминантного анализа между
сливой домашней, алычой и тёрном

Дискриминантная функция	Собственное значение	Критерий Уилкса	χ -квадрат	Уровень значимости
1	0,28	0,72	102,11	0,000
2	0,08	0,92	25,07	0,002

Как видно из табл. 7, разделение сортов сливы, алычи и тёрна проходит успешно и достоверными оказываются обе дискриминантные функции. Несмотря на это величина критерия лямбда Уилкса, приближающаяся к 1, свидетельствует о сложности разделения, что иллюстрирует график распределения сортов сливы, алычи и тёрна (рис. 14).

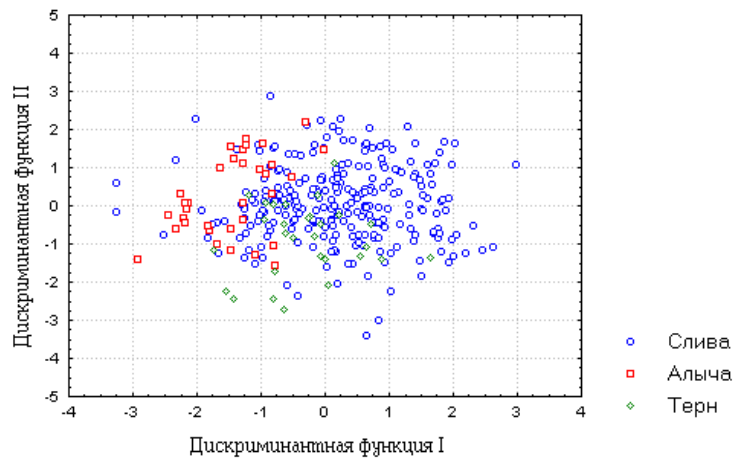


Рис. 14. Разделение сортов сливы, алычи и тёрна
в пространстве дискриминантных функций

На рис. 14 показано расположение сортов сливы, алычи и тёрна в пространстве первой и второй дискриминантных функций, которые, судя по результатам, показанным в табл. 7, должны хорошо разделять эти виды. Видно, что некоторые сорта разных видов находятся на большом расстоянии друг от друга, а некоторые, видимо, очень сходные – на небольшом расстоянии и даже смешиваются в центре рисунка.

В результате дискриминантного анализа можно также получить функции классификации, позволяющие относить новые сортообразцы к одной из групп (слива, алыча, тёрн) по признакам морфологии листа и плода (табл. 8).

Таблица 8

Функции классификации сортов сливы, алычи и тёрна

Признак	Слива	Алыча	Тёрн
ODL	-0,53	-0,35	-0,70
SL	2,60	2,40	2,80
SP	1,11	0,79	1,03
DCH	-0,18	-0,17	-0,10
U1	0,72	0,79	0,82
U2	0,73	0,62	0,72
PD	-0,98	-1,23	-0,92
ZL	3,16	3,45	3,40
ZP	1,05	0,97	1,37
Константа	-44,15	-42,45	-50,06

С помощью этих функций, представленных в табл. 8, можно вычислить классификационные значения

для отнесения сортообразцов на основании измерений листа и плода к сливе, алыче или тёрну по следующим формулам:

$$\begin{aligned} \text{Слива} = & -0,53 \times \text{ODL} + 2,60 \times \text{SL} + 1,11 \times \text{SP} - 0,18 \times \\ & \times \text{DCH} + 0,72 \times \text{U1} + 0,73 \times \times \text{U2} - 0,98 \times \text{PD} + \\ & + 3,16 \times \text{ZL} + 1,05 \times \text{ZP} - 44,15. \end{aligned}$$

$$\begin{aligned} \text{Алыча} = & -0,35 \times \text{ODL} + 2,40 \times \text{SL} + 0,79 \times \text{SP} - 0,17 \times \\ & \times \text{DCH} + 0,79 \times \text{U1} + 0,62 \times \times \text{U2} - 1,23 \times \text{PD} + \\ & + 3,45 \times \text{ZL} + 0,97 \times \text{ZP} - 42,45. \end{aligned}$$

$$\begin{aligned} \text{Тёрн} = & -0,70 \times \text{ODL} + 2,80 \times \text{SL} + 1,03 \times \text{SP} - 0,10 \times \\ & \times \text{DCH} + 0,82 \times \text{U1} + 0,72 \times \text{U2} - 0,92 \times \text{PD} + 3,40 \times \\ & \times \text{ZL} + 1,37 \times \text{ZP} - 50,06. \end{aligned}$$

Сортообразец будет относиться к тому виду, для которого классификационное значение максимально.

7. МЕТОД РАСПОЗНАВАНИЯ СОРТОВ И/ИЛИ ГИБРИДОВ ПЛОДОВЫХ И ЯГОДНЫХ КУЛЬТУР, ОБЛАДАЮЩИХ ВЫСОКИМ И СТАБИЛЬНЫМ УРОЖАЕМ

Цель предлагаемого метода – разработка способа распознавания сортов и гибридов плодовых и ягодных культур, обладающих высоким и стабильным урожаем.

