

ПРИМЕНЕНИЕ ДИСПЕРСИОННОГО И ФАКТОРНОГО АНАЛИЗА В ОБРАБОТКЕ ДАННЫХ ЭКОЛОГИЧЕСКОГО СОРТОИСПЫТАНИЯ КАРТОФЕЛЯ

Оплеухин А.А., Стрельцова Т.А.

Целью работы является подбор методов статистической обработки способных представить данные по экологическому сортоиспытанию в максимально информативной форме и помочь выявить закономерности влияния экологических факторов на генотип. Для этой цели использовали дисперсионный анализ и метод главных компонент.

ВВЕДЕНИЕ

Биологические данные являются, как правило, сложными совокупностями, для адекватной оценки которых, зачастую, необходимо проводить статистическую обработку. При этом ставятся цели более достоверного предоставления данных, во-первых, а во-вторых, для удобства интерпретации.

Экологическое сортоиспытание требует от статистической обработки: как типичный эксперимент — проверить статистическую значимость действия факторов; в связи с тем что фиксируются как правило несколько признаков — наиболее адекватно и компактно представить все признаки для возможности интерпретации их изменчивости под действием многообразия факторов.

МАТЕРИАЛЫ И МЕТОДЫ

При проведении статистической обработки использованы экспериментальные данные, полученные в лаборатории экологической генетики и селекции растений ГАГУ [1-3].

Исходными для дисперсионного анализа были данные по экологическому сортоиспытанию и генетическому мониторингу (2005-2007 гг.) четырех среднеспелых сортов картофеля в трех пунктах Горного Алтая — в с. Майме (предгорье), с. Усть-Коксе (среднегорье) и с. Улагане (высокогорье). Испытуемая коллекция в этой серии экспериментов включала генотипы – Накра, Аспия, Никулинский, Кетский.

Испытания проводились согласно "Методическим указаниям по экологическому сортоиспытанию картофеля", с использованием и других современных методик [4-10]. Полевые опыты размещались синхронно по вертикальной зональности, площадь делянок 7,5 м, повторность 4-х-кратная, размещение рендомизированное. Выбор для анализа этих данных связан с необходимостью повторения каждого варианта фактора для просчета взаимодействия факторов в дисперсионном анализе. Данные прошли процедуру нормализации.

В методе главных компонент происходит редукция переменных, в результате две-три главные компоненты представляют большую часть изменчивости признаков [11], понятно, что чем больше исходных признаков введено в анализ, тем информативнее будут полученные данные. В связи с этим, исходными для применения многомерных методов были данные экологического сортоиспытания картофеля по пяти ранним сортам (Новосибирский, Корине, Приекульский ранний, Алмаатинский, Уральский сувенир, так же в исследование был включен сорт Приекульский ранний, оздоровленный методом апикальной меристемы).

В исследовании (2-я серия экспериментов) за три года (1994-1996 гг.), в двух пунктах (с. Бирюля - среднегорье и с. Усть-Уба - низкогорье), учитывались восемь признаков: масса клубней с одного куста, число клубней с куста, масса одного клубня, число стеблей на куст, высота стеблей, степень пораженности клубней гнилями, паршой, фитофторой.

Полученные данные обрабатывались с помощью компьютерных программ «SNEDEKOR», «Excel» и «STATISTICA».

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Рассмотрим применение дисперсионного анализа. Данные первой серии экспериментов (2005-2007 гг.) первоначально подвергались многофакторному дисперсионному анализу с разными подходами – общим и дробным, для выделения влияния высокогорья и среднегорья (табл. 1).

Влияние, всех факторов и их взаимодействий оказалось статистически значимым.

Таблица 1

Относительная доля влияния факторов и их взаимодействий на изменчивость массы клубней с 1 куста, %

Источник варьирования	Общий	Майма — Усть-кокса	Майма — Улаган
Пункт испытания (А)	15,2	16,8	3,7
Годы (В)	6,8	5,8	33,2
Генотипы (С)	13,5	8,8	18,0
Взаимодействие (АхВ)	34,1	38,0	13,3
Взаимодействие (ВхС)	3,6	4,4	12,0
Взаимодействие (АхС)	2,4	2,0	1,0
Взаимодействие (АхВхС)	14,5	14,8	6,6
Случайные отклонения	9,9	9,4	12,2

Как при общей обработке, так и при обработке лишь двух вариантов фактора «пункт испытания (А)» по Майме и Усть-Коксе, мы получили сходные данные доли дисперсии. Однако при сравнении «общего» - 3х градаций и 2х значений (Майма, Улаган) кардинально меняется распределение доли дисперсии по факторам: резко падает доля фактора А (пункт испытания), так же резко возрастает доля влияния фактора В (годы), и существенно (до 38%) - взаимодействия АхВ. Для выяснения причины обратимся к средним (рис. 1), так как дисперсионный анализ и есть сравнение средних по вариантам факторов с учетом как дисперсии между ними, так и внутри вариантов (случайная).

