

Т. Литтл, Ф. Хиллз

СЕЛЬСКО -
ХОЗЯЙСТВЕННОЕ
ОПЫТНОЕ
ДЕЛО

**Планирование
и анализ**

*Перевод с английского
кандидата сельскохозяйственных наук
Б. Д. Кирюшина
Под редакцией и с предисловием
кандидата сельскохозяйственных наук
Д. В. Васильевой*



Москва Колос 1981

AGRICULTURAL EXPERIMENTATION

Design and Analysis

Thomas M. Little

Extension Biometrician, Emeritus
University of California, Riverside

F. Jackson Hills

Extension Agronomist
University of California, Davis

John Wiley and Sons

New York Chichester Brisbane Toronto

Сельскохозяйственное опытное дело. Планирование и анализ. Т. М. Литтл, Ф. Дж. Хиллз. / Пер. с англ. Б. Д. Кирюшина; Под ред. и с предисловием Д. В. Васильевой. — М.: Колос, 1981. — 320 с.

Авторы на конкретных примерах в логической последовательности излагают математические приемы и принципы статистической обработки данных, необходимых как при выборе схемы опыта, так и при интерпретации полученных результатов. Подробно обсуждаются корреляция и регрессия (линейная, криволинейная и множественная). Учитывая возможности биометрии в наши дни, все математические уравнения решаются с помощью портативных калькуляторов, последовательность операций на которых авторы подробно объясняют. Большое значение придается выявлению существенно различающихся средних величин в эксперименте. Подчеркивается, что в сельскохозяйственной научно-исследовательской литературе в настоящее время очень часто, из-за неправильной оценки средних, результаты опытов истолковываются без учета всех факторов. Знание метода дисперсионного анализа и метода ортогональных коэффициентов, а также других методов, обсуждаемых в книге, улучшит постановку, проведение и оценку результатов опытов. В приложении даны таблицы, необходимые для расчетов.

Иллюстраций 42, таблиц 99, список литературы — 8 названий.

Л 40100—284
035(01)—81 26—81. 3802010000

Copyright © 1978, by John Wiley and Sons, Inc.

© Перевод на русский язык, «Колос», 1981

В настоящее время наблюдается все возрастающий интерес к использованию математической статистики представителями самых различных отраслей науки. Статистические методы находят широкое применение в планировании экспериментов и обследований в сельском хозяйстве и биологии, при оценке параметров и проверке гипотез, при принятии решений и изучении сложных систем.

Заметно возросла квалификация специалистов-опытников и их требовательность к выбору математических методов. Исклчительно быстро растет спрос на специальную литературу по постановке опытов и анализу получаемых данных. Повышение комплексности проблем, решаемых постановкой сложных опытов, подчеркивает актуальность выхода в свет книги Т. Литтла и Ф. Хиллза «Сельскохозяйственное опытное дело. Планирование и анализ».

Курсы повышения квалификации специалистов по методике полевого опыта, начавшие свою работу с 1973 г. на базе кафедры земледелия и методики опытного дела Московской сельскохозяйственной академии имени К. А. Тимирязева, получили широкое признание. Настоящая книга послужит хорошим дополнением пособий Б. А. Доспехова по методике полевого опыта и будет ценным пополнением библиотеки экспериментатора.

Сочетание познаний широко известного математика Т. Литтла и специалиста агрономического профиля Ф. Хиллза обусловило доступность, лаконичность и простоту излагаемого материала, достаточно полно иллюстрированного примерами из сельскохозяйственной практики. Авторы предостерегают от упования на математические выводы из результатов опытов и указывают на возможность ошибочного заключения. Однако они полностью отвергают отрицательный результат как таковой, обращают внимание на необходимость предварительного обсуждения данных с коллегами и опытными специалистами, прежде чем принять окончательное решение.

Знание современных статистических методов необходимо не только для обработки полученных в опыте данных, когда многое уже нельзя исправить, но и на всех этапах эксперимента — от планирования до обработки полученных величин. Предлагаемая в переводе книга Т. Литтла и Ф. Хиллза поможет экспериментаторам правильно планировать опыты и обрабатывать результаты.

В книге математическая теория дается сжато в расчете на то, что исследователи, работающие в области сельского хозяйства, знакомы с основами математической статистики. Правильно оценивая роль статистики как важного вспомогательного средства исследования сложных биологических объектов, авторы подчеркивают, что современный полевой опыт должен иметь статистическое обоснование и базироваться на случайной выборке, то есть на рендомизации.

Использование способов случайного размещения вариантов на опытном участке — характерная особенность современного периода развития опытного дела и практического применения выборочного метода исследования, научные основы которого устанавливаются рядом теорем теории вероятности, доказанных русскими математиками П. Л. Чебышевым, А. М. Ляпуновым, А. А. Марковым, А. Н. Колмогоровым и другими.

В предисловии к изданию книги Д. У. Снедекора «Статистические методы в применении к исследованиям в сельском хозяйстве и биологии» В. Н. Перегудов пишет: «Английские и американские авторы руководств по применению статистики к биологии и агрономии, как и Д. У. Снедекор, мало уделяют внимания разъяснению вопроса о рендомизации». Далее он указывает, что нельзя ожидать, что абстрактные и в известной мере идеализированные условия, на которых строится математическая теория, могут быть полностью осуществлены в агрономической практике. Рендомизация является именно тем средством, которое в доступной степени приближает нас к абстрактной случайной выборке и которое дает нам известные гарантии в правильности практических выводов, полученных в результате статистической обработки экспериментальных данных.

Много сил и энергии для внедрения рендомизированных методов в практику приложил советский ученый Б. А. Доспехов — автор учебника «Методика полевого опыта» (с основами статистической обработки результатов исследований), выдержавшего 4 издания.

В настоящее время рендомизированные методы в опытной работе используют многие научно-исследовательские учреждения нашей страны.

В книге достаточно полно рассмотрена сущность дисперсионного анализа и применение критерия F . К сожалению, авторы дают лишь одну схему дисперсионного анализа. Подробно изложены основные современные схемы полевых опытов: полная рендомизация, рендомизированные повторения, латинский квадрат, расщепленные делянки и блоки (2- и 3-факторные опыты). Отдельной главой рассматриваются особенности опытов с многолетними травами, которые ежегодно в течение ряда лет скашивают неоднократно. Повторные учеты укосов в этом случае рассматриваются в качестве субделянок.

Дан критический анализ различий, связи и взаимозаменяемости разных методов. Приводятся конкретные примеры предпочтительного использования того или иного метода или критерия для анализа данных. Особого внимания заслуживают основательно разработанные способы оценки существенности различий между средними.

Детально обсуждается корреляция и регрессия, описаны сокра-

щенные статистические методы, в основе которых лежит анализ регрессии для равновеликой изменчивости переменных. Этот прием известен как метод ортогональных полиномов.

Для расчета линейной корреляции предлагается два способа — стандартный и сокращенный (с использованием коэффициента Спирмана). Дана подробная оценка существенности коэффициента корреляции, способы построения линии регрессии и определения ее доверительной зоны. Криволинейная зависимость представлена полиномами 2, 3, 4 и 5 степеней. Множественная корреляция и регрессия рассмотрены подробно на примере трех переменных и коротко — с большим числом переменных. Описано построение поверхности отклика.

Важное значение имеет глава, где рассмотрены необходимость и способы преобразования исходных величин, даны предпосылки дисперсионного анализа и его особенности в этом случае.

Должное внимание уделено анализу с использованием хи-квадрата и способам выделения существенных средних. В книге указано 3 способа: НСР, множественно-линейные тесты (критерий Дункана) и F-критерий.

Представляют также интерес шкалы для учета засоренности посевов и пораженности культур. Несомненно подобные шкалы могут найти широкое применение в сельскохозяйственных исследованиях.

Большое значение авторы придают аккуратности и точности проведения исследования. Основными моментами улучшения точности опыта авторы считают: повторность, тщательный отбор вариантов, совершенствование техники опыта, выбор экспериментального материала, увеличение единицы эксперимента (до оптимальных пределов), выполнение дополнительных измерений и использование эффективных схем опытов. Достаточно подробно описан анализ ковариации как прием, снижающий ошибку опыта.

Книга имеет в приложении много вспомогательных таблиц, в составлении которых авторы принимали непосредственное участие. Следует, однако, отметить, что в целом книга не может полностью удовлетворить всем требованиям начинающего специалиста сельскохозяйственного опытного дела. Использование греческого алфавита и начальных букв тех или иных терминов в качестве символики не всегда упрощает расчеты.

Учитывая, что в биологической литературе греческие буквы применяются сравнительно мало, считаем необходимым дать здесь греческий алфавит (только строчные буквы):

α — альфа	η — эта	ν — ню	τ — тау
β — бэта	θ — тэта	ξ — кси	υ — ипсилон
γ — гамма	ι — иота	$\omicron$ — омикрон	φ — фи
δ — дельта	κ — капта	π — пи	χ — хи
ε — эпсилон	λ — лямбда	ρ — ро	ψ — пси
ζ — зета	μ — мю	σ — сигма	ω — омега

Как уже указывалось выше, работа Т. Литтла и Ф. Хиллза написана как практическое пособие и предназначена для исследователей, знакомых с современными статистическими методами и совершенствующих свои познания в опытном деле. В связи с этим для полноценного усвоения книги, особенно начинающими исследователями, можно рекомендовать краткий список литературы, где специально рассматриваются агробиологические аспекты применения статистических методов и методики полевого опыта.

- Бейли Н. Статистические методы в биологии. Перевод с англ. М., «Мир», 1964.
Вольф В. Г. Статистическая обработка опытных данных. М., «Колос», 1966, 254 с.
Доспехов Б. А. Методика полевого опыта (с основами статистической обработки результатов исследований). М., «Колос», 1979, 416 с.
Переудов В. Н. Планирование многофакторных полевых опытов с удобрениями и математическая обработка их результатов. М., «Колос», 1978, 183 с.
Рокитский П. Ф. Биологическая статистика. Минск, «Вышэйшая школа», 1967, 326 с.
Снедекор Д. У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. Перевод с англ. М., «Сельхозгиз», 1961, 503 с.
Урбах В. Ю. Биометрические методы. М., «Наука», 1964, 415 с.

Д. В. ВАСИЛЬЕВА

Лишь немногие сельскохозяйственные исследователи могут найти время для глубокого изучения абстрактной и сложной математической статистики, однако безусловно они хотели бы иметь общее представление о той ее части, которая относится к планированию и проведению экспериментов. Другими словами, они хотели бы знать основные принципы математической статистики в такой мере, чтобы надлежащим образом планировать опыты и делать соответствующие выводы по результатам эксперимента. Именно этому и посвящается данная книга. Необходимая техническая часть математики ограничена в книге простыми арифметическими расчетами. Многие математические закономерности подаются просто, без формальных доказательств. Однако были предприняты все усилия для того, чтобы сохранить математическую логику и избежать опасности переупрощения излагаемого материала.

Отдельные главы посвящаются каждой из семи схем экспериментов, что, вероятно, составляет более 90% от всех схем, используемых в сельскохозяйственных исследованиях. Четыре главы посвящены корреляции и регрессии (линейная, криволинейная и множественная). Регрессия представлена также в контексте дисперсионного анализа как статистический метод, используемый при обследованиях (наблюдениях). Особого внимания заслуживает глава по использованию сокращенных методов в анализе регрессии, когда мы имеем одинаковый интервал (шаг варьирования) между экспериментальными вариантами или наблюдениями.

В каждом анализе большое значение придается детальной последовательности ведения расчетов необходимых статистических характеристик. Каждый анализ сопровождается объяснением его логики и аргументации. Целая глава посвящена предпосылкам, лежащим в основе дисперсионного анализа, и путям преобразования данных, выходящих за пределы предполагаемых закономерностей.

Особое внимание уделено выбору и определению достоверного различия одного из показателей в ряду нескольких величин. Как следует из имеющейся сельскохозяйственной литературы по опытному делу, с этим вопросом сопряжено много недоразумений. В результате этого не принимаются во внимание некоторые важные выводы, вытекающие из получаемых данных. В большей мере это относится к методу, известному под названием функционального дисперсионного анализа или метода ортогональных коэффициентов. Этот исключительно простой и

очень сильный («чувствительный») анализ неизвестен многим сельскохозяйственным исследователям. В книге подробно излагаются эти и другие методы выбора показателей. Предшественница этой книги — «Статистические методы в сельскохозяйственных исследованиях» — в течение нескольких лет составляла основу курса для сельскохозяйственных консультантов и специалистов, изучающих статистические методы. Исправленное и дополненное издание этой книги повышает эффективность ее использования как учебного пособия. В значительной степени улучшено изложение методов выявления существенных средних; дополнительно включены рекомендации по использованию калькуляторов с программой, упрощающей проведение дисперсионного анализа; даны объяснение и пример поверхности отклика и глава «Повышение точности эксперимента», в которой обсуждается ковариация и определение необходимой повторности в опытах.

Мы выражаем благодарность соавтору покойного сэра Рональда А. Фишера доктору Ф. Иейтсу и их издателям — Longman Group Ltd., London, за разрешение использовать таблицы: A.2, A.6, A.7 и A.8 из их книги «Statistical Tables for Biological, Agricultural and Medical Research» (6th edition, 1974).

Т. М. ЛИТТЛ,
Ф. ДЖ. ХИЛЛЗ

Сентябрь, 1977 г.

ЛОГИКА, ИССЛЕДОВАНИЕ И ЭКСПЕРИМЕНТ

«Цель статистической науки заключается в обеспечении объективной базы для анализа проблем, данные которых отступают от законов строгой обусловленности. Разработанная общая логическая система индуктивного мышления применима для данных такого рода и в настоящее время широко используется в научном исследовании. Поэтому некоторое понимание ее принципов важно как для исследователей, так и для тех, чьи интересы связаны с использованием технологических преимуществ, открываемых исследованием. Это особенно верно для сельскохозяйственной и биологической наук» *

Д. Финни

Вышеприведенная цитата кратко излагает важность статистической науки в сельском хозяйстве. Чтобы понять полностью смысл «логической системы индуктивного мышления», мы должны рассмотреть некоторые элементарные понятия логики. Классифицируя проблемы в зависимости от системы аргументации, используемой в их решении, мы находим, что имеется лишь два вида проблем.

ДЕДУКТИВНОЕ И ИНДУКТИВНОЕ МЫШЛЕНИЕ

Во-первых, имеется вид проблем, в которых нам дан некоторый общий принцип или ряд принципов, и нужно определить, что произойдет при определенных условиях. Используемый вид рассуждений от общего к частному и называется дедуктивным мышлением. Несколько примеров для его пояснения.

Дана общая формула площади круга: $A = \pi r^2$, какова площадь круга с радиусом в 15 см?

Даны определитель растений и описание сорняков Калифорнии. К какому виду относится тот или иной сорняк?

Дан закон Бойля; как изменится определенный объем газа при конкретном изменении давления и температуры?

Дан общий принцип борьбы с болезнью; как изменится урожайность конкретной культуры при внесении установленной дозы фунгицида?

Дана нормальная (без внешних дефектов монета), при подбрасывании которой вероятность появления решки равна $1/2$. Что получится при подбрасывании монеты 10 раз?

* Финни Д. Введение в теорию планирования экспериментов. Перевод с английского. М., «Наука», 1970, 280 с.— Прим. пер.

Почти все проблемы, рассматриваемые в процессе нашего обучения, были вышесказанного типа, для решения которых требовалось дедуктивный вид мышления. Часто говорят, что аграрники должны быть «хорошо» подкованы в отношении фундаментальных правил. Они должны иметь в своем распоряжении емкий багаж общих принципов и способность дедуктивно мыслить, чтобы применять это в конкретных случаях.

Второй вид проблем — противоположность первому. Мы имеем некоторые определенные данные и нужно вывести общие принципы, применимые ко всем членам класса, представляемого этими данными. Используемый вид рассуждений от конкретного к общему и называется индуктивным мышлением. Следующие примеры проблем с индуктивной аргументацией аналогичны выше приведенным, иллюстрировавшим дедуктивный тип проблем.

Даны площади и радиусы нескольких кругов; какой вид примет общая формула, выражающая связь между площадью и радиусом круга?

Дано несколько образцов сорняков, описания которых нет в определителе; как мы охарактеризуем виды и межвидовую связь их с другими сорняками, перечисленными в определителе растений?

Дан ряд результатов наблюдений за объемом газа при различных давлении и температуре; какие общие законы объясняют эти наблюдения?

Даны результаты серии испытаний по борьбе с болезнью; какие общие рекомендации можно сделать в отношении методов борьбы?

Даны результаты десятикратного подбрасывания монеты; какое заключение можно сделать относительно дефекта или отсутствия его у монеты?

Заметьте, что все проблемы этого вида сходны в одном — они начинаются с ряда наблюдений. В некоторых случаях, например при описании новых видов, результаты наблюдений просто регистрируются как явление, имеющееся в естественных условиях. Однако обычно результаты наблюдения получают в контролируемых условиях. Изучаемые факторы заставляют варьировать с соблюдением некоторой закономерности, путем наложения вариантов *. Все другие неизучаемые факторы, влияющие на показатели наблюдений, стараются свести до практически возможного минимума. Таким образом, мы имеем эксперимент.

Проблема исследователя. Как упоминалось выше, почти все проблемы, с которыми мы сталкиваемся при обучении, требуют дедуктивного мышления. Теперь мы можем дополнить, что почти все проблемы, с которыми сталкивается специалист сельского хозяйства, требуют индуктивного мышления.

Какая же типичная проблема стоит перед сельскохозяйственным исследователем? Она может быть представлена следующими общими выражениями: воздействует ли тот или иной прием на результаты

* Вариант опыта — изучаемое растение, сорт, удобрение, агротехнический прием, условие возделывания или их сочетание. — *Прим. ред.*

определенной сферы деятельности сельскохозяйственного производства, и если да, то до каких пределов? Поскольку этот вопрос не может быть разрешен с уверенностью на 100%, то мы должны принять во внимание риск и цену неправильно сделанного решения. Это положение будет понятнее в дальнейшем.

Чтобы ответить на такой вопрос, обычно требуется исследование. В простейшем опыте может быть только два варианта — новый прием и старый.

Более сложный эксперимент может включать несколько уровней дозировки или способов нового приема. Еще более сложными являются такие эксперименты, в которых изучается действие нескольких приемов одновременно.

Независимо от схемы опыта его цель — обеспечить условия проведения наблюдений (возможность отбора образцов), которые могут быть использованы для достоверных обобщений об изучаемом приеме. Получение данных для обобщений является типичным примером индуктивного мышления.

У читателя не должно сложиться мнение, что индуктивное мышление представляет независимый ход мыслей, далекий от дедуктивного мышления. Индуктивные обобщения следует постоянно контролировать точными дедуктивными методами.

Элемент случайности. Другая фраза из цитаты в начале главы требует некоторого разъяснения. Что значит «проблемы, данные которых отступают от законов строгой обусловленности?». Возвращаясь к ранее приведенным примерам, мы отмечаем некоторые важные различия среди них. В задаче по определению площади круга нет никакой неясности в отношении ответа. Для любого заданного радиуса может быть единственный ответ.

Задача с подбрасыванием монеты совершенно иного вида. Общее предположение строится на том, что монета недеформированная, однако даже при одном подбрасывании нам неясно, каким будет результат. Один из двух исходов может быть получен, причем оба одинаково возможны. Ответ на вопрос: «Что получится, если подбросить монету 10 раз? — еще более неопределенный, поскольку возможны 11 случаев выпадения решки, и эти результаты различаются по вероятности их наступления. Очевидно, здесь проявляется непостоянство усилия при каждом подбрасывании, поскольку причина и следствие связаны сложной функциональной связью.

Такая ситуация является почти всеобщей в сельском хозяйстве. Независимо от того, как много знают ученые о питании и физиологии, они не могут предсказать точно, каким будет прирост массы бычка или прибавка урожая картофеля на делянке при определенных условиях выращивания. Случайные отклонения, вызываемые множеством причин, являются источником дополнительной вариации результатов независимо от прилагаемых усилий при контроле всех известных факторов.

Понятие термина «случайность» трудно определить, но даже без четкого определения подразумеваемое значение достаточно хорошо понятно, чтобы оценить важность ее воздействия на биологические

результаты. Когда элемент случайности входит в проблему, создаются **реальные трудности**, которые усложняют в большей степени индуктивное, чем дедуктивное мышление.

Рассмотрим дедуктивную проблему, возникаемую с подбрасыванием монеты 10 раз. С помощью дедуктивных методов мы можем перечислить все 11 возможных исходов и довольно легко рассчитать вероятность каждого из них. Например, на вопрос: «Какова вероятность выпадения пяти решек и пяти орлов?», — ответ можно найти из формулы

$$\frac{10!}{5! (5!) (2^{10})}$$

что дает 0,246, или 24,6%. С увеличением количества подбрасываний или с изменением начального предположения и допущением о некоторой деформации монеты расчеты становятся более трудоемкими, хотя и остаются закономерными, а результаты простыми и конкретными. К счастью, мы имеем разработанную математиками теорию вероятностей, а также доступные практические методы и таблицы, которые позволяют в сложных случаях значительно сократить необходимые расчеты.

Теперь рассмотрим индуктивную проблему. Получив выпадение пяти решек и пяти орлов после десятикратного подбрасывания монеты, возникает вопрос: «Что можно сказать о наличии или отсутствии ее деформации?» Все что мы можем сказать с уверенностью, — это то, что монета не имела ни двух решек, ни двух орлов. Если бы она не была деформирована, мы ожидали бы получить результат, составляющий 25% от случаев произведенных подбрасываний. Мы можем сказать, причем с высокой степенью вероятности оказаться правыми, что монета не очень сильно деформирована ни со стороны решки, ни со стороны орла.

Необходимо помнить, что мы никогда не можем сделать такое утверждение с полной уверенностью. Даже в случае сильно деформированной монеты (такой, когда решка выпадает в 90% случаев) упомянутый результат выпадения пяти решек и пяти орлов возможен, хотя мало вероятен.

Единственным другим утверждением, которое мы можем сделать о монете, причем с довольно высокой уверенностью, так это то, что степень деформированности монеты занимала промежуточное положение между легкой деформированностью со стороны решки и легкой деформированности со стороны орла. Заметьте, что в этом интервале имеется бесконечное количество возможностей и что нулевая деформация является одной из них. Важно понять, что, не имея других данных о монете, кроме результатов 10 подбрасываний, мы не можем утверждать, что эта монета была недеформированной. Увеличивая число подбрасываний, мы можем сузить этот интервал влияния деформированности монеты, что делает возможным получение упомянутого результата, однако мы никогда не сможем утверждать с уверенностью, что монета не была деформирована.

Мы преднамеренно для упрощения оставили **без объяснения** выражение «сильная» и «слабая» (легкая) деформация. Однако с помощью статистических методов можно определить, какую **степень деформации** мы допускаем или отвергаем в зависимости от **уровня вероятности**, которую мы желаем иметь в наших выводах.

Таким образом, ответ на наш вопрос: «Что можно **сказать о деформированности монеты?**» — был довольно неопределенным. Читатель, привыкший к точным ответам дедуктивной математики, получит разочарование от этой неопределенности. Эта неопределенность является самой сутью индуктивного мышления, так что данный ответ **лучший** из всех возможных. В этой связи Альфред Норт Вайтхед, великий философ-математик, сказал: «Теория индукции — источник **огорчения** философов, однако вся наша деятельность основана на ней».

Исследователь не должен отчаиваться, пытаясь найти ответы на вопросы путем наблюдений и экспериментов. Однако необходимо твердо усвоить, что ответы не могут быть абсолютными, а обобщения нужно делать осторожно и только после аккуратно выполненных наблюдений и использования лучших систем аргументации под чьим-то руководством.

Необходимость статистической оценки. Большинство исследователей охотно признают необходимость статистического анализа в целях обеспечения объективной базы для оценки результатов, однако некоторые примеры могут оказаться полезными. Урожай зерна с двух одинаковых частей поля пшеницы независимо от того, будут ли они составлять ряды длиной в 1 род (5,03 м) или две половины поля, редко будет одинаковым. Масса фруктов с соседних деревьев сада оказывается редко одинаковой; показатели прироста массы тела двух любых животных одинакового вида и линии редко совпадают. Различия такого рода среди отдельных экземпляров культур и животных связаны с генетическими особенностями и условиями внешней среды, находящимися вне контроля исследователя. Подобные отклонения представляют вариабельность между экспериментальными единицами и называются экспериментальной ошибкой*.

Наличие такой вариабельности (изменчивости) затрудняет оценку нового приема на основе сравнения данных одной экспериментальной единицы опытного и такой же единицы контрольного варианта, не подвергавшейся воздействию приема. Действие нового приема смешивается с неучтенной вариабельностью. Таким образом, эксперимент с единственной повторностью дает приблизительную меру эффекта опытного варианта. Отсутствие двух экспериментальных единиц одного варианта не позволяет рассчитать его среднюю величину воздействия и экспериментальную ошибку. Математическая статистика, опирающаяся на выборочную совокупность экспериментальных единиц, преодолевает эти трудности, позволяя сделать несмещенную оценку действия вариантов, оценивает различия между ними с помощью крите-

* Экспериментальная ошибка не является ошибкой расчета, а характеризует меру отклонения измеряемой величины от истинной.— *Прим. ред.*

риев существенности, основанных на измерении экспериментальной ошибки.

Эффект вариантов оценивается на основе анализа их воздействия по крайней мере, на две (обычно больше) экспериментальные единицы и расчета средних результатов по каждому из них. Критерии существенности (значимости) действия вариантов дают долю различий по вариантам, обусловленных только случайным фактором.

Схемы опытов должны учитывать три важных принципа, обеспечивающих объективную основу применения статистики.

1. Повторность. Она означает, что одноименный вариант повторяется два или большее число раз. Ее назначение — дать оценку экспериментальной ошибки и обеспечить более точное определение действия вариантов. Повторность, необходимая для конкретного опыта, зависит от величины различий, которую вы желаете иметь, и от variability материала, с которым вы работаете. Учет этих двух моментов в начале эксперимента в значительной мере избавит от лишних усилий.

2. Рендомизация (случайность). Рендомизация — это такое наложение вариантов в эксперименте, когда все экспериментальные единицы имеют равный шанс получить тот или иной вариант воздействия. Она должна обеспечить несмещенную оценку средних величин и экспериментальной ошибки.

3. Локальный контроль *. Этот принцип схемы опытов позволяет снизить экспериментальную ошибку за счет частичного ограничения рендомизации. Например, в опыте, заложенном по схеме рендомизированных блоков, варианты группируют в блоки, ожидая, что они будут действовать неодинаково, и эффект блоков можно будет устранить из общей вариации в опыте.

ИССЛЕДОВАНИЕ, НАУЧНЫЙ МЕТОД И ЭКСПЕРИМЕНТ

Исследование, как общее понятие, представляет систематическое изыскание в предмете новых факторов или закономерностей. Приемы проведения исследования, как известно, воплощены в научном методе, который, хотя и не поддается точному определению, обычно включает многие этапы.

1. Формулирование гипотезы — предполагаемое объяснение или решение.

2. Планирование эксперимента с целью объективной проверки гипотезы.

3. Тщательное проведение опыта и сбор его результатов.

4. Интерпретация результатов эксперимента. Сопоставление их с уже известными фактами по данной проблеме приводит к подтверждению, отрицанию или изменению гипотезы.

Эксперимент является важным инструментом исследования. Некоторые основные характеристики хорошо спланированного опыта приводятся ниже.

* Эту роль в опыте выполняет организованное (рендомизированное) повторение (полный набор вариантов схемы опыта) или блок. — *Прим. пер.*

1. Простота. Выбор вариантов и схемы их размещения должны быть просты, насколько возможно, и соответствовать цели эксперимента.

2. Точность. Уровень вероятности должен обеспечить существенность различий по вариантам со степенью точности, необходимой экспериментатору. Это предполагает соответствующую схему и достаточную повторность.

3. Отсутствие систематической ошибки. При планировании эксперимента необходимо устранить систематические различия между отдельными экспериментальными единицами, что позволит в последующем сделать несмещенную оценку действия каждого варианта.

4. Достоверность заключений. Заключение должны иметь настолько широкий интервал надежности, насколько возможно. Эксперимент, повторенный во времени и на территории, увеличивает достоверность заключений, вытекающих из него. Другой путь увеличения достоверности опыта — факториальность схемы. В факториальном опыте действие одного фактора оценивается на фоне различных уровней воздействия другого фактора.

5. Расчет допустимой степени риска. Любой эксперимент всегда оставляет некоторое сомнение в отношении достоверности сделанных заключений. Схема эксперимента должна обеспечить возможность проверки действия и случайного фактора на получаемые результаты.

ЭТАПЫ ЭКСПЕРИМЕНТА

Выбор методики исследования зависит в значительной степени от предмета и цели исследований. Исследование может быть описательным и представлять выборочное обследование или оно может быть контролируемым экспериментом или серией экспериментов. Если мы имеем дело с экспериментом, необходимо продумать целый ряд моментов с целью его успешного проведения. Некоторые из наиболее важных этапов перечислены ниже.

1. Определение проблемы. Первый шаг к решению проблемы — это ее ясное и четкое изложение. Если проблема не может быть сформулирована, то мало шансов для ее решения вообще. Уяснив проблему, следует поставить задачи, ответы на которые приведут к ее решению.

2. Формулирование цели. Оно может быть в форме вопросов, подлежащих ответу, гипотез, подлежащих проверке, или эффектов, подлежащих оценке. Задачи должны быть ясно сформулированы. Это позволит более эффективно спланировать методику исследований. В случае нескольких задач их следует расположить в порядке важности, так как без этого трудно выбрать правильную схему эксперимента. При постановке задач не должно быть неуверенности или, наоборот, чрезмерной самоуверенности.

3. Выбор вариантов. Успех эксперимента основывается на тщательном выборе вариантов, оценка действия которых дает ответ на поставленные задачи исследования.

4. Выбор экспериментального материала. При выборе экспериментального материала следует иметь в виду поставленные задачи

и объекты наблюдений, по которым будут сделаны выводы. Экспериментальный материал должен достаточно полно представлять всю совокупность изучаемых объектов.

6. Выбор схемы эксперимента. Наряду с учетом задач опыта схема должна отличаться простотой и обеспечивать заданный уровень точности.

6. Определение экспериментальной единицы и повторности. В полевом опыте с растениями задача сводится к определению размера и формы делянки; в опытах с животными — числа животных, которые и составят экспериментальную единицу. Особенно ценную помощь в решении указанной задачи имеют результаты других подобных экспериментов. Как размеры делянки, так и повторность должны обеспечить заданную точность оценки избранных вариантов.

7. Контроль краевого воздействия соседних делянок друг на друга. Обычно это достигается путем введения защитных полос и рендомизированного размещения вариантов.

8. Определение показателей, подлежащих учету. Эти показатели должны надлежащим образом характеризовать действие вариантов в соответствии с поставленными задачами эксперимента. Дополнительно нужно учитывать также другие данные, которые помогут объяснить, почему эти варианты действуют именно так, а не другим путем.

9. Выбор метода статистического анализа и табулирование результатов. Выпишите источники варьирования и соответствующие степени свободы для дисперсионного анализа. Внесите в таблицу дисперсионного анализа все F -критерии *, которые необходимо использовать для установления наличия существенных различий по вариантам опыта. Представьте, как можно использовать полученные в опыте результаты, и подготовьте возможные итоговые таблицы или графики, которые будут характеризовать ожидаемые результаты. Сравните эти ожидаемые результаты с поставленными задачами эксперимента, чтобы убедиться, что этот эксперимент дает те ответы, которые ожидают получить.

На этом этапе необходимо ознакомить статистику и одного или нескольких сотрудников с намеченными планами, в выполнении которых они будут принимать участие. Подобная консультация может восполнить то, что вами было не учтено. Некоторые изменения или дополнения значительно обогатят эксперимент и позволят получить больше информации из исследования, планируемого к проведению.

10. Проведение эксперимента. Проводя исследования, используйте методику, лишенную субъективизма. Используйте схему эксперимента при сборе образцов так, чтобы индивидуальные различия или разница в зависимости от очередности их получения не отражались на экспериментальной ошибке. Избегайте трудностей при получении и обработке первичных данных. Немедленно проверяйте данные,

* Критерии Фишера (критерии существенности) — теоретический и фактический; первый находят по таблице (приложение к книге) для заданного числа степеней свободы изучаемой и случайной дисперсий, а второй определяют как частное от деления большей дисперсии на меньшую. — *Прим. пер.*

резко отличающиеся от остальных. Организуйте сбор и обработку данных так, чтобы облегчить их анализ и избежать ошибки при записи. При дублировании данных точно же сверяйте копию с оригиналом.

11. Анализ данных и интерпретация результатов. Анализ всех данных производится в соответствии с планом, а объяснение результатов — с учетом условий эксперимента. Проверяются гипотезы и устанавливается связь полученных результатов с уже известными фактами. Помните, что статистические методы ничего не подтверждают и всегда возможно, что ваши выводы могут быть ошибочными. Поэтому допускайте возможность принятия неправильного решения. Не делайте поспешных выводов даже при их статистической достоверности, если результаты не совпадают с ранее установленными фактами. В подобном случае продолжайте исследование отмеченного явления.

12. Подготовка полного, правильного, легко читаемого научного отчета. Необходимо помнить, что отрицательного результата не бывает. Если нулевая гипотеза не отвергается, то это является положительным свидетельством отсутствия реальных различий между изучаемыми вариантами опыта. В этом случае также необходимо ознакомить с полученными вами результатами других сотрудников и допустить возможность пересмотра выводов.

Хотя большинство из упомянутых выше этапов исследований являются нестатистическими, но статистический анализ — важная часть исследования. Математическая статистика помогает исследователю правильно выбрать схему эксперимента и объективно оценить получаемый цифровой материал. Лишь немногие ученые имеют время и желание стать компетентными биометриками, однако все могут познать и применять в исследованиях значения трех R .

1. Имейте повторность (Replicate). Это единственный путь, позволяющий определить достоверность выводов, вытекающих из эксперимента.

2. Используйте рендомизацию (Randomize). Статистический анализ предполагает размещение изучаемых вариантов по делянкам совершенно объективным, случайным образом.

3. Консультируйтесь (Request help). Обращайтесь за консультацией при сомнениях в выборе оптимальной схемы опыта, правильности его проведения или анализа полученных данных. От вас не требуется быть экспертом в статистике, но вы должны иметь достаточно знаний, чтобы понять главные принципы научного экспериментирования, быть способным преодолеть обычные заблуждения и обратиться за помощью, когда в этом есть необходимость.

ЗАКЛЮЧЕНИЕ

Процесс мышления, проделывающий путь от общей закономерности до конкретного вывода, является дедуктивным. Индуктивное мышление приводит к общему принципу от конкретного вывода. Исследования проводятся с целью нагроуления конкретных фактов, на основе которых устанавливаются общие закономерности или принципы, и, та-

ИЗБ. № 44589

Библиотека и института

ботаники 9/21

ким образом, представляют индуктивный тип мышления или аргументации.

Изменчивость — характерная особенность биологического материала, что затрудняет решение вопроса: являются ли различия между экспериментальными единицами результатом неучтенной варибельности или действительным результатом применяемых вариантов? Математическая статистика, основанная на выборочном методе, позволяет преодолеть эти трудности и дает несмещенную оценку действия вариантов и значимости различий между ними с помощью критериев существенности, основанных на измерении неучтенной вариации.

Повторность, рендомизация и локальный контроль (организованное повторение) являются тремя основными принципами схемы эксперимента.

Научный метод представляет процесс непрерывного познания от известных факторов к гипотезе и далее к эксперименту, что позволяет получать новые факты, которые будут отвергать, усиливать или изменять первоначальную (нулевую) гипотезу.

Хорошо продуманный и надлежащим образом спланированный эксперимент должен быть прост, насколько возможно, иметь высокую вероятность решения поставленных задач и не допускать систематических и грубых ошибок *. Заключения должны иметь широкие пределы достоверности, а получаемые данные — пригодными для соответствующей статистической обработки.

Этапы исследования включают: определение проблемы и формулирование задач, выбор вариантов, экспериментального материала, схемы эксперимента, экспериментальной единицы и повторности, контроль взаимного воздействия соседних (краевых) единиц, сбор данных, анализ и интерпретацию результатов, а также подготовку научного отчета.

* Классификация ошибок дана в книге Б. А. Доспехова «Методика полевого опыта». М., «Колос», 1979, 415 с.— *Прим. пер.*

НЕКОТОРЫЕ ОСНОВНЫЕ ПОНЯТИЯ

Экспериментальная единица означает единицу экспериментального материала, на которую распространяется действие варианта. Это может быть единственный лист растения, целое растение, некоторая площадь земли с растениями, сосуд или чашка в вегетационном домике, одно животное, несколько животных или целое стадо. Выражение «делянка» является синонимом экспериментальной единицы и часто используется в качестве таковой в полевых опытах с растениями. Понятие «делянка» иногда неправильно распространяют на весь эксперимент, который на самом деле включает несколько делянок. Любое свойство (характеристика) экспериментальной единицы, поддающееся измерению, называется *переменной*. Переменная может быть дискретной (прерывной) величиной, принимающей только определенные значения, например количество больных растений на делянке, или она может быть непрерывной величиной и принимать любые значения в определенном интервале, например урожай зерна ячменя с делянки. Отдельное наблюдение переменной называют *вариантой*.

Вариант опыта — это доза материала или метод, которые подлежат проверке в эксперименте. Сорт культуры является одним из видов вариантов. Когда воздействию одного варианта подвергается больше чем одна экспериментальная единица, то мы имеем *повторность* этого варианта. Две экспериментальные единицы, подвергающиеся одинаковому воздействию, дают две *повторности* (или *повторения*). Несколько экспериментальных единиц, подвергающихся воздействию различных вариантов, имеющих *повторность* и размещающихся по определенной схеме, составляют *эксперимент* (опыт, испытание).

В статистическом смысле выражение «совокупность» означает серию измерений или учетов одной переменной, сделанных на всех единицах, относящихся к данной популяции. Совокупность может быть относительно малой, как, например, урожайность зерна ячменя с одного акра (0,4 га) на всех полях определенного региона в конкретном году; или она может быть большой, как, например, рост всех мужчин США в возрасте свыше 20 лет или урожай всех возможных делянок определенной формы, которые могут быть размещены на территории эксперимента. Даже в малой совокупности производят измерения очень большого числа индивидуумов или экспериментальных единиц.

Мы можем иметь совокупность значений переменной индивидуальных экспериментальных единиц, совокупность выборочных средних переменной или совокупность разностей между парами выборочных средних.

Выборка — это серия измерений, составляющая часть совокупности. Мы получаем информацию и делаем заключения о совокупности на основе выборки. В этой связи важно, чтобы выборка была репрезентативной или соответствовала совокупности.

Для получения репрезентативной выборки мы используем принцип случайности. Случайная выборка — это такая выборка, когда любые индивидуальные наблюдения имеют равновеликую возможность быть в ее составе.

Совокупности списывают с помощью характеристик, называемых параметрами. Параметры — это строго определенные величины. Например, арифметическая средняя всех наблюдений совокупности является параметром. Она имеет лишь одно значение, хотя мы редко знаем, что это за значение. Выборочные совокупности описывают теми же характеристиками, однако в данном случае их называют статистическими. Средняя выборки является статистической характеристикой. Мы рассчитываем статистические характеристики выборки, чтобы оценить параметры совокупности. Каждая выборка имеет конкретные значения статистических характеристик.

Различные значения переменной имеют различную частоту их появления в совокупности. Для удобства описания (характеристики) совокупности данные большой выборки принято группировать на

Таблица 2.1. Группировка данных анализа 10 тыс. луковиц лука на рефрактометре

Классовый интервал	Срединное значение	Разноска дат по классам (накопление частот)	Частота
6,8 - 7,2	7,0	III	10
7,3 - 7,7	7,5	III III III III	19
7,8 - 8,2	8,0		60
...			
10,8 - 11,2	11,0		1600
11,3 - 11,7	11,5		1700
...			
14,3 - 14,7	14,5		65
14,8 - 15,2	15,0		50
15,3 - 15,7	15,5	III III III III III	25
15,8 - 16,2	16,0	III III III III	20
16,3 - 16,7	16,5	III III II	12

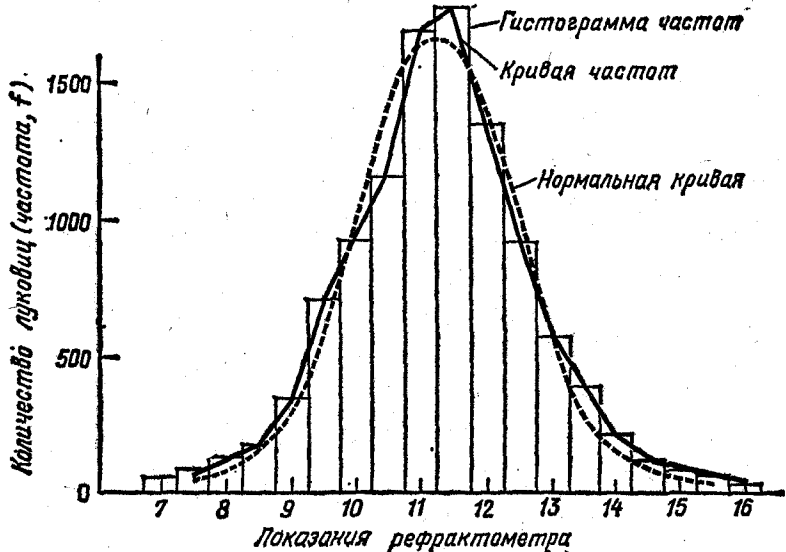


Рис. 2.1. Распределение частот 10 000 луковиц лука по данным отсчетов на рефрактометре и теоретическая кривая нормального распределения.

классы и представлять в виде таблицы, гистограммы или кривой распределения частот. В таблице 2.1 значения переменной разносят по нескольким классам. Их общее количество для каждого класса может быть выражено на графике частотой встречаемости и на этой основе построена гистограмма распределения частот (рис. 2.1). Соединив срединные значения каждого класса, мы получим кривую распределения частот.

Если бы мы нанесли на графике частоты распределения: урожая зерна ячменя с нескольких делянок, процента жирности молока, полученного от ряда коров, прироста массы большого числа групп телят, числа точек поражения паршой клубня картофеля из 1000 обследованных клубней или данных анализа большого количества луковиц лука на рефрактометре, то получили бы результаты, сходные по ряду важных показателей. Все кривые имели бы приблизительно форму колокола с вершиной вблизи середины, которая представляет самый типичный класс. Кривые довольно симметрично спадают в сторону двух исключительно редко встречающихся классов, расположенных по краям.

Большинство биологических данных (фактически и объекты других областей применения статистики), представленное в виде кривой распределения частот, в значительной мере соответствует математически рассчитанной кривой, называемой кривой нормального распределения частот. На рисунке 2.1 кривая нормального распределения нанесена точками на гистограмму и кривую распределения луковиц по данным анализа на рефрактометре. Обратите внимание, насколько велико соответствие двух кривых.

...а, описывающая кривую нормального распределения, имеет следующий вид:

$$f = \frac{N}{(\sigma \sqrt{2\pi})} e^{-(y-\mu)^2/2\sigma^2},$$

где f — частота встречаемости любого данного значения переменной, y — любое данное значение переменной, N — число объектов совокупности (объем совокупности), μ — средняя совокупности и σ — стандартное отклонение совокупности. Заметьте, что нормальную кривую, описывающую частоту появления дат различной величины, можно нанести на графике с помощью всего лишь двух параметров μ и σ .

Кривые нормального распределения различаются между собой лишь их средними значениями и (или) стандартными отклонениями. Средняя определяет положение кривой на горизонтальной оси. Стандартное отклонение определяет разброс или дисперсию дат. Рисунок 2.2а показывает две кривые нормального распределения с одинаковым значением стандартного отклонения, но различным значением средних. Две кривые нормального распределения на рисунке 2.2б имеют одинаковые средние, но различные стандартные отклонения.

СТАТИСТИЧЕСКИЕ СИСТЕМЫ ОФОРМЛЕНИЯ РАСЧЕТОВ, СРЕДНИЕ И СТАНДАРТНЫЕ ОТКЛОНЕНИЯ

Для математического исчисления вариантов, средних и сумм необходимо иметь систему для выражения процессов и закономерностей. В этой книге сложные расчеты во многих случаях не приведены, поскольку они вызывают затруднение у большинства студентов. Тем не менее, если вы продолжаете изучение математической статистики с помощью других книг, вам может быть полезным краткое знакомство с более или менее стандартной системой расчетов. Мы применили выражение

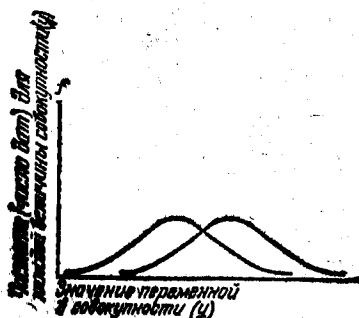


Рис. 2.2а. Кривые нормального распределения с одинаковыми стандартными отклонениями.

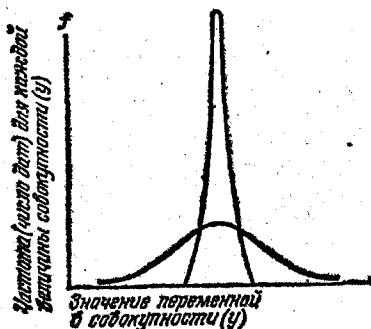


Рис. 2.2б. Кривые нормального распределения с одинаковыми средними и разными стандартными отклонениями.

«более или менее стандартной системой расчетов» потому, что в специальной литературе имеются значительные расхождения по этому вопросу, к великому огорчению всех студентов. Во-первых, мы разберем расчеты средней и стандартного отклонения и таким образом познакомимся с некоторыми простыми системами исчисления.

Общепринятой и, как правило, лучшей мерой центральной тенденции является арифметическая средняя. Символом, обозначающим арифметическую среднюю (далее будем называть просто среднюю), является греческая буква μ для всей совокупности и $\bar{Y}$ или $\bar{X}$ для выборки. μ — это параметр, строго определенная величина, значение которой мы редко знаем, и $\bar{Y}$ — статистическая характеристика, своя для каждой выборки, взятой из одной совокупности.

Средняя совокупности рассчитывается по следующей формуле:

$$\mu = \frac{Y_1 + Y_2 + Y_3 + \dots + Y_N}{N},$$

где Y_1 , Y_2 и т. д. — варианты совокупности и N — число вариантов в совокупности.

Таким образом, Y_N есть N -ая дата совокупности.

Во многих книгах для обозначения переменной чаще используют X , чем Y . Однако это вызывает некоторое заблуждение, когда дело доходит до регрессии. В анализе регрессии значения переменной обозначают буквой Y , поскольку их наносят на ось Y (ось ординат) графика; значения X , откладываемые на оси X (оси абсцисс), являются вариантами вашего эксперимента, например дозы удобрений. Таким образом, обозначив значения переменной через Y , мы предупредим возможные заблуждения в последующем.

Формула средней (μ) может быть записана в общем виде.

$$\mu = \frac{\sum_{i=1}^N Y_i}{N}.$$

Греческая буква Σ в этой формуле указывает на необходимость суммирования всех значений Y_i . Индекс при знаке суммирования $i=1 \dots N$ означает, что значения Y_i начинаются от Y_1 и заканчиваются Y_N .

Поскольку мы практически не знаем величину μ , то оцениваем ее с помощью средней выборки, которая определяется по следующей формуле:

$$\bar{Y} = \frac{\sum_{i=1}^r Y_i}{r},$$

где r представляет количество дат в выборке. Если уже известно число дат, то запись числителя формулы сокращается до вида ΣY_i , или даже ΣY .

Таблица 2.2. Сухая масса пяти растений, $\bar{Y}=3$

Y , г/растение	$Y - \bar{Y}$	$(Y - \bar{Y})^2$
3	0	0
4	1	1
5	2	4
2	-1	1
1	-2	4
Σ 15	0	10

Для выборки, представленной в таблице 2.2, $\bar{Y} = \sum Y_i / r = (3 + 4 + 5 + 2 + 1) / 5 = 15 / 3 = 3$ г/растение. Часто мы желаем установить отклонение отдельных вариантов (Y) от их средней ($\bar{Y}$). Такие отклонения часто обозначают курсивом малых букв y или x . Таким образом, $y = Y - \bar{Y}$, а $x = X - \bar{X}$.

Выделяют два важных свойства средней. Сумма отклонений отдельных вариантов от нее равна нулю (колонка 2 в табл. 2.2), а сумма квадратов отклонений (колонка 3 в табл. 2.2) является минимальной, т. е. сумма квадратов отклонений от любого другого значения переменной будет больше.

В других приемах (которые вы не встретите в нашей книге) используют медиану — величину, занимающую срединное положение и делящую ранжированный ряд значений переменной на две одинаковые половины (если количество дат равно четному числу, то медиана будет составлять среднюю двух центральных значений), и моду — наиболее часто встречающуюся величину переменной. В нормальном распределении средняя, медиана и мода принимают одну и ту же величину.

Общепринятой мерой разброса, лучшим образом отвечающей большинству целей статистики, является стандартное отклонение и ее квадрат, называемый дисперсией (дисперсией). Стандартное отклонение и дисперсию для совокупности обозначают как σ и σ^2 соответственно, а для выборки — s и s^2 соответственно.

Дисперсия совокупности определяется по формуле

$$\sigma^2 = \frac{\sum (Y_i - \mu)^2}{N},$$

где N — число дат в совокупности (ее объем).

Для расчета оценки дисперсий (σ^2) малых выборок (где $r < 60$) лучше использовать следующую формулу:

$$s^2 = \frac{\sum (Y_i - \bar{Y})^2}{r - 1},$$

где r — объем выборки. Почему в качестве делителя чаще используют $r - 1$, чем r ? Если мы знаем значение величины μ , то тогда для расчета дисперсии выборки лучше взять формулу $s^2 = \frac{\sum (Y_i - \mu)^2}{r}$, где r — объем

выборки. Однако мы практически никогда не знаем значения величины μ , так что в числителе мы заменяем ее на $\bar{Y}$. В среднем $\bar{Y}$ равняется μ , хотя она варьирует от выборки к выборке и редко совпадает точно со значением μ . Ранее мы установили, что $\sum (Y_i - \bar{Y})^2$ меньше суммы квадратов отклонений от любого другого значения, чем от $\bar{Y}$. Поэтому если $\bar{Y}$ не совпадает точно со значением μ , то $\sum (Y_i - \bar{Y})^2$ меньше чем $\sum (Y_i - \mu)^2$. Это означает, что формула $\sum (Y_i - \bar{Y})^2 / r$ дает заниженное значение дисперсии (σ^2). Заменяя r в знаменателе на $r-1$, мы получаем более правильную оценку дисперсии. Другими словами, мы можем записать, что

$$\frac{\sum (Y_i - \bar{Y})^2}{r-1} \cong \frac{\sum (Y_i - \mu)^2}{r} \cong \sigma^2.$$

Числитель формулы $\sum (Y_i - \bar{Y})^2$ — сумма квадратов, а для данного случая — сумма квадратов отклонений индивидуальных вариантов от их средней. Знаменатель формулы $r-1$ степени свободы для выборки, число которых обычно на единицу меньше числа наблюдений.

Мы используем малую выборку, представленную в таблице 2.2, для демонстрации расчета величин s^2 и s .

$$\begin{aligned} s^2 &= \frac{\sum (Y_i - \bar{Y})^2}{r-1} = \frac{(3-3)^2 + (4-3)^2 + (5-3)^2 + (2-3)^2 + (1-3)^2}{(5-1)} = \\ &= \frac{(0)^2 + (1)^2 + (2)^2 + (-1)^2 + (-2)^2}{4} = \frac{0+1+4+1+4}{4} = \frac{10}{4} = 2,5; \\ s &= \sqrt{2,5} = 1,58 \text{ г/растение.} \end{aligned}$$

Для малых выборок, выражаемых целыми числами, где средняя иногда бывает также целым числом, значения s^2 и s можно легко рассчитать по приведенной формуле. Для больших выборок их расчет значительно проще осуществлять по ускоренному методу, особенно при наличии счетных машин. При желании вы можете убедиться в справедливости приводимого ниже равенства:

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{r}.$$

Таким образом, удобной рабочей формулой для расчета s^2 является

$$s^2 = \frac{\sum Y_i^2 - \frac{(\sum Y_i)^2}{r}}{r-1}.$$

Правая часть числителя называется корректирующим выражением, или корректирующим фактором, что будем в последующем обозначать буквой C . $C = (\sum Y_i)^2 / r$. Знаменатель $(r-1)$ называют числом степеней свободы (будем обозначать через df), на которых основана дисперсия; в данном случае их число на единицу меньше объема выборки.

Подставляя данные таблицы 2.2 в эту формулу, мы получим

$$s^2 = \frac{3^2 + 4^2 + 5^2 + 2^2 + 1^2 - \frac{(3+4+5+2+1)^2}{5}}{5-1} = \frac{55 - \frac{15^2}{5}}{4} =$$

$$= \frac{55-45}{4} = \frac{10}{4} = 2,5, \text{ т. е. как и в первом случае.}$$

Многие настольные и карманные калькуляторы приспособлены для расчета стандартного отклонения путем нажатия соответствующей клавиши после ввода данных выборки. Калькулятор с таким приспособлением значительно облегчает расчеты при дисперсионном анализе. Обратите внимание, что является делителем в вашем калькуляторе: r или $r-1$. Делитель r следует использовать лишь для большой выборки, по крайней мере, если она содержит 60 наблюдений. Другими характеристиками разброса (дисперсии) являются размах варьирования и среднее отклонение. Однако не будем касаться их в данной книге, ввиду значительно большего практического значения величин s^2 и s .

Вариабельность экспериментальных единиц в опыте при различной их размерности (единицах измерения) и (или) различных размерах делянок можно сравнивать с помощью коэффициентов вариации. Коэффициенты вариации определяют путем деления стандартного отклонения экспериментальных единиц на общую среднюю экспериментала и выражают в процентах. Таким образом, CV (коэффициент вариации) равен $\frac{s}{\bar{Y}} 100$.

Например, рассмотрим данные двух экспериментов: первый касался урожайности сахарной свеклы, где $s=3$ т/га и средняя урожайность со всех делянок равнялась 76,2 т/га; в другом изучали густоту всходов лимской фасоли на делянке при $s=5,8$ и $\bar{Y}=82,7$. Коэффициенты вариации составили соответственно $\frac{3}{76,2} 100=3,9\%$ и $\frac{5,8}{82,7} 100=7,0\%$. Сравнение этих данных показывает, что вариация густоты всходов

Таблица 2.3. Пример таблицы с двумя входами

Ряды (i , варианты)	Столбцы (j , повторения) ^a			Суммы $Y_{i.}$	Средние $\bar{Y}_{i.}$
	1	2	3... r		
1	Y_{11}	Y_{12}	 Y_{1r}	$Y_{1.}$	$\bar{Y}_{1.}$
2	Y_{21}	Y_{22}	 Y_{2r}	$Y_{2.}$	$\bar{Y}_{2.}$
.	.	.	.	.	.
.	.	.	.	.	.
n	Y_{n1}	Y_{n2}	 Y_{nr}	$Y_{n.}$	$\bar{Y}_{n.}$
Суммы $Y_{.j}$	$Y_{.1}$	$Y_{.2}$	 $Y_{.r}$	$Y_{..}$	
Средние $\bar{Y}_{.j}$	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	 $\bar{Y}_{.r}$		$\bar{Y}_{..}$

Таблица 2.4. Урожайность корнеплодов сахарной свеклы (т/га) в опыте с пятью вариантами в четырех повторениях

Варианты (ряды)	Повторения (столбцы)				Суммы $Y_{i.}$	Средние $\bar{Y}_{i.}$
	1	2	3	4= r		
1	15	18	17	18	68	17,0
2	16	15	13	16	60	15,0
3	23	25	22	24	94	23,5
4	20	16	14	16	66	16,5
5= n	20	17	15	16	68	17,0
Суммы $Y_{.j}$	94	91	81	90	$356 = Y_{..}$	
Средние $\bar{Y}_{.j}$	18,8	18,2	16,2	18,0	$17,8 = \bar{Y}_{..}$	

лимской фасоли в 1,8 раза (7,0 : 3,9) превышает вариацию урожайности сахарной свеклы.

Размещение значений переменной в таблицах с двумя входами. В связи с требованиями схемы эксперимента или в целях облегчения расчетов получаемые в опытах данные обозначают соответствующими символами и представляют в виде таблицы с двумя входами (табл. 2.3). Символом любого значения переменной в таких таблицах является Y_{ij} , а в других книгах — X_{ij} . Индекс i обозначает строки таблицы от единицы до n , а j — колонки (или столбцы) от единицы до r .

Частное значение переменной указывается на пересечении соответствующего ряда и столбца, например Y_{23} является датой 2-го ряда и 3-го столбца.

Обратите внимание, что точка при индексе принята в качестве обозначения определенной операции, производимой со всеми датами ряда или столбца. $Y_{1.}$ обозначает сумму всех дат 1-го ряда. Для обозначения операции, производимой с суммами всех рядов, мы используем символ $Y_{.j}$; например, выражение $\sum Y_{.j}^2$ означает, что вы должны возвести в квадрат сумму каждого ряда и сложить получаемые квадраты сумм. Средней 1-го ряда является величина $Y_{1.}$, которая равняется $Y_{1.}/r$, или в общем виде $\bar{Y}_{1.} = Y_{1.}/r$. Последняя формула дополнительно под-
сказывает: сложите отдельные значения j переменной 1-го ряда и сумму разделите на r — число дат в ряду.

Использование такой системы исчисления в виде таблицы (если вы, наконец, освоились с ней) дает возможность опустить дополнительную запись математических операций и закономерностей. Мы будем использовать ее умеренно и почти всегда сопровождать конкретными цифрами в качестве иллюстрации.

Для приобретения некоторого навыка давайте сопоставим символы таблицы 2.3 и реальные даты таблицы 2.4 и сделаем необходимые расчеты.

Вычисление суммы квадратов отклонений отдельных дат от средней для таблицы с двумя входами будем производить по формуле

$$SS^* = \sum_{i=1}^n \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2.$$

Индексы при знаках суммирования часто опускают, более того, иногда для сокращения выражения оставляют один знак суммирования

$$SS = \sum (Y_{ij} - \bar{Y}_{..})^2.$$

Первая запись формулы является более полной, поскольку она указывает полный ряд слагаемых как для рядов, так и столбцов, однако, усвоив ее суть, в дальнейшем можно ограничиться второй записью формулы.

Для расчета SS по данной формуле возьмем данные таблицы 2.4:

$$SS = (15 - 17,8)^2 + (18 - 17,8)^2 + \dots + (15 - 17,8)^2 + (16 - 17,8)^2 = 223,2.$$

Точки в таблице означают, что нужно продолжить операцию со всеми датами таблицы и закончить на двух последних значениях переменной 15 и 16.

Расчет сумм квадратов отклонений средних от общей средней по столбцам мы производим по формуле $SSC = n \sum (\bar{Y}_{.j} - \bar{Y}_{..})^2$. Это означает, что мы берем каждую среднюю столбца ($\bar{Y}_{.j}$), вычитаем из нее общую среднюю $\bar{Y}_{..}$, возводим каждую разницу в квадрат $[()]^2$, суммируем квадраты ($\sum$) и умножаем сумму на число дат (n) в столбце. Оставим пока без объяснения необходимость указанного умножения суммы на n и продолжим эти, не совсем понятные пока, короткие расчеты по таблице 2.4:

$$SSC = 5[(18,8 - 17,8)^2 + (18,2 - 17,8)^2 + (16,2 - 17,8)^2 + (18,0 - 17,8)^2] = 5(3,76) = 18,8.$$

SSC можно рассчитать, используя не средние, а суммы по столбцам $SSC = (\sum Y_{.j}^2 / n) - (Y^2_{..} / nr)$. Для данных упомянутой таблицы получаем

$$SSC = \frac{94^2 + 91^2 + 81^2 + 90^2}{5} - \frac{356^2}{5(4)} = 6355,8 - 6336,8 = 18,8.$$

Теперь мы вернемся к нормальному распределению и другим взаимосвязям, составляющим основу математической статистики.

ВЫБОРКА ИЗ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ (СОВОКУПНОСТИ)

Мы обычно подвергаем воздействию конкретного варианта несколько делянок или животных. Действие варианта определяется на основе расчета средней выборки. Мы знаем, что расширение эксперимента

* Символ SS обозначает первые буквы двух английских слов — сумма квадратов, ранее встречались символы аналогичного происхождения: C — корректирующий фактор, r — повторность и CV — коэффициент вариации: подобные обозначения являются типичными для данной книги. — *Прим. пер.*

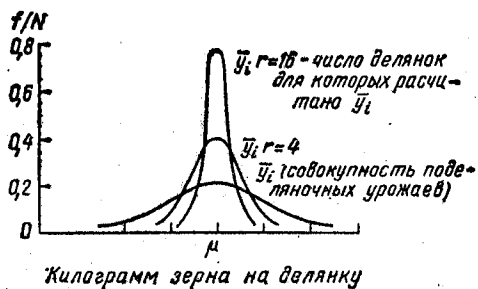
(распространение его на другие выборки) дает ряд различных средних. В этой связи возникает вопрос: как верно отражает единственная средняя действительный эффект варианта? Одним из подходов к решению этой проблемы является расчет доверительного интервала, т. е. размаха значений переменной, внутри которого будет находиться истинная средняя эффекта варианта, до тех пор, пока мы не возьмем необычную (резко отличающуюся от ранее сделанных) выборку. Прежде чем рассчитывать доверительный интервал для средней варианта, обратите внимание на сходство определенных параметров совокупностей объектов и совокупностей средних, обусловленное их отбором из одной родительской совокупности.

Распределение выборочных средних. Если все возможные выборки данного размера составлены из нормальной совокупности объектов, то средние таких выборок будут давать значительно большую совокупность, чем исходная совокупность. Средняя новой совокупности останется равной средней исходной совокупности, а стандартное отклонение уменьшится. При таком виде составления выборки все объекты исходной совокупности принимают за идентичные. После составления выборки и определения ее среднего значения они возвращаются в исходную совокупность, и берется другая выборка. Этот процесс повторяется до тех пор, пока все возможные сочетания объектов не будут представлены в выборках.

Стандартное отклонение для совокупности выборочных средних называют стандартной ошибкой средней или просто стандартной ошибкой и обозначают как $\sigma_{\bar{y}}$. Если $\sigma_{\bar{y}}$ определена для выборки, то ее обозначают как $S_{\bar{y}}$.

Между дисперсией исходной совокупности и дисперсией совокупности выборочных средних имеется важная и очень нужная математическая связь, которая выражается следующим равенством: $\sigma_{\bar{y}}^2 = \sigma^2/r$, где r — объем выборки, на котором основана совокупность средних. Рисунок 2.3 иллюстрирует эту связь. С увеличением объема выборок (r) кривая распределения средних становится уже и выше, т. е. стандартное отклонение уменьшается, а общая средняя остается прежней. Имея такое равенство ($\sigma_{\bar{y}}^2 = \sigma^2/r$), мы можем определить $\sigma_{\bar{y}}^2$ на основе одной выборки по формуле $s_{\bar{y}}^2 = s^2/r$. Мы используем эту математическую закономерность при расчете доверительного интер-

Рис. 2.3. Распределение частот совокупности выборочных средних, различающихся объемом выборок, составленных из одной нормальной совокупности поделяночных урожаев зерна. Кривые распределения (все нормальные) становятся уже и выше по мере увеличения объема выборок в соответствии с зависимостью $\sigma_{\bar{y}}^2 = \sigma^2/r$.



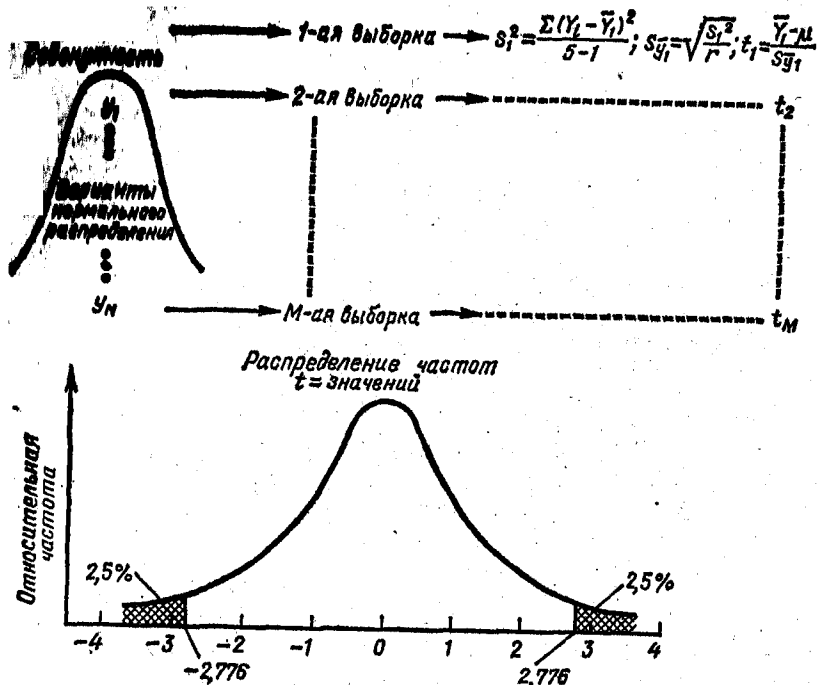


Рис. 2.4. Построение кривой t -распределения для выборок с объемом, равным 5. Значение t рассчитывается для каждой из всех возможных выборок с пятью датами. Нанесение частот одинаковых значений t на графике дает кривую распределения, которая покрывает меньше значений вблизи центра и больше по краям по сравнению с кривой нормального распределения.

вала выборочных средних. Эта формула неоднократно используется в ходе дисперсионного анализа (ANOVA), когда нам надо определить поделеночную дисперсию (s^2), составить ряд средних, если мы принимаем, что средняя рассчитана для выборок, взятых из одной совокупности. В этом случае мы рассчитываем s_y^2 на основе выборочных средних по формуле $s_y^2 = \Sigma (\bar{Y}_i - \bar{Y}_{..})^2 / (n-1)$ и далее определяем s^2 . Используя равенство $s_y^2 = s^2/r$, откуда $s^2 = r s_y^2$. Мы еще вернемся к этому позднее.

t -распределение и доверительный интервал. Рассмотрим другой ряд выборок заданного размера, например $r=5$ (рис. 2.4). Для каждой выборки вычислим $\bar{Y}$, s , s_y и другую статистическую характеристику t из формулы $t = (\bar{Y} - \mu) / (s / \sqrt{r})$. Теперь представим большую совокупность величин t в виде кривой распределения частот. Кривая будет иметь форму, показанную на рисунке 2.4.

Для каждого объема выборок характерно типичное t -распределение. Для объема, равного 5, только 2,5% всех значений будет равно или больше 2,776 и 2,5% равно или меньше 2,776. Таблица А.2 в приложении является двухсторонней (симметричной) таблицей, где указаны

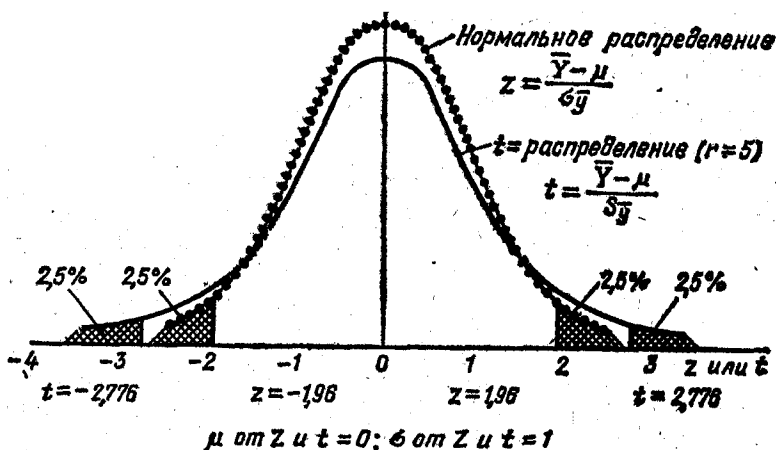


Рис. 2.5. Распределение z в сравнении с t -распределением для выборок объемом, равным 5. По мере увеличения объема выборки t -распределение приближается к нормальному z -распределению. (Заштрихованы значения t и z , составляющие краевые 5% от всей площади под кривыми.)

вероятности получения $\pm t$ -величин для различного числа степеней свободы в зависимости от объема выборки. Например, при $df=10$ находим, что вероятность появления $\pm t=3,169$ равна 0,01 (1%).

На рисунке 2.5 t -распределение для объема выборки, равного 5, сравнивается с нормальным распределением. Отметим, что t -распределение подвержено большей вариабельности, чем нормальное распределение.

Чем больше объем выборки, тем ближе приближается t -распределение к нормальному распределению. Если значения t рассчитаны на основе выборок, содержащих 60 или большее количество вариантов, их распределение практически принимается за нормальное, поскольку оно оценивает статистическую характеристику Z , подчиняющуюся нормальному распределению по соответствующей формуле $Z = \frac{\bar{Y} - \mu}{\sigma_y / \sqrt{n}}$; формулы для расчета t и Z различаются лишь знаменателем.

Величина s_y малых выборок сильно варьирует и поэтому t подвержено большей вариации, чем Z , поскольку величина σ_y остается константной. Для больших выборок значения s_y менее вариабельны и поэтому величина t более тесно соответствует величине Z . В последнем ряду большинства таблиц t -распределения при числе степеней свободы, приближающемся к бесконечности, $t=Z$ (табл. А.2). Таблица распределения площади под кривой нормальной величины не включена в эту книгу, поскольку мы редко имеем дело с выборками, объем которых достаточно велик для ее использования.

Доверительный интервал. На основе любой случайной выборки можно рассчитать доверительный интервал или доверительные пределы, внутри которых будет находиться μ для обусловленного уровня доверия. Из формулы $\pm t = (\bar{Y} - \mu) s_y$ находят μ и ее

два крайних значения, которые называют доверительными пределами $CL = \bar{Y} \pm t s_{\bar{y}}$. Если мы хотим быть на 95% уверены в том, что данный интервал содержит μ , то необходимо величину $s_{\bar{y}}$ умножить на табличное значение t при $n-1$ число степеней свободы и 5%-ном уровне вероятности (табл. А.2 — приложение). Для выборки объемом (r), равным 5, величину $s_{\bar{y}}$ умножают на 2,776.

В качестве примера рассмотрим выборку таблицы 2.2, где $r=5$, $\bar{Y}=3$ и $s^2=2,5$, откуда получаем

$$s_{\bar{y}} = \sqrt{\frac{2,5}{5}} = 0,707 \text{ и } CL_{.95} = 3 \pm 2,776(0,707) = 4,96 \div 1,04 \text{ г/растение.}$$

Таким образом, с уверенностью на 95% мы можем утверждать что μ находится в этом интервале. Неправильно утверждать о 95% вероятности нахождения μ в этом интервале, поскольку вычисленная на основе статистических характеристик конкретной выборки μ может быть как в пределах рассчитанного интервала, так и за его пределами*. Мы можем взять выборку, $\bar{Y}$ которой и (или) s^2 настолько отклоняются от μ и (или) s^2 , что μ не будет входить в $CL_{.95}$. Однако такая выборка возможна в 5 случаях из 100.

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ СУЩЕСТВЕННОСТИ

Статистическая процедура сравнения средних двух или большего числа вариантов сводится к проверке так называемой нулевой гипотезы, суть которой составляет предположение об отсутствии действия вариантов. Далее мы переходим к проверке вероятности того, что различия между средними наших выборок являются случайными, если объекты для выборок были действительно случайно отобраны из совокупностей, имеющих одинаковые значения средних и дисперсий. Если наш анализ приводит к заключению, что такие различия, обусловленные случайностью, можно ожидать довольно часто, то мы не отвергаем нулевую гипотезу и делаем вывод об отсутствии весомых доказательств фактического эффекта вариантов. Если анализ покажет, что наблюдаемые различия будут редко проявляться в случайных выборках, взятых из совокупностей с одинаковыми значениями средних и дисперсий, то мы отвергаем нулевую гипотезу и заключаем, что выявлено реальное действие по крайней мере одного варианта. В данном случае принято утверждать, что одна из средних будет существенно отличаться от других.

Если вероятность того, что наблюдаемая вариация между средними, обусловленная действием случайного фактора, составляет 5% или меньше, то мы утверждаем о наличии существенных различий

* Выражения: доверие, уверенность и надежность в данной книге отображают условную вероятность, поскольку вероятность — математически обусловленная категория. — Прим. пер.

между средними. Если же вероятность упомянутой вариации составляет 1% или меньше, то такие различия принято называть очень существенными.

Тот факт, что нулевая гипотеза не отвергнута и мы делаем вывод об отсутствии существенных различий между средними, еще не доказывает невозможность проявления действия некоторых вариантов. Всегда допустима определенная вероятность того, что реальный эффект имел место, но данный эксперимент имел слишком грубые допуски (недостаточно тонко поставлен), чтобы выявить различия по вариантам на желаемом уровне вероятности.

На данном этапе вы должны осознать, что 5%-ный уровень существенности не таит в себе ничего магического. Выводы, сделанные вами относительно результатов эксперимента, являются вашими. Они не могут исходить от статистика и должны опираться на нечто большее, чем статистическая достоверность. Идея ваших заключений должна быть рассмотрена в свете уже известного исследования о предмете. Не спешите принять существенный результат, если он теряет смысл при сравнении с другими известными фактами. Всегда допустима вероятность того, что ваш существенный результат был обусловлен случайностью и вы сделали ошибку, отвергнув нулевую гипотезу.