Для решения этой задачи предлагается использовать данные по урожайности сортов и/или гибридов за три или более последовательных года плодоношения. Каждому сорту и/или гибриду присваивается номер. По каждому сорту и/или гибриду вычисляют: разность урожая на втором и первом году плодоношения, третьем и втором и т.д. за весь период исследований; сумму полученных разностей урожаев (СР) по каждому сорту и/или гибриду; среднее значение урожая за учтенные годы плодоношения (СУ); коэффициент вариации урожая за учтенные годы плодоношения (КВ).

Рассчитывают среднее значение СУ по всей совокупности сортов и/или гибридов. Далее сорта и/или гибриды ранжируют по убыванию СУ и исключают из них давшие урожай ниже среднего по всей совокупности сортов и/или гибридов.

Затем по каждому из оставшихся сортов и/или гибридов в декартовой системе координат, на оси абсцисс которой откладывают сумму разностей урожаев (СР), а на оси ординат – коэффициент вариации (КВ), отмечают точки с соответствующими координатами, вычисляют средние значения СР и КВ и полученные координаты наносят на график в виде точки. Через эту точку проводят перпендикулярные линии, разделяющие график на 4

квадранта с нумерацией по часовой стрелке, начиная с левого нижнего. В первый и второй квадранты попадают сорта и/или гибриды с низким урожаем и снизившие урожай за годы плодоношения; в третий квадрант – с высоким, но не стабильным урожаем; в четвертый квадрант – сорта и/или гибриды с высоким и стабильным урожаем. Эти сорта и/или гибриды пригодны для селекции. Полученные данные по каждому из исследуемых сортов позволяют определить наиболее пригодные из них для данных климатических условий.

Покажем работу метода на примере плодовых культур. Объектом исследования послужили данные по урожайности 50 сортов и гибридов яблони за период исследования, равный 7 годам. Каждому сорту и гибриду был присвоен номер.

Средний урожай с дерева за 7 лет всех 50 сортов и гибридов оказался 36,4 кг. После проведения ранжирования значений СУ каждого сорта и гибрида по убыванию и исключения из них давших значение СУ ниже 36,4 кг осталось 32 сорта и гибрида.

По каждому из этих 32 сортов и гибридов на декартовой системе координат на оси абсцисс отложим сумму разностей урожаев (СР) каждого сорта и гибрида, а на оси ординат – коэффициент вариации (КВ) и отметим точки с соответствующими координатами, проставляя соответствующий номер сорта и гибрида.

Вычислим средние значения СР (70,2 кг) и КВ (60,5%). Полученные координаты отметим в виде точки. Через эту точку проведем перпендикулярные линии, разделяющие график на 4 квадранта с нумерацией по часовой стрелке, начиная с левого нижнего (рис. 15).

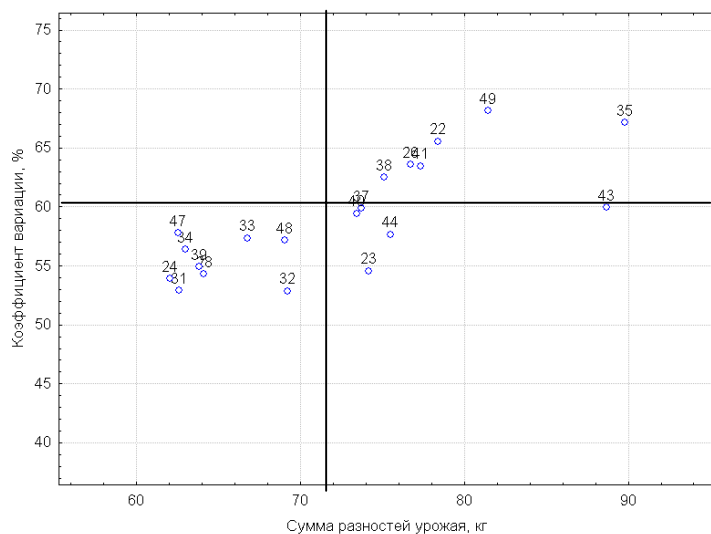


Рис. 15. Распределение сортов яблони

В первый и второй квадрант попали сорта и гибриды с низким урожаем и снизившие урожай за годы плодоношения, в третий квадрант – с высоким, но не стабильным урожаем, в четвертый квадрант – с высоким и стабильным урожаем (Джонатан 34-31, Джонатан 34-23, Джонатан 34-80, Джонатан 34-34, Джонатан 11-5).

Теперь проиллюстрируем работу предлагаемого метода на ягодных культурах. Объектом исследования послужили данные по урожайности 50 сортов и гибридов земляники садовой за период исследования, равный 3 годам. Каждому сорту и гибриду присвоили номер.

За 3 года средний урожай с куста всех 50 сортов и гибридов оказался 246,4 г. После проведения ранжирования значений СУ каждого сорта и гибрида по убыванию

и исключения из них давших значения СУ ниже 246,4 г осталось 23 сорта и гибрида.