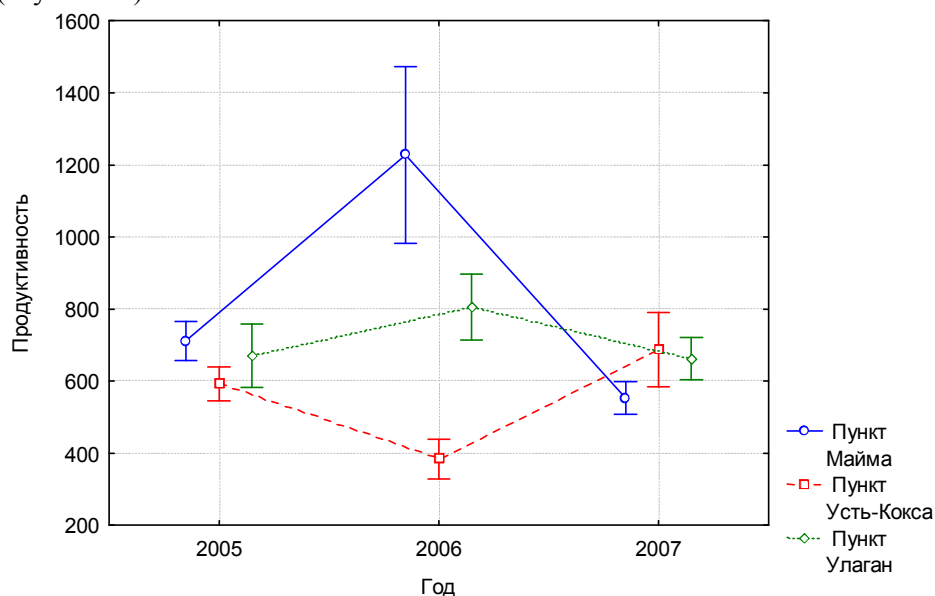


Рис. 1. Взвешенные средние. Смещение средних в зависимости от года и пункта. Вертикальные планки показывают 0,95 доверительный интервал.

Как видно из показателей рисунка 1, сдвиг доли дисперсии вызван средними по факторам, а фактор А и В взаимонезависимы, т.е. вследствие удаленности пунктов в один и тот же год может наблюдаться разная продуктивность обусловленная не столько экологическими условиями самих пунктов, сколько флуктуациями погодных условий.

Это можно представить как если бы фактор «пункты (А)» был генеральной совокупностью, а конкретные годы выборками из нее. Такой вариант в общем и рассматривается, но, однако, малое число выборок (годы) делает практически бессмысленным параметр доли вариации, так как он становится крайне подвержен сдвигам за счет сильного независимого (от фактора А) колебания фактора В. Дополнительно к этому устраняется условие того, что фактор должен иметь одно значение для всех групп, так как в один год в *разных пунктах* этот фактор будет иметь неоднозначное влияние.

Вследствие этого более рационально предварительно проводить анализ дисперсии факторов В (годы) и С (генотипы) в пределах одного значения фактора А, т.е. рассматривать вначале по двум факторам (В и С) в пределах каждого значения фактора А и только затем проводить трехфакторный анализ [12].

Это даст более достоверную картину за счет удаления возможной компенсации/усиления факторов А и В (табл. 2).