Подумайте о значении ошибочно сделанного заключения. В случае серьезных последствий, как, например, при ошибочных рекомендациях об изменении процесса, технологии всего производства и т. д., что потребует значительных вложений и даст относительно малую прибыль, вы можете сомневаться, надо ли отвергать нулевую гипотезу на основе одного эксперимента (одной проверки), даже если результаты и существенны на 5%-ном уровне. В данной ситуации дополнительная проверка становится совершенно необходимой.

С другой стороны, если последствия ошибочно сделанного заключения не будут серьезными, вы можете отвергнуть нулевую гипотезу, даже если статистический анализ показал, что такой результат обусловлен случайным фактором в одном случае из 15 или даже в одном из 10 испытаний. Рассмотрим, к примеру, проверку нового недорогого приема предпосевной обработки семян, если комплексное исследование в нескольких полевых опытах дало результат, близкий к существенному на 5%-ном уровне. Кроме того, предположим, что результаты нескольких вегетационных опытов показали значительно большую эффективность нового приема по защите проросших семян от основных болезней, поражающих данную культуру. В такой ситуации, отвергая нулевую гипотезу, вы можете оказаться правыми вплоть до того, чтобы рекомендовать этот прием в производство, одновременно проверяя сделанные рекомендации в дополнительных полевых опытах.

F-распределение. F-критерий определяется как отношение двух дисперсий и используется для выявления возможности допустить, что две величины дисперсий, установленные независимо одна от другой, являются оценками одной дисперсии. Как F-критерий это отношение было названо Дж. У. Снедекором в честь Р. А. Фишера — пионера в области использования математической статистики в сельскохозяйственном опытно-деле. В дисперсионном анализе F-критерий исполь-

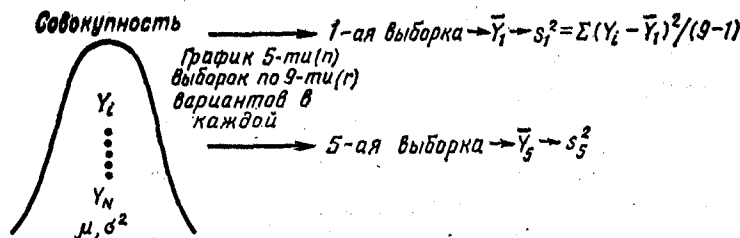


Рис. 2.6а. Чтобы рассчитать F -критерий для 5 выборок по 9 вариантам в каждой после 1-го тура их составления, нужно:

а) оценить σ^2 по $s_y^2 = \Sigma (\bar{Y}_i - \bar{Y}_{..})^2 / (5-1)$, где $\bar{Y}_{..} = \Sigma Y_i / 5$;

б) оценить σ^2 на основе вариальности выборочных средних по $s_b^2 = r s_y^2 = 9 s_y^2$;

в) оценить σ^2 на основе средней внутри выборочной вариальности по $s_w^2 = (s_1^2 + \dots + s_5^2) / 5$. Итак, $F = \frac{s_b^2}{s_w^2}$ при $df = \frac{5-1}{5(9-1)} = \frac{4}{40}$. Прodelав большое число указанных

туров составления выборок, вы получите совокупность F -значений, которые на графике представят кривую, изображенную на рисунке 2.6б.

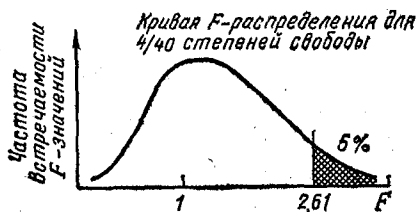


Рис. 2.6б. Построение кривой F -распределения на основе 5 (n) выборок по 9 (r) вариантам в каждой, составляемой из совокупности значений переменной ($Y_1 \dots Y_N$), подчиняющихся нормальному распределению. Пять процентов F -значений составляют 2,61, или больше, чем 2,61 (см. текст).

зуется для проверки отсутствия различий между средними, т. е. для ответа на вопрос: «Можно ли допустить, что средние по вариантам получены из выборочных совокупностей, имеющих одинаковые средние?». Это можно продемонстрировать на примере расчета F -критерия (фактического и табличного).

Воспользуемся данными рисунка 2.6а: из совокупности, подчиняющейся нормальному распределению, взято пять выборок ($n=5$), каждая содержит конкретное количество дат, к примеру девять ($r=9$). Вычислите средние этих пяти выборок. Оцените значение σ^2 на основе расчета дисперсии s^2 для каждой выборки от s_1^2 до s_5^2 . Сложите эти оценки σ^2 и получите среднюю (единую) оценку $s^2 = (s_1^2 + \dots + s_5^2) / 5$.

Теперь оцените дисперсию средних (σ_y^2) на основе средних пяти выборок $s_y^2 = (\bar{Y}_i - \bar{Y}_{..})^2 / (5-1)$. На основе s_y^2 снова оцените σ^2 , используя равенство $s^2 = r s_y^2$, где r — число дат в каждой выборке. Используйте отношение дисперсий F следующим образом:

$$F = \frac{s^2, \text{ дисперсия выборочных средних (дисперсия между выборками)}}{s^2, \text{ дисперсия выборочных совокупностей (дисперсия внутри выборок)}}$$

* Расчет дисперсий проводят по формуле $s_1^2 (Y_1 - \bar{Y}_1)^2 + \dots + (Y_5 - \bar{Y}_5)^2$, аналогично s_2^2 и т. д., где $\bar{Y}$ — среднее значение каждой выборки. — Прим. пер.

Число степеней свободы для числителя будет равно $n-1=4$ (где n — число выборок) и для знаменателя $n(r-1)=5(8)=40$ (где r — число дат в каждой выборке). Теперь представьте, что процедура взятия выборок продолжается до тех пор, пока все возможные сочетания объектов в выборках и необходимые расчеты статистических характеристик не будут осуществлены и записаны; кроме того, будут установлены частоты получения F -критериев разной величины и кривая распределения частот будет нанесена на графике. F -критерий, равный 2,61, — это предельная величина, вне которой будет находиться лишь 5% рассчитанных значений F . Эта величина для 5%-ного уровня взята из таблицы F -критерия для 4 и 40 степеней свободы (табл. А.3, приложение). Подобным образом значения F (табличные или теоретические) могут быть установлены и для других уровней вероятности.

Поскольку обе дисперсии в F отношении являются оценками одной дисперсии σ^2 , то их величины будут оставаться близкими к 1 до тех пор, пока не будет взята необычная (смещенная) выборка. Распределение F -критерия для указанных размеров выборок ($n=5$, $r=9$) представлено на рисунке 2.6б. Площадь под кривой представляет частоту любого данного значения F . Для любой серии выборок при $n=5$ и $r=9$ вероятность получения F -критерия, равного или большего, чем 2,61, равна 5%. Или, другими словами, вероятность получения значения F , меньшего, чем 2,61, равна 95%. Отметим, что F -критерий является односторонним, т. е. нас не интересуют его значения меньше 1.

Вышеприведенные гипотетические эксперименты с выборками предназначались для того, чтобы показать, каким образом, составляя выборки из объектов совокупностей, подчиняющихся нормальному распределению, можно получить t - и F -распределения. Таблицы теоретических значений t и F подготовлены не таким трудоемким путем составления всех возможных выборок, а на основе точных и довольно сложных математических закономерностей. Использование F -критерия в ANOVA (дисперсионном анализе) будет рассмотрено в следующей и дальнейших главах.

ЗАКЛЮЧЕНИЕ

Экспериментальная единица (делянка в полевых опытах). Единица экспериментального материала, на которую распространяется действие варианта.

Переменная. Экспериментальную единицу, обладающую свойством изменяться на определенную величину, которую можно измерить, называют *переменной*.

Варианта. Конкретное значение переменной.

Совокупность. Серия измерений (или учетов) переменной, выполненная на всех индивидуумах, составляющих данную популяцию.

Выборка. Серия измерений вариант, выполненная на части популяции.

Параметр. Характеристика совокупности (например, средняя). Параметр — это устойчивая величина, которую мы редко знаем. Параметры оценивают на основе выборки. Параметры обычно обозначают греческими буквами (μ , σ и т. д.).

Статистическая характеристика. Это характеристика выборки, часто используется для оценки параметра; символика латинская ($\bar{Y}$, s и т. д.).

Нормальное распределение. Математически определенная закономерность, кривая, имеющая форму колокола; получают путем нанесения на графике частот появления конкретных значений переменной для всего размаха этих значений. Для описания нормального распределения достаточно иметь его среднее значение и стандартное отклонение.

Средняя совокупности индивидуальных вариантов, μ .

$$\mu = \frac{\sum Y_i}{N}, \text{ где } N — \text{число индивидуумов в совокупности.}$$

Оценка значения μ на основе выборки, $\bar{Y}$.

$$\bar{Y} = \frac{\sum Y_i}{r}, \text{ где } r — \text{число индивидуумов в выборке.}$$

Дисперсия совокупности индивидуальных вариантов, σ^2 .

$$\sigma^2 = \frac{\sum (Y_i - \mu)^2}{N}.$$

Стандартное отклонение совокупности индивидуальных вариантов, σ .

$$\sigma = \sqrt{\sigma^2}.$$

Оценка значения σ^2 на основе выборки, s^2 .

$$s^2 = \frac{\sum (Y_i - \bar{Y})^2}{r-1} \text{ (общий вид); } s^2 = \frac{\sum Y_i^2 - \frac{(\sum Y_i)^2}{r}}{r-1} \text{ (рабочая формула).}$$

Корректирующий фактор, используется в рабочей формуле, C .

$$C = \frac{(\sum Y_i)^2}{r}.$$

Оценка значения σ на основе выборки, s .

$$s = \sqrt{s^2}.$$

Коэффициент вариации, CV .

$$CV = \frac{s}{\bar{Y}} 100.$$

Совокупность средних. Совокупность всех возможных выборочных средних ($\bar{Y}$) для определенного объема выборок (r), взятых из совокупности индивидуумов.

Средняя совокупности выборочных средних, $\mu_{\bar{Y}}$.

$$\mu_{\bar{Y}} = \frac{\sum \bar{Y}_i}{M}, \text{ где } M — \text{число выборочных средних.}$$

Дисперсия совокупности средних, $\sigma_{\bar{Y}^2}$.

$$\sigma_{\bar{Y}}^2 = \frac{\sum (\bar{Y}_i - \mu)^2}{M}.$$

Стандартное отклонение совокупности средних или стандартная ошибка, $\sigma_{\bar{y}}$.

$$\sigma_{\bar{y}} = \sqrt{\sigma_y^2}.$$

Зависимость между σ^2 и σ_y^2 .

$$\sigma_y^2 = \frac{\sigma^2}{r},$$

где r — число дат в каждой выборке средних (объем выборки).

Оценка значения σ_y^2 на основе n -выборок, s_y^2 .

$$s_y^2 = \frac{\sum (\bar{Y}_i - \bar{Y}_{..})^2}{n-1}.$$

Оценка значения σ_y^2 на основе одной выборки, объемом в r .

$$s_y^2 = \frac{s^2}{r} = \frac{\sum (Y_i - \bar{Y})^2}{r-1} \left(\frac{1}{r} \right).$$

Оценка значения σ^2 при известном s_y^2 .

$$\sigma^2 \cong r s_y^2,$$

где r — число дат в каждой выборке, t — статистическая характеристика выборки, выражающая различие между выборочной средней и средней совокупности в долях стандартной ошибки выборки.

$$t = (\bar{Y} - \mu) / s_{\bar{y}}.$$

Доверительный интервал для μ , малая выборка.

$$CL = \bar{Y} \pm t s_{\bar{y}}.$$

F -отношение двух оценок значения σ^2 .

$$F = \frac{s^2, \text{ дисперсия выборочных средних}}{s^2, \text{ дисперсия выборочных совокупностей}}.$$

ДИСПЕРСИОННЫЙ АНАЛИЗ И *t*-КРИТЕРИЙ

Итак, мы имеем представление об общих статистических понятиях, необходимых для усвоения дисперсионного анализа. Однако прежде, чем перейти к сложным экспериментам, необходимо рассмотреть применение этих понятий при анализе данных простейшего опыта из двух вариантов, каждый из которых случайно размещен на 5 из 10 экспериментальных единиц. Сначала мы объясним суть дисперсионного анализа, а затем сделаем обычную процедуру выполнения расчетов.

ANOVA (ДИСПЕРСИОННЫЙ АНАЛИЗ) ДВУХ ВЫБОРОК

Для иллюстрации ANOVA используем данные таблицы 3.1.

Для того чтобы определить вариабельность, называемую экспериментальной ошибкой, мы вычислим дисперсии двух выборок (s_1^2 и s_2^2); предположим, что они оценивают общую дисперсию (σ)² и затем оценим эту общую дисперсию с помощью средней выборочных дисперсий.

$$s_1^2 = \frac{\sum (Y_{1j} - \bar{Y}_1)^2}{r-1} = \frac{(19-17)^2 + \dots + (20-17)^2}{5-1} = \frac{26}{4} = 6,5;$$

$$s_2^2 = \frac{\sum (Y_{2j} - \bar{Y}_2)^2}{r-1} = \frac{(23-20)^2 + \dots + (18-20)^2}{5-1} = \frac{16}{4} = 4,0.$$

Взяв среднюю значений s_1^2 и s_2^2 , получим оценку σ^2 , основанную на вариабельности внутри выборок, которую обозначим, как s_w^2 .

$$s_w^2 = \frac{s_1^2 + s_2^2}{2} = \frac{6,5 + 4,0}{2} = 5,25.$$

Приняв в качестве нулевой гипотезы предположение о том, что эти две выборки являются случайными выборками из одной совокупности,

Таблица 3.1. Поделяночная урожайность (ц/га) двух случайно размещенных сортов пшеницы

Сорта	Повторность	$Y_{i.}$	$\bar{Y}_{i.}$	
1	19 14 15 17 20	85	17	$\bar{Y}_1.$
2	23 19 19 21 18	100	20	$\bar{Y}_2.$
		185	18,5	$= \bar{Y}_{..}$

а потому обе выборочные средние будут ($\bar{Y}_1$ и $\bar{Y}_2$) оценками одного значения средней всей совокупности (μ), мы оцениваем дисперсию средних (σ_y^2) на основе средних 1-й и 2-й выборки

$$s_y^2 = \frac{\sum (\bar{Y}_i - \bar{Y}..)^2}{n-1} = \frac{(17-18,5)^2 + (20-18,5)^2}{2-1} = \frac{(-1,5)^2 + (1,5)^2}{1} = 4,5.$$

Снова оцениваем σ^2 , используя равенство $s_y^2 = s^2/r$, откуда получаем оценку s^2 . Помните, что r — это число дат, на которые опирается выборочная средняя. Мы обозначим эту оценку σ^2 , как s_b^2 .

$$s_b^2 = r s_y^2 = 5 (4,5) = 22,5.$$

Итак, мы имеем две оценки σ^2 : s_w^2 , основанную на средней вариабельности внутри выборок, и s_b^2 , основанную на вариабельности между выборками. Приняв нулевую гипотезу, мы ожидаем получить близкие значения s_w^2 и s_b^2 , поскольку они обе являются оценками одной дисперсии (σ^2). Мы можем установить вероятность получения различных оценок σ^2 путем расчета фактического значения F -критерия и сравнения его с величиной табличного F . Для расчета фактического F мы всегда берем дисперсию выборочных средних (вариантов) в числитель, а дисперсию отдельных вариантов (объектов) в знаменатель. Таким образом, получаем $F = s_b^2/s_w^2$.

Если два варианта (две выборки) взяты из совокупностей, имеющих различные средние, то s_b^2 будет характеризовать дополнительную вариацию, обуславливающую это различие, и будет иметь большее значение, чем s_w^2 . Для нашего опыта $F = 22,5/5,25 = 4,29$.

Числитель s_b^2 основывается на 1-й степени свободы, поскольку имеется лишь две выборочные средние, а знаменатель s_w^2 — на сумме степеней свободы для каждой выборки. Каждая выборка имеет 5 дат и поэтому 4 df , а число степеней свободы для s_w^2 будет $4+4=8$.

Из таблицы F -критериев (табл. А.3 — приложение) берем те значения F , которые мы будем ожидать для заданного уровня вероятности, если нулевая гипотеза принята и различия между выборочными средними обусловлены только случайным фактором. Для 1-й (числитель) и 8-й (знаменатель) степеней свободы мы будем ожидать значение F , равное 4,29 или большее с вероятностью около 7%. Выразаясь по-другому, если фактическая разность между средними равна нулю ($\mu_1 - \mu_2 = \mu_d = 0$), то шанс получения оценки разности средних ($\mu_d = 3$ ц/га) составляет 7%. Обычно мы не рискуем утверждать, что это событие (вероятность появления которого равна 7%) не произойдет; поэтому будет неразумным отвержение нулевой гипотезы и заключение о том, что средняя 1-го сорта действительно отличается от средней 2-го сорта. С другой стороны, различие между сортами в 3 ц/га может представлять на практике значительный экономический эффект. Поэтому возможным выводом будет наше решение о продолжении оценки двух сортов в дополнительных опытах.

Этапы дисперсионного анализа. Ниже приводятся этапы дисперсионного анализа (ANOVA) данных таблицы 3.1 с использованием настольного или карманного калькулятора.

Таблица 3.2. Сводная таблица дисперсионного анализа (исходные данные из таблицы 3.1)

Виды варьирования *	Степени свободы, df	Суммы квадратов, SS	Средний квадрат, MS	$F_{\text{факт}}$	$F_{\text{табл}}$	
					10%	5%
Общее	9	64,5				
Сортов	1	22,5	22,5	4,29	3,46	5,32
Ошибки	8	42,5	5,25			

* В отечественной литературе имеются разные переводы этого названия, оставим данное, а в тексте частично используем «источники вариации». — *Прим. пер.*

1. Подготовить сводную таблицу дисперсионного анализа с источниками (видами) варьирования и соответствующими степенями свободы (табл. 3.2).

Это общее число степеней свободы расчленяется на компоненты в соответствии со схемой опыта. В опыте два варианта, поэтому $df=2-1=1$. Число степеней свободы для ошибки можно всегда определить путем вычитания: $9-1=8$, кроме того, в данном случае путем сложения степеней свободы для двух выборок. В каждой выборке 5 сортов, откуда df для одной выборки будет равно $5-1=4$, а df ошибки $4+4=8$.

2. Вычислить сумму квадратов для сортов (SSV) и средний квадрат * для сортов (MSV).

$$SSV = \frac{\sum Y_i^2}{r} - \frac{Y_{..}^2}{nr} = \frac{85^2 + 100^2}{5} - \frac{185^2}{2(5)} = 3445,0 - 3422,5 = 22,5;$$

$$MSV = \frac{SSV}{df(V)} = \frac{22,5}{1} = 22,5.$$

Обратите внимание, что при расчете SSV мы используем суммы, а не средние. На больших счетных машинах за основу берут отклонения от средних, а для настольных калькуляторов значительно легче и с меньшей ошибкой будет использование сумм, поскольку не нужно брать отклонения и округлять цифры при расчете средних. Следующие алгебраические расчеты покажут, почему можно использовать суммы вместо средних при вычислении сумм квадратов отклонений.

Предположив, что наши два сорта не различаются и оба являются выборками из одной совокупности, мы узнали, что вторая оценка σ^2 получается из уравнения $s_b^2 = rs_y^2$, где s_y^2 — дисперсия средних двух сортов и r — число дат, лежащих в основе каждой средней. Средним квадратом вариантов таблицы 3.2 (в данном случае сортов) является s_b^2 , т. е. $MSV = rs_y^2$. Учитывая, что $s_y^2 = \sum (\bar{Y}_i - \bar{Y}_{..})^2 / (n-1)$, получим

$$MSV = r \left[\frac{\sum (\bar{Y}_i - \bar{Y}_{..})^2}{n-1} \right].$$

Поскольку $MSV = SSV / (n-1)$, то $SSV = r [\sum (\bar{Y}_i - \bar{Y}_{..})^2]$.

* Следует понимать «квадраты отклонений», но в данной книге слово «отклонения» опущено. — *Прим. пер.*

По второй главе мы знаем формулу $\sum (Y_{it} - Y_{..})^2 = \sum Y_{it}^2 - (\sum Y_{it})^2/n$, так что теперь можем записать, что

$$SSV = r \left[\sum \bar{Y}_{it}^2 - \frac{(\sum \bar{Y}_{it})^2}{n} \right].$$

Теперь заменим средние на суммы, используя равенства $Y_{it} = \bar{Y}_{it}/r$ и $\sum \bar{Y}_{it} = Y_{..}/r$, таким образом

$$SSV = r \left[\frac{\sum Y_{it}^2}{r^2} - \frac{Y_{..}^2}{r^2} \left(\frac{1}{n} \right) \right].$$

Преобразуем выражение, умножив r на его значение, получим формулу $SSV = (\sum Y_{it}^2/r) - (Y_{..}^2/rn)$, которая уже была дана раньше. Познакомьтесь с некоторыми требованиями этой формулы, прежде чем вычислять суммы квадратов на основе исходных сумм.

а) Первое выражение — $\sum Y_{it}^2$ указывает: нужно сложить квадраты сумм (в данном случае суммы по сортам) и разделить на r , то есть на число слагаемых, вошедших в каждую сумму числителя. Наиболее часто студенты ошибаются в выборе делителя и делят на число квадратов сумм, а не на число дат, вошедших в каждую сумму.

б) Второе выражение — $Y_{..}^2/nr = (\sum Y_{it})^2/nr$ известно под названием корректирующего члена или корректирующего фактора. Это квадрат суммы всех дат, составляющих суммы первого выражения, поделенный на число дат, вошедших в эту сумму ($Y_{..}$).

в) Если варианты имеют неодинаковое число повторений, сумму по каждому варианту возводят в квадрат и делят на число слагаемых этой суммы. Таким образом,

$$SST = \sum \frac{Y_{it}^2}{r_i} - \frac{Y_{..}^2}{\sum r_i} = \left(\frac{Y_{1.}^2}{r_1} + \dots + \frac{Y_{n.}^2}{r_n} \right) - \left(\frac{Y_{..}^2}{r_1 + \dots + r_n} \right).$$

Например, если дата $Y_{25} = 18$ (2-й ряд и 5-я колонка таблицы 3.1) выпала, то сумма для 2-го сорта будет $(100 - 18) = 82$ и $Y_{..} = 85 + 82 = 167$. Тогда

$$SSV = \frac{85^2}{5} + \frac{82^2}{4} - \frac{167^2}{5+4} = 3126 - 3099 = 27.$$

Теперь перейдем к третьему этапу дисперсионного анализа.

3. Вычислить общую сумму квадратов (SS). Этот этап выполняется перед расчетом суммы квадратов для ошибки. Оставив SS в памяти калькулятора, сумму квадратов для ошибки находят путем вычитания:

$$SS = \sum Y_{it}^2 - \frac{Y_{..}^2}{nr} = (19^2 + 14^2 + \dots + 18^2) - \frac{185^2}{2(5)} = 3487,0 - 3422,5 = 64,5.$$

4. Вычислить сумму квадратов и средний квадрат для ошибки (SSE и MSE):

$$SSE = SS - SSV = 64,5 - 22,5 = 42,0$$

$$MSE = \frac{SSE}{dfE} = \frac{42,0}{8} = 5,25.$$

5. Вычислить F -критерий для сортов:

$$F = \frac{MST}{MSE} = \frac{22,5}{5,25} = 4,29.$$

6. Возьмите из таблицы А.3 (приложение) теоретическое значение F для уровня значимости, который вы желаете иметь в опыте. Соответствующие степени свободы для числителя отношения двух дисперсий расположены в «шапке» таблицы, а для знаменателя — с левой стороны таблицы.

Клавиша стандартного отклонения. Настольные и карманные калькуляторы с программой для расчета стандартного отклонения или дисперсии упрощают дисперсионный анализ, причем в случае одинаковой повторности вариантов отпадает необходимость в использовании корректирующего фактора. Перед началом анализа убедитесь, что калькулятор вычисляет s и s^2 путем деления суммы квадратов на число дат, которые вы вводите, за минусом единицы, т. е. $s = \sqrt{(Y_i - \bar{Y})^2 / (r - 1)}$. Используйте для контроля следующий ряд значений переменной: 19, 14, 15, 17, 20. Поочередно вводите каждую дату путем нажатия клавиши, часто обозначенной ΣX . Закончив на последней дате (20), нажмите клавишу стандартного отклонения, обычно обозначенную σ . Если калькулятор делит на $n - 1$, ответ будет равен 2,5495. Если делителем является n , ответ будет 2,2803.

С использованием клавиши « σ » вычисления MST и SST по данным сумм таблицы 3.1 проводят следующим образом.

Вводите 85 и затем 100.

Нажмите клавишу « σ » (ответ = 10,6066+), возведите в квадрат (ответ = 112,5). Разделите на число дат, которые вы ввели для получения каждой суммы, т. е. на 5, ответ = 22,5 = MST .

Умножьте MST на число степеней свободы для вариантов, т. е. на 1. Ответ = 22,5 = SST .

Общая SS вычисляется путем ввода каждой даты (19, 14... 21, 18), нажатия « σ » (ответ = 2677+), возведения в квадрат « σ » (ответ = 7,166+) и умножения полученного квадрата на общее число степеней свободы (9), ответ = 64,5.

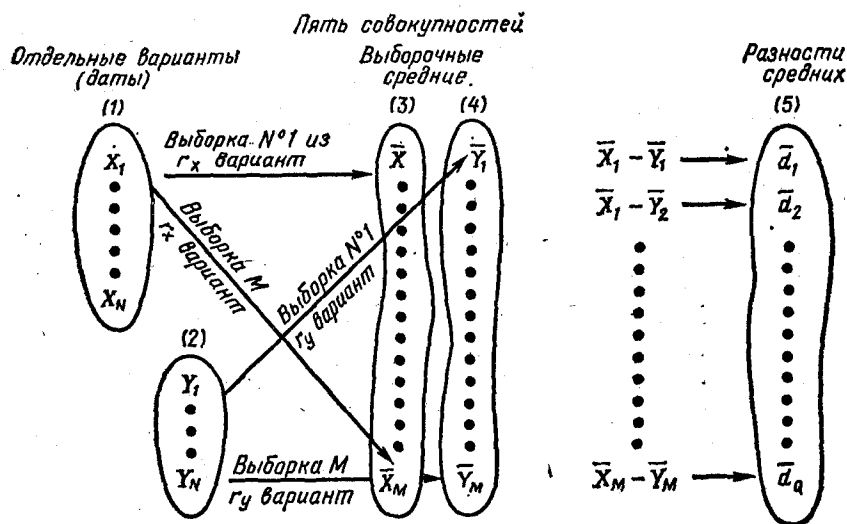
Проделав несколько примеров, вы усвоите эти простые требования, после чего легко и быстро будете выполнять дисперсионный анализ с помощью настольного или карманного калькулятора. Помните, после ввода суммы по вариантам или других сумм (или индивидуальных дат для расчета общей суммы квадратов) надо нажать клавишу стандартного отклонения, возвести σ в квадрат для получения σ^2 , разделить на число дат, вошедших в каждую сумму (для расчета общей суммы квадратов делить на 1), после чего получают средний квадрат для того вида варьирования, который вы рассчитываете.

СОВОКУПНОСТЬ РАЗНОСТЕЙ МЕЖДУ СРЕДНИМИ

В дополнение к F -критерию для проверки факта о наличии существенных различий между двумя средними мы можем использовать t -критерий. Во-первых, рассмотрим, каким образом составляется совокуп-

ность разностей между средними из нормальной совокупности объектов; особенно нам необходимо узнать связь параметров этой новой совокупности с параметрами исходной совокупности и совокупности средних, также искусственно составленной для получения совокупности разностей средних.

Если из двух нормальных совокупностей $X_1, X_2, \dots, X_N$ и $Y_1, Y_2, \dots, Y_N$ мы возьмем все возможные выборки заданного объема и рассчитаем их средние, то получим две дополнительные совокупности $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_M$ и $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_M$. Теперь если мы возьмем все возможные пары средних и найдем разность внутри них следующим образом: $\bar{X}_1 - \bar{Y}_1, \bar{X}_1 - \bar{Y}_2, \dots, \bar{X}_1 - \bar{Y}_M, \bar{X}_2 - \bar{Y}_1, \dots, \bar{X}_2 - \bar{Y}_M, \dots, \bar{X}_M - \bar{Y}_M$, то получим пятую совокупность, совокупность разностей средних (рис. 3.1). Количество разностей средних (Q) этой совокупности будет значительно больше объема совокупностей $\bar{X}_i$ и $\bar{Y}_i$. Если число средних в этих обеих совокупностях равно M , тогда $Q = M^2$.



		<i>Средние</i>		
(1)	(2)	(3)	(4)	(5)
μ_x	μ_y	$\mu_{\bar{x}}$	$\mu_{\bar{y}}$	$\mu_{\bar{d}}$
$\mu_x - \mu_y$		$\mu_{\bar{x}} - \mu_{\bar{y}}$		$\mu_{\bar{d}}$
		$(\mu_x = \mu_y \quad \mu_{\bar{d}} = 0)$		

Дисперсии

$$\sigma_x^2 \quad \sigma_y^2 \quad \sigma_{\bar{x}}^2 \quad \sigma_{\bar{y}}^2 \quad \sigma_{\bar{d}}^2 = \frac{\sum (\bar{d}_i - \mu_{\bar{d}})^2}{Q}$$

$$\sigma_{\bar{d}}^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 = \frac{\sigma_x^2}{r_x} + \frac{\sigma_y^2}{r_y} \text{ если } \sigma_x^2 = \sigma_y^2 = \sigma^2 \text{ и } r_x = r_y = r, \text{ то } \sigma_{\bar{d}}^2 = \frac{2\sigma^2}{r}$$

Рис. 3.1. Составление совокупностей средних и разностей этих средних на основе двух исходных совокупностей объектов и установление связи между их параметрами (см. текст).

Имеющиеся закономерности между средними и стандартными отклонениями этих совокупностей могут быть доказаны математическими выкладками, однако ограничимся пока лишь упоминаниями о них. Средняя разностей средних равняется разнице между средними выборочными средними совокупностей X и Y , и эта разница равняется разнице между средними совокупностей X и Y :

$$\mu_{\bar{d}} = \mu_{\bar{x}} - \mu_{\bar{y}} = \mu_x - \mu_y. \text{ Если } \mu_x = \mu_y, \text{ то } \mu_{\bar{d}} = 0.$$

Дисперсия совокупностей разностей средних будет определяться по формуле

$$\sigma_{\bar{d}}^2 = \frac{\sum (\bar{d}_i - \mu_{\bar{d}})^2}{Q}$$

и будет равна сумме дисперсий соответствующих средних. Таким образом, $\sigma_{\bar{d}}^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2$. При двух выборках $\sigma_{\bar{d}}^2$ оценивается величиной $s_{\bar{d}}^2$, дисперсией выборочных средних $s_{\bar{d}}^2 = s_{\bar{x}}^2 + s_{\bar{y}}^2$.

Поскольку $s_{\bar{x}}^2 = s_x^2/r_x$ и $s_{\bar{y}}^2 = s_y^2/r_y$, то $s_{\bar{d}}^2 = (s_x^2/r_x) + (s_y^2/r_y)$.

Корень квадратный из дисперсии разностей средних принято называть стандартной ошибкой разности. Часто в статистических анализах одна дисперсия оценивается при помощи другой, рисунок 3.1.

Ниже приводятся важные зависимости между дисперсиями, которые вы будете часто использовать:

$$s_{\bar{y}}^2 = \frac{s_y^2}{r}; \quad s_{\bar{d}}^2 = s_{\bar{x}}^2 + s_{\bar{y}}^2; \quad s_{\bar{d}}^2 = \frac{s_x^2}{r_x} + \frac{s_y^2}{r_y} \text{ и если } r_x = r_y = r \text{ и } s_x^2 = s_y^2 = s^2, \\ \text{тогда } s_{\bar{d}}^2 = \frac{2s^2}{r}.$$

t -критерий для оценки существенности различий. Формула расчета t для совокупности разностей средних будет иметь вид: $t = (\bar{d} - \mu_{\bar{d}})/s_{\bar{d}}$. На основе данных опыта, представленного в таблице 3.1, проверим вероятность того, что 1-я и 2-я выборки взяты из совокупностей, имеющих идентичные средние ($\mu_1 = \mu_2$). Это предположение аналогично выше сформулированному в отношении совокупностей X и Y , только теперь мы назовем их Y_1 и Y_2 . Средняя разность между нашими выборочными средними составит $\bar{d} = 17 - 20 = 3$ ц/га.

Стандартная ошибка разности будет равна:

$$s_{\bar{d}} = \sqrt{s_{y_1}^2 + s_{y_2}^2} = \sqrt{\frac{s_1^2}{r_1} + \frac{s_2^2}{r_2}} = \sqrt{\frac{6,5}{5} + \frac{4}{5}} = \\ = \sqrt{\frac{10,5}{5}} = \sqrt{2,10} = 1,449.$$

Приняв нулевую гипотезу о том, что $\mu_1 = \mu_2$ ($\mu_{\bar{d}} = 0$), получим t из формулы

$$t = \frac{\bar{d} - \mu_{\bar{d}}}{s_{\bar{d}}} = \frac{3 - 0}{1,449} = 2,07.$$

Из таблицы А.2 (приложение) мы можем найти самое низкое значение t , вероятность получения которого равна 5%. Допустив, что $\sigma_1^2 = \sigma_2^2$, мы берем из таблицы значение t , основанное на общем числе степеней свободы для выборок, в данном случае на $4+4=8$. Ожидаемое значение t для 5%-ного уровня вероятности будет равно 2,306; итак, действие наших вариантов снова оценено как незначительное. Обратите внимание на равенство $t^2 = F$, т. е. $2,07^2 = 4,285$. Округлив данное значение, мы получим ранее рассчитанную величину $F=4,29$.

Необходимо подчеркнуть, что дисперсионный анализ и расчет величины F ведет к такому же заключению, как и оценка с помощью t -критерия. Исследователи часто высказываются в пользу t -критерия, считая его уникальным и более весомым, чем F -критерий в дисперсионном анализе. Однако оба критерия эквивалентны, более того, дисперсионный анализ обычно легче осуществляется.

Одно дополнение необходимо сделать в отношении использования t -критерия: t -оценка становится особенно уместной, если $\sigma_1 \neq \sigma_2$. Критерий F дисперсионного анализа в этом случае непригоден. Если $\sigma_1 \neq \sigma_2$ и $r_1 = r_2 = r$, то значение t , необходимое для оценки существенности, основывается на $r-1$ степенях свободы. В нашем примере $r=5$ и необходимая величина t для 5%-ного уровня составит табличное значение при 4-х степенях свободы, т. е. 2,776. Если $r_1 \neq r_2$, то необходимая величина t должна быть вычислена, так как она будет находиться между табличными значениями t при r_1-1 и r_2-1 степенях свободы. Если $\sigma_1 \neq \sigma_2$ и $r_1 \neq r_2$, то необходимое значение t рассчитывается приблизительно, по формуле

$$t = \frac{t_1 s_{y_1}^2 + t_2 s_{y_2}^2}{s_{y_1}^2 + s_{y_2}^2},$$

где t_1 и t_2 — табличные значения t для r_1-1 и r_2-1 степеней свободы соответственно.

Доверительный интервал для средней разности. В нашем примере мы имеем оценку разности между средними двух совокупностей, а именно 3 ц/га и, возможно, пожелаем вычислить доверительный интервал, в котором будет находиться истинная разность между средними совокупности, за исключением случая выемки смещенной выборки (резко отличающейся от других). С уверенностью на 95% мы можем утверждать, что $\mu_{\bar{d}}$ будет в пределах $\bar{d} \pm t_{0.05} s_{\bar{d}}$, где $t_{0.05}$ — табличное значение t , взятое из таблицы А.2 (приложение) для числа степеней свободы ошибки (табл. 3.2). Таким образом, 95%-ный доверительный интервал будет иметь следующие границы:

$$CL_{95} = 3 \pm 2,306 (1,449) = 3 \pm 3,34 = -0,34 \text{ при } 6,34 \text{ ц/га.}$$

Обратите внимание, что этот доверительный интервал включает нулевое значение, что является еще одним свидетельством отсутствия существенных различий между средними 1-го и 2-го сортов.

Наименьшая существенная разность. Наименьшая существенная разность (НСР) освещена более подробно в шестой главе «Оценка частных различий между средними»; ее упоминание

здесь связано с тем, что она является одной из форм обсуждаемой нами оценки по t -критерию. Формула расчета HCP для двух средних имеет следующий вид: $HCP = t \cdot s_{\bar{d}}$, что представляет вторую часть формулы рассмотренного выше доверительного интервала. Для опытов, включающих только два варианта, нет необходимости рассчитывать HCP , поскольку там мы имеем дело лишь с одной разностью средних, и F - или t -анализ покажет, существенна ли эта разность.

t -критерий для поделочночных сравнений. Предположив, что даты двух сортов таблицы 3.1 разбиты на пары, мы можем установить различия внутри каждой пары и оценить эти различия. Вычитая даты первого варианта из дат второго, мы получим разности внутри пар делянок: 4,5, 4,4, —2. Средняя разность равна 3 т. е. $\bar{d}=3$; дисперсия разностей будет равна

$$s_d^2 = \frac{(4-3)^2 + (5-3)^2 + \dots + (-2-3)^2}{5-1} = \frac{32}{4} = 8.$$

Дисперсия средней разности оценивается значением $s_{\bar{d}}^2 = s_d^2/r = 8/5 = 1,6$, откуда стандартная ошибка средней разности будет равна $s_{\bar{d}} = 1,265$. Формула для проверки существенности средней разности

$$t = \frac{\bar{d} - \mu_{\bar{d}}}{s_{\bar{d}}} = \frac{3-0}{1,265} = 2,37.$$

После того как вы ознакомитесь с пятой главой и поймете схемы рендомизированных блоков, предположите, что повторения таблицы 3.1 являются такими блоками, и выполните дисперсионный анализ, в результате которого вы должны получить следующие результаты: $MST=22,5$, $MSE=4$, F сортов = 5,62 и стандартная ошибка средней разности будет равна $s_{\bar{d}} = \sqrt{\frac{2s^2}{r}} = \sqrt{\frac{2(4)}{5}} = 1,265$, как и выше при t -оценке. Учитывая, что $t^2=F$, получим $2,37^2=5,62$. Важно понять, что t -оценка средней разности внутри пар делянок ведет к идентичному статистическому выводу, как и F -оценка данных опыта с двумя вариантами, заложенного по схеме рендомизированных блоков. Последний (F -анализ) обычно выполняется легче.

ОКРУГЛЕНИЕ ОПЫТНЫХ ДАННЫХ

Термины «точность» и «аккуратность» часто используют как синонимы, однако на языке статистиков они имеют разное значение. Точность выражает ширину различия между двумя вариантами, которую может выявить эксперимент при заданном уровне значимости, тогда как аккуратность показывает, насколько тщательно проводится измерение. В следующей главе мы рассмотрим методы повышения точности экспериментов, а здесь коротко укажем на аккуратность при получении данных и проведении расчетов.

По возможности необходимо иметь данные в оригинале, т. е. избегать их копирования. Если планируется использование электронных машин, то можно организовать дело таким образом, что уже при

сборе данных подготовить таблицы с исходными цифрами для перфокарт. Это избавит от ошибок, обусловленных копированием данных. Если цифры необходимо передавать, то копию нужно тотчас же сверять с оригиналом.

При получении данных их нужно контролировать на возможность появления выпадающих из общего ряда цифр, которые следует проверять для недопущения возможных ошибок.

Помните, что биологические данные достаточно переменны и без дополнительной вариации, связанной с ошибками, которых можно избежать.

При проведении взвешиваний или других измерений на экспериментальных единицах редко имеет смысл записывать цифры с точностью до знака, меньшего, чем дает одна четвертая стандартного отклонения отдельных наблюдений. Если s экспериментальной единицы равно 6,96 кг, то получим $6,96/4=1,74$. Поскольку первой значащей цифрой этого числа является единица, то данные можно округлять до 1 кг. Если s было бы равно 2,5 кг на единицу ($2,5/4=0,625$), то первой значащей цифрой были бы десятые и данные можно округлять до десятых килограмма.

Инструменты, используемые для взвешивания и других измерений, не обязательно должны иметь более высокую точность, чем требуют задачи эксперимента. Например, если ряд предстоящих взвешиваний будет округлен до одного килограмма, то предпочтительнее взять весы, имеющие шкалу в целых килограммах, чем в его долях.

Нельзя считать излишними усилиями проведение измерений с большей точностью (иметь больше значащих цифр), чем того допускает вариация экспериментальных данных; при наличии современной счетной техники вся необходимая первичная обработка данных легко осуществима, однако в научных отчетах такую излишнюю «точность» (цифры после запятой) следует опустить. Примените указанное правило округления для средних по вариантам и округлите их до первой значащей цифры, которую дает одна четвертая стандартной ошибки средних. Если стандартное отклонение отдельных наблюдений составляет 6,96 кг и в основе каждой средней лежит пять наблюдений, то $s_y=6,96/\sqrt{5}=3,11$ и $3,11/4=0,68$, следовательно, средние нужно округлять до десятых.

В дисперсионном анализе при расчете сумм квадратов лучше брать дополнительную значащую цифру; например, если исходные данные имеют десятые, то сумму квадратов нужно вычислять до сотых. Не округляйте полученные данные более точно, пока не приступите к подготовке научного отчета. При округлении последняя значащая цифра числа увеличивается на единицу, если за ней стоит цифра больше 5, или когда после 5 следуют цифры больше нуля. Если за цифрой следует нуль, то последняя нечетная цифра увеличивается на единицу, а четная остается неизменной. Например, округляя числа 21,550 и 21,450 до десятых, получим 21,6 и 21,4 соответственно.

ФАКТОРИАЛЬНЫЕ ЭКСПЕРИМЕНТЫ

В факториальных экспериментах исследуются одновременно эффекты двух или большего числа факторов. Если имеется предположение, что изменения первого фактора будут связаны с изменениями другого, то эта гипотеза может быть проверена с помощью факториальной группы вариантов, заложенных по соответствующей эксперименту схеме.

Если два или большее число факторов (каждый может иметь две или больше градаций) исследуются во всех возможных сочетаниях, то варианты в данном случае называют факториальными. Различные воздействия одного фактора на другой дают взаимодействия. Наличие взаимодействия в опыте повышает его информативность. Интервал достоверности опытных данных возрастает, что является необходимым условием хорошо спланированного эксперимента. Результаты факториального эксперимента могут быть использованы шире, чем обычного, даже если в нем не было установлено взаимодействия, поскольку главные эффекты выдержали испытание широкого круга условий.

Примерами сочетания факторов в эксперименте являются: испытание сортов при различных уровнях почвенного плодородия; оценка действия гормона на прирост массы тела ягнят (баранчиков и ярочек).

Факториальный комплект вариантов представлен в таблице 3.3.

Девять вариантов составили все возможные сочетания трех доз инсектицида и трех доз фунгицида, использованных в качестве препаратов по обработке семян лимской фасоли.

Таблица 3.3. Факториальное сочетание трех доз фунгицида и инсектицида при обработке семян лимской фасоли

Доза фунгицида	Доза инсектицида		
	J_0 (без обработки)	J_1	J_2
F_0 (без обработки)	F_0J_0	F_0J_1	F_0J_2
F_1	F_1J_0	F_1J_1	F_1J_2
F_2	F_2J_0	F_2J_1	F_2J_2

Таблица 3.4. Влияние обработки семян различными дозами фунгицида и инсектицида (г/кг) на всхожесть лимской фасоли (число проростков на 100 семян)

Фунгицид	Инсектицид			Средний эффект фунгицида
	0 (J_0)	10 (J_1)	20 (J_2)	
Средние взаимодействия $J \times F$				
0 (F_0)	68	58	48	59
84 (F_1)	94	93	90	92
165 (F_2)	89	92	92	91

Такой комплект вариантов позволяет оценить относительное влияние фунгицида и инсектицида на появление всходов лимской фасоли. Средние по вариантам в таблице 3.4 и график рисунка 3.2 иллюстрируют значение взаимодействия.

Кривые графика показывают снижение прорастания семян с увеличением дозы инсектицида на фоне без применения фунгицида. Применение фунгицида устраняло отрицательное влияние инсектицида.

Различный эффект, получаемый в результате применения фунгицида или без использования его, и называют явлением взаимодействия. В случае отсутствия взаимодействия факториальное расположение вариантов увеличивает повторность отдельных исследуемых факторов. Обратите внимание на отсутствие заметного различия в действии инсектицида при обоих уровнях (F_1 и F_2) фунгицида. Другими словами, в опыте не установлено взаимодействия $J \times F$ в отношении доз (F_1 и F_2) фунгицида. В этом случае лучшей оценкой эффекта обоих доз фунгицида будет их средняя по всем грациям инсектицида. Итоговые средние, $F=92\%$ и $F_2=91\%$ (табл. 3.4) основаны на количестве наблюдений, равном произведению, полученному от умножения цифры 3 на число повторностей отдельного варианта. Из данных таблицы следует, что в опыте не установлено эффекта увеличения дозы фунгицида на всхожесть семян.

Иногда вы можете встретить в литературе выражение «факториальные схемы». Если строго подходить, то такая терминология не совсем правильна; сочетание вариантов является факториальным, но не схема.

СХЕМА ЭКСПЕРИМЕНТА И ДИСПЕРСИОННЫЙ АНАЛИЗ

Принципиальным различием экспериментальных схем является способ объединения или группировки экспериментальных единиц. Во всех схемах экспериментальные единицы объединяют по вариантам, но в некоторых варианты группируют еще и по блокам, главным делянкам и т. п. В дисперсионном анализе используют значения этих групп под названием источников вариации (видов варьирования) для определения средних квадратов (дисперсий). Величину среднего квадрата, определяющего междуделяночную дисперсию, обусловленную действием случайного фактора, также рассчитывают, и ее называют экспериментальной ошибкой. При отсутствии фактических различий между средними вариантов, блоков или других источников вариации эти средние квадраты будут приблизительно равны. Лишь изредка случайный фактор может стать причиной значительного отклонения одного среднего квадрата от другого. Если оценка по F -критерию пока-

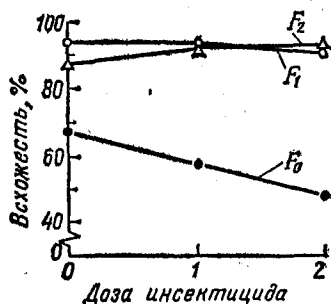


Рис. 3. 2. График средних по вариантам таблицы 3.4. F_0 — без фунгицида; F_1 — первая доза фунгицида; F_2 — вторая доза фунгицида.

вала, что средний квадрат одного из источников закономерной вариации существенно больше, чем средний квадрат случайной вариации, то мы утверждаем о наличии фактических различий между средними этого конкретного источника вариации. Но помните, что ваше заключение всегда сопряжено с определенным уровнем риска. Однако это дело экспериментатора — выбирать уровень вероятности при оценке действия вариантов. Общепринято обозначать результаты, вероятность случайного появления которых равна 5% или меньше, как существенные, а 1% или меньше, — как очень существенные. Если экспериментатор заключает: «Варианты различаются существенно», то это значит, что в случае правильности нулевой гипотезы вероятность получения таких различий между средними вариантов составляет лишь 5%. Экспериментатор рискует, утверждая, что указанные существенные различия в опыте обусловлены не случайным фактором, а фактическим действием вариантов.

В последующих главах объясняются принципиальные особенности каждой из экспериментальных схем, наиболее широко распространенных в полевых опытах, даются примеры каждой из них и показана последовательность анализа получаемых данных. Для двух схем: схемы полной рендомизации и схемы рендомизированных блоков используются одинаковые данные. Это позволяет выявить возможное преимущество одной схемы над другой и упростить расчеты настолько, что вы можете сосредоточить свое внимание на их сути и назначении.

ЗАКЛЮЧЕНИЕ

ANOVA (дисперсионный анализ) простейшего опыта из двух вариантов, случайно размещенных на одинаковом количестве экспериментальных единиц, включает следующие этапы.

1. Вычислить экспериментальную ошибку как среднюю дисперсию двух выборок, например $MSE = (s_1^2 + s_2^2)/2$.

2. Рассчитать средний квадрат вариантов (MST), основанный на нулевой гипотезе: обе выборочные средние являются оценками средней общей совокупности, т. е. $MST = rs, \frac{2}{y}$, где r — число дат, приходящихся на один вариант.

3. Рассчитать отношение дисперсий (F) — MST/MSE и сравнить фактическое и табличное значения F , что покажет вероятность случайного получения рассчитанного значения F , если нулевая гипотеза окажется верной и обе выборочные средние будут представлять общую среднюю.

Статистическая существенность разности между двумя выборочными средними может быть оценена с помощью F -критерия в дисперсионном анализе или t -критерия. Оба критерия статистически эквивалентны, $t^2 = F$. Однако дисперсионный анализ и F -оценку обычно легче выполнить.

Средняя разностей, взятых для всех возможных пар выборочных средних из двух совокупностей X и Y , обозначаемая символом μ_d , связана со средними совокупностей выборочных средних и средних

отдельных дат исходных совокупностей следующим равенством:

$$\mu_d = \mu_x - \mu_y = \mu_x - \mu_y.$$

Дисперсия средней разности σ_d^2 оценивается на основе двух выборок с помощью s_d^2

$$s_d^2 = s_x^2 + s_y^2 = \frac{s_x^2}{r_x} + \frac{s_y^2}{r_y} \text{ и если } s_x^2 = s_y^2 \text{ и } r_x = r_y = r, \\ \text{то } s_d^2 = \frac{2s^2}{r}.$$

Избегайте лишних знаков после запятой в данных при подготовке научного отчета. Округляйте средние до той значащей цифры, которую дает одна четвертая стандартной ошибки средних.

Факториальный эксперимент — это эксперимент, в котором изучают все возможные сочетания по крайней мере двух градаций двух или большего числа факторов.

Экспериментальные схемы представляют способ объединения или группировки экспериментальных единиц.

СХЕМА ПОЛНОЙ РЕНДОМИЗАЦИИ

Эта схема, наиболее простая из всех возможных, предусматривает случайное размещение вариантов по всем экспериментальным единицам установленного количества. Она особенно эффективна в условиях незначительной вариации между экспериментальными единицами, связанной с их местоположением, возрастом, темпераментом и другими поддающимися распознаванию (учету) свойствами. Схема полной рендомизации обладает гибкостью в отношении практического расположения экспериментальных единиц, дает максимальное число степеней свободы для экспериментальной ошибки и минимальное значение F , необходимое для оценки статистической достоверности (значимости). Ее недостатком часто является наличие дополнительной, поддающейся учету вариации, которая входит в данном случае в так называемую экспериментальную ошибку. Другие схемы при их правильном использовании способны сократить дополнительную вариацию, что делает возможным вскрыть значимость более тонких различий по вариантам.

РЕНДОМИЗАЦИЯ

Номер может быть произвольно закреплен за каждым числом делянок или группой животных, используемых в эксперименте.

Количество экспериментальных единиц в опыте будет равно произведению числа вариантов на повторность. Для решения вопроса о том, на какие экспериментальные единицы распределить каждый вариант, удобно пользоваться таблицей случайных чисел. Если каждый вариант должен повториться четыре раза, то первые номера, взятые из таблицы, относят к варианту 1, следующие четыре случайных числа — к варианту 2 и т. д. Например, предположим, что мы хотим изучить действие трех гормонов (каждый при однократной дозе) на прирост массы тела ягнят. Таким образом, включая контроль, мы будем иметь четыре варианта. Если мы возьмем за экспериментальную единицу одного ягненка и примем 4-кратную повторность каждого варианта, то нам понадобится 16 ягнят. Каждый из 16 ягнят, отобранных для исследования, получит бирку с номером от 1 до 16. Используя таблицу А.1 (приложение), начнем с двух однозначных цифр, например в 5-й и 6-й колонках. Пойдем вниз по этим колонкам однозначных цифр,

Таблица 4.1. Приросты массы (в фунтах *) тела ягнят за 100 дней откорма в зависимости от вида гормонов (цифры в скобках означают номера на бирках 16 исследуемых ягнят, случайно отобранных для получения указанного гормона)

Варианты (гормоны)	Повторность				Варианты	
	1	2	3	4	суммы ($Y_{i..}$)	среднее ($\bar{Y}_{i..}$)
1 (контроль)	47 (14)	52 (13)	62 (9)	51 (8)	212	53
2	50 (12)	54 (11)	67 (6)	57 (5)	228	57
3	57 (2)	53 (7)	69 (1)	57 (15)	236	59
4	54 (3)	65 (4)	75 (10)	59 (16)	252	63
					$928 = Y_{..}$	$58 = \bar{Y}_{..}$

* 1 фунт=0,45 кг. Перевод данных этой и ряда последующих таблиц в метрическую систему измерений изменяет выводы на противоположные. — *Прим. пер.*

затем вверх по 6-й 7-й и вниз по 7-й и 8-й сдвоенным колонкам, откуда возьмем первые четыре номера (ягненка) между 1-м и 16-м для варианта 1 (14, 13, 9, 8), другие четыре — для варианта 2 (12, 11, 6, 5), третьи четыре — для варианта 3 (2, 7, 1, 15). Оставшиеся четыре ягненка (3, 4, 10, 6) относят к варианту 4. После откормочного периода данные по приросту массы тела ягнят были сведены в таблицу (табл. 4.1).

ДИСПЕРСИОННЫЙ АНАЛИЗ

Источники вариации и степени свободы. Анализ дисперсий начинается после заполнения первых двух колонок сводной таблицы (табл. 4.2). В опытах, заложенных по схеме полной рендомизации, имеется лишь два вида варьирования: между экспериментальными единицами внутри варианта, это то, что мы называем экспериментальной ошибкой, и между средними вариантов.

Число степеней свободы на единицу меньше числа наблюдений для каждого вида варьирования: в опыте 4 варианта, следовательно, будет $3df$; в каждый вариант входят четыре экспериментальные единицы, следовательно, $3df$ каждого варианта, умноженное на 4 варианта, дает $12df$ для ошибки. Выражение R/T означает повторения внутри вариантов. Число степеней свободы для общей вариации в эксперименте будет на единицу меньше числа экспериментальных единиц: $16-1=15df$. Обратите внимание на аддитивность степеней сво-

Таблица 4.2. Сводная таблица дисперсионного анализа

Виды варьирования	Степени свободы (df)	Суммы квадратов (ss)	Средний квадрат (ms)	$F_{\text{факт}}$	$F_{\text{табл}}$	
					5%	1%
Общее	15	854				
Вариантов	3	208	69,3	1,29	3,49	5,95
Ошибки (R/T)	12	648	53,8			

боды, т. е. сумма по видам варьирования дает их общее число. Это обеспечивает возможность определения степеней свободы для ошибки путем вычитания другого компонента из их общего числа: $15-3=12$.

Для удобства расчетов числа степеней свободы и суммы квадратов для ошибки мы помещаем в сводной таблице дисперсионного анализа сначала общую вариацию, хотя вычисляем ее сумму квадратов после суммы квадратов для вариантов.

Корректирующий фактор (C).

$$C = \frac{(Y^2)}{rn} = \frac{928^2}{4(4)} = 53\,824.$$

Сумма квадратов и средние квадраты вариантов SST и MST . $SST = (\sum Y_i^2 / r) - C$, где Y_i — суммы по вариантам и r — повторность каждого варианта.

$$SST = \frac{212^2 + 228^2 + \dots + 252^2}{4} - C = 54032 - 53824 = 208.$$

Вносим SST в таблицу 4.2. Средний квадрат для вариантов определяем путем деления SST на число степеней свободы для вариантов. $MST = SST/df(T) = 208/3 = 69,3$, вносим в таблицу 4.2.

Для калькуляторов с клавишей стандартного отклонения $MST = s_T^2/r$, где s_T^2 — дисперсия сумм по вариантам (в данном случае 212... 252) и r — число дат в каждой сумме. Тогда $SST = df(MST)$. Для данного примера $s_T^2 = 277,33$; $MST = 277,333/4 = 69,3$, как и выше, и $SST = 3(69,3) = 208$, как и выше.

Общая SS . Нам не надо определять средний квадрат для общего варьирования, поскольку он содержит дисперсии всех видов варьирования.

$$SS = \sum Y_{ij}^2 - C = 47^2 + 50^2 + \dots + 59^2 - C = 54\,678 - 53\,824 = 854.$$

Ошибки: $SSE = SS - SST = 854 - 208 = 646$; $MSE = SSE/df(E) = 646/12 = 53,8$.

При работе на таком калькуляторе используется другая формула для расчета общего варьирования: $SS = s_{ij}^2(nr-1)$, где s_{ij}^2 — дисперсия всех дат эксперимента (47... 59) будет равна 73,80; $nr-1 = 4(4)-1 = 15$, что составляет число степеней свободы для общего варьирования. Таким образом, получим $SS = 56,933(15) = 854$, как и выше.

Величина F . Значение F для вариантов получают путем деления MST на MSE , т. е. $F = MST/MSE = 75,3/73,4 = 1,02$. Теоретическое значение F для оценки существенности различий берут из таблицы А.3 (приложение) для числа степеней свободы вариантов — «шапка» и для ошибки — левая сторона таблицы. Поскольку фактическое значение F составляет 1,29, что значительно меньше табличного F для 5%-ного уровня значимости, нам следует принять нулевую гипотезу и сделать вывод об отсутствии фактических различий между вариантами. Однако учтите, что такое заключение не доказывает отсутствия различий между вариантами. Возможно, что различия, обусловленные действием вариантов, существуют, однако наш эксперимент недостаточно «чувствителен», чтобы вскрыть их при заданном уровне значимости.

СУТЬ И НАЗНАЧЕНИЕ АНАЛИЗА

Проверяя нулевую гипотезу, мы предполагаем отсутствие действия вариантов и что средние вариантов различаются настолько, на сколько могли бы различаться выборки из одной совокупности. Таким образом, дисперсия экспериментальных единиц (наблюдений), называемая экспериментальной ошибкой (σ^2), может быть определена на основе вариации выборочных средних путем использования зависимости $s_y^2 = s^2/r$, откуда $s^2 = r s_y^2 = MST$.

Дисперсия средних определяется по формуле:

$$s_y^2 = \frac{\sum (\bar{Y}_{t.} - \bar{Y}_{..})^2}{n-1} = \frac{(53-58)^2 + \dots + (63-58)^2}{4-1} = 17,33,$$

где n — число вариантов. Тогда $MST = r s_y^2 = 4(17,33) = 69,3$ является оценкой вариабельности отдельных наблюдений (σ^2), рассчитанной на основе вариации между средними вариантами.

Дисперсии внутри каждого варианта дают независимую оценку σ^2 , и взвешенная средняя этих дисперсий является лучшей оценкой дисперсии отдельных наблюдений, основанной на вариабельности внутри вариантов. Таким образом,

$$s^2 = \frac{(r_1-1)s_1^2 + (r_2-1)s_2^2 + (r_3-1)s_3^2 + (r_4-1)s_4^2}{(r_1-1) + (r_2-1) + (r_3-1) + (r_4-1)}.$$

Обратите внимание, что «весами» (частотами) каждой оценки $\sigma^2 (s_1^2, s_2^2 \text{ и т. д.})$ являются их степени свободы. Если все варианты имеют одинаковую повторность, т. е. $r_1 = r_2 = \dots = r_4$, то

$$s^2 = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2}{n},$$

где n — число вариантов. В нашем опыте дисперсии внутри вариантов будут составлять следующие величины:

$$s_1^2 = \frac{\sum (Y_{1j} - \bar{Y}_{1.})^2}{r_1-1} = \frac{(47-53)^2 + \dots + (51-53)^2}{4-1} = 40,67;$$

$$s_2^2 = \frac{(50-57)^2 + \dots + (57-57)^2}{4-1} = 52,67;$$

$$s_3^2 = \frac{(57-59)^2 + \dots + (57-59)^2}{4-1} = 48,0;$$

$$s_4^2 = \frac{(54-63)^2 + \dots + (59-63)^2}{4-1} = 74,0.$$

Средняя этих дисперсий дает оценку σ^2 , называемую экспериментальной ошибкой,

$$s^2 = \frac{(s_1^2 + s_2^2 + s_3^2 + s_4^2)}{n} = \frac{40,67 + 52,67 + 48,0 + 74,0}{4} = 53,8.$$

Итак, мы имеем две оценки дисперсии экспериментальных единиц (σ^2): средняя дисперсия внутри вариантов (MSE) и полученная на основе вариабельности средних по вариантам (s_y^2) величина $MST = r s_y^2$.

В случае правильности нулевой гипотезы, т. е. если все четыре выборки являются случайными выборками из одной совокупности, то величины MST и MSE будут близкими, а отношение MST/MSE (F -критерий) будет около 1, до тех пор, пока мы не возьмем необычную (смещенную) выборку. В этом опыте указанное отношение равно величине 1,29, вероятность появления которой превышает 25% для того случая, когда не установлено фактического действия вариантов. Таким образом, мы не отвергаем нулевую гипотезу и делаем вывод об отсутствии существенных различий между вариантами. Если мы установим наличие существенного различия между средними вариантов, то следующий вопрос, который необходимо решать, — какие средние существенно различаются. Этот этап анализа называют оценкой частных различий между средними. Ему посвящена глава 6.

ЗАКЛЮЧЕНИЕ

Схема полной рендомизации заслуживает внимания для тех условий, когда в эксперименте нет дополнительной вариабельности среди экспериментальных единиц, кроме вариации, обусловленной действием вариантов. Она является самой гибкой в отношении практического расположения экспериментальных единиц. Она дает максимально возможное число степеней свободы для оценки дисперсии экспериментальных единиц (экспериментальная ошибка) и максимальное значение F для оценки статистической существенности.

СХЕМА РЕНДОМИЗИРОВАННЫХ (организованных) БЛОКОВ (повторений)

В этой схеме варианты размещаются случайно внутри группы экспериментальных единиц, называемых блоком или повторением. Выражение «блок» предпочтительнее, чем «повторение», поскольку оно устраняет путаницу этого понятия с повторностью вариантов схемы полной рендомизации. Цель схемы — сведение вариабельности экспериментальных единиц внутри блоков до возможного минимума и обеспечение максимальных различий между блоками. В случае отсутствия различий между блоками эта схема не будет способствовать повышению точности сравнения вариантов. Блоки должны составить, насколько это возможно, однородные экспериментальные единицы. Для достижения однородности экспериментальные единицы необходимо сгруппировать на основе возраста, массы, общего поведения, их предшествующей продуктивности или урожайности или некоторых других свойств, которые будут обеспечивать однородность внутри группы. В полевых опытах соседние делянки обычно различаются меньше по уровню плодородия почвы, чем делянки, удаленные друг от друга. Блоки можно разместить компактно при плотном расположении делянок, обычно вытянутой формы. В зависимости от цели эксперимента необходимо иметь как можно меньшее количество вариантов. С увеличением размеров блока возрастает внутриблоковая вариация. Не обязательно иметь блоки одинаковой формы, однако в полевых опытах с растениями это требование желательно соблюдать, поскольку различия по форме обычно увеличивают внутриблоковую вариацию. Если ожидается изменение плодородия почвы внутри опытного участка, то блоки следует размещать поперек направления варьирования почвенного плодородия, а делянки внутри блоков — параллельно этому направлению (рис. 5.1). Каждый из вариантов размещается внутри блока по экспериментальным единицам обычно один раз, однако внутриблоковая повторность всех или некоторых вариантов может составить два или большее число повторностей. Наибольшая эффективность обычно достигается

I	II	III	IV
D	A	C	C
A	D	D	B
B	C	B	D
C	B	A	A

Низкое плодородие → *Высокое плодородие*

Рис. 5.1. Четыре варианта (A, B, C, D) повторяются четыре раза (I, II, III, IV) в схеме рендомизированных блоков.

при однократной повторности вариантов внутри блока. Для сведения экспериментальной ошибки до минимума необходимо направить все усилия на поддержание одинаковых условий для экспериментальных единиц внутри блоков.

РЕНДОМИЗАЦИЯ

После того как экспериментальные единицы сгруппированы по блокам, внутри блоков случайно размещают варианты, причем рендомизация вариантов осуществляется внутри каждого блока отдельно. Например, четыре варианта (рис. 5.1) можно разместить случайно следующим образом. Произвольно возьмем 15-й ряд таблицы А.1 (приложение) и выберем вдоль него однозначные цифры от 1 до 4, которые будут представлять варианты от А до Д: 4, 1, 2, 3...— этот ряд цифр и составит порядок размещения вариантов в I-м блоке. Продолжим выбор цифр вдоль 15-го ряда до конца (справа налево) и назад вдоль 16-го ряда, из которого возьмем цифры 1, 4, 3, 2, которые составят порядок размещения вариантов в блоке II. Подобным образом рендомизация выполняется для III и IV блоков.

ДИСПЕРСИОННЫЙ АНАЛИЗ

Данные для анализа возьмем те же, что и в 4-й главе. Эксперимент заключался в установлении действия гормона (стилбестрола) при имплантации подкожно в области уха на прирост массы тела ягнят (ба-

Таблица 5.1. Действие стилбестрола на прирост массы тела валухов и ярок

Пол	Стилбестрол (S)	
	0	3 мг/животное
Самки (F)	FS_0	FS_3
Самцы (M)	MS_0	MS_3

Таблица 5.2. Прирост массы (в фунтах) тела ягнят за 100 дней откорма

Варианты	Блоки				Варианты	
	I	II	III	IV	суммы (ΣY_i)	средние ($\bar{Y}_i$)
A FS_0	47	52	62	51	212	53
B MS_0	50	54	67	57	228	57
C FS_3	57	53	69	57	236	59
D MS_3	54	65	74	59	252	63
Суммы ($\Sigma Y_{.j}$)	208	224	272	224	928 = $\Sigma Y_{..}$	
Средние ($\bar{Y}_{.j}$)	52	56	68	56		58 = ($\bar{Y}_{..}$)

Таблица 5.3. Сводная таблица дисперсионного анализа

Варьирование	df	SS	MS	F _{факт}	F _{табл}	
					5%	1%
Общее, в том числе:	15	854				
блоков	3	576	192,0	24,69	3,86	6,99
вариантов	3	208	69,3	8,91		
ошибки (BT) *	9	70	7,78			

* BT — означает взаимодействие блоков с вариантами, которое обуславливает разное действие вариантов по блокам. Случай равновеликого эффекта одноименных вариантов по всем блокам и отсутствия указанного взаимодействия является редким исключением.

ранчиков и ярочек). Таким образом, варианты составили факториальную группу с двумя факторами: пол и гормон, каждый на двух градациях (табл. 5.1). В этом опыте блоками служили четыре различных пастбища.

Итак, повторности таблицы 4.1 стали блоками, а варианты составили факториальный комплекс таблицы 5.1. Данные таблицы 4.1 размещены в таблице 5.2 в соответствии с требованиями новой схемы. Итоги дисперсионного анализа сведены в таблице 5.3.

Источники вариации и степени свободы. В новой экспериментальной схеме мы имеем дополнительный источник вариации — варьирование блоков. Поскольку каждый вариант встречается одинаковое число раз в каждом блоке, то различия между блоками связаны не с действием вариантов, а с причинами внутриблокового характера. Этот компонент вариации можно вычесть из общей суммы квадратов и соответствующим образом снизить неучтенную вариацию (экспериментальную ошибку).

Число степеней свободы будет на единицу меньше числа наблюдений, связанных с каждым источником вариации. Поскольку имеется 16 экспериментальных единиц (групп ягнят), то общее число степеней свободы будет равно 15. В опыте четыре блока и четыре варианта, поэтому оба вида варьирования получают по 3 df. Число степеней свободы для ошибки можно определить по разности 15—3—3=9 или путем умножения степеней свободы для блоков на степени свободы для вариантов, т. е. 3×3=9. В такой схеме, где каждый вариант встречается один раз в каждом блоке, число степеней свободы для ошибки всегда равно df блоков, умноженной на df вариантов.

Корректирующий фактор.

$$C = \frac{Y^2}{rn},$$

где r — повторность, а n — число вариантов, тогда

$$C = \frac{928^2}{4(4)} = 53\,824.$$

Суммы квадратов и средние квадраты.

Б ло ки. $SSB = \frac{\sum Y_{.j}^2}{n} - C;$

$$SSB = \frac{208^2 + \dots + 224^2}{4} - 53\,824 = 54\,400 - 53\,824 = 576.$$

Обратите внимание, что делитель и в выражении $\sum Y_{.j}^2/n$ является числом дат, слагающих сумму в числителе; в данном случае число вариантов

$$MSB = \frac{SSB}{df(B)} = \frac{576}{3} = 192,0.$$

Аналогично этому, используя калькулятор с программой для расчета стандартного отклонения, получим средний квадрат блоков по формуле $MSB = s_b^2/n$, где s_b^2 — дисперсия блоковых сумм: 208 ... 224, а n — число дат, слагающих суммы.

$$MSB = \frac{678}{4} = 192 \text{ и } SSB = 3(192) = 576.$$

В а р и а н т ы. $SST = \frac{\sum Y_i^2}{r} - C;$

$$SST = \frac{212^2 + \dots + 252^2}{4} - 53\,824 = 54\,032 - 53\,824 = 208;$$

$$MST = \frac{SST}{df(T)} = \frac{208}{3} = 69,3.$$

Используя клавишу стандартного отклонения, вводим в калькулятор суммы по вариантам (212 .. 252), получаем 16,653 и $MST = 16,653^2/4 = 69,3$.

Общ е е. $SS = \sum Y_{ij}^2 - C;$

$$SS = 47^2 + 52^2 + \dots + 59^2 - C = 54\,678 - 53\,824 = 854.$$

Снова через клавишу стандартного отклонения вводим даты 47... 59, получаем $s = 7,545$, или $7,545^2(15) = 854$.

О ш и б к и. $SSE = SS - SST - SSB$ или

$$SSE = 854 - 208 - 576 = 70.$$

Если различные суммы квадратов вычислены в указанном выше порядке, SSE легко получают из разности после того, как рассчитана общая сумма квадратов

$$MSE = \frac{SSE}{df(E)} = \frac{70}{9} = 7,78.$$

СУТЬ И НАЗНАЧЕНИЕ АНАЛИЗА

Прежде чем продолжить дисперсионный анализ, будет полезно разобраться в том, что было сделано и почему так сделано при расчете каждого среднего квадрата.

Средний квадрат блоков. Предположив отсутствие фактических различий между средними блоков (снова нулевая гипотеза), рассчита-

ем оценку вариабельности экспериментальных единиц на основе дисперсии блоковых средних. Таким образом, $s^2 = MSB = ns_{yb}^2$, где n — число вариантов и s_{yb}^2 — дисперсия блоковых средних. Обратите внимание, что в этой формуле используется связь дисперсии средних с дисперсией экспериментальных единиц, т. е. $s^2 = ns_y^2$. Поскольку $s_{yb}^2 = \frac{\sum (\bar{y}_{.j} - \bar{y}_{..})^2}{(r-1)}$, то формула для среднего квадрата блоков примет следующий вид:

$$MSB = n \left[\frac{\sum (\bar{y}_{.j} - \bar{y}_{..})^2}{r-1} \right],$$

где $\bar{y}_{.j}$ — средние блоков, $\bar{y}_{..}$ — общая средняя и r — число блоковых средних. Подставляя исходные данные в формулу, получим

$$MSB = \frac{4 [(52-58)^2 + (56-58)^2 + \dots + (56-58)^2]}{4-1} = \frac{4 (144)}{3} = 192,0.$$

Средний квадрат вариантов. Используя снова нулевую гипотезу и предположив, что между средними вариантов нет фактических различий, рассчитаем дисперсию $s^2 = MST = rs_y^2$, где r — повторность и s_y^2 — дисперсия средних вариантов. Эта дисперсия является второй оценкой дисперсии экспериментальных единиц, основанной на вариабельности между средними вариантов. Снова и в этой формуле используется связь между оценкой дисперсии средних (s_y^2) и оценкой дисперсии отдельных наблюдений исходной совокупности (s^2). Раскрыв значение дисперсии средних, получим следующую формулу:

$$MST = \frac{r \sum (\bar{y}_{i.} - \bar{y}_{..})^2}{n-1},$$

где $\bar{y}_{i.}$ — любая средняя вариантов и n — число вариантов. SST представляет числитель, а степени свободы для вариантов — знаменатель. Подставляя соответствующие данные в формулу, получим

$$MST = \frac{4 [(53-58)^2 + (57-58)^2 + \dots + (63-58)^2]}{4-1} = \frac{4 (52)}{3} = 69,3.$$

Средний квадрат ошибки (MSE) представляет вариабельность между экспериментальными единицами, которая остается после вычитания варьирования других источников из общей вариации. На данном этапе полезно рассмотреть, что подразумевается под вычлениением эффекта блоков и вариантов из общей вариации.