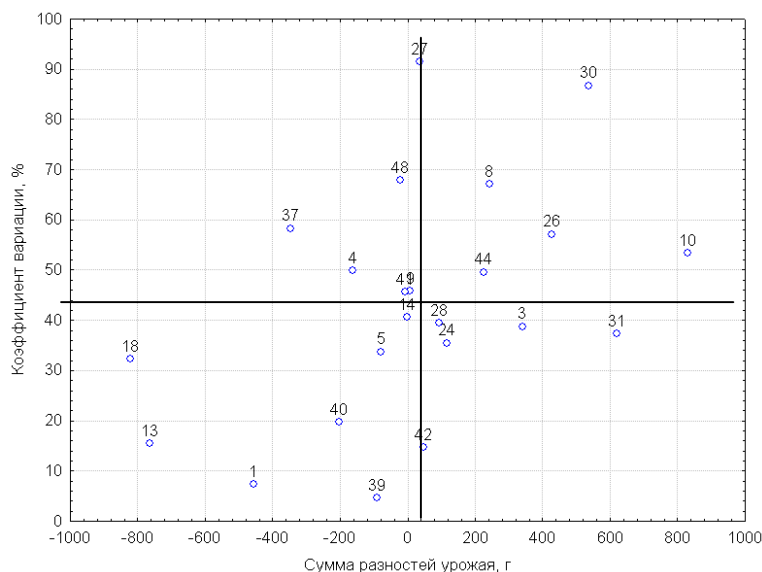


Рис. 16. Распределение сортов и гибридов земляники

По каждому из этих 23 сортов и гибридов на декартовой системе координат на оси абсцисс отложим сумму разностей урожаев (СУ) каждого сорта и гибрида, а на оси ординат – коэффициент вариации (КВ). Отметим точки с соответствующими координатами и проставим соответствующий номер сорта.

Вычислим средние значения СУ (7,5 г) и КВ (40,5%) и полученные координаты отметим в виде точки. Через эту точку проведем перпендикулярные линии, разделяющие график на 4 квадранта, с нумерацией по часовой стрелке, начиная с левого нижнего (рис. 16).

В первый и второй квадрант попали варианты с низким урожаем и снизившие урожай за годы плодоношения, в третий квадрант – варианты с высоким, но не стабильным урожаем, в четвертый квадрант – варианты с высоким и стабильным урожаем (Гера × Белруби (10-1-39), Зефир × св.оп. (3-8), Редчиф, Белруби × Трубадур (4-2-5), Веспер × св.оп. (5-2-4).

8. МЕТОД ОТБОРА КУСТОВ ЗЕМЛЯНИКИ ДЛЯ УСКОРЕННОЙ СЕЛЕКЦИИ

Известные способы селекции земляники предполагают использование при отборе кустов в основном селекционно ценных признаков, например, урожайности, сроков созревания.

Все эти способы предусматривают прямой отбор кустов, причем не учитываются молодые кусты, т.е. не проводится селекционная работа на ранних стадиях.

Предлагаемый нами метод с помощью морфологических признаков листа позволит проводить отбор земляники для ускоренной селекции ее кустов, обладающих повышенной урожайностью.

Чтобы решить эту задачу, необходимо сформировать обучающую выборку кустов земляники (не менее 30), присвоив каждому кусту индивидуальный номер. Отобранные кусты описывают по морфометрическим характеристикам центрального листочка типичного листа, а именно: длине (DSD); наибольшей ширине (SSD); длине собственного черешочка (CSD); расстоянию по центральной жилке от основания листочка до места, где листочек имеет наибольшую ширину (RSD); количеству зубчиков (ZSD).

Проводят кластерный анализ кустов обучающей выборки на основе оценки их морфометрических характеристик в евклидовой плоскости главных компонент. Кластеризацию осуществляют исходя из морфометрических признаков листа методом Уорда.

Вычисляют по кластерам средние значения морфометрических характеристик центрального листочка ти-

пичного листа, а именно: длину (DSD); наибольшую ширину (SSD); длину собственного черешочка (CSD); расстояние по центральной жилке от основания листочка до места, где листочек имеет наибольшую ширину (RSD); количество зубчиков (ZSD). Далее находят индексы, являющиеся отношением этих средних значений величин, как то: расстояния по центральной жилке от основания листочка до места, где листочек имеет наибольшую ширину, к его длине (I1); числа зубчиков к длине листочка (I2); длины собственного черешочка к длине листочка (I3).

Дождавшись плодоношения, в обучающей выборке определяют среднюю урожайность в каждом кластере (UR). Полученные кластеры сравнивают по средней урожайности кустов и определяют кластер с наибольшей средней урожайностью.

Все кусты обучающей выборки ранжируют по урожайности и полученный ряд делят по медиане, т.е. значению признака, разделяющему выборку на две части. Отбирают кусты из части ряда с наибольшей урожайностью.

Затем вновь производят ранжирование кустов всей обучающей выборки по признаку DSD и определяют его медиану. Сравнивают средние значения этого признака во всех 4 кластерах, чтобы определить, какую часть ранжированного ряда выбрать – в сторону увеличения или уменьшения значений от медианы. Выбрав наиболее подходящую часть ряда по признаку DSD, определяют номера кустов, совпавших с отобранными по урожайности.

Вновь ранжируют кусты всей обучающей выборки по индексу I1 и находят его медиану. Вновь сравнивают средние значения этого индекса во всех 4 кластерах, чтобы определить, какую часть ранжированного ряда выбрать – в сторону увеличения или уменьшения значений от медианы. Выбрав наиболее подходящую часть ряда по индексу I1, определяют номера кустов, совпавших с отобранными по DSD.