Таблица 2

Относительная доля влияния двух факторов и их взаимодействия на изменчивость массы клубней с 1 куста у сортов среднеспелой группы, %

Факторы	Пункты испытания		
	Майма	Усть-Кокса	Улаган
В (год)	53,7	47,9	16,7
С (генотип)	13,1	12,6	60,7
ВхС (взаимодействие)	21,3	31,2	8,3
Случайное отклонение	11,9	8,3	14,3

Таким образом, в многолетнем эксперименте, проводимом над несколькими сортами, выращиваемыми в различных экологических условиях, с учетом одного признака, следует проводить статистическую обработку данных по модели многофакторного дисперсионного анализа со случайным фактором [13]. Поэтому, вначале проводим общий трехфакторный анализ, включающий все данные, однако, от типа фактора зависит формула вычисления критерия Фишера-Снедекора. Потом, разбиваем данные для нескольких двухфакторных анализов, выясняя влияние факторов внутри отдельного пункта, т.е. для каждой градации фактора А (пункт испытания). Это позволяет устранить флуктуации, вызываемые независимыми метеорологическими условиями в разных пунктах, и более четко увидеть силу влияния факторов на исследуемый признак. Модель со случайными значениями фактора выявила недостоверность влияния факторов В (годы) и ВхС (взаимодействие метеоусловий и генотипа).

Рассмотрим применение факторного анализа (метод главных компонент, поворот осей методом варимакс) на данных второй серии экспериментов (1994-1996 гг.). Первоначально было проведено исследование нормальности распределения данных в пределах каждого признака по критерию Шапиро-Вилка и Колмогорова-Смирнова [14]. В результате признаки пораженности клубней болезнями были исключены из анализа, данные признаков числа стеблей и числа клубней – логарифмированы.

Факторные нагрузки (табл. 3), полученные на основе факторного анализа, позволили выявить корреляции (значимые при $p < 0,05$), между: массой клубня и массой клубней с куста ($r = 0,61$); массой клубня и количеством клубней ($r = -0,4520$); массой клубня и количеством стеблей ($r = -0,5726$); массой клубней с одного куста и количеством стеблей ($r = -0,4768$); массой клубня и высотой стебля ($r = -0,3477$).

Таблица 3

Таблица факторных нагрузок главных компонент

Признаки	1й фактор (47% общей дисперсии)	2й фактор (27% общей дисперсии)
Масса клубней с куста	-0,905535	0,260847
Масса клубня	-0,816296	-0,410709
Высота стебля	0,119295	0,679215
Количество клубней	0,064176	0,923875
Количество стеблей	0,744447	0,293536

Используя метод главных компонент, мы получили распределение объектов в пространстве первой и второй главных компонент, прежде всего по годам (рис. 2).

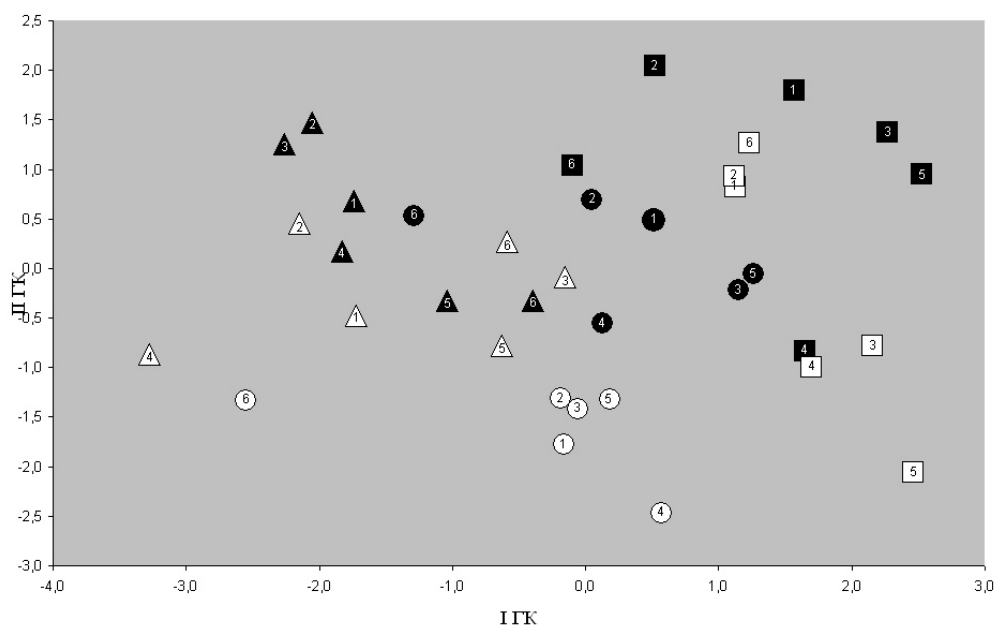


Рис. 2. Распределение объектов в пространстве главных компонент.