Модель, соответствующая схеме рендомизированных блоков, включает следующие компоненты: $Y_{ij} = \bar{Y}_{..} + T_i + B_j + e_{ij}$. Это означает, что клетки таблицы с двумя входами, такой, как таблица 5.2, заполнены средним значением всех наблюдений (дат), т. е. $\bar{Y}_{..}$, эффектом вариантов, т. е. T_i , эффектом блоков (B_j) и остаточным компонентом (e_{ij}), представляющим неучтенное варьирование, называемое экспериментальной ошибкой. Каждый вариант и каждый блок имеют свой эффект, который определяется по разности между средними вариантов или бло-

ков и общей средней опыта. Например, эффект варианта FS_0 является одним для всех повторений этого варианта и составляет: $53-58=-5$. В общем виде эффект этого варианта определяется из равенства $FS_0 = -\bar{Y}_1 - \bar{Y}_{..}$, а для определения эффектов всех вариантов формула примет вид $\bar{Y}_i - \bar{Y}_{..}$, где $\bar{Y}_i$ — любое значение средней по вариантам из всех имеющихся. Подобным образом определяют блокные эффекты $B_j = \bar{Y}_{.j} - \bar{Y}_{..}$.

Заменяя T и B в первоначальной модели на их установленные эффекты, получим $\bar{Y}_{ij} = \bar{Y}_{..} + (\bar{Y}_i - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) + e_{ij}$. Теперь мы можем, записав модель в измененном виде, определить эффект ошибки для любой клетки таблицы с двумя входами следующим образом: $e_{ij} = \bar{Y}_{ij} - \bar{Y}_{..} - (\bar{Y}_i - \bar{Y}_{..}) - (\bar{Y}_{.j} - \bar{Y}_{..})$. Например, определим e_{13} (для 1-го ряда и 3-й колонки):

$$e_{13} = 62 - 58 - (53 - 58) - (68 - 58) = 62 - 58 + 5 - 10 = -1.$$

Формула для e_{ij} может быть упрощена и дальше путем раскрытия скобок и соответствующего сокращения $+\bar{Y}_{..}$ на $-\bar{Y}_{..}$ в исходном уравнении

$$e_{ij} = Y_{ij} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}_{..}$$

откуда $e_{13} = 62 - 68 - 53 + 58 = -1$, как и выше.

Проведя указанную операцию по вычислению эффектов вариантов и блоков для всех клеток таблицы 5.2, мы получим таблицу ошибок (неучтенных эффектов), таблица 5.4.

Сумма квадратов этих ошибок, деленная на общее число степеней свободы, за минусом степени свободы для других поддающихся учету источников вариации — блоков и вариантов, дает средний квадрат ошибки (MSE), который и представляет s^2 — неучтенную вариацию экспериментальных единиц. Таким образом,

$$s^2 = MSE = \frac{0^2 + 1^2 + \dots + (-2)^2}{15 - 3 - 3} = \frac{70}{9} = 7,78.$$

ВЕЛИЧИНА F

Отношение дисперсий, или F -критерий, используются для определения вероятности получения таких средних по вариантам и блокам, которые варьируют одинаково со средними нашего эксперимента в случае

Таблица 5.4. Таблица неучтенных эффектов (ошибок)

Варианты	Блоки			
	I	II	III	IV
FS_0	0	1	-1	0
MS_0	-1	-1	0	2
FS_1	4	-4	0	0
MS_1	-3	4	1	-2

отсутствия фактических различий по вариантам и блокам. Мы произвели оценку σ^2 — дисперсии экспериментальных единиц совокупности тремя способами: 1) на основе вариации между средними вариантов (MST); 2) на основе вариации между средними блоков (MSB) и 3) на основе вариации между экспериментальными единицами за вычетом эффектов блоков и вариантов (MSE). В случае отсутствия различий между блоками и вариантами все три средних квадрата должны быть приблизительно равны:

$$F_{\text{(блоки)}} = \frac{MSB}{MSE} = \frac{192,0}{7,78} = 24,69; \quad F_{\text{(варианты)}} = \frac{MST}{MSE} = \frac{69,3}{7,78} = 8,91.$$

F -значение, необходимое для оценки статистической существенно-сти при 3-х (числитель) и 9-ти (знаменатель) степенях свободы, берут из таблицы А.3 (приложение) и вносят в сводную таблицу дисперсионного анализа (табл. 5.3).

Поскольку фактическое (рассчитанное) значение F как для блоков, так и вариантов превышает теоретическое (табличное) на 1 %-ном уровне значимости, то в случае правильности нулевых гипотез мы можем сказать, что шанс случайного получения таких выборок для блоков и вариантов в нашем эксперименте составляет менее 1 из 100. Мы рискуем, утверждая, что такой шанс не появится, отвергаем нулевые гипотезы и делаем вывод о наличии фактических различий по блокам и вариантам. Далее надо установить, какие варианты существенно различаются. Этому вопросу посвящена глава 6. Прежде чем закончить со схемой рендомизированных блоков, мы должны указать на ее более высокую эффективность по сравнению со схемой полной рендомизации. Точность нашего опыта возросла благодаря вычлениению значительных различий между блоками из общего варьирования, что позволило вскрыть более тонкие различия по вариантам, неуловимые для схемы полной рендомизации.

ЗАКЛЮЧЕНИЕ

В схеме рендомизированных блоков: блоками являются группы экспериментальных единиц, расположенные или выбранные с учетом требований к размещению на них вариантов таким образом, что внутриблоковое варьирование сводится до минимума, тогда как варьирование между блоками максимально увеличивается. Варианты случайно размещают одинаковое число раз (обычно 1 раз) по экспериментальным единицам внутри блока. Рендомизация проводится независимо для каждого блока. По сравнению со схемой полной рендомизации в данной схеме степени свободы для экспериментальной ошибки уменьшаются на число степеней свободы для блоков. Варьирование блоков устраняется из экспериментальной ошибки. Таким образом, с увеличением вариабельности между блоками повышается эффективность этой схемы, т. е. ее способность по выявлению имеющихся различий по вариантам.

ВЫДЕЛЕНИЕ СУЩЕСТВЕННЫХ СРЕДНИХ

(оценка частных различий)

Как мы уже убедились, эксперимент ставится с целью ответа на конкретные вопросы, заранее поставленные исследователем. Эти вопросы играют важную роль для выбора необходимых вариантов, схемы эксперимента и соответствующего способа сравнения средних по вариантам.

Обычно правильный и соответствующий выбор вариантов дает возможность привлечь для ответа на важные вопросы исследования планируемые F -критерии. Например, при изучении двух или большего числа факторов факториальная группа вариантов позволяет ответить на вопрос о том, каким будет взаимодействие факторов. Даже в случае отсутствия взаимодействия выводы относительно средних эффектов факторов имеют более широкие пределы применения, поскольку каждый фактор был проверен в интервале условий. Градации вариантов можно спланировать так, чтобы определить не только наличие действия вариантов, но также лучший способ характеристики этого действия и количественно оценить связь дозы варианта и ее действия. Варианты могут быть разбиты по группам с общими свойствами, что позволит провести важное сравнение групп по F -критериям. Такие запланированные оценки по F -критериям позволяют более точно сравнить средние, чем критерии множественного сравнения. Последние следует применять лишь для случаев, где не выявлено логической связи между вариантами.

Вывод о существенности различий, сделанный на основе F -оценки, ставит вопрос: «Какие средние по вариантам различаются существенно?» Ниже дается краткое объяснение трех из наиболее распространенных способов выделения существенных средних.

НАИМЕНЬШАЯ СУЩЕСТВЕННАЯ РАЗНОСТЬ ($НСР$)

Эту оценку используют лишь после того, как установлено наличие существенных различий в опыте по F -критерию. Если подходить строго, то $НСР$ следует использовать только для сравнения соседних средних ранжированного ряда (расположение средних в порядке возрастания их значений). В случае ее применения для оценки всех возможных различий между средними ряд разностей может быть существенным,

но не на заданном уровне значимости. Вместо оценки различий на 5%-ном уровне, сравнения между средними, удаленными друг от друга в ранжированном ряду больше чем на два места, проводят на более низких уровнях значимости. *НСР* можно использовать для сравнения соседних средних и, когда ее используют для проведения сравнений, запланированных до анализа, то возможность ошибочного заключения уменьшается. Большим достоинством *НСР* является то, что она легко вычисляется и представляет единственную величину для проведения сравнений.

Как указывалось ранее, *НСР* является одной из форм *t*-оценки. Ее формулу получают из формулы для *t*-критерия, оценивающего статистическую существенность разности между двумя средними $t = (\bar{d} - \mu_{\bar{d}}) / s_{\bar{d}}$. Пусть разность между двумя средними ($\bar{Y}_1 - \bar{Y}_2 = \bar{d}$) будет составлять самую малую величину, вероятность случайного получения которой равна 5% или больше, если мы возьмем выборки из совокупности разностей между средними, где их средняя равна нулю ($\mu_{\bar{d}} = 0$). Заменив $\bar{d}$ на *НСР* и $\mu_{\bar{d}}$ нулем, получим новую формулу $t = \text{НСР} / s_{\bar{d}}$, откуда $\text{НСР} = t s_{\bar{d}}$, где $s_{\bar{d}} = (s_1^2/r_1) + (s_2^2/r_2)$; s_1^2 и s_2^2 — оценки дисперсий между деланками 1-го и 2-го вариантов; r_1 и r_2 — число экспериментальных единиц, составляющих 1-й и 2-й варианты соответственно. Дисперсионный анализ предполагает равенство оценок дисперсий $s_1^2 = s_2^2$ и обычно, $r_1 = r_2$, поэтому $\text{НСР} = t \sqrt{2s^2/r}$, где s^2 — средний квадрат ошибки, r — повторность и t — табличная величина для числа степеней свободы ошибки.

В случае сравнения двух вариантов, имеющих разную повторность, $\text{НСР} = t \sqrt{(s^2/r_1) + (s^2/r_2)}$, где r_1 и r_2 — повторность каждого варианта.

В качестве примера использования *НСР* возьмем наш эксперимент по оценке действия гормона стилбестрола на прирост массы тела ягнят (табл. 5.2). Средние этого опыта составили: $FS_0 = 53$; $MS_0 = 57$; $FS_3 = 59$ и $MS_3 = 63$ фунта за 100 дней откорма.

$$\begin{aligned} \text{НСР}_{.05} &= t_{.05} \sqrt{\frac{2s^2}{r}} = 2,262 \sqrt{\frac{2(7,78)}{4}} = \\ &= 2,262 (1,972) = 4,46 \text{ фунта/животное за 100 дней.} \end{aligned}$$

После сравнения пересчитанных средних с *НСР* мы убеждаемся в отсутствии существенных различий, однако ранее проведенная *F*-оценка показала их наличие в опыте. Сравнив все средние, мы устанавливаем, что стилбестрол улучшает прирост массы тела ягнят, как ярок (59—53=6), так и баранчиков (63—57=6). В зависимости от пола различия в действии препарата несущественны.

Оценка различий с помощью *НСР*, а вернее *t*-оценка каждой разности, ведет к такому же статистическому заключению, как и *F*-оценка тех же разностей. Исследователи часто заблуждаются по этому вопросу, перепроверяя существенность различий на основе всех трех оценок. Не делайте этого, так как все оценки дадут одинаковый результат. Для примера возьмем разность $FS_0 - MS_0 = 53 - 57 = -4$ и оценим с помощью трех критериев.

1. Поскольку $HCP=4,5$, то разность незначительна.

2. $t = (\bar{d} - \mu_d) / s_d = 4 / 1,972 = 2,028$ (фактическое t). Табличное t для 5%-ного уровня значимости и 9-ти степеней свободы будет равно 2,262. Поэтому разность снова незначительна.

3. $F = MS(FS_0 - MS_0) / MSE$. Поскольку $MS(FS_0 - MS_0)$ — средний квадрат, основанный на 1-й степени свободы, то он будет равен сумме квадратов $SS(MS_0 - FS_0)$, а $SS(MS_0 - FS_0) = (228^2 + 212^2) / 4 - (228 + 212)^2 / 8 = 24\,232 - 24\,200 = 32$, $F = 32 / 7,78 = 4,11$. Табличное значение F для 5%-ного уровня, 1-й и 9-й степеней свободы равно 5,32. Поэтому еще раз получаем, что разность незначительна. Фактическое значение F можно получить, используя равенство $t^2 = F = 2,028^2 = 4,11$. Эта связь двух критериев справедлива во всех случаях, и они оба ведут к одинаковому статистическому заключению.

HCP является критерием фиксированного ранга (точечным критерием), поскольку она дает одно значение для оценки всех разностей. К другим, менее используемым (устаревшим) точечным критериям относят тесты Тьюки и Шеффи [2].

МНОГОРАНГОВЫЕ КРИТЕРИИ

Эти критерии называют так потому, что они обеспечивают множество значений для оценки парных сравнений между несколькими средними. Для ранжированного ряда средних многоранговые критерии дают предельную величину существенной разности средних, которая возрастает по мере изменений от наименьшего до наибольшего значения средних. Давно известным многоранговым методом оценки разностей, предусматривающим существенные уровни для отклонений между всеми средними, является анализ Стьюдента — Ньюмена — Кьюлза (Bancroft T. A., 1968). Из всех многоранговых методов в данной книге рассматривается лишь критерий Дункана, поскольку он и (или) правильное применение HCP после того, как по F -критерию установлены существенные различия между вариантами, представляют адекватный анализ логической оценки парных сравнений средних *.

Критерий Дункана. Этот критерий является самым распространенным из всех имеющихся многоранговых методов. Он не допускает ошибок, возникающих при оценке всех пар средних с помощью HCP . Этот критерий идентичен HCP при оценке разностей соседних средних ранжированного ряда, а при увеличении удаленности средних друг от друга порог существенной разности критерия Дункана прогрессивно возрастает. Использование этого критерия особенно эффективно, если в опыте изучаются независимые друг от друга варианты, например, для оценки всех возможных пар сравнений продуктивности нескольких сортов. В качестве примера рассмотрим уже известный опыт с ягнятами.

* Более подробно различные критерии для оценки любых парных сравнений рассмотрены в соответствующей литературе. См. S. G. Carmer, M. R. Swanson "An Evaluation of ten Pairwise Multiple Comparison Procedures by Monte Carlo Methodes" Journal of the American Statistical Association, 68, 66—74, 1973.

Основу этого метода составляет расчет ближайших существенных разностей (D) для всех относительных позиций средних по вариантам, расположенных в ранжированном ряду. Далее величины D используются по порядку для статистической оценки разностей между средними. В большинстве книг для расчета D приводится следующая формула: $D = Qs_y$, где Q — табличная величина [8], зависящая от избранного уровня значимости, числа степеней свободы для ошибки и относительной удаленности средних (количество мест) в ранжированном ряду, и s_y — стандартная ошибка средних, определяемая по формулам $\sqrt{MSE/r} = \sqrt{s^2/r}$. В данной книге $D = R$ (НСП), где R — табличное значение (табл. А.4 и табл. А.5, приложение) для заданного уровня значимости, числа степеней свободы для ошибки и положения места средних в ранжированном ряду; $НСП = t\sqrt{2s^2/r}$. В таблице А.4 и А.5 значения R получены путем преобразования величины Q с целью упрощения расчета D на основе НСП.

Воспользуемся данными эксперимента с ягнятами и сделаем следующие этапы расчетов.

1. Вычислим наименьшую существенную разность.

$$НСП_{0.05} = t \sqrt{\frac{2s^2}{r}} = 2,262 \sqrt{\frac{2(7,78)}{4}} = 4,46.$$

2. Вычислим D для относительных позиций средних в ранжированном ряду. Поскольку в опыте имелось четыре средних, они могут быть удалены друг от друга на 2, 3 или 4 места. Учтите, что для соседних средних принимается удаленность, равная 2 (2 места).

Относительная позиция в ранжированном ряду (по табл. А. 4 приложения)	2	3	4
Значения R для 5%-ного уровня (табл. А. 4)	1,00	1,04	1,07
$D = R$ (НСП)	4,5	4,6	4,8

3. Построим ранжированный ряд средних для проверки существенных разностей.

Варианты	FS_0	MS_0	FS	MS_3
Средние	53	57	59	63

Сначала сравним самую большую среднюю и наименьшую, используя D -оценку для их относительных друг к другу позиций в ранжированном ряду (в данном случае $r=4$, поэтому $D=4,8$). Если разность между этими средними равна или больше, чем величина D , то средние будут различаться существенно ($63-53=10$, $D=4,8$, поэтому 63 существенно больше, чем 53). Сравнив самую большую среднюю со второй средней ранжированного ряда ($63-57=6$, $D=4,6$), находим, что 63 существенно больше, чем 57. Далее при сравнении с третьей величиной средней ($63-59$, $D=4,5$) находим, что 63 несущественно больше, чем 59. После того как будет найдена несущественная разность, можно провести линию, связывающую (и выделяющую) эти средние. Затем повторите процесс, начав сравнение уже со второй по величине средней, среди всех средних, расположенных до нее.

Многоранговый критерий Дункана имеет одно особое правило. Оно гласит, что разницу между двумя средними нельзя считать существенной, если эти средние занимали промежуточное положение между двумя другими, несущественно различающимися средними ряда. Так, если среди пяти средних ранжированного ряда установлено несущественное отличие A от D , т. е. $ABCDE$, и существенное отличие B от E , то отпадает необходимость проверки отклонения B от D и C , поскольку они попали в интервал несущественной разницы между A и D . Следующим шагом будет сравнение C и E ; в случае несущественной разницы между ними средние C и E соединяют линией $ABCDE$ и дальнейшая проверка разностей теряет смысл. Таким образом, упомянутое правило позволяет не проводить оценку разностей между средними, уже соединенными линией.

4. Укажем статистическую существенность с помощью линий или букв.

MS_s	FS_s	MS_0	FS_0		MS_s	FS_s	MS_0	FS_0
				или				
63	59	57	53		63 a	59 ab	57 bc	53 c

Средние, соединенные одной линией или сопровождаемые общими буквами, различаются несущественно на 5%-ном уровне. В случае использования букв существенные разницы могут быть показаны и без ранжирования средних.

Обратите внимание, что в нашем примере сравнение средних с помощью многогранового критерия Дункана, или HCP ведет к одинаковому заключению о существенном повышении массы тела ягнят под действием гормона ($MS_s > MS_0$ и $FS_s > FS_0$) и о несущественной разнице изучаемого показателя в зависимости от пола ягнят как с применением гормона, так и без него ($MS_s \nrightarrow FS_s$ и $MS_0 \nrightarrow FS_0$).

ПЛАНИРУЕМЫЕ F-КРИТЕРИИ

При планировании эксперимента мы можем часто получить ответы на поставленные вопросы с помощью F -оценки. Такой подход вызывает необходимость расчленения числа степеней свободы и суммы квадратов для вариантов на соответствующие компоненты. Компонентами могут быть величины групповых сравнений или реакций отзывчивости. Эти показатели оцениваются путем расчленения степеней свободы и суммы квадратов для вариантов на самостоятельные составные части: отдельные (одиночные) степени свободы и соответствующие им суммы квадратов. Умело выбранные варианты могут дать ответ на столько независимых вопросов, сколько степеней свободы они имеют. Если проводимые сравнения являются независимыми, то их называют ортогональными; ортогональность — ценное свойство, поскольку такие сравнения ведут к получению четко выраженных обобщений.

Исследователи еще не оценили полностью весомость и простоту этого метода проверки существенных различий между средними.

Этот метод предусматривает выбор ортогональных коэффициентов и, вероятно, такое название (термин) создает впечатление о его сложности и трудоемкости. Однако это далеко от истины. В действительности этот метод имеет три важных достоинства: 1) позволяет дать ответы на специфические, важные вопросы о действии вариантов; 2) отличается простотой и 3) обеспечивает необходимую проверку правильности расчета суммы квадратов для вариантов.

Ортогональные коэффициенты. Построение таблицы коэффициентов сравнения обеспечивает контроль ортогональности и правильности расчета соответствующих сумм квадратов. Коэффициенты для сравнения реакций отзывчивости получают из таблиц ортогональных полиномов, например, таких, как таблица А.11 в приложении. Коэффициенты для межклассовых сравнений подготавливают, соблюдая следующие простые правила:

1) при сравнении двух групп одинакового размера коэффициент одной группы получает обозначение «+1», а коэффициент другой — «-1». При этом не имеет значения, какая из групп будет иметь положительный коэффициент;

2) при сравнении групп с разным количеством вариантов первая группа получает коэффициенты, равные числу вариантов второй группы, а вторая группа получит коэффициенты с обратным знаком, абсолютная величина которых будет равна числу вариантов первой группы. Так, если в опыте изучают пять вариантов и первые два нужно сравнить с тремя другими, то искомые коэффициенты составят следующие величины: +3, +3 и -2, -2, -2 соответственно;

3) по возможности сокращайте коэффициенты до наименьшего целого числа. Например, при сравнении двух вариантов одной группы с четырьмя из другой, в соответствии с требованием пункта 2, получим коэффициенты: +4, +4 и -2, -2, -2, -2, которые можно сократить в два раза, т. е. до: +2, +2 и -1, -1, -1, -1;

4) коэффициенты для взаимодействия факторов могут быть найдены путем перемножения соответствующих коэффициентов главных эффектов.

Для проверки независимости сравнений применяют два правила. Сравнения будут независимыми и, следовательно, ортогональными, если: 1) сумма коэффициентов для каждого сравнения равна нулю и 2) сумма произведений соответствующих коэффициентов двух любых сравнений равна нулю.

Для пояснения указанных правил возьмем практический пример построения таблицы ортогональных коэффициентов. Например, мы планируем полевой опыт по оценке эффективности трех способов внесения в почву фосфорных удобрений: поверхностный разбросной (В), ленточный с мелкой (S) и глубокой (D) заделкой. Любой способ внесения в почву удобрений будем оценивать на фоне двух доз фосфорного удобрения (P_1 и P_2). Также для выявления отзывчивости (эффекта) на фосфорные удобрения включим вариант без удобрений (NT). В «шапке» таблицы (табл. 6.1) поместим перечень вариантов. Коэффициенты будем вносить в таблицу по мере выполнения сравнений.

1. Есть ли ответная реакция на внесение в почву фосфора (P)?

Таблица 6.1. Коэффициенты для расчленения суммы квадратов шести вариантов на шесть независимых (ортогональных) сравнений *

Сравнения для выявления эффективности	Варианты						
	NT	P ₁ B	P ₁ S	P ₁ D	P ₂ B	P ₂ S	P ₂ D
1. Удобрения (P)	6	-1	-1	-1	-1	-1	-1
2. Дозы P ₁ и P ₂	0	1	1	1	-1	-1	-1
3. Способы внесения B, S и D	0	2	-1	-1	2	-1	-1
4. Глубина заделки S и D	0	0	1	-1	0	1	-1
5. Дозы и способы (P ₁ и P ₂) (B, S и D)	0	2	-1	-1	-2	1	1
6. Дозы и глубина (P ₁ и P ₂) (S и D)	0	0	1	-1	0	-1	1

* NT — без удобрений; P₁ и P₂, где 1 и 2 обозначают дозы фосфорного удобрения; B, S и D — способы внесения: разбросной и ленточный с мелкой и глубокой заделкой соответственно.

Это можно установить путем сравнения варианта без удобрений (NT) со всеми другими вариантами, имеющими фосфор. Поскольку таких других вариантов насчитывается шесть, то NT получает коэффициент 6, а другие —1, так как они сравниваются с одним вариантом. Сделав такое сравнение, мы не можем использовать вариант NT еще раз, поскольку будет нарушена ортогональность сравнений и поэтому NT получает коэффициент «0» для всех последующих сравнений.

2. Будет ли средний эффект дозы P₂ выше, чем дозы P₁, или нет? Для ответа на данный вопрос необходимо сравнить группу вариантов: P₁B+P₁S+P₁D с другой группой: P₂B+P₂S+P₂D. Поскольку мы имеем две группы одинакового размера, обозначим их +1 и —1 соответственно.

3. При использовании двух уровней внесения фосфорного удобрения в почву проявляется ли преимущество ленточного способа внесения удобрения по сравнению с разбросным? Для этого сравним P₁S+P₁D+P₂S+P₂D и P₁B+P₂B. Теперь мы сравниваем группу из четырех вариантов с группой из двух вариантов, поэтому варианты второй группы получат коэффициенты 4, а варианты первой группы —2. Сократив эти коэффициенты до наименьшего кратного числа, получим 2 и —1 соответственно.

4. Рассмотрим теперь способ только ленточного внесения. Различается ли эффективность мелкой и глубокой заделки удобрений? Имеем две группы P₁S+P₂S и P₁D+P₂D. Они получают коэффициенты 2 и —2; после сокращения 1 и —1 соответственно.

5. Различаются ли прибавки урожая при увеличении дозы фосфорного удобрения с P₁ до P₂ для разбросного и ленточного способов внесения? В этом случае имеем взаимодействие двух факторов, поэтому коэффициенты находят путем перемножения соответствующих коэффициентов для 2-го и 3-го сравнения: 0(0)=0, 1(2)=2, 1(—1)=—1, 1(—1)=—1, —1(2)=—2, —1(—1)=1, —1(—1)=1.

Таблица 6.2. Сравнения, варианты, суммы по вариантам и коэффициенты для расчленения суммы квадратов вариантов

Сравнения для выявления эффектов	Варианты и суммы по вариантам			
	FS_2 212	FS_3 236	MS_2 228	MS_3 252
1. Гормона (J)	+1	-1	+1	-1
2. Пола (S)	+1	+1	-1	-1
3. Взаимодействия (J×S)	+1	-1	-1	+1

6. И наконец, различаются ли прибавки урожая с увеличением дозы фосфорного удобрения с P_1 до P_2 при мелкой и глубокой заделке удобрения? При этом наблюдается взаимодействие 2-го и 4-го сравнений, и коэффициенты будут определены путем перемножения коэффициентов 2 и 4 этих сравнений (табл. 6.1).

Обратите внимание, что суммы коэффициентов по всем рядам таблицы 6.1 равны нулю и суммы произведений коэффициентов одних и тех же вариантов для любых двух сравнений дают нуль. К примеру, возьмем 1-е и 5-е сравнения: $6(0) + (-1)2 + (-1)(-1) + (-1)(-1) + (-1)(-2) + (-1)1 + (-1)1 = 0$. Таким образом, мы можем быть уверены, что сравнения ортогональны, и суммы квадратов отдельных сравнений дадут при их сложении сумму квадратов для шести вариантов.

В качестве простого примера использования коэффициентов групповых сравнений воспользуемся данными эксперимента по выяснению влияния имплантации гормона на развитие ягнят.

Групповые сравнения. Обратите внимание, что при выборе вариантов этого эксперимента требовалось получить ответ на три специфических вопроса: 1) влияет ли гормон на прирост массы тела ягнят в целом?, 2) имеются ли различия между баранчиками и ярочками в прибавках массы тела и 3) одинаков ли эффект гормона в зависимости от принадлежности к полу? Ответ на каждый вопрос включает единственную степень свободы. Коэффициенты для трех необходимых сравнений приводятся в таблице 6.2.

Для выявления эффекта гормона в целом мы сравниваем ягнят обоих полов при воздействии на них стилбестрола, а с другой стороны без применения этого гормона. Такое сравнение является правильным, поскольку варианты с гормоном и без него составили одинаковые группы баранчиков и ярочек.

Сравнивая прибавки массы тела ягнят в зависимости от пола, мы имеем с одной стороны среднюю прибавку массы тела у всех ярочек, а с другой стороны — у всех баранчиков для двух доз этого гормона. Это сравнение также является правильным, поскольку в опыте подвергались воздействию гормона одинаковые группы ягнят обоих полов.

Если бы применение гормона вызывало существенно большую прибавку массы тела ягнят одного пола по сравнению с другим, то можно было бы утверждать о наличии существенного взаимодействия между полом и гормоном. Коэффициенты подобного сравнения (JS или

Таблица 6.3. Ортогональное расчленение вариантов в эксперименте при изучении воздействия гормона на ягнят

Источники варьирования	df	SS	MS	F _{факт}	F _{теор}	
					5%	1%
Варианты	3	208	69,33	8,91	3,86	6,99
Гормон	1	144	144	18,51	5,12	10,56
Пол	1	64	64	8,23		
Взаимодействие (J × S)	1	0	0	0		
Ошибка	9	70	7,78			

гормон/пол) получают путем перемножения коэффициентов для каждого из двух вариантов в соответствии с данными первых двух строчек таблицы 6.2.

Для расчета сумм квадратов, средних квадратов и F -критериев проделаем нижеуказанные этапы и сведем результаты в таблицу 6.3.

Расчет сумм квадратов компонентов, составляющих варианты, проведем сначала с помощью корректирующего фактора, а затем продемонстрируем использование коэффициентов сравнения, представленных в таблице 6.2. Последний способ расчета сумм квадратов ограничен в том случае, когда эти суммы имеют единственную степень свободы.

Сумма квадратов в опыте при имплантации гормона.

$$SSJ = \frac{(212+228)^2 + (236+252)^2}{8} - \frac{(928)^2}{16} = 144.$$

Расчет суммы квадратов с использованием коэффициентов сравнения проводится по следующей формуле:

$$SS = \frac{(\sum c_i Y_{i.})^2}{r \sum c_i^2},$$

где c_i — коэффициенты сравнения из таблицы 6.2; $Y_{i.}$ — суммы по вариантам и r — число повторностей.

$$SSI = \frac{[1(212) - 1(236) + 1(228) - 1(252)]^2}{4[(1)^2 + (-1)^2 + (+1)^2 + (-1)^2]} = \frac{(-48)^2}{4(4)} = 144.$$

Сумма квадратов для пола:

$$SSS = \frac{(212+236)^2 + (228+252)^2}{8} - \frac{(928)^2}{16} = 64.$$

Или с помощью метода коэффициентов:

$$SSS = \frac{[1(212) + 1(236) - 1(228) - 1(252)]^2}{4(4)} = \frac{(-32)^2}{16} = 64.$$

Сумма квадратов для взаимодействия $I \times S$.

$$SS(I \times S) = SST - SSI - SSS = 208 - 144 - 64 = 0$$

или с коэффициентами:

$$SS(I \times S) = \frac{[1(212) - 1(236) - 1(228) + 1(252)]^2}{4(4)} = \frac{0^2}{16} = 0.$$

Сравните, насколько проще проводятся расчеты методом коэффициентов, чем в использовании корректирующего фактора. При расчете каждой суммы квадратов методом коэффициентов лишь одно число возводится в квадрат, тогда как в первом случае наряду со сложением квадратов двух больших чисел требуется использование корректирующего фактора. Обратите внимание, что сумма трех компонентов дает сумму квадратов для вариантов, вычисленную обычным путем (в ходе дисперсионного анализа), что дает возможность произвести проверку расчетов.

Поскольку каждая сумма квадратов компонентов имеет единственную степень свободы, то средние квадратов будут такими же, как и их суммы.

Ф-оценки находят путем деления величины каждого среднего квадрата изучаемых эффектов на средний квадрат ошибки (*MSE*). Таким образом, используя эти чувствительные *F*-критерии, мы узнали нечто большее, чем нам дала оценка с помощью *HSP* или применение много-рангового критерия Дункана. Теперь нам стало совершенно ясно, что баранчики развиваются лучше, чем ярочки.

Другие примеры расчленения суммы квадратов для вариантов на субкомпоненты и использования планируемых *F*-критериев будут рассмотрены в разделе «Выделение существенных средних» в главе 7.

Сравнения величин, характеризующих ответные реакции (отзывчивость). Исследователи часто желают изучить изменения на нескольких уровнях, например нарастающие дозы удобрений, сроки уборки, дозы фунгицида или гербицида. В этих случаях экспериментатора интересует характер ответной реакции экспериментальных единиц на меняющуюся дозу воздействующего варианта. Предназначение статистического анализа сводится к определению направления (тренда) или отзывчивости.

По возможности лучше использовать арифметическую серию градаций фактора. Равновеликие интервалы (шаги) между дозами варианта или сроками воздействия дают оценку реакции для всего размаха избранных градаций и обеспечивают лучшую основу построения соответствующей кривой, чем в случаях неодинаковых промежутков между соседними градациями вариантов. Кроме того, как вы увидите далее, эта равномерность дает большие преимущества при расчете сумм квадратов и подборе уравнений регрессии.

В качестве простого примера мы взяли данные эксперимента с обработкой семян лимской фасоли, представленные на рисунке 3.2. Обратите внимание на равновеликие градации дозы: 0, 10, 20 г инсектицида на 100 кг семян. Одной из задач эксперимента являлось установление характера реакции семян на внесение инсектицида на фоне фунгицида и без него. Часть исходных данных этого опыта приведена в таблице 6.4, а данные дисперсионного анализа в таблице 6.5. Прежде чем начать сравнение реакций отзывчивости (трендов), ознакомьтесь, каким обра-

Таблица 6.4. Действие фунгицида и инсектицида на всхожесть семян лимской фасоли (число проросших из 100 семян, посеянных на делянке). $F_0=0$ и $F_2=166$ г фунгицида на 100 кг семян. $J_0=0$, $J_1=10$ г и $J_2=20$ г инсектицида на 100 кг семян

Варианты	Блоки					$Y_{i.}$	$\bar{Y}_{i.}$
	I	II	III	IV	V		
$F_0 J_0$	55	69	71	78	68	341	68,2
$F_0 J_1$	65	47	55	64	59	290	58,0
$F_2 J_2$	47	37	58	48	54	244	48,8
$F_0 J_0$	91	76	92	92	95	446	89,2
$F_2 J_1$	85	93	97	88	96	459	91,8
$F_2 J_2$	84	94	94	96	92	460	92,0
$Y_{.j}$	427	416	467	466	464	2240 = $Y_{..}$	74,7 = $\bar{Y}_{..}$

зом в таблице 6.5 факториальные варианты расчленены на главные эффекты и взаимодействия. Вычисления суммы квадратов для вариантов и последующее расчленение проводят по формулам:

$$SST = \frac{341^2 + 290^2 + \dots + 460^2}{5} - C, \text{ где } C = \frac{2240^2}{6(5)} = 176\,222,8 - 167\,253,33 = 8969,47;$$

$$SSF = \frac{(341 + 290 + 244)^2 + (446 + 459 + 460)^2}{15} - C = 8003,33;$$

$$SSI = \frac{(341 + 446)^2 + (290 + 459)^2 + (244 + 460)^2}{10} - C = 345,27;$$

$$SS(F \times I) = SST - SSF - SSI = 8969,47 - 8003,33 - 345,27 = 620,87.$$

Наличие существенного взаимодействия (FJ) указывает, что реакция семян на инсектицид зависит также и от обработки их фунгицидом. Средние вариантов могут быть статистически проверены на наличие существенно различающихся линейных трендов всхожести по мере увеличения дозы инсектицида на фоне двух уровней фунгицида. По-

Таблица 6.5. Итоги дисперсионного анализа данных всхожести семян лимской фасоли

Источник варьирования	df	SS	MS	$F_{\text{факт}}$	$F_{\text{теор}}$	
					5%	1%
Общий	29	10 140,67				
Блоки	4	401,00	100,25			
Варианты	5	8 969,47	1793,89			
Фунгицид (F)	1	8 003,33	8003,33	207,82	4,35	8,10
Инсектицид (J)	2	345,27	172,64	4,48	3,49	5,85
Взаимодействие (FJ)	2	620,87	310,44	8,05		
Ошибка	20	770,20	38,51			

Таблица 6.6. Коэффициенты сравнения для определения зависимости всхожести семян лимской фасоли от уровней дозы инсектицида

Сравнение факторов воздействия	Варианты и суммы по вариантам					
	F_0J_0 841	F_0J_1 290	F_0J_2 244	F_1J_0 446	F_1J_1 459	F_1J_2 460
Фунгицида (F)	-1	-1	-1	1	1	1
Инсектицида (линей- ное) (JL)	1	0	-1	1	0	-1
Инсектицида (нелиней- ное) (JNL)	1	-2	1	1	-2	1
Взаимодействия (FJL — линейное)	-1	0	1	1	0	-1
Взаимодействие (FJNL — нелиней- ное)	-1	2	-1	1	-2	1

сколько для взаимодействия FJ имеется две степени свободы, то мы можем поставить два независимых вопроса. Этими вопросами будут: есть ли существенное различие в линейной отзывчивости на инсектицид двух уровней фунгицида (F_0 и F_1) и есть ли существенная разница между некоторыми нелинейными реакциями семян на их обработку препаратами? Для упрощения расчетов сумм квадратов подготовим таблицу коэффициентов сравнения (табл. 6.6).

Сравнение по выявлению эффекта фунгицида аналогично расчету, приведенному в таблице 6.5, и заключается в сопоставлении средней всех делянок с F_0 и средней всех делянок с F_1 . Это простой пример сравнения классов (групп), и поскольку обе сравниваемые группы одинакового размера, то коэффициент -1 получают члены первой группы, а коэффициент $+1$ — второй. Коэффициенты для эффектов: «инсектицид (линейное)» и «инсектицид (нелинейное)» берут из таблицы А.11 (приложение) под $n=3$ (три градации дозы инсектицида). Необходимо указать, что данные этой таблицы можно использовать лишь в случае равновеликих интервалов (одинакового шага) между дозами

Таблица 6.7. Средние квадраты и F -критерии для оценки существенности отзывчивости семян на обработку инсектицидом

Источники вариации	df	MS	$F_{\text{факт}}$	$F_{\text{теор}}$	
				5%	1%
Фунгицид	1	8003,33	207,8	4,35	8,10
Инсектицид (линейное)	1	344,45	8,9		
Инсектицид (нелинейное)	1	0,82	0,02		
Взаимодействие:					
(FJL)	1	616,05	16,0		
(FJNL)	1	4,82	0,1		
Ошибки	20	38,51			

вариантов. Коэффициенты для взаимодействий FJL и $FJNL$ получают путем умножения коэффициентов для фунгицида на коэффициенты для инсектицида (линейное или нелинейное).

Каждое сравнение таблиц 6.6 включает единственную степень свободы, так что мы вычисляем суммы квадратов по формуле:

$$SS = (\sum c_i Y_{i.})^2 / (r \sum c_i^2), \text{ поэтому}$$

$$SSF = \frac{(-341 - 290 - 244 + 446 + 459 + 460)^2}{5(6)} = 8003,33;$$

$$SS(IL) = \frac{(341 - 244 + 446 - 460)^2}{5(4)} = 344,45;$$

$$SS(INL) = \frac{[341 - 2(290) + 244 + 446 - 2(459) + 460]^2}{5(12)} = 0,82;$$

$$SS(FIL) = \frac{(-341 + 244 + 446 - 460)^2}{5(4)} = 616,05;$$

$$SS(FINL) = \frac{[-341 + 2(290) - 244 + 446 - 2(459) + 460]^2}{5(12)} = 4,82.$$

Все средние квадраты будут равны их суммам, поскольку каждая сумма имеет единственную степень свободы и фактические значения F вычисляют путем деления средних квадратов соответствующих компонентов на MSE (средний квадрат ошибки), таблица 6.7.

Отметьте, что суммы квадратов для пяти компонентов вариантов таблицы 6.7 дают при их сложении сумму квадратов для вариантов таблицы 6.5. Это служит дополнительной проверкой наших арифметических расчетов; поскольку компоненты составляют ортогональную группу, то их общая сумма квадратов должна быть равна сумме рассчитанных квадратов.

Данные величины таблицы 6.7 показывают очень существенное взаимодействие инсектицида и фунгицида для линейной реакции и взаимодействие всхожести семян и отсутствие эффекта взаимодействия для реакции, отличной от линейной. Таким образом, результаты эксперимента можно представить в виде четкого и соответствующего заключения; так на рисунке 6.1 с помощью линий линейной регрессии определяется эффективность увеличения дозы инсектицида на фоне обработки семян фунгицидом и без обработки.

Расчет линий с анализом регрессии рекомендуется провести самостоятельно, после того как вы ознакомитесь с ним в 13-й и 14-й главах.

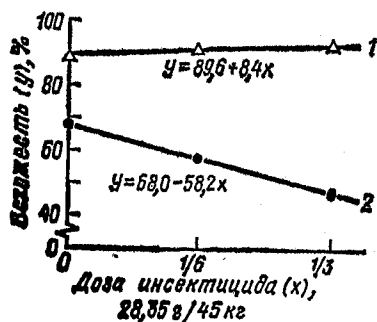


Рис. 6.1. Действие обработки семян инсектицидом на фоне фунгицида и без него на всхожесть лимской фасоли. Семена, обработанные фунгицидом (1) и необработанные (2).

ЗАКЛЮЧЕНИЕ

Решение вопроса о том, средние каких вариантов существенно различаются друг от друга, составляет сущность приема, называемого выделением существенных средних. Это достигается тремя способами: ис-

пользованием наименьших существенных разностей, многогранговых критериев или планируемых F -критериев.

Наименьшая существенная разность определяется по формуле

$$HCP = t \sqrt{\frac{2(MSE)}{r}},$$

где t — табличная величина для числа степеней свободы ошибки и избранного уровня значимости, MSE — средний квадрат ошибки и r — число дат, на которых базируется выбор оцениваемых средних.

Для оценки средних, имеющих в своей основе разное количество дат, формула приобретает вид:

$$HCP = t \sqrt{\frac{MSE}{r_1} + \frac{MSE}{r_2}}.$$

Критерий Дункана, самый распространенный из всех известных многогранговых критериев, рассчитывается по формуле $D = R(HCP)$, где R — табличная величина для степеней свободы ошибки, определенного уровня значимости и расстояния между двумя средними ранжированного ряда средних; HCP — наименьшая существенная разность.

Процедура выделения существенных средних с помощью планируемых F -критериев является наиболее значительной. При этом способе оценки средних можно поставить столько независимых вопросов и получить на них ответы, сколько степеней свободы имеют варианты. Вопросы следует планировать до постановки эксперимента. В случае единственного числа степеней свободы сумма квадратов может быть вычислена с помощью группы коэффициентов, сумма которых дает нуль, по следующей формуле:

$$SS = \frac{(\sum c_i Y_{i.})^2}{r \sum c_i^2},$$

где c_i — коэффициенты, $Y_{i.}$ — суммы по вариантам и r — число дат, составляющих каждую сумму. Два сравнения будут ортогональными (независимыми), если их коэффициенты и сложение произведений соответствующих коэффициентов дают нуль. Если количество ортогональных сравнений равно числу степеней свободы для вариантов, то их суммы квадратов будут давать при сложении сумму квадратов для вариантов.

Коэффициенты для измерения трендов берут из таблицы А.11 (приложение) в случае одинаковых интервалов между градациями вариантов и из таблицы А.11а в случае неравновеликих интервалов.

СХЕМА ЛАТИНСКОГО КВАДРАТА

В этой схеме рендомизация дополнительно ограничивается группировкой вариантов не только по рядам, но и столбцам. Такое расположение вариантов позволяет устранить вариабельность, связанную с эффектом столбцов и рядов, снизить экспериментальную ошибку. Каждый вариант встречается одинаковое число раз (обычно однажды) в каждом ряду и столбце. Эта схема будет обеспечивать большую точность сравнения вариантов, чем схема рендомизированных блоков, лишь в случае значительной вариации по столбцам.

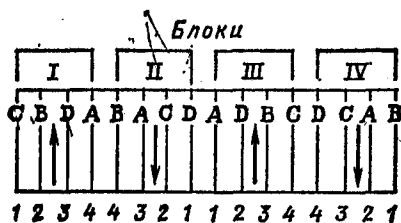
Ряды и столбцы могут выражать пространственное размещение экспериментальных единиц или порядок группировки вариантов. На рисунке 7.1 даны варианты: А, В и С — три марки исследуемых счетных машин, столбцы — три оператора и ряды — шесть различных сроков испытания этих машин. Каждый оператор испытывает машину два раза, кроме того, все три машины испытывают в каждый период времени испытаний. Таким образом, влияние сроков испытаний и операторов на результаты эксперимента являются контролируруемыми источниками вариации, независимыми от машин, и они могут быть устранены из общей вариабельности в опыте, что снижает экспериментальную ошибку. Данная схема является примером двойного латинского квадрата. Когда в опыте мало вариантов и ожидается значительный эффект рядов и столбцов, эта вариация может быть устранена по двум направлениям с использованием схемы двух латинских квадратов (при независимой рендомизации).

Латинский квадрат может иметь преимущество и в тех случаях, когда делянки расположены на одной линии (одноярусное расположе-

Ряды (сроки испытаний)	Столбцы (операторы)		
	I	II	III
I	В	А	С
II	С	В	А
III	А	С	В
IV	В	С	А
V	С	А	В
VI	А	В	С

Рис. 7.1. Три варианта размещены по схеме двойного латинского квадрата. Степени свободы по соответствующим источникам вариации: для рядов — 5, для столбцов — 2, для вариантов — 2 и для ошибки — 8. Варианты (А, В, С) представлены тремя марками счетных машин.

Рис. 7.2. Схема латинского квадрата для размещения четырех вариантов предпосевной обработки семян (*A*, *B*, *C* и *D*) по четырем высевальным аппаратам (1, 2, 3 и 4). Стрелки указывают направление движения сеялки. Степени свободы распределены по источникам вариации следующим образом: блоки — 3, высевальные аппараты — 3, варианты семян — 3, ошибка — 6.



ние). Рассмотрим, к примеру, эксперимент по оценке четырех вариантов предпосевной обработки семян, где индивидуальными делянками являются отдельные рядки, расположенные в один ярус. Используется сеялка с четырьмя семявысевающими аппаратами. Высевальные аппараты могут различаться по скорости высева семян. Для устранения влияния этого эффекта на результаты опыта каждый вариант высева может быть произведен одним высевальным аппаратом в каждом из четырех блоков, так что в опыте все семена будут распределены по всем высевальным аппаратам одинаковое число раз, как показано на рисунке 7.2.

Повторность вариантов в схеме латинского квадрата должна по меньшей мере быть равна их числу, поэтому использование этой схемы для опытов с большим количеством вариантов будет непрактично. По такой схеме обычно закладывают опыты, имеющие от четырех до восьми вариантов, причем на один вариант в каждом ряду и столбце приходится одна экспериментальная единица.

РЕНДОМИЗАЦИЯ

Возьмем любую схему латинского квадрата (систематическую или рендомизированную) с количеством вариантов, необходимых для вашего эксперимента. Например, мы хотим случайно разместить шесть вариантов: *A*, *B*, *C*, *D*, *E* и *F*. Начнем с латинского квадрата, представленного на рисунке 7.3; затем используем данные таблицы случайных чисел (табл. А.1, приложение); произвольно выберем, например, 5-й ряд;

Ряды	Столбцы					
	4	2	5	1	3	6
1	<i>B</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>A</i>	<i>C</i>
3	<i>C</i>	<i>E</i>	<i>A</i>	<i>D</i>	<i>F</i>	<i>B</i>
5	<i>A</i>	<i>F</i>	<i>C</i>	<i>B</i>	<i>E</i>	<i>D</i>
4	<i>D</i>	<i>A</i>	<i>F</i>	<i>C</i>	<i>B</i>	<i>E</i>
2	<i>F</i>	<i>B</i>	<i>D</i>	<i>E</i>	<i>C</i>	<i>A</i>
6	<i>E</i>	<i>C</i>	<i>B</i>	<i>A</i>	<i>D</i>	<i>F</i>

Рис. 7.3. Подготовка схемы латинского квадрата 6×6 к повторной рендомизации. Рендомизация вариантов по рядам и столбцам должна быть выполнена в соответствии с последовательностью, полученной из таблицы случайных чисел. Результаты отображены на рисунке 7.4.

Ряды	Столбцы						Суммы по рядам $Y_i...$
	I	II	III	IV	V	VI	
I	F 28,2	D 29,1	A 32,1	B 33,1	E 31,1	C 32,4	186,0
II	E 31,0	B 29,5	C 29,4	F 24,8	D 33,0	A 30,6	178,2
III	D 30,6	E 28,8	F 21,7	C 30,8	A 31,9	B 30,1	173,9
IV	C 33,1	A 30,4	B 28,8	D 31,4	F 26,7	E 31,9	182,3
V	B 29,9	F 25,8	E 30,3	A 30,3	C 33,5	D 32,3	182,1
VI	A 30,8	C 29,7	D 27,4	E 29,1	B 30,7	F 21,4	169,1
Суммы по рядам Y_f	183,8	173,3	169,7	179,5	186,9	178,7	1071,7=Y
Варианты							
	A(1) (NH ₄) ₂ SO ₄	B(2) NH ₄ NO ₃	G(3) CO(NH ₂) ₂	D(4) Ca(NO ₃) ₂	E(5) NaNO ₃	F(6) No N	
Суммы $Y_{...k}$	186,1	182,1	188,9	183,8	182,2	148,6	
Средние $\bar{Y}_{...k}$	31,0	30,4	31,5	30,6	30,4	24,8	

Рис. 7.4. Схема латинского квадрата 6×6. Каждый вариант встречается один раз в ряду и в столбце. Вариантами являются пять видов азотного удобрения из расчета 45 кг/га и контроль. Датыми является урожайность сахарной свеклы (корнеплоды), т/0,4 га.

пойдем вдоль этого ряда до конца, повернем назад по 6-му ряду; в результате ряды схемы опыта с 1-го по 6-й получают соответствующие номера: 1, 3, 5, 4, 2, 6. Продолжим движение по 6-му ряду таблицы случайных чисел и назад по 7-му ряду, получим соответствующие номера для столбцов с 1-го по 6-й: 4, 2, 5, 1, 3, 6. Размещение вариантов в новой схеме латинского квадрата выполняется путем перестройки рядов и столбцов старой схемы согласно полученным случайным числам, рисунок 7.4.

ДИСПЕРСИОННЫЙ АНАЛИЗ

Проанализируем данные рисунка 7.4, где даты можно классифицировать по трем группам, т. е. по рядам, столбцам и вариантам. Ряды от 1 до r имеют предстрочный индекс i , столбцы от 1 до c — j и варианты от 1 до n — k . В обычной схеме латинского квадрата $r=c=k$. Начнем с заполнения первых двух колонок таблицы 7.1.

Источники вариации и степени свободы. Число степеней свободы, как обычно, на единицу меньше числа наблюдений для каждого источника вариации: $df_{\text{общее}} = rc - 1 = 6(6) - 1 = 35$; $df_{\text{рядов}} = r - 1 = 6 - 1 = 5$; $df_{\text{столбцов}} = c - 1 = 6 - 1 = 5$; $df_{\text{вариантов}} = n - 1 = 6 - 1 = 5$. Степени сво-

Таблица 7.1. Итоги дисперсионного анализа данных опыта по изучению действия азотного удобрения на урожайность сахарной свеклы

Источники варьирования	Степени свободы <i>df</i>	Сумма квадратов <i>SS</i>	Средние квадраты <i>MS</i>	<i>F</i> _{факт}	<i>F</i> _{теор}	
					5%	1%
Общее	35	281,88				
Ряды (<i>R</i>)	5	32,19	6,438	4,26	2,71	4,10
Столбцы (<i>C</i>)	5	33,67	6,734	4,45		
Варианты (<i>T</i>)	5	185,77	37,154	24,56		
Ошибка (<i>RC-T</i>)	20	30,25	1,513			

боды для ошибки находят путем вычитания: $35 - 5 - 5 - 5 = 20$, или по формуле $(r-1)(c-1) - (n-1) = 5(5) - 5 = 20$.

Корректирующий фактор,

$$C = \frac{Y^2 \dots}{rc} = \frac{1071,7^2}{6(6)} = 31\,903,91.$$

Суммы квадратов и средние квадраты,

Рядов

$$SSR = \frac{\sum Y_{i..}^2}{c} - C = \frac{186,0^2 + \dots + 169,1^2}{6} - 31\,903,91 = 32,19,$$

где *c* — количество делянок в каждом ряду.

$$MSR = \frac{SSR}{df(R)} = \frac{32,19}{5} = 6,438.$$

Столбцов

$$SSC = \frac{\sum Y_{.j.}^2}{r} - C,$$

где *r* — количество делянок в каждом столбце.

$$SSC = \frac{183,6^2 + \dots + 178,8^2}{6} - 31\,903,91 = 33,67;$$

$$MSC = \frac{SSC}{df(C)} = \frac{33,67}{5} = 6,734.$$

Вариантов

$$SST = \frac{\sum Y_{..k}^2}{r} - C,$$

где *r* — повторность каждого варианта.

$$SST = \frac{148,6^2 + \dots + 182,2^2}{6} - 31\,903,91 = 185,77;$$

$$MST = \frac{SST}{df(T)} = \frac{185,77}{5} = 37,154.$$

Общая

$$SS = \sum y_{ijk}^2 - C = 28,2^2 + 32,1^2 + \dots + 27,4^2 + 29,1^2 - 31\,903,91 = 32\,185,79 - 31\,903,91 = 281,88.$$

Ошибки

$$SSE = SS - SSR - SSC - SST = 281,88 - 32,19 - 33,67 - 185,77 = 30,25$$

$$MSE = \frac{SSE}{df(E)} = \frac{30,25}{20} = 1,513.$$

Калькуляторы с программой для расчета стандартного отклонения. Суммы квадратов и средние квадраты можно также вычислить на основе стандартных отклонений для сумм. Например, для расчета SSR вводим в машину сумму каждого ряда (186,0...169,1) с помощью соответствующей клавиши вашего калькулятора и находим стандартное отклонение для сумм по рядам — 6,21496. Возводим стандартное отклонение в квадрат и делим на число дат, вошедших в каждую сумму: $(6,21496)^2/6 = MSR = 6,4376$. Умножаем на число степеней свободы для рядов, получаем $SSR = 6,4376(5) = 31,19$, как и ранее.

F-величины (фактические).

$$F_{\text{рядов}} = \frac{MSR}{MSE} = \frac{6,438}{1,513} = 4,26;$$

$$F_{\text{столбцов}} = \frac{MSC}{MSE} = \frac{6,734}{1,513} = 4,45;$$

$$F_{\text{вариантов}} = \frac{MST}{MSE} = \frac{37,154}{1,513} = 24,56.$$

Все три F -критерия базируются на 5 (для числителя) и 20 (для знаменателя) степенях свободы. Необходимая величина F для утверждения о статистической существенности берется из приложения (табл. А.3) и вводится в таблицу дисперсионного анализа. Все три источника вариации классифицируются как очень существенные. На основе этого мы делаем заключение о наличии фактических различий как между рядами и столбцами, так и между вариантами.

ВЫДЕЛЕНИЕ СУЩЕСТВЕННЫХ СРЕДНИХ

Планируя эксперимент по оценке действия различных азотных удобрений на урожайность сахарной свеклы, исследователь поставил несколько вопросов, ответы на которые должны быть получены в результате расчленения суммы квадратов для вариантов на ортогональную группу сравнений, как показано в таблице 7.2. Коэффициенты для контроля ортогональности сравнений и для завершения таблицы 7.2 даны в таблице 7.3.

Обратите внимание, что в таблице 7.3 суммы по рядам и суммы произведений соответствующих коэффициентов любых двух сравнений дают нуль и поэтому сравнения вариантов ортогональны. Коэффициен-

Таблица 7.2. Ортогональное расчленение вариантов схемы эксперимента, изображенного на рисунке 7.4

Источники вариации (сравнения)	df	SS	MS	F _{факт}	F _{теор}	
					5%	1%
Варианты	5	185,77	37,154	24,56	2,71	4,10
Без N и с N	1	180,200	180,200	119,10	4,35	8,10
Органический N и неорганический N	1	3,816	3,816	2,52		
Аммиачный N и нитратный N	1	0,202	0,202	0,13		
(NH ₄) ₂ SO ₄ и NH ₄ NO ₃	1	1,334	1,334	0,88		
NaNO ₃ и Ca(NO ₃) ₂	1	0,213	0,213	0,14		
Ошибка	20	30,25	1,513			

ты для проверки ортогональности этих сравнений и для построения таблицы 7.2 приведены в таблице 7.3.

Суммы квадратов для сравнений могут быть вычислены на основе сумм по вариантам следующим образом:

$$SS(\text{без } N \text{ и с } N) = \frac{148,6^2}{6} + \frac{(186,1 + \dots + 182,2)^2}{30} - \frac{1071,7^2}{36} = 3680,327 + 28\,403,787 - 31\,903,914 = 180,200.$$

Если сравнение опирается на одну степень свободы, то кратчайшим путем расчета будет метод с использованием ортогональных полиномов таблицы 7.3: $SS = (\sum c_j Y_{..k})^2 / (r \sum c_i^2)$, где c_i — коэффициенты таблицы 7.3; $Y_{..k}$ — суммы по вариантам и r — число дат, вошедших в суммы по вариантам. Таким образом,

$$SS(\text{без } N \text{ и с } N) = \frac{[5(148,6) - 186,1 - 182,1 - 188,9 - 183,8 - 182,2]^2}{6(30)} = \frac{-180,1^2}{180} = 180,200.$$

Знаменатель 6(30) определен в результате сложения квадратов коэффициентов, стоящих перед отдельными выражениями числителя,

Таблица 7.3. Коэффициенты вариантов для контроля ортогональности сравнений и облегчения расчетов сумм квадратов

Сравнения	Варианты и суммы по вариантам					
	Без N 148,6	(NH ₄) ₂ SO ₄ 186,1	NH ₄ NO ₃ 182,1	CO(NH ₂) ₂ 188,9	Ca(NO ₃) ₂ 183,8	NaNO ₃ 182,2
Без N и с N	+5	-1	-1	-1	-1	-1
Органический N и неорганический N	0	-1	-1	+4	-1	-1
NH ₄ -N и NO ₃ -N	0	+1	+1	0	-1	-1
(NH ₄) ₂ SO ₄ и NH ₄ NO ₃	0	+1	-1	0	0	0
Ca(NO ₃) ₂ и NaNO ₃	0	0	0	0	+1	-1

и умножения суммы на число дат, составляющих каждое выражение (член) числителя, т. е.

$$6(5)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 = 6(30).$$

$$SS(\text{органический } N \text{ и неорганический } N) = \frac{188,9^2}{6} + \\ + \frac{(186,1 + 182,1 + 183,8 + 182,2)^2}{24} - \frac{(188,9 + 186,1 + 182,1 + 183,8 + 182,2)^2}{30} = \\ = \frac{188,9^2}{6} + \frac{734,2^2}{24} - \frac{923,1^2}{30}.$$

Заметьте, что третье выражение является новым корректирующим фактором: $= 5947,202 + 22\,460,402 - 28\,403,787 = 3,816$.

При кратком методе расчета с использованием коэффициентов получаем:

$$= \frac{[4(188,9) - 186,1 - 182,1 - 183,8 - 182,2]^2}{6(20)} = 3,186;$$

$$SS(NH_4-N \text{ и } NO_3-N) = \frac{(186,1 - 182,1)^2 + (183,8 + 182,2)^2}{12} - \\ - \frac{(186,1 + 182,1 + 183,8 + 182,2)^2}{24} = \frac{368,2^2 + 366,0^2}{12} - \frac{734,2^2}{24} = \\ = 22\,460,603 - 22\,460,402 = 0,201 \\ \text{или} \quad = \frac{(186,1 + 182,1 - 183,8 - 182,2)^2}{6(4)} = \frac{(2,2)^2}{24} = 0,202.$$

$$SS[(NH_4)_2SO_4 \text{ и } NH_4NO_3] = \frac{186,1^2 + 182,1^2}{6} - \frac{(186,1 + 182,1)^2}{12} = \\ = 11\,298,937 - 11\,297,603 = 1,334 \\ \text{или} \quad = \frac{(186,1 - 182,1)^2}{2(6)} = \frac{4,0^2}{12} = 1,333.$$

$$SS[Ca(NO_3)_2 \text{ и } NaNO_3] = \frac{183,8^2 + 182,2^2}{6} - \frac{(183,8 + 182,2)^2}{12} = \\ = 11\,163,213 - 11\,163,000 = 0,213 \\ \text{или} \quad = \frac{(183,8 - 182,2)^2}{2(6)} = \frac{1,6^2}{12} = 0,213.$$

Средние квадраты (MS) получают путем деления сумм квадратов на их степени свободы; поскольку в данном случае каждое сравнение опирается на единственную степень свободы, то $SS=MS$.

Фактическое значение F -критерия получают путем деления каждого MS на MS ошибки. Теоретическое значение F берут из приложения (табл. А.3) для 1-й и 20-ти степеней свободы. Теперь мы имеем F -критерий для ответа на каждый из вопросов, поставленных при планировании эксперимента. Единственно существенное значение F -критерия состоит в применении его для сравнения варианта без азота со всеми остальными; все другие разности незначительны, что позволяет заключить о наличии эффекта азота для сахарной свеклы, тогда как ее отзывчивость на различные виды азотного удобрения была одинаковой.

ЗАКЛЮЧЕНИЕ

В схеме латинского квадрата экспериментальные единицы, кроме вариантов, сгруппированы еще по двум другим категориям. Эти две категории обычно обозначают как ряды и столбцы, согласно оформлению данных в таблице с двумя входами.

Каждый вариант встречается одинаковое число раз (обычно однажды) внутри каждой категории, так что различия между категориями не могут быть обусловлены действием вариантов. Повторность должна по меньшей мере равняться числу вариантов. Схема латинского квадрата становится непрактичной для опытов с числом вариантов, превышающим восемь.

Только в случае значительной вариации по обеим категориям (рядам и столбцам) схема латинского квадрата позволит вскрыть различия по вариантам, неуловленные схемой рендомизированных блоков.

СХЕМА РАСЩЕПЛЕННЫХ ДЕЛЯНОК — СПЛИТ-ПЛОТ (двухфакторный опыт)

Схема расщепленных делянок или ее разновидность — схема расщепленных блоков часто используется для факториальных экспериментов, в которых природа экспериментального материала или операции, связанные с его проведением, затрудняют возможность одинакового воздействия на все сочетания факторов или в тех случаях, когда исследователь желает увеличить точность оценки одних эффектов, жертвуя точностью оценки других. Основу схемы расщепленных делянок составляет размещение вариантов одного фактора по главным делянкам (делянкам 1-го порядка), расположенным, в свою очередь, по схеме полной рендомизации, рендомизированных блоков или латинского квадрата. Варианты другого фактора размещаются по субделянкам (делянки 2-го порядка) каждой главной делянки. Схема обычно жертвует точностью оценки средних эффектов вариантов главных делянок. Часто она повышает точность сравнения средних эффектов вариантов, расположенных по субделянкам, и при наличии взаимодействия сравнения эффектов вариантов субделянок одной главной делянки. Это связано с тем, что экспериментальная ошибка для главных делянок обычно больше, чем экспериментальная ошибка, используемая для сравнения вариантов субделянок. Обычно величина ошибки для вариантов субделянок меньше, чем ошибка, получаемая в случае размещения всех сочетаний вариантов по схеме рендомизированных блоков.

Ознакомьтесь с экспериментом, представленным на рисунке 8.1. Он включает два фактора: азотное удобрение (N) в двух градациях ($n=2$) и зеленое удобрение (G) четырех видов ($g=4$). Общее количество вариантов этого опыта составит $n \times g = 8$. Обратите внимание, что все восемь вариантов встречаются однажды в каждом из трех блоков, а внутри одного блока мы находим все варианты принятых уровней азота вместе. При использовании разных источников азота необходимо применять схему рендомизированных блоков с двумя вариантами в трех блоках. Степени свободы для этих шести главных делянок распределяются аналогично схеме рендомизированных блоков (колонка «расщепленных делянок» таблицы 8.1), где дано сравнение этих двух схем.

Ограничение рендомизации при размещении вариантов внутри блока дает в результате две величины ошибок для схемы расщепленных делянок. Ошибка для главных делянок больше, поскольку она вклю-

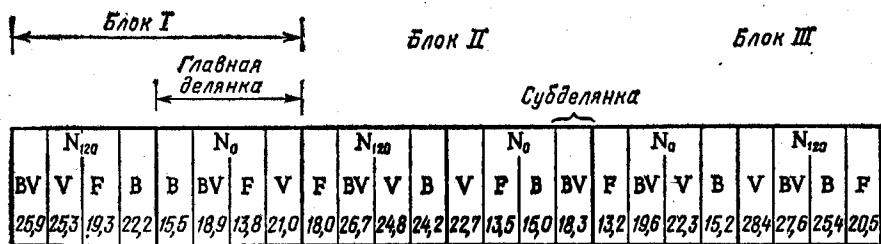


Рис. 8.1. Схема расщепленных делянок (сплит-плот). Главные делянки (N_{120} и N_0) представляют уровни азотного удобрения в фунтах/акр (135 кг/га). Субделянки BV, V, F и B — виды зеленого удобрения. Все делянки расположены полосами в один ярус по трем блокам. Урожайность корнеплодов сахарной свеклы дана по вариантам зеленого удобрения в тоннах на 0,4 га.

чает вариабельность между большими по размеру и удлиненными делянками, а ошибка для субделянок обычно меньше, поскольку она включает вариабельность между близко расположенными субделянками внутри главных делянок.

Схема расщепленных делянок (сплит-плот) может быть использована и для большего, чем два, числа факторов, поскольку нет необходимости в дополнительном расщеплении делянок для каждого нового фактора. Например, в опыте по оценке двух сортов на фоне двух уровней азотного удобрения, вносимого в два срока, на главных делянках можно разместить четыре сочетания как сортов, так и удобрений, причем эти делянки могут быть расщеплены на субделянки для оценки влияния сроков внесения азота. Одно повторение этого опыта показано на рисунке 8.2. Для трех факторов (A, B, C) при двух или большем числе

Таблица 8.1. План дисперсионного анализа двухфакторного опыта (рис. 8.1) по схеме расщепленных делянок и схеме рандомизированных блоков. (Квадратные скобки и стрелки указывают соответствующие величины ошибки для оценки действия вариантов в двух схемах)

Источники вариации	Степени свободы	Схема	
		расщепленных делянок	рандомизированных блоков
Субделянки (SP)	$nbg - 1$	23	23
Главные делянки (MP)	$nb - 1$	5	—
Блоки (B)	$b - 1$	2	2
Азотное удобрение (N)	$n - 1$	1 ←	1 ←
Ошибка (MP)	$(b - 1)(n - 1)$	2	—
Зеленые удобрения (G)	$(g - 1)$	3 ←	3 ←
Взаимодействие (N × G)	$(n - 1)(g - 1)$	3 ←	3 ←
Ошибка (SP)	$(b - 1)[(g - 1) + (n - 1)(g - 1)]$	12	—
Ошибка (RCB) *	$(b - 1)[(n - 1) + (g - 1) + (n - 1) \times (g - 1)]$	—	14

* Рандомизированные блоки.

$V_1 N_1$	$V_1 N_2$	$V_1 N_1$	$V_2 N_1$	Блок I
T_1	T_2	T_2	T_1	

Рис. 8.2. Одно повторение (блок) схемы расщепленных делянок. Главные делянки ($V_1 N_2$, $V_2 N_2$, $V_1 N_1$ и $V_2 N_1$) представляют сочетания двух сортов и уровней азотного удобрения. Субделянки (T_1 , T_2) — сроки внесения азотных удобрений.

градаций каждого из них имеется шесть различных возможностей для выбора главных делянок: A, B, C, AB, AC, BC.

Всякая модификация схемы расщепленных делянок вносит дополнительные значения величины экспериментальной ошибки, кото-

рые следует использовать для оценки тех или иных эффектов вариантов. Поэтому важно расположить факторы таким образом, чтобы получить самую высокую точность оценки взаимодействия и средних эффектов вариантов, представляющих для экспериментатора наибольший интерес. Чтобы заложить эксперимент по схеме расщепленных делянок, требуется определенное мастерство и поэтому обязательно необходимо проконсультироваться у специалиста, имеющего опыт применения этой схемы.

РЕНДОМИЗАЦИЯ

Рендомизация вариантов, размещаемых на главных делянках, проводится в соответствии с требованиями схемы, избранной для вариантов главных делянок. После этого производится случайное размещение вариантов субделянок внутри каждой главной делянки, причем рендомизация проводится отдельно для каждой главной делянки.

ДИСПЕРСИОННЫЙ АНАЛИЗ

Для иллюстрации расчетов мы воспользуемся данными эксперимента, представленного на рисунке 8.1. Опыт проводился с целью выявления действия трех видов зеленого удобрения при использовании двух видов азотного удобрения на продуктивность сахарной свеклы. При закладке опыта предполагалось (рабочая гипотеза), что сахарная свекла будет по-разному реагировать на зеленое удобрение в зависимости от уровня обеспеченности азотом, и в этой связи ставилась задача с максимальной точностью сравнить влияние зеленых удобрений на фоне применения разных уровней азотного удобрения.

Поэтому для оценки двух уровней азотного удобрения, вносимого в период прорезывания сахарной свеклы, были использованы главные делянки при их трехкратной повторности в соответствии со схемой рендомизированных блоков. На субделянках до посева сахарной свеклы размещали разные культуры на зеленое удобрение, которые выращивались в течение осени и зимы. Культурами на зеленое удобрение были: ячмень (B), вика (V), смешанный посев ячменя с викой (BV) и пар (F). На участке под паром (контроль) ничего не высевалось вплоть до посева сахарной свеклы. Делянки были расположены по схеме, представленной на рисунке 8.1. Данные по урожайности са-

Таблица 8.2. Исходная таблица урожайности корнеплодов сахарной свеклы для дисперсионного анализа, т/0,4 га *

Варианты		Блоки (j)			Суммы	Средние
кг N/га, (i)	зеленое удобрение (k)	I	II	III		
N ₀	Пар (F — 0)	13,8	13,5	13,2	40,5	13,5
	Ячмень (B)	15,5	15,0	15,2	45,7	15,2
	Вика (V)	21,0	22,7	22,3	66,0	22,0
	Ячмень+вика (BV)	18,9	18,3	19,8	56,8	18,9
Суммы главных деленок (Y _{1j})		69,2	69,5	70,3	209,0 = Y _{1..}	17,4
N ₁₃₅	Пар (F)	19,3	18,0	20,5	57,8	19,3
	Ячмень (B)	22,2	24,2	25,4	71,8	23,9
	Вика (V)	25,3	24,8	28,4	78,5	26,2
	Ячмень+вика (BV)	25,9	26,7	27,6	80,2	26,7
Суммы главных деленок (Y _{2j})		92,7	93,7	101,9	288,3 = Y _{2..}	24,0
Суммы блоков (Y _{.j})		161,9	163,2	172,2	497,3 = Y _{...}	20,7

	Зеленые удобрения			
	F (0)	B	V	BV
Суммы (Y _{..k})	98,3	117,5	144,5	137,0
Средние (Y _{..k})	16,4	19,6	24,1	22,8

* Символы факторов и их градаций: N — азотное удобрение, n=2; G — зеленое удобрение, g=4; B — блоки, b=3.

Таблица 8.3. Сводная таблица дисперсионного анализа эксперимента с сахарной свеклой

Источники вариации	df	SS	MS	F _{факт}	F _{теор}	
					5%	1%
Субделанки (SP)	23	516,12				
Главные деланки (MP)	5	274,92				
Блоки (B)	2	7,87	3,935			
Азотное удобрение (N)	1	262,02	262,020	104,18	18,51	98,49
Ошибка (MP)	2	5,03	2,515			
Зеленое удобрение (G)	3	215,26	71,753	118,99	3,49	5,95
Взаимодействие (N×G)	3	18,70	6,233	10,34		
Ошибка (SP) BG+B (N×G)	12	7,24	0,603			

харной свеклы в зависимости от вида использованных зеленых удобрений по всем субделянкам сведены для анализа в таблице 8.2.

При дисперсионном анализе в первую очередь определяют источники вариации и степени свободы для них, т. е. используют (заполняют) первые две колонки таблицы 8.3.

Источники вариации и степени свободы. Общее число степеней свободы в опыте будет на единицу меньше, чем количество субделянок, $ngb-1=2\cdot4\cdot3-1=23$. Главные делянки выделяют в качестве одного из источников вариации, поскольку при их расчленении получают степени свободы для ошибки главных делянок. Степени свободы главных делянок $[df(MP)=nb-1=(2\cdot3)-1=5]$ распределяют в соответствии со схемой их расположения. В случае схемы рендомизированных блоков имеем степени свободы для: блоков $(b-1)=3-1$; азотного удобрения $(n-1)=2-1$; ошибки главных делянок, часто называемые ошибкой «а» $= (b-1)(n-1) = (3-1)(2-1)$. Степени свободы для зеленых удобрений $(g-1)=4-1$ и для взаимодействия факторов $N \times G = (n-1)(g-1) = (2-1)(4-1)$. Степени свободы для ошибки субделянок, обычно называемые ошибкой «b», можно получить путем вычитания, обращая внимание на промежуточные источники вариации, т. е. $23-5-3-3=12$, или путем сложения степеней свободы для взаимодействий: блоки $\times$ зеленые удобрения и блоки $\times$ азотное удобрение $\times$ зеленые удобрения, т. е. $(b-1)(g-1) + (b-1)(n-1)(g-1) = (3-1)(4-1) + (3-1)(2-1)(4-1) = 12$.