Аналогично проводят ранжирование кустов всей обучающей выборки и определяют медиану с последующим отбором подходящей части ряда по индексам I2 и I3, каждый раз отбирая кусты, номера которых совпали с предыдущим отбором.

В завершающей стадии отбора сравнивают номера оставшихся кустов с номерами кустов кластера с наибольшей средней урожайностью. Те кусты, номера которых совпали, пригодны для селекции.

На основании морфометрических признаков центрального листочка типичного листа и его индексов последней оставшейся выборки кустов составляют описание каждого сорта земляники, которое используют при массовом отборе ее кустов в полевых условиях для дальнейшей селекции.

Все известные способы отбора кустов не учитывали морфологические признаки листа и не позволяли отбирать кусты земляники для селекции на ранних стадиях.

С разработкой нашей методики это стало возможным. Например, сформирована обучающая выборка из 86 кустов земляники сорта Трубадур. Проведен кластерный анализ методом Уорда кустов обучающей выборки на основе оценки морфометрических характеристик между

ними в евклидовой плоскости главных компонент. В результате обнаружено четыре группы кустов, сходных по структуре изменчивости морфологии листа. Средние значения морфометрических признаков и средняя урожайность (UR) в выделенных кластерах приведены в табл. 9.

Таблица 9

Средние значения морфометрических признаков, их индексов и урожайности кустов земляники сорта Трубадур в выделенных кластерах

№	UR	DSD	SSD	CSD	ZSD	RSD	I1	I2	I3
1	105,9	42,2	38,2	6,1	15,4	21,6	0,513	0,366	0,145
2	101,3	37,6	31,2	4,1	14,1	19,9	0,530	0,407	0,111
3	201,9	54,0	48,3	9,0	16,3	28,8	0,534	0,303	0,168
4	99,8	48,5	43,2	7,0	16,0	24,3	0,502	0,332	0,145

В табл. 9 кластер с наиболее урожайными кустами фигурирует под номером 3. Критерий отбора кустов обучающей выборки осуществляют ранжированием кустов по урожайности, DSD и индексам I1, I2, I3 с последующим делением по медиане и сравнением номеров кустов с номерами кустов предыдущего отбора.

В обучающей выборке по урожайности медиана равна 100 г. Отбирают кусты с наибольшей урожайностью: 1, 2, 9, 10, 11, 12, 13, 16, 18, 21, 22, 24, 27, 28, 29, 33, 34, 39, 40, 41, 43, 44, 45, 46, 49, 50, 55, 67, 70, 73, 74, 75, 77, 78, 79, 80, 81, 82, 84, 85, 86. В итоге отобран 41 куст.

В обучающей выборке по признаку DSD медиана равна 50 мм. Так как кусты кластера 3 относятся к наиболее крупнолистным, отбирают кусты со значением

признака $DSD \geq 50$ мм и сравнивают их номера с номерами наиболее урожайных кустов. Отобранные кусты имели номера: 1, 2, 12, 13, 18, 22, 39, 40, 44, 45, 46, 55, 70, 73, 74, 75, 78, 81, 82, 84, 85. В результате отобран 21 куст.

В обучающей выборке по индексу $I1$ медиана равна 0,526. Так как кусты кластера 3 имеют наибольшее значение этого индекса, отбирают кусты со значением индекса $I1 \geq 0,526$ и сравнивают номера кустов с предыдущим отбором. После сравнения отобранные кусты имели номера: 1, 2, 40, 44, 45, 73, 74, 78, 81, 82, 84, 85. В итоге отобрано 12 кустов.

В обучающей выборке по индексу $I2$ медиана равна 0,314. Так как кусты кластера 3 имеют наименьшее значение этого индекса, отбирают кусты со значением индекса $I2 \leq 0,314$ и сравнивают номера кустов с предыдущим отбором. Отобранные кусты имели номера: 44, 45, 73, 78, 81, 82, 84, 85. В результате отобрано 8 кустов.

В обучающей выборке по индексу $I3$ медиана равна 0,152. Так как кусты кластера 3 имеют наибольшее значение этого индекса, отбирают кусты со значением индекса $I3 \geq 0,152$ и сравнивают номера кустов с предыдущим отбором. Отобранные кусты имели номера: 44, 45, 73, 78, 81, 82, 84, 85. Номера отобранных 8 кустов совпали с номерами кустов, отобранных по индексу $I2$.

При сравнении номеров отобранных 8 кустов с номерами кустов кластера 3 обнаружено их полное совпадение, следовательно, эти кусты пригодны для селекции.

Было составлено описание отобранных 8 кустов, отличающихся высокой урожайностью, они идентифицированы визуально по следующим характеристикам: листья крупные, центральный листочек типичного листа

овальный, с крупными зубцами и длинным собственным черешочком. Эти легко узнаваемые признаки были использованы при массовом отборе кустов земляники сорта Трубадур в полевых условиях для их дальнейшей селекции.