Числами обозначены сорта (1 - Новосибирский; 2 - Корине; 3 - Прикульский ранний; 4 - Алмаатинский; 5 - Уральский сувенир; 6 - Прикульский ранний; оздоровленный), формой годы (▲- 1994; ●- 1995; ■-1996), цветом пункты (Бирюля – черный; Усть-Уба - белый).

Пункты в основном разнятся по пространству второй главной компоненты. Таким образом, обращаясь к факторным нагрузкам, мы можем сказать, что растения, произрастающие в Усть-Убе (низкогорье), имеют более крупные клубни, меньшую высоту и меньшее число клубней. Облака в пределах одного года и пункта (т.е. все сорта) характеризуются направленностью: слева-сверху, вправо-вниз. Однако, распределение сортов не закономерно и эта форма, видимо, обусловлена указанными корреляционными зависимостями.

В дальнейшем планируется применить метод главных компонент для выявления различий между пунктами с разной высотной поясностью, по всей совокупности признаков, на возможно большей выборке, а так же регрессионный анализ для выявления возможной зависимости между механическим повреждением клубней и их дальнейшим поражением болезнями, дискриминантный анализ для выявления признаков наиболее отличительных у сортов разных групп спелости, происхождения и мест выращивания.

Литература

1. *Стрельцова Т.А.* Генотипическая и паратипическая изменчивость продуктивности картофеля в условиях Горного Алтая / Т.А. Стрельцова, О.В. Сафонова. // Сиб. вести, с.-х. науки. 2000. № 3-4. - С. 23-30.
2. *Стрельцова Т.А.* Рекомендации по внедрению адаптированных сортов картофеля в Горном Алтае / Т.А. Стрельцова - Горно-Алтайск: РИО ГАГУ, 2009. – 47 с.
3. *Стрельцова Т.А.* Картофель в Горном Алтае / Т.А. Стрельцова // Монография. – Новосибирск: Универсальное книжное издательство, 2007. – 200 с.
4. Методические указания по экологическому сортоиспытанию картофеля / сост. С.Н. Карманов [и др.] - М.: Изд-во ВАСХНИЛ, 1982. - 14 с.
5. Методические указания по оценке и поддержанию мировой коллекции картофеля / ВИР., Сост. С.М. Букасов, А.Г. Зыкина, А.Я. Камераз и др. – Л., 1976. – 30 с.
6. Международный классификатор СЭВ. – Л., 1984. – 43 с.
7. Методические рекомендации по проведению исследований с картофелем. – УААН, Немешаево, 2002. – 182 с.
8. Методика исследования по культуре картофеля. - М.: Колос, 1976. – 225 с.
9. Методические указания по оценке отличимости, однородности, стабильности сортов картофеля и отбору отечественных сортов – эталонов. М.: Изд-во Россельхозакадемия, ВНИИКХ, 2000. – 28 с.
10. *Рокицкий П.Ф.* Биологическая статистика / П.Ф. Рокицкий. - Минск.: Высшая шк., 1994. - 328 с.
11. *Ефимов, В.М.* Многомерный анализ биологических данных: учебное пособие / В.М. Ефимов, В.Ю. Ковалева - 2-е изд., испр. и доп. - СПб.: ВИЗР, 2008. - 86 с.
12. *Оплеухин А.А.* Сравнительная оценка разных подходов к математическому анализу силы влияния экологических факторов на продуктивность растений / А.А. Оплеухин, Н.А. Нургаллиманова // Экология России и сопредельных территорий: материалы XIV МЭСК. - Новосибирск: Новосибирский гос. ун-т., 2009. - С. 241-242.
13. *Сорокин О.Д.* Прикладная статистика на компьютере / О.Д. Сорокин - 2-е изд., - Краснообск: ГУП РПО СО РАСХН, 200. — С. 15-18.
14. *Лемешко Б.Ю.* Сравнительный анализ критериев проверки отклонения распределения от нормального закона / Б.Ю. Лемешко, С.Б. Лемешко // Метрология. - 2005. № 2. - С. 3-23.

USE OF ANOVA AND FACTORIAL ANALYSIS IN DATA PROCESSING ECOLOGICAL TEST OF GRADES'S A POTATO

Opleuhin A.A., Streltsova T.A.

The purpose of this work - finding methods statistical processing capable to submit data of ecological test of a grades as much as possible informative form and help to reveal laws of influence ecological factors on a genotype. For this purpose used analysis of variance and method of principal components.