Корректирующий фактор.

$$C = \frac{Y^2}{ngb} = \frac{497,3^2}{2(4)(3)} = 10\,304,47.$$

Суммы квадратов и средние квадраты.

$$\text{Для блоков: } SSB = \frac{\sum Y_{i.}^2}{ng} - C = \frac{161,9^2 + 163,2^2 + 172,2^2}{2(4)} - C = 7,87.$$

Отметьте, что знаменатель (8) — это число дат, составляющих каждое выражение (член) числителя.

$$MSB = \frac{SSB}{df(B)} = \frac{7,87}{2} = 3,935.$$

Для азотного удобрения:

$$SSN = \frac{\sum Y_{.j}^2}{gb} - C = \frac{209,0^2 + 288,3^2}{3(4)} - C = 262,02;$$

$$MSN = \frac{SSN}{df(N)} = \frac{262,02}{1} = 262,02.$$

Для главных делянок:

$$SS(MP) = \frac{\sum Y_{ij}^2}{g} - C = \frac{69,2^2 + \dots + 101,9^2}{4} - C = 274,92.$$

Для ошибки (главных делянок):

$$SS(MPE) = SS(MP) - SSB - SSN = 274,92 - 7,87 - 262,02 = 5,03;$$

$$MS(MPE) = \frac{SS(MPE)}{(MPE)} = \frac{5,03}{2} = 2,515.$$

Для зеленых удобрений:

$$SSG = \frac{\sum Y_{..k}^2}{nb} - C = \frac{98,3^2 + \dots + 137,0^2}{2(9)} - C = 215,26;$$

$$MSG = \frac{SSG}{df_i(0)} = \frac{215,26}{3} = 71,753.$$

Для взаимодействия ($N \times GM$):

$$SS(N \times G) = \frac{\sum Y_{i.k}^2}{b} - C - SSN - SSG = \\ = \frac{40,5^2 + \dots + 80,0^2}{3} - C - 262,02 - 215,26 = 18,70.$$

$$MS(N \times G) = \frac{(N \times G)}{df(N \times G)} = \frac{18,70}{3} = 6,233.$$

Для субделянок:

$$SS(SP) = \sum Y_{ijk}^2 - C = 13,8^2 + 15,5^2 + \dots + 27,6^2 - C = 516,12.$$

Для ошибки (субделянок):

$$SS(SPE) = SS(SP) - SS(MP) - SSG - SS(N \times G) = \\ = 516,12 - 274,92 - 215,26 - 18,70 = 7,24;$$

$$MS(SPE) = \frac{SS(SPE)}{df(SPE)} = \frac{7,24}{12} = 0,603.$$

Клaviша стандартного отклонения. На калькуляторе с программой для расчета $s = \sqrt{\sum (Y_i - \bar{Y})^2 / (r-1)}$ сумму квадратов получают по следующему правилу. Вводят в машину каждую сумму посредством соответствующей клавиши. После ввода всех сумм надо нажать клавишу стандартного отклонения, возвести стандартное отклонение в квадрат, полученное s^2 разделить на число дат, вошедших в каждую сумму, и умножить на число степеней свободы. Например, при расчете $SS(MP)$: вводим 69,2; 69,5; ...; 101,9; нажимаем клавишу стандартного отклонения, получаем = 14,830; возводим в квадрат = 219,9377; делим на 4 = 54,9844 = $MS(MP)$; умножаем на 5 = 274,92 = $SS(MP)$.

F-величины (фактические). Действие азотного удобрения проверяется на основе среднего квадрата ошибки главных делянок, т. е. $MS(MPE)$, зеленых удобрений и взаимодействий азотного и зеленых удобрений на основе $MS(SPE)$. F для N равно $262,02/2,515 = 104,18$. F для $N \times G$ равно $6,233/0,603 = 10,34$. Очень существенная фактическая величина F свидетельствует о различной отзывчивости сахарной

Свеклы на зеленые удобрения в зависимости от уровня ее обеспеченности азотом. Трудностью эксперимента является выявление и объяснение взаимодействия, что составляет проблему при оценке частных различий и агрономических условий.

ВЫЯВЛЕНИЕ СУЩЕСТВЕННЫХ СРЕДНИХ. ОЦЕНКА ЧАСТНЫХ РАЗЛИЧИЙ

F-оценка. Расчленив сумму квадратов для $N \times G$ взаимодействия, мы постигаем природу взаимодействия. Для этого имеется несколько способов, однако лишь расчленение с целью ответа на три следующих вопроса представляется логичным. Различается ли отзывчивость сахарной свеклы при двух уровнях обеспеченности азотом за счет зеленых удобрений, например, при использовании: вики и без нее; пара и ячменя; вики и ячменя с викой? В таблице 8.4 даны суммы по вариантам и комплект ортогональных коэффициентов для расчета компонентов взаимодействия, а также других сравнений, имеющих единственную степень свободы.

Для распределения 3-х степеней свободы взаимодействия $N \times G$ мы должны прежде записать коэффициенты для расчленения главных эффектов на компоненты с единственной степенью свободы. Коэффициенты для N будут простыми, поскольку сравниваются лишь две группы. Четыре варианта зеленых удобрений расчленяются с целью ответа на три поставленных к взаимодействию вопроса, причем это осуществляется для обоих уровней азота; получаем следующие группы сравнений: $V+B$ и $F+B$, F и B , V и BV . Суммы квадратов этих трех сравнений при их сложении дадут сумму квадратов для зеленых удобрений, указанных в таблице 8.3. Подобные сравнения не имеют большого значения, но поскольку мы уже знаем различный эффект зеленых удобрений в зависимости от обеспеченности растений азотом, то наша задача сводится к более детальной оценке, достигаемой в результате вычисления F-критериев для 5, 6 и 7 сравнений таблицы 8.4. Коэффициенты для

Таблица 8.4. Ортогональные коэффициенты для указанных сравнений

Сравнения для выявления эффектов	Варианты, суммы по вариантам и средние							
	N_0				$N_{1,25}$			
	F $\bar{Y}_{i.k}$ 40,5 $\bar{V}_{i.k}$ 13,5	B 45,7 15,2	V 66,0 22,0	BV 56,8 18,9	F 57,8 19,3	B 71,8 23,9	V 78,5 26,2	BV 80,2 26,7

1. N	-1	-1	-1	-1	1	1	1	1
2. V и без V	-1	-1	1	1	-1	-1	1	1
3. F и B	-1	1	0	0	-1	1	0	0
4. V и BV	0	0	-1	1	0	0	-1	1
5. $N \times (V$ и без $V)$	1	1	-1	-1	-1	-1	1	1
6. $N \times (V$ и $B)$	1	-1	0	0	-1	1	0	0
7. $N \times (V$ и $BV)$	0	0	1	-1	0	0	-1	1

Таблица 8.5. Компоненты дисперсий для взаимодействия

Источники вариации	df	SS	MS	F _{факт}	F _{теор}	
					5%	1%
N×G	3	18,70	6,233	10,34	3,49	5,95
N×(V и без V)	1	2,344	2,344	3,88	4,75	9,33
N×(F и B)	1	6,453	6,453	10,70		
N×(V и BV)	1	9,901	9,901	16,42		
SPE (ошибка субделянок)	12	7,24	0,603			

этих сравнений получают путем перемножения коэффициентов для 1-го и 2-го, 1-го и 3-го, 1-го и 4-го сравнений.

Вычисления сумм квадратов трех компонентов взаимодействия приводятся ниже; результаты сведены в таблице 8.5.

$$SS[N \times (V \text{ и без } V)] = \frac{(40,5 + 45,7 - 66,0 - 56,8 - 57,8 - 71,8 + 78,5 + 80,2)^2}{3(8)} = \frac{7,5^2}{24} = 2,344.$$

Отметьте, что снова используется формула для расчета сумм квадратов, имеющих единственную степень свободы:

$$SS = (\sum c_i Y_{i..})^2 / (r \sum c_i^2).$$

В таких случаях все коэффициенты (c_i) равны ± 1 и поэтому отпадает необходимость их записи в числителях.

$$SS[N \times (F \text{ или } B)] = \frac{(40,5 - 45,7 - 57,8 + 71,8)^2}{3(4)} = 6,453;$$

$$SS[N \times (V \text{ или } BV)] = \frac{(66,0 - 56,8 - 78,5 + 80,2)^2}{3(4)} = 9,901.$$

Три F-критерия таблицы 8.5, каждый из которых имеет одну степень свободы для дисперсии числителя, дают ответы на три вопроса, поставленных выше.

N×(V против без N). Отзывчивость сахарной свеклы на применение вики в качестве зеленого удобрения различается несущественно по фону N_0 и N_{135} . Обращаясь к средним таблицы 8.4, мы находим, что разность в урожаях с делянок, получавших и не получавших вику в качестве зеленого удобрения на фоне N_0 , несущественно отличается от подобной разности на фоне N_{135} , т. е. $(22,0 + 18,9 - 13,5 - 15,2)/2 = 6,1$ сравниваем с $(26,2 + 26,7 - 19,3 - 23,9)/2 = 4,85$, получаем $6,1 - 4,85 = 1,25$, что несущественно отличается от нуля. Делитель 2 используют для установления разности средних на одну делянку.

Для оценки этого сравнения можно использовать t-критерий, однако он приведет к такому же статистическому заключению. Мы используем его здесь, чтобы показать эквивалентность двух оценок и продемонстрировать простоту расчета F-критерия в сравнении с t-критерием. Мы оцениваем разность двух разностей: $6,1 - 4,85 = 1,25$. Необходи-

данный критерий определяется из формулы

$$t = (\bar{d}_1 - \bar{d}_2) / s_{\bar{d}_1 - \bar{d}_2},$$

где $\bar{d}_1 = 6,1$; $\bar{d}_2 = 4,85$ и $s_{\bar{d}_1 - \bar{d}_2}$ — стандартная ошибка разности разностей, определяемая следующим образом:

$$s_{\bar{d}_1 - \bar{d}_2} = \sqrt{s_{\bar{d}_1}^2 + s_{\bar{d}_2}^2} = \sqrt{\frac{2s_1^2}{r_1} + \frac{2s_2^2}{r_2}}.$$

Если s_1^2 и s_2^2 оценивают общую дисперсию и $r_1 = r_2$, то

$$s_{\bar{d}_1 - \bar{d}_2} = \sqrt{4s^2/r} = \sqrt{4s^2/6} = \sqrt{4(0,603)/6} = 0,634.$$

Делитель 6 означает число дат, составляющих сравниваемые средние разности, в данном случае каждая средняя разность является средней двух средних по вариантам, имеющим трехкратную повторность.

Подставляем это значение в вышеприведенную формулу и получаем $t = (6,1 - 4,85) / 0,634 = 1,25 / 0,634 = 1,97$, то есть несущественную величину, поскольку необходимая величина t_{05} для 12 степеней свободы равна 2,179. Воспользуемся известным равенством $t^2 = F = (1,97)^2 = 3,88$, т. е. величину F , данную в таблице 8.5.

$N \times (F \text{ против } B)$. Разность в урожайности сахарной свеклы по пару и ячменю существенно ниже на фоне N_0 , чем на фоне $N_{13,5}$, т. е. $15,2 - 13,5 = 1,7$, что значительно меньше, чем $23,9 - 19,3 = 4,6$. Прибавка в урожайности от ячменя на зеленое удобрение по сравнению с паром составила на $4,6 - 1,7 = 2,9$ т/0,4 га больше на фоне азотного удобрения, чем без него. Доверительный интервал этой разности разностей может быть вычислен из $CL_{95} = \bar{d}_1 - \bar{d}_2 \pm ts_{\bar{d}_1 - \bar{d}_2}$, где t — табличная величина для степеней свободы и 5%-ного уровня значимости; $s_{\bar{d}_1 - \bar{d}_2} = \sqrt{4s^2/r} = \sqrt{4(0,603)/3} = 0,897$. В ответе запишем $CL_{95} = 2,9 \pm 2,179(0,897) = 2,9 \pm 2,0 = 0,9 - 4,9$ т/0,4 га. Таким образом, с 95%-ной уверенностью мы можем сказать, что для условий опыта положительный эффект использования ячменя на зеленое удобрение был на $0,9 - 4,9$ т/0,4 га больше на фоне дополнительного внесения азотного удобрения под сахарную свеклу, чем без азотного удобрения.

$N \times (V \text{ против } BV)$. Совместный посев вики с ячменем на зеленое удобрение существенно снижает урожайность сахарной свеклы в сравнении с посевом на удобрение одной вики: $22,0 - 18,9 = 3,1$ т/0,4 га; подобного снижения урожайности не было в случае дополнительного применения под сахарную свеклу азотного удобрения: $26,2 - 26,7 = -0,5$. Доверительный интервал для разности разностей составляет $CL_{95} = 3,1 - (-0,5) \pm 2,179(0,897) = 3,6 \pm 2,0 = 1,6$ до $5,6$ т/0,4 га.

Результаты этого эксперимента можно представить в виде данных таблицы 8.6 и 8.7. Таблица 8.6 дает средние урожаи по вариантам, а таблица 8.7 — необходимую статистическую информацию для оценки существенности взаимодействия. Принято обозначать статистическую существенность одной, двумя или тремя звездочками для 5,1 и 0,1% уровней значимости соответственно. Общие средние для уровней воз-

Таблица 8.6. Влияние зеленых удобрений и азота на урожайность (т/0,4 га) корнеплодов сахарной свеклы

Азотное удобрение, кг N/га	Зеленые удобрения			
	пар (0)	ячмень	вика	ячмень + вика
0	13,5	15,2	22,0	18,9
135	19,3	23,9	26,2	26,7
<i>HCP</i> ₀₅ для зеленых удобрений и такого же уровня N—1,4. То же и различных уровней N—2,9.				

действия азота или зеленых удобрений не приводятся, поскольку при значительном взаимодействии они не имеют значения (нелогичны).

HCP, приведенные в таблице 8.6, не являются существенно необходимыми, однако они помогают при интерпретации результатов.

Стандартные ошибки и *HCP*. Иногда для оценки результатов эксперимента необходимо определять *HCP* или многограновые критерии. Стандартные ошибки для этих оценок вычисляют на основе вариабельности между экспериментальными единицами, по которым размещают варианты. В опытах, заложенных по схеме расщепленных делянок, расчеты стандартной ошибки для необходимых сравнений вариантов усложняются в связи с тем, что (как видно из таблицы 8.8) мы имеем два источника экспериментальной ошибки: одна из них относится к основной делянке, а другая — к субделянкам.

Необходимо подчеркнуть, что стандартная ошибка для сравнения средних по вариантам субделянок внутри главных делянок включает только ошибку субделянок, однако если сравнение средних субделянок производится для различных главных делянок, то стандартная ошибка включает как ошибку субделянок, так и ошибку главных делянок.

Не затрагивая глубоко алгебраической стороны вопроса, находим, что стандартная ошибка для субделянок является средней взве-

Таблица 8.7. Средние квадраты взаимодействия и взаимодействие компонентов при расчете влияния зеленых удобрений и азота на урожайность корней сахарной свеклы

Источники вариации *	<i>df</i>	Средние квадраты
$N \times G$	3	6,233 **
$N \times (V \text{ и без } V)$	1	2,344
$N \times (F \text{ и } B)$	1	6,453 **
$N \times (V \text{ и } BV)$	1	9,901 **
Ошибка субделянок	12	0,603
Ошибка главных делянок	2	2,515

* N — азот; G — зеленые удобрения; V — вика; F — пар; B — ячмень и BV — ячмень с викой.

** Две звездочки — математический знак; существенно на 1%-ом уровне.

Таблица 8.8. Формулы расчета стандартных ошибок для опыта, заложенного по схеме расщепленных делянок

Пары сравниваемых средних	Стандартная ошибка средних *, $s_{\bar{y}}$
Варианты главных делянок: $A_1 - A_2$	$\sqrt{\frac{Ea}{rb}}$
Варианты субделянок: $B_1 - B_2$	$\sqrt{\frac{Eb}{ra}}$
Варианты субделянок внутри одной главной делянки: $B_1A_1 - B_2A_1$	$\sqrt{\frac{Eb}{r}}$
Варианты субделянок для разных главных делянок: $B_1A_1 - B_1A_2$ или $B_1A_1 - B_2A_2$	$\sqrt{\frac{(b-1)Eb + Ea}{rb}}$

* Обратите внимание на использование $s_{\bar{y}}$ при определении НСР или D (НСР = $t \sqrt{2}$ * $s_{\bar{y}}$; $D = R$ (НСР). $Ea = MS$ (MPE); $Eb = MS$ (SPE); a — число вариантов главных делянок; b — число вариантов субделянок; r — повторность; A — варианты главных делянок и B — варианты субделянок.

шенной величин Ea и Eb ; величина средней взвешенной для Ea и Eb составляют 1 и $b-1$ соответственно. Поскольку $1+b-1=b$, знаменателем будет br , где b — число вариантов субделянок и r — повторность.

Последовательность расчетов НСР для всех возможных сравнений средних урожаев сахарной свеклы (табл. 8.2) в двухфакторном опыте с использованием зеленых удобрений и азотного удобрения приводится ниже.

НСР для оценки различий между вариантами главных делянок (между средними уровнями азотного удобрения):

$$НСР_{05} = t_a \sqrt{\frac{2(Ea)}{rb}},$$

где t_a — табличная величина t для числа степеней свободы Ea .

$$НСР_{05} = 4,303 \sqrt{\frac{2(2,515)}{3(4)}} = 4,303(0,647) = 28 \text{ т/0,4 га.}$$

НСР для оценки различий между вариантами субделянок (между средними зелеными удобрениями):

$$НСР_{05} = t_b \sqrt{\frac{2(Eb)}{ra}},$$

где t_b — табличная величина t для степеней свободы Eb .

$$НСР_{05} = 2,179 \sqrt{\frac{2(0,603)}{3(2)}} = 2,179(0,448) = 1 \text{ т/0,4 га.}$$

НСР для оценки различий между вариантами субделянок внутри одной главной делянки (между средними зеленых удобрений на одном уровне азотного удобрения):

$$HCP_{05} = t_b \sqrt{\frac{2(Eb)}{r}} = 2,179 \sqrt{\frac{2(0,603)}{3}} = 2,179 (0,634) = 1,4 \text{ т/0,4 га.}$$

НСР для оценки различий между вариантами субделянок разных главных делянок (сравнение разных средних зеленых удобрений при различных уровнях азотного удобрения или сравнение средних одного вида зеленого удобрения при разных уровнях азотного удобрения):

$$HCP_{05} = t_{ab} \sqrt{\frac{2[(b-1)Eb + Ea]}{rb}},$$

где t_{ab} — взвешенная величина t , занимающая промежуточное положение между значениями t_a и t_b , вычисляется по формуле

$$\begin{aligned} t_{ab} &= \frac{(b-1)(Eb)(t_b) + Ea(t_a)}{(b-1)Eb + Ea} = \\ &= \frac{(4-1)(0,603)(2,179) + 2,515(4,303)}{(4-1)(0,603) + 2,515} = \frac{14,764}{4,324} = 3,414; \end{aligned}$$

$$HCP_{05} = 3,414 \sqrt{\frac{2[(4-1)(0,603) + 2,515]}{3(4)}} = 3,414 (0,849) = 2,9 \text{ т/0,4 га.}$$

Если бы все восемь сочетаний этих факторов были случайно размещены внутри каждого блока, то мы имели бы схему рендомизированных блоков. Средний квадрат ошибки составил бы тогда следующую величину:

$$EMS = \frac{SS(MP) + SS(SP)}{df(MP) + df(SP)} = \frac{5,03 + 7,24}{2 + 12} = 0,876$$

и *НСР* для всех сравнений вариантов равнялась бы

$$HCP = t \sqrt{\frac{2(0,876)}{3}} = 2,145 (0,764) = 1,6 \text{ т/0,4 га,}$$

где t — табличная величина для 14 степеней свободы и 5%-ного уровня значимости.

Сравнение рассчитанных выше *НСР* показывает относительную эффективность двух схем по выявлению эффектов вариантов. Отметим повышенную «чувствительность» (меньшие величины *НСР*) схемы расщепленных делянок для оценки средних вариантов субделянок и срав-

вариантов субделянок внутри одной главной делянки, но понижая точность (большие величины HCP) при сравнении вариантов главных делянок и вариантов субделянок, расположенных поперек главных делянок.

ВЫВОДЫ

Схема расщепленных делянок часто отвечает требованиям факториальной группы вариантов. Схема включает случайное размещение вариантов одного фактора или сочетания факторов на главных делянках, которые затем расщепляются на субделянки для случайного размещения вариантов другого фактора или сочетания факторов. В сравнении со схемой рендомизированных блоков точность сравнения вариантов главных делянок и вариантов субделянок разных главных делянок снижается, однако повышается точность сравнения вариантов субделянок и вариантов субделянок, расположенных внутри одной главной делянки.

СХЕМА РАСЩЕПЛЕННЫХ ДЕЛЯНОК. СПЛИТ-СПЛИТ-ПЛОТ (трехфакторный опыт)

Добавление третьего фактора в схему расщепленных делянок (сплит-плот) путем расщепления субделянок дает схему сплит-сплит-плот. Эта схема очень часто используется для трехфакторных экспериментов с целью облегчения полевых работ или в случае необходимости компактного расположения всех сочетаний вариантов. Однако дополнительное ограничение рендомизации вызывает необходимость расчета третьей величины ошибки, используемой при оценке главных эффектов фактора, размещаемого на делянках второго расщепления, и для всех взаимодействий, включающих этот фактор. В отношении технической стороны расположения экспериментальных единиц мы можем иметь определенные преимущества, однако необходимость использовать специальный термин для обозначения третьей ошибки может значительно усложнить выявление существенных средних. Необходимо получить консультацию у специалиста по опытному делу, прежде чем использовать эту схему для расчета.

Принципы рендомизации остаются теми же, что и для сплит-плот схемы, но проводится расщепление субделянок по числу градаций третьего фактора, на суб-субделянки, на которых проводят его случайное размещение, т. е. используют новую рендомизацию вариантов для каждого комплекта суб-субделянок. Рисунок 9.1 иллюстрирует общий вид сплит-сплит-плот схемы эксперимента по изучению влияния сроков посева, обработки против тли и сроков уборки на эффективность борьбы с вирусом тли на сахарной свекле. Процедура необходимой обработки данных этого эксперимента будет сопровождаться оценкой действия изучаемых вариантов на урожайность корнеплодов.

ТАБУЛИРОВАНИЕ ДАННЫХ

В таблице 9.1 данные сгруппированы по вариантам и блокам. В таблице 9.2 определяют суммы для выявления двусторонних взаимодействий и главных эффектов.

ДИСПЕРСИОННЫЙ АНАЛИЗ

Результаты дисперсионного анализа приводятся в таблице 9.3. Ниже дается последовательность выполнения анализа.

I			Основная делянка			II					
Блок											
			P ₁			P ₃			P ₂		
субделянка	III					суб-субделянка	IV				
P ₂ S ₂	P ₁ S ₁	P ₃ S ₂	P ₃ S ₁ H ₂	P ₃ S ₁ H ₃	P ₃ S ₁ H ₁						
субделянка						суб-субделянка					
P ₂ S ₁	P ₁ S ₂	P ₃ S ₁	P ₃ S ₂ H ₁	P ₃ S ₂ H ₂	P ₃ S ₂ H ₃						

Рис. 9.1. Общий вид сплит-сплит-плот схемы эксперимента по оценке эффективности борьбы с тлей на сахарной свекле. Главные делянки, предназначенные для определения сроков посева (P₁, P₂, P₃), расположены по схеме рандомизированных блоков (I, II, III, IV). На субделянках определяли влияние тли на рост сахарной свеклы: s₁ — без опрыскивания и s₂ — с применением опрыскивания против тли. Суб-суб-делянки были предназначены для определения влияния сроков уборки с 4-недельным интервалом (H₁, H₂, H₃). На суб-субделянках одной главной делянки (P₃) IV блока дана урожайность сахарной свеклы. Полные данные этого опыта приводятся в таблице 9.1.

Источники вариации и степени свободы. Степени свободы для источников вариации, указанных в таблице 9.3, определялись следующим образом:

$$\text{суб-субделянки} = pshb - 1 = 3(2)(3)(4) - 1 = 71,$$

$$\text{субделянки} = psb - 1 = 23,$$

$$\text{главные делянки} = pb - 1 = 11,$$

$$\text{блоки} = b - 1 = 3,$$

$$\text{сроки посева} = p - 1 = 2.$$

$$\text{Ошибка главных делянок} = (b-1)(p-1) = 6, \text{ или } 11 - 3 - 2 = 6.$$

$$\text{Обработка против тли} = s - 1 = 1.$$

$$P \times S = (p-1)(s-1) = 2.$$

$$\text{Ошибка субделянок} = (b-1)(s-1) + (b-1)(p-1)(s-1) = 3 + 6 = 9, \text{ или } 23 - 11 - 2 - 1 = 9.$$

$$\text{Сроки уборки} = h - 1 = 2,$$

$$P \times H = (p-1)(h-1) = 4,$$

$$S \times H = (s-1)(h-1) = 2,$$

$$P \times S \times H = (p-1)(s-1)(h-1) = 4.$$

$$\text{Ошибка суб-субделянок} = (b-1)(h-1) + (b-1)(p-1)(h-1) + (b-1)(s-1)(h-1) + (b-1)(p-1)(s-1)(h-1) = 6 + 12 + 6 + 12 = 36, \text{ или } 71 - 23 - 2 - 4 - 2 - 4 = 36.$$

Таблица 9.1. Данные урожайности (т/0,4 га), сгруппированные по вариантам и блокам для сплит-сплит-плот схемы

Варианты			Блоки (j)				Суммы	Сред- ние
P (i)	S (k)	H (l)	I	II	III	IV		
1	1	1	25,7	25,4	23,8	22,0	96,9	24,2
		2	31,8	29,5	28,7	26,4	116,4	29,1
		3	34,6	37,2	29,1	23,7	124,6	31,2
	SP суммы $Y_{1/j1}$		92,1	91,2	81,6	72,1	337,9 = $Y_{1.1}$	28,2
	2	1	27,7	30,3	30,2	33,2	121,4	30,0
		2	38,0	40,6	34,6	31,0	144,2	36,0
		3	42,1	43,6	44,6	42,7	173,0	43,2
	SP суммы $Y_{1/j2}$		107,8	114,5	109,4	106,9	438,6 = $Y_{1.2}$	36,6
	MP суммы $Y_{1/j..}$		199,9	206,6	191,0	179,0	776,5 = $Y_{1...}$	
	2	1	28,9	24,7	27,8	23,4	104,8	26,2
		2	37,5	31,5	31,0	27,8	127,8	32,0
		3	38,4	32,5	31,2	29,8	131,9	33,0
	SP суммы $Y_{2/j1}$		104,8	88,7	90,0	81,0	364,5 = $Y_{2.1}$	30,4
	2	1	38,0	31,0	29,5	30,7	129,2	32,3
		2	36,9	31,9	31,5	35,9	136,2	34,0
		3	44,2	41,6	38,9	37,6	162,3	40,6
	SP суммы $Y_{2/j2}$		119,1	104,5	99,9	104,2	427,7 = $Y_{2.2}$	35,6
	MP суммы $Y_{2/j..}$		223,9	193,2	189,9	185,2	792,2 = $Y_{2...}$	
3	1	1	23,4	24,2	21,2	20,9	89,7	22,4
		2	25,3	27,7	23,7	24,3	101,0	25,2
		3	29,8	29,9	24,3	23,8	107,8	27,0
	SP суммы $Y_{3/j1}$		78,5	81,8	69,2	69,0	298,5 = $Y_{3.1}$	24,9
	2	1	20,8	23,0	25,2	23,1	92,1	23,0
		2	29,0	32,0	26,5	31,2	118,7	29,7
		3	36,6	37,8	34,8	40,2	149,4	37,4
	SP суммы $Y_{3/j2}$		86,4	92,8	86,5	94,5	360,2 = $Y_{3.2}$	30,0
	MP суммы $Y_{3/j..}$		164,9	174,6	155,7	163,5	658,7 = $Y_{3...}$	
	B суммы $Y_{j..}$		588,7	574,4	536,6	527,7	2227,4 = $Y_{...}$	

$$C = (2227,4)^2 / 72 = 68\,907,0939, \quad \sum Y_{ijk1}^2 = 71\,747,70$$

Символы обозначения вариантов и градаций факторов: P — сроки посева, p — 3 срока; S — аэрозольная обработка против тлей; s=2; H — сроки уборки, h=3; B — блоки, b=4.

Таблица 9.2. Суммы для двухсторонних взаимодействий и главных эффектов

Суммы для двухсторонних взаимодействий								
$P \times S (Y_{t..k})$			$P \times H (Y_{t..1})$			$S \times H (Y_{...1})$		
	S_1	S_2	H_1	H_2	H_3		S_1	S_2
P_1	337,9 ^a	438,6	218,3 ^b	260,6	297,6	H_1	291,4 ^c	342,7
P_2	364,5	427,7	234,0	264,0	294,2	H_2	345,2	399,1
P_3	298,5	360,2	181,8	219,7	257,2	H_3	364,3	484,7

Суммы для главных эффектов

Сроки посева ($Y_{t...}$)			Афицид ($Y_{...k}$)		Сроки уборки ($Y_{...1}$)		
P_1	P_2	P_3	S_1	S_2	H_1	H_2	H_3
776,5	792,2	658,7	1000,9	1226,5	634,1	744,3	849,0

^a Из таблицы 9.1 суммы для P, S по всем срокам уборки и блокам.

^b Суммы для P, H по всем обработкам против тлей и блокам: $96,9 + 121,4 = 218,3$.

^c Суммы для S, H по всем срокам посева и блокам: $96,9 + 104,6 + 89,7 = 291,4$.

Если бы 18 вариантов этого эксперимента были размещены по схеме рендомизированных блоков, то мы имели бы одну величину ошибки с числом степеней свободы $df = (b-1)(t-1) = 3(17) = 51$, что составляет

Таблица 9.3. Сводная таблица дисперсионного анализа

Источники вариации	df	SS	MS*	$F_{\text{факт}}$	$F_{\text{теор}}$	
					5%	1%
Суб-субделянки	71	2840,6061				
Субделянки	23	1542,8128				
Главные делянки	11	698,9028				
Блоки, B	3	143,4561	47,8187			
Сроки посева, P	2	443,6886	221,8443	11,91	5,14	10,92
Ошибка главных делянок, BP	6	111,7581	18,6264			
Обработка против тлей, S	1	706,8800	706,8800	81,21	5,12	10,56
$P \times S$	2	40,6875	20,3438	2,34	4,26	8,02
Ошибка субделянок, $BS + B(P \times S)$	9	78,3425	8,7047			
Сроки уборки, H	2	962,3353	481,1676	102,80	3,26	5,25
$P \times H$	4	13,1097	3,2774	0,70	2,63	3,89
$S \times H$	2	127,8308	63,9154	18,66	3,26	5,25
$P \times S \times H$	4	44,0192	11,0048	2,35	2,63	3,89
Ошибка суб-субделянок $BH + B(P \times H) + B(S \times H) + B(P \times S \times H)$	36	168,4983	4,6805			

* Стрелками указаны пары дисперсий для определения F -критерия.

сумму степеней свободы для трех ошибок таблицы 9.3, т. е. $6+9+36=51$. Таким образом, расщепление делянок приводит к расчленению степеней свободы и суммы квадратов для ошибки на компоненты, имеющие меньшее число степеней свободы и обычно меньший средний квадрат для каждой последующей величины ошибки. Сравните средние квадраты для трех ошибок таблицы 9.3.

Корректирующий фактор.

$$C = \frac{Y^2}{pshb} = \frac{2227,4^2}{3(2)(3)(4)} = 68\,907,0939.$$

Суммы квадратов и средние квадраты.

$$SSB = \frac{\sum Y_{j..}^2}{psh} - C = \frac{588,7^2 + \dots + 527,7^2}{(3)(2)(3)} - C = 142,4561.$$

$$SSP = \frac{\sum Y_{i..}^2}{shb} - C = \frac{776,5^2 + \dots + 658,7^2}{(2)(3)(4)} - C = 443,6886.$$

$$SS(MP) = \frac{\sum Y_{ij.}^2}{sh} - C = \frac{199,9^2 + \dots + 163,5^2}{(2)(3)} - C = 698,9028.$$

$$SS(MPE) = SS(MP) - SSB - SSP = 111,7581.$$

$$SSS = \frac{\sum Y_{..k.}^2}{phb} - C = \frac{1000,9^2 + 1226,5^2}{(3)(3)(4)} - C = 706,8800.$$

$$SS(P \times S) = \frac{\sum Y_{i.k.}^2}{hb} - C - SSP - SSS = \frac{337,9^2 + \dots + 360,2^2}{(3)(4)} - C - SSP - SSS = 40,675.$$

$$SS(SP) = \frac{\sum Y_{ijk.}^2}{h} - C = \frac{92,1^2 + \dots + 94,5^2}{4} - C = 1524,8128.$$

$$SS(SPE) = SS(SP) - SS(MP) - SSS - SS(P \times S) = 78,3425.$$

$$SSH = \frac{\sum Y_{...1}^2}{psb} - C = \frac{634,1^2 + \dots + 849,0^2}{(3)(2)(4)} - C = 962,3353.$$

$$SS(P \times H) = \frac{\sum Y_{i..1}^2}{sb} - C - SSP - SSH = \frac{218,3^2 + \dots + 257,2^2}{(2)(4)} - C - SSP - SSH = 13,1097.$$

$$SS(S \times H) = \frac{\sum Y_{...k1}^2}{pb} - C - SSS - SSH = \frac{291,4^2 + \dots + 484,7^2}{(3)(4)} - C - SSS - SSH = 127,8308.$$

$$SS(P \times S \times H) = \frac{\sum Y_{i.k.1}^2}{b} - C - SSP - SSS - SSH - SS(P \times S) - SS(P \times H) - SS(S \times H) = \frac{96,9^2 + \dots + 149,4^2}{4} - C - SSP - SSS - SSH - SS(P \times S) - SS(P \times H) - SS(S \times H) = 44,0192.$$

$$SS(SSP) = \sum Y_{ijkl}^2 - C = 25,7^2 + \dots + 40,2^2 - C = 2840,6061.$$

$$SS(SSPE) = SS(SSP) - SS(SP) - SSH - SS(P \times H) - SS(S \times H) - SS(P \times S \times H) = 168,4983.$$

Средние квадраты определяют обычным путем деления суммы квадратов (SS) на соответствующее число степеней свободы, например $MS(SSPE) = \frac{168,4983}{36} = 4,6805$.

Калькулятор стандартного отклонения. На калькуляторе с программой для расчета стандартного отклонения $s = \sqrt{(Y_i - \bar{Y})^2 / (r - 1)}$ суммы дат можно вводить в машину и получать средние квадраты и суммы квадратов. Например, для расчета SSB вводим блоковые суммы: 588,7; 574,4; 536,6 и 527,7; нажимаем клавишу стандартного отклонения, получаем $s = 29,3383$, возводим s в квадрат, получаем $s^2 = 860,7367$; делим на число дат (18), возведших в каждую сумму, возводимую в квадрат, и получаем $47,8187 = MSB$; умножаем на $df B(3)$, получаем $143,4561 = SSB$.

F-величины. Средний квадрат ошибки главных делянок используется для: 1) оценки эффектов сроков посева; 2) определения ошибки для субделянок (MS) при проверке эффектов обработки против тлей и взаимодействия $P \times S$ и 3) определения ошибки суб-субделянок для оценки оставшихся источников вариации, относящихся к вариантам суб-субделянок.

ОЦЕНКА СУЩЕСТВЕННЫХ СРЕДНИХ

Выбор метода выявления существенных средних будет зависеть от характера вариантов поставленных экспериментатором вопросов и результатов начальных анализов. В нашем примере эти анализы позволили утверждать, что влияние обработок против тлей и сроков уборки урожая было сходным для всех сроков посева (несущественное значение F -факторов для $P \times S$; $P \times H$ и $P \times S \times H$), однако реакция растений (подвергавшихся опрыскиванию) по отношению к срокам уборки была совершенно иной по сравнению с растениями, не обработанными против тлей (очень существенное значение F -фактора для $S \times H$).

Расчленение взаимодействия. Таблица 9.4 подготовлена с целью более детальной оценки $S \times H$ -взаимодействия. Средние этой таблицы указывают на увеличение урожая корнеплодов при более поздней уборке, причем прибавки урожая на фоне опрыскивания (S_2) будут выше, чем без него (S_1). Поскольку разрыв в сроках уборки составлял четыре недели, мы можем использовать коэффициенты таблицы А.11 (приложения) при $n=3$ для упрощения расчленения суммы квадратов для сроков уборки на величину, используемую для расчета линейной зависимости увеличения урожайности от сроков уборки, и величину, характеризующую долю суммы квадратов, не учтенную линейным трендом. Имея 2 степени свободы для сроков уборки, мы можем расчленить $2df$ для $S \times H$ на линейный и остаточный эффекты.

Используя средний квадрат ошибки суб-субделянок ($MSSPE$, табл. 9.3), мы вычислим F -фактор для данных таблицы 9.4, характеризующий очень существенное различие в линейной реакции сахарной свеклы при вариантах s_1 и s_2 в зависимости от сроков уборки.

Таблица 9.4. Коэффициенты для расчленения суммы квадратов вариантов на эффекты опрыскивания, сроков уборки и взаимодействия $S \times H$; средние квадраты и F -значения

Сравнения для выявления эффектов	Суммы	$S \times H$ -сочетания						Средний квадрат	F_a
		$S_1 H_1$	$S_1 H_2$	$S_1 H_3$	$S_2 H_1$	$S_2 H_2$	$S_2 H_3$		
		291,4	345,2	364,3	342,7	399,1	484,7		
	Сред-ние	24,3	28,8	30,4	28,6	33,3	40,4		

S	—1	—1	—1	1	1	1	706,8800	
H -линейное (L)	—1	0	1	—1	0	1	962,1252	205,6
H -остаточное (R)	1	—2	1	1	—2	1	0,2101	<1
$S \times HL$	1	0	—1	—1	0	1	99,4752	21,25
$S \times HR$	—1	2	—1	1	—2	1	28,3556	6,06

^a Рассчитаны путем деления средних квадратов эффектов на средний квадрат ошибки SSP таблицы 9.3. Табличное значение F_1 для проверки статистической существенности опирается на 1 и 36 df и равно для 5% — 4,11, а для 1% — 7,39.

В опыте также установлен существенный эффект остаточной величины (компонента) $S \times H$ -взаимодействия из-за меньшего увеличения в прибавках урожая корнеплодов сахарной свеклы при переходе от второго s_1 (28,8—30,4 т/0,4 га) к третьему сроку уборки s_2 (33,3—40,4 т/0,4 га) на фоне без инсектицида или с применением инсектицида соответственно. С агрономической точки зрения отмеченный факт объясняется тем, что не обработанные против тли растения сахарной свеклы постепенно замедляют свой рост по мере приближения срока уборки, тогда как темп роста обработанных растений был более или менее постоянным к моменту уборки в три разных срока. Эта интерпретация иллюстрируется на рисунке 9.2, на котором показана прибавка урожая от первого к третьему сроку уборки: прямой линией на фоне опрыскивания для s_2 и кривой (квадратичная зависимость) — для s_1 без обработки против тлей. Экстраполяция (продолжение) линий за пределы указанных сроков уборки будет неуместной, поскольку промежуток между линиями будет уменьшаться по мере приближения зимы, т. е. линия для s_2 не будет подниматься выше, а кривая для s_1 не будет снижаться. Однако в пределах трех сроков уборки уравнения обеих линий графически иллюстрируют повышение повреждаемости корнеплодов сахарной свеклы вирусами тли и указывают на необходимость определения урожайности по всем срокам осенней уборки.

Расчет уравнений регрессии (рис. 9.2) можно выполнять после того, как ознакомитесь с сокращенными методами анализа регрессии, приведенными в главе 15. Вычисления средних квадратов таблицы 9.4, опирающихся на одну степень свободы, приводятся ниже.

$$SS (\text{опрыскивание}) = \frac{(-291,4 - 345,2 - 364,3 + 342,7 + 399,1 + 484,7)^2}{4 (3) (6)} = 706,8800.$$

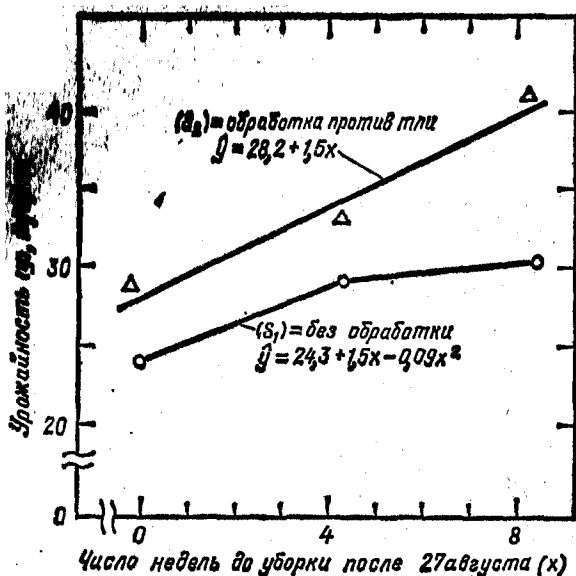


Рис. 9.2. Влияние обработки против тли на рост сахарной свеклы осенью. Различия в наклоне двух линий графика показывают характер $S \times H$ -взаимодействия. Для определения прироста урожая сахарной свеклы в течение осеннего периода на фоне опрыскивания против тли и без опрыскивания можно использовать уравнения регрессии, приведенные на рисунке.

В данном случае использовалась формула расчета сумм квадратов с одной степенью свободы $SS = (\sum c_i Y_i)^2 / (r \sum c_i^2)$. Величина c_i — коэффициент групповых (классовых) сравнений таблицы 9.4 и r — число дат, составляющих каждое выражение числителя. В данном примере $r = br = 4(3)$.

$$SS(H \text{ линейное}) = \frac{(-291,4 + 364,3 - 342,7 + 484,7)^2}{4(3)(4)} = 962,1252;$$

$$SS(H \text{ остаточное}) = \frac{[291,4 - 2(345,2) + 364,3 + 342,7 - 2(399,1) + 484,7]^2}{4(3)(12)} = 0,2101;$$

$$SS(S \times HL) = \frac{[291,4 - 364,3 - 342,7 + 484,7]^2}{4(3)(4)} = 99,4752;$$

$$SS(S \times HR) = \frac{[-291,4 + 2(345,2) - 364,3 + 342,7 - 2(399,1) + 484,7]^2}{4(3)(12)} = 28,3556.$$

В качестве проверки арифметических расчетов можно использовать два равенства: $SS(H \text{ линейное}) + SS(H \text{ остаточное}) = SSH$ (табл. 9.3) и $SS(S \times HL) + SS(S \times HR) = SS(S \times H)$ (табл. 9.3).

Стандартные ошибки и НСР. Для некоторых экспериментов, заложенных по сплит-сплит-плот схеме, желательно делать оценку частных различий между определенными средними с помощью НСР или много-ранговых критериев, поэтому необходимо знать соответствующие стандартные ошибки, используемые для этих оценок. Стандартные ошибки для оценки главных эффектов факторов, изучаемых на главных делянках и субделянках, и эффектов их взаимодействия определяют по тем же формулам, которые даны в таблице 8.8, за исключением того, что

Таблица 9.5. Формулы для расчета стандартных ошибок и t -критериев для выделения существенных средних, включая варианты с фактором C

Сравниваемые средние	Стандартные ошибки	t -значения *
Средние фактора C	$\sqrt{\frac{E_c}{rab}}$	t_c
C -средние для одного уровня A	$\sqrt{\frac{E_c}{rb}}$	t_c
C -средние для одного уровня B	$\sqrt{\frac{E_c}{ra}}$	t_c
B -средние для одного или разных уровней C	$\sqrt{\frac{(c-1)E_c + Eb}{rac}}$	$t_{bc} = \frac{(c-1)Ect_c + Ebt_b}{(c-1)E_c + Eb}$
A -средние для одного или разных уровней C	$\sqrt{\frac{(c-1)E_c + Ea}{rbc}}$	$t_{ac} = \frac{(c-1)Ect_c + Eat_a}{(c-1)E_c + Ea}$
C -средние для одного уровня A и B	$\sqrt{\frac{E_c}{r}}$	t_c
B -средние для одного A и одного или разных уровней C	$\sqrt{\frac{(c-1)E_c + Eb}{rc}}$	$t_{bc} = \frac{(c-1)Ect_c + Ebt_b}{(c-1)E_c + Eb}$
A -средние для одного или разных уровней B и C	$\sqrt{\frac{b(c-1)E_c + (b-1)Eb + Ea}{rbc}}$	$t_{abc} = \frac{b(c-1)t_c + (b-1)Ebt_b + Eat_a}{b(c-1) + (b-1)Eb + Ea}$

* t_a, t_b, t_c — табличные величины (табл. А.2., приложение) для степеней свободы Ea, Eb, Ec соответственно. A, B и C — варианты, размещаемые на главных делянках, субделянках и суб-субделянках с грациями a, b, c соответственно; r — повторность; Ea, Eb, Ec — средние квадраты ошибок для главных делянок, субделянок и суб-субделянок соответственно.

c (число градаций фактора C) является множителем для всех знаменателей. Стандартные ошибки для оценки средних третьего фактора, изучаемого на суб-субделянках, и эффектов взаимодействия его с двумя другими факторами даны в таблице 9.5 наряду с величинами для t , значения которого необходимо находить для определенных сравнений.

Обычно хорошим завершением оформления результатов эксперимента считается выделение средних эффектов сочетания факторов самого высокого порядка и средних сочетаний факторов, которые в большей мере относятся к результатам опыта, наряду с частью расчетов по оценке частных различий. Так, для нашего примера в таблице 9.6 даны средние эффектов тройного взаимодействия: 1) сроков посева $\times$ обработки против тлей $\times$ сроков уборки, 2) средние сроков посева и 3) средние очень существенного взаимодействия ($S \times H$). В примечании к таблице даются значения HCP для приблизительной оценки частных различий. Расчеты этих HCP (все на 5%-ном уровне) с использованием

Таблица 9.6. Влияние сроков посева, обработки против тлей и сроков уборки на урожайность корнеплодов сахарной свеклы, т/0,4 га

Сроки посева	Обработка против тлей (опрыскива- ние)	Сроки уборки			Средние для сро- ков по- сева ^a
		27/VIII	24/IX	22/X	
P×S×H (средние ^б)					
2/III	Нет	24,2	29,1	31,2	32,3
	Да	30,4	36,0	43,2	
2/IV	Нет	26,2	32,0	33,0	33,0
	Да	32,3	34,0	40,6	
2/V	Нет	22,4	25,2	27,0	27,4
	Да	23,0	29,7	37,4	

Обработка против тлей $\times$ сроки уборки (средние в)

Без опрыскивания	24,3	28,8	30,4
Опрыскивание	28,6	33,3	40,4

^a $HCP_{0.1} = 3,0$.

^б $HCP_{0.1}$ для оценки разностей между средними: для сроков уборки при одних сроках посева и фонах опрыскивания=3,1; для фонов опрыскивания при одних сроках посева и одних или разных сроках уборки=3,7; для сроков посева для одних или разных уровней опрыскивания или сроках уборки=4,4. Взаимодействие $P \times S \times H$ несущественно на 5%-ном уровне.

^в $HCP_{0.1}$ между сроками уборки на одном фоне опрыскивания=1,8; между фонами опрыскивания при одних или разных сроках уборки=2,1. Взаимодействие $S \times H$ существенно на 0,1%-ном уровне значимости.

стандартных ошибок таблиц 8.8 и 9.5 приводятся ниже. Для расчета HCP и D используйте формулы $HCP = t\sqrt{2s_y}$ и $D = R(HCP)$.

HCP для средних сроков посева. $HCP = t_a \sqrt{(2Ea)/rbc}$. Обратите внимание, что в знаменатель этой формулы (по сравнению с формулой из таблицы 8.8) включена величина c для сохранения уровня стандартной ошибки на уровне величин суб-субделянок.

$$HCP = 2,447 \sqrt{[2(18,6264)]/4(2) \times 3} = 2,447(1,246) = 3,0 \text{ т/0,4 га.}$$

HCP для средних сроков уборки (H) для одних сроков посева (P) и фонов опрыскивания (S). Например, разность двух средних $P_1S_1H_1 - P_1S_1H_2$. $HCP = t_c \sqrt{2Ec/r} = 2,028 \sqrt{[2(4,6805)]/4} = 2,028(1,530) = 3,1 \text{ т/0,4 га.}$ Учтите, что t_c основывается на 36df и определяется путем линейной интерполяции между табличными величинами t (табл. А.2, приложение) для 35 и 40df.

HCP для средних фонов опрыскивания при одних P , одних или разных H .

Например, $P_1S_1H_1 - P_1S_2H_1$ или $P_1S_1H_1 - P_1S_2H_2$.

$$HCP = t_{bc} = \sqrt{\frac{2[(c-1)Ec + Eb]}{rc}}; \quad t_{bc} = \frac{(c-1)Ect_c + Ebt_c}{(c-1)Ec + Eb};$$

$$t_{bc} = \frac{(3-1)(4,6805)2,028 + 8,7047(2,262)}{(3-1)4,6805 + 8,7047} = 2,141;$$

$$HCP = 2,141 \sqrt{\frac{2[(3-1)4,6805 + 8,7047]}{4(3)}} = 2,141(1,735) = 3,7 \text{ т/0,4 га.}$$

HCP для средних сроков уборки при одних или разных S и H . Например, $P_1S_1H_1 - P_2S_1H_1$ или $P_1S_2H_1 - P_2S_1H_1$.

$$HCP = t_{abc} \sqrt{\frac{2 [b(c-1)Ec + (b-1)Eb + Ea]}{rbc}},$$

где $t_{abc} = 2,242$ (см. формулу из таблицы 9.5).

$$\begin{aligned} HCP &= 2,242 \sqrt{\frac{2 [2(3-1)4,6805 + (2-1)7,7047 + 18,6264]}{4(2)3}} = \\ &= 2,242 (1,959) = 4,4 \text{ т/0,4 га.} \end{aligned}$$

HCP для H -средних при одном уровне S . Например, $S_1H_1 - S_1H_2$.

$$HCP = t_c \sqrt{\frac{2Ec}{ra}} = 2,028 \sqrt{\frac{2(4,6805)}{4(3)}} = 2,028 (0,883) = 1,8 \text{ т/0,4 га.}$$

HCP для S -средних при одном или разном уровне H . Например, $S_1H_1 - S_2H_1$ или $S_1H_1 - S_2H_2$.

$$\begin{aligned} HCP &= t_{bc} \sqrt{\frac{2[(c-1)Ec + Eb]}{rac}} = 2,141 \sqrt{\frac{2[(3-1)4,6805 + 8,7047]}{4(3)3}} = \\ &= 2,141 (1,002) = 2,1 \text{ т/0,4 га.} \end{aligned}$$

ЗАКЛЮЧЕНИЕ

Сплит-сплит-плот схема представляет дальнейшее развитие схемы расщепленных делянок (сплит-плот), где субделянки расщепляют на суб-субделянки для размещения третьего фактора. Дисперсионный анализ данных опытов с такой схемой усложняется, поскольку для оценки эффектов вариантов требуется определить величины трех ошибок. Точность оценки действия фактора, размещаемого на суб-субделянках, а также эффектов взаимодействий, включающих этот фактор, обычно выше, чем оценка по другим компонентам вариантов. Три величины ошибки усложняют оценку частных различий.

СХЕМА РАСЩЕПЛЕННЫХ БЛОКОВ

В этой модификации схемы расщепленных делянок варианты второго фактора накладывают полосами поперек целого повторения вариантов главных делянок. Если главные делянки расположены по схеме латинского квадрата, то варианты второго фактора накладывают полосами поперек целого ряда или столбца главной делянки. Такое расположение часто облегчает проведение технологических операций (полевых работ) на делянках-полосах, но снижает точность сравнения главных эффектов фактора B . Эта схема часто улучшает сравнение AB -взаимодействий, особенно сравнений B -средних для заданного уровня фактора A . Эта схема наиболее эффективна для проведения предварительных исследований в целях выявления первоначального эффекта вариантов. Прежде чем применять ее в своих исследованиях, следует получить консультацию у специалиста, знакомого с ней.

На рисунке 10.1 сравнивается одно повторение схемы расщепленных делянок и схемы расщепленных блоков. Отметьте, что в последней схеме делянки второго фактора расположены поперек главных делянок целого блока, и, таким образом, каждая делянка второго порядка расщепляет блок. Другим термином, применимым к этой схеме, может быть схема — «полоса — делянки», поскольку оба фактора A и B накладывают полосами. Для вариантов обоих факторов проводят независимую рендомизацию в каждом повторении.

Повторение!

	A_3	A_2	A_1	A_5	A_4
B_2	B_3	B_2	B_1	B_5	B_4
B_3	B_1	B_5	B_4	B_2	B_3
B_4	B_5	B_4	B_3	B_1	B_2

Расщепленные делянки

Повторение!

	A_3	A_2	A_1	A_5	A_4
B_2	B_2	B_2	B_2	B_2	B_2
B_3	B_3	B_3	B_3	B_3	B_3
B_4	B_4	B_4	B_4	B_4	B_4
B_1	B_1	B_1	B_1	B_1	B_1

Расщепленный блок

Рис. 10.1. Сравнение одного повторения схемы расщепленных делянок и схемы расщепленного блока. Эксперимент включает пять вариантов (градаций) фактора A и четыре варианта фактора B . В схеме расщепленных делянок для B -вариантов проводят независимую рендомизацию внутри каждой делянки фактора A , тогда как в схеме расщепленных блоков B -варианты занимают полосы поперек целого блока делянок фактора A .

Таблица 10.1. Степени свободы для источников вариации схемы расщепленных делянок и расщепленных блоков, представленных на рисунке 10.1

Источники вариации *	Степени свободы		
	формула	расщепленные	
		делянки	блоки
Субделянки	$rab-1$	79	79
Главные делянки	$ra-1$	19	19
Блоки	$r-1$	3	8
Фактор A	$a-1$	4←	4←
MP-ошибка	$(r-1)(a-1)$	12←	12←
Фактор B	$b-1$	3←	3←
$A \times B$	$(a-1)(b-1)$	12←	12←
Ошибка полос-делянок	$(r-1)(b-1)$	↓	9←
Ошибка субделянок	$(r-1)(a-1)(b-1) \rightarrow 45$	ошибка b**	36←

* $r=4$ повторения, $a=5$ градаций фактора A, $b=4$ градаций фактора B. Скобки и отстрелки указывают необходимые ошибки (соответствующие степени свободы) для расчетов F-критериев.

** В данном случае объединяет df для ошибок b и c схемы расщепленных блоков.

В таблице 10.1 показано расщепление общего числа степеней свободы для обеих схем рисунка 10.1 с учетом четырех повторений для каждой из них. Отметим требование схемы расщепленных блоков с разделением ошибки B схемы расщепленных делянок на две ошибки, что уменьшает число степеней свободы для оценки главных эффектов вариантов фактора B. Однако наложение второго фактора полосами через главные делянки устраняет часть вариации из ошибки b схемы расщепленных делянок, поскольку из нее выделяется еще ошибка c для схемы расщепленных блоков, которая вследствие своей меньшей величины часто обеспечивает более точную F-оценку эффектов взаимодействий.

На рисунке 10.2 представлена схема опыта по изучению действия доз азотного удобрения на урожайность корнеплодов сахарной свеклы при различных сроках уборки. Главные делянки, расположенные по схеме латинского квадрата 4×4 , предназначены для испытания четырех доз азотного удобрения. Варианты второго фактора предназначены для изучения пяти сроков уборки. Субделянки каждого срока уборки расположены полосами по всей длине главных делянок. Для каждого столбца делянок приводится независимая рендомизация сроков уборки по полосам. Уборку производить легче, когда делянки, предназначенные для определенного срока уборки, образуют одну непрерывную полосу столбца. Такое расположение, однако, вызывает необходимость расчета отдельной ошибки для оценки главных эффектов сроков уборки. Урожайность корнеплодов для каждой субделянки приводится на рисунке 10.2 вместе с суммами для главных делянок, рядов, столбцов и полос, предназначенных для разных сроков уборки. Эти данные наряду с суммами по вариантам таблицы 10.2 требуются для вычисления сумм квадратов при дисперсионном анализе таблицы 10.3. Расчеты следуют непосредственно за таблицей.

Столбцы																						Y _i
																						Ряды
																						↓
																						439,8
																						422,3
																						384,5
																						353,4
																						Y = 1600,0

Рис. 10.2. Общий вид схемы эксперимента с сахарной свеклой, урожайность (в тоннах корнеплодов на 0,4 га) по делянкам и суммы. Варианты главных делянок, предназначенные для установления влияния на урожайность доз азотного удобрения (из расчета 1,12 кг/га), расположены по схеме латинского квадрата 4×4 . Варианты субделянок — пять сроков уборки с трехнедельным интервалом. Отметьте, что уборку урожая на всех удобряемых азотом (N) делянках предусмотрено проводить в одинаковые сроки; таким образом, каждый столбец главных делянок становится «расщепленным блоком».

Период от посева до уборки с первого (H_1) по пятый (H_5) сроки составляет 20, 23, 26, 29 и 32 недели соответственно. Любая субделянка может быть обозначена символом Y_{ijkl} , где i — ряд ($r=4$); j — столбец ($c=4$), k — дозы азота ($n=4$) и l — сроки уборки ($h=5$). На схеме цифры в скобках обозначают сумму урожая при указанном уровне внесения азота (N).

ДИСПЕРСИОННЫЙ АНАЛИЗ

Степени свободы для источников вариации таблицы 10.3 приводятся ниже. Обратите внимание, что число рядов обозначено как ($r=4$), столбцов ($c=4$), уровней азота ($n=4$) и сроков уборки ($h=5$).

Таблица 10.2. Суммы по вариантам и средние для эксперимента (рис. 10.2) с сахарной свеклой

N-дозы азота, кг/0,4 га	Сроки уборки					
	1	2	3	4	5	$Y_{..k}$
Суммы ($Y_{..h}$)						
0	22,0	47,4	81,4	69,8	78,1	276,4
80	39,4	67,9	85,6	105,0	110,1	408,0
160	40,7	74,4	91,9	120,1	129,3	456,4
320	37,9	77,5	96,6	122,1	125,1	459,2
$Y_{...1}$	140,0	267,2	336,2	417,0	440,6	$Y_{...} = 1600,0$
Средние						
0	5,5	11,8	15,3	17,4	19,0	
80	9,8	17,0	21,4	26,2	27,5	
160	10,2	18,6	23,0	30,0	32,3	
320	9,5	19,4	24,2	30,5	31,3	

Субделянки — $(rch-1)=4(4)(5)-1=79$.

Главные делянки — $(rc-1)=4(4)-1=15$.

Ряды — $(r-1)=4-1=3$.

Столбцы — $(c-1)=4-1=3$.

Уровни азота $(n-1)=4-1=3$.

Ошибка $a=(r-1)(c-1)-(n-1)=3(3)-3=6$.

Сроки уборки — $(h-1)=(5-1)=4$.

Ошибка — $b=(c-1)(h-1)=3(4)=12$.

Совокупность уровней азота и сроков уборки — $(N \times H) = (n-1)(h-1)=3(4)=12$.

Ошибка $c=(c-1)(n-1)(h-1)=3(3)(4)=36$.

Таблица 10.3. Сводная таблица дисперсионного анализа. Схема расщепленных блоков

Источники вариации	df	SS	MS	$F_{\text{факт}}$	$F_{\text{теор}}$	
					5%	1%
Субделянки	79	5542,680				
Главные делянки	15	1503,720				
Ряды	3	224,657	74,886			
Столбцы	3	58,063	19,354			
N-дозы азота	3	1101,328	367,109	18,41	4,76	9,78
Ошибка a, RC-N	6	119,672	19,945			
H-сроки уборки	4	3710,765	927,691	111,92	3,26	5,41
Ошибка b, CH	12	99,467	8,289			
$N \times H$	12	157,147	13,096	6,59	2,03	2,72
Ошибка c, C($N \times H$)	36	71,581	1,988			

Корректирующий фактор

$$C = \frac{Y^2}{rnh},$$

r — число повторений, n — число градаций азота и h — число уборок.

$$C = \frac{1600^2}{4(4)(5)} = 32\,000,00.$$

Суммы квадратов.

$$SSR = \frac{Y^2_{i..}}{nh} - C = \frac{439,8^2 + \dots + 353,4^2}{4(5)} - C = 32\,224,657 - C = 224,657;$$

$$SSC = \frac{Y^2_{.j.}}{nh} - C = \frac{379,9^2 + \dots + 391,2^2}{4(5)} - C = 32\,058,063 - C = 58,063;$$

$$SSN = \frac{\Sigma Y^2_{..k}}{rh} - C = \frac{276,4^2 + \dots + 459,2^2}{4(5)} - C = 33\,101,328 - C = 1101,328;$$

$$\begin{aligned} SS(\text{главные делянки}) &= \frac{\Sigma Y^2_{ijk}}{h} - C = \\ &= \frac{107,1^2 + \dots + 92,1^2}{5} - C = 33\,503,720 - C = 1503,720; \end{aligned}$$

$$\begin{aligned} SS(Ea) &= SS(\text{главные делянки}) - SSR - SSC - SSN = \\ &= 1503,720 - 224,657 - 58,063 - 1101,328 = 119,672; \end{aligned}$$

$$\begin{aligned} SSH &= \frac{\Sigma Y^2_{...1}}{rn} - C = \\ &= \frac{140,0^2 + \dots + 440,6^2}{4(4)} - C = 35\,710,765 - C = 3610,765; \end{aligned}$$

$$\begin{aligned} SS(Eb) &= \frac{\Sigma Y^2_{.j.1}}{r} - C - SSC - SSH = \\ &= \frac{95,7^2 + \dots + 69,7^2}{4} - C - SSC - SSH = \\ &= 35\,868,295 - C - SSC - SSH = 99,467; \end{aligned}$$

$$\begin{aligned} SS(N \times H) &= \frac{\Sigma Y^2_{..k1}}{r} - C - SSN - SSH = \\ &= \frac{22,0^2 + \dots + 125,1^2}{4} - C - SSN - SSH = \\ &= 36\,969,240 - C - SSN - SSH = 157,147; \end{aligned}$$

$$\begin{aligned} SS(\text{субделянки}) &= \Sigma Y^2_{ijk1} - C = \\ &= 26,4^2 + 29,3^2 + \dots + 15,9^2 - C = 37\,542,68 - C = 5542,680; \\ SS(Ec) &= SS(\text{субделянки}) - SS(M\text{-делянки}) - SSH - SS(Eb) - \\ SS(N \times H) &= 5542,680 - 1503,720 - 3710,765 - 99,467 - \\ &\quad - 157,147 = 71,581. \end{aligned}$$

Средние квадраты. Средние квадраты получают путем деления суммы квадратов на соответствующее число степеней например:

$$MS(Ec) = \frac{SS(Ec)}{df(Ec)} = \frac{71,581}{36} = 1,988.$$

Клавиша стандартного отклонения. Необходимость в корректирующей величине (члене) отпадает при использовании калькуляторов с программой расчета стандартного отклонения по формуле:

$$s = \sqrt{\Sigma(Y_i - \bar{Y})^2 / (r - 1)}.$$

При вычислении любого среднего квадрата в машину вводят соответствующие суммы, получают s^2 и делят на число дат, вошедших в каждую сумму. Например, чтобы вычислить SSR , вводим 439,8; 422,3; 384,5 и 353,4; нажимаем клавишу, получаем $s=38,7003$; квадрат этого числа дает 1497,7133; делим на 20, получаем $74,886=MSR$; умножаем на 3 (df для R), получаем $224,657=SSR$.

F-критерии и оценка частных различий. F -значения получают путем деления средних квадратов на соответствующие величины ошибок: Ea — для азота; Eb — для сроков уборки и Ec — для $N \times H$ -взаимодействия.

Очень существенное значение $F_{\text{фактического}}$ для $N \times H$ -взаимодействия указывает на различную отзывчивость сахарной свеклы на азотное удобрение в зависимости от сроков уборки. Понимание характера этого взаимодействия дает основу интерпретации результатов опыта. При расчленении доз азота и сроков уборки на полиномиальные (многочленные) сравнения трендов мы можем также расчленить суммы квадратов $N \times H$ -взаимодействия для познания природы взаимодействия. Получение полиномиальных коэффициентов для расчленения вариации разноинтервальных доз азота — не легкая задача, что является одной из причин предпочтительного планирования доз вариантов с одинаковыми интервалами (шагом) между ними. Некоторые из таких коэффициентов даны в таблице А.11 (приложение), причем коэффициенты для наших четырех доз азота берем из-под серий 0, 1, 2, 4. Коэффициенты для расчленения вариации пяти сроков уборки с равновеликими интервалами находим в таблице А.11 (приложение) под $n=5$. Коэффициенты N -доз азота и сроков уборки распределены по вариантам нашего эксперимента в таблице 10.4. Коэффициенты для взаимодействия получают путем перемножения. Например, коэффициентами для $NL \times HL$ будут: $7(-2)=14$, $-7(-1)=7$, $-7(0)=0$ и т. д. Примечательно, что коэффициенты для всех сравнений отвечают двум правилам ортогональности, данным в 6-й главе, и, таким образом, суммы квадратов сравнений, опирающихся на одну степень свободы, будут дополнять каждую из сумм квадратов, имеющих множество степеней свободы, что контролирует достоверность расчетов.

При одной степени свободы суммы квадратов вычисляют по формулам $SS = \Sigma(c_i Y_i)^2 / (r \Sigma c_i^2)$, где c_i — коэффициенты сравнений, Y_i — суммы по вариантам и r — число дат каждой суммы. Например, сумма квадратов для линейной реакции на азот будет равна:

$$SS(NL) = \frac{[-7(22,0) - 7(47,4) + \dots + 9(125,1)]^2}{4(700)} = 730,7301.$$

Таблица 10.4. Ортогональные коэффициенты для расчленения суммы квадратов квадратов и их статистическая существенность

Сравнения трендов ^a	N_0					N_{80}				
	1	2	3	4	5	1	2	3	4	5
	22,0	47,4	61,1	69,8	76,1	39,4	67,9	85,6	105,0	110,1

По N -дозам азота:

линейное	-7	-7	-7	-7	-7	-3	-3	-3	-3	-3
квадратичное	7	7	7	7	7	-4	-4	-4	-4	-4
остаточное	-3	-3	-3	-3	-3	8	8	8	8	8

По H -срокам уборки:

линейное	-2	-1	0	1	2	-2	-1	0	1	2
квадратичное	2	-1	-2	-1	2	2	-1	-2	-1	2
кубичное	-1	2	0	-2	1	-1	2	0	-2	1
остаточное	1	-4	6	-4	1	1	-4	6	-4	1

По величинам $N \times H$

$NL \times HL$	14	7	0	-7	-14	6	3	0	-3	-6
$NL \times HQ$	-14	7	14	7	-14	-6	3	6	3	-6
$NL \times HC$	7	-14	0	14	-7	3	-6	0	6	-3
$NL \times HR$	-7	28	-42	28	-7	-3	12	-18	12	-3
$NQ \times HL$	-14	-7	0	7	14	8	4	0	-4	-8
$NQ \times HQ$	14	-7	-14	-7	14	-8	4	8	4	-8
$NQ \times HC$	-7	14	0	-14	7	4	-8	0	8	-4
$NQ \times HR$	7	-28	42	-28	7	-4	16	-24	16	-4
$NR \times HL$	6	3	0	-3	6	-16	-8	0	8	16
$NR \times HQ$	-6	3	6	3	-6	16	-8	-16	-8	16
$NR \times HC$	3	-6	0	6	-3	-8	16	0	-16	8
$NR \times HR$	-3	12	-18	12	-3	8	-32	48	-32	8

^a L — линейное, Q — квадратичное, R — остаточное и C — кубичное.

^b Для определения F -критериев используйте соответствующие средние квадраты ошибок действий.

Другие суммы квадратов с одной степенью свободы вычисляют подобным образом и вносят в таблицу 10.4.

F -критерии получают с использованием соответствующих средних квадратов ошибок из таблицы 10.3: ошибка a для N -сравнений по азоту, ошибка b — для H -сравнений по срокам уборки и ошибка c — для $N \times H$ -сравнений взаимодействий. Например, F -критерий для выявления эффекта взаимодействия $NL \times HL = 98,889/1,988 = 49,75$, что значительно превосходит величину 7,39 из таблицы А.3 (приложение) на 1%-ном уровне значимости ($df=1$ и 36). Следует указать, что лишь два компонента: $NL \times HL$ и $NQ \times HL$ взаимодействия $N \times H$ имеют статистическую значимость.

Для того чтобы продолжить рассуждения и соответствующим образом представить результаты этого эксперимента, необходимо неко-

N_{180}					N_{320}					df	Суммы квадратов	Статистическая существенность сравнений, опирающихся на 1 df 6
1	2	3	4	5	1	2	3	4	5			
40,7	74,4	91,9	120,1	129,3	37,9	77,5	96,6	122,1	125,1			
1	1	1	1	1	9	9	9	9	9	3	1101,328	
-8	-8	-8	-8	-8	5	5	5	5	5	1	730,730	**
-6	-6	-6	-6	-6	1	1	1	1	1	1	359,593	**
											11,005	Несущественно
										4	3710,765	
-2	-1	0	1	2	-2	-1	0	1	2	1	3525,006	**
2	-1	-2	-1	2	2	-1	-2	-1	2	1	166,980	**
-1	2	0	-2	1	-1	2	0	-2	1	1	0,006	Несущественно
1	-4	6	-4	1	1	-4	6	-4	1	1	18,772	
										12	157,147	
-2	-1	0	1	2	-18	-9	0	9	18	1	98,899	**
2	-1	-2	-1	2	18	-9	-18	-9	18	1	5,194	Несущественно
-1	2	0	-2	1	-9	18	0	-18	9	1	1,015	То же
1	-4	6	-4	1	9	-36	54	-36	9	1	5,043	»
16	8	0	-8	-16	-10	-5	0	5	10	1	38,660	**
-16	8	16	8	-16	10	-5	-10	-5	10	1	2,595	Несущественно
8	-16	0	16	-8	-5	10	0	-10	5	1	1,359	То же
-8	32	-48	32	-8	5	-20	30	-20	5	1	1,557	»
12	6	0	-6	-12	-2	-1	0	1	2	1	1,536	»
-12	6	12	6	-12	2	-1	-2	-1	2	1	0,755	»
6	-12	0	-12	6	-1	2	0	-2	1	1	0,384	»
-6	24	-36	24	-6	1	-4	6	-4	1	1	0,151	»

из таблицы 10.3: Ea — для N -доз азота; Eb — для H -сроков уборки и Ec — для $N \times H$ -взаим.

торое знание полиномиальной и множественной регрессии, поэтому отложим это до главы 16. На данном этапе отметьте, что общая сумма шести существенных сумм квадратов с одной степенью свободы составляет 99% от суммы квадратов всех 19 сочетаний вариантов, т. е. $(730,730 + 359,593 + 3525,006 + 166,980 + 98,899 + 38,659)/(1101,328 + 3710,765 + 157,147) = 0,990$.

СТАНДАРТНЫЕ ОШИБКИ

Стандартные ошибки даны в таблице 10.5 с целью их использования при оценке частных различий с помощью HSP и многоранговых критериев.

Таблица 10.3. Стандартные ошибки средних для схемы расщепленных блоков

Сравниваемые средние	Стандартные ошибки	<i>t</i> -значения
А-средние	$\sqrt{\frac{Ea}{rb}}$	t_a
В-средние	$\sqrt{\frac{Eb}{ra}}$	t_b
А-средние для одного или разных уровней В	$\sqrt{\frac{(b-1) Ec + Ea}{rb}}$	$t_{ac} = \frac{(b-1) Ec (t_c) + Ea (t_a)}{(b-1) Ec + Ea}$
В-средние для одного уровня А	$\sqrt{\frac{(a-1) Ec + Eb}{ra}}$	$t_{bc} = \frac{(a-1) Ec (t_c) + Eb (t_b)}{(a-1) Ec + Eb}$

* А — варианты фактора, изучаемого на главных делянках, В — варианты фактора, изучаемого на субделянках; а, b и r — число вариантов (градаций) главных делянок, субделянок и повторность соответственно. Ea, Eb, Ec — средние квадраты ошибок; t_a , t_b , t_c — табличные величины для d_f , Ea, Eb, Ec соответственно. НСР вычисляются по формуле $НСР = t \sqrt{2} s_{\bar{y}}$ и многогранговый критерий Дункана — $D = R (НСР)$.

Для иллюстрации назначения таблицы 10.5 мы вычислим НСР для сравнения средних взаимодействий таблицы 10.2. Однако ни НСР, ни многогранговые критерии не будут соответствовать требованиям оценки частных различий этого эксперимента в той мере, в какой это можно сделать с помощью более сильного метода, описанного в 16-й главе.

НСР (5%) при сравнении N-средних азота для одного или разных сроков уборки (H):

$$НСР_{05} = t_{ac} \sqrt{\frac{2 [(b-1) Ec + Ea]}{rb}};$$

$$t_{ac} = \frac{(b-1) Ec (t_c) + Ea (t_a)}{(b-1) Ec + Ea} = \frac{(5-1) 1,988 (2,028) + 19,945 (2,477)}{(5-1) 1,988 + 19,945} = \frac{64,9321}{27,897} = 2,328;$$

$$НСР = 2,328 \sqrt{\frac{2 [(5-1) 1,988 + 19,945]}{4 (5)}} = 2,328 (1,670) = 3,4 \text{ т/0,4 га.}$$

НСР (5%) при сравнении H-средних сроков уборки с одним уровнем азота (N):

$$НСР_{05} = t_{bc} \sqrt{\frac{2 [(a-1) Ec + Eb]}{ra}};$$

$$t_{bc} = \frac{(a-1) Ec (t_c) + Eb (t_b)}{(a-1) Ec + Eb} = \frac{(4-1) 1,988 (2,028) + 8,289 (2,179)}{(4-1) 1,988 + 8,289} = \frac{30,1567}{14,253} = 2,116;$$

$$НСР = 2,116 \sqrt{\frac{2 [(4-1) 1,988 + 8,289]}{4 (4)}} = 2,116 (1,335) = 2,8 \text{ т/0,4 га.}$$

ЗАКЛЮЧЕНИЕ

В схеме расщепленных блоков блоки делянок фактора A расщепляются таким образом, что каждый вариант фактора B размещается на непрерывных полосах, проходящих через весь блок. Для каждого блока делянок фактора A проводится независимая рендомизация вариантов фактора B . Преимущества этой схемы заключаются в уменьшении затрат труда на технологические операции, связанные с проведением эксперимента, а также в большей точности оценки $A \times B$ -взаимодействия.