В приведенном примере отбор по индексам I2 и I3 продемонстрировал идентичность отобранных кустов, что является вовсе необязательным. Проведенные эксперименты на землянике сортов Белруби, Санрайз, Львовская ранняя, Гера, 50 лет Октября показали, что важно совпадение номеров кустов, отобранных по индексу I3, с номерами кустов из кластера с наибольшей средней урожайностью, что повышает точность отбора кустов для ускоренной селекции на ранних стадиях.

9. ИЗМЕНЧИВОСТЬ РАЗМЕРНО-ВЕСОВЫХ ПРИЗНАКОВ И ХИМИЧЕСКОГО СОСТАВА У ПЛОДОВ ЗЕМЛЯНИКИ

Наряду с высокой урожайностью, выносливостью к неблагоприятным условиям, устойчивостью к наиболее опасным болезням и вредителям сорта земляники должны обладать высокими товарными, вкусовыми, технологическими и целебными свойствами ягод.

Ягоды должны быть крупными, правильной формы, относительно одномерными, с привлекательной равномерной интенсивной окраской, обладать высокой транспортабельностью, хорошо храниться, содержать большое количество биологически активных веществ. Поэтому требуется совершенствование методов оценки этой категории признаков.

Материалом для работы послужило описание 42 сортов и гибридов земляники по следующему комплексу признаков: SVB – средняя высота ягоды, мм; SHB – средняя ширина ягоды, мм; SVES – средний вес ягоды, г; SV – содержание сухих веществ, мг/%; OS – содержание общих сахаров, мг/%; ОК – общая кислотность, мг/%; VC – содержание витамина С, мг/%; КАТ – содержание катехинов, мг/%; АНТ – содержание антоцианов, мг/%.

Работа была начата с количественной оценки изменчивости между сортами и гибридами земляники по всем учтенным признакам. Эту задачу мы решили с помощью однофакторного дисперсионного анализа. В результате трехлетних наблюдений обнаружены статистически достоверные различия между сортами и гибридами практически по всем признакам, за исключением содер-

жания сухих веществ. Доля учтенного фактора составила от 19,9 (витамин С) до 49,9% (средняя ширина ягоды).

Результат дисперсионного анализа по годам выращивания дал отрицательный результат, не обнаруживший статистически достоверных различий. Это можно интерпретировать как хорошую стабильность учтенных признаков.

Так как в исследование были вовлечены размерно-весовые признаки и признаки химического состава следовало решить вопрос объединения их в комплекс. Как известно, комплекс признаков состоит из связанных друг с другом переменных. Эту связь и требовалось обнаружить. Традиционный подход с помощью корреляционного анализа ничего не дал. Достоверных корреляций между признаками из разных групп обнаружить не удалось.

Для решения этой проблемы мы обратились к одному из многомерных статистических методов – анализу канонических корреляций, позволяющему находить связи между двумя группами признаков. В результате обнаружено, что размерно-весовые признаки и признаки химического состава имеют положительную корреляцию средней силы ($R = 0,42$; $p < 0,05$).

Полученный результат дал основание для объединения всех учтенных признаков в комплекс. Это осуществлено с помощью метода главных компонент. Установлено, что первые 7 главных компонент учитывают в совокупности 95% общей дисперсии. Эти компоненты были использованы для группировки сортов и гибридов с помощью кластерного анализа (рис. 17).

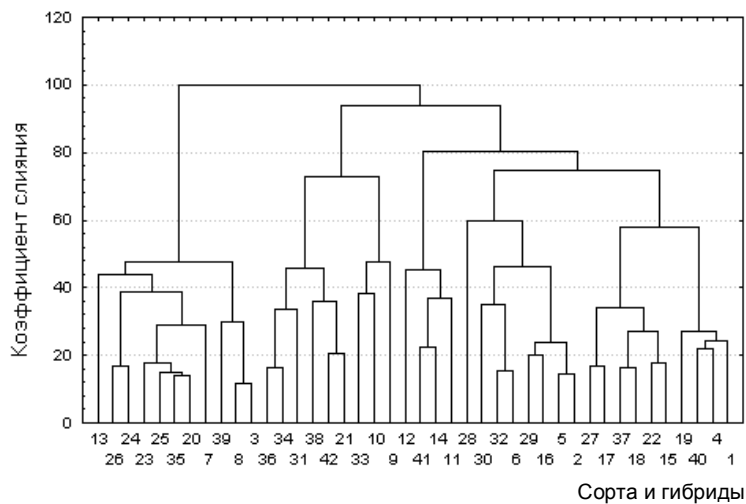


Рис. 17. Кластерный анализ сортов и гибридов

В результате выделено 6 кластеров сортов и гибридов в составе:

1-й кластер – 13, 26, 24, 23, 25, 35, 20, 7, 8, 3;

2-й кластер – 36, 34, 31, 38, 42, 21;

3-й кластер – 33, 10, 9;

4-й кластер – 12, 41, 14, 11;

5-й кластер – 28, 30, 32, 29, 16, 5, 2;

6-й кластер – 27, 17, 37, 18, 22, 15, 19, 40, 4, 1.

Правильность кластерного решения проверена с помощью дисперсионного анализа с фактором «кластер». Установлены статистически достоверные различия между группами практически по всем признакам, за исключением среднего веса ягоды, сухих веществ и содержания катехинов. Доля соответствующей дисперсии колебалась от 16,5 (средняя высота ягоды) до 56,6% (антоцианы).

Средние по кластерам значения признаков, показавших статистически достоверные различия, представлены в табл. 10.

Таблица 10

Среднее значения признаков в выделенных группах сортов и гибридов земляники садовой

Кластер (группа)	Количество сортов	SVB	SHB	OS	OK	VS	ANT
1	10	30,8	25,59	7,35	1,13	64,08	101,8
2	6	26,2	20,91	7,42	0,9	64,37	89,5
3	3	26,6	22,59	9,30	1,17	70,46	134,2
4	4	31,7	24,82	7,25	1,18	65,37	198,4
5	7	25,3	23,67	7,28	1,23	78,30	25,1
6	10	26,7	24,69	7,49	1,24	62,19	125,8

Обнаруженные различия позволяют сделать описание каждой группы сортов и гибридов по максимальному и минимальному значениям признака в группе:

группа 1 – ягоды широкой формы;

группа 2 – ягоды узкой формы, с низкой кислотностью и слабой окраской;

группа 3 – ягоды сладкого вкуса;

группа 4 – ягоды удлиненной формы с интенсивной окраской и низким содержанием сахаров;

группа 5 – небольшие ягоды с повышенным содержанием витамина С;

группа 6 – ягоды кислого вкуса с низким содержанием витамина С.

Полученное описание можно использовать для характеристики сортов и гибридов при проведении сортоиспытания и селекции.

ЗАКЛЮЧЕНИЕ

Современного селекционера невозможно представить без знания математических методов, применяемых для описания и анализа экспериментальных данных. Эти знания позволяют полнее и в более обобщенном виде описывать закономерности биологических явлений, понимать их суть и, самое главное, повышать доказательность умозаключений и выводов, поскольку современная культура экспериментального исследования требует убедительных статистических подтверждений.

Нельзя забывать, что сама по себе статистика – только инструментарий, помогающий селекционеру эффективно разобраться в сложном экспериментальном материале. Наиболее важным в любом эксперименте является четкая постановка задачи, тщательное планирование эксперимента, построение непротиворечивых гипотез.

Следует обратить особое внимание на возможность использования современным исследователем пакетов статистических программ, освобождающих его от непосредственных трудоемких вычислений. К наиболее распространенным и универсальным относятся следующие программные продукты: Statistica, StatGraphics, NCSS, SPSS. В основном все пакеты совпадают, но существенно различаются в деталях. В разных пакетах одни и те же методы представлены в несколько отличных версиях и с разной полнотой. Пакеты отличаются по максимальному объему данных, которые можно включить в анализ, по способу построения изображений и т.п. Опыт показывает, что пакет статистических программ Statistica позволяет успешно решить примерно 90% типичных задач. Од-

нако в некоторых случаях приходится обращаться и к другим программным продуктам. Кроме того, в Интернете можно найти множество частных статистических программ, адаптированных для решения относительно узких задач. Обычно эти задачи решаются с помощью пакета общего назначения, но такое решение более трудоемко.

Математическая статистика в руках селекционера может и должна быть мощным инструментарием, позволяющим не только успешно ориентироваться в большом числе экспериментальных данных, но и способствовать становлению его объективного мышления. Правильное применение статистики позволит селекционеру:

- 1) доказывать правильность и обоснованность используемых методических приемов и методов;
- 2) обобщать данные эксперимента;
- 3) находить зависимости между экспериментальными данными;
- 4) выявлять наличие существенных различий между группами растений (например, испытуемыми и контрольными);
- 5) строить статистические предсказания;
- 6) избегать логических и содержательных ошибок и др.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Прикладная статистика. Классификация и снижение размерности. М., 1989. 342 с.
2. Андерсон Т. Введение в многомерный статистический анализ. М., 1963. 500 с.
3. Андрукович П.Ф. Применение метода главных компонент в практических исследованиях. М., 1973. 234 с.
4. Волчков Ю.А., Кузнецова А.П., Алексеенко А.В. Статистический метод оценки устойчивости сортов плодовых культур к грибным болезням по данным многолетних наблюдений // Методики опытного дела и методические рекомендации Северо-Кавказского зонального НИИ садоводства и виноградарства. Краснодар, 2002. С. 116–118.
5. Горшкова В.В., Щеглов С.Н. Экоэлементная структура сортов земляники и ее воспроизведение при вегетативном размножении. Сообщение I: Оценка устойчивости передачи комплекса признаков морфологических признаков листа // Актуальные вопросы экологии и охраны природы экосистем южных регионов России и сопредельных территорий. Краснодар, 1997. С. 276–277.
6. Джениферс Дж. Введение в системный анализ: применение в экологии. М., 1981. 250 с.
7. Доспехов В.А. Методика полевого опыта. М., 1979. 416 с.
8. Драгавцев В.А., Драгавцева И.А., Лопатина Л.М. Управление продуктивностью сельскохозяйственных культур на основе закономерностей их генетических и

фенотипических изменений при смене лимитов внешней среды. Краснодар, 2003. 210 с.