К недостаткам этой схемы эксперимента относят: пониженную точность определения эффектов фактора B , большую комплексность расчетов и затруднения в оценке частных различий.

ИСПОЛЬЗОВАНИЕ ПОВТОРНЫХ УЧЕТОВ В КАЧЕСТВЕ СУБДЕЛЯНОК

Принцип схемы расщепленных делянок может быть использован в опытах, где проводят последовательные наблюдения за одной и той же целой экспериментальной единицей в течение определенного периода времени. Например, в опытах с удобрениями или сортами многолетних трав уборка может производиться несколько раз в год и (или) в течение двух или большего числа лет. Делянки, на которых размещаются варианты, могут быть названы главными делянками, а несколько уборок, производимых на них, субделянками. Однако такая субделянка отличается от обычной субделянки тем, что данные получают со всей площади главной делянки, а не с ее части, как в случае схемы расщепленных делянок.

При анализе данных единичных учетов или сумм учетов по всем срокам с главных делянок дополнительных проблем не возникает. Однако F -критерии для оценки эффектов последовательных наблюдений и взаимодействия вариантов главных делянок и отдельных наблюдений, вариация которых отклоняется от типичной для F -оценки распределения, могут дать завышенное число существенных эффектов.

Ниже приводится последовательность расчетов и разъяснение интерпретации результатов таких опытов. В качестве примера рассматриваются данные эксперимента по урожайности сена различных сортов люцерны. В опыте изучаются четыре сорта, случайно размещенных внутри пяти организованных блоков. Для упрощения анализа мы возьмем лишь четыре укоса: два в первой половине и два во второй половине вегетации и только за два года.

СТАТИСТИЧЕСКИЙ АНАЛИЗ КАЖДОЙ СЕРИИ НАБЛЮДЕНИЙ

Необходимо выполнить дисперсионный анализ ($ANOVA$) для каждого укоса. Даты и их суммы в исходной таблице дисперсионного анализа (табл. 11.1) позволяют выполнить анализ как по данным за каждый укос, так и по всем укосам за весь год.

Итоги дисперсионного анализа каждого укоса приводятся в таблице 11.2. Степени свободы и сумма квадратов для сортов расчленены так потому, что сорт 1 и сорт 2 генетически тесно связаны, поскольку

Таблица 11.1. Урожайность сена люцерны (т/0,4 га) первого года в опыте, заложенном по схеме рендомизированных блоков с четырьмя сортами ($v=4$), пятью блоками ($b=5$) и четырьмя укосами ($h=4$) при изучении сортов люцерны

Сорта (i)	Укосы (k)	Блоки (j)						$Y_{i..}$	$\bar{Y}_{i..}$
		1	2	3	4	5	$Y_{i.k}$		
1	1	2,69	2,40	3,23	2,87	3,27	14,46	2,89 ^a	
2	1	2,87	3,05	3,09	2,90	2,98	14,89	2,98	
3	1	3,12	3,27	3,41	3,48	3,19	16,47	3,29	
4	1	3,23	3,23	3,16	3,01	3,05	15,68	3,14	
$Y_{.j1}$		11,91	11,95	12,89	12,26	12,49	61,50 = $Y_{..1}$		
1	2	2,74	1,91	3,47	2,87	3,43	14,42	2,88	
2	2	2,50	2,90	3,23	2,98	3,05	14,66	2,93	
3	2	2,92	2,63	3,67	2,90	3,25	15,37	3,07	
4	2	3,50	2,89	3,39	2,90	3,16	15,84	3,17	
$Y_{.j2}$		11,66	10,33	13,76	11,65	12,89	60,29 = $Y_{..2}$		
1	3	1,67	1,22	2,29	2,18	2,30	9,66	1,93	
2	3	1,47	1,85	2,03	1,82	1,51	8,68	1,74	
3	3	1,67	1,42	2,81	1,51	1,76	9,17	1,83	
4	3	2,60	1,92	2,36	1,92	2,14	10,94	2,19	
$Y_{.j3}$		7,41	6,41	9,49	7,43	7,71	38,54 = $Y_{..3}$		
1	4	1,92	1,45	1,63	1,60	1,96	8,56	1,71	
2	4	2,00	2,03	1,71	1,60	1,96	9,30	1,86	
3	4	2,03	1,96	1,85	1,82	2,40	10,06	2,01	
4	4	2,07	1,89	1,92	1,82	1,78	9,48	1,90	
$Y_{.j4}$		8,02	7,33	7,11	6,84	8,10	37,40 = $Y_{..4}$		
Сорта $\times$ суммы блоков (главные делянки $Y_{i..}$)									
		$Y_{i..}$					$Y_{i..}$	$\bar{Y}_{i..}$	
1		9,02	6,98	10,62	9,52	10,96	47,10	9,42 ^b	
2		8,84	9,83	10,06	9,30	9,50	47,53	9,51	
3		9,74	9,28	11,74	9,71	10,60	51,07	10,21	
4		11,40	9,93	10,83	9,65	10,13	51,94	10,39	
$Y_{.j.}$		39,00	36,02	43,25	38,18	41,19	197,64 = $Y_{...}$		

^a В т/0,4 га за укос.

^b В т/0,4 га за год.

	df	Укос 1		Укос 2		Укос 3		Укос 4	
		SS	MS	SS	MS	SS	MS	SS	MS
19	1	1,1801		3,1045		3,3016		0,8376	
4	0,1651	0,0413	1,7249	0,4312	1,2562	0,3140	0,3113	0,0778	
3	0,4729	0,1576	0,2547	0,0849	0,5660	0,1887	0,2295	0,0765	
и 3+4	1	0,3920	0,3920 *	0,2268	0,2268	0,1567	0,1567	0,1411	0,1411 *
и 2	1	0,0185	0,0185	0,0058	0,0058	0,0960	0,0960	0,0548	0,0548
и 4	1	0,0624	0,0624	0,0221	0,0221	0,3133	0,3133	0,0336	0,0336
Ошибка	12	0,5421	0,0452	1,1249	0,0937	1,4794	0,1233	0,2969	0,0247

* Отношение MS к MS ошибки превышает табличные значения F на 5%-ном уровне значимости, табличное $F_{0,05}$ (при 1 и 12 df) = 4,75. V — сорта: v=4; G — укосы, c=4; B — блоки, b=5.

ку сорт 2 выведен на основе сорта 1. Процедура расчетов при анализе данных одного срока уборки приводится под таблицей 11.2 на примере 1-го укоса.

$$C = \frac{Y_{..1}^2}{vnb} = \frac{61,50^2}{4(4)5} = 189,1125;$$

$$SSB = \frac{\sum Y_{i..1}^2}{v} - C = \frac{11,91^2 + \dots + 12,49^2}{4} - C = 0,1651;$$

$$SSV = \frac{\sum Y_{i..1}^2}{b} - C = \frac{14,46^2 + \dots + 15,68^2}{5} - C = 0,4729;$$

$$SS(V1+2 \text{ и } 3+4) = (14,46 + 14,89 - 16,47 - 15,68)^2 / 5(4) = 0,3920.$$

Отметьте, что расчеты этой суммы квадратов с единичной степенью свободы и двух последующих проводятся по формуле $SS = (\sum c_i Y_i)^2 / (r \sum c_i^2)$. Для этих трех расчетов все значения c_i равны +1 или -1.

$$SS(V1 \text{ и } V2) = \frac{(14,46 - 14,89)^2}{5(2)} = 0,0185;$$

$$SS(V3 \text{ и } V4) = \frac{(16,47 - 15,68)^2}{5(2)} = 0,0624;$$

$$SS(\text{общая}) = 2,69^2 + \dots + 3,05^2 - C = 1,1801;$$

$$SS(\text{ошибки}) = SS(\text{общая}) - SSB - SSV = 0,5421.$$

Средние квадраты получают путем деления SS на соответствующее число степеней свободы, например, $MSV = 0,4729/3 = 0,1576$. F-критерии (фактические) для оценки эффектов сортов находят как частное от деления MS одного из компонентов сортов на средний квадрат ошибки для конкретного укоса, например, для 1-го укоса. F-оценка для сравнения сортов 1+2 и 3+4 = $0,3920/0,0452 = 8,67$. Средние урожаи 3-го и 4-го сортов больше, чем урожаи 1-го и 2-го сортов, по всем укосам — от 1-го до 4-го — на 0,28; 0,27; 0,18 и 0,17 т/0,4 га соответственно. Поскольку разность (V1+2) — (V3+4) статистически существенна для 1-го и 4-го укосов, то будет логичным предположение о

Таблица 11.3. Результаты дисперсионного анализа урожая сортов люцерны первого года

Источники вариации	df	SS	MS	F _{факт}	F _{табл}	
					5%	1%
Субделянки, SP	79	34,8690				
Главные делянки, MP	19	5,0769				
Блоки, B	4	1,9386	0,4846			
Сорта, V	3	0,9014	0,3005	1,61	3,49	5,95
1+2 и 3+4	1	0,8778	0,8778	4,71	4,75	9,33
1 и 2	1	0,0046	0,0046			
3 и 4	1	0,0189	0,0189			
MP-ошибка, BV	12	2,2369	0,1864			
Укосы, C	3 (1)	26,4452	8,8151	155,2	4,49 ^a	8,53 ^a
V×C	9 (3)	0,6217	0,0690	1,21	3,24 ^a	5,29 ^a
SP-ошибка, BC+B(V×C)	48 (16)	2,7252	0,0568			

^a F_{табличный} определяется по числу степеней свободы, указанному в скобках.

наличии реальных различий этих сортов также для 2-го и 3-го укосов, даже если их $F_{факт}$ незначительна на 5%-ном уровне.

Анализ данных за год. Результаты дисперсионного анализа годовых данных приведены в таблице 11.3. Последовательность расчетов, идентичная дисперсионному анализу данных опытов, заложенных по схеме расщепленных делянок, приводится ниже.

$$C = \frac{Y_{...}^2}{vnb} = \frac{197,64^2}{4(4)5} = 488,2696.$$

Обратите внимание, что использование величины h в знаменателе приводит данные к средним на один укос.

$$SSB = \frac{\sum Y_{.j}^2}{vh} - C = \frac{39,00^2 + \dots + 41,19^2}{4(4)} - C = 420,2082 - C = 1,9386.$$

Если вы имеете калькулятор с программой расчета $s = \sqrt{(Y_i - \bar{Y})^2 / (r-1)}$, то SSB и другие суммы квадратов определяют следующим образом: вводят соответствующие суммы, получают s , возводят s в квадрат и делят на число сортов введенной суммы. В результате получают величину MS, умножение которой на соответствующие степени свободы дает SS. Например s^2 блоковых сумм = 7,75437. Делим на 16 = 0,48465 = MSB. Умножаем на 4 = 1,9386 = SSB.

$$SSV = \frac{\sum Y_{i..}^2}{hb} - C = \frac{47,10^2 + \dots + 51,94^2}{4(5)} - C = 489,1710 - C = 0,9014.$$

$$SS(V1+2 \text{ и } 3+4) = \frac{(\sum c_i Y_{i..})^2}{bh \sum c_i^2}.$$

Отметьте, что все коэффициенты для сравнений (c_i) равны +1 или -1.

$$SS(V1+2 \text{ и } 3+4) = \frac{(47,10+47,53-51,07-51,94)^2}{5(4)4} = 0,8778;$$

$$SS(MP) = \left(\frac{\sum Y_{ij}^2}{h} \right) - C = \frac{9,02^2 + \dots + 10,13^2}{4} - C = 5,0769;$$

$$SS(MP \text{ ошибки}) = SS(MP) - SSB - SSV - 2,2369;$$

$$SSC = \frac{\sum Y_{.k}^2}{bh} - C = \frac{61,50^2 + \dots + 37,40^2}{5(4)} - C = 26,4452;$$

$$SS(V \times H) = \frac{\sum Y_{i.k}^2}{b} - C - SSV - SSH = \\ = \frac{14,46^2 + \dots + 9,48^2}{5} - C - SSV - SSH = 0,6217;$$

$$SS(\text{субделянок}) = \sum Y_{ijk}^2 - C = (2,69^2 + \dots + 1,78^2) - C = \\ = 523,1386 - C = 34,8690.$$

Отметьте также, что $SS(SP) = SS(\text{общая})$ для $H_1 + \dots + SS(\text{общее}) H_4 + SSH = 1,1801 + \dots + 0,8376 + 26,4452 = 34,8690;$

$$SS(SP \text{ ошибки}) = SS(SP) - SS(MP) - SSH - SS(V \times H) = \\ = 34,8690 - 5,0769 - 26,4452 - 0,6217 = 2,7252.$$

Средние квадраты получают путем деления сумм квадратов на степени свободы. Например, $MSV = 0,9014/3 = 0,3005$.

F-величины и оценка частных различий (выделение существенных средних). Квадратные скобки со стрелками, соединяющие средние квадраты таблицы 11.3, выделяют используемые ошибки для расчета $F_{\text{факт}}$. F-значения для укусов и взаимодействия $V \times H$ должны быть достаточно велики, чтобы делать заключение о наличии реальных различий. Испытанный способ, рекомендуемый многими статистиками в качестве проверки существенных эффектов, требует больших значений F. Степени свободы для сроков укусов надо расчленить на степени свободы для H , $V \times H$ и ошибки субделянок (величины таблицы 11.3, заключенные в скобки) и брать табличное значение F на основе оставшихся степеней свободы (в таблице 11.3 величины подчеркнуты). Большие значения F для укусов не оставляют сомнений в отношении реальных различий между средними сроками уборки. В опыте не выявлено существенного влияния взаимодействия $V \times H$ на урожайность.

Следует указать, что наибольшие величины вариабельности между сортами получены при сравнении сортов $V1+2$ и $V3+4$ и что разность этого сравнения почти существенна на 5%-ном уровне ($F_{\text{факт}} \approx F_{\text{теор}}$).

Стандартные ошибки. Стандартные ошибки, используемые для НСР и многогранговых критериев, являются теми же, что и для обычной схемы расщепленных делянок в отношении средних эффектов фактора, изучаемого на главных делянках (в данном случае сорта), однако стандартные ошибки, используемые для оценки средних повторных учетов (сроки укусов) и взаимодействия фактора главных делянок и повторных учетов ($V \times H$), отличаются от применяемых в схеме расщепленных делянок. В последующем мы в большей или меньшей

мере будем придерживаться ранее описанной схемы расчетов (Steel and Torrie, 1960). Использование стандартной ошибки при вычислении HCP и многогранового критерия Дункана подробно изложено в главе 6. Конечные формулы имеют следующий вид: $HCP = ts_{\bar{d}}$ и $D = R(HCP)$.

1. Сравнение двух средних фактора A , $V_1 - V_2$:

а) на основе данных за один укос: $s_{\bar{d}} = \sqrt{\frac{2(MP \text{ ошибка})}{bh}}$,

$$s_{\bar{d}} = \sqrt{\frac{2(0,1864)}{5(4)}} = 0,1363, \quad HCP_{0,05} = 2,179(0,1363) = 0,30 \text{ т/0,4 га};$$

б) на основе суммы укосов за год $s_{\bar{d}} = \sqrt{\frac{2h(MP \text{ ошибка})}{b}}$,

$$s_{\bar{d}} = \sqrt{\frac{2(4)(0,1864)}{5}} = 0,5461, \quad HCP = 2,179(0,5461) = 1,19 \text{ т/0,4 га}.$$

2. Сравнение двух средних фактора B , $H_1 - H_2$:

$$s_{\bar{d}} = \sqrt{\frac{2(SP \text{ ошибка})}{bv}}$$

$$s_{\bar{d}} = \sqrt{\frac{2(0,0568)}{5(4)}} = 0,0754, \quad HCP = 2,120(0,0754) = 0,16 \text{ т/0,4 га},$$

где $t_{0,05}$ берется для 16df.

3. Сравнение двух средних фактора A на одном уровне B , $V_1H_1 - V_2H_1$:

$$s_{\bar{d}} = \sqrt{\frac{2(E_1)}{b}},$$

где E_1 — ошибка для анализа рассматриваемого укоса.

$$\text{Для } H_1s_{\bar{d}} = \sqrt{\frac{2(0,0452)}{5}}, \quad s_{\bar{d}} = 0,1345, \quad HCP = 2,179(0,1345) = 0,29 \text{ т/0,4 га}.$$

4. Сравнение двух средних фактора B для одного или разных уровней фактора A , $V_1H_1 - V_1H_2$ или $V_1H_1 - V_2H_2$:

$$s_{\bar{d}} = \sqrt{\frac{2(E_1 + E_2)}{2b}},$$

где E_1 и E_2 — средние квадраты ошибок для двух укосов усредняются.

$$s_{\bar{d}} \sqrt{\frac{2(0,0452 + 0,0937)}{2(5)}} = 0,1667 \text{ и}$$

$$HCP = 2,179(0,1667) = 0,36 \text{ т/0,4 га}.$$

ОБЪЕДИНЕНИЕ ДАННЫХ ЗА ДВА ИЛИ БОЛЬШЕ ЧИСЛО ЛЕТ

В дополнение к анализу продуктивности сортов за каждый год исследования интересуют их продуктивность в течение ряда лет и возможное взаимодействие сортов с особенностями лет.

Результаты ряда лет при нескольких укосах в год могут быть объединены по схеме сплит-сплит-плот (трехфакторный опыт), причем сорта

Таблица 11.4. Поделяночная урожайность (т/0,4 га) сена за 1-й и 2-й годы в зависимости от сочетания сортов люцерны (отметьте, что данные первого года совпадают с данными нижней части таблицы 11.1)

Годы (i)		Блоки (j)					$Y_{i..}$	$\bar{Y}_{i..}$
		1	2	3	4	5		
1	1	9,02	6,98	10,62	9,52	10,96	47,10	9,42
2	1	8,84	9,83	10,06	9,30	9,50	47,53	9,51
3	1	9,74	9,28	11,74	9,71	10,60	51,07	10,21
4	1	11,40	9,93	10,83	9,65	10,13	51,94	10,39
$Y_{.j1}$		39,00	36,02	43,24	38,18	41,19	$197,64 = Y_{...1}$	
1	2	11,88	11,33	11,81	12,22	10,65	57,89	11,58
2	2	12,15	10,98	12,20	11,30	12,54	59,15	11,83
3	2	12,92	11,95	12,05	11,88	13,19	61,99	12,40
4	2	11,74	11,62	11,54	12,00	11,74	58,64	11,73
$Y_{.j2}$		48,69	45,86	47,60	47,40	48,12	$237,67 = Y_{...2}$	
Сорта $\times$ блоковые суммы (главные делянки Y_{ij})								$\bar{Y}_{ij}$
1		20,90	18,31	22,43	21,74	21,64	104,99	10,50
2		20,99	20,79	22,26	20,60	22,04	106,68	10,67
3		22,66	21,23	23,79	21,59	23,79	113,06	11,31
4		23,14	21,55	22,37	21,65	21,87	110,58	11,06
$Y_{.j.}$		86,69	81,88	90,85	85,58	89,31	$435,31 = Y_{...}$	

будут представлять главные делянки, годы — субделянки, а укусы — суб-субделянки. При этом тройное взаимодействие сорта $\times$ годы $\times$ укусы обычно не имеет первостепенной важности. Анализ в течение года и анализы годовых сумм целых делянок в течение ряда лет — это все, что обычно требуется для заключения о пригодности сортов.

Для иллюстрации процесса объединения годовых сумм поделяночных урожаев сортов за ряд лет воспользуемся лишь двухгодичными данными. Процедура анализа такая же, как и для укусов одного года. Из таблицы 11.4 берут необходимые данные. Обратите внимание, что суммы взаимодействия сорта $\times$ блоки одного года в таблице 11.1 соответствуют данным первого года в таблице 11.4.

Годовые анализы. Годовые анализы — это анализы данных главных делянок за каждый год (табл. 11.3). Поскольку теперь мы желаем иметь поделяночные данные за годы, а не за укусы, как в таблице 11.3, следует умножить суммы квадратов таблицы 11.3 на число укусов, чтобы заполнить таблицу 11.5. Таким образом, дисперсионный анализ первого года будет дополнен следующими суммами квадратов:

$$SSB = (SSB \text{ за 1-й год}) 4 = (1,9386) 4 = 7,7544,$$

$$SSV = (0,9014) 4 = 3,6056,$$

$$SS_{\text{ошибки}} = (2,2369) 4 = 8,9476.$$

Таблица 11.5. Дисперсионный анализ общих поделяночных урожаев за каждый год

Источники вариации	df	1-й год		2-й год	
		SS	MS	SS	MS
Блоки	4	7,7544	1,9386	1,1261	0,2815
Сорта	3	3,6054	2,2018	1,9254	0,6418
1+2 и 3+4	1	3,5112	3,5112	0,6444	0,6444
1 и 2	1	0,0184	0,0184	0,1588	0,1588
3 и 4	1	0,0756	0,0756	0,1122	0,1122
Ошибка	12	8,9476	0,7456	3,7462	0,3120

Аналогичная процедура производится для заполнения второго года таблицы 11.5. Отметим, что во второй год не было установлено существенных эффектов сортов, а большая часть вариации вариантов (сорт) падает на различия между сортами 1+2 и 3+4.

Сложение данных годовых анализов. Таблица 11.6, заполненная на основе данных таблицы 11.4, объединяет суммы по годам за два года по аналогии объединения укосов в течение года (табл. 11. 3). Расчеты

Таблица 11.6. Дисперсионный анализ годовых урожаев за два года

Источники вариации	df	SS	MS	F _{факт}	F _{табл}	
					5%	1%
Субделянки	39	67,1654				
Главные делянки	19	14,0138				
Блоки, B	4	6,1058	1,5264			
Сорта, V	3	4,0323	1,3441	4,16	3,49	5,95
1+2 и 3+4	1	3,5820	3,5820	11,09	4,75	9,33
1 и 2	1	0,1428	0,1428			
3 и 4	1	0,3075	0,3075			
MP-ошибка, VB	12	3,8757	0,3230			
Годы, Y	1	40,0600	40,0600	55,29	4,49	8,53
V×Y	3	1,4985	0,4995			
(V 1+2 и 3+4)×Y	1	0,5736	0,5736			
(V 1 и 2)×Y	1	0,0344	0,0344			
(Y 3 и 4)×Y	1	0,8904	0,8904			
SP-ошибка, BY + B (V×Y)	16	11,5931	0,7246			

V—сорта, v=4; B—блоки, b=4; Y—годы, y=2.

сумм квадратов приведены ниже. Средние квадраты получены как частные от деления сумм квадратов на их степени свободы.

$$C = \frac{\sum Y_{ij}^2}{vby} = \frac{435,31^2}{4(5)2} = 4737,3699;$$

$$SSB = \frac{\sum Y_{.j}^2}{vy} - C = \frac{86,69^2 + \dots + 89,31^2}{4(2)} - C = 6,1058;$$

$$SSV = \frac{\sum Y_{i.}^2}{by} - C = \frac{104,99^2 + \dots + 110,58^2}{5(2)} - C = 4,0323;$$

$$SS(V1+2 \text{ и } 3+4) = \frac{(\sum c_i Y_{i.})^2}{by \sum c_i^2},$$

где c_i — коэффициенты классовых сравнений, которые в этом случае равны +1 и -1.

$$SS(V1+2 \text{ и } 3+4) = \frac{(104,99 + 106,68 - 113,06 - 110,58)^2}{5(2)4} = 3,5820;$$

$$SS(V1 \text{ и } V2) = \frac{(104,99 - 106,68)^2}{5(2)2} = 0,1428;$$

$$SS(V3 \text{ и } V4) = \frac{113,06 - 110,58}{5(2)2} = 0,3075;$$

$$SS(MP) = \frac{\sum Y_{ij}^2}{vb} - C = \frac{20,90^2 + \dots + 21,87^2}{4(5)} - C = 14,0138;$$

$$SS(MP\text{-ошибки}) = SS(MP) - SSB - SSV = 3,8757;$$

$$SSY = \frac{\sum Y_{...k}^2}{vb} - C = \frac{197,64^2 + 237,67^2}{4(5)} - C = 40,0600;$$

$$SS(V \times Y) = \frac{\sum Y_{i.k}^2}{b} - C - SSV - SSY = \\ = \frac{47,10^2 + \dots + 58,64^2}{5} - C - SSV - SSY = 1,4985.$$

Эта сумма квадратов взаимодействия расчленяется с учетом коэффициентов классовых сравнений, которые даны в таблице 11.7.

$$SS(V1+2 \text{ и } 3+4) \times Y = \\ = \frac{(47,10 + 47,53 - 51,07 - 51,94 - 57,89 - 59,15 + 61,99 + 58,64)^2}{5(8)} = 0,5736;$$

$$SS(V1 \text{ и } 2) \times Y = \frac{(47,10 - 47,53 - 57,89 + 59,15)^2}{5(4)} = 0,0344;$$

$$SS(V3 \text{ и } 4) \times Y = \frac{51,07 - 51,94 - 61,99 + 58,64^2}{5(4)} = 0,8904;$$

$$SS(\text{субделянок}) = \frac{\sum Y_{ijk}^2}{vby} - C = \frac{9,02^2 + \dots + 11,74^2}{4(5)2} - C = 67,1654;$$

$$SS(SP\text{-ошибки}) = SS(SP) - SS(MP) - SSY - SS(V \times Y) = 11,5931.$$

F-величины определяют путем деления соответствующих средних квадратов на средний квадрат ошибки, которые показаны скобками

Таблица 11.7. Ортогональные коэффициенты для расчленения годовых и сортовых сравнений

Сравнения	Годовые суммы по сортам							
	$V_1 Y_1$ 47,10	$V_2 Y_1$ 47,53	$V_3 Y_1$ 51,07	$V_4 Y_1$ 51,94	$V_1 Y_2$ 57,89	$V_2 Y_2$ 59,15	$V_3 Y_2$ 61,99	$V_4 Y_2$ 58,64
Y	+	+	+	+	-	-	-	-
V 1+2 и 3+4	+	+	-	-	+	+	-	-
V 1 и 2	+	-	0	0	+	-	0	0
V 1 и 4	0	0	+	-	0	0	+	-
(V 1+2 и 3+4) × Y	+	+	-	-	-	-	+	+
(V 1 и 2) × Y	+	-	0	0	-	+	0	0
(V 3 и 4) × Y	0	0	+	-	0	0	-	+

и стрелками в таблице 11.6. В опыте не выявлено взаимодействия сортов с годами, а большие значения F для годов указывают на реальное влияние особенностей лет, полученное в результате использования для F -оценки (сомнительной на первый взгляд важности) среднего квадрата ошибки субделянок. Тот факт, что MS -ошибки субделянок больше, чем MS -ошибки главных делянок, подтверждает справедливость вывода, сделанного выше.

Следует подчеркнуть, что остается мало сомнений в отношении преимуществ 3-го и 4-го сортов над 1-м и 2-м сортом, поскольку $F_{\text{факт}}$ превышает табличную величину на 1%-ном уровне. В опыте не выявлено очевидного преимущества 2-го сорта над 1-м сортом или 3-го сорта над 4-м сортом.

Стандартные ошибки. Расчеты стандартных ошибок являются аналогичными расчетам для анализов по годам. Стандартные ошибки и HSP , необходимые для выявления существенных эффектов в этом анализе, приводятся ниже.

1. Сравнение двух сортовых средних,

$$s_{\bar{d}} = \sqrt{\frac{2 (MP\text{-ошибка})}{by}} :$$

$$s_{\bar{d}} = \sqrt{\frac{2 (0,3230)}{5 (2)}} = 0,2542, HSP = 2,179 (0,2542) = 0,55.$$

2. Сравнение сортов 1-го и 2-го с 3-м и 4-м,

$$s_{\bar{d}} = \sqrt{\frac{2 (MP\text{-ошибка})}{by2}} :$$

$$s_{\bar{d}} = \sqrt{\frac{2 (0,3230)}{5 (2) 2}} = 0,1797, HSP = 2,179 (0,1797) = 0,39.$$

При расчете стандартных ошибок следует соблюдать требование о равенстве знаменателя числу сортов, составляющих сравниваемые средние. Так, мы помещаем 2 в знаменатель, потому что сравниваются средние 1-го и 2-го сортов со средней 3-го и 4-го сортов.

ЗАКЛЮЧЕНИЕ

Наиболее подходящей схемой статистического анализа периодических учетов урожаев с главных делянок, как, например повторные укосы многолетних трав с делянок по изучению сортов, повторные сборы фруктов с одних и тех же деревьев или повторных отборов образцов почвы с делянок на содержание питательных элементов — является схема расщепленных делянок. Анализ данных проводят по схеме расщепленных делянок, но выводы о наличии реальных эффектов для проведения повторных учетов и взаимодействия этих эффектов с вариантами главных делянок следует делать осторожно до тех пор, пока не будут получены большие значения.

ПРЕОБРАЗОВАНИЯ ИСХОДНЫХ ДАННЫХ (что делать, если данные нарушают установленные закономерности)

Исследователи, которые при изучении дисперсионного анализа ограничиваются знакомством с его «рецептами» и не проявляют попытки понять его основополагающие принципы, могут столкнуться с серьезными недоразумениями. Осознавая это или нет, они строят определенные предположения о получаемых данных, выполняя дисперсионный анализ. Если данные не подчиняются этим предположениям, то дисперсионный анализ может привести к ошибочному заключению. Они могут также не учесть важные дополнения к выводам, чего не бывает при надлежащем анализе данных.

ПРЕДПОСЫЛКИ, СОСТАВЛЯЮЩИЕ ОСНОВУ ДИСПЕРСИОННОГО АНАЛИЗА

Вкратце эти предпосылки могут быть сформулированы следующим образом.

1. Случайное, независимое и нормальное распределение компонентов ошибки.
2. Однородность дисперсий различных выборок.
3. Отсутствие корреляции между дисперсиями и средними различных выборок.
4. Слагаемость (аддитивность) главных эффектов.

Теперь обсудим эти четыре предположения более подробно.

Нормальность распределения компонентов ошибки. К счастью, отклонения данных от нормального распределения не слишком серьезно влияют на пригодность дисперсионного анализа. В статистике имеются критерии контроля нормальности, однако оснований для применения их нет до тех пор, пока объемы наших выборок не будут достаточно большими. Независимость предполагает отсутствие связи между величинами компонентов ошибок и группой экспериментального материала, к которой они относятся. Поскольку соседние делянки в полевом опыте склонны к большей зависимости между собой, чем с другими случайно разбросанными делянками, то важно избежать размещения одноименных вариантов на соседних участках поля. Это положение является одним из главных доводов в пользу требования о недопустимости деления делянки с определенным вариантом на части

и последующего использования их в качестве повторности. Лучшей гарантией против серьезных нарушений первого предположения дисперсионного анализа будет проведение рендомизации в соответствии с конкретно используемой экспериментальной схемой.

Однородность дисперсий. Первое практическое использование дисперсионного анализа в этой книге (глава 3) касалось простого опыта с двумя вариантами и пятикратной повторностью. Вы обратили внимание на наше предположение о том, что дисперсии внутри каждого варианта оценивают общую дисперсию. Поэтому было оправданным использование нами средней двух дисперсий как лучшей оценки σ^2 по сравнению с любой из них в отдельности. В главе 4 мы также использовали «усредненный средний квадрат ошибки», т. е. среднюю четырех дисперсий, обеспечившую нам лучшую оценку общей дисперсии.

Если дисперсии внутри различных вариантов были разными, то их усреднение было бы неоправданным. Предположим, что повторения двух вариантов фактически были выборками из совокупностей с большими дисперсиями, тогда как выборки двух других вариантов происходили из совокупностей со значительно меньшими дисперсиями. Очевидно, для того чтобы разность между средними была существенной, она должна иметь большую величину для сильно варьирующих дат первых двух вариантов, чем для двух других, имеющих меньшую вариабельность. Усреднение большей и меньшей дисперсий может привести к большому заблуждению. Разность между средними двух вариантов с большими дисперсиями может быть принята за существенную, тогда как в действительности она часто обуславливается случайным фактором. С другой стороны, разность между средними двух вариантов с меньшими дисперсиями может быть принята как случайная, хотя фактически она была реальной. Нижеприведенные данные гипотетического эксперимента с четырьмя вариантами и пятикратной повторностью каждого иллюстрируют такую ситуацию:

Варианты	Повторность					Суммы	Средние	s ²
	1	2	3	4	5			
A	3	1	5	4	2	15	3	2,5
B	6	8	7	4	5	30	6	2,5
C	12	6	9	3	15	45	9	22,5
D	20	14	11	17	8	70	14	22,5

Проведя дисперсионный анализ обычным путем, мы получим:

Источник вариации	df	S	MS	F
Варианты	3	330	110	8,8 **
Ошибка	16	200	12,5	

Отметим, что средний квадрат ошибки является средней четырех индивидуальных дисперсий внутри вариантов. Фактическое значение F очень существенно. Рассчитаем теперь HCP :

$$HCP_{05} = t \sqrt{2EMS/r} = 2,12 \sqrt{5} = 4,74.$$

Поскольку средняя разность между вариантами A и B равна 3, мы делаем заключение об ее незначительности. Средняя разность между C и D равна 5, и она будет названа существенной на 5%-ном уровне. Мы подчеркиваем, однако, что дисперсии вариантов C и D в 9 раз превышают дисперсии A и B . Предположение об однородности дисперсий будет в данном случае крайне сомнительным. Поэтому целесообразно делать анализ вариантов для A и B отдельно от C и D .

Источники вариации	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Анализ вариантов для <i>A</i> и <i>B</i>				
Варианты	1	22,5	22,5	9 *
Ошибка	8	20,0	2,5	
Для <i>C</i> и <i>D</i>				
Варианты	1	62,5	62,5	2,78 Несущественно
Ошибка	8	180	22,5	

Теперь мы пришли к совершенно противоположному заключению в отношении существенности разностей между вариантами A и B и C и D . Позже мы покажем способы проверки данных на однородность дисперсий. Если дисперсии данных окажутся неоднородными, то для этого случая имеется несколько разных дополнений к анализу. Во-первых, мы можем сгруппировать данные на основании однородности дисперсий внутри каждой группы. Затем каждая группа отдельно подвергается анализу, как в вышеприведенном примере. Во-вторых, можно использовать метод, описанный в пособиях математической статистики повышенной сложности, который основывается на взвешивании средних в соответствии с их дисперсиями. В-третьих, мы можем преобразовать данные таким образом, что они будут однородными. Мы рассмотрим этот метод подробнее в настоящей главе.

Независимость средних и дисперсий. Некоторые данные имеют определенную связь выборочных средних и их дисперсий. Этот частный случай наиболее типичный пример неоднородности дисперсий. Положительная корреляция часто имеет место при наличии большого размаха выборочных средних.

Предположим, что экспериментатор изучал эффективность нескольких инсектицидов по борьбе с тлей, подсчитывая количество тлей на листе после опрыскивания. Если средние двух довольно неэффективных вариантов составили 305 и 315, то естественными были бы его сомнения в отношении большого значения этой разницы. С другой стороны, если средние двух других вариантов составили 5 и 15, он был бы склонен оценить эту разность как значительную, будучи под

впечатлением того факта, что одна из средних превышает другую в три раза. Предполагая однородность дисперсий и отсутствие их связи со средними, он придал бы одинаково большое значение как разности между 305 и 315, так и разности между 5 и 15, поскольку абсолютная величина этих разностей одна и та же в обоих случаях. Он, вероятно, усомнился бы в правильности выводов. Проверка различных выборок почти наверняка устранит причину сомнений; как правило, выборки с большими значениями средних будут также иметь большую дисперсию, а выборки с малыми средними — малую дисперсию. В результате предположение об отсутствии корреляции между средними и дисперсиями будет признано неправильным, а следовательно, обычный дисперсионный анализ исходных данных станет непригодным.

Давайте возьмем более экстремальный пример. Допустим, группа исследователей решила изучить действие новых витаминов на рост животных. Для эксперимента были взяты различные виды животных: мыши, цыплята и овцы. Здравый смысл подскажет нам, что разность в полфунта (225 г) между средними массами тела двух групп овец может быть признана незначительной и легко отнесена за счет случая. Разность в полфунта между средними массами тела двух групп цыплят будет признана очень большой, но не выходящей за пределы возможного. Подобная разность в средней массе двух групп мышей была бы рассмотрена как совершенно нереальная. Хотя это экстремальный и почти абсурдный пример, но он помогает осознать то, что предположение о независимости дисперсий и средних нельзя принимать слепо. Мы должны проверить данные, а при необходимости — уточнить и правильность предположения, прежде чем приступать к дисперсионному анализу.

Другими видами данных, в которых проявляется зависимость между дисперсиями и средними, являются величины, основанные на подсчетах, а также величины, выражаемые в долях или процентах. Предположим, что мы установили связь между дисперсиями и средними. Означает ли это, что мы вынуждены отказаться от дисперсионного анализа как метода статистической обработки данных? К счастью, это случается редко. Чаще всего мы можем преобразовать данные таким образом, что предположение о независимости дисперсий и средних будет иметь силу. Далее можно провести дисперсионный анализ на основе преобразованных дат.

Слагаемость (аддитивность). Для каждой экспериментальной схемы имеется математическое уравнение, называемое линейной моделью слагаемости. Для схемы полной рендомизации модель имеет вид $Y_i = \bar{Y} + t_i + e_i$, которая гласит, что каждая дата эксперимента складывается из эффекта общей средней, эффекта вариантов и величины ошибки (эффекта случайного фактора). Соответствующая модель для схемы рендомизированных блоков $Y_{ij} = \bar{Y} + t_i + b_j + e_{ij}$ гласит, что каждая дата эксперимента складывается из эффектов общей средней, вариантов, блоков и ошибки. Важным моментом этих моделей является сложение всех выражений (эффектов), что и объясняет термин «слагаемость».

Модель схемы рендомизированных блоков, например, предполагает

одинаковый эффект вариантов по всем блокам и одинаковый эффект блоков для всех вариантов. Другими словами, если установлено, что вариант дает определенную прибавку урожая по сравнению с общей средней, то она должна быть одинаковой для блоков с высоким и низким плодородием почвы.

Можно привести много ситуаций, где такое предположение будет необоснованным. Например, в опыте по оценке действия азотного удобрения на урожай некоторые блоки могут дать меньшую урожайность, чем другие, ввиду низких запасов почвенного азота в них. Мы можем ожидать больший эффект от азотного удобрения на делянках этих блоков, чем на делянках, где содержание азота в почве было на таком же уровне и без применения удобрений. С другой стороны, предположим, что низкая урожайность была обусловлена недостаточной влагообеспеченностью почвы. Тогда применение азотного удобрения может оказать лишь незначительное влияние на урожайность блоков с низким плодородием почвы и обеспечить существенную прибавку урожая блоков с достаточной влагообеспеченностью. Другая ситуация может быть такой, что эффект варианта выражается в увеличении урожая на определенный процент или долю. Такой случай обозначают как умножающийся (мультипликативный) эффект варианта.

Во всех вышеприведенных] ситуациях предположение о слагаемости эффектов будет неправильным; это надо учитывать при анализе дат. Для случая умножающихся эффектов вариантов также имеются способы преобразования исходных дат, обеспечивающих слагаемую модель.

ПРОВЕРКА СООТВЕТСТВИЯ ДАННЫХ ПРЕДПОЛОЖЕНИЯМ ДИСПЕРСИОННОГО АНАЛИЗА

Теперь вы подготовлены для того, чтобы рассмотреть некоторые специфические примеры данных, отклоняющихся от одного или большего числа предположений дисперсионного анализа. Мы покажем, как проверять эти предположения и пути преобразования исходных данных, устраняющие отмеченное несоответствие. В таблице 12.1 даны некоторые гипотетические данные, которые могли быть получены в эксперименте по изучению действия нового витамина на мышей, цыплят и овец.

Очень существенная разность между видами животных совершенно не удивляет нас. И очень странным кажется то, что мы не находим существенного эффекта витамина, особенно потому, что каждое животное каждого повторения, получавшее витамин, имело большую массу, чем соответствующее животное контрольного варианта. Также странным кажется то, что мы не находим доказательства взаимодействия между эффектами витамина и видов, тем более, что различия в отзывчивости на введение витамина у разных животных были столь очевидными. Если принять результаты этого анализа такими, как есть, то мы должны признать эксперимент неудачным. Все, что он нам дал, ограничивается подтверждением различия в массе тела мышей, цыплят и овец. Однако даже на этой основе при расчленении видового

Таблица 12.1. Действие витамина на рост животных (масса тела в фунтах) в опыте со схемой рендомизированных блоков

Варианты	Блоки				Суммы	Средние
	I	II	III	IV		
Мыши, контроль	0,18	0,30	0,28	0,44	1,2	0,3
Мыши, получавшие витамин	0,32	0,40	0,42	0,46	1,6	0,4
Субсуммы	0,50	0,70	0,70	0,90	2,8	0,35
Цыплята, контроль	2,0	3,0	1,8	2,8	9,6	2,40
Цыплята, получавшие витамин	2,5	3,3	2,5	3,3	11,6	2,90
Субсуммы	4,5	6,3	4,3	6,1	21,2	2,65
Овцы, контроль	108	140	135	165	548	137
Овцы, получавшие витамин	127	153	148	176	604	151
Субсуммы	235	293	283	341	1152	144
Общие суммы	240	300	288	348	1176	49

Проведя статистическую обработку данных по методам, используемым в главах 5 и 6, получим следующие результаты дисперсионного анализа:

Источник вариации	df	SS	MS	F
Блоки	3	984,00	328,00	2,63
Варианты	5	108713,68	21742,74	174,43 **
Виды животных	2	108321,16	54160,58	434,51 **
Витамин	1	142,11	142,11	1,14
Виды × витамин	2	250,41	125,20	1,00
Ошибка	15	1869,72	124,65	

эффекта на два сравнения, сопоставляя в первом овец с цыплятами и мышами, а в другом — цыплят с мышами, мы сможем выявить существенную разность в массе тела цыплят и мышей.

Давайте посмотрим на данные с позиций предпосылок, основополагающих дисперсионный анализ, и решим, что надо предпринять, если некоторые предположения будут нарушены. Во-первых, мы можем оценить, распределяются ли эффекты ошибки случайным, независимым или нормальным образом. Для этого из каждой даты таблицы исходных данных мы вычтем значение общей средней, эффекты вариантов и блоков по способу, описанному в главе 5. Таким образом, мы получим таблицу компонентов (эффектов) ошибки (табл. 12.2).

Эффекты ошибки распределяются явно не случайным образом. Они, очевидно, не являются независимыми, поскольку в каждом блоке эффекты ошибки для двух животных одного вида тесно связаны. Наконец, их распределение выглядит далеко не соответствующим нормальному, поскольку имеется два модальных класса (включающих наибольшее количество значений): один составляет интервал от 8,5 до 9,0, а другой — от —8,5 до —9,0. Первое предположение дисперсионного анализа не выдержало строгой проверки.

Таблица 12.2. Компоненты ошибки в опыте с витамином

Варианты	Блоки				Суммы
	I	II	III	IV	
Мыши, контроль	8,88	-1,00	0,98	-8,86	0
Мыши, получавшие витамин	8,92	-1,00	1,02	-8,94	0
Цыплята, контроль	8,60	-0,40	0,40	-8,60	0
Цыплята, получавшие витамин	8,60	-0,60	0,60	-8,60	0
Овцы, контроль	-20,00	2,00	-1,00	19,00	0
Овцы, получавшие витамин	-15,00	1,00	-2,00	16,00	0
Суммы	0	0	0	0	0

Теперь мы проверим предположение об однородности дисперсий. Однако прежде следует познакомиться с критерием, известным под названием «критерий однородности дисперсий Бартлетта».

Во-первых, нам надо вычислить дисперсию каждого варианта на основе четырех повторений. Для контроля мышей получим:

$$\frac{0,18^2 + 0,30^2 + 0,28^2 + 0,44^2 - (1,2^2 / 4)}{\text{количество повторений} - 1} = 0,0115.$$

После того, как все дисперсии будут вычислены подобным образом, результаты заносят в таблицу (табл. 12.3).

В целях избежания отрицательных логарифмов дисперсии кодируются. Мы можем, не изменив результата анализа, умножить дисперсии на любое одинаковое для всех число. Желательно, чтобы все кодируемые величины были не меньше 1, поэтому для получения кодов мы умножим каждую дисперсию на 1000. Проще использовать десятичные логарифмы, при этом в мантиссе достаточно иметь две значащие цифры. Среднюю кодируемых дисперсий определяют путем деления их суммы на число вариантов, кроме этого, в таблицу заносят логарифмы

Таблица 12.3. Дисперсии и их логарифмы для вариантов эксперимента с витамином

Варианты	df	s_i^2	Коды s_i^2	Логарифмы кодов s_i^2
Мыши, контроль	3	0,0115	11,5	1,06
Мыши, получавшие витамин	3	0,0035	3,5	0,54
Цыплята, контроль	3	0,3467	346,7	2,54
Цыплята, получавшие витамин	3	0,2133	213,3	2,33
Овцы, контроль	3	546,0	546000	5,74
Овцы, получавшие витамин	3	425,0	425000	5,63
Суммы	18		971875	17,84
Средняя			161979	
Логарифм средней			5,209	

рифм этой средней. Теперь мы готовы для того, чтобы рассчитать так называемый неприведенный хи-квадрат (χ^2).

Общая формула для выборок неодинакового объема имеет вид

$$\chi^2 = 2,3026 (\lg s^{-2} \times \Sigma df) - \Sigma (df \times \lg s_i^2).$$

Когда выборки одинакового размера, как в нашем примере, формула сокращается до:

$$\chi^2 = 2,3026 df (n \lg s^{-2} - \Sigma \lg s_i^2) = 2,3026 (3) [6 (5,209) - 17,84] = 92,66.$$

Коэффициент 2,3026 служит в этих формулах фактором перевода десятичных логарифмов в натуральные, n — количество выборок (вариантов), df — число степеней свободы, приходящихся на выборку.

Неприведенный хи-квадрат должен быть приведен путем деления на корректирующий фактор C . При разных объемах выборок используют следующую формулу:

$$C = 1 + \frac{1}{3(n-1)} \left(\Sigma \frac{1}{df} - \frac{1}{\Sigma df} \right);$$

при одинаковых объемах выборок формула сокращается до

$$C = 1 + \frac{n+1}{3n(df)}.$$

В нашем примере

$$C = 1 + \frac{7}{3(6)(3)} = 1,13.$$

Тогда χ^2 приведенный = χ^2 неприведенный / $C = 92,66 / 1,13 = 82,00$.

Теперь обратимся к таблице хи-квадратов (приложение, табл. А.6) и на основе 5 степеней свободы (на единицу меньше числа вариантов-выборок) находим, что фактическое значение хи-квадрата (82) значительно превышает табличную величину (20,517) на 0,1%-ном уровне значимости. Это служит очень убедительным доказательством неоднородности дисперсий.

Следующее предположение, подлежащее проверке, — независимость между средними и дисперсиями. Беглого осмотра данных достаточно, чтобы убедиться наверняка в нарушении этого предположения, ибо высокие значения средних соответствуют очень большим дисперсиям и малые значения средних — очень малым дисперсиям.

При выборе способа преобразования данных важно уяснить, что в большей степени пропорционально средним — дисперсии или стандартные отклонения? Для этого мы построим таблицу указанных отклонений (табл. 12.4).

Мы видим, что отношения дисперсий к средним заметно возрастают, тогда как отношения стандартных отклонений к средним остаются довольно постоянными величинами. Другими словами, стандартные отклонения приблизительно пропорциональны средним. Между прочим, если дисперсии и средние не были связаны, то ожидалось бы уменьшение обоих отношений по мере увеличения средних.

Таблица 12.4. Отношения дисперсий и стандартных отклонений к средним в опыте с витамином

Варианты *	$\bar{Y}$	s_i^2	s_i	$s_i^2/\bar{Y}$	$s_i/\bar{Y}$
М—К	0,3	0,01147	0,107	0,04	0,36
М—В	0,4	0,00347	0,059	0,01	0,15
Ц—К	2,4	0,3467	0,589	0,14	0,24
Ц—В	2,9	0,2133	0,462	0,07	0,16
О—К	137,0	546,0	23,367	3,98	0,17
О—В	151,0	425,3	20,624	2,82	0,14

* Сокращения М—К, М—В, Ц—К, Ц—В, О—К и О—В обозначают: мыши—контроль; мыши, получавшие витамин; цыплята—контроль; цыплята, получавшие витамин; овцы—контроль и овцы, получавшие витамин, соответственно.—Прим. ред.

Последнее предположение, подлежащее проверке,—это слагаемость. В основе этого предположения лежит приблизительное равенство блоковых эффектов для всех вариантов. По данным таблицы 12.1, средняя разность между блоком I и блоком 4 составила 18 фунтов *. Однако средние разности между этими блоками по видам таких животных, как мыши, цыплята и овцы, составили 0,2; 0,8 и 53 фунта соответственно.

Общепринятым критерием слагаемости служит критерий Тьюки. Этот критерий применим для любых группировок с двумя входами, таким, как данные опытов, заложенных по схеме рендомизированных блоков, где производят группировку по блокам и вариантам.

Далее необходима таблица, подобная таблице 12.5, которая наряду с исходными данными таблицы 12.1 содержит по краям рассчитанные эффекты блоков и вариантов.

Отметьте, что суммы отдельных эффектов, как блоков, так и вариантов, равны нулю. Для проверки слагаемости по критерию необходимо

Таблица 12.5. Эффекты блоков и вариантов

Варианты (T)	Блоки (B)				Средние ($\bar{Y}_i$)	Эффекты вариантов ($\bar{Y}_i - \bar{Y}_{..}$)
	I	II	III	IV		
М—К	0,18	0,30	0,28	0,44	0,3	—48,7
М—В	0,32	0,40	0,42	0,46	0,4	—48,6
Ц—К	2,00	3,00	1,80	2,80	2,4	—46,6
Ц—В	2,50	3,30	2,50	3,30	2,9	46,1
О—К	108,0	140,0	135,0	165,0	137,0	88,0
О—В	127,0	153,0	148,0	176,0	151,0	102,0
Средние	40,0	50,0	48,0	58,0	49,0	
$\bar{Y}_{.j} - \bar{Y}_{..}$	—9,0	1,0	—1,0	9,0		

* Фунт=0,45 кг.—Прим. ред.

провести расчеты по следующей формуле:

$$Q = \sum Y_{ij} (\bar{Y}_{i.} - \bar{Y}_{..}) (\bar{Y}_{.j} - \bar{Y}_{..}),$$

которая означает, что мы должны умножить каждую дату таблицы на соответствующий эффект варианта и блока и сложить полученные произведения.

В нашем примере:

$$Q = 0,18(-48,7)(-9,0) + \dots + 176,0(102,0)(9,0) = 90\ 140,56.$$

Тогда сумму квадратов для неслагаемости находят следующим образом:

$$SS \text{ неслагаемости} = \frac{(Q^2 \times \text{общее число экспериментальных единиц})}{(SST_r \times SSB)}$$

Используя это уравнение для нашего примера, получим

$$SS \text{ неслагаемости} = \frac{(90\ 140,56^2 \times 24)}{(108\ 713,68 \times 984,0)} = 1822,94.$$

Это составляет часть суммы квадратов взаимодействия $B \times T$ или сумму квадратов для ошибки, которая расчленяется на:

Источник вариации	df	SS	MS	F
Ошибка ($B \times T$)	15	1869,72		
Неслагаемость	1	1822,94	1822,94	545,79
Остаток	14	46,78	3,34	

Фактическое значение F далеко превосходит требуемое (8,86 из таблицы А.3) на 1%-ном уровне при 1-й и 14-й степенях свободы, что служит сильным доказательством неправильности предположения о слагаемости.

Итак, мы проверили все предположения дисперсионного анализа и нашли, что наши данные не удовлетворяют ни одному из них. Не удивительно, что результаты дисперсионного анализа обманули все ожидания. Возможно, раздельный анализ по видам животных будет наиболее разумным методом статистической обработки данных этого эксперимента. Результаты таких анализов приведены ниже.

Виды животных	Источник вариации	df	SS	MS	F
Мыши	Блоки	3	0,0400	0,0133	8,31
	Витамин	1	0,0200	0,0200	12,50 *
	Ошибка	3	0,0048	0,0016	
Цыплята	Блоки	3	1,64	0,547	41,00 **
	Витамин	1	0,50	0,500	37,5 **
	Ошибка	3	0,04	0,013	
Овцы	Блоки	3	2834,0	944,7	157,4 *
	Витамин	1	392,0	392,0	66,3 **
	Ошибка	3	18,0	6,0	

Эти результаты в значительно большей степени удовлетворяют экспериментатора, чем ранее полученные на основе общего дисперсионного анализа.

Эти анализы обоснованны, потому что данные по любым видам животных вполне соответствуют основным предположениям дисперсионного анализа. Единственно слабым местом этих анализов является то, что они мало проясняют вопрос об идентичности или различиях в реакции разных видов животных на витамин. Возможно, этот вопрос не имеет большого значения для производства, и научный работник, несомненно, будет вполне доволен тем, что уже получено. Однако мы рассмотрим преобразование данных, чтобы показать, какие удивительные результаты можно получить.

ЛОГАРИФМИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ

Во-первых, следует указать, как преобразуют данные. Если стандартные отклонения (не дисперсии) данных выборок приблизительно пропорциональны средним (все отношения дают близкие величины), то наиболее эффективным способом преобразования данных является их логарифмирование. Другим мерилом при выборе этого способа преобразования является большая вероятность наличия умножающихся главных эффектов, чем складывающихся. Оба из указанных доводов являются очевидными для наших данных, поэтому мы преобразуем данные в логарифмы и посмотрим, что получится.

Таблица 12.6. Логарифмическое преобразование данных эксперимента с витамином (предварительно исходные даты умножили на 10)

Варианты (виды животных)	Блоки				Суммы	Средние
	I	II	III	IV		
Мыши, контроль	0,26	0,48	0,45	0,64	1,83	0,4575
Мыши, получавшие витамин	0,51	0,60	0,62	0,66	2,39	0,5975
Субсуммы	0,77	1,08	1,07	1,30	4,22	0,5275
Цыплята, контроль	1,30	1,48	1,26	1,45	5,49	1,3725
Цыплята, получавшие витамин	1,40	1,52	1,40	1,52	5,84	1,4600
Субсуммы	2,70	3,00	2,66	2,97	11,33	1,41625
Овцы, контроль	3,03	3,15	3,13	3,22	12,53	3,1325
Овцы, получавшие витамин	3,10	3,18	3,17	3,25	12,70	3,1750
Субсуммы	6,13	6,33	6,30	6,47	25,23	3,15375
Суммы	9,60	10,41	10,03	10,74	40,78	
Средние	1,60	1,735	1,672	1,790		1,69917

Таблица 12.6а. Результаты дисперсионного анализа (преобразованные данные)

Источник вариации	df	SS	MS	F
Блоки	3	0,12075	0,04025	13,77 **
Варианты	5	28,60738	5,72148	1959,41 **
Витамин	1	0,04860	0,04860	16,62 **
Виды животных	2	28,54926	14,27463	4883,00 **
Виды животных × витамин	2	0,00952	0,00476	1,63
Ошибка	15	0,04385	0,00292	

Однако прежде необходимо сделать несколько общих замечаний о применении этого преобразования. Величины с отрицательными знаками не могут быть преобразованы таким путем. Если среди данных встречаются нулевые значения, то возникает новая проблема, поскольку логарифм нуля дает минус бесконечность. Для того чтобы избежать этого, рекомендуется до логарифмирования прибавить ко всем датам единицу.

Данные, содержащие большое количество нулевых значений, лучше обрабатывать каким-то другим способом. Можно использовать логарифмы с любым основанием, однако обычно легче даются десятичные логарифмы. Вполне обоснованным до преобразования данных будет умножение их на постоянное число, поскольку это не изменяет последующий анализ. Это особенно желательно в тех случаях, когда имеются величины меньше 1, поскольку такой прием исключает применение отрицательных логарифмов.

Данные, с которыми мы работаем, не имеют нулевых значений, однако минимальная величина равна 0,18, поэтому мы прежде умножаем все даты на 10, а затем берем их логарифмы. Это позволит нам заполнить таблицу преобразованных данных (табл. 12.6, 12.6а).

Несомненно, эти результаты более весомы, чем результаты анализов исходных дат, в отношении положительных выводов. Мы все еще не установили существенного взаимодействия между видами животных и витамином, однако теперь постановка вопроса иная. Прежде мы спрашивали: «Различается ли прибавка в массе тела, обусловленная применением витамина, по видам животных?» Теперь мы спрашиваем: «Различается ли доля или процент прибавки в массе тела, обусловленные применением витамина, по видам животных?»

Имеем ли мы больше положительных результатов в этот раз по причине простой «игры с цифрами» при получении нужного нам результата? Или оправдало себя использованное преобразование и новый анализ оказался обоснованным? Для уверенности ответа на эти вопросы необходимо проверить соответствие новых данных предположениям дисперсионного анализа.

Как и прежде, мы построим таблицу компонентов ошибки путем вычитания из каждой даты таблицы преобразованных величин данных: средней по опыту и эффектов соответствующих вариантов и блоков (табл. 12.7).

Таблица 12.7. Компоненты ошибки преобразованных данных (обозначения вариантов даны в сноске к табл. 12.4)

Варианты	Блоки			
	I	II	III	IV
М—К	—0,10	—0,01	0,02	0,09
М—В	0,01	—0,03	0,05	—0,03
Ц—К	0,03	0,07	—0,08	—0,01
Ц—В	0,04	0,02	—0,03	—0,03
О—К	0,00	—0,02	0,02	0,00
О—В	0,02	—0,03	0,02	—0,02

Эти величины ошибки кажутся в большей мере случайно и нормально распределенными, чем ошибки исходных данных.

Для проверки однородности дисперсий вновь воспользуемся критерием Бартлетта (табл. 12.8).

$$\chi^2 = 2,3026[(18 \times 0,9614) - (3 \times 5,11)] = 4,548;$$

$C = 1,13$, как и прежде, тогда

$$\chi^2 \text{ приведенный} = \frac{4,548}{1,13} = 4,03.$$

χ^2 приведенный, согласно табличной величине χ^2 (приложение, табл. А.6), будет случайно превзойден более чем в 50 случаях из 100. Достаточно беглого взгляда на таблицу 12.8, чтобы убедиться в отсутствии какой-либо связи между средними и дисперсиями.

Для проведения проверки на слагаемость мы рассчитаем эффекты блоков и вариантов на основе преобразованных данных (табл. 12.9), подобно расчетам для исходных дат, приведенным в таблице 12.5.

Таблица 12.8. Использование критерия Бартлетта для проверки преобразованных дат опыта с витамином (обозначения вариантов даны в сноске к табл. 12.4)

Варианты	Средние	s_i^2	Коды s_i^2	Логарифмы кодов s_i^2
М—К	0,4575	0,0243	24,3	1,39
М—В	0,5975	0,0040	4,0	0,60
Ц—К	1,3725	0,0118	11,8	1,07
Ц—В	1,4600	0,0048	4,8	0,68
О—К	3,1325	0,0062	6,2	0,79
О—В	3,1750	0,0038	3,8	0,58
Суммы			54,9	5,11
Средняя			9,15	
Логарифм средней			0,9614	

Таблица 12.9. Расчеты эффектов блоков и вариантов для преобразованных данных
(обозначения вариантов даны в сноске к табл. 12.4)

Варианты	Блоки				Средние ($\bar{Y}_{i.}$)	Эффекты вариантов ($\bar{Y}_{i.} - \bar{Y}_{..}$)
	I	II	III	IV		
M—K	0,26	0,48	0,45	0,64	0,46	—1,24
M—B	0,51	0,60	0,62	0,66	0,60	—1,10
II—K	1,30	1,48	1,26	1,45	1,37	—0,33
II—B	1,40	1,52	1,40	1,52	1,46	—0,24
O—K	3,03	3,15	3,13	3,22	3,13	1,43
O—B	3,10	3,18	3,17	3,25	3,18	1,48
Средние	1,60	1,74	1,67	1,79	1,70	
$\bar{Y}_{.j} - \bar{Y}_{..}$	—0,10	0,04	—0,03	0,09		

Как и прежде, используем формулу

$$Q = \sum Y_{ij} (\bar{Y}_{i.} - \bar{Y}_{..}) (\bar{Y}_{.j} - \bar{Y}_{..}) = 0,26 (-1,24) (-0,10) + \dots + 3,25 (1,48) (0,09) = -0,023768;$$

$$SS \text{ неслагаемости} = \frac{Q^2 \times \text{суммы экспериментальных единиц}}{SST_r \times SSB} = \\ = \frac{-0,023768^2 \times 24}{28,60738 \times 0,12075} = 0,00392.$$

Сумма квадратов для ошибки может быть расчленена следующим образом:

Источник вариации	df	SS	MS	F
Ошибка	15	0,04385		
Неслагаемость	1	0,00392	0,00392	1,37
Остаток	14	0,03993	0,00285	

Величина F не достигает даже 10%-го уровня существенности при 1-й и 14-ти степенях свободы (требуемое $F_{.10} = 3,10$).

Теперь мы уверены в обоснованности нового анализа, поскольку преобразованные данные отвечают всем предположениям дисперсионного анализа. Исходные даты не соответствовали ни одному из предположений.

ПРЕОБРАЗОВАНИЯ ЧЕРЕЗ КВАДРАТНЫЙ КОРЕНЬ

Когда мы имеем дело с подсчетом численностей редких событий, то получаемые данные подчиняются особому распределению, называемому распределением Пуассона. Под редким событием понимают со-

бытие, вероятность наступления которого для любого индивидуума составляет очень малую величину. Предположим, что 0,1% большой массы семян салата-латука имели вирус мозаичности. Тогда вероятность того, что какое-то отдельное семя содержит вирус составит лишь 1/1000, так что в отношении индивидуальных семян пораженность вирусом является редким событием. Если мы возьмем 100 выборок из такой большой массы по 1000 семян в каждой, то получим приблизительно следующие результаты:

37	выборок	будут	содержать	0	пораженных	семян
37	»	»	»	1	пораженное	семя
18	»	»	»	2	пораженных	семян
6	»	»	»	3	»	»
2	выборки	»	»	4	»	»

Явно, это очень мало напоминает нормальное распределение. Распределение Пуассона имеет очень примечательное свойство — дисперсия равна средней. Практически дисперсия обычно несколько больше, чем средняя, поскольку в дополнение к вариации, обусловленной выборкой, на появление учитываемого события воздействуют другие факторы. По меньшей мере дисперсия стремится быть пропорциональной средней, что нарушает предположение об отсутствии корреляции между дисперсиями и средними.

Другой пример данных такого рода можно встретить при подсчетах численности насекомых, пойманных с помощью стандартной ловушки (сачка). В этом случае довольно трудно определить, что мы обозначаем индивидуальным наблюдением. Это может быть индивидуальным участком, на котором встречаются насекомые. Используя сеть ловушек, мы берем пробы тысячи таких участков и находим лишь несколько насекомых. Таким образом, случайное нахождение насекомого в заданном месте и в конкретное время будет редким событием.

Данные такого рода можно сделать близкими к нормальному, с одновременным достижением относительной независимости дисперсий от средних, путем извлечения квадратных корней. В действительности

Таблица 12.10. Число травяных клопов на 50 ловушек

Варианты	Блоки				Суммы	Средние	s_i^2
	I	II	III	IV			
A	7	5	4	1	17	4,25	6,25
B	6	1	2	1	10	2,50	5,67
C	6	2	1	0	9	2,25	6,92
D	0	1	2	0	3	0,75	0,92
E	1	0	1	2	4	1,00	0,67
F	5	14	9	15	43	10,75	21,58
G	8	6	3	6	23	5,75	4,25
H	3	0	5	9	17	4,25	14,25
I	4	10	13	5	32	8,00	18,00
J	6	11	5	2	24	6,00	14,00
K	8	11	2	6	27	6,75	14,25

лучше использовать такое выражение: $\sqrt{Y + \frac{1}{2}}$, особенно если численность не превышает 10.

В таблице 12.10 приведены данные числа травяных клопов, попавших в 50 ловушек каждой делянки, при оценке 10 инсектицидов в опыте, проведенном по схеме рендомизированных блоков с четырехкратной повторностью. Итоги дисперсионного анализа таковы:

Источник вариации	df	SS	MS	F
Блоки	3	12,25	4,08	0,40
Варианты	10	380,00	38,00	3,70 **
Ошибка	30	308,00	10,27	

Преобразование всех данных путем извлечения квадратного корня ($\sqrt{Y + \frac{1}{2}}$) дает таблицу 12.11.

Дисперсионный анализ дат из таблицы 12.11 дает следующие результаты:

Источник вариации	df	SS	MS	F
Блоки	3	0,532	0,177	0,36
Варианты	10	19,993	1,999	4,04 **
Ошибка	30	14,841	0,495	

Результаты этих двух анализов мало различаются, поскольку оба показали очень существенный эффект вариантов. Величина F возросла после преобразования данных почти на 10%. Некоторые важные различия появятся при оценке частных различий (выделение существенных средних), как показано в таблице 12.12.

Таблица 12.11. Преобразованные данные численности травяных клопов

Варианты	Блоки				Суммы	Средние	s_i^2
	I	II	III	IV			
A	2,74	2,35	2,12	1,22	8,43	2,11	0,41
B	2,55	1,22	1,58	1,22	6,57	1,65	0,39
C	2,55	1,58	1,22	0,71	6,06	1,52	0,60
D	0,71	1,22	1,58	0,71	4,22	1,06	0,18
E	1,22	0,71	1,22	1,58	4,73	1,18	0,13
F	2,35	3,81	3,08	3,94	13,18	3,29	0,54
G	2,92	2,55	1,87	2,55	9,89	2,45	0,19
H	1,87	0,71	2,35	3,08	8,01	2,00	0,99
I	2,12	3,24	3,67	2,35	11,38	2,84	0,53
J	2,55	3,39	2,35	1,58	9,87	2,47	0,55
K	2,92	3,39	1,58	2,55	10,44	2,61	0,59

Таблица 12.12. Оценка исходных и преобразованных дат по многоранговому критерию Дункана (8%-ный уровень)

Выделение существенных средних из	Варианты и средние по вариантам										
	D	E	G	B	H	A	G	J	K	I	F
Исходных дат	0,75	1,00	2,25	2,50	4,25	4,25	5,75	6,00	6,75	8,00	10,75
Уравновешенных средних преобразованных дат	0,62	0,89	1,81	2,22	3,50	3,95	5,50	5,60	6,31	7,57	10,32

Обратите внимание, что после преобразования данных установлены существенные различия между вариантами *G* и *D*, *G* и *E*, *J* и *D* и *J* и *E*, что не было выявлено на основе исходных данных.

Взвешенные средние, показанные в таблице 12.12, получены в результате «обратного преобразования» средних преобразованных дат в первоначальные величины. Это сделано путем возведения преобразованных средних в квадрат и вычитания $1/2$ -й. Средние, полученные таким образом, будут меньше по сравнению со средними исходных данных, потому что произошло перераспределение значений переменной. Именно это и требовалось сделать, поскольку в распределении Пауссона малые значения переменной соразмеряются меньшей выборочной ошибкой, чем большие значения.

Фактически при подготовке научного отчета по такому эксперименту лучше использовать эти взвешенные средние с примечанием о способе их получения.

В целом эффект преобразования данных через квадратный корень сводится к увеличению точности измерения различий между малыми средними. Это в высшей степени желательно при борьбе с вредными насекомыми, поскольку для нас представляют меньший интерес различия среди относительно неэффективных вариантов, чем различия между приемами, обеспечивающими хорошую защиту.

Беглый просмотр дисперсий двух таблиц покажет, что до преобразования данных имела сильная положительная связь между средними и дисперсиями. Коэффициент линейной корреляции между ними (равный 0,89) существен на 0,1%-ном уровне значимости. После преобразования корреляция, составившая лишь 0,37, была несущественна даже на 10%-ном уровне. Таким образом, одно из предположений дисперсионного анализа было нарушено в исходных данных, а преобразование позволило это устранить.

Относительно других предположений дисперсионного анализа серьезных нарушений отмечено не было. Проверка компонентов ошибки не показывает четкого отклонения их от случайного нормального распределения. Проверка однородности дисперсий исходных дат по Бартлетт-критерию дает приведенное значение χ^2 -квадрата (равное 12,56), вероятность случайного превышения этой величины составляет 25%. После преобразования значение χ^2 -квадрата сократилось до

4,81, вероятность превышения которого только из-за случайного фактора составляла 90%. Итак, преобразование уменьшило неоднородность дисперсий по сравнению с исходными датами, однако в обоих случаях неоднородность дисперсий не была существенной.

Величина F для неслагаемости при проверке на слагаемость по критерию Тьюки была меньше единицы даже для исходных дат.

В целом мы можем заключить, что данные, требующие преобразования через квадратный корень, не нарушают столь глубоко предположений дисперсионного анализа, как данные, требующие логарифмического преобразования. Во всяком случае изменения в результатах анализа, связанные с этим преобразованием, не так выразительны.

ПРЕОБРАЗОВАНИЯ ЧЕРЕЗ АРКСИНУС ИЛИ УГОЛ

Другим видом данных, подлежащих преобразованию, могут быть результаты подсчетов, выражаемые в процентах или долях к объему общей выборки. Такие данные скорее имеют так называемое биномиальное распределение, чем нормальное распределение. Одной из особенностей этого распределения является то, что дисперсии связаны со средними, однако совершенно по-иному, чем уже рассмотренные другие виды данных. До сих пор мы обсуждали те случаи, когда большие средние стремятся иметь большие дисперсии, и наоборот. В данных, подчиняющихся биномиальному распределению, дисперсии стремятся принять малые величины по краям всего размаха значений переменной (приближающихся к нулю и 100%), а большие дисперсии сосредотачиваются в центре (около значения в 50%). Фактически эта особенность представляется естественной даже для нематематиков. Мы склонны придать большее значение разности между нулем и 6% или между 94% и 100%, чем разности между 47% и 53% хотя величина разности одинакова во всех случаях.

Соответствующее преобразование данных такого рода называют угловым преобразованием или преобразованием через арксинус. Оно достигается путем нахождения угла, синус которого равен квадратному корню из доли (процент/100). Математически это может быть выражено следующим образом: арксинус $\sqrt{Y}$, или синус $^{-1}\sqrt{Y}$. Таблица А.8 приложения может быть использована для преобразованных дат непосредственно из процентов.

Данные следует преобразовывать, если размах между отдельными датами превышает 40%. В других случаях это едва ли целесообразно. В таблице 12.13 приведены данные эксперимента с 24 вариантами и тремя повторениями по изучению всхожести семян салата-латука. Эксперимент проведен по схеме рендомизированных блоков. Варианты расположены в таблице в порядке возрастания их средних. Отметьте сильную тенденцию дисперсий принимать меньшую величину по краям, чем в центре ранжированного ряда. Это типично для биномиальных данных. Поскольку мы имеем логарифмы дисперсий (умноженных на 10), то можно провести проверку однородности дисперсий по Бартлетт-критерию.

Таблица 12.13. Число проросших семян салата-латука из 50 посеянных

Варианты	Повторения			Средние	s_i^2	Log $(10 \times s_i^2)$
	1	2	3			
1	0	0	1	0,33	0,33	0,519
2	0	1	0	0,33	0,33	0,519
3	0	0	1	0,33	0,33	0,519
4	0	2	0	0,67	1,33	1,124
5	2	0	0	0,67	1,33	1,124
6	0	2	3	1,67	2,33	1,367
7	7	10	7	8,00	3,00	1,477
8	11	12	15	12,67	4,33	1,637
9	13	18	18	16,33	8,33	1,921
10	22	16	13	17,00	21,00	2,322
11	24	13	18	18,33	30,33	2,482
12	23	21	29	24,33	17,33	2,239
13	24	29	29	27,33	8,33	1,921
14	37	28	27	30,67	30,33	2,482
15	42	41	40	41,00	1,00	1,000
16	39	41	45	41,67	9,33	1,970
17	41	45	40	42,00	7,00	1,845
18	47	41	43	43,67	9,33	1,970
19	45	42	48	45,00	9,00	1,954
20	46	42	48	45,33	9,33	1,970
21	49	46	48	47,67	2,33	1,367
22	48	49	48	48,33	0,33	0,519
23	50	49	48	49,00	1,00	1,000
24	49	49	50	49,33	0,33	0,519
Суммы					178,00	35,767
10 × средняя					74,167	
Log (10 × средняя)					1,8702	

Неприведенный $\chi^2 = 2,3026 [(\lg \text{средний} \times \Sigma df) - (df \text{ одной выборки} \times \Sigma \lg \text{кодов } s_i^2)] = 2,3026 [(1,8702 \times 48) - (2 \times 35,767)] = 41,99$;

$$C = 1 + \frac{1}{3(\text{число вариантов} - 1)} - \left(\frac{\text{число вариантов}}{df \text{ одного варианта}} - \frac{1}{\Sigma df} \right) =$$

$$= 1 + \frac{1}{3 \times 23} \left(\frac{24}{2} - \frac{1}{48} \right) = 1,1736.$$

Приведенный $\chi^2 = \chi^2 / C = 35,78$. Эта величина как раз существенна на 5%-ном уровне (табличное значение равно 35,172), и мы имеем довольно весомое доказательство неоднородности дисперсий.