9. Драгавцева И.А., Дьяков А.Б. Возможности использования комплекса биометрических методов для оценки адаптивности сортов плодовых культур // *Формы и методы повышения экономической эффективности регионального садоводства и виноградарства. Организация исследований и их кооперация*. Краснодар, 2001. Ч. 1. С. 57–65.

10. Драгавцева И.А., Дьяков А.Б. Метод сравнительной оценки экологической адаптации видов и сортов плодовых культур // *Методики опытного дела и методические рекомендации Северо-Кавказского зонального НИИ садоводства и виноградарства*. Краснодар, 2002. С. 82–89.

11. Драгавцева И.А., Дьяков А.Б. Проблемы измерения экологической стабильности оценок урожайности сортов возделываемых растений // *Организационно-экономический механизм инновационного процесса и приоритетные проблемы научного обеспечения развития отрасли*. Краснодар, 2003. С. 64–69.

12. Драгавцева И.А., Запорожец Н.М., Луценко Е.В., Луценко Н.Е. Новые подходы к районированию плодовых культур на юге России с применением компьютерного моделирования // *Организационно-экономический механизм инновационного процесса и приоритетные проблемы научного обеспечения развития отрасли*. Краснодар, 2003. С. 74–76.

13. Драгавцева И.А., Лопатина Л.М. Изучение отклика плодовых культур на изменение условий среды // *Системообразующие и экологические факторы и крите-*

рии зон устойчивого развития плодового хозяйства на Северном Кавказе. Краснодар, 2001. С. 232–233.

14. *Иберла К.* Факторный анализ. М., 1980. 400 с.

15. *Кендалл М., Стьюарт А.* Многомерный статистический анализ и временные ряды. М., 1976. 302 с.

16. *Ким Дж., Мьюлеер Ч.* Факторный анализ: статистические методы и практические вопросы // Факторный, дискриминантный и кластерный анализ. М., 1989. С. 5–74.

17. *Клекка У.Р.* Дискриминантный анализ // Факторный, дискриминантный и кластерный анализ. М., 1989. С. 78–137.

18. *Ключникова Г.Н., Улитин В.О.* К методике оценки реакции сортов винограда на погодные условия // Организационно-экономический механизм инновационного процесса и приоритетные проблемы научного обеспечения развития отрасли. Краснодар, 2003. С. 411–415.

19. *Кольцов Ю.В., Коршиков А.Ф., Матвиенко В.Ю.* Обзор методов моделирования в экологических исследованиях // Наука Кубани. 2001. №1. С. 10–17.

20. *Лакин Г.Ф.* Биометрия. М., 1990. 352 с.

21. *Лапшин В.И.* Изменчивость морфологических признаков листа в гибридных семьях земляники // Пути интенсификации и кооперации в селекции садовых культур и винограда. Краснодар, 2002. С.155–157.

22. *Лопатина Л.М.* О создании компьютерно-методологического комплекса для селекционно-производственного мониторинга возделывания плодовых культур // Формы и методы повышения экономической эффективности регионального садоводства и виноградарства. Организация исследований и их кооперация. Краснодар, 2001. Ч. 1. С. 65–69.

23. *Лопатина Л.М.* Планирование экологических испытаний и оценка экологической пластичности сортов и гибридов сельскохозяйственных культур с помощью регрессионных моделей // *Вестн. с.-х. науки.* 1986. №4. С. 71–79.

24. *Лопатина Л.М., Клочков С.А.* Оценка генетической детерминации гомеостаза сортов яровой пшеницы // *Докл. РАСХН.* 1995. №1. С. 9–12.

25. *Лопатина Л.М., Кудряшов И.Н.* Роль теории планирования в многофакторных полевых экспериментах // *Научные труды Краснодарского НИИСХ им. П.П. Лукьяненко.* 1996. С. 156–161.

26. *Ляпунов А.А., Багриновская Г.П.* О методологических вопросах математической биологии. М., 1975. 130 с.

27. *Меттлер Л., Грегг Т.* Генетика популяций и эволюция. М., 1972. 324 с.

28. *Олдендерфер М.С., Блэшфилд Р.К.* Кластерный анализ // *Факторный, дискриминантный и кластерный анализ.* М., 1989. С. 139–201.

29. *Рокицкий П.Ф., Савченко В.К., Добина А.И.* Генетическая структура популяций и ее изменения при отборе. М., 1977. 200 с.

30. *Савченко В.К.* Генетический анализ и синтез в практической селекции. Минск, 1986. 95 с.

31. *Сажин Ю.В., Басова В.А.* Многомерные статистические методы. М., 2002. 163 с.

32. *Сергеева Н.Н., Улитин В.О.* Влияние различных факторов на урожайность яблони // *Организационно-экономический механизм инновационного процесса и*

приоритетные проблемы научного обеспечения развития отрасли. Краснодар, 2003. С. 117–122.

33. *Смиряев А.В., Гохман М.В.* Биометрические методы в селекции растений. М., 1985. 180 с.

34. *Теренько Г.Н., Шафоростова Н.К., Мазурик И.А.* Оценка урожайности плодового дерева // Методики опытного дела и методические рекомендации Северо-Кавказского зонального НИИ садоводства и виноградарства. Краснодар, 2002. С. 60–67.