Дисперсионный анализ исходных дат дает следующие результаты:

Источник вариации	df	SS	MS	F
Варианты	23	25266,0	1098,52	148,12 **
Ошибка	48	356,0	7,42	

Таблица 12.14. Данные таблицы 12.13, преобразованные с помощью арксинуса

Варианты	Повторения			Средние	s_i^2	$\text{Log } s_i^2$
	1	2	3			
1	0,0	0,0	8,1	2,70	21,870	1,34
2	0,0	8,1	0,1	2,70	21,870	1,34
3	0,0	0,0	8,1	2,70	21,870	1,34
4	0,0	11,5	0,0	3,83	44,083	1,64
5	11,5	0,0	0,0	3,83	44,083	1,64
6	0,0	11,5	14,2	8,57	56,863	1,75
7	22,0	26,6	22,0	23,53	7,053	0,85
8	28,0	29,3	33,2	30,17	7,323	0,86
9	30,7	36,9	36,9	34,83	12,813	1,11
10	41,6	34,4	30,7	35,57	30,723	1,49
11	43,9	30,7	36,9	37,17	43,613	1,64
12	42,7	40,4	49,6	44,23	22,923	1,36
13	43,9	49,6	49,6	47,70	10,830	1,03
14	59,3	48,4	47,3	51,67	44,003	1,64
15	66,4	64,9	63,4	64,90	2,250	0,35
16	62,0	64,9	71,6	66,17	24,243	1,38
17	64,9	71,6	63,4	66,63	19,063	1,28
18	75,8	64,9	68,0	69,57	31,543	1,50
19	71,6	66,4	78,5	72,17	36,843	1,57
20	73,6	66,4	78,5	72,83	37,043	1,57
21	81,9	73,6	78,5	78,00	17,410	1,24
22	78,5	81,9	78,5	79,63	3,853	0,59
23	90,0	81,9	78,5	83,47	34,903	1,54
24	81,9	81,9	90,0	84,60	21,870	1,34
Суммы					618,941	31,39
Средняя					25,789	
Log средней					1,411	

Преобразованные данные приводятся в таблице 12.14. Поскольку данные таблицы 12.13 основываются на выборках с объемом, равным 50, то каждую дату следует умножить на 2, чтобы получить проценты. Определенная система расположения дисперсий в ряду, наблюдаемая для исходных дат, полностью нарушается после преобразования.

Выполним проверку однородности дисперсий после преобразования дат по Бартлетт-критерию.

Неприведенный $\chi^2 = 2,3026(1,411 \times 48) - (31,39 \times 2) = 11,3933$;
 $C = 1,1736$, как и прежде. Приведенный $\chi^2 = \chi^2 / C = 9,708$.

В таблице А.6 (приложение) против числа степеней свободы, равном 23, находим, что вероятность случайного превышения полученной величины хи-квадрата составляет более 99%.

Дисперсионный анализ преобразованных дат, возможно, не изменяет выводов, сделанных на основе анализа исходных дат:

Источник вариации	df	SS	MS	F
Варианты	23	59487,8	2586,43	100,29 **
Ошибка	48	1237,9	25,79	

Важное отличие выясняется не после общего анализа, а в результате оценки частных разностей. Многогранговый критерий Дункана показывает, что:

1) пять существенных разностей, выявленных до преобразования, стали несущественными после преобразования: 7—8, 8—11, 10—12, 11—12 и 12—14;

2) после преобразования выявлено пять существенных разностей, которые не были таковыми до преобразования: 18—22, 19—23, 19—24, 20—23 и 20—24.

Встает вопрос: «Какие выводы взять за основу?» Ответ простой: мы должны взять выводы, вытекающие из более обоснованного анализа, в данном случае из анализа преобразованных дат.

Помните, мы преобразуем даты не для того, чтобы получать желательные нам результаты. Мы преобразуем даты с целью обеспечения обоснованного анализа и получения правильных выводов.

Другой момент, который нужно помнить при выполнении преобразований, — это то, что все критерии проверки существенности и выделения существенных средних следует проводить в большей мере с преобразованными данными, чем с исходными. Далее необходимо указать, что лучше вычислить прежде средние преобразованных дат, а затем сделать обратный переход к исходным единицам. Это позволит получить правильно взвешенные средние.

ПРЕДВАРИТЕЛЬНО ПРЕОБРАЗОВАННЫЕ ШКАЛЫ

Часто случается так, что нам хотелось бы выражать исследуемые явления в процентах, но мы находим, что связанные с ними точные измерения будут очень трудными и займут много времени. Рассмотрим, например, оценку пораженности клубней картофеля паршой. Надлежащей единицей измерения было бы выражение площади пораженной поверхности клубня в процентах, что очень трудно точно вычислить. Другим примером может служить определение процента поверхности листа, пораженного болезнью. Еще одним примером может быть процент погибших сорняков при применении различных гербицидов. Во всех перечисленных примерах мы могли бы, приложив максимум усилий, измерить довольно точно требуемые проценты, однако эта чрезвычайно кропотливая работа сильно ограничила бы количество обследованных делянок. Практически в целях повышения производительности таких учетов выполняют грубые визуальные оценки процентов вместо точных измерений.

Обычно подготавливают рабочую шкалу, например, от 0 до 10, часто используемую при учете эффективности борьбы с сорняками, где 0 означает нулевой и 10—100%-ный эффект. Если мы выберем шкалу с одинаковыми интервалами процентов между делениями, то получаемые данные следует подвергать угловому преобразованию, как и проценты точных измерений.

Возникает вопрос: «А почему бы заранее не преобразовать нам шкалу?» Другими словами, мы могли бы выбрать деления в процентах таким образом, чтобы после преобразования через арксинус получить

шкалу с равноинтервальными отметками, которые можно округлить до целых чисел.

Предположим, что мы создали такую шкалу от нуля до пяти. Одинаковые интервалы в угловом выражении составят 18° , поскольку нужно разделить 90° на пять. Поэтому необходимо найти такие величины процентов, которые после преобразования дадут углы в 0° , 18° , 36° , 54° , 72° и 90° .

Из таблицы А.8 (приложение) находим ближайшую к 18° величину, равную $18,4^\circ$, которая представляет угол, преобразованный из 10%. Следующее деление шкалы кажется проблематичным. Подыскивая по таблице величину, соответствующую углу в 36° , мы находим табличные значения углов в $35,7^\circ$ и $36,3^\circ$, преобразованные из 34 и 35% соответственно. Мы были бы склонны к тому, чтобы сделать на шкале отметку в 34,5%, однако это искажало бы смысл точности. В конце концов мы планируем выполнение лишь грубых визуальных оценок процентов. В этой связи использование для шкалы дробных процентов становится неоправданным, за исключением ее крайних отметок — ниже 5% или выше 95%.

В таблице 12.15 приводятся соответствующие проценты для всех шкал, используемых на практике.

Эти шкалы учитывают тот факт, что малые различия обычно легче выявляются в непосредственной близости от нуля или от 100%, чем

Таблица 12.15. Предварительно преобразованные разрядные шкалы с делениями от 0 до 100%

Разряды (классы)	4	5	6	8	10	15	18	20	24
0	0	0	0	0	0	0	0	0	0
1	15	10	7	4	2,5	1	0,75	0,7	0,5
2	50	35	25	15	10	4	3	2,5	2
3	85	65	50	30	21	10	7	5	4
4	100	90	75	50	35	17	12	10	7
5		100	93	70	50	25	18	15	10
6			100	85	65	35	25	20	15
7				96	79	45	33	27	20
8				100	90	55	42	35	25
9					97,5	65	50	42	31
10					100	75	58	50	37
11						83	67	58	43
12						90	75	65	50
13						96	82	73	57
14						99	88	80	63
15						100	93	85	69
16							97	90	75
17							99,25	95	80
18							100	97,5	85
19								99,3	90
20								100	93
21									96
22									98
23									99,5
24									100

около 50%. Фактически некоторые шкалы, используемые в прошлом, преднамеренно или подсознательно строились в соответствии с этими классами процентов.

Для картофеля использовалась шкала с делениями от 0% до 10%, шаблонами для которой послужили фотоснимки различной степени пораженности клубней; в приближенном виде она показана процентами в таблице 12.15. При обследовании пораженности яблок используется шкала на содержание крахмала, которая тесно соответствует шкале от 0% до 8%. При учете борьбы с сорняками, где применяют шкалу от 0% до 10%, отмечается тенденция предпочтительного использования разряда (класса) 1, чем оценки в 10% для малоэффективного приема и разряда 9 для случая почти полного уничтожения сорняков.

Для выбора требуемой шкалы следует решить, какое разумно обоснованное количество делений мы можем использовать. Слишком много делений усложняет шкалу и дает высокую точность, не требуемую для учетов. В случае использования шкалы с незначительным числом делений приходится выделять дробные разряды.

При анализе данных, полученных на основе предварительно преобразованных разрядных шкал, не следует проводить их преобразования. Кроме того, средние необходимо вычислять на основе разрядов до обратного перевода в проценты. Для обратного преобразования умножают среднее значение разряда на угловой интервал и находят соответствующую величину процентов в таблице А.8 (приложение). Например, если вариант имеет средний разряд, равный 1,4 при шкале от нуля до пяти, то угловой интервал будет равен $90^\circ/5=18^\circ$, а искомая величина в угловом выражении равна $18^\circ \times 1,4=25,2^\circ$. По таблице А.8 угол в $25,1^\circ$ соответствует 18%, что и будет искомой величиной взвешенной средней для научного отчета. Использование дробных величин процентов в отчетах едва ли оправдано, за исключением экстремально высоких или низких значений.

Несколько слов необходимо уделить разрядам для контрольных делянок. Следует различать два момента: 1) когда эти делянки включены в эксперимент как нулевой уровень некоторого фактора и подвергаются такой же вариации, как все другие градации фактора, и 2) когда они включены в качестве контроля для сравнения с ними всех других делянок. Во втором случае им дают часто нулевой разряд, а другие делянки блока сравниваются с ними. В этом случае данные контрольных делянок не следует включать в комплекс дисперсионного анализа. Контрольные делянки, произвольно получившие нулевые разряды, не имеют контролируемой дисперсии. Следовательно, их дисперсия отличается от дисперсий вариантов, что автоматически нарушает предположение об однородности дисперсий.

ЗАКЛЮЧЕНИЕ

1. Главные предположения, лежащие в основе дисперсионного анализа: случайность и нормальность распределения компонентов ошибки, однородность дисперсий, независимость дисперсий и средних и слагаемость главных эффектов.

2. Когда эти предположения серьезно нарушены, применение дисперсионного анализа становится необоснованным.

3. Для устранения несоответствия данных предположениям можно использовать преобразования.

4. Когда стандартные отклонения линейно сходны со средними, а главные эффекты не однозначны (мультипликативны), то логарифмическое преобразование устраняет оба из этих отклонений.

5. Данные учетов редких явлений, где дисперсии связаны со средними, следует преобразовывать через квадратные корни.

6. Данные, выражаемые в долях или процентах, следует преобразовывать через арксинус или угол.

7. Разрядные шкалы могут быть предварительно преобразованы на основе переменной шкалы процентов.

8. Если в расчетах используют преобразование, то все оценки существенности и выбор средних необходимо проводить, используя преобразованные данные.

9. При необходимости обратного перехода к исходным единицам это следует делать лишь после вычисления средних из преобразованных данных.

ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

ИДЕЯ

Термины «корреляция» и «регрессия» могут показаться трудноусвояемыми, однако основные идеи, заключенные в этих понятиях, настолько просты, что мы все используем их в каждодневной речи. Рассмотрим, к примеру, следующие известные поговорки:

- «Чем выше взлетишь, тем больше падать».
- «Чем меньше слов, тем больше дела».
- «Чем легче досталось, тем легче утрата».
- «Чем лучше настроение, тем легче работать».
- «Ветка колыхнется — дерево зашумит».

Все эти поговорки имеют несколько общих идей. Каждая включает две количественные переменные, причем величина одной зависит от величины другой. Статистики обозначают их как независимая и зависимая переменные. Более того, в этих конкретных поговорках скрыта идея о том, что по мере увеличения одной переменной возрастает другая. В статистике эту связь называют прямой или положительной корреляцией.

Рассмотрим другую группу поговорок:

- «Тише едешь — дальше будешь».
- «У малой лжи — длинные уши (ноги)».
- «Чем дороже подарок, тем меньше упаковка».

Здесь также имеется общая идея двух переменных, зависимость одной от другой, однако в данной связи есть небольшое отклонение. Увеличение одной переменной сопровождается уменьшением другой. Эту связь называют обратной или отрицательной корреляцией. Смысл корреляции не ограничивается этим простым определением. Подумайте о тех вопросах, встречающихся время от времени в сельскохозяйственной практике, которые имеют отношение к двум переменным. Какая зависимость существует между дозой вносимых удобрений и урожайностью культуры? Какая связь имеется между количеством потребленного корма и прибавкой массы тела животного? Как влияет количество произведенных продуктов на их цену? Как доза инсектицида связана с процентом погибших насекомых или с остаточным количеством препарата? Какова связь между размером фермы и ее доходом? Список подобных вопросов можно продолжать до бесконечности, однако теперь

стало понятнее, что в каждом из них подразумевается корреляция, независимо от того, упоминается ли этот термин или нет.

Другой пример корреляции, с которой мы сталкиваемся почти каждый день,— это обычные графики. Почти каждый график в сущности представляет картину корреляции между двумя переменными. Горизонтальная шкала, или абсцисса,— это обычно размах значений независимой переменной. Величины, откладываемые на вертикальной шкале, или ординате, являются значениями зависимой переменной. Представление данных в виде графика, как правило, является очень полезным начальным этапом их корреляционного анализа.

Итак, рассмотрев некоторые часто встречающиеся примеры корреляции, вы должны быть в состоянии понять абстрактное определение этого термина: «корреляция — это тенденция двух переменных к определенно выраженной зависимости между ними». Фактически связь может распространяться и на большее число переменных, так, например, закон спроса и предложения включает три переменных: цена, предложение и спрос. Для упрощения объяснений мы ограничимся пока примерами корреляции между двумя переменными.

Общепринято рассматривать одну из переменных в качестве зависимой от другой. Вопрос о том, какую переменную назвать зависимой и какую независимой, обычно не вызывает сомнений. Например, при изучении связи урожайности с удобрениями будет логичным рассмотреть зависимость урожайности от удобрений. В отношении связи цены и предложения мы рассматриваем обычно цену в качестве зависимой от предложения. С другой стороны, рынок дает ситуации, в которых цена становится зависимой от предложения. Часто измерение одной переменной и соответствующее измерение другой переменной разделяет промежуток времени. В таких случаях переменная, измеренная первой, называется независимой переменной. Иногда полезно исследовать корреляцию между парами измерений одной переменной. Например, изучение корреляции между ценами на продукты питания последовательного ряда лет и соответствующими ценами предшествующего года может позволить выявить циклические тренды в ценах.

В некоторых случаях нас не интересует, какая из переменных обозначена зависимой. Мы можем просто исследовать общее распределение двух переменных, где каждая переменная распределяется нормально. Такое распределение называется нормальным распределением, варьирующим в двух направлениях. Для описания этого распределения нужна оценка ρ (ρ), последняя величина является одним из параметров совокупности. Коэффициент корреляции r является лучшей оценкой ρ . Изучение корреляции между длиной предплечья и высотой человека может послужить примером ситуации, когда выбор той или другой переменной в качестве зависимой не будет иметь значения.

ИЗМЕРЕНИЕ КОРРЕЛЯЦИИ

До сих пор наше обсуждение корреляции касалось общей идеи двух переменных, связанных определенным образом. Это не потребовало больших знаний математики или статистики. Факт предполагаемой

связи двух переменных мало что даст. Для описания связи требуется дать ответ на два важных вопроса: как тесно связаны две переменные и является ли установленная зависимость действительной или случайной? Для ответа на первый вопрос требуется определенное измерение степени (тесноты) связи между двумя переменными. Для этого применяют коэффициент корреляции, обозначаемый символом r . После определения нескольких других терминов мы покажем, как вычисляют и объясняют эту величину. Ответ на второй вопрос можно получить из соответствующих таблиц вероятностей.

РЕГРЕССИЯ

Термин «регрессия» не был еще использован при обсуждении материала этой главы, за исключением начальной фразы. Что он означает? Помощь толкового словаря будет неэффективной, поскольку «регрессия» — это одно из тех неудачных выражений (как, например, термин «ошибка»), которые претерпели эволюцию, и их настоящее значение имеет малое сходство с первоначальным. Выражаясь кратко, регрессия — это количественное изменение одной переменной, приходящееся на единицу изменения другой. Это определение может вызвать критические замечания, поскольку с математической точки зрения оно обобщено или недостаточно точно, однако позволяет подчеркнуть главное различие между корреляцией и регрессией, что отвечает нашим целям. Отметим, что корреляция указывает на факт наличия связи и степень ее близости. Регрессия, с другой стороны, определяет природу этой связи.

Обратимся вновь к известным поговоркам и посмотрим, каким образом понятие регрессии сопровождает наши повседневные размышления:

«Пенс сбереженный равен пенсу заработанному».

«Синица в руках лучше, чем две в кустах».

«Лучше один раз сделать, чем девять раз пообещать».

«Лучше один раз увидеть, чем тысячу раз услышать».

Отметим, что все эти поговорки не только основаны на корреляции двух переменных, но даже больше — они характеризуют связь двух переменных в цифровом выражении. На основе этих поговорок можно построить таблицу (табл. 13.1).

Мы последовали общепринятой традиции: называть независимую переменную X , а зависимую переменную Y .

Таблица 13.1. Математическое выражение поговорок

Независимая переменная (X)	Зависимая переменная (Y)	Уравнение регрессии	Коэффициент регрессии
Сбереженный пенс	Заработанный пенс	$Y = X$	1
Пойманная синица	Две пойманных	$Y = 2X$	2
Один раз сделать	Девять раз пообещать	$Y = 9X$	9
Один раз увидеть	Тысячу раз услышать	$Y = 1000X$	1000

Третья колонка таблицы 13.1 озаглавлена как уравнение регрессии. Все они являются уравнениями прямой линии. Уравнение прямой линии имеет общий вид: $Y=a+bX$. Символ a обозначает отрезок оси ординат, поскольку когда $X=0$, то $Y=a$; таким образом, линия регрессии пересекает ось Y выше начала координат на величину a . Когда a равно нулю, линия проходит через начало координат, и если X равен нулю, то Y также будет равен нулю. Символ b обозначает наклон, поскольку эта величина определяет крутизну наклона линии. Из уравнения достаточно очевидно, что b — это количественное изменение Y , приходящееся на единицу изменения X . Это совпадает с упомянутым определением регрессии. Поэтому будет логичным называть b коэффициентом регрессии.

СРАВНЕНИЕ КОРРЕЛЯЦИИ И РЕГРЕССИИ

Какой вид анализа применять для этой или иной заданной проблемы? Некоторые статистики настаивают на строгом разграничении двух видов анализа. Такой подход зависит от того, относят ли имеющиеся данные к модели I, где X имеет фиксированные значения, или к модели II, где значения X распределяются случайно или подвержены ошибке.

Рассмотрим эксперимент, в котором преднамеренно используют несколько уровней некоторого варианта, повторяя каждый уровень несколько раз. В этом случае наибольший интерес будет представлять количественное изменение Y в связи с изменением уровня варианта (X). Это будет регрессия. С другой стороны, коэффициент корреляции (r), как оценка параметра совокупности (ρ), не будет иметь значения. Мы имеем дело с совокупностью, которая не обладает таким параметром. Тем не менее квадрат этого коэффициента (r^2), известный под названием «коэффициент детерминации», имеет действительное значение в такой ситуации. Он дает долю общей суммы квадратов вариантов, приходящейся на регрессию.

Рассмотрим случаи, в которых мы имеем дело с нормальным двумерным распределением и когда ни одна из переменных не может быть обозначена зависимой от другой. Эти случаи как раз соответствуют модели I, и нас прежде всего интересует степень связи двух переменных, измеряемая коэффициентом корреляции. Уравнение регрессии для определения величины одной переменной на основе другой не представляет большего интереса. Тем не менее мы вычислим два таких уравнения в зависимости от переменной, названной независимой. Фактически коэффициент корреляции будет представлять геометрическую среднюю двух коэффициентов регрессии, получаемых из уравнений.

На примере двух вышеприведенных случаев мы видим, что вид данных определяет первостепенность корреляции или регрессии, однако нельзя полностью разделить два вида анализов.

Между этими довольно четко отличающимися двумя случаями имеется много других, когда не возникает вопроса при выборе зависимой переменной, но необходимо решить вопрос о том, считать

ли независимую переменную случайной или фиксированной. Даже в экспериментах, варианты которых составляют конкретные количества некоторого материала, мы не можем утверждать, что каждая деланка или каждая ее повторность получают одинаковое количество материала. Тем не менее эти ошибки измерений составляют очень малые величины по сравнению с экспериментальной ошибкой случайной выборки из совокупности сильно варьирующих значений переменной. Поэтому величины X в таком эксперименте принимают за фиксированные.

Ситуация становится менее ясной, когда мы имеем дело не со стационарным экспериментом, а с парами измерений, сделанных на группе индивидуумов, взятых из совокупности. Если выборка отдельных единиц производилась совершенно случайно, то у нас не возникает других вопросов, кроме вопроса о том, имеем ли мы дело с проблемой регрессии модели II, где X -ы представляют случайные величины. С другой стороны, систематическая выборка, произведенная с целью получения заданного размаха значений X , позволяет нам рассматривать X -ы фиксированными величинами.

Теперь мы видим, что различие между моделью I и моделью II проблемы регрессии не является очень строгим. В примерах, подлежащих нашему обсуждению, а фактически в большинстве сельскохозяйственных исследований при описании связи между переменными нас прежде всего интересуют хорошо обоснованные уравнения регрессии. Кроме того, нас интересует, насколько близко уравнения регрессии соответствуют данным наблюдений; с этой целью мы вычисляем коэффициент детерминации (квадрат коэффициента корреляции).

Для иллюстрации общих методов анализа линейной корреляции и регрессии возьмем вначале пример с группой парных наблюдений.

РАСЧЕТЫ ЛИНЕЙНОЙ КОРРЕЛЯЦИИ

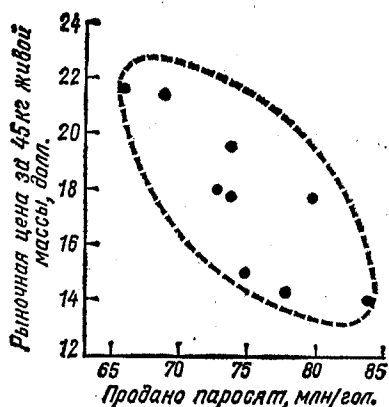
Доступным примером корреляции является зависимость цены от предложения. В таблице 13.2 показано производство поросят для рынка и цены на них с 1950 по 1959 г.

Имеется ли действительная связь между предложением и ценой за этот период? Первое, что мы отмечаем в таблице, это соответствие

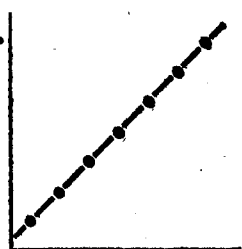
Таблица 13.2. *Поголовье поросят (млн. гол.), поступивших на рынок, и соответствующие цены за 100 фунтов (45 кг) их массы в долларах*

Годы	Поросята (X)	Цены (Y)	Годы	Поросята (X)	Цены (Y)
1950	73	18,0	1955	75	15,0
1951	79	20,0	1956	78	14,4
1952	80	17,8	1957	74	17,8
1953	69	21,4	1958	74	19,6
1954	66	21,6	1959	84	14,1

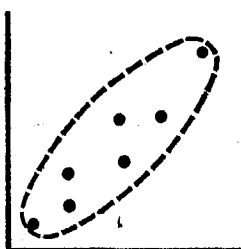
Рис. 13.1. Точечная диаграмма связи между ценой на поросят и их поголовьем, поступившим на рынок за год.



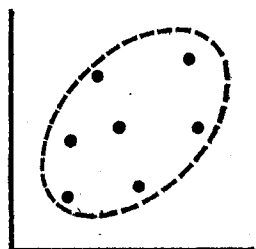
высоких цен меньшему количеству поступивших на рынок поросят, и наоборот. Эта взаимосвязь является убедительным доказательством отрицательной корреляции, которую мы можем ожидать. Представим эту закономерность в виде графика. Это легко достигается нанесением точек на миллиметровой бумаге, положение которых по отношению к оси X дает цены, а удаленность от оси Y вправо дает число поросят для соответствующего года (рис. 13.1).



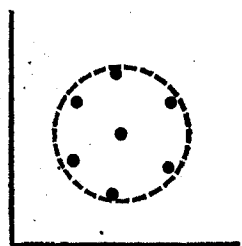
Полная положительная зависимость
 $r = +1$



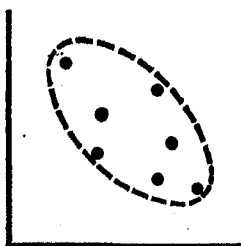
Сильная положительная зависимость
 $r = +0.8$



Слабая положительная зависимость
 $r = +0.1$



Нулевая
 $r = 0$



Средняя отрицательная
 $r = -0.5$



Полная отрицательная
 $r = -1$

Рис. 13.2. Различные виды точечных диаграмм с соответствующими им коэффициентами и названиями корреляции.

Таблица 13.3. Производство поросят для рынка и цены на них в ранговом выражении

Ранг предложения	Ранг цены	Разность между рангами (d)	d^2
8	5	3	9
3	3	0	0
2	6,5	-4,5	20,25
9	2	7	49
10	1	9	81
5	8	-3	9
4	9	-5	25
6,5	6,5	0	0
6,5	4	2,5	6,25
1	10	-9	81
Суммы		0,0	280,5

График подобного вида называется рассеянной (точечной) диаграммой. Если мы предполагали очень тесную корреляцию между предложением и ценой, то довольно случайный разброс этих точек может вызвать разочарование. Однако на графике можно отметить общую тенденцию более высокого расположения точек левой стороны по сравнению с правосторонними точками.

Точки кажутся падающими внутри растянутого эллипса (рис. 13.1), что является типичной диаграммой среднесильной корреляции. Другие виды точечных диаграмм, представленных на рисунке 13.2, могут оказаться полезными при анализе подобных графиков. Направление оси эллипса в нашем примере указывает на отрицательную корреляцию. Теперь мы готовы к тому, чтобы рассчитать степень взаимосвязи. Вначале используем сокращенный метод расчета.

Сокращенный метод. Он также известен под названием метода Спирмана для расчета ранговых различий и коэффициента ранговой корреляции *. Ни одно из этих названий не отображает легкости расчетов, предлагаемых методом Спирмана, даже без помощи счетных машин. Хотя метод и имеет серьезные недостатки, он очень удобен для быстрого получения приближенной оценки коэффициента корреляции. В таблице 13.3 приводятся расчеты корреляции на примере данных с ценами на поросят.

Во-первых, проранжируем наблюдения в каждой колонке от наибольшего к наименьшему значению. При наличии группы одинаковых значений каждый член группы получает усредненный ранг. Например, в обеих колонках вышеприведенной таблицы ранги 6 и 7 связаны одинаковым числом, поэтому оба числа получают номер 6,5.

Во-вторых, произведем вычитание второго из первого ранга в каждом ряду и внесем разность в колонку, озаглавленную d . Сумма

* Существует два коэффициента ранговой корреляции — Спирмана и Кендэла. (См. Ван дер Варден Б. Л. Математическая статистика. Юл Дж. Э., Кендэл М. Дж. Теория статистики.) — *Прим. ред.*

по этой колонке должна быть всегда равна нулю. Этот показатель используют при проверке расчетов.

В-третьих, возведем цифры колонки d в квадрат и полученную величину внесем в колонку d^2 . Второй этап можно опустить, поскольку квадраты числа легко запоминаются *, и их сразу заносят в колонку d^2 .

В-четвертых, подведем итог колонки d^2 . Этот итог обозначается суммой Σd^2 .

В-пятых, вычисляем коэффициент корреляции r по формуле

$$r = 1 - \left[\frac{6 \Sigma d^2}{n(n-1)(n+1)} \right],$$

где n — число пар наблюдений.

В нашем примере

$$r = 1 - \frac{6 \times 280,5}{10 \times 9 \times 11} = 1 - 1,70 = -0,70.$$

Величина, полученная в ответе, всегда должна находиться между $+1$ и -1 . Плюс или минус единица свидетельствует о полной корреляции, тогда как нуль указывает на отсутствие корреляции. Так, в нашем примере проявляется довольно сильная отрицательная связь, поэтому мы произведем более тщательные расчеты коэффициента, используя стандартный метод.

Стандартный метод. Он известен как метод вычисления коэффициента линейной корреляции по способу произведений.

В главе 2 мы указывали, что отклонение индивидуального значения Y от средней Y -ов ($Y - \bar{Y}$) можно обозначать малой буквой y . Подобно этому можно использовать символ x для ($X - \bar{X}$). Использование этих малых символов значительно упрощает многие из выражений в данной книге, и они будут использованы в этой и последующих главах.

Формула коэффициента корреляции имеет несколько форм записи. Вначале будет удобнее записать их для квадрата коэффициента (r^2), затем получить r путем извлечения квадратного корня из окончательного выражения.

$$r^2 = \frac{[\Sigma (X - \bar{X})(Y - \bar{Y})]^2}{\Sigma (X - \bar{X})^2 \Sigma (Y - \bar{Y})^2}. \quad (1)$$

Поскольку $x = X - \bar{X}$ и $y = Y - \bar{Y}$, то можно записать формулу (1) в сокращенном виде

$$r^2 = \frac{(\Sigma xy)^2}{\Sigma x^2 \Sigma y^2}. \quad (2)$$

Хотя эти формулы и являются простыми, прямые расчеты обычно затруднены возведением в квадрат громоздких десятичных дробей.

* Большинство d составляют малые значения целых чисел. Если они заканчиваются на 0,5, то возведение в квадрат можно производить в уме, используя следующее равенство: $(X + 0,5)^2 = X(X+1) + 0,25$; так, $4,5^2 = 4 \times 5 + 0,25 = 20,25$; $7,5^2 = 7 \times 8 + 0,25 = 56,25$ и т. д.

Чтобы избежать этих трудностей, мы воспользуемся следующим равенством:

$$\Sigma x^2 = \Sigma (X - \bar{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}.$$

Заменяя y на x там, где это возможно, мы можем записать формулу (3) в ином виде

$$r^2 = \left[\Sigma XY - \frac{\Sigma X \Sigma Y}{n} \right]^2 / \left[\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n} \right) \left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{n} \right) \right]. \quad (3)$$

В этом виде она называется «формулой для компьютеров».

Особое внимание следует уделить выражению в квадратных скобках, являющемуся числителе уравнения (3). Это выражение представляет сумму внутривариационных произведений. В отличие от знакомых сумм квадратов в знаменателе (они всегда должны быть положительными) суммы произведений могут быть либо положительными, либо отрицательными.

Используя формулу (3), мы можем теперь вычислить коэффициент корреляции для данных нашего примера по стандартному методу. Нам нужно получить ΣX , ΣY , ΣX^2 , ΣY^2 и ΣXY . В нашем примере находим $\Sigma X = 752$; $\Sigma Y = 179,7$; $\Sigma X^2 = 56\,804,0$; $\Sigma Y^2 = 3297,53$ и $\Sigma XY = 13\,420,40$. Отсюда

$$\begin{aligned} r^2 &= \left[13\,420,40 - \frac{752 \times 179,7}{10} \right]^2 / \left[\left(56\,804,0 - \frac{752^2}{10} \right) \left(3297,53 - \frac{179,7^2}{10} \right) \right] = [13\,420,40 - 13\,513,44]^2 / [(56\,804,0 - 56\,550,4) \times \\ &\times (3297,53 - 3229,21)] = (-93,04)^2 / (253,6 \times 68,32) = 0,4996; \\ r &= \sqrt{r^2} = \sqrt{0,4996} = -0,707. \end{aligned}$$

Отметьте, что знак перед r должен быть аналогичным знаку перед Σxy , в данном случае отрицательный. Сокращенный метод дал результат $(-0,70)$, очень близкий к результату стандартного метода $(-0,707)$. Не слишком радуйтесь этому совпадению. Обычно результаты двух методов не бывают столь близкими.

В главе 14 мы продемонстрируем тот случай, когда сокращенный метод дает полную корреляцию и приводит к крайнему заблуждению. Можно привести другие примеры, когда сокращенный метод слишком занижает коэффициент корреляции.

Исследователь может использовать сокращенный метод для быстрого получения ответа без помощи калькулятора или в том случае, когда считает приблизительный ответ достаточным. Для получения более точной оценки коэффициента корреляции и проверки его существенности необходимо использовать стандартный метод.

СТАТИСТИЧЕСКАЯ СУЩЕСТВЕННОСТЬ

В последнем абзаце мы упомянули выражение «существенность». Общая идея, заключенная в ней, является аналогичной той, которая указывалась в дисперсионном анализе. Мы выдвигаем гипотезу об

отсутствии корреляции между двумя переменными и что любое проявление связи обуславливается простой случайностью. Это обычно называется нулевой гипотезой. Далее мы ставили вопрос: «Если бы нулевые гипотезы оказались верными, то какова вероятность получения r , равного или превышающего его фактическую величину?». Если эта вероятность равна 5%, то мы называем корреляцию существенной. При утверждении о реальности данной связи мы рискуем оказаться в 5 случаях из 100 неправыми. Если вероятность равна 1% или меньше, то мы называем корреляцию очень существенной и отвергаем нулевую гипотезу с риском оказаться неправыми в 1 случае из 100.

К счастью, трудоемкие расчеты, необходимые для нахождения требуемых вероятностей, уже проделаны и сведены в таблицу (табл. А.7, приложение). Посмотрев на строчку таблицы, расположенную против 8 степеней свободы, мы находим, что вероятность случайного получения коэффициента корреляции, равного 0,7, находится в промежутке между 1 и 5%. Поэтому мы можем назвать корреляцию существенной. Однако следует проявлять осторожность при истолковании данных такого рода. Даже если корреляция существенна, необходимо соблюдать осторожность при заявлении, что изменения в сфере производства вызывают изменения в ценах. Как цена, так и предложение могут быть связаны со временем, третьей переменной, не учтенной проведенными вычислениями. В конце этой главы мы обсудим некоторые из ошибок, к которым может приводить анализ корреляции, и будет дан пример, показывающий, насколько рискованна интерпретация корреляции между двумя переменными, дополнительно связанными с фактором времени.

Почему использовалось 8 степеней свободы? Обычно число степеней свободы бывает на единицу меньше числа объектов, однако в данном примере, имея 10 пар наблюдений, для получения числа степеней свободы вычитают 2 из 10. Только теперь становится понятной та осторожность, проявленная в нашем утверждении, что обычно число степеней свободы на единицу меньше числа объектов. Здесь мы встречаем первое исключение из правила. Вычитание двух единиц принято объяснять тем, что одна степень свободы теряется при вычислении средней, а другая — при расчете регрессии.

Чтобы было понятнее, давайте посмотрим на это с другой стороны. Предположим, что мы имеем две пары наблюдений, любые две неидентичные пары. Они могут быть представлены на графике двумя точками, через которые можно провести линию. Мы называем эту линию линией регрессии, и две точки вполне обеспечивают такое построение. Поскольку это положение является верным для любых двух пар наблюдений, даже когда не наблюдается какой-либо связи между переменными, то было бы нелепо придавать значение коэффициенту корреляции, основанному лишь на двух парах. Точно так же как одно наблюдение не может сказать нам что-либо о вариабельности, так и две пары наблюдений ничего не говорят о корреляции.

Обратимся к простой иллюстрации этих положений: например, утренняя газета сообщает, что заводы фирмы «Додж» выпустили

за прошедший день 8 партий автомобилей, а ставка на определенные акции зафиксирована на высоте 51. Днем раньше фирма «Додж» сделала 4 партии, а ставка на те же акции составила 49. На основе этих данных можно заключить, что как число выпускаемых автомобилей, так и цена акций подвержены вариации. Мы можем даже оценить вариацию в обоих случаях, однако оценка будет очень грубой, поскольку в каждом случае она основывается лишь на 1-й степени свободы ($n-1$). Что можно сказать в отношении связи этих двух переменных? Легко установить, что

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} = \frac{4^2}{(8)(2)} = 1,$$

следовательно, $r=1$.

Разве не будет абсурдом утверждать в данном случае о полной зависимости между числом партий автомобилей и ценой определенных акций за один и тот же день? Однако это то, что со всей очевидностью предлагает коэффициент корреляции. Мы обойдем такой абсурдный результат, если скажем, что эта корреляция была основана на $n-2$ или нулевым числе степеней свободы и поэтому не имеет смысла.

Как часто слышали вы людей, делающих огульные заключения относительно корреляций, основанных на очень малом количестве наблюдений? Представьте обобщения человека, впервые в своей жизни совершившего перелет из Сан-Франциско в Денвер: «Чем дальше продвигаешься на восток, тем холоднее становится» (или, выразившись более глубокомысленно, этот человек может сказать: «Я наблюдал положительную корреляцию между температурой и географической долготой»). Этот пример не является преувеличением, поскольку нетрудно встретить людей, делающих пространные обобщения на основе скудных наблюдений. Мы должны стараться избегать таких ошибок, и статистическая наука помогает нам обходить эти ловушки.

ЛИНИЯ РЕГРЕССИИ

До сих пор в нашем примере с предложением и ценой мы определили лишь насколько связаны эти величины и вероятность того, что эта связь обусловлена случайностью. Мы ничего не узнали о том, как связаны эти две переменные.

Если мы согласимся, что эта связь линейная, и она лучше всего описывается прямой линией, то вопрос будет сводиться к нахождению конкретной прямой линии, обеспечивающей самое близкое соответствие данным наблюдений. Что подразумеваем мы под самым тесным соответствием? Из графика видно, что провести прямую линию через все точки данных наблюдений является невыполнимой задачей. Какую бы линию мы не построили, несколько точек будут находиться вне ее. Мы измеряли вариации внутри одной группы наблюдений путем расчета суммы квадратов отклонений от средней. В данном случае будет логичным измерение вариации линии путем расчета суммы квадратов отклонений от нее. Используя эту меру в качестве критерия

близости точек, мы попытаемся найти прямую линию, которая сделает сумму квадратов отклонений настолько малой, насколько это возможно. Такая процедура называется методом наименьших квадратов. Те из вас, кто знаком с вычислениями по этому методу, сразу примут эту проблему как типичный пример нахождения минимального значения функции.

Решение этой задачи оказывается очень простым. Уравнение, составленное на основе отклонений X и Y от их средних и наилучшим образом соответствующее линии, будет иметь вид

$$\hat{y} = \left(\frac{\sum xy}{\sum x^2} \right) x$$

(символ $\hat{y}$ читается как «оценочная величина y »).

Выражение $\sum xy / \sum x^2$ представляет коэффициент регрессии, поскольку оно дает нам оценочное изменение y на каждую единицу изменения x . Это соответствует нашему определению регрессии. Поскольку мы уже обозначали коэффициент регрессии символом b , то теперь мы можем записать $b = \sum xy / \sum x^2$. Более точно мы должны называть это выражение как «коэффициент регрессии Y на X » и использовать символ b_{yx} . Если b не имеет подстрочного указателя, то обычно подразумевается именно этот коэффициент.

Вышеприведенное уравнение можно записать с помощью символов фактических наблюдений вместо символов отклонений от средних. Мы можем записать $(\hat{Y} - \bar{Y}) = b(X - \bar{X})$ и получить после преобразования $\hat{Y} = (\bar{Y} - b\bar{X}) + bX$.

Если мы обозначим $\bar{Y} - b\bar{X}$ через a , то уравнение примет вид $\hat{Y} = a + bX$, что представляет (уже упомянутое в начале обсуждения регрессии) уравнение прямой линии с указанием наклона к оси абсцисс и отрезка, отсекаемого на оси ординат.

Теперь покажем, как применить это уравнение к нашим данным. Мы имеем уже все суммы, необходимые для расчета r , коэффициента корреляции. Мы установили, что:

$$\sum X = 752, \text{ тогда } \bar{X} = \frac{752}{10} = 75,2,$$

$$\sum Y = 179,7; \text{ откуда } \bar{Y} = 17,97,$$

$$\sum xy = -93,04,$$

$$\sum x^2 = 253,6; \text{ тогда } b = -\frac{93,04}{253,6} = -0,367.$$

Заменяв буквенные выражения на цифры в уравнении

$$\hat{Y} = (\bar{Y} - b\bar{X}) + bX,$$

мы получим:

$$\hat{Y} = [17,97 - (-0,367) 75,2] + (-0,367) X;$$

$$\hat{Y} = 45,57 - 0,367 X.$$

Это уравнение можно выразить в следующих словах: начиная с базовой цены в 45,57 доллара за хандредвейт (американский центнер,

Таблица 13.4. Фактические и расчетные (ожидаемые) цены на поросят

X	Y	$\hat{Y}=45,57-0,367X$	$d=Y-\hat{Y}$	d^2
73	18,0	18,8	-0,8	0,64
79	20,0	16,6	3,4	11,56
80	17,8	16,2	1,6	2,56
69	21,4	20,2	1,2	1,44
66	21,6	21,4	0,2	0,04
75	15,0	18,1	-3,1	9,61
78	14,4	16,9	-2,5	6,25
74	17,8	18,4	-0,6	0,36
74	19,6	18,4	1,2	1,44
84	14,1	14,7	-0,6	0,36
Суммы			0,0	34,26

равный 45,36 кг), каждая единица (миллион голов) ежегодного увеличения продажи поросят сопровождается средним снижением цены на 0,367 доллара за хандредвейт. В таблице 13.4 сравниваются фактические величины Y с оценочными величинами ($\hat{Y}$ -ов), рассчитанными на основе уравнения регрессии.

Сумма отклонений, равная нулю, служит проверкой правильности расчетов (без учета погрешностей округления). Сумма квадратов отклонений может быть вычислена значительно проще по формуле

$$\Sigma d^2 = (1-r^2)\Sigma y^2.$$

В нашем примере

$$\Sigma d^2 = (1-0,4996) \times 68,32 = 34,19.$$

Величина, полученная в ответе, очень близка к цифре 34,26, имеющейся в таблице 13.4. Небольшая разница обусловлена погрешностью округлений.

Эта сумма квадратов называется суммой квадратов отклонения от регрессии, а корень квадратный из выражения $\Sigma d^2/(n-2)$ называется стандартной ошибкой оценки. Эта ошибка как раз представляет стандартную ошибку другого вида, подобную ранее встречавшейся. Она является мерой отклонения от линии регрессии.

Обычно не требуется обязательно проходить через все трудности построения таблицы, подобной таблице 13.4, с целью проверки правильности линии регрессии. Построение линии регрессии на основе точечной диаграммы обычно позволяет избежать грубых ошибок. Построение линии — очень простая задача, поскольку для изображения любой линии необходимо иметь лишь две точки. Одна из них может находиться на оси Y , на определенном расстоянии (в данном случае 45,57) от начала координат. Другая точка может быть представлена $\bar{X}$ (средним значением X) и $\bar{Y}$ (средним значением Y). Линия, проходящая через эти две точки, и будет искомой линией регрессии. На рисунке 13.3 показана линия, построенная по данным нашего

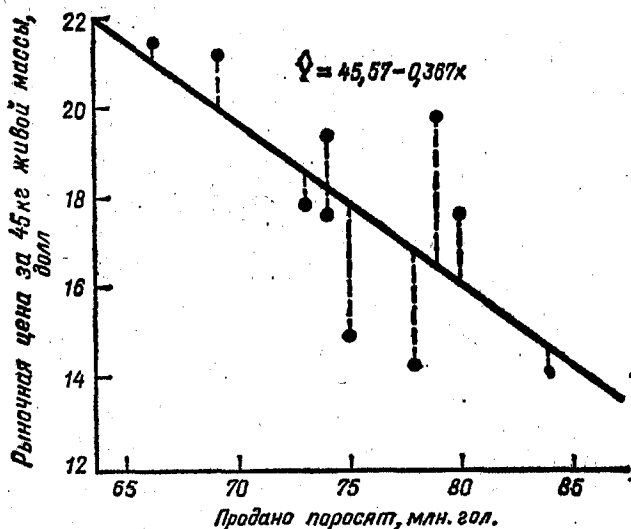


Рис. 13.3. Линия регрессии и отклонения от регрессии на основе данных о продаже поголовья поросят и цены на них.

примера и проведенная через поле полученных точек. Пунктирные линии, соединяющие точки с линией регрессии, показывают величины отклонений. Обратите внимание, что шкалы (масштаб) графика для оси ординат и абсцисс не начинаются с нуля. Они включают лишь несколько большее количество точек, чем фактический размах наблюдений.

Вы обратили внимание, что отклонения представлены вертикальными линиями. Именно сумму квадратов этих отклонений мы свели к минимуму для получения самого близкого соответствия линии точкам. Например, мы желаем построить такую линию, когда сумма квадратов горизонтальных отклонений от точек до линии будет минимальной. Будет ли построенная линия такой же, как первая? Ответ останется отрицательным до тех пор, пока мы не будем иметь функциональную зависимость. Эта новая линия будет иметь уравнение

$$\hat{x} = \left(\frac{\sum xy}{\sum y^2} \right) y.$$

Выражение $\sum xy / \sum y^2$ называется коэффициентом регрессии X на Y и обозначается символом b_{xy} . Теперь стало понятнее, почему мы соблюдали осторожность, подчеркивая, что под символом b понимают b_{yx} (регрессию Y на X) до тех пор, пока не рассматривается обратный случай.

По этой причине следует упомянуть, что имеются две наиболее подходящие линии, в соответствии с которыми избирается направление отклонений. Отметим, что

$$b_{yx} \cdot b_{xy} = \frac{\sum xy}{\sum x^2} \cdot \frac{\sum xy}{\sum y^2} = r^2.$$

Таблица 13.5. Результаты регрессионного анализа, представленные в виде таблицы дисперсионного анализа

Источник вариации	Степени свободы	Сумма квадратов	Средний квадрат	F
Общий	9	$\Sigma y^2 = 68,32$	—	
Регрессия	1	$r^2 \Sigma y^2 = 34,13$	34,13	7,99 *
Отклонение от регрессии	8	$(1-r^2) \Sigma y^2 = 34,19$	4,27	

* См. пояснения к таблице 11.2.

Это равенство открывает связь между коэффициентами регрессии и коэффициентом корреляции.

Теперь мы можем ответить на вопросы, поставленные при анализе данных в нашем примере:

1) насколько близкой была связь между предложением и ценой? Ответ: очень близкой. Коэффициент корреляции был равен минус 0,7, а ± 1 соответствовала бы функциональной зависимости;

2) какова вероятность, что такая корреляция может быть обусловлена случаем? Ответ: вероятность случайного получения коэффициента корреляции такой величины на основании 10 пар наблюдений будет находиться между 5 и 1%;

3) какое уравнение лучше описывает связь между ценой (Y) и предложением (X), исходя из этих данных?

Ответ: $\bar{Y} = 45,57 - 0,367 X$;

4) как точно соответствует эта линия данным наблюдениям? Ответ: сумма квадратов отклонений наблюдаемых точек от линии была равной 34,19, или составляла около половины общей вариации цен. Таким образом, лишь половина вариации цен была некоторым образом обусловлена вариацией предложения. Это показано в простой таблице дисперсионного анализа (табл. 13.5).

Отметьте в этой таблице, что r^2 составляет часть общей суммы квадратов, приходящейся на регрессию, а выражение $(1-r^2)$, иногда называемое коэффициентом отклонения, составляет неучтенную регрессией долю.

Тот факт, что полученная величина $F=7,99$ занимает промежуточное положение между величинами, требуемыми для 5% (5,32) и 1% (11,26) уровней при 1 и 8 степенях свободы, подтверждает ранее данный нами ответ на вопрос 2.

Фактически не будет иметь значения, используем ли мы величину F из таблицы А.3. или величину r из таблицы А.7 (приложение). Оба критерия идентичны, что легко доказывается. Это можно показать на примере получения F с помощью выражений таблицы дисперсионного анализа (табл. 13.5):

$$F = \frac{r^2 \Sigma y^2}{(1-r^2) \Sigma y^2 / (n-2)} = r^2 (n-2) / (1-r^2).$$

Используя это уравнение, можно рассчитать, что

$$r^2 = \frac{F}{F + n - 2}.$$

Мы можем подставить в формулу требуемую величину F , извлечь квадратный корень и получить требуемую величину r . Например, требуемая величина F для 1%-ного уровня при 1 и 8 степенях свободы равна +11,26. Подставив эту величину в вышеприведенную формулу, мы получим

$$r^2 = \frac{11,26}{11,26 + 10 - 2} = 0,5846,$$

$$r = 0,7646.$$

Полученный результат соответствует табличной величине (табл. А.7) для 8 степеней свободы и 1%-ного уровня значимости.

ДОВЕРИТЕЛЬНЫЕ ПРЕДЕЛЫ

Средний квадрат отклонения (DMS) обеспечивает основу для количественного расчета нескольких доверительных пределов. Дисперсия коэффициента регрессии равна

$$s_b^2 = \frac{DMS}{SSX},$$

а доверительные пределы составляют

$$b \pm t(s_b).$$

В нашем примере с производством поросят для рынка и ценами на них

$$s_b^2 = \frac{4,27}{253,6} = 0,0168,$$

$$s_b = 0,1298.$$

Табличная величина t для 8-ми степеней свободы на 5%-ном уровне равна 2,306, так что 5%-ные доверительные пределы будут равны $-0,367 \pm 2,306(0,1298) = -0,367 \pm 0,299 = -0,666$ и $+0,068$.

Величина t на 1%-ном уровне равна 3,355, и 1%-ные доверительные пределы будут равны

$$-0,367 \pm 3,355(0,1298) = -0,367 \pm 0,435 = -0,802 \text{ и } +0,068.$$

Отметьте, что 5%-ный доверительный интервал не включает нулевое значение, а 1%-ный интервал включает. Это находится в соответствии с ранее сделанными заключениями о том, что регрессия существенна на 5%-ном уровне и несущественна на 1%-ном уровне.

Оценки Y , обозначаемые Y , подвержены ошибкам двух видов, обусловленных дисперсией средней и дисперсией коэффициента ре-

Таблица 13.6. Дисперсии, стандартные ошибки и 5%-ные доверительные пределы $\hat{Y}$ и $\hat{Y}$ в зависимости от конкретных значений X (по данным корреляции между ценой и производством поросят для рынка) *

		$\hat{Y}$	$s_{\hat{Y}}^2$	$s_{\hat{Y}}$	$t(s_{\hat{Y}})$	Нижний предел	Верхний предел	s_Y^2	s_Y	$t(s_Y)$	Нижний предел	Верхний предел
65,2	-10	21,64	2,11	1,45	3,35	18,29	24,99	6,28	2,53	5,82	15,82	27,46
67,2	-8	20,91	1,50	1,23	2,83	18,08	23,74	5,77	2,40	5,54	15,37	26,45
69,2	-6	20,17	1,03	1,02	2,34	17,83	22,51	5,30	2,30	5,31	14,86	25,48
71,2	-4	19,44	0,70	0,83	1,92	17,52	21,36	4,97	2,23	5,14	14,30	24,58
73,2	-2	18,71	0,49	0,70	1,61	17,10	20,32	4,76	2,18	5,03	13,68	23,74
75,2	0	17,97	0,43	0,65	1,51	16,46	19,48	4,70	2,17	5,00	12,97	22,97
77,2	2	17,23	0,49	0,70	1,61	15,62	18,84	4,76	2,18	5,03	12,20	22,26
79,2	4	16,50	0,70	0,83	1,92	14,58	18,42	4,97	2,23	5,14	11,36	21,64
81,2	6	15,77	1,03	1,02	2,34	13,43	18,11	5,30	2,30	5,31	10,46	21,08
83,2	8	15,04	1,50	1,23	2,83	12,21	17,87	5,77	2,40	5,54	9,50	20,58
85,2	10	14,30	2,11	1,45	3,35	10,95	17,65	6,28	2,53	5,82	8,48	20,12

* В расчетах использована величина $t=2,306$ (табличное значение для 5%-ного уровня и восьми степеней свободы).

грессии. Дисперсия $\hat{Y}$ равна

$$s_{\hat{Y}}^2 = DMS \left(\frac{1}{n} + \frac{x^2}{SSX} \right).$$

Отметьте, что размер этой дисперсии зависит от величины x (отклонения X от среднего значения $\bar{X}$). Доверительные пределы для $\hat{Y}$ равны $\hat{Y} = \pm t(s_{\hat{Y}})$.

В нашем примере

$$s_{\hat{Y}}^2 = 4,27 \left(\frac{1}{10} + \frac{x^2}{253,6} \right) = 0,427 + 0,0168 x^2.$$

Доверительные интервалы для нескольких значений x даны в таблице 13.6. Нанесение этих значений на графике дает «доверительный пояс» вокруг линии регрессии, ограниченный двумя кривыми, как показано на примере внутреннего пояса на рисунке 13.4.

Доверительные пределы, только что рассчитанные нами, применяют для средних значений совокупностей Y -величин, связанных с конкретными величинами X . Часто больший интерес представляет определение доверительных пределов для прогноза отдельных величин Y на основе конкретных величин x . Здесь мы должны принять во внимание дополнительный источник ошибки. В дополнение к ошибке коэффициента регрессии и ошибки средней мы имеем вариацию индивидуальных значений вокруг выборочной средней.

Общая вариация Y для конкретной величины $X = \bar{X}$ или x равна

$$s_{\hat{Y}}^2 = DMS \left(1 + \frac{1}{n} + \frac{x^2}{SSX} \right).$$

а доверительные пределы равны

$$\bar{Y} \pm t(s_Y).$$

Пределы для избранных величин x приводятся в таблице 13.6 и нанесены на графике (рис. 13.4) в виде внешнего доверительного пояса.

РЕГРЕССИОННЫЙ АНАЛИЗ ДАННЫХ ЭКСПЕРИМЕНТА, ИМЕЮЩЕГО ПОВТОРНОСТЬ

В главах 6, 9 и 10 было показано использование ортогональной группы коэффициентов при определении суммы квадратов по линейной регрессии. Этот метод применим лишь для определенных сочетаний градаций вариантов, а общие методы этой главы можно применять для любых сочетаний градаций вариантов.

Для иллюстрации общих методов будут использованы примеры из главы 10, а позднее в главе 15 те же самые данные будут проанализированы с помощью сокращенных методов.

В таблице 13.7 символом Y обозначены суммы пяти сроков уборки. Использование сумм вместо средних снижает погрешность округления данных. Мы оценим их существенность с помощью прямолинейной регрессии:

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{5} = 55 - \frac{15^2}{5} = 10;$$

$$\Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{5} = 571\,372,24 - \frac{1600^2}{5} = 59\,372,24;$$

$$\Sigma xy = \Sigma XY - \frac{\Sigma X \Sigma Y}{5} = 5551,0 - \frac{15(1600)}{5} = 751,0;$$

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{751}{10} = 75,1;$$

$$a = \bar{Y} - b\bar{X} = 320 - 75,1(3) = 94,7;$$

$$r^2 = \frac{(\Sigma XY)^2}{(\Sigma x^2)(\Sigma y^2)} = \frac{751^2}{(10)(59\,372,24)} = 0,94994;$$

$$SSY \text{ (на поделяночной основе)} = \frac{\Sigma y^2}{16} = \frac{59\,372,24}{16} = 3710,765;$$

$$SS \text{ (регрессия)} = r^2(SSY) = 0,94994(3710,765) = 3525,004;$$

$$SS \text{ (отклонения)} = (1 - r^2)(SSY) = 0,05006(3710,765) = 185,761.$$

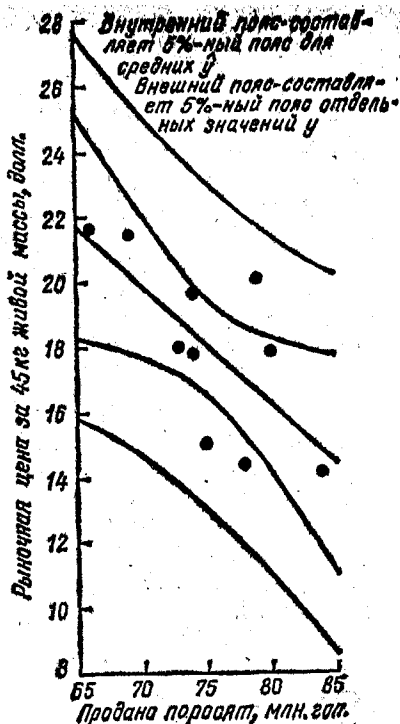


Рис. 13.4. Доверительные пояса вокруг линии регрессии.

Таблица 13.7. Суммы (ΣY) по срокам уборки (X) эксперимента с сахарной свеклой (см. главу 10)

X	ΣY	X^2	$X \Sigma Y$	ΣY^2	$\bar{Y}$	$\Sigma Y - \bar{Y}$	$(\Sigma Y - \bar{Y})^2$
1	140,0	1	140,0	19 600,0	169,8	-29,8	888,04
2	267,2	4	534,4	71 395,84	244,9	22,3	497,29
3	335,2	9	1005,6	112 359,04	320,0	15,2	231,04
4	417,0	16	1668,0	173 889,00	395,1	21,9	479,61
5	440,6	25	2203,0	194 128,36	470,2	-29,6	876,16
Итого 15	1600,0	55	5551,0	571 372,24	1600,0	0,0	2972,14

Сумма квадратов отклонений (последняя колонка таблицы 13.7) может быть приведена к поделяночной основе путем деления на 16 ($2972,14/16=185,759$), что соответствует вышеприведенной величине без учета погрешности округлений.

Уравнение регрессии было рассчитано на основе сумм по вариантам, уравнение для оценки средних получают после деления значений a и b на 16.

$$\hat{Y} = 5,91875 + 4,69375X.$$

Суммы квадратов могут быть сведены в итоговой таблице дисперсионного анализа (табл. 13.8).

Обратите внимание, что при анализе данных о поголовье поросят, где мы имели дело с индивидуальными парами наблюдений, для оценки среднего квадрата регрессии был использован средний квадрат отклонений. В стационарном эксперименте, имеющем повторность, для оценки обоих средних квадратов регрессии и отклонений можно использовать средний квадрат ошибки.

Очень существенное значение среднего квадрата регрессии в нашем примере указывает на выраженную тенденцию роста урожаев сахарной свеклы по мере движения от первого срока уборки к последующему (имеются в виду сроки в данном опыте). F -величина для отклонения от регрессии очень существенна, хотя и не настолько велика, как для регрессии. Это указывает на наличие очень существенного дополни-

Таблица 13.8. Дисперсионный анализ регрессии урожаев сахарной свеклы на сроки уборки по данным таблицы 13.7

Источник вариации	df	SS	MS	$F_{\text{факт}}$	$F_{\text{теор}}$	
					5%	1%
Сроки уборки	4	3710,765	927,691	111,92	3,26	5,41
Регрессия	1	3525,004	3525,004	425,26	4,75	9,33
Отклонения	3	185,761	61,920	7,47	3,49	5,95
Ошибка	12	99,467	8,289			

тельного источника вариации, помимо положительного линейного тренда, воздействующего на урожай.

Мы исследуем некоторые из возможных источников вариации в следующей главе.

ЗАБЛУЖДЕНИЯ (ЛОВУШКИ)

Вероятно, ни один из разделов статистики не подвержен в большей мере неправильному истолкованию, чем корреляция и регрессия. Утверждение о том, что «исследователь может доказать все с помощью статистики», является верным лишь для того случая, когда исследователь игнорирует некоторые из основных принципов. Наиболее часто забывают два основных принципа корреляции:

1) полное название коэффициента корреляции — коэффициент линейной корреляции;

2) ничто в определении корреляции не указывает или не подразумевает, что связь между двумя переменными является причинно-следственной. Следующие примеры показывают, как легко можно попасть впросак.

Слабая корреляция не всегда означает отсутствие связи. Посмотрите на следующие пары цифр:

X	0	1	2	3	4	5	6	7	8	9	10
Y	0	144	256	336	394	400	384	336	256	144	0

Если мы вычислим коэффициент корреляции между X и Y , то найдем, что он равен нулю (можете проверить и убедиться). Однако заключение об отсутствии связи между X и Y будет совершенно неправильным в данном случае. X — время (в секундах) нахождения стрелы в воздухе после вертикального выстрела с первоначальной скоростью, равной 160 фут/сек (48,8 м/сек). Y — высота подъема стрелы в футах (1 фут=30,48 см) за данное время. Разумеется, будет абсурдным заключение об отсутствии связи между высотой и временем полета стрелы. Где была допущена ошибка?

Важное слово «линейная», всегда подразумеваемое, когда речь идет о коэффициенте корреляции, в данном примере было не принято во внимание. Верно, что ни одна из прямых линий не обеспечит близкого соответствия этим данным, однако уравнение $Y=160X-16X^2$ дает полное соответствие. Это уравнение является уравнением параболы.

Этот пример указывает на необходимость учета криволинейных зависимостей, которые могут лучше соответствовать данным по сравнению с простой линейной зависимостью. Способы анализа данных подобного рода будут рассмотрены позже.

Сильная корреляция не обязательно означает причинно-следственную зависимость. Рассмотрим данные таблицы 13.9, на основе которых мы можем вычислить коэффициент корреляции и уравнение регрессии.

Высокая величина коэффициента корреляции (0,937) указывает на тесную связь X и Y . Исследователь может утверждать, что изме-

Таблица 13.9 *Пятидесятый пар сильно коррелируемых данных*

X	Y	$\Sigma X =$	5 669
395	73	$\bar{X} =$	377,9
389	78	$\Sigma X^2 =$	2 163 935
343	85	$(\Sigma X)^2/15 =$	2 142 504
344	91	$\Sigma \hat{X}^2 =$	21 431
357	100	$\Sigma Y =$	1 768
359	109	$\bar{Y} =$	117,9
368	119	$\Sigma Y^2 =$	218 482
395	125	$(\Sigma Y)^2/15 =$	208 388
414	129	$\Sigma y^2 =$	10 094
406	135	$\Sigma XY =$	681 962
385	142	$(\Sigma X \Sigma Y)/15 =$	668 186
394	139	$\Sigma xy =$	13 776
404	140		
420	147		
446	156		

$$r^2 = (\Sigma xy)^2 / \Sigma x^2 \Sigma y^2 = (13\,776)^2 / (21\,431 \times 10\,094) = 189\,778\,176 / 216\,324\,514 = 0,8773$$

$$r = \sqrt{0,8773} = 0,937 \text{ (коэффициент корреляции)}$$

$$b = \Sigma xy / \Sigma x^2 = 13\,776 / 21\,431 = 0,643 \text{ (коэффициент регрессии)}$$

$$a = \bar{Y} - b\bar{X} = 117,9 - 243,0 = -125,1 \text{ (свободный член)}$$

$$\hat{Y} = -125,1 + 0,643 X \text{ (уравнение регрессии)}$$

нение X на единицу измерения вызывает изменение Y на 0,643. Давайте посмотрим, что представляет X и Y . X — количество сигарет, используемых ежегодно в США (в миллиардах) с 1944 по 1958 г. Величины Y обозначают числовые индексы производства сена и фуражных культур на человека в час за тот же период. Потребуется большая перегрузка воображения, чтобы увидеть какую-либо прямую причинно-следственную связь между потреблением сигарет и эффективностью производства сена. Просто так случилось, что обе эти переменные устойчиво возрастали в течение рассмотренного периода времени.

Этот пример указывает, что коэффициент корреляции измеряет степень связи между двумя переменными, но ничего не говорит о том, является ли эта связь причинно-следственной. Решение этого вопроса является делом исследователя и должно опираться на глубокое знание изучаемых переменных.

Корреляция между частью и целым. Несколько лет назад на симпозиуме метеорологов была представлена работа по исследованию длины вегетационного периода, ограничиваемого устойчивыми заморозками. Отмечалась слабая связь или ее отсутствие между последним весенним и первым осенним заморозками за длительный период времени. Вместе с тем отмечалась довольно сильная корреляция между последним заморозком весной и длиной вегетационного периода.

Если мы рассмотрим это второе заключение, то отметим, что длина вегетационного периода полностью определяется двумя показателями:

началом (последний весенний заморозок) и концом (первый осенний заморозок). Если переменная состоит из двух или большего числа независимых частей, то произвольно появляется связь между частью и целым, что можно легко доказать. Эта зависимость проста: $r = (\text{стандартное отклонение от части}) / (\text{стандартное отклонение от целого})$. Для примера с заморозками в случае приблизительно одинаковой вариабельности сроков весенних и осенних заморозков мы будем ожидать, что корреляция между сроками весенних заморозков и длиной вегетационного периода составит около $\sqrt{0,5} = 0,707$. Сделанное заключение о наличии корреляции между последним весенним заморозком и длиной вегетационного периода было хотя и правильным, но тривиальным.

Экстраполирование не только заманчиво, но и опасно. Часто ряд наблюдений попадает в ограниченный интервал значений двух изучаемых переменных. Если получен высокий коэффициент корреляции, то появляется сильное желание продолжить линию регрессии за пределы наблюдаемых значений и попытаться предсказать, что произойдет с величиной Y , если взять значения X выше или ниже фактически наблюдаемых. Это и составляет основу экстраполяции. Этот прием является рискованным потому, что многие переменные, связанные между собой криволинейной зависимостью, могут давать высокую линейную корреляцию на коротких отрезках кривой при малых выборках.

В таблице 13.10 даны результаты измерений диаметра 10 луковиц лука в пределах от 50 до 70 мм и их соответствующие массы в граммах.

Расчеты r , коэффициента корреляции и уравнения регрессии выполняются в такой последовательности:

$$\begin{aligned} \Sigma X &= 614,9, & \Sigma Y &= 1031,9, \\ \bar{X} &= 61,49; & \bar{Y} &= 103,19; \\ \Sigma X^2 &= 38,192,1 & \Sigma Y^2 &= 113\,247,79, & \Sigma XY &= 65\,014,60, \\ (\Sigma X)^2/n &= 37\,810,20, & (\Sigma Y)^2/n &= 106\,481,76, & \Sigma X \Sigma Y/n &= 63\,451,53, \\ \Sigma yx^2 &= 381,97; & \Sigma y^2 &= 6766,03; & \Sigma xy &= 1563,07, \end{aligned}$$

где

$$\begin{aligned} r^2 &= (1563,07)^2 / (381,97 \times 6766,03) = 0,9454; \\ r &= \sqrt{0,9454} = 0,97 \text{ (коэффициент корреляции);} \\ b &= 1563,07 / 381,97 = 4,092 \text{ (коэффициент регрессии);} \\ a &= 103,19 - (4,092)(61,49) = -148,43 \text{ (свободный член) и} \\ \hat{Y} &= 4,092X - 148,43 \text{ (уравнение регрессии).} \end{aligned}$$

Корреляция между диаметром и массой, равная 0,97, является очень сильной. (Это и не удивительно.) Внутри интервала диаметров от 50 до 70 мм уравнение прямой линии очень хорошо описывает связь между двумя переменными.

Теперь давайте сделаем экстраполяцию и посмотрим, что изменится. Луковица диаметром в 92,4 мм имела массу 300,2 г, а наша оценка массы, вычисленная на основе уравнения регрессии, равнялась 229,7.

Таблица 13.10. Результаты измерений 10 луковиц

Диаметр (X)	Масса (Y)	Диаметр (X)	Масса (Y)
51,0	63,4	67,1	125,6
66,2	115,3	58,1	80,0
69,2	146,6	53,9	78,7
69,5	132,6	63,0	112,8
56,9	80,7	60,0	96,2

Экстраполяция привела к ошибке нашей оценки на 70,5 г. С другой стороны, луковица диаметром в 37,8 мм имела массу в 27,8 г, тогда как экстраполяция дала оценку массы, равную 6,2 г. Экстраполирование еще меньших значений X дает нам совершенно абсурдные оценки Y . Например, луковица диаметром в 36,27 мм оценивается нулевой массой, а все луковицы меньшего диаметра получают оценки массы меньше нуля. На рисунке 13.5 показаны линия, соответствующая данным наблюдений, и эффекты экстраполирования. Из графика четко видно, почему в данном случае экстраполяция ведет нас к столь большому заблуждению. Уравнение линейной регрессии предполагает, что определенная прибавка в диаметре луковиц увеличит их массу на конкретное фиксированное число. Однако практическая невозможность получения подобного результата должна быть очевидной. Один сантиметр, добавленный к луковице диаметром в 9 см,

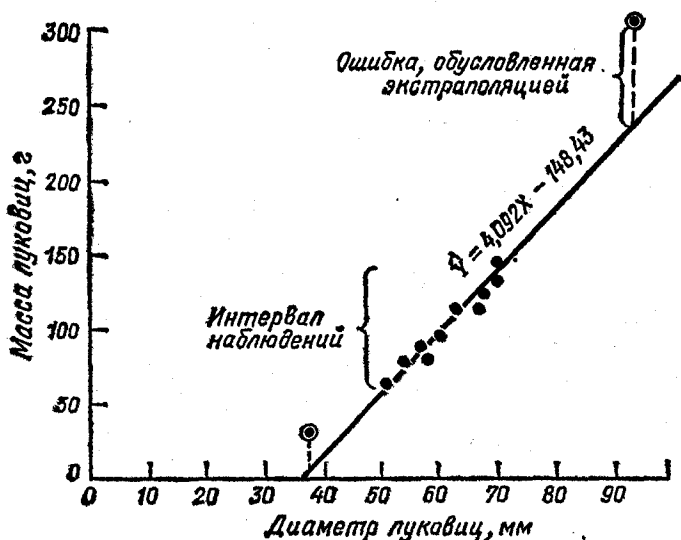


Рис. 13.5. Линия регрессии, проведенная через данные учета луковиц за пределы фактического интервала значений, показывает опасность экстраполяции на основе ограниченного числа наблюдений.

приведет к значительно большей прибавке массы, чем 1 см, добавленный к луковице диаметром 2 см.

Если исследователь желает установить характер связи двух переменных за пределами его наблюдений, то наиболее надежным приемом будет проведение дополнительных наблюдений в интересующей его части явления.

ЗАКЛЮЧЕНИЕ

Корреляция — это тенденция двух переменных быть связанными определенным образом.

Эти две переменные называют независимой и зависимой, исходя из того, какая из них рассматривается зависимой от другой. Независимую переменную обозначают X , а зависимую переменную Y .

Коэффициент корреляции измеряет тесноту (степень, силу) связи. Регрессия — это количественное изменение зависимой переменной, обусловленное изменением независимой переменной на единицу измерения. Уравнение линейной регрессии записывается следующим образом: $\hat{Y} = a + bX$, где $\hat{Y}$ — оценочная величина Y , a — свободный член или отрезок оси Y , отсекаемый линией, и b — наклон линии или коэффициент регрессии.

Нанесение данных, представленных парными значениями переменных, на графике дает точечную диаграмму. Этот удобный прием представляет обычно первый этап регрессионного анализа. Ускоренный сокращенный метод корреляции, известный под названием метода ранговых различий (ранговой корреляции), дает легко вычисляемое приближенное значение коэффициента корреляции.

Формула имеет вид $r = 1 - \frac{6\sum d^2}{n(n-1)(n+1)}$, где r — коэффициент корреляции, d — разность между рангами каждой пары наблюдений и n — число пар.

Стандартный метод или метод сумм произведений может быть записан с помощью формул:

$$r^2 = \frac{[\sum (X - \bar{X})(Y - \bar{Y})]^2}{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2} \text{ (на основе фактических наблюдений);}$$

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} \text{ (на основе отклонений от средних значений);}$$

$$r^2 = \left[\sum XY - \frac{\sum X \sum Y}{n} \right]^2 / \left[\left(\sum X^2 - \frac{(\sum X)^2}{n} \right) \left(\sum Y^2 - \frac{(\sum Y)^2}{n} \right) \right] \text{ (при использовании калькуляторов);}$$

$$r^2 = b_{YX} b_{XY} \text{ (на основе коэффициентов регрессии);}$$

$r = \pm \sqrt{r^2}$ (знак коэффициента корреляции), который будет соответствовать знаку числа, получаемому в скобках числителей уравнений с 1-го по 3-е. Он будет также соответствовать коэффициенту b_{XY} в уравнении 4.

Существенность коэффициента корреляции может быть определена после сопоставления с табличным значением r при $(n-2)$ степенях свободы, где n — число пар наблюдений.

Корреляции, основанные лишь на двух парах наблюдений, будут всегда равны $+1$ или -1 , однако они не имеют смысла.

Коэффициент регрессии получают из формулы $b = \Sigma xy / \Sigma x^2$.

Свободный член $a = \bar{Y} - b\bar{X}$.

Определив значения a и b , мы можем записать уравнение регрессии $\hat{Y} = a + bX$.

Величину неполного соответствия между наблюдаемыми значениями и расчетной оценкой $\hat{Y}$ измеряют суммой квадратов отклонений от регрессии по формуле $\Sigma d^2 = (1 - r^2) \Sigma y^2$.

Сумма квадратов отклонений, деленная на соответствующие им степени свободы ($n-2$), дает средний квадрат отклонений (DMS). Квадратный корень из среднего квадрата отклонений называется стандартной ошибкой оценки.

Сумма квадратов, приходящаяся на регрессию, может быть получена прямо $SSR = r^2 \Sigma y^2$ или путем вычитания $SSR = \Sigma y^2 - \Sigma d^2$. Средний квадрат регрессии будет равен сумме квадратов, поскольку он имеет лишь одну (1) степень свободы. Проверка существенности, основанная на критерии F - (средний квадрат регрессии/средний квадрат отклонений), осуществляется путем сопоставления с табличным F , взятым в колонке 1, и при $n-2$ степенях свободы. Это будет та же проверка, что и проведенная на основе r -величин из таблицы.

Дисперсия коэффициента регрессии равна $s_b^2 = DMS/SSX$. Доверительные пределы для b равны $b \pm t(s_b)$. Дисперсию оценочной величины $\hat{Y}$ получают из формулы $s_{\hat{Y}}^2 = DMS(1/n + x^2/SSX)$. Доверительные пределы для $\hat{Y}$ равны $\hat{Y} \pm t(s_{\hat{Y}})$. Дисперсия расчетной оценки индивидуального наблюдения $\hat{Y}$ равна $s_{\hat{y}}^2 = DMS(1 + 1/n + x^2/SSX)$. Доверительные пределы единичной оценки $\hat{Y}$ равны $\hat{Y} \pm t(s_{\hat{y}})$.

В эксперименте, имеющем повторность, средний квадрат регрессии и средний квадрат отклонения от регрессии можно оценивать с помощью среднего квадрата ошибки, используемого для проверки общего среднего квадрата вариантов.

Необходимо всегда помнить, что обычный коэффициент корреляции предполагает линейную связь между двумя переменными. Он не может помочь нам также в решении вопроса: соответствует ли связь причинно-следственной зависимости?

Низкая величина коэффициента корреляции не всегда означает отсутствие связи; в таком случае мы можем иметь дело с криволинейной зависимостью.

Высокий коэффициент корреляции не подразумевает прямую причинно-следственную связь. Две переменные могут быть просто связаны с третьей переменной, такой, как время.

Избегайте расчета корреляции переменной с одной из ее составных частей. Выводы, основанные на такой корреляции, являются тривиальными.

Избегайте экстраполяции линии регрессии за пределы наблюдаемых значений.

КРИВОЛИНЕЙНЫЕ ЗАВИСИМОСТИ

В предшествующей главе мы неоднократно напоминали о том, что обычные коэффициенты корреляции и регрессии основаны на линейной зависимости между двумя переменными. Линейная зависимость является самым простым типом связи между переменными. Даже в случае заметного отклонения от линейности экстремальных значений X и Y часто случается так, что в пределах важных или практических значений переменных прямая линия оказывается достаточной для характеристики связи. Например, при изучении удобрений мы часто отмечаем наличие устойчивого повышения урожайности по мере увеличения дозы некоторого питательного элемента до определенного уровня. Выше этого уровня прибавка урожая будет менее выраженной, и, наконец, урожайность пойдет на убыль, если мы будем вносить избыточное количество удобрений. Если для нас представляют интерес дозы удобрений в пределах от низкого до среднего уровня, то прямая линия может быть использована для описания связи между урожайностью и удобрениями. Если мы желаем дать описание подобной связи по всем уровням применения удобрений от нуля до необычно большой дозы, то в таком случае мы скорее всего должны использовать кривую линию, которая достигает максимума и затем снижается.

ВЫБОР КРИВОЙ ДЛЯ ОПИСАНИЯ ЗАВИСИМОСТИ

Поскольку в математической статистике имеется так много различных видов кривых, которые можно использовать для выражения связи между двумя переменными, то в начале анализа необходимо решить, какую кривую мы проверим на соответствие нашим данным. Желательно найти кривую, выражающую естественную зависимость между двумя переменными, однако это не всегда возможно. Иногда на основе тщательного изучения и опыта работы с переменными нам удается выбрать кривую, отличающуюся большей логичностью, чем другие. По мере дальнейшего изложения материала мы приведем некоторые примеры подобного характера. Иногда, наоборот, другие кривые оказываются более логичными. Выбор кривой, тесно соответствующий данным, может дать нам важный ключ к выявлению характера естественной связи, существующей между двумя перемен-

ными. Многие из законов естествознания были открыты подобным образом, например закон Бойля — Мариотта и закон падающих тел.

В отношении биологических объектов связь между двумя переменными может быть настолько комплексной, что ни одно из простых уравнений не сможет удовлетворить ее описанию. Часто мы довольствуемся получением уравнения, достаточно хорошо соответствующего данным наблюдений, и не требуем обязательного выражения им какой-либо естественной связи. Всегда можно подобрать кривую, которая будет хорошо соответствовать экспериментальным данным, однако такая кривая может быть сугубо искусственной и полностью лишенной физического или биологического смысла.

Из множества кривых мы выбрали для обсуждения пять видов. Такой выбор основывается на том, что, во-первых, эти кривые являются общепринятыми при анализе биологических и экономических данных, и, во-вторых, для их обсуждения требуются лишь элементарные математические знания.

Степенная кривая. Это — кривая, где Y является функцией некоторой степени X . Общий вид уравнения кривой такого типа записывается следующим образом: $Y = aX^b$.

Прологарифмировав обе части этого уравнения, мы получим

$$\log Y = \log a + b \log X^*.$$

Если взять $\log X$ и $\log Y$ в качестве переменных, обозначив их через X и Y , и постоянную величину $\log a$ — через a' , то уравнение примет следующий вид: $Y' = a' + bX'$.

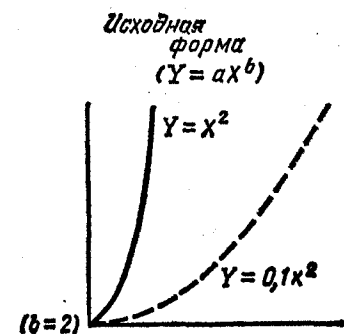
Это выражение хорошо известно как общий вид уравнения для прямой линии, обсуждавшееся в предшествующей главе. Следовательно, все, что требуется для математической обработки данных такого рода, сводится к преобразованию исходных наблюдений в логарифмы и последующему анализу линейной корреляции и регрессии.

Величина b может быть положительной или отрицательной, целым числом или дробью. На рисунке 14.1 представлены некоторые из широкого разнообразия форм кривых, обусловленных различной величиной b . После преобразования X и Y в логарифмы все из этих кривых становятся прямыми линиями с наклоном, равным b , как показано на правой половине рисунка.

Действие a на исходные кривые выражается в сжатии или увеличении масштаба одной из осей координат, тогда как ее действие на преобразованную через логарифмы линию выражается в простом передвижении ее вверх или вниз без изменения ее наклона.

Поскольку только положительные числа имеют логарифмы, то логарифмическая форма уравнений ничего не означает для отрицательных значений X . Таким образом, преобразованию через логарифмы следует подвергать только те данные, в которых X и Y имеют положительные значения. Фактически это не очень серьезное огра-

* Для тех, кто забыл правила применения логарифмов и степеней, необходимо повторить этот материал по любому элементарному учебнику алгебры.



Логарифмическая преобразованная форма
($\log Y = \log a + b \log X$)

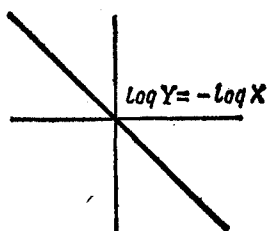
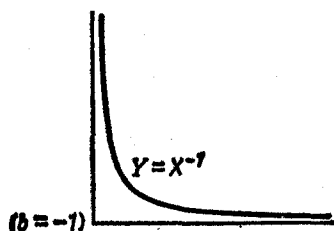
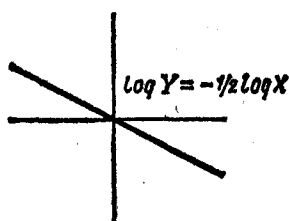
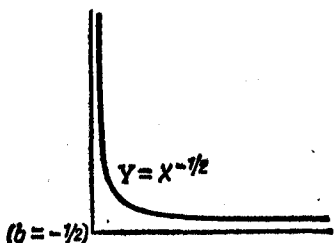
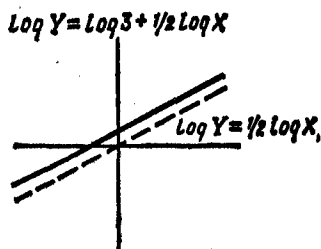
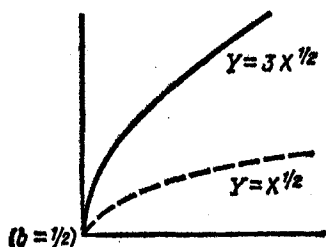
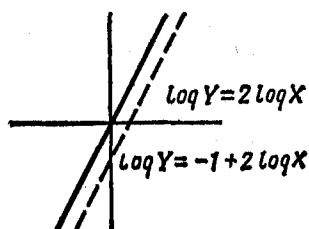


Рис. 14.1. Кривые различной формы и их логарифмические преобразования, показывающие возможность превращения кривых в прямые с помощью логарифмов.

нение, поскольку многие физические измерения, такие, как масса, длина, площадь и т. д., дают только положительные значения.

Как узнать, уместно ли применение логарифмического преобразования? В этом случае использование графика будет снова хорошим началом анализа. График может быть построен двумя способами. Исходные значения X и Y можно преобразовать в логарифмы и от-

ложить на миллиметровке. Еще более простой способ — это отложение исходных величин на логарифмической бумаге (графленой в логарифмах). При любом способе в результате получаем точечную диаграмму. Если эта точечная диаграмма имеет вид продолговатого узкого эллипса, типичного для линейно-коррелируемых данных, то можно продолжить анализ логарифмов X и Y .

Логически мы будем ожидать, что графическому изображению данных измерений, представляющих два вида пространственной мерности, будет соответствовать кривая следующей формы: $Y = aX^b$. Например, рост — одномерная величина, в то время как масса, связанная с объемом, — трехмерная. Поэтому при определении корреляции роста с массой будет логичным использовать логарифмическое преобразование. То же самое относится к данным измерений ширины и площади, длины и объема, поверхности и диаметра и т. п.

В предыдущей главе при обсуждении опасности экстраполяции линии регрессии мы использовали некоторые данные по диаметру и массе луковиц лука. Мы подчеркнули, что прямая линия довольно хорошо описывала связь этих показателей, если рассматривался короткий интервал значений диаметра. Продолжение линии в том или другом направлении за пределы интервала наблюдений не обеспечивало хорошего отражения связи между диаметром и массой. Осмыслив это явление, мы признаем, что фактически его и следовало ожидать. Один сантиметр, добавленный к диаметру большой луковицы, даст большую прибавку массы, чем увеличение диаметра маленькой луковицы на 1 см. Более того, в случае сферической формы луковиц связь между диаметром и объемом выражалась бы следующим уравнением:

$$V = \frac{\pi d^3}{6}.$$

Если удельная масса луковиц всех размеров оставалась бы постоянной величиной, то масса была бы прямой линейной функцией объема. Поэтому мы будем ожидать, что масса (Y) является функцией диаметра (X) в кубе.

В действительности в отношении луковиц дело обстоит совершенно не так просто, поскольку их форма редко бывает сферической, а чаще приближается к сфероидальной и на продольном срезе переходит в эллиптическую. Более того, по мере того как луковицы растут, они постоянно меняют форму: маленькие имеют форму вытянутого сфероида, средние — почти шара, а большие — сжатого сфероида. Это постоянное изменение формы объясняется тем фактом, что их диаметр растет быстрее длины. Несмотря на эти сложности роста, логарифмическое преобразование может значительно упростить данные такого рода.

В таблице 14.1 приводятся величины диаметра и массы 30 обследованных луковиц, расположенные в порядке увеличения величины диаметров.

Во-первых, мы вычислим коэффициенты корреляции и уравнения регрессии для исходных дат.

Таблица 14.1. Диаметр и масса луковиц

Диаметр (X)		Масса (Y)	Log X (X')	Log Y (Y')
	35,1	24,3	1,54531	1,38561
	35,3	24,1	1,54777	1,38202
	35,5	24,4	1,55023	1,38739
	37,8	27,8	1,57749	1,44404
	37,8	28,7	1,57749	1,45788
	41,4	42,0	1,61700	1,62325
	41,7	34,5	1,62014	1,53782
	44,8	56,1	1,65128	1,74896
	44,9	49,0	1,65225	1,69020
	47,9	58,4	1,68034	1,76641
	51,0	63,4	1,70757	1,80209
	53,9	78,7	1,73159	1,89597
	56,9	80,7	1,75511	1,90687
	58,1	80,0	1,76418	1,90309
	60,0	96,2	1,77815	1,98318
	63,0	112,8	1,79934	2,05231
	66,2	115,3	1,82086	2,06183
	67,1	125,6	1,82672	2,09899
	69,2	146,6	1,84011	2,16613
	69,5	132,6	1,84198	2,12254
	70,7	142,8	1,84942	2,15473
	73,1	137,1	1,86392	2,13704
	73,1	163,2	1,86392	2,21272
	77,4	180,0	1,88874	2,25527
	81,7	198,0	1,91222	2,29667
	81,7	207,8	1,91222	2,31765
	82,3	190,8	1,91540	2,28058
	83,1	225,5	1,91960	2,35315
	84,6	237,0	1,92737	2,37475
	92,4	300,2	1,96567	2,47741
Итого	1817,2	3383,6	52,90339	58,27655
Суммы квадра- тов	118 958,58	542 675,26	93,80268806	116,4541216
Суммы произ- ведений	241 772,67		104,0495715	

$$\Sigma x^2 = 118\,958,58 - \frac{(1817,2)^2}{30} = 8884,72;$$

$$\Sigma y^2 = 542\,675,26 - \frac{(3383,6)^2}{30} = 161\,050,29;$$

$$\Sigma xy = 241\,772,67 - \frac{1817,2 \cdot 3383,6}{30} = 36\,816,74;$$

$$r^2 = \frac{(\Sigma xy)^2}{\Sigma x^2 \Sigma y^2} = \frac{(36\,816,74)^2}{8884,72 \cdot 161\,050,29} = 0,9473;$$

$$r = \sqrt{0,9473} = 0,973 \quad b = \frac{\Sigma xy}{\Sigma x^2} = \frac{36\,816,74}{8884,72} = 4,144;$$

$$a = \bar{Y} - b\bar{X} = \frac{3383,6}{30} - \left[4,144 \left(\frac{1817,2}{30} \right) \right] = -138,20;$$

$$\hat{Y} = -138,20 + 4,144X$$

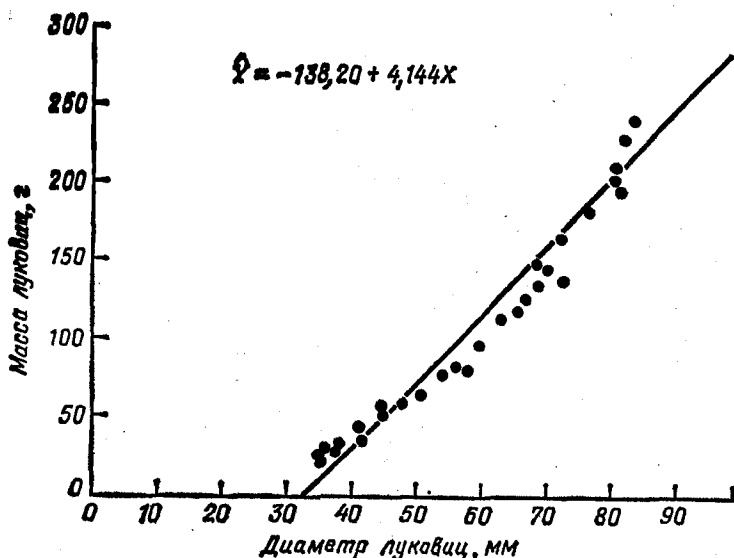


Рис. 14.2. Данные более широкого, чем на рис. 13.5, интервала результатов измерений луковиц, показывающие неслучайные отклонения от линии регрессии.

На первый взгляд кажется, что прямая линия обеспечила нам великолепное соответствие данным измерений. Коэффициент корреляции, равный 0,973, является очень высоким. Однако если мы взглянем на график данных с наложенной на них линией регрессии (рис. 14.2), то отметим одно нарушение соответствия. Все отклонения от линии на концах интервала имеют положительный знак, тогда как откло-

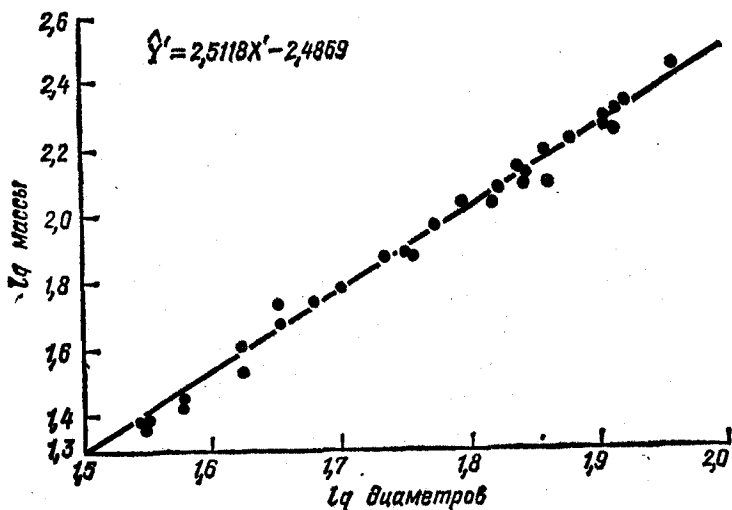


Рис. 14.3. Данные рисунка 14.2, преобразованные через логарифмы, показывают более тесное соответствие прямой линии.

нения, находящиеся в середине интервала,— отрицательный. Если бы отклонения носили случайный характер, мы были бы удовлетворены, однако эта систематическая группировка отклонений показывает нам, что кривая описала бы фактические измерения лучше, чем прямая. В данном случае имеется еще более веская причина необходимости обращения к кривой линии. Прямая линия, рассчитанная для данных измерений, просто не имеет смысла для диаметров менее 34 мм, поскольку она дает для них отрицательные величины массы.

Теперь мы рассчитаем прямую линию для логарифмов X и Y и посмотрим, преодолены ли отмеченные трудности. Порядок вычислений остается прежним, за исключением замены X на $X' = \log X$ и Y на $Y' = \log Y$.

$$\Sigma x'^2 = 93,802\,688\,06 - \frac{(52,90339)^2}{30} = 0,51039894;$$

$$\Sigma y'^2 = 116,4541216 - \frac{(58,27655)^2}{30} = 3,2489123;$$

$$\Sigma x'y' = 104,0495715 - \frac{52,90339(58,27655)}{30} = 1,2820031;$$

$$r^2 = \frac{1,2820031^2}{0,51039894(3,2489123)} = 0,991129;$$

$$r = \sqrt{0,9911} = 0,996;$$

$$b = \frac{1,2820031}{0,51039894} = 2,5118;$$

$$a' = \frac{58,27655}{30} - 2,5118 \left(\frac{52,90339}{30} \right) = -2,4869;$$

$$\hat{Y}' = 2,4869 + 2,5118X'.$$

Коэффициент корреляции, равный 0,996, говорит об экстремально близком соответствии, даже большем, чем при использовании необработанных дат. Усиление корреляции, однако, не является главной причиной предпочтительного использования преобразованных дат в этом случае. Как видно из рисунка 14.3, отклонения точек от линии регрессии распределяются по направлению более или менее случайно. Более того, из связи между X и Y , выражаемой новым уравнением, следует, что по мере приближения диаметра к нулю масса также приближается к нулю.

Уравнение регрессии, выраженное в логарифмах, может быть снова переведено в исходные единицы измерений путем взятия антилогарифма a' (находят величину a) и подстановки a в уравнение $Y = aX^b$.

Логарифмическая форма $Y' = -2,4869 + 2,5118X'$.

Исходная форма $Y = 0,00326 \cdot (X^{2,5118})$.

Показатель степени, приблизительно равный 2,5, интересен тем, что вскрывает закономерности роста луковиц. Если бы луковицы росли одинаково по всем направлениям, то их форма оставалась бы постоянной, и масса была бы функцией диаметра в кубе, или X^3 . Если бы их высота оставалась постоянной и увеличивался только диаметр, то масса должна была бы быть функцией диаметра в квадрате, или X^2 . Если бы высота луковиц увеличивалась с меньшей

скоростью, чем их диаметр, то форма последовательно изменялась бы от вытянутой до сферической и приплюснутой, и масса должна была быть функцией диаметра в степени между 2 и 3. Последнее положение точно соответствует фактическому результату. Полученное нами уравнение не только близко соответствует точечному графику данных измерений, но также выражает естественную связь между диаметром и массой, что соответствует другим фактам геометрии роста.

Экспоненциальная кривая (показательная кривая роста или убыли). В этой кривой X выступает в качестве показателя степени, и коэффициент b выражает скорость роста или убыли. Уравнение такой кривой имеет общий вид $Y = ab^x$.

Если мы прологарифмируем обе части уравнения, то получим

$$\log Y = \log a + (\log b)X,$$

с учетом следующих замен: $\log Y = Y'$, $\log a = a'$ и $\log b = b'$:

$$Y' = a' + b'X.$$

Снова после преобразования имеем уравнение линии, однако в этом виде кривой используются логарифмы Y и исходные величины X вместо логарифмов обеих переменных. По этой причине ее называют полулогарифмическим видом. Имеется специальная бумага с полулогарифмическим графлением — шкалой логарифмов на оси Y и простой шкалой на оси X . Данные можно наносить на полулогарифмическую бумагу или предварительно логарифмировать значения Y и наносить на простую миллиметровку. В обоих случаях при соответствии точечной диаграммы линейно расположенным данным необходимо вычислить коэффициенты линейной корреляции и регрессии Y на X .

Значения X могут быть положительными или отрицательными, дробными или целыми числами, а величина b может быть только целым положительным числом. На рисунке 14.4 приводятся две типичные экспоненциальные кривые, значение величины b в одном случае равно 2, а в другом $1/2$. На рисунке также даны прямые линии, полученные после преобразования Y в $\log Y$.

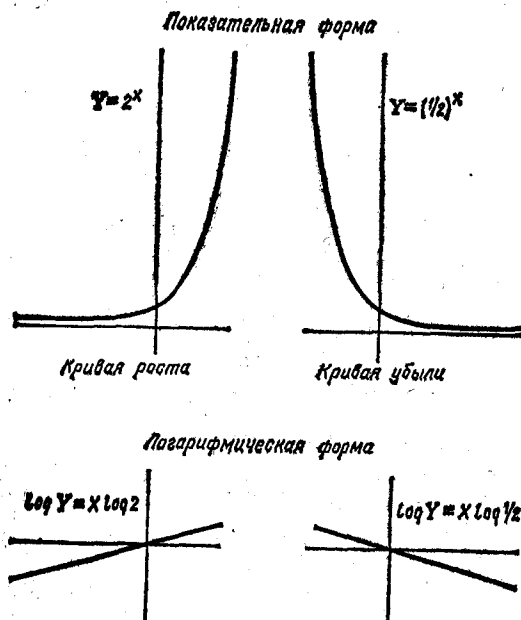


Рис. 14.4. Типичные показательные кривые и их логарифмические преобразования.

С наибольшей вероятностью такому типу кривых будет соответствовать распределение данных, связанных с процентами денежных вложений. Формула изменения капитальной (исходной) суммы, вложенной при постоянной, ежегодно складывающейся норме процента, имеет вид

$$A = P(1+r)^t,$$

где A — сумма вклада по истечении промежутка времени t ; P — исходная сумма вклада; r — ежегодная норма процента и t — время в годах.

Где найти подобную закономерность в сельском хозяйстве? Многие организмы имеют довольно постоянный рост, по крайней мере в течение ранних стадий развития, и поэтому отвечают закономерности складывающейся нормы процента. Если мы изучаем связь между интервалом времени и размером организма или популяции, то целесообразно проверить, соответствуют ли распределения данных такому типу кривой.

Другая ситуация, в которой может оказаться полезным такой тип кривой, связана с физическими законами, экспоненциальными (показательными) по своему характеру. Рассмотрим, к примеру, закон Ван-Гоффа, который гласит о том, что скорость реакций почти удваивается при каждом увеличении температуры на 10°C . Известны многие виды реакций растений, довольно хорошо подчиняющиеся этому закону, по крайней мере в пределах ограниченного интервала температур. Так, температура и скорость порчи фруктов и овощей легко поддаются изучению, если принять экспоненциальный характер их связи.

Скорость охлаждения продуктов, помещенных в холодильник, характеризуется кривой этого вида. В этом случае мы имеем дело не

Таблица 14.2. Рост численности населения Сан-Диего (Калифорния) с 1860 по 1960 г.

Годы переписи	Номер десятилетия с 1860 г. (X)	Численность населения (Y)	Log Y
1860	0	731	2,864
1870	1	2 300	3,362
1880	2	2 638	3,421
1890	3	16 159	4,208
1900	4	17 700	4,248
1910	5	39 578	4,597
1920	6	74 361	4,871
1930	7	147 995	5,170
1940	8	203 341	5,308
1950	9	334 387	5,524
1960	10	573 224	5,758
Итого	55		49,331
Суммы квадратов	385		230,393503
Сумма $X \log Y$			277,981

с ростом, а с уменьшением или убылью. Кривая убыли имеет величину b меньше единицы, тогда как b кривой роста больше единицы. Другими примерами кривых убыли служат кривые разложения определенных пестицидов в почве и распада радиоактивных изотопов.

В качестве примера данных, анализируемых путем преобразования Y в $\log Y$, мы возьмем зависимость численности населения (Y) города Сан-Диего (Калифорния) от времени (X), полученную по итогам 11 переписей, таблица 14.2.

Точечный график зависимости численности населения от времени (рис. 14.5) сразу же указывает на бесполезность расчета уравнения линейной регрессии для этих данных. Этот случай является убедительным примером того, какие крайне заблуждающие результаты может давать сокращенный метод регрессии. Поскольку ранги численности населения имеют те же самые номера, что и ранги лет, то сокращенный метод даст коэффициент корреляции, равный $+1$. Он не обнаружит тот факт, что данные твердо подчиняются криволинейной зависимости. Однако если наносить на график логарифмы численности населения для соответствующих лет, то прямая линия будет обоснованно характеризовать эту связь.

Расчеты принимают прямолинейный характер, если заменить Y на $Y' = \log Y$ в качестве одной из переменных.

$$\Sigma x^2 = 385 - \frac{(55)^2}{11} = 110 = \Sigma X^2 - \frac{(\Sigma X)^2}{n};$$

$$\Sigma y'^2 = 230,393503 - \frac{(49,331)^2}{11} = 9,161907 = \Sigma Y'^2 - \frac{(\Sigma Y')^2}{n};$$

$$\Sigma xy' = 277,982 - \frac{55(49,331)}{11} = 31,326 = \Sigma XY' - \frac{(\Sigma X \Sigma Y')}{n};$$

$$r^2 = \frac{(\Sigma xy')^2}{\Sigma x^2 \Sigma y'^2} = \frac{31,326^2}{110(9,161907)} = 0,9737;$$

$$r = \sqrt{0,9737} = 0,987;$$

$$b' = \frac{\Sigma xy'}{\Sigma x^2} = \frac{31,326}{110} = 0,2848;$$

$$a' = \bar{Y}' - b' \bar{X} = \frac{49,331}{11} - 0,2848 \left(\frac{55}{11} \right) = 3,0606.$$

Уравнение регрессии: $\hat{Y}' = 3,0606 + 0,2848X$. Преобразование через антилогарифмы обеих частей дает экспоненциальное уравнение $\hat{Y} = 1150(1,927)^x$. Это уравнение говорит нам, что через каждые 10 лет численность населения увеличилась в среднем на 92,7%.

Несомненно, экспоненциальная кривая соответствует данным значительно лучше, чем любая возможная прямая линия. Однако даже в случае использования логарифмов численности населения в зависимости от времени и расчета прямой линии ее соответствие будет неидеально, и точки на графике проявляют незначительную, но определенную тенденцию к образованию кривой. Отклонения от центральной части линии принимаются как положительные, а отклонения на ее

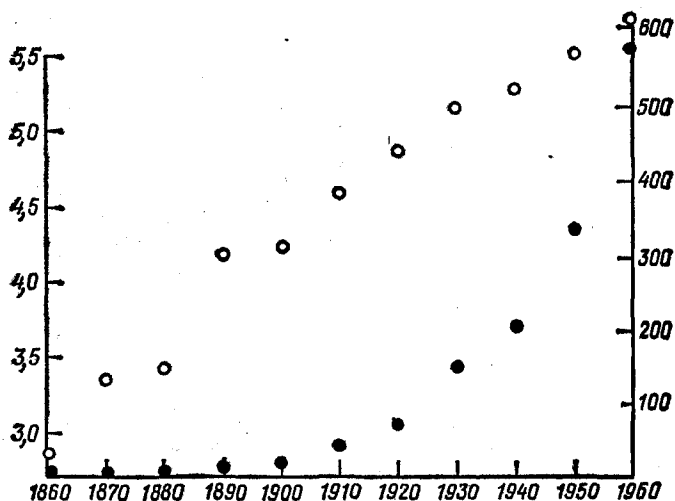


Рис. 14.5. График исходных и полулогарифмически-преобразованных дат численности населения Сан-Диего по результатам 11 переписей по десятилетиям. Ось абсцисс — годы; слева на оси ординат даны $\lg$ численности населения (незаштрихованные кружочки) и справа — численность населения в тысячах (черные кружочки).

концах — как отрицательные. Из графика становится очевидно, что скорость роста не была постоянной, а имела тенденцию к замедлению.

Если линию экстраполировать до 1970 г., то расчетная (ожидаемая) численность населения будет равна 1 561 000. Позже мы покажем, как можно подобрать еще более подходящее уравнение для выражения связи численности населения со временем.

Асимптотические кривые. Это особые разновидности экспоненциальной кривой, упомянутые в последнем абзаце. Если коэффициент b в уравнении $Y=ab^x$ меньше единицы, то Y приближается к нулю по мере бесконечного увеличения X . Линия, к которой таким образом приближается кривая, называется асимптотой. В вышеупомянутом случае асимптотой является ось X .

Имеются случаи, где асимптотой служит некоторая, отличная от нуля, величина Y . Например, температура корзины с продуктами, помещенной в холодильник, достигнет температуры воздуха в холодильнике. Поглощение катиона растениями очень заметно возрастает при небольших добавках его в питательную среду в условиях низкой концентрации раствора. По мере того как концентрация раствора достигает оптимального для развития растений уровня, повышение поглощаемости, обусловленное дополнительным внесением катиона в питательную среду, становится незначительным. Поглощение достигает верхнего предела, который может рассматриваться асимптотой.

Если Y уменьшается с увеличением X и достигает асимптоты сверху, то уравнение вида $Y=c+ab^x$ может обеспечить хорошее соответствие. Если Y увеличивается с увеличением X и достигает асимптоты снизу, то соответствующим уравнением кривой будет $Y=c-ab^x$.

Асимптотой в обоих случаях является $Y=c$. В математике нет простого прямого метода перерасчета данных для этих уравнений. Трудности заключаются в определении величины c . В некоторых случаях эта величина является довольно очевидной, как в случае кривой, характеризующей охлаждение, где асимптотой будет температура охлаждающей среды. В других случаях все, что представляется возможным сделать, сводится к определению ее обоснованной оценки.

В случае ниспадающей кривой мы можем записать уравнение в ином виде $(Y-c)=ab^x$. После логарифмирования получаем линейное уравнение $\log(Y-c)=\log a + X \log b$. Для любого выбранного значения c мы можем определить прямую линию такого вида, соответствующую данным наблюдений. Мы можем проверить различные значения c и, сравнивая величины r^2 , попытаться получить максимальную близость соответствия.

Необходимо подчеркнуть, что c должна быть не больше наименьшего значения Y , поскольку разность $(Y-c)$ должна быть положительной, доступной для логарифмирования.

Аналогично обстоит дело для случая восходящей кривой. Здесь уравнение можно записать в виде $(c-Y)=ab^x$, а в логарифмической форме как $\log(c-Y)=\log a + X \log b$. В этом случае c должно быть больше самого большого наблюдаемого значения Y . Можно составить программы для ЭВМ и легко проверить все последовательные значения c и найти значение, обеспечивающее наименьшую сумму квадратов отклонений от рассчитанной линии.

Этот довольно простой способ приближения может подвергаться критике по причине сходства его названия с методом наименьших квадратов, который сводит сумму $(Y-\hat{Y})^2$ к минимуму. В данном случае речь идет о сведении к минимуму суммы квадратов разностей между наблюдаемыми и расчетными значениями $\log(c-Y)$ или $\log(Y-c)$.

Может случиться, что эти логарифмы будут более однородными в пределах интервала значений X , чем значения исходных Y . Это можно проверить только тогда, когда мы будем иметь несколько значений Y для каждой величины X , как в случае эксперимента с повторностью (см. главу 12). Если дисперсии логарифмов в большей степени приближаются к однородности, чем дисперсии исходных значений Y , то будет разумным ограничиться подбором прямой линии для преобразованных дат $\log(c-Y)$ или $\log(Y-c)$, вместо того, чтобы рассчитывать кривую наименьших квадратов, основанную на непреобразованных значениях Y .

Если желательно определить уравнение, сводящее сумму $(Y-\hat{Y})^2$ к минимуму, то следует обратиться к детальному методу, описанному в 6-м издании книги Снедекора и Кохрана «Статистические методы» (стр. 467—471). Фактически результаты, получаемые на основе подбора прямой линии для преобразованных дат: $\log(c-Y)$ или $\log(Y-c)$, обычно дают уравнения, очень близкие к получаемым на основе более трудоемкого «истинного» метода наименьших квадратов.

Полиномиальный тип кривых. Этот тип кривых имеет общее урав-

нение вида $Y = a + bX + cX^2 + dX^3 \dots$. Ряд точек означает, что мы можем иметь столько членов в уравнении, сколько пожелаем. Если уравнение имеет только два первых члена в правой части, то мы будем иметь уравнение прямой линии. Если оно заканчивается третьим членом (cX^2), то это будет уравнение второй степени (порядка), или квадратное уравнение. Кривая, представляемая квадратным уравнением, называется параболой. График уравнения вида $y = a + bX + cX^2 + \dots + mX^n$ называется параболой степени n . Уравнение, заканчивающееся на dX^3 , называется уравнением третьей степени. Самый высокий показатель степени X , имеющийся в уравнении, определяет его степень, причем для наиболее встречающихся степеней даны специальные названия. В соответствии с первыми пятью степенями различают следующие названия: линейное, квадратное, кубическое, квартичное и квинтичное * соответственно.

Полином — это наиболее широко используемое выражение для описания связи между двумя переменными. Иногда это может быть не совсем «естественное» выражение, т. е. характеризующее причинно-следственную связь между переменными. Однако полином легко поддается расчету, что является очень удачным с математической точки зрения.

На рисунке 14.6 приводятся некоторые из многих кривых, которые могут описываться полиномиальным уравнением. Поразительным свойством этого типа уравнений является то, что независимо от количества пар сделанных наблюдений они позволяют рассчитать полиномиальную кривую, лучше всего соответствующую каждой точке, обеспечивая при этом только одно значение Y для каждой величины X . Степень полинома, требуемая для этого уравнения чаще всего, на единицу меньше числа пар наблюдений. Практически исследователь редко имеет дело с уравнениями третьей, четвертой или большей степени. Расчеты уравнений больших степеней — невероятно трудная задача, а результатом обычно бывает труднообъяснимая извилистая кривая.

Мы отметили, что прямая линия обычно представляет частный случай общего полиномиального уравнения, выраженного уравнением первой степени или линейным полиномом. Для того чтобы найти выражение криволинейной зависимости для двух переменных, мы попытаемся сделать то же самое, что делали при подборе прямой линии. То есть мы найдем кривую заданной степени, которая сделает сумму квадратов отклонений минимальной.

Проблема сводится к нахождению коэффициентов a, b, c, d и т. д., которые дадут полином, отвечающий требованию минимальности суммы квадратов отклонений. Для этого мы воспользуемся известными нормальными уравнениями. Нам требуется столько уравнений, сколько коэффициентов у полинома или на одно больше, чем степень искомого полинома.

* В отечественной литературе уравнения, начиная с четвертой степени, не имеют специального названия. — *Прим. пер.*

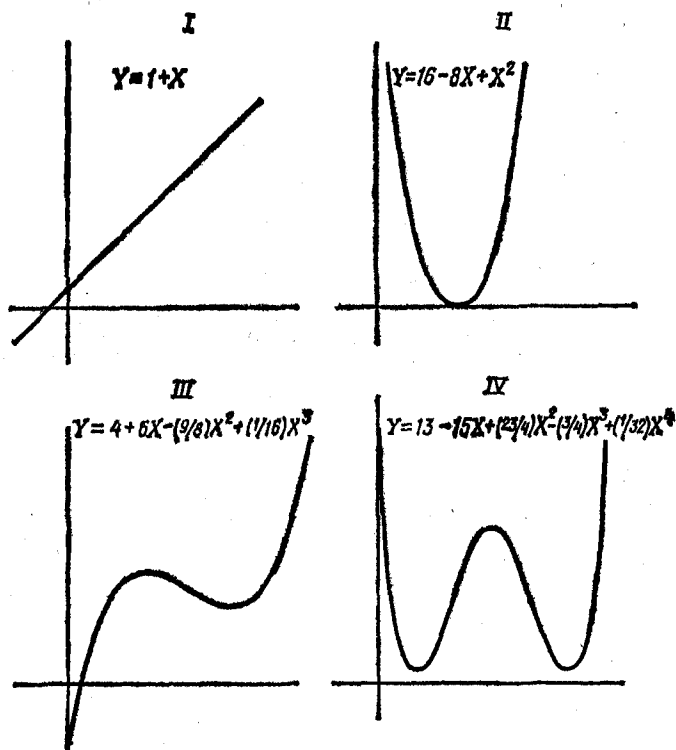


Рис. 14.6. Типичные формы полиномиальных кривых первых четырех степеней (I, II, III, IV).

К нормальным уравнениям относятся следующие уравнения:

$$\begin{aligned} a_n + b\Sigma X + c\Sigma X^2 + d\Sigma X^3 + \dots &= \Sigma Y; \\ a\Sigma X + b\Sigma X^2 + c\Sigma X^3 + d\Sigma X^4 + \dots &= \Sigma XY; \\ a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 + d\Sigma X^5 + \dots &= \Sigma X^2Y; \\ a\Sigma X^3 + b\Sigma X^4 + c\Sigma X^5 + d\Sigma X^6 + \dots &= \Sigma X^3Y. \end{aligned}$$

Точки означают, что ряд продолжается до тех пор, пока с левой стороны от знака равенства не будет столько членов и столько уравнений, сколько требуется вычислить коэффициентов. Так, для прямой линии нужно иметь только два первых члена первых двух уравнений. Для квадратичной, или кривой второго порядка, нам требуются первые три члена первых трех уравнений и т. д.

На основе исходных данных мы должны вычислить суммы X каждой степени и суммы произведений, необходимые для составления уравнений. Для уравнения степени n нам требуется иметь суммы всех степеней от X до X^{2n} и суммы произведений до $X^n Y$. С математической точки зрения это простая задача, однако арифметические расчеты становятся труднопреодолимыми, когда дело доходит до полиномов высоких порядков.

Таблица 14.3. Урожайность зеленых бобов лимской фасоли (Y), в фунтах, по шести срокам сбора (X — число дней, начиная с даты первого сбора)

X	Y	X^2	X^3	X^4	X^5	X^6	XY	X^2Y	X^3Y
0	27,4	0	0	0	0	0	0	0	0
4	39,3	16	64	256	1 024	4 096	157,2	628,8	2 515,2
7	46,2	49	343	2 401	16 807	117 649	323,4	2 263,8	15 846,6
10	47,8	100	1000	10 000	100 000	1 000 000	478,0	4 780,0	47 800,0
13	44,5	169	2197	28 561	371 293	4 826 809	578,5	7 520,5	97 766,5
18	24,5	324	5832	104 976	188 568	34 012 224	441,0	7 938,0	142 884,0
Итого	52 229,7	658	9436	146 194	2 378 692	39 960 778	1978,1	23 131,1	306 812,3

Для примера мы используем некоторые данные по урожайности зеленых бобов лимской фасоли разной стадии развития по полям к периоду сбора (табл. 14.3). Дата самого раннего сбора взята за базовую дату, а величина X принята за нулевое значение. За величины X последующих сроков сбора принято число дней, прошедших от базовой даты. Урожайность в фунтах является зависимой переменной, обозначенной через Y . Ожидается криволинейное распределение данных, поскольку вначале должен быть рост урожайности с появлением новых бобов, однако по мере поспевания окраска бобов переходит от зеленой к бледной и далее к белой. Поэтому урожайность зеленых бобов, достигнув максимума, будет снижаться.

Теперь мы имеем все суммы, необходимые для расчета нормальных уравнений, вплоть до третьего порядка. Вначале для характеристики данных мы рассчитаем прямую линию, используя нормальные уравнения:

$$\begin{aligned} an + b\Sigma X &= \Sigma Y; \\ a\Sigma X + b\Sigma X^2 &= \Sigma XY. \end{aligned}$$

Подставив в эти уравнения известные величины, мы будем иметь:

$$\begin{aligned} 6a + 52b &= 229,7, \\ 52a + 658b &= 1978,1. \end{aligned} \quad (1) \quad (2)$$

Умножив уравнение (1) на 52 и уравнение (2) на 6, мы получим:

$$\begin{aligned} 312a + 2704b &= 19\,944,4 \\ 312a + 3948b &= 11\,868,6 \\ \hline 1244b &= -75,8 \end{aligned} \quad (3) \quad (4)$$

вычтем (3) из (4)

$$b = -75,8/1244 = -0,06093.$$

Подставив эту величину b в уравнение (1), мы получим:

$$\begin{aligned} 6a &= 229,7 + 52(0,0609) = 232,868, \\ a &= 38,8114. \end{aligned}$$

Поэтому уравнение регрессии примет вид

$$\hat{Y} = 38,81 - 0,0609X.$$

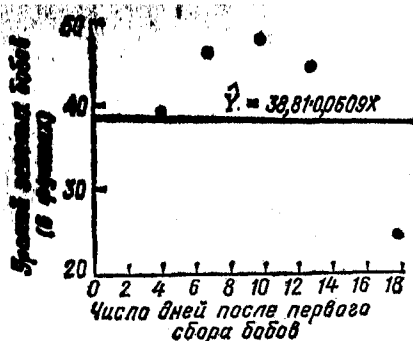


Рис. 14.7. График данных по лимской фасоле, демонстрирующий полную несостоятельность использования линейной регрессии для выражения связи между урожаем и стадией развития бобов ко времени их сбора.

Мы могли бы получить это уравнение с помощью стандартных формул:

$$b = \frac{\sum xy}{\sum x^2} \text{ и } a = \bar{Y} - b\bar{X}.$$

Процедура расчетов с использованием нормальных уравнений была выполнена с целью приобретения вами некоторых навыков для последующего расчета кривых более высоких степеней.

Мы можем видеть на графике (рис. 14.7), что эта линия обеспечивает слабое соответствие данным наблюдений. Для расчета коэффициента корреляции необходимо иметь величину $\sum Y^2$, которая равна в данном случае 9295,03. Тогда

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} =$$

$$= \left[1978,1 - \left(\frac{52(229,7)}{6} \right) \right]^2 / \left(658 - \frac{(52)^2}{6} \right) \left(9295,03 - \frac{(229,7)^2}{6} \right) =$$

$$= (-12,6)^2 / (207,33)(501,35) = 0,00153; r = \sqrt{0,00153} = -0,039.$$

Коэффициент близок к нулю и, очевидно, несуществен. Мы имеем хороший пример одной из ловушек (заблуждений), описанных в главе 13: «Низкий коэффициент корреляции необязательно означает отсутствие связи». Хотя в данном примере коэффициент почти равен нулю, вывод об отсутствии связи между урожаем зеленых бобов и их фазой развития в период сбора был бы нелепым.

Теперь мы подберем уравнение второй степени или квадратичное уравнение для характеристики данных. Нам необходимо иметь три нормальных уравнения:

$$\begin{aligned} an + b\sum X + c\sum X^2 &= \sum Y; \\ a\sum X + b\sum X^2 + c\sum X^3 &= \sum XY; \\ a\sum X^2 + b\sum X^3 + c\sum X^4 &= \sum X^2 Y. \end{aligned}$$

Подставив имеющиеся в таблице величины, мы получим:

$$\begin{aligned} 6a + 52b + 658c &= 229,7 & (1) \\ 52a + 658b + 9436c &= 1978,1 & (2) \\ 658a + 9436b + 146194c &= 23131,1 & (3) \end{aligned}$$

Умножим (1) на 52, (2) на 6 и произведем вычитание (1) из (2):

$$\begin{array}{r} 312a + 2704b + 34216c = 11944,4 \\ 312a + 3948b + 56\,616c = 11868,6 \\ \hline 1244b + 22400c = -75,8. \end{array} \quad (4)$$

Теперь умножим (1) на 658 и (3) на 6 и вычтем:

$$\begin{array}{r} 3948a + 34\,216b + 432\,964c = 151\,142,6 \\ 3948a + 56\,616b + 877\,164c = 138\,786,6 \\ \hline 22\,400b + 444\,200c = -12\,356,0. \end{array} \quad (5)$$

Два предшествующих этапа позволили исключить (элиминировать) a и дали нам два новых уравнения с двумя неизвестными. Теперь умножим (4) на 22400 (5) на 1244 и произведем вычитание:

$$\begin{array}{r} 27\,865\,600b + 501\,760\,000c = -1\,697\,920 \\ 27\,865\,600b + 552\,584\,800c = -15\,370\,864 \\ \hline 50\,824\,800c = -13\,672\,944 \\ c = -0,2690. \end{array}$$

Подставим значение c в (4):

$$\begin{array}{r} 1244b - 6025,6 = -75,6, \\ 1244b = 5949,8, \\ b = 4,7828. \end{array}$$

Подставим значения b и c в (1):

$$\begin{array}{r} 6a + 248,7056 - 177,0020 = 229,7, \\ 6a = 157,9964, \\ a = 26,3327. \end{array}$$

Теперь мы можем записать уравнение второго порядка

$$\hat{Y} = 26,3327 + 4,7828X - 0,2690X^2.$$

Давайте посмотрим, насколько это уравнение лучше линейного уравнения соответствует данным. Мы обозначим линейную оценку (расчетное уравнение) через $\hat{Y}_L$ и квадратичную оценку $\hat{Y}_Q$. В таблице 14.4 эти две оценки сравниваются с исходными величинами.

Результаты могут быть сведены в таблице дисперсионного анализа следующим образом:

Источник вариации	SS	df
Общий	501,35	5
Линейный	0,83	1
Отклонения от линейного	500,52	4
Квадратичный компонент	492,76	1
Отклонения от квадратичного	7,76	3

Таблица 14.4. Физические и расчетные урожаи лимской фасоли

X	Y	$\hat{Y}_L$	$d_L = Y - \hat{Y}_L$	d_L^2	$\hat{Y}_Q$	$d_Q = Y - \hat{Y}_Q$	d_Q^2
0	27,4	38,81	-11,41	130,19	26,33	1,07	1,14
4	39,3	38,57	0,73	0,53	41,16	-1,86	3,46
7	46,2	38,38	7,82	61,15	46,63	-0,43	0,18
10	47,8	38,20	9,60	92,16	47,26	0,54	0,29
13	44,5	38,02	6,48	41,99	43,05	1,45	2,10
18	24,5	37,71	-13,21	174,50	25,27	-0,77	0,59
Итого			0,01	500,52		0,00	7,76

Таким образом, в случае использования прямой линии в ней укладывалось лишь около 0,2% вариабельности Y (0,83/501,35), а в случае квадратичной кривой — (492,76+0,83)/501,35, или 98,5%.

Доля вариабельности Y , приходящаяся на линейный плюс квадратичный компоненты (0,985), обозначается через R^2 и называется «множественным коэффициентом детерминации». Более детально он будет обсуждаться в главе 16.

Если квадратичное уравнение очень хорошо соответствует описанию данных, как в примере с лимской фасолью, то тогда часто бывает полезным сделать расчет тех значений X , которые дадут максимальную (или минимальную) величину Y . Решение этой проблемы сводится к простым вычислениям $X_{\max} = \frac{-b}{2c}$. В нашем примере $X_{\max} = -4,7828/2(-0,2690) = 8,9$, или приблизительно 9 дням после базовой даты. Подстановка этой величины X в квадратичное уравнение дает цифру 47,59, что и будет расчетным максимумом величины Y .

Поскольку только 1,5% вариабельности Y остается неучтенной после использования квадратичного уравнения, то на практике мы обычно на этом и заканчиваем регрессионный анализ. Однако для иллюстрации возможностей метода мы рассчитаем кривую третьего порядка. Воспользуемся следующими нормальными уравнениями:

$$\begin{aligned} an + b\Sigma X + c\Sigma X^2 + d\Sigma X^3 &= \Sigma Y; \\ a\Sigma X + b\Sigma X^2 + c\Sigma X^3 + d\Sigma X^4 &= \Sigma XY; \\ a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 + d\Sigma X^5 &= \Sigma X^2Y; \\ a\Sigma X^3 + b\Sigma X^4 + c\Sigma X^5 + d\Sigma X^6 &= \Sigma X^3Y. \end{aligned}$$

Подставив известные величины, мы получили уравнения, которые дадут значения коэффициентов a , b , c и d :

$$6a + 52b + 658c + 9436d = 229,7; \quad (1)$$

$$52a + 658b + 9436c + 146\,194d = 1978,1; \quad (2)$$

$$658a + 9436b + 146\,194c + 2\,378\,692d = 23\,131,1; \quad (3)$$

$$9436a + 146\,194b + 2\,378\,692c + 39\,960\,778d = 306\,812,3. \quad (4)$$

Вначале мы элиминируем a путем умножения уравнения (2) на 6 минус уравнение (1), умноженное на 52, откуда получим

$$1244b + 22\,400c + 38\,692d = -75,8. \quad (5)$$

Уравнение (3), умноженное на 6, минус уравнение (1), умноженное на 658, дает

$$22\,400b + 44\,200c + 8\,063\,264d = -12\,536,0. \quad (6)$$

Уравнение (4), умноженное на 6, минус уравнение (1), умноженное на 9436, дает

$$386\,392b + 8\,063\,264c + 150\,726\,572d = -326\,575,4. \quad (7)$$

Теперь мы элиминируем b путем следующих этапов: уравнение (6), умноженное на 1244, минус уравнение (5), умноженное на 22 400, дает

$$50\,824\,800c + 1\,373\,279\,616d = -13\,672\,944. \quad (8)$$

Уравнение (7), умноженное на 1244, минус уравнение (5), умноженное на 386 492, дает

$$1\,373\,279\,616c + 38\,127\,789\,500d = -276\,963\,704. \quad (9)$$

Для исключения c мы возьмем уравнение (8), умноженное на 1 373 279 616, минус уравнение (9), умноженное на 50 824 800, разделим обе стороны нового уравнения на 10 000 000 и округлим до десяти значащих цифр. Это даст

$$5\,194\,037\,206d = -38\,232\,948$$

$$d = -0,00736.$$

Подставив полученное значение d в уравнение (8), решим его относительно c :

$$c = -0,07015.$$

Подстановка значений d и c в уравнение (5) даст

$$b = 3,48886.$$

Наконец, подстановка d , c и b в уравнение (1) даст

$$a = 27,31449.$$

Уравнение третьей степени, или кубическое, в этом случае будет равно:

$$\hat{Y}_c = 27,31449 + 348886X - 0,07015X^2 - 0,00736X^3.$$

Вычислив оценочные значения $\hat{Y}_c$, мы получим более близкое соответствие данных кубической кривой, чем квадратичной кривой.

x	y	$\hat{y}_c$	$d = y - \hat{y}_c$	d^2
0	27,4	27,31	0,09	0,1
4	39,3	39,68	-0,38	0,14
7	46,2	45,78	0,42	0,18
10	47,8	47,83	-0,03	0,00
13	44,5	44,64	-0,14	0,02
18	24,5	24,46	-0,04	0,00
Суммы			0,00	0,35

Сумма квадратов отклонений от квадратичной кривой может быть расчленена следующим образом.

Источник вариации	SS	df	MS	$F_{\text{факт}}$	$F_{\text{теор}}$	
					5%	1%
Отклонение от квадратичного	7,76	3				
Кубичный компонент	7,41	1	7,41	42,3	18,51	98,49
Отклонение от кубического	0,35	2	0,175			

Улучшенное соответствие данным, достигнутое с помощью расчета кубичного уравнения, хотя и заметное, но существенно лишь на 5%-ном уровне. Имея так мало степеней свободы, этот итог и неудивителен, тем более, что величина F , требуемая для существенности на 1%-ном уровне, равна 98,49.

Рисунок 14.8 дает изображение квадратичной и кубической кривых, опирающихся на более широкий интервал значений, чем фактический интервал; рисунок приведен с целью выявления различий в их форме. В пределах фактического интервала наблюдений эти две кривые различаются не очень сильно, однако лучшее соответствие кубической кривой очевидно.

Вы, вероятно, заметили, как все более громоздкими становятся вычисления при последовательном переходе от линейной к квадратичной и кубической кривым. Разрабатывались различные методы систематизации этих вычислений; самыми эффективными считаются Дулиттл-метод и сокращенный Дулиттл-метод. Описание этих методов в данной книге не приводится, но может быть найдено в некоторых более сложных книгах по статистике. Для научных целей разработаны

программы расчета коэффициентов кривой почти любой желаемой степени на ЭВМ.

При равновеликих промежутках между отдельными значениями X можно использовать крайне простые сокращенные методы, которые будут рассмотрены в следующей главе.

Полиномы в экспериментах с повторностью. В том случае, когда мы имеем лишь одно значение Y для каждой величины X , единственно возможным путем проверки существенности регрессии будет сопоставление ее среднего квадрата с остаточным средним квадратом. В

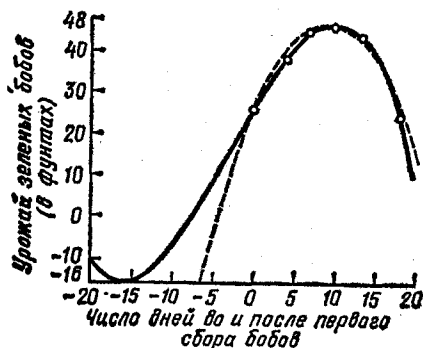


Рис. 14.8. График данных по лимской фасоле рисунка 14.7, демонстрирующий хорошее соответствие их квадратичной кривой (пунктирная линия) и еще более тесное соответствие кубической кривой (сплошная линия).

Таблица 14.5. Подбор квадратного уравнения для сроков уборки сахарной свеклы (расчеты сумм для нормальных уравнений)

x	y	x^2	x^4	x^3y	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$
1	140,0	1	1	140,0	142,1714	-2,1714	4,7150
2	267,2	8	16	1 068,8	258,7142	-8,4858	72,0088
3	335,2	27	81	3 016,8	347,6284	-12,4284	154,4651
3	417,0	64	256	6 672,0	408,9140	-8,0680	65,3834
5	440,6	125	625	11 015,0	442,5710	-1,9710	3,8848
15	1600,0	225	979	21 912,6	1599,9990	0,0010	300,4571

экспериментах с повторностью, с другой стороны, мы имеем средний квадрат ошибки, который можно использовать не только для проверки существенности каждого компонента регрессии, но также для проверки остаточного среднего квадрата.

В предшествующей главе мы рассчитывали прямую линию для урожая сахарной свеклы по пяти срокам уборки, взятых из таблицы 10.1. Хотя величина среднего квадрата линейной регрессии и была установлена очень существенной, средний квадрат отклонения от линейности регрессии был также существен.

Теперь мы вычислим квадратичное уравнение для этих данных и посмотрим, приходится ли большая часть отклонения от прямой линии на кривую второй степени. Некоторые из сумм, необходимые для нормальных уравнений, были уже рассчитаны в таблице 13.7. Другие мы найдем в таблице 14.5.

Итак, мы имеем все суммы, необходимые для записи нормальных уравнений:

$$5a + 15b + 55c = 1600,0; \quad (1)$$

$$15a + 55b + 225c = 5551,0; \quad (2)$$

$$55a + 225b + 979c = 21912,6. \quad (3)$$

Уравнение (1), умноженное на 3 и вычтенное из уравнения (2), и уравнение (1), умноженное на 11 и вычтенное из уравнения (3), дадут нам два новых уравнения с двумя неизвестными:

$$10b + 60c = 751,0; \quad (4)$$

$$60b + 374c = 4312,6. \quad (5)$$

Уравнение (4), умноженное на 6 и вычтенное из уравнения (5), даст

$$14c = -193,4; \quad c = -13,8143.$$

Подстановка значения c в уравнение (4) даст

$$b = 157,9857,$$

а подстановка b и c в уравнение (1) даст

$$a = -2,0000.$$

В итоге получим квадратное уравнение вида

$$\hat{Y} = -2 + 157,9857X - 13,8143X^2.$$

В таблицу 14.5 мы внесли величины $\hat{Y}$, разности между этими (расчетными) и фактическими величинам и квадраты разностей. Сумма отклонений по существу равна нулю, как и должно быть, а сумма квадратов отклонений равна 300,4571; эта величина должна быть приведена к поделяночной, основе, так как мы имеем дело с суммами. Поскольку в каждую сумму сроков уборки вошли урожаи 16 делянок, то частное от деления $300,4571/16=18,7786$ и будет суммой квадратов отклонений от квадратичной кривой. Поскольку сумма квадратов отклонений от линейной регрессии равнялась 185,7587, то сумма квадратов для квадратичной регрессии будет равна

$$185,7587 - 18,7786 = 166,9802.$$

Все это может быть сведено в таблице дисперсионного анализа.

Источник вариации	df	SS	MS	F
Сроки уборки	4	3710,7650	927,691	111,92
Линейная	1	3525,0062	3525,006	425,26
Квадратичная	1	166,9802	166,980	20,14
Остаточная	2	18,7786	9,389	1,13
Ошибка	12	99,4670	8,289	

Мы видим, что на квадратичную регрессию приходится большая часть существенного отклонения от линейной зависимости. Остаточная сумма квадратов незначительна и фактически она не была бы существенной и в случае единственной степени свободы, поэтому отпадает необходимость в дальнейшем продолжении регрессионного анализа.

Мы использовали довольно долгий и трудоемкий путь для определения квадратичного уравнения и сумм квадратов, приходящихся на квадратичную регрессию и отклонения от регрессии. В следующей главе мы познакомимся с сокращенным методом определения квадратичного уравнения. Мы уже имеем некоторые навыки нахождения суммы квадратов для регрессии с использованием коэффициентов таблицы A.11 (приложение). В соответствующей части таблицы под $n=5$ мы находим коэффициенты для квадратичной зависимости: 2, -1, -2, -1 и 2.

$$SS = \frac{(\sum c_i T_i)^2}{r (\sum c_i^2)} = \frac{[(2) 140,0 - 267,2 - (2) 335,2 - 417,0 + (2) 440,6]^2}{16 (14)} = \frac{(-193,4)^2}{224} = 166,9802,$$

что равняется величине, полученной косвенным путем.

Сочетание кривых. Мы обсудили четыре общих типа кривых и показали, каким образом осуществляется их подбор для данных

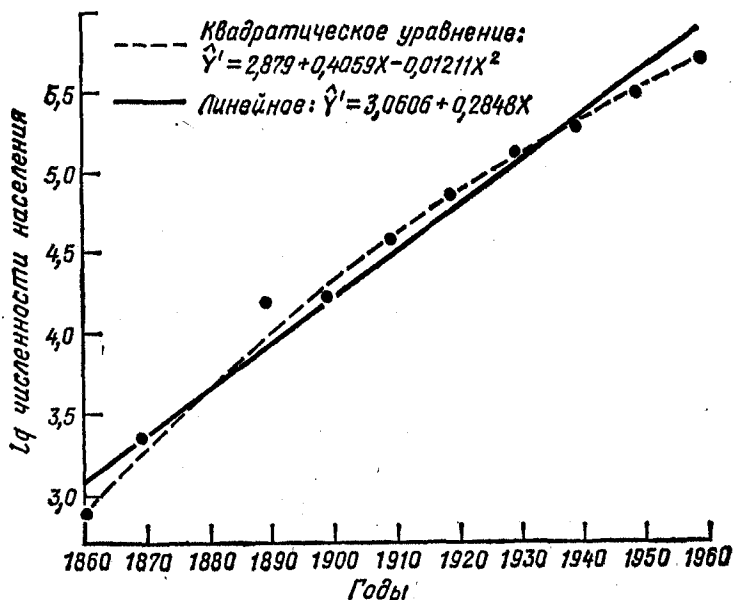


Рис. 14.9. График логарифмов данных численности населения Сан-Диего, выраженный линейным уравнением (сплошная линия), и уточнение этой зависимости с помощью квадратичного уравнения (пунктирная линия).

наблюдений. Иногда бывает полезным использование сочетания двух типов. На примере данных численности населения Сан-Диего мы установили, что нанесение на графике логарифмов численности населения для соответствующих лет обеспечивало более тесное приближение к прямой линии, чем отложение исходных дат. Однако, как видно из рисунка 14.5, даже нанесение преобразованных данных не дает правильной прямой линии, а проявляется довольно четкая тенденция к образованию кривой. Темпы роста населения замедляются со временем.

Мы можем легко рассчитать кривую второй степени для этих данных, используя снова $Y' = \log Y$ в качестве зависимой переменной вместо Y . Расчеты возлагаются на заинтересованного читателя, что явится для него хорошим упражнением по подбору кривой второй степени. В итоге должно быть получено уравнение следующего вида:

$$Y' = 2,87906 + 0,40590X - 0,01211X^2.$$

На рисунке 14.9 приводится уравнение прямой линии и кривой второй степени для логарифмов численности населения. Мы уже указывали, что экстраполяция прямой линии дала прогноз численности населения на 1970 г., равный 1 561 000. Экстраполяция квадратичной кривой дает прогноз в 756 800 *. Исходя из более тесного

* Численность населения Сан-Диего по результатам переписи 1970 г. составила 697 000 человек, что на 8% ниже расчетной цифры.

соответствия кривой второй степени намечаемым трендам вторая цифра прогноза кажется более обоснованной.

Периодическая кривая. Эта кривая связывает некоторую переменную со временем и повторяется через фиксированные промежутки времени. В книгах по математике она известна под названием кривой Фурье и хорошо описывает любой вид данных, проявляющих тенденцию к размеренным колебаниям. Лишь немногие пособия по статистике рассматривают проблему подбора кривой к таким данным, однако мы посчитали ее настолько нужной для многих видов сельскохозяйственных данных, что представляем краткое описание общего метода решения этой проблемы. В следующей главе мы рассмотрим сокращенный метод, используемый в особых случаях.

Общее уравнение периодической кривой имеет вид

$$Y = a_0 + a_1 \cos CX + b_1 \sin CX + a_2 \cos 2CX + b_2 \sin 2CX + a_3 \cos 3CX + b_3 \sin 3CX \dots,$$

где X — фактическое количество единиц времени к данному моменту, начиная от произвольно выбранного стартового времени, а C — константная величина, равная 360° , деленным на количество единиц времени в цикле.

Например, мы изучаем часовые колебания некоторой переменной в 24-часовых циклах и берем полночь за стартовое время. Наблюдение, сделанное в 9 часов утра, будет иметь X , равный 9, а C будет равно $360^\circ/24$, или 15° . Отсюда величина $CX = 9 \times 15$, или 135° .

Многоточие с правой стороны общего уравнения означает, что ряд парных выражений (членов) может быть продолжен до тех пор, пока общее число выражений не превысит число периодов времени, для которых имеются наблюдения.

Эта кривая имеет много общего с полиномиальной кривой. Она имеет то же примечательное свойство, согласно которому при соответствии каждого значения X только одному значению Y может быть рассчитано уравнение, обеспечивающее прохождение кривой через каждую точку наблюдений.

Вы помните, что полиномом первой степени является прямая линия с уравнением $Y = a + bX$. Эта линия полностью характеризуется двумя величинами — отрезком, отсекаемым на оси ординат, — a и наклоном — b . Кривая Фурье первого порядка — это простая волнообразная кривая с уравнением $Y = a_0 + a_1 \cos CX + b_1 \sin CX$. Для построения этой линии необходимо иметь три числа. Величина a_0 дает центральное значение, вокруг которого колеблется волна. Ее можно рассматривать в качестве взвешенной средней. Вторая характеристика $A = \sqrt{a_1^2 + b_1^2}$ называется полуамплитудой и дает величину отклонений (выше и ниже) кривой от центральной точки. Общий размах волны от самой высокой точки до самой низкой точки равен $2A$ и называется амплитудой. Третьей величиной, необходимой для описания волны, является фаза (угол). Она дает нам ту точку в цикле, где волна достигает своего максимума. При ее определении вначале находят величину θ' (тета) $= \arctan(b_1/a_1)$, которая читается как «угол,

тангенс которого равен b_1/a_1 . Далее мы находим фазу, применяя следующие правила.

Если b_1 и a_1 имеют положительные значения, то $\theta = \theta'$.

Если b_1 положительно, а a_1 отрицательно, то $\theta = 180^\circ - \theta'$.

Если b_1 и a_1 имеют отрицательные значения, то $\theta = 180^\circ + \theta'$.

Если b_1 отрицательно, а a_1 положительно, то $\theta = 360^\circ - \theta'$.

Рассматривая полиномы, мы получили более сложные кривые путем добавления выражений с последовательно возрастающей степенью X , таких, как cX^2 , dX^3 и т. д. В случае кривой Фурье мы получаем более сложную форму волны путем добавления пар выражений, таких, как $a_2 \cos 2CX + b_2 \sin 2CX$, $a_3 \cos 3CX + b_3 \sin 3CX$ и т. д. Эффект пары второго порядка заключается в наложении на первую волну второй волны с двумя полными колебаниями за цикл. Пара третьего порядка накладывает следующую кривую с тремя полными колебаниями за цикл и т. д.

Метод расчета (подбора) кривой Фурье также во многом подобен методу подбора полиномиальной кривой. Мы используем группу нормальных уравнений, куда подставляем вычисленные суммы данных наблюдений, и после решения их получаем требуемые коэффициенты.

Для упрощения нормальных уравнений целесообразно ввести два символа — U и V :

$$U_i = \cos i(CX)$$

$V_i = \sin i(CX)$, тогда $\sum U_i V_i$ будет означать

$$\sum \cos 2(CX) \sin(CX).$$

Нормальные уравнения примут следующий вид:

$$a_0 n + a_1 \sum U_i + b_1 \sum V_i + a_2 \sum U_i^2 + b_2 \sum V_i^2 + \dots = \sum Y$$

$$a_0 \sum U_i + a_1 \sum U_i^2 + b_1 \sum U_i V_i + a_2 \sum U_i^3 + b_2 \sum U_i V_i^2 + \dots = \sum U_i Y$$

$$a_0 \sum V_i + a_1 \sum U_i V_i + b_1 \sum V_i^2 + a_2 \sum U_i V_i^2 + b_2 \sum V_i^3 + \dots = \sum V_i Y$$

$$a_0 \sum U_i^2 + a_1 \sum U_i^3 + b_1 \sum U_i^2 V_i + a_2 \sum U_i^4 + b_2 \sum U_i^2 V_i^2 + \dots = \sum U_i^2 Y$$

$$a_0 \sum V_i^2 + a_1 \sum U_i V_i^2 + b_1 \sum V_i^3 + a_2 \sum U_i V_i^3 + b_2 \sum V_i^4 + \dots = \sum V_i^2 Y.$$

Как и в случае полиномов, нам нужно иметь столько членов с левой стороны этих уравнений и столько самих уравнений, сколько потребуется вычислить коэффициентов. Для расчета полинома n -го порядка нам нужно составить $n+1$ уравнений, каждое из них с $n+1$ членами в левой стороне уравнения. Для кривой Фурье потребуется $2n+1$ уравнений с $2n+1$ членами в каждом.

Для иллюстрации мы рассчитаем кривую Фурье первого порядка для средних температур воздуха, зарегистрированных за девять месяцев в городе Стоктоне (Калифорния).

В таблице 14.6 имеются фактические результаты наблюдений и необходимые колонки для расчета выражений нормальных уравнений.

Теперь мы можем записать три из нормальных уравнений, необходимых для определения a_0 , a_1 и b_1 :

$$9a_0 + 2,366a_1 - 1,366b_1 = 518,9; \quad (1)$$

$$2,366a_0 + 4a_1 + 0,866b_1 = 77,894; \quad (2)$$

$$-1,366a_0 + 0,866a_1 + 5b_1 = 103,202. \quad (3)$$

Таблица 14.6. Среднемесячные температуры воздуха (по шкале Фаренгейта) за девять месяцев (Стоктон, Калифорния)

(Цикл=12 месяцев, $C=360^\circ/12=30^\circ$)										
У (темпе- рату- ра)	Месяц	X	CX	$U_1=\cos$ (CX)	$V_1=\sin$ (CX)	U_1^2	V_1^2	U_1V_1	YU_1	YV_1
44,7	Январь	0	0°	1,000	0,000	1,00	0,00	0,000	44,700	0,000
49,0	Февраль	1	30°	0,866	0,500	0,75	0,25	0,433	42,434	24,500
53,7	Март	2	60°	0,500	0,866	0,25	0,75	0,433	26,850	46,504
59,7	Апрель	3	90°	0,000	1,000	0,00	1,00	0,000	0,000	59,700
76,2	Август	7	210°	-0,866	-0,500	0,75	0,25	0,433	-65,989	-38,100
72,7	Сентябрь	8	240°	-0,500	-0,866	0,25	0,75	0,433	-36,350	-62,958
64,0	Октябрь	9	270°	0,000	-1,000	0,00	1,00	0,000	0,000	-64,000
53,0	Ноябрь	10	300°	0,500	-0,866	0,25	0,75	-0,433	26,500	-45,898
45,9	Декабрь	11	330°	0,866	-0,500	0,75	0,25	-0,433	39,749	-22,950
Суммы				2,366	-1,366	4,00	5,00	0,866	77,894	-103,202

Умножим уравнение (1) на 0,866, а уравнение (2) на 1,366 и сложим их, откуда получим

$$11,026a_0 + 7,513a_1 = 555,771. \quad (4)$$

После умножения уравнения (1) на 5 и уравнения (3) на 1,366 и последующего их сложения получим

$$43,134a_0 + 13,013a_1 = 2453,526. \quad (5)$$

После умножения уравнения (4) на 13,013 и уравнения (5) на 7,513 и вычитания получим

$$-180,584a_0 = -11\,201,093 \text{ и } a_0 = 62,027.$$

Подстановка этого значения a_0 в уравнение (4) дает:

$$(11,026 \times 62,027) + 7,513a_1 = 555,771;$$

$$7,513a_1 = -128,139;$$

$$a_1 = -17,056.$$

Подстановка a_0 и a_1 в уравнение (3) дает:

$$(-1,366 \times 62,027) + [0,866(-17,057)] + 5b_1 = -103,202;$$

$$-84,729 - 14,770 + 5b_1 = -103,202;$$

$$5b_1 = 84,729 + 14,770 - 103,202 = -3,703;$$

$$b_1 = -0,741.$$

Теперь мы можем записать искомое уравнение в окончательном виде

$$Y = 62,027 - 17,056 \cos(CX) - 0,741 \sin(CX).$$

Таблица 14.7. Фактические и расчетные температуры воздуха (по шкале Фаренгейта) девяти месяцев (Стоктон, Калифорния)

Месяцы	Y (фактические)	$\hat{Y}$ (расчетные)	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
Январь	44,7	44,97	-0,27	0,0729
Февраль	49,0	46,89	2,11	4,4521
Март	53,7	52,86	0,84	0,7056
Апрель	59,7	61,29	-1,59	2,5281
(Май)	(66,2)	(69,91)	(-3,71)	
(Июнь)	(72,8)	(76,43)	(-3,63)	
(Июль)	(78,2)	(79,08)	(-0,88)	
Август	76,2	77,17	-0,97	0,9409
Сентябрь	72,7	71,20	1,50	2,2500
Октябрь	64,0	62,77	1,23	1,5129
Ноябрь	53,0	54,14	-1,14	1,2996
Декабрь	45,9	47,63	-1,73	1,9929
Суммы	518,9	518,92	-0,02	16,7550

Подстановка значений $\cos(CX)$ и $\sin(CX)$ по каждому месяцу в уравнение даст нам прогнозируемые (расчетные) величины, которые можно будет сравнить с фактическими данными.

Цифры в скобках в таблице 14.7 представляют данные, которыми мы не располагали к моменту вычисления кривой, и поэтому они не вошли в расчеты. Следует подчеркнуть, что кривая, рассчитанная на основе имеющихся данных, превышала фактические средние, не вошедшие в расчеты по месяцам.

Кривая очень тесно соответствует данным наблюдений. Общую сумму квадратов фактических температур, равную 1032,942, можно расчлениить на составляющие части в таблице дисперсионного анализа:

Источники вариации	df	SS	MS	F
Общая	8	1032,942		
Регрессия	2	1016,187	508,094	181,85 ***
Отклонения от регрессии	6	16,755	2,794	

Регрессия имеет две степени свободы, поскольку мы вычисляли два параметра a_1 и b_1 в дополнение к средней. Сумму квадратов для регрессии получают путем вычитания суммы квадратов отклонений из общей суммы. Доля общей суммы квадратов, приходящаяся на регрессию, равна $1016,187/1032,942 = 0,98838$ и обозначается через R .