35. *Тюрин В.В., Морев И.А., Волчков Ю.А.* Дискриминантный анализ в селекционно-генетических исследованиях. Краснодар, 2003. 23 с.

36. *Щеглов Н.И., Щеглов С.Н.* Статистические методы, применяемые в селекции плодовых и ягодных культур // Садоводство и виноградарство 21 века. Ч. 3. Итоги и перспективы селекционных исследований: Матер. Междунар. науч.-практ. конф. Краснодар, 1999. С. 157–160.

37. *Щеглов С.Н.* Изменчивость морфологических признаков листа в различных популяциях земляники лесной (*Fragaria vesca* L.) // Актуальные вопросы экологии и охраны природы экосистем южных регионов России и сопредельных территорий. Краснодар, 2000. С. 51–52.

38. *Щеглов С.Н.* Статистические методы оценки селекционного материала плодовых и ягодных культур // Программа (проект) селекционных работ по плодовым, ягодным, орехоплодным и цветочно-декоративным культурам союза селекционеров Северного Кавказа на период до 2010 г. Краснодар, 2002. С. 241–257.

39. *Щеглов С.Н., Ефимова И.Л., Павлова С.В.* Использование многомерных статистических методов для

изучения экологической пластичности сортов и гибридов яблони // Актуальные вопросы экологии и охраны природы экосистем южных регионов России и сопредельных территорий. Краснодар, 2003. С. 64–66.

40. Яковенко В.В., Подорожный В.Н., Щеглов С.Н., Лапшин В.И. Меж- и внутрисемейная изменчивость комплекса селекционно ценных признаков земляники // Пути интенсификации и кооперации в селекции садовых культур и винограда. Краснодар, 2002. С.151–154.

41. Eberhart S.A., Russel W.A. Stability parameters for comparing varieties // Crop Science. 1966. Vol.6. №1. P.36–40.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
1. МЕТОДЫ МОДЕЛИРОВАНИЯ В ГЕНЕТИКО-СЕЛЕКЦИОННЫХ ИССЛЕДОВАНИЯХ.....	6
2. МНОГОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ.....	28
2.1. Метод главных компонент и факторный анализ.....	28
2.2. Кластерный анализ.....	36
2.3. Дискриминантный анализ.....	42
3. ИСПОЛЬЗОВАНИЕ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МЕТОДОВ ДЛЯ ИЗУЧЕНИЯ ЭКОЛОГИЧЕСКОЙ ПЛАСТИЧНОСТИ СОРТОВ И ГИБРИДОВ ЯБЛОНИ.....	50
4. КЛАССИФИКАЦИЯ КОЛЛЕКЦИИ СОРТОВ ЯБЛОНИ ПО КОМПЛЕКСУ ПРИЗНАКОВ.....	57
5. МИНИМИЗАЦИЯ ЭФФЕКТОВ УСЛОВИЙ ГОДА ПРИ ОЦЕНКЕ ГЕНЕТИЧЕСКИ ОБУСЛОВЛЕННЫХ РАЗЛИЧИЙ В УСТОЙЧИВОСТИ ЧЕРЕШНИ К МОНИЛИОЗУ МЕТОДОМ ДИСКРИМИНАНТНОГО АНАЛИЗА.....	65
6. ИДЕНТИФИКАЦИЯ ПРЕДСТАВИТЕЛЕЙ РОДА PRUNUS L. ПО МОРФОЛОГИЧЕСКИМ ПРИЗНАКАМ ЛИСТА И КОСТОЧКИ ПЛОДА.....	76
7. МЕТОД РАСПОЗНАВАНИЯ СОРТОВ И/ИЛИ ГИБРИДОВ ПЛОДОВЫХ И ЯГОДНЫХ КУЛЬТУР, ОБЛАДАЮЩИХ ВЫСОКИМ И СТАБИЛЬНЫМ УРОЖАЕМ....	83
8. МЕТОД ОТБОРА КУСТОВ ЗЕМЛЯНИКИ ДЛЯ УСКОРЕННОЙ СЕЛЕКЦИИ.....	88
9. ИЗМЕНЧИВОСТЬ РАЗМЕРНО-ВЕСОВЫХ ПРИЗНАКОВ И ХИМИЧЕСКОГО СОСТАВА У ПЛОДОВ ЗЕМЛЯНИКИ.....	94
ЗАКЛЮЧЕНИЕ.....	98
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	100

Научное издание

Щеглов Сергей Николаевич

**ПРИМЕНЕНИЕ БИОМЕТРИЧЕСКИХ МЕТОДОВ
ДЛЯ УСКОРЕНИЯ СЕЛЕКЦИОННОГО ПРОЦЕССА
ПЛОДОВЫХ И ЯГОДНЫХ КУЛЬТУР**

Подписано в печать 11.01.05. Печать офсетная.
Формат 60×84 1/16. Гарнитура «Таймс». Бумага офсетная.
Уч.-изд.л. 6,7. Тираж 700 экз. Заказ № 38

Кубанский государственный университет
350040 г. Краснодар, ул. Ставропольская, 149.
Типография Кубанского государственного университета.
350023 г. Краснодар, ул. Октябрьская, 25.