Факт равенства параметра a величине 62,027 заслуживает внимания. Ранее этот параметр упоминался в качестве взвешенной средней. Эта величина является оценкой того значения средней, какое она приняла бы при наличии данных за весь год. Она и в самом деле очень близка к истинной средней за год, равной 61,34 и основанной

на полных данных. Очевидно, средняя фактических наблюдений равная $518,9/9=57,656$, будет очень слабой оценкой среднегодовой величины, поскольку все отсутствующие данные приходились на теплые месяцы. Однако величина a , получаемая при расчете кривой Фурье, позволяет нам получить близкую оценку, несмотря на отсутствующие даты.

Параметры a и b могут быть использованы для нахождения полуамплитуды и фазы (угла).

$$\text{Полуамплитуда} = A = \sqrt{a_1^2 + b_1^2} = \sqrt{(-17,056)^2 + (-0,741)^2} = 17,1;$$

$\theta = \text{tg}^{-1} b_1/a_1$ характеризует угол, тангенс которого равен $-0,741/17,056 = 2,5^\circ$.

Согласно правилам знаков, $\theta = 180^\circ + \theta' = 182,5^\circ$.

Поскольку 1 месяц $= 30^\circ$, то $182,5^\circ$ будет эквивалентно 6,1 месяца. Это свидетельствует о том, что кривая достигает максимума через 6,1 месяца после исходной даты. Среднюю января мы приняли за нашу стартовую дату, так что назовем эту дату 15 января. Поэтому наш вычисленный максимум придется на число, отстоящее от 15 января на 6,1 месяца, что соответствует примерно 18 июля.

Мы прошли через все этапы подбора для данных простой кривой Фурье первого порядка. Если появляется необходимость подобного подбора для данных кривой 2-го или большего порядка, то вычисления становятся крайне трудными, поскольку надо использовать в расчетах два дополнительных уравнения на каждую добавляемую степень. Такие проблемы легко решаются с помощью ЭВМ. На рисунке 14.10 дается кривая связи сроков посева сельдерея с количеством дней от посева до уборки, которая была рассчитана и нанесена

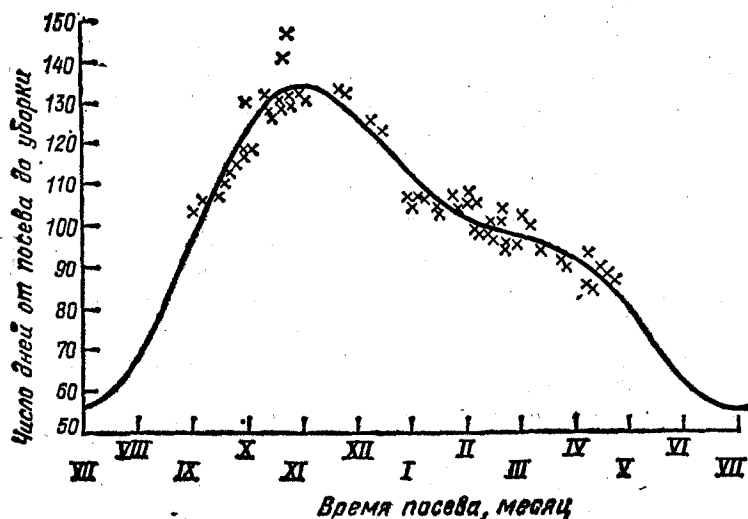


Рис. 14.10. График связи сроков посева сельдерея с числом дней до уборки (округ Вентура, Калифорния).

на график с помощью ЭВМ. Для расчета этой кривой были использованы десятилетние данные.

К счастью, если мы имеем наблюдения, взятые через равные промежутки внутри полного цикла, то вычисления значительно упрощаются. В следующей главе мы опишем сокращенный метод обработки данных подобного рода.

ЗАКЛЮЧЕНИЕ

Если точечная диаграмма связи двух переменных указывает на тенденцию точек быть рассеянными больше вокруг кривой, чем вокруг прямой линии, то желательнее исследовать криволинейную связь между переменными. В противном случае анализ приведет к большим заблуждениям.

Если логарифмы значений двух переменных образуют точечную диаграмму, соответствующую прямой линии, то кривая, характеризующая связь переменных, будет иметь форму $Y=aX^b$ и будет называться степенной кривой. Для описания связи переменных, включающих различные пространственные измерения, этот тип кривой будет наиболее подходящим.

Для анализа таких данных следует преобразовать исходные переменные X и Y в новые переменные $X'=\log X$ и $Y'=\log Y$. Далее надо продолжить анализ аналогично линейной корреляции и регрессии и определить уравнение регрессии для прямой линии $Y'=a'+b'X'$.

Если график логарифмов Y и исходных значений X образует прямолинейную точечную диаграмму, то соответствующая кривая связи переменных будет иметь форму $Y=ab^x$ и будет называться экспоненциальной (показательной) кривой. Можно ожидать, что данным, переменная Y которых стремится иметь довольно постоянные темпы возрастания или снижения, будет отвечать этот тип кривой.

Для анализа следует преобразовать лишь Y в $Y'=\log Y$ и продолжить расчеты аналогично линейной регрессии до получения уравнения $Y'=a'+b'X$.

Особая разновидность экспоненциальной кривой, в которой Y приближается к некоторому значению, отличному от нуля, называется асимптотической кривой. Она описывается уравнением: $Y=c+ab^x$, где c — асимптота. Это уравнение может быть преобразовано в уравнение прямой линии путем преобразования Y в $Y'=\log(Y-c)$ или $Y'=\log(c-Y)$. Однако наиболее подходящее значение c надо находить методом проб и ошибок.

Для криволинейно распределенных данных, которые не достигают прямолинейности как с помощью логарифмического, так и полулогарифмического преобразования, может быть подобрана полиномиальная форма кривой

$$Y=a+bX+cX^2+dX^3+\dots,$$

используя при этом столько членов, сколько потребуется для получения удовлетворительного соответствия данным.

Для определения неизвестных коэффициентов a, b, c, d и т. д. необходимо решить систему уравнений, известных под названием нормальных уравнений:

$$\begin{aligned} an + b\Sigma X + c\Sigma X^2 + d\Sigma X^3 + \dots &= \Sigma Y, \\ a\Sigma X + b\Sigma X^2 + c\Sigma X^3 + d\Sigma X^4 + \dots &= \Sigma XY, \\ a\Sigma X^2 + b\Sigma X^3 + c\Sigma X^4 + d\Sigma X^5 + \dots &= \Sigma X^2Y, \\ a\Sigma X^3 + b\Sigma X^4 + c\Sigma X^5 + d\Sigma X^6 + \dots &= \Sigma X^3Y. \end{aligned}$$

Число уравнений и число членов левой стороны от знака равенства должно быть равно числу требуемых коэффициентов или на одно больше, чем степень уравнения регрессии.

Уравнения первых пяти степеней имеют специальные названия, как и некоторые из кривых.

Степень	Название уравнения	Название кривой
Первая	Линейное	Прямая линия
Вторая	Квадратичное	Парабола
Третья	Кубичное	Кубическая парабола
Четвертая	Четвертой степени	Парабола четвертой степени
Пятая	Пятой степени	Парабола пятой степени

Если отклонения наблюдений от рассчитанной кривой носят более или менее случайный характер, то обычно подбор кривой большей степени теряет смысл. Если отклонения носят систематический характер или образуют определенные группы по отношению к знаку, то расчет уравнения следующего порядка, как правило, оказывается полезным.

В экспериментах, имеющих повторность, оценка существенности среднего квадрата отклонений от регрессии может быть проведена с помощью среднего квадрата ошибки.

Расчеты коэффициентов для уравнений более высоких степеней, чем кубические, следует проводить лишь после того, как овладеете специальными методами (такими, как Дулиттл-метод) или используйте ЭВМ. В случае равновеликих интервалов между значениями X большую экономию времени дают сокращенные методы, описанные в главе 15. Сочетание логарифмического и полиномиального методов приводит иногда к значительно лучшему соответствию данным, чем тот или другой метод в отдельности.

Данным, варьирующим (вверх и вниз) во времени (внутри довольно равных периодов), может соответствовать периодическая кривая (кривая Фурье) следующей формы:

$$Y = a_0 + a_1 \cos CX + b_1 \sin CX + a_2 \cos 2CX + b_2 \sin 2CX + \dots$$

Нормальными уравнениями для нахождения неизвестных коэффициентов будут уравнения:

$$\begin{aligned} a_0 n + a_1 \Sigma U_1 + b_1 \Sigma V_1 + a_2 \Sigma U_2 + b_2 \Sigma V_2 + \dots &= \Sigma Y; \\ a_0 \Sigma U_1 + a_1 \Sigma U_1^2 + b_1 \Sigma U_1 V_1 + a_2 \Sigma U_1 U_2 + b_2 \Sigma U_1 V_2 + \dots &= \Sigma U_1 Y; \\ a_0 \Sigma V_1 + a_1 \Sigma U_1 V_1 + b_1 \Sigma V_1^2 + a_2 \Sigma U_2 V_1 + b_2 \Sigma V_1 V_2 + \dots &= \Sigma V_1 Y; \\ a_0 \Sigma U_2 + a_1 \Sigma U_1 U_2 + b_1 \Sigma U_2 V_1 + a_2 \Sigma U_2^2 + b_2 \Sigma U_2 V_2 + \dots &= \Sigma U_2 Y; \\ a_0 \Sigma V_2 + a_1 \Sigma U_1 V_2 + b_1 \Sigma V_1 V_2 + a_2 \Sigma U_2 V_2 + b_2 \Sigma V_2^2 + \dots &= \Sigma V_2 Y, \end{aligned}$$

где $U_i = \cos i (CX)$ и $V_i = \sin i (CX)$.

Если наблюдения получают через одинаковые промежутки времени внутри полного цикла, то можно использовать сокращенные методы, рассмотренные в главе 15.

СОКРАЩЕННЫЕ МЕТОДЫ РЕГРЕССИИ

Часто случается так, что мы проводим наблюдения за изменением зависимой переменной Y через равновеликие интервалы величины независимой переменной X . Например, если независимой переменной является время, и мы делаем отсчеты значений Y через день, неделю, месяц или год, то между отдельными X -ми или сроками будут равновеликие интервалы. Другим примером, когда мы часто имеем равновеликие интервалы между значениями X , будут эксперименты с дозами фунгицидов, инсектицидов, удобрений и т. п. Эксперимент с равновеликим шагом градаций вариантов предпочтителен с точки зрения математической обработки.

Помимо легкости расчетов, использование равновеликого шага градаций дает и другие преимущества. Если мы пожелаем узнать что-либо о тренде отклика на уровни воздействующего варианта, то лучшим источником необходимой информации будет эксперимент, равномерно распределенный по всему размаху уровней варианта. Например, вариант с серией уровней 0, 1, 2, 4 имеет очень слабое обоснование, хотя и широко используется в опытном деле. Такие серии не подчиняются ни арифметической, ни геометрической закономерности. Информация, получаемая в нижней (левой) части этой серии, отличается большей полнотой, чем информация верхней части. Предположим, что мы установили повышение урожайности с ростом уровня X от 0 до 2 и заметное снижение урожайности на уровне 4. Было бы полезно узнать, где между уровнями 2 и 4 произошло изменение направления тренда на обратное. В этом случае уровень варианта, равный 3, окажет наибольшую помощь.

Сокращенный метод, с которым мы собираемся вас познакомить, обсуждался в разделе сравнения трендов главы 6. Этот метод заслуживает, чтобы было продолжено обсуждение его с точки зрения связи с криволинейной регрессией, рассмотренной в предшествующей главе. Статистики обычно называют его методом ортогональных полиномов. Те из вас, кто трудно воспринимает это внушительное название, могут называть его «сокращенным методом определения трендов». Вы найдете технику расчетов по этому методу легкой и обеспечивающей огромную экономию времени.

ПОДБОР ПОЛИНОМИАЛЬНОЙ КРИВОЙ

Стержнем метода подбора полиномов является таблица А.11 (приложение) *, использование которой избавляет от выполнения многих трудоемких расчетов, обычно необходимых при анализе криволинейной регрессии. Эта таблица может быть использована для: 1) нахождения линейного, квадратичного, кубического и четвертой степени уравнений регрессии для любого числа наблюдений равновеликого шага до 25 и 2) расчленения суммы квадратов вариантов при дисперсионном анализе на линейный, квадратичный, кубический, четвертой степени и остаточный компоненты в опыте с числом вариантов или наблюдений равновеликого шага, не превышающем 25.

В «шапке» таблицы даны значения n , числа наблюдений или вариантов. Для любого данного эксперимента нам требуется использовать лишь часть таблицы под соответствующим значением n . Первая колонка коэффициентов, озаглавленная c_1 , в дополнение к использованию для различных расчетов дает кодированные величины X . Кодирование выполнено таким образом, чтобы получить по возможности самые наименьшие целые числа. Независимо от величины X равновеликого шага, если n нечетное число, мы можем взять за код $X' = (X - \bar{X})/L$, где L — интервал (шаг) между последовательными величинами X . Если n четное число, то мы берем $X' = (X - \bar{X})2/L$. Эти преобразования дают величины для заполнения колонки c_1 .

Для того чтобы использовать другие коэффициенты таблицы, не требуется знать, каким образом они получены. Однако для любознательного студента представляют интерес следующие равенства.

Коэффициенты в колонке c_2 могут быть получены из следующего равенства: $c_{2i} = (c_{1i}^2 n - \sum c_{1i})^2 / GCD$. После того, как будут рассчитаны числители для всех значений i от 1 до n , следует определить наибольший общий знаменатель (GCD), который позволит по возможности сократить коэффициенты до группы наименьших целых чисел. Коэффициенты колонки c_3 находят из следующего равенства: $c_{3i} = (c_{1i}^4 \sum c_{1i}^4 - c_{1i} \sum c_{1i}^4) / GCD$ и c_4 из колонки по равенству

$$c_{4i} = (c_{1i}^4 n \sum c_{1i}^4 c_{2i} - c_{1i}^2 n \sum c_{1i}^4 c_{2i} - \sum c_{1i}^4 \sum c_{1i}^4 c_{2i} + \sum c_{1i}^4 \sum c_{1i}^4 c_{2i}) / GCD.$$

Как видно из формул, расчеты становятся громоздкими, особенно для больших значений n , так что таблица обеспечивает большую экономию времени.

Принцип расчета величин K легко объясняется с помощью некоторых положений теории чисел, что выходит за рамки материала этой книги.

При определении линейного, квадратичного, кубического и четвертичного уравнений регрессии необходимо провести обработку по этапам:

* Эту таблицу, подготовленную авторами, рекомендуется использовать в большей мере, чем любую из многих подобных таблиц, опубликованных в других пособиях. Насколько известно авторам, величины K не встречаются ни в одной из других опубликованных таблиц.

1) последовательно разместить величины Y в колонке по мере возрастания соответствующих величин X , начав с Y , равного самому наименьшему значению X ;

2) умножить величины Y на коэффициенты для c_1, c_2, c_3 и c_4 , приведенные в таблице, что даст четыре колонки;

3) найти сумму по каждой колонке, учитывая знаки плюс и минус. Эти суммы обозначают через $\Sigma Y, P_1, P_2, P_3$ и P_4 ;

4) на основании полученных величин P и табличных величин K можно записать линейное, квадратичное, кубичное и четвертичное уравнения.

$$\text{Линейное: } \hat{Y}_L = \bar{Y} + (K_2 P_1) X'.$$

$$\text{Квадратичное: } \hat{Y}_Q = (\bar{Y} - K_1 P_2) + (K_2 P_1) X' + (K_4 P_2) X'^2.$$

$$\text{Кубичное: } \hat{Y}_C = (\bar{Y} - K_1 P_2) + (K_2 P_1 - K_3 P_3) X' + (K_4 P_2) X'^2 + (K_5 P_3) X'^3.$$

$$\text{Четвертой степени: } \hat{Y}_4 = (\bar{Y} - K_1 P_2 + K_3 P_3) + (K_2 P_1 - K_3 P_3) X' + (K_4 P_2 - K_5 P_3) X'^2 + (K_5 P_3) X'^3 + (K_6 P_4) X'^4.$$

Отметьте, что эти уравнения записаны с помощью кодированных величин X ;

5) если величины Y на этапе 1 представляли суммы нескольких наблюдений или повторений на каждом уровне X и мы хотим записать уравнения с помощью средних, то следует разделить каждое выражение в уравнениях на число повторений. (Повторность должна быть одинаковой для всех уровней X .)

В таблице 15.1 приводятся валовые удои молока за сутки (в фунтах) от 37 коров; эти данные регистрировались ежемесячно в течение 10 месяцев с момента отела до конца лактации. Мы проведем обработку этих данных через пять указанных этапов.

Коэффициенты c_1, c_2, c_3 и c_4 были взяты из таблицы А.11 (приложение) и умножены на соответствующие величины Y (надой молока). Суммы по этим колонкам дадут величины $\Sigma Y, P_1, P_2, P_3$ и P_4 . Теперь мы готовы перейти к этапу 4 и записать уравнения:

$$\hat{Y}_L = 1959,48 + (1/330)(-22\,266,6) X' = 1959,48 - 67,475 X';$$

$$\hat{Y}_Q = [1959,48 - (1/32)(-1048,8)] - 67,475 X' + (1/1056)(-1048,8) X'^2 = 1992,26 - 67,475 X' - 0,9932 X'^2;$$

$$\hat{Y}_C = 1992,26 + [-67,475 - (293/205\,920)(4798,2)] X' - 0,9932 X'^2 + (1/46184)(4798,2) X'^3 = 1992,26 - 74,302 X' - 0,9932 X'^2 + 0,11651 X'^3;$$

$$\hat{Y}_4 = [1992,26 - (9/1280)(-5384,6) - 74,302 X' + [-0,9932 - (41/54\,912)(-5384,6)] X'^2 + 0,11651 X'^3 + (1/109\,824)(-5384,6) X'^4 = 1954,40 - 74,302 X' + 3,0272 X'^2 + 0,11651 X'^3 - 0,049029 X'^4.$$

Эти уравнения основаны на валовой продуктивности 37 коров. Если мы хотим рассчитать их, чтобы получить величину средней

Таблица 15.1. Надои молока от 37 коров в течение 10 месяцев

Надои (Y)	Месяц (X)	X' (c_1)	$c_1 Y$	c_2	$c_2 Y$
2442,3	1	—9	—21 980,7	6	14 653,8
2517,6	2	—7	—17 623,2	2	5 035,2
2334,4	3	—5	—11 672,0	—1	—2 334,4
2166,1	4	—3	—6 498,3	—3	—6 498,3
2030,0	5	—1	—2 030,0	—4	—8 120,0
1903,9	6	1	1 903,9	—4	—7 615,6
1779,5	7	3	5 338,5	—3	—5 338,5
1630,6	8	5	8 153,0	—1	—1 630,6
1485,7	9	7	10 399,9	2	2 971,4
1304,7	10	9	11 742,3	6	7 828,2
Суммы 19 594,8			$P_1 = -22 266,6$		$P_2 = 1048,8$

σ_3	$c_3 Y$	σ_4	$c_4 Y$
-42	-102 576,6	18	43 961,4
14	35 246,4	-22	-55 387,2
35	81 704,0	-17	-39 684,5
31	67 149,1	3	6 498,3
12	24 360,0	18	36 540,0
-12	-22 846,8	18	34 270,2
-31	-55 164,5	3	5 338,5
-35	-57 071,0	-17	-27 720,2
-14	-20 799,8	-22	-32 685,4
42	54 797,4	18	23 484,6
$P_3 = 4798,2$		$P_4 = -5384,6$	

продуктивности одной коровы, то просто поделим каждое выражение на 37 и получим:

$$\hat{Y}_1 = 52,959 - 1,8236X';$$

$$\hat{Y}_Q = 53,845 - 1,8236X' - 0,02684X'^2;$$

$$\hat{Y}_C = 53,845 - 2,0082X' - 0,02684X'^2 + 0,003149X'^3;$$

$$\hat{Y}_4 = 52,822 - 2,0082X' + 0,08182X'^2 + 0,003149X'^3 - 0,0013251X'^4.$$

Практически не возникает необходимости построения таблицы подобной таблице 15.1, поскольку требуемые P -величины могут быть получены путем накопления произведений в памяти счетных машин без записи каждого отдельного произведения. Особое внимание следует обращать на знаки коэффициентов. Там, где коэффициент отрицательный, произведение его и соответствующей величины Y нужно вычитать из накопленной суммы.

Очень важно помнить о том, что вычисленные нами уравнения основываются на выражениях X' , т. е. кодированных величинах X которые идентичны коэффициентам c_1 . Предположим, что мы хотим рассчитать заранее с помощью квадратичного уравнения среднюю продуктивность на корову за третий месяц. Из таблицы 15.1 мы знаем, что X' за третий месяц равно -5 , поэтому мы заменяем X на -5 в квадратичном уравнении:

$$\begin{aligned}\hat{Y}_Q &= 53,845 - 1,8236(-5) - 0,02684(-5)^2 = \\ &= 53,845 + 9,118 - 0,671 = 62,292.\end{aligned}$$

Типичной ошибкой студентов является то, что они подставляют c_1 -коэффициенты в линейное уравнение, c_2 -коэффициенты — в квадратичное уравнение и т. д., тогда как только c_1 -коэффициенты представляют кодированные величины X в каждом уравнении независимо от его степени.

Обычно с такой формой уравнений легче всего работать, однако в случае публикации результатов в научной статье они должны быть записаны с помощью исходных величин X . Для этого следует заменить в уравнениях X' на $(X - \bar{X})/L$ или $(X - \bar{X})2/L$ в зависимости от того, является ли n нечетным или четным числом. Для того чтобы показать, каким образом это делается, мы запишем наше квадратичное уравнение в таком виде: $\hat{Y}_Q = 53,845 - 1,8236X' - 0,02684X'^2$, используя выражения X .

В данном случае $n=10$ было четным, так что мы заменяем X' на $(X - \bar{X})2/L$. Интервал между последовательными величинами X равнялся 1, поэтому $L=1$. Величина $\bar{X}$ равнялась 5,5, поэтому мы будем иметь $X' = (X - 5,5)2/1$ или $2X - 11$. Подстановка этого выражения в наше уравнение дает

$$\begin{aligned}\hat{Y}_Q &= 53,845 - 1,8236(2X - 11) - 0,02684(2X - 11)^2 = \\ &= 53,845 - 1,8236(2X - 11) - 0,02684(4X^2 - 44X + 121) = \\ &= 53,845 - 3,6472X + 20,0596 - 0,10736X^2 + 1,18096X - 3,23764.\end{aligned}$$

Таблица 15.2. Фактические и расчетные месячные надои молока от 37 коров

$Y_{\text{факт}}$	$\hat{Y}_L$	$Y - \hat{Y}_L$	$\hat{Y}_Q$	$Y - \hat{Y}_Q$	$\hat{Y}_C$	$Y - \hat{Y}_C$	$\hat{Y}_4$	$Y - \hat{Y}_4$
2442,3	2566,8	-124,5	2519,1	-76,8	2495,6	-53,3	2461,7	-19,4
2517,6	2431,8	85,8	2415,9	101,7	2423,7	93,9	2465,2	52,4
2334,4	2296,9	37,5	2304,8	29,6	2324,4	10,0	2356,4	-22,0
2166,1	2161,9	4,2	2185,7	-19,6	2203,1	-37,0	2197,4	-31,3
2030,0	2027,0	3,0	2058,7	-28,7	2065,5	-35,5	2031,6	-1,6
1903,9	1892,0	11,9	1923,8	-19,9	1917,1	-13,2	1883,2	20,7
1779,5	1757,1	22,4	1780,9	-1,4	1763,6	15,9	1757,9	21,6
1630,6	1622,1	8,5	1630,1	0,5	1610,5	20,1	1642,5	-11,9
1485,7	1487,2	-1,5	1471,3	14,4	1463,4	22,3	1504,9	-19,2
1304,7	1352,2	-47,5	1304,5	0,2	1328,0	-23,3	1294,1	10,6
Σ отклонения		-0,2		0,0		-0,1		-0,1
Σ (отклонения) ²		27 268,0		18 930,76		16 258,59		6105,23
Σ (отклонения) ² /37		737,0		511,64		439,42		165,01

Упростив выражение, мы получим

$$\hat{Y}_Q = 70,65696 - 2,46624X - 0,10736X^2.$$

Давайте воспользуемся этим уравнением для расчета $\hat{Y}_Q$ за третий месяц. Замена X на 3 в этом новом уравнении даст $\hat{Y} = 70,65696 - 2,46624(3) - 0,10736(3)^2 = 62,292$, как и прежде.

Давайте посмотрим, насколько это сокращает расчеты. Используя методы главы 14 (которые нужно использовать в случае неравновеликих интервалов между значениями X -ов) для определения четырех уравнений регрессии, потребуется найти: ΣX , ΣX^2 , $-\Sigma X^3$, ΣY , ΣXY , ΣX^2Y , ΣX^3Y и ΣX^4Y . Значения этих величин необходимо подставить в нормальные уравнения и затем решить группу систем уравнений, из которых два — для линейных коэффициентов и до пяти для уравнений четвертой степени. Если вы решили примеры главы 14, то могли оценить, насколько трудоемка эта работа. Сопоставьте все эти расчеты с анализом по сокращенному методу.

При использовании последнего нам требуются только ΣY , P_1 , P_2 , P_3 и P_4 . Подстановка значений этих величин в стандартные уравнения 4-го этапа сразу дает нам четыре требуемых уравнения регрессии. Нам потребуется вычислить только пять сумм вместо 13 и не надо решать ни одной сложной системы уравнений.

Теперь, имея четыре уравнения, мы можем сравнить рассчитанные на их основе надои с фактической продуктивностью коров за каждый месяц. Лучше работать с данными валового производства молока, чем со средними, поскольку будет меньше общая погрешность округлений.

В таблице 15.2 приводятся расчетные величины, полученные на основе каждого уравнения, и их отклонения от фактических дат. В отношении этой таблицы следует запомнить несколько моментов. Сумма отклонений для каждой из кривых должна давать нуль без

учета небольшой погрешности округления. Это обеспечивает проверку расчетов. Сумма квадратов отклонений от кривой является показателем соответствия: чем меньше эта сумма, тем сильнее кривая соответствует распределению данных. Каждая добавляемая степень уравнения приводит к снижению этой суммы квадратов. Это положение должно всегда подтверждаться, в противном случае ищите ошибку в расчетах.

Остается выяснить: нужно ли добиваться улучшения соответствия? В дальнейшем мы покажем несложный способ оценки существенности, а пока лишь обратите внимание на умеренное снижение суммы квадратов при переходе от линейной к квадратичной кривой, на очень слабое снижение при переходе от квадратичной к кубичной и на очень сильное снижение при переходе от кубичной к четвертой степени кривой. Наконец, отметьте, что знаки отклонений имеют определенное сходство распределения для первых трех степеней, тогда как знаки отклонений от четвертичной кривой распределяются более или менее случайно. Также мы видим, что только кривая четвертой степени показывает увеличение надоев от первого ко второму месяцу. Эта особенность характерна для большинства кривых, отображающих продуктивность коров.

Расчленение суммы квадратов. Нахождение всех рассчитанных величин и их отклонений от фактических дат и последующее определение сумм квадратов этих отклонений явилось трудоемкой процедурой. Второй характерной чертой сокращенного метода анализа данных равновеликого шага является нетрудоемкость расчета этих сумм квадратов. Посмотрев на любую колонку величины n в таблице А.11 (приложение), вы согласитесь, что s -величины действительно составляют ортогональные группы вариантов. Коэффициенты каждой колонки дают в сумме нуль так же, как и произведения соответствующих коэффициентов любых двух колонок. Мы знаем из главы 6, что сумма квадратов, опирающаяся на одну степень свободы, может быть найдена при помощи группы коэффициентов по общей формуле

$$SS = \frac{(\sum c_i T_i)^2}{r \sum c_i^2}.$$

Как показали ранее проведенные расчеты, P_1 будет той же величиной, что и $\sum c_i T_i$, если s -коэффициенты являются линейными. По той же причине $P_2 = \sum c_i^2 T_i$, если используются квадратичные коэффициенты, и т. д. Делители, показанные в таблице А.11 (приложение), являются суммами квадратов коэффициентов. Поэтому сумма квадратов благодаря линейной регрессии представляет просто $P_1^2 / (\text{делитель} \times \text{на число повторностей})$. Подобно этому сумма квадратов квадратичной регрессии равна $P_2^2 / (\text{делитель} \times \text{на число повторностей})$ и т. д. до четвертичного компонента. После расчета сумм квадратов каждого компонента мы можем определить остаточную сумму квадратов путем вычитания компонентных сумм квадратов из общей суммы квадратов. Эта остаточная сумма квадратов является той же самой величиной, что и сумма квадратов отклонений фактических дат от кривой.

Давайте применим этот метод расчленения к данным продуктивности коров. Величина P_1 , которую мы нашли, равнялась —22266,6, поэтому линейная SS равна

$$\frac{-22\,266,6^2}{330 \times 37} = 40\,606,18.$$

Общая сумма квадратов для Y равнялась 41 343,01, поэтому остаточная сумма квадратов равна разности 41 343,01—40 606,18=736,83. Это то же самое (без учета небольшой погрешности округлений), что и сумма квадратов отклонений прямой линии, но полученная значительно более сложным методом по таблице 15.2.

Поскольку величина P_2 равнялась —1048,8, то сумма квадратов квадратичной регрессии равна

$$\frac{-1048,8^2}{132 \times 37} = 225,22.$$

Остаток после вычитания этой величины из 736,83 равен 511,62. Величина, рассчитанная по таблице 15.2, равнялась 511,64. Величина P_3 была равна 4798,2, а сумма квадратов кубической составляет при делении

$$\frac{4798,2^2}{8580 \times 37} = 72,52$$

и дает остаток, равный 439,09 (в таблице 15.2 эта величина составляет 439,42).

Наконец, P_4 составляла —5384,6, отсюда сумма квадратов четвертой степени регрессии будет равна

$$\frac{-5384,6^2}{2860 \times 37} = 273,99$$

и дает остаток, равный 165,10.

Все эти результаты можно свести в таблице дисперсионного анализа (табл. 15.3), где суммы квадратов для величин продуктивности

Таблица 15.3. Дисперсионный анализ данных продуктивности коров

Источник вариации	df	SS	MS	F
Общий	369	76 167,74		
Коровы	36	24 364,56	651,79	18,59 **
Месяцы	9	41 343,01	4 593,67	131,02 **
Линейная	1	40 606,18	40 606,18	1158,19 **
Отклонение от линейной	8	736,83	92,10	2,63 *
Квадратное	1	225,22	225,22	1,42 *
Отклонение от квадратной	7	511,61	73,09	2,08 *
Кубичное	1	72,52	72,52	2,07 несущ.
Отклонение от кубической	6	439,09	73,18	2,09 несущ.
Четвертой степени	1	273,99	273,99	7,81 **
Отклонение от четвертой степени	5	165,10	33,02	0,94 несущ.
Ошибка	324	11 360,17	35,06	

коров и ошибки были получены из учетных данных индивидуальной продуктивности коров.

В опыте установлены очень существенные различия в надоях молока между коровами и по месяцам. Ни один из этих результатов не удивляет, однако мы хотим знать больше о закономерности изменения надоев по месяцам. Очень высокая величина F для линейного компонента говорит нам об очень существенном тренде снижения надоев. Существенное отклонение от линейности указывает на то, что прямая линия не отображает полностью вариацию по месяцам. Существенность квадратичного компонента свидетельствует о том, что простая кривая лучше соответствует данным, чем прямая линия, однако остаточная вариация все еще существенна. Кубичная кривая не дает существенного улучшения соответствия, а остаток вариации несуществен. На этом многие исследователи склонны закончить анализ. Часто, как в данном случае, это решение будет ошибочным. Регрессия четвертой степени составила настолько высокую долю оставшейся суммы квадратов, что оказалась очень существенной величиной. Отклонение от кривой четвертой степени несущественно. Вероятность нахождения еще одного существенного компонента очень мала, поскольку даже если бы на найденный компонент приходилось 80% оставшейся вариабельности, то все равно он не был бы существен. Наше решение закончить анализ в этом месте оправдано.

Сравнение сокращенного и стандартного методов. В главе 14 мы подбирали квадратичную кривую для данных урожайности сахарной свеклы пяти сроков уборки. Для этого мы сначала нашли семь сумм по степеням и произведениям (табл. 14.5). Затем из этих сумм мы получили три системы уравнений, которые предстояло решить для нахождения трех неизвестных. Теперь мы сопоставим стандартный и сокращенный методы.

Используя коэффициенты таблицы A.11 (приложение) под $n=5$, мы вначале найдем:

$$P_1 = (-2)140,0 + (-1)267,2 + (1)417,0 + (2)440,6 = 751;$$

$$P_2 = (2)140,0 + (-1)267,2 + (-2)335,2 + (-1)417,0 + (2)440,6 = -193,4.$$

На основании этих величин и значений K из таблицы A.11 мы можем сразу же записать квадратичное уравнение

$$\hat{Y}_Q = 320 - (1/7)(-193,4) + (1/10)751X' + (1/14)(-193,4)X'^2 = 347,6286 + 75,1X' - 13,8143X'^2.$$

Для перехода к исходным X -единицам мы заменяем X' на $(X-3)$ и X'^2 на $(X-3)^2 = X^2 - 6X + 9$. Это дает

$$\hat{Y} = -2,0 + 157,9857X - 13,8143X^2,$$

т. е. то же самое уравнение, что и полученное более длинным путем с использованием стандартного метода.

Варианты неравновеликого интервала. Мы указали на преимущества вариантов равновеликого интервала, однако если наш экспери-

мент имеет варианты с разными интервалами между ними, то все еще можно найти группу ортогональных коэффициентов для расчета сумм квадратов регрессий. Формулы определения этих коэффициентов становятся значительно более сложными, чем в случае равновеликого шага между вариантами. Здесь также нет простого пути вывода уравнений через непосредственное использование K -величин.

В таблице А.11а (приложение) даны серии ортогональных коэффициентов и делителей для наиболее часто встречающихся градаций вариантов. Они помогают, по крайней мере, определить суммы квадратов регрессий в указанных случаях.

ПОДБОР ПЕРИОДИЧЕСКОЙ КРИВОЙ

В таблице А.12 в приложении приводятся группы ортогональных коэффициентов для периодических данных, когда между наблюдениями имеются равновеликие интервалы по всему циклу. Таблица построена для избранных величин n , наиболее часто встречающихся при работе с дневными, недельными или годовыми циклами.

В отличие от ранее рассмотренных коэффициентов эти не могут быть сокращены до малых целых чисел. По этой причине расчеты P несколько затруднены, однако в других отношениях расчеты уравнений и расчленение сумм квадратов являются даже более легкими, чем для полиномиальных кривых, поскольку специальные делители и K -величины не требуются. Причина значительно более простых расчетов в случае равновеликих интервалов между наблюдениями по сравнению с неравномерно распределенными датами кроется в том, что большинство выражений нормальных уравнений, данных в главе 14, не используют. В соответствии с этим $\sum U_i = \sum V_i = 0$, где i — любой индекс. Также $\sum U_i^2 = \sum V_i^2 = n/2$. Поэтому первое нормальное уравнение

$$na_0 + a_1 \sum U_i + b_1 \sum V_i + a_2 \sum U_i^2 + b_2 \sum V_i^2 + \dots = \sum Y$$

можно сократить до $na_0 = \sum Y$ или $a_0 = \sum Y/n = \bar{Y}$.

$$a_1 \left(\frac{n}{2} \right) = \sum U_i Y \quad \text{или} \quad a_1 = \frac{2 \sum U_i Y}{n};$$

$$b_1 = \frac{2 \sum V_i Y}{n};$$

$$a_2 = \frac{2 \sum U_i^2 Y}{n};$$

$$b_2 = \frac{2 \sum V_i^2 Y}{n}$$

и т. д., следуя аналогичной системе, за исключением случая, где n — четное число и когда последний коэффициент рассчитывают по другой формуле:

$$a_{(n/2)} = \frac{\sum U_{(n/2)} Y}{n}$$

(Нам редко придется выполнять анализ до полного использования остаточной суммы квадратов. Другими словами, решение уравнения

Таблица 15.4. Среднемесячные температуры воздуха (по шкале Фаренгейта) в Стоконе (Калифорния) с расчетами подбора периодической кривой второй степени ($C = 1/12 \times 360^\circ = 30^\circ$)

Месяц (X)	Температура (Y)	$\cos CX (U_1)$	$U_1 Y$	$\sin 3X (V_1)$	$V_1 Y$	$\cos 2CX (U_2)$	$U_2 Y$	$\sin 2CX (V_2)$	$V_2 Y$
0	44,7	1,0	44,7000	0,0	0,0000	1,0	44,7000	0,0	0,0000
1	49,0	0,866	42,4340	0,5	24,5000	0,5	24,5000	0,866	42,4340
2	53,7	0,5	26,8500	0,866	46,5042	-0,5	-26,8500	0,866	46,5042
3	59,7	0,0	0,0000	1,0	59,7000	-1,0	-59,7000	0,0	0,0000
4	66,2	-0,5	-33,1000	0,866	57,3292	-0,5	-33,1000	-0,866	-57,3292
5	72,8	-0,866	-63,0448	0,5	36,4000	0,5	36,4000	-0,866	-63,0448
6	78,2	-1,000	-78,2000	0,0	0,0000	1,0	78,2000	0,0	0,0000
7	76,2	-0,866	-65,9892	-0,5	-38,1000	0,5	38,1000	0,866	65,9892
8	72,7	-0,5	-36,3500	-0,866	-62,9582	-0,5	-36,3500	0,866	62,9582
9	64,0	0,0	0,0000	-1,0	-64,0000	-1,0	-64,0000	0,0	0,0000
10	53,0	0,5	26,5000	-0,866	-45,8980	-0,5	-26,5000	-0,866	-45,8980
11	45,9	0,866	39,7494	-0,5	-22,9500	0,5	22,9500	-0,866	-39,7494

Суммы 736,1
 $a_0 = 61,34$

$$PU_1 = -96,4506$$

$$a_1 = -16,0751$$

$$\hat{Y} = 61,34 - 16,0751 \cos CX - 1,5788 \sin CX - 0,275 \cos 2CX + 1,9774 \sin 2CX$$

$$PV_1 = -9,4728$$

$$b_1 = -1,5788$$

$$PU_2 = -1,6500$$

$$a_2 = -0,2750$$

$$PV_2 = 11,8642$$

$$b_2 = 1,9774$$

Таблица 15.5. Фактические и расчетные среднемесячные температуры воздуха (Стоктон, Калифорния)

Месяцы	Y фактические	$\hat{Y}_1$ 1-я степень	$(Y - \hat{Y}_1)$	$\hat{Y}$ 2-я степень	$(Y - \hat{Y}_2)$
Январь	44,7	45,26	-0,56	44,99	-0,29
Февраль	49,0	46,63	2,37	48,20	0,80
Март	53,7	51,94	1,76	53,79	-0,09
Апрель	59,7	59,76	0,06	60,04	-0,34
Май	66,2	68,01	-1,81	66,44	-0,24
Июнь	72,8	74,47	-1,67	72,62	0,18
Июль	78,2	77,42	0,78	77,14	1,06
Август	76,2	76,0	0,15	77,63	-1,43
Сентябрь	72,7	70,74	1,96	72,59	0,11
Октябрь	64,0	62,92	1,08	63,19	0,81
Ноябрь	53,0	54,67	-1,67	53,09	-0,09
Декабрь	45,9	48,21	-2,31	46,36	-0,46
Сумма отклонений			0,02		0,02
Σd^2			28,86		4,99

до полного использования остаточной суммы квадратов будет давать кривую, проходящую через все точки наблюдений, что аналогично подбору прямой линии для двух точек.)

Мы примем символы, аналогичные используемым при подборе полиномов, обозначив $\Sigma U_i Y$ через PU_i и $\Sigma V_i Y$ через PV_i . Отметим, что в случае полиномов мы имели единственную величину P на каждую степень, а при подборе периодической кривой нам требуются две величины P , обозначаемые PU и PV , на каждую степень * подбираемой кривой.

Общими выражениями (членами) уравнений являются:

$$a_0 = \bar{Y}; \quad a_i = \frac{2PU_i}{n}; \quad b_i = \frac{2PV_i}{n}.$$

Давайте применим этот метод для подбора периодической кривой к полным данным среднемесячных температур воздуха в Стоктоне (Калифорния), что показано в таблице 15.4.

В уравнение, рассчитанное нами и являющееся основным, можно подставить любое значение X и выбрать соответствующие синусы и косинусы в тригонометрической таблице. Однако, если нас интересует только расчет величин, соответствующих точкам фактических дат, мы можем просто заменить в уравнении U_1 на $\cos CX$, V_1 на $\sin CX$, U_2 на $\cos 2CX$ и V_2 на $\sin 2CX$. Например, чтобы найти $\hat{Y}$ для марта (месяц номер 2, поскольку январь был принят за ноль), мы проводим

* Мы назвали каждую пару выражений, добавляемых к общему уравнению периодической регрессии, степенью с целью сохранения аналогии с общим видом полиномов. В математике их называют гармониками.

следующие расчеты:

$$\hat{Y}_1 = 61,34 - 16,0751(0,5) - 1,5788(0,866) - 0,275(-0,5) + 1,9774(0,866) = 53,79.$$

Если надо получить расчетную величину только для кривой первой степени, то мы просто используем три первых члена вышеприведенного уравнения

$$\hat{Y}_1 = 61,34 - 16,0751(0,5) - 1,5788(0,866) = 51,94.$$

Расчетные величины для уравнений первой и второй степени показаны в таблице 15.5 наряду с отклонениями фактических величин от этих двух кривых.

Расчленение суммы квадратов. Как и для полиномов, в этом случае имеется очень простой способ расчленения общей суммы квадратов без построения таблицы, подобной таблице 15.5. Сумма квадратов для регрессии первой степени равна $2(PU_1^2 + PV_1^2)/n$, для второй степени $2(PU_2^2 + PV_2^2)/n$ и т. д. В отличие от полиномов здесь не требуются различные делители для каждой степени. Суммы квадратов отклонений от фактических дат могут быть получены по разности. Мы нашли, что $PU_1 = -96,4506$ и $PV_1 = -9,4728$ (табл. 15.4). Поэтому сумма квадратов регрессии первой степени будет равна

$$\frac{2[(-96,4506)^2 + (-9,4728)^2]}{12} = 1565,41.$$

Общая сумма квадратов для Y равнялась 1594,33, так что сумма квадратов отклонений будет равна

$$1594,33 - 1565,41 = 28,92,$$

что отличается от величины 28,86, данной в таблице 15,5, вследствие погрешностей округлений.

Аналогично этому сумма квадратов регрессии второй степени равна

$$\frac{2[(-1,65)^2 + (11,8642)^2]}{12} = 23,91.$$

Остаточная или сумма квадратов отклонений от регрессии второй степени равна разности $28,92 - 23,91 = 5,01$ (в сопоставлении с величиной 4,99 таблицы 15.5). Эти результаты сведены в таблице 15.6.

Таблица 15.6. Дисперсионный анализ данных по величинам температур воздуха

Источник вариации	df	SS	MS	F-значения
Месяцы	11	1594,33		
1-я степень	2	1565,41	782,705	243,61 **
Отклонения	9	28,92	3,213	
2-я степень	2	23,91	11,955	16,70 **
Отклонения	7	5,01	0,716	

Отметьте, что кривая каждой степени имеет 2 степени свободы. Это связано с тем, что необходимо вычислить два коэффициента: a и b для каждой степени. Для F -оценки средний квадрат каждой степени сопоставляется со средним квадратом остаточного компонента. В данном случае обе (и первая, и вторая) степени оказались очень существенными.

Мы подобрали (рассчитали) кривую среднемесячных температур и расчленили сумму квадратов для месяцев на несколько компонентов. Если мы пожелаем рассмотреть отдельно учеты за каждый год, на основе которых были рассчитаны средние по месяцам, то дисперсионный анализ значительно усложняется. Подробности этого анализа можно получить из статьи К. И. Блисса, помещенной в бюллетене № 615, опубликованном в 1958 г. Коннектикутской сельскохозяйственной опытной станцией *.

Кривая второй степени, которую мы рассчитали, в действительности состоит из двух простых синусоид, сложенных одна с другой. Первая имеет полуамплитуду $A = \sqrt{a_1^2 + b_1^2}$, поэтому

$$A = \sqrt{(-16,0751)^2 + (-1,5788)^2} = 16,13.$$

Фазой колебаний является $\text{tg}^{-1}(b_1/a_1) + 180^\circ$, то есть угол, тангенс которого равен $0,0982 + 180^\circ = 185^\circ 36'$, что в переводе на время составляет около шести месяцев и пяти дней после начала цикла. Поскольку наш цикл начинается со средней температуры января, то за начало мы можем принять 15 января, поэтому максимум нашей кривой будет падать на 20 июля, а минимум на 6 месяцев раньше, т. е. на 20 января.

Обратив внимание на сплошную кривую нижней половины рисунка 15.1, мы видим, что фактические температуры имеют тенденцию располагаться выше кривой в первой и третьей четверти и ниже кривой — во второй и четвертой четвертях цикла. Эти расхождения в значительной мере соответствуют кривой второго порядка. Ее полуамплитуда равна

$$A = \sqrt{a_2^2 + b_2^2} = \sqrt{(-0,275)^2 + (1,9774)^2} = 2,00,$$

а фаза

$$180^\circ - \text{tg}^{-1}\left(\frac{b_2}{a_2}\right) = 180^\circ - \text{tg}^{-1} 7,1905 = 180^\circ - 85^\circ 5' = 94^\circ 5'.$$

Этот результат надо поделить на 2, поскольку теперь мы имеем дело с двухцикловой кривой, поэтому максимум будет при $48^\circ 32,5'$, или через 1 месяц и 18 дней после 15 января. Другой максимум цикла будет на 6 месяцев позже, а минимума кривая будет достигать спустя 3 месяца после каждого максимума. Графически это изображено в виде пунктирной кривой на нижней половине рисунка 15.1.

* C. I. Bliss. «Periodic regression in biology and climatology», Connect. Agr. Exp. Sta., bul. 615, 1958.

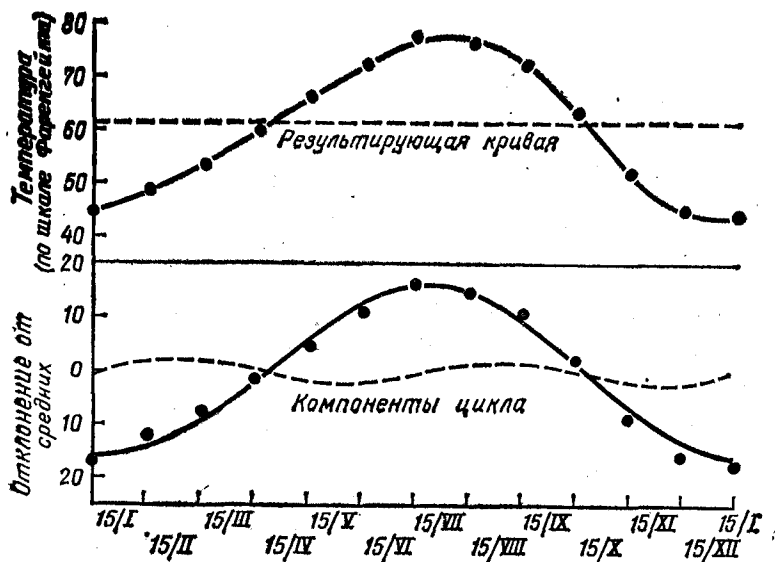


Рис. 15.1. Среднемесячные температуры воздуха в Стоктоне, Калифорния. Кривая Фурье второго порядка и ее составные части (компоненты).

Сложение этих двух кривых со средней температурой (61,34) дает результирующую кривую, представленную на верхней половине рисунка 15.1.

ЗАКЛЮЧЕНИЕ

Для наблюдений или вариантов равновеликого интервала в приложении дана таблица А.11, которая значительно упрощает расчеты регрессии линейного, квадратичного, кубического и уравнения четвертой степени или расчленения сумм квадратов на компоненты трендов. Каждая колонка числа наблюдений от 3 до 25 в этой таблице состоит из трех частей: c -коэффициенты, делители и K -величины.

P -величины получают из уравнения $P_i = \sum c_i T_i$. После расчета P -величин на основе фактических дат линейное, квадратичное, кубическое и четвертой степени уравнения регрессий получают из следующих равенств:

$$\hat{Y}_L = \bar{Y} + (K_2 P_1) X';$$

$$\hat{Y}_Q = (\bar{Y} - K_1 P_2) + (K_2 P_1) X' + (K_4 P_3) X'^2;$$

$$\hat{Y}_C = (\bar{Y} - K_1 P_2) + (K_2 P_1 - K_3 P_3) X' + (K_4 P_2) X'^2 + (K_5 P_3) X'^3;$$

$$\hat{Y}_4 = (\bar{Y} - K_1 P_2 + K_6 P_4) + (K_2 P_1 - K_3 P_3) X' + (K_4 P_2 - K_7 P_4) X'^2 + (K_5 P_3) X'^3 + (K_6 P_4) X'^4.$$

Величины X' в уравнениях регрессии являются кодированными величинами X , равными c_1 -коэффициентам. Уравнения в форме X -выражений можно получить путем замены X' на $(X - \bar{X})/L$, когда n

Нечетное, или $(X - \bar{X})^2/L$, если n — четное число. L — интервал между последовательными значениями величины X .

После расчленения сумма квадратов для вариантов может быть: линейная $SS = P_1^2/(\text{делитель} \times \text{повторность})$, квадратичная $SS = P_2^2/(\text{делитель} \times \text{повторность})$, кубичная $SS = P_3^2/(\text{делитель} \times \text{повторность})$ и четвертой степени $SS = P_4^2/(\text{делитель} \times \text{повторность})$.

Остаточная $SS = SS$ вариантов — линейная SS — квадратичная SS — кубичная SS — четвертой степени SS .

В приложении (в таблице А.12) приводятся группы коэффициентов для расчета периодичных кривых распределения данных, имеющих равновеликие интервалы на всем протяжении временного цикла. Эта таблица содержит две группы коэффициентов, обозначенных U и V , для каждой из первых четырех степеней (гармоник), необходимых значений n . Две P -величины вычисляются для каждой степени соответствия с помощью уравнений

$$PU_i = \sum U_i Y \text{ и } PV_i = \sum V_i Y.$$

После установления P -величины уравнение любой желаемой степени вплоть до четвертой можно вывести непосредственно из следующего равенства:

$$\begin{aligned} \hat{Y} = \bar{Y} + \left(\frac{2PU_1}{n} \right) \cos CX + \left(\frac{2PV_1}{n} \right) \sin CX + \dots \\ \dots + \left(\frac{2PU_i}{n} \right) \cos iCX + \left(\frac{2PV_i}{n} \right) \sin iCX, \end{aligned}$$

где X — число интервалов (единиц) времени от начала цикла и C — величина каждого отрезка в градусах.

Сумма квадратов кривой любого порядка имеет 2 степени свободы и определяется по формуле

$$SS \text{ для степени } i = \frac{2(PU_i^2 + PV_i^2)}{n},$$

а сумма квадратов отклонений от кривой может быть получена путем вычитания сумм квадратов компонентов регрессии из общей SS .

Методы этой главы применимы лишь в том случае, когда величины X равно удалены друг от друга, за исключением тех случаев, когда анализируют наиболее часто встречающиеся группы вариантов неравновеликого интервала, для которых в приложении (в таблице А.11а) приводятся ортогональные коэффициенты, используемые только для вычисления сумм квадратов регрессии, но не уравнений регрессии.

МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

В предыдущих главах мы обсуждали связи только между двумя переменными. Часто нас интересует связь одной зависимой переменной с более чем одной независимой переменной. Например, закон спроса и предложения подразумевает связь между ценой (зависимая переменная) и двумя переменными — спросом и предложением. В опытах на животных мы можем быть заинтересованы в прибавках массы тела в зависимости от использования различных составных частей корма. Цель полевых опытов может сводиться к изучению действия различных доз азота, фосфора и калия на урожайность культур.

КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

Корреляция между двумя переменными независимо от любых других переменных, варьирующих одновременно с рассматриваемыми, называется простой, или общей корреляцией. Корреляция между двумя переменными, когда одна или большее число других переменных удерживаются на постоянном уровне, называется частной корреляцией. Объединенная связь одной переменной с двумя или большим числом других переменных, варьирующих одновременно, называется множественной корреляцией.

Предположим, что у нас есть зависимая переменная Y и для каждого значения Y имеются соответствующие значения двух других переменных X_1 и X_2 . Простую, или общую, корреляцию между Y и X_1 выражает коэффициент линейной корреляции, который мы рассматривали в главе 13. Вы помните, что он вычислялся по формуле *

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2}.$$

Для того чтобы подчеркнуть, что мы имеем дело с простой корреляцией между Y и X , обычно используют пояснительные подписи, так что формула примет следующий вид:

$$r^2_{YX_1} = \frac{(\sum x_1 y)^2}{\sum x_1^2 \sum y^2}.$$

* Как и прежде, в этой формуле был использован коэффициент r^2 , а не r . Необходимо помнить, что r является коэффициентом корреляции, получаемой путем извлечения корня квадратного из r^2 .

Подобным образом записывается простая корреляция между Y и X_2 :

$$r_{YX_2}^2 = \frac{(\sum x_2 y)^2}{\sum x_2^2 \sum y^2}.$$

Наконец, для расчета частной и множественной корреляции требуется третья простая корреляция между X_1 и X_2 :

$$r_{X_1 X_2}^2 = \frac{(\sum x_1 x_2)^2}{\sum x_1^2 \sum x_2^2}.$$

Частная корреляция между Y и X_1 при фиксированном X_2 обозначается как $r_{YX_1 \cdot X_2}$ и вычисляется на основе простых корреляций следующим образом:

$$r_{YX_1 \cdot X_2}^2 = \frac{(r_{YX_1} - r_{YX_2} r_{X_1 X_2})^2}{(1 - r_{YX_2}^2)(1 - r_{X_1 X_2}^2)}.$$

Подобно этому

$$r_{YX_2 \cdot X_1}^2 = \frac{(r_{YX_2} - r_{YX_1} r_{X_1 X_2})^2}{(1 - r_{YX_1}^2)(1 - r_{X_1 X_2}^2)}.$$

Коэффициент множественной корреляции, обозначаемый через $R_{Y \cdot X_1 X_2}$, измеряет объединенную связь X_1 и X_2 с Y . Его находят путем извлечения квадратного корня.

$$R_{Y \cdot X_1 X_2}^2 = \frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1} r_{YX_2} r_{X_1 X_2}}{1 - r_{X_1 X_2}^2}.$$

Точно так же, как r^2 , был назван коэффициентом детерминации (определения), так и R^2 называется множественным коэффициентом детерминации. Он дает долю вариации Y , приходящуюся на вариацию двух или большего числа независимых переменных.

Отметьте, как добавление всего лишь одной переменной повысило комплексность корреляции. При двух переменных X и Y мы имеем только один коэффициент корреляции. При трех переменных (X_1 , X_2 и Y) мы имеем три простых коэффициента, три частных и множественный коэффициент.

Проблема изображения связи трех переменных также становится более трудной, чем для двух переменных. При расчете двух переменных мы можем отложить величины наблюдения на двумерном графике. В таком случае связь описывается линией регрессии, а при большем числе переменных точечная диаграмма представляет эллипс. Чем уже эллипс, тем сильнее корреляция. С тремя переменными связь должна описываться плоскостью в трехмерном пространстве. Разброс точек вокруг этой плоскости будет представлять форму эллипсоида. Проекция эллипсоида на плоскость $X_1 Y$ показывает простую корреляцию X_1 и Y . Сечение эллипсоида, параллельное плоскости $X_1 Y$ и спроектированное на плоскость $X_1 Y$, покажет частную корреляцию X_1 и Y при фиксированном X_2 , обозначаемую $r_{YX_1 \cdot X_2}$.

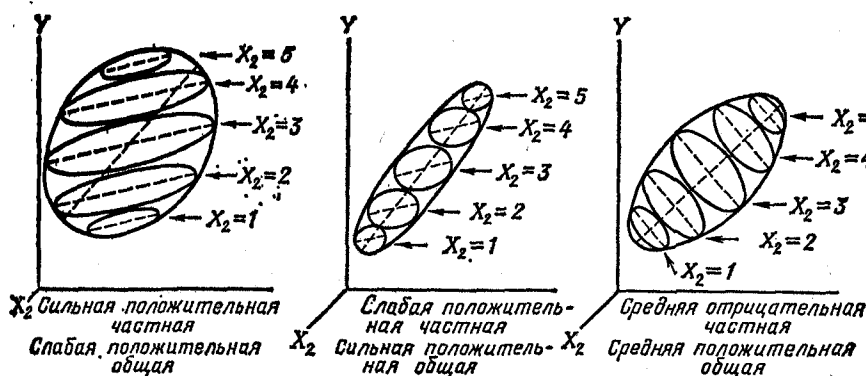


Рис. 16.1. Диаграммы различных сочетаний частной и общей корреляций в случае трех переменных.

Различные ситуации показаны диаграммами на рисунке 16.1. Отметим, что простая корреляция может быть слабой, а частная корреляция сильной, и наоборот. Они могут различаться даже знаками.

Множественный коэффициент корреляции R показывает, как тесно точки эллипсоида собраны вокруг плоскости регрессии. Величина R всегда положительна и варьирует от нуля до единицы. Более того, эта величина почти всегда равна самым большим величинам простого и частного коэффициентов. Это условие используется в качестве хорошей проверки расчетов.

КОЭФФИЦИЕНТЫ РЕГРЕССИИ

До сих пор мы обсуждали только коэффициенты корреляции, определяющие силу связи между переменными. Интересно знать также и природу связи. Насколько изменяется Y при изменении независимых переменных на единицу измерений? Для ответа на этот вопрос требуется уравнение следующей формы:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2.$$

Выражения b_1 и b_2 называются коэффициентами частной регрессии. Лучшим уравнением такой формы будет уравнение, преобразующее сумму квадратов отклонений фактических значений Y от расчетных величин $\hat{Y}$ к минимальным. Для нахождения величин a , b_1 и b_2 , отвечающих этому требованию, необходимо решить нормальные уравнения, очень похожие на нормальные уравнения, решаемые при анализе криволинейной регрессии:

$$\begin{aligned}
 an + b_1 \sum X_1 + b_2 \sum X_2 + \dots &= \sum Y; \\
 a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + \dots &= \sum X_1 Y; \\
 a \sum X_2 + b_1 \sum X_2 X_1 + b_2 \sum X_2^2 + \dots &= \sum X_2 Y. \\
 \dots \dots \dots
 \end{aligned}$$

Точки указывают, что ряд уравнений может быть продолжен для включения более трех переменных.

Вычисления могут быть сокращены путем замены в уравнениях исходных величин X на отклонения от средних. Поскольку сумма отклонений любой переменной от ее среднего значения равна нулю, то $\Sigma x_1 = \Sigma x_2 = \Sigma y = 0$. Поэтому первое нормальное уравнение выпадает, как и все первые члены оставшихся уравнений, и остаются в виде

$$b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + \dots = \Sigma x_1 y;$$

$$b_1 \Sigma x_2 x_1 + b_2 \Sigma x_2^2 + \dots = \Sigma x_2 y.$$

.....

Решение этих уравнений относительно значений b дает уравнение регрессии следующей формы: $\hat{y} = b_1 x_1 + b_2 x_2 + \dots$. Если мы желаем иметь уравнение в исходных единицах наблюдений, то можно вычислить $a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 \dots$. Тогда уравнение примет вид $\hat{Y} = a + b_1 X_1 + b_2 X_2 + \dots$.

ПРИМЕР КОРРЕЛЯЦИИ С ТРЕМЯ ПЕРЕМЕННЫМИ

Для иллюстрации частной и множественной корреляции и регрессии мы проанализируем некоторые данные удельной массы картофеля (Y), содержания азота (X_1) и содержания фосфора (X_2). Для упрощения расчетов фактические даты кодируют (табл. 16.1).

Вначале мы вычислим различные коэффициенты корреляции. Простые или общие корреляции примут следующие значения:

$$r_{YX_1}^2 = (\Sigma y x_1)^2 / \Sigma y^2 \Sigma x_1^2 = \frac{-29\,218,35^2}{(51\,172,95)(21\,240,55)} = 0,7854,$$

$$r_{YX_1} = \sqrt{r_{YX_1}^2} = -0,8862 \text{ (отметьте, что отрицательный знак}$$

коэффициента обусловлен отрицательным знаком выражения $\Sigma y x_1$);

$$r_{YX_2}^2 = (\Sigma y x_2)^2 / \Sigma y^2 \Sigma x_2^2 = \frac{-6611,8^2}{(51\,172,95)(1663,2)} = 0,5136;$$

$$r_{YX_2} = \sqrt{r_{YX_2}^2} = -0,7167;$$

$$r_{X_1 X_2}^2 = (\Sigma x_1 x_2)^2 / \Sigma x_1^2 \Sigma x_2^2 = \frac{2584,4^2}{(21\,240,55)(1663,2)} = 0,1891;$$

$$r_{X_1 X_2} = \sqrt{r_{X_1 X_2}^2} = 0,4348.$$

Коэффициенты частной корреляции примут следующие значения:

$$\begin{aligned} r_{YX_1 \cdot X_2}^2 &= \frac{(r_{YX_1} - r_{YX_2} r_{X_1 X_2})^2}{(1 - r_{YX_2}^2)(1 - r_{X_1 X_2}^2)} = \frac{[-0,8862 - (-0,7167)(0,4338)]^2}{(1 - 0,5136)(1 - 0,1891)} = \\ &= \frac{-0,5746^2}{(0,4864)(0,8109)} = 0,8371; \end{aligned}$$

$$r_{YX_1 \cdot X_2} = \sqrt{r_{YX_1 \cdot X_2}^2} = -0,9149;$$

$$\begin{aligned} r_{YX_2 \cdot X_1}^2 &= \frac{(r_{YX_2} - r_{YX_1} r_{X_1 X_2})^2}{(1 - r_{YX_1}^2)(1 - r_{X_1 X_2}^2)} = \frac{[-0,7167 - (-0,8862)(0,4348)]^2}{(1 - 0,7854)(1 - 0,1891)} = \\ &= \frac{-0,3314^2}{(0,2146)(0,8109)} = 0,6310; \end{aligned}$$

$$r_{YX_2 \cdot X_1} = \sqrt{r_{YX_2 \cdot X_1}^2} = -0,7944.$$

Таблица 16.1. Удельная масса, содержание азота и фосфора в 20 выборках картофеля

У удельная масса (-1,07) 10 ⁴	X ₁ (азот-1)× ×100	X ₂ (фосфор)× ×100	У удельная масса (-1,07) 10 ⁴	X ₁ (азот-1)× ×100	X ₂ (фосфор)× ×100
2	96	40	83	73	15
14	82	36	91	58	26
15	121	30	97	31	25
15	88	42	98	38	24
16	100	28	101	56	11
27	114	26	128	24	22
48	71	33	140	37	11
54	94	26	163	10	24
58	74	15	179	14	10
68	36	35			
82	36	25			
			Суммы 1,479	1,253	504
ΣY ² = 160545;			ΣX ₁ ² = 99 741;	ΣX ₂ ² = 14 364;	
(ΣY) ² /20 = 109 372,05;			(ΣX ₁) ² /20 = 78 500,45;	(ΣX ₂) ² /20 = 12 700,8;	
Σy ² = 51 172,95;			Σx ₁ ² = 21 240,55;	Σx ₂ ² = 1 663,2;	
ΣYX ₁ = 63 441;			ΣYX ₂ = 30 659;	ΣX ₁ X ₂ = 34 160;	
ΣYΣ X ₁ /20 = 92 659,35;			ΣYΣ X ₂ /20 = 37 270,8;	ΣX ₁ ΣX ₂ /20 = 31 575,6;	
Σyx ₁ = -29 218,35;			Σyx ₂ = -6 611,8;	Σx ₁ x ₂ = 2 584,4.	

Наконец, мы рассчитаем R — множественный коэффициент:

$$\begin{aligned}
 R_{Y.X_1X_2}^2 &= \frac{r_{YX_2}^2 + r_{YX_1}^2 - 2r_{YX_1}r_{YX_2}r_{X_1X_2}}{1 - r_{X_1X_2}^2} = \\
 &= \frac{0,5136 + 0,7854 - 2(-0,8862)(-0,7167)(0,434)}{1 - 0,1891} = \\
 &= \frac{0,7467}{0,8109} = 0,9208; \\
 R_{Y.X_1X_2} &= \sqrt{0,9208} = 0,9596.
 \end{aligned}$$

Простые корреляции как азота, так и фосфора по отдельности с удельной массой не очень велики, однако если две переменные рассматриваются одновременно, то связь их с удельной массой становится очень тесной. В процентах азот отдельно обуславливает вариативность удельной массы на 78,54% ($100 \times r_{YX_1}^2$). На фосфор приходится 51,36%. Азот и фосфор вместе обуславливают 92,08% вариативности удельной массы.

Теперь нам необходимо описать связь путем расчета уравнения регрессии. Нормальные уравнения, выраженные в форме отклонений от средних, примут вид:

$$\begin{aligned}
 b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 &= \Sigma x_1 y; \\
 b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 &= \Sigma x_2 y.
 \end{aligned}$$

Подставим фактические даты из данных:

$$21\,240,55b_1 + 2584,4b_2 = -29\,218,35;$$

$$2584,40b_1 + 1663,2b_2 = -6661,8.$$

Умножив первое уравнение на 2584,4 и второе уравнение на 21 240,55 и произведя вычитание, мы получим

$$28\,648\,159,4b_2 = -64\,926\,364,75;$$

$$b_2 = -2,666.$$

Подставив эту величину b_2 в каждое из исходных уравнений и решив их относительно b_1 , мы найдем, что $b_1 = -1,100$.

Для того чтобы получить уравнение регрессии, выраженное в исходных величинах, необходимо найти a :

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2, \text{ тогда получим}$$

$$\frac{1479}{20} - \left(-1,100 \frac{1253}{20}\right) - \left(-2,666 \frac{504}{20}\right) = 199,968$$

Теперь мы можем записать уравнение регрессии $\hat{Y} = 199,968 - 1,100X_1 - 2,266X_2$.

На основе этого уравнения мы можем вычислить величины $\hat{Y}$ и сравнить их с фактическими датами (табл. 16.2).

Сумма отклонений равна нулю, как и должно быть. Это обеспечивает хорошую проверку расчетов. Сумма квадратов отклонений равна 4051,16. Эта сумма представляет вариацию удельной массы (Y), не связанную с вариацией в содержании азота (X_1) или фосфора (X_2). Она может быть вычислена без предварительного расчета каждого значения Y с помощью следующей формулы: $(1-R^2)\Sigma y^2$, что дает

$$(1-0,9208)51172,95 = 4052,90.$$

Оба ответа находятся в близком соответствии (небольшая разница обусловлена погрешностью округлений).

Таблица 16.2. Фактические и расчетные удельные массы 20 образцов картофеля

Y	$\hat{Y}$	$d = Y - \hat{Y}$	Y	$\hat{Y}$	$d = Y - \hat{Y}$
2	3,7	-1,7	82	103,7	-21,7
14	28,2	-14,2	83	85,7	-2,7
15	-1,1	16,1	91	77,2	13,9
15	8,0	7,0	97	109,3	-12,2
16	26,5	-10,5	98	103,8	-5,8
27	-15,7	11,3	101	113,4	-12,4
48	47,1	0,9	128	123,7	4,3
54	37,7	16,3	140	134,3	5,7
58	84,6	-26,6	163	134,6	28,4
68	81,1	-13,1	179	161,9	17,1

Полученные нами результаты могут быть сведены в таблице дисперсионного анализа следующим образом.

Источник вариации	Метод расчета	SS	df	MS	F
Общий	Σy^2	51 172,95	19		
Регрессия на X_1	$r^2_{YX_1} (\Sigma y^2)$	40 191,23	1	40 191,23	65,9 **
Отклонения от регрессии	$(1 - r^2_{YX_1}) \Sigma y^2$	10 981,72	18	610,10	
Дополнительная регрессия на X_2	$r^2_{YX_1 \cdot X_2} (1 - r^2_{YX_1}) \Sigma y^2$	6 929,47	1	6 929,47	29,07 **
Отклонения от множественной регрессии	$(1 - R^2_{Y \cdot X_1 X_2}) \Sigma y^2$	4 052,90	17	238,41	

Последняя сумма квадратов может быть получена путем вычитания $10\,981,72 - 6\,929,47 = 4\,052,25$. Расхождение между этой величиной и приведенной в таблице есть результат погрешности округлений и не будет иметь важного значения для F -оценки. Корень квадратный из 238,41, или 15,44, называется стандартной ошибкой оценки и обозначается символом $S_{Y \cdot X_1 X_2}$.

В статистике имеется другой путь составления таблицы дисперсионного анализа, который дает совершенно другие значения F .

Источник вариации	Метод расчета SS	SS	df	MS	F
Общий	Σy^2	51 172,95	19		
Регрессия на X_1	$r^2_{YX_1} (\Sigma y^2)$	26 282,43	1	26 282,43	19,01 **
Отклонения от простой регрессии	$(1 - r^2_{YX_1}) \Sigma y^2$	24 890,52	18	1 382,81	
Дополнительная регрессия на X_2	$r^2_{YX_1 \cdot X_2} (1 - r^2_{YX_1}) \Sigma y^2$	20 835,85	1	20 835,85	87,40 **
Отклонения от множественной регрессии	$(1 - R^2_{Y \cdot X_1 X_2}) \Sigma y^2$	4 052,90	17	238,41	

В первой из этих двух таблиц мы рассматривали общий эффект азота и затем дополнительное действие фосфора. Во второй таблице рассматривался общий эффект фосфора и затем дополнительное действие азота. Последовательность проведения анализа переменных значительно влияет на результаты общего анализа и может привести в замешательство исследователя, впервые применившего анализ множественной регрессии.

Простой пример может прояснить эту путаницу. Хорошо известно, что на урожайность многих культур влияют как температура воздуха, так и долгота дня. Предположим, что у нас есть результаты

многочисленных учетов урожайности культуры, выращиваемой в разное время года. Для каждого значения урожайности мы имеем соответствующие значения средней долготы дня и средней температуры воздуха за вегетационный период. Мы предполагаем, что долгота дня и температура воздуха тесно коррелируют друг с другом. Поскольку это верно, то не удивительно, если мы установили, что урожайность тесно коррелировала с температурой, а дополнительный анализ эффекта долготы дня мало объяснил не учтенную вначале вариацию урожайности. В то же самое время долгота дня отдельно может тесно коррелировать с урожайностью, тогда как температура может оказывать малый дополнительный эффект. Уместно заключить, что продолжительные теплые дни скорее ассоциируются с высокими урожаями, чем короткие холодные дни. У нас мало сведений о том, какой из факторов — температура или долгота дня — был более важен. Для ответа на этот вопрос необходимо поставить эксперимент, в котором долгота дня и (или) температура контролировались бы так, что между ними была бы менее тесная корреляция, чем в естественных условиях.

В главе 13 мы привели пример «ложной» корреляции между потреблением сигарет и производством сена. Сильная корреляция в этом случае была обусловлена, вероятно, тем фактором, что обе переменные были тесно связаны с третьей переменной — временем. Анализ множественной регрессии вскроет разительное различие между двумя анализами в зависимости от того, какая из независимых переменных рассматривается в них первой, таблица 16.3.

Во втором анализе, где регрессия на время рассматривается первой, установлена несущественная дополнительная регрессия производства сена на потребление сигарет.

Таблица 16.3. Анализ множественной регрессии производства сена (Y), потребления сигарет (X_1) и время (X_2)

Источник вариации	d	SS	MS	F
X_1 рассматривается первой				
Общий	14	10 094,00		
Регрессия на X_1	1	8 855,31	8855,31	92,94 **
Отклонения от простой регрессии	13	1 238,69	95,28	
Дополнительная регрессия на X_2	1	918,01	918,01	84,35 **
Отклонения от множественной регрессии	12	320,67	26,72	
X_2 рассматривается первой				
Общий	14	10 094,00		
Регрессия на X_2	1	9 723,21	9723,21	320,90 **
Отклонения от простой регрессии	13	370,79	28,52	
Дополнительная регрессия на X_1	1	50,11	50,11	1,88 несущественна
Отклонения от множественной регрессии	12	320,67	26,72	

КОРРЕЛЯЦИЯ ДЛЯ ТРЕХ И БОЛЕЕ ПЕРЕМЕННЫХ

В целях простоты большая часть текста и иллюстрирующих примеров этой главы основывалась на трех переменных — одной зависимой двух независимых. Фактически коэффициенты множественной частной корреляции и уравнения регрессии могут быть рассчитаны для любого числа переменных. Недавнее исследование, проведенное в Калифорнийском университете, включало 35 переменных. В данном пособии мы можем показать лишь общий способ применения описанных методов для трех и более переменных и указать на некоторые из трудностей, сопутствующих анализу.

Мы уже показывали, каким образом может быть продолжен ряд нормальных уравнений для расчета коэффициентов регрессии b_1 , b_2 и т. д., включающих столько переменных, сколько мы пожелаем. Каждая новая переменная требует включения следующего выражения с левой стороны каждого уравнения и включения нового уравнения с таким же количеством членов, как и все предшествующие. Для m -переменных последнее нормальное уравнение будет представлять

$$b_1 \sum x_1 x_m + b_2 \sum x_2 x_m + b_3 \sum x_3 x_m + \dots + b_m \sum x_m^2 = \sum x_m y.$$

По мере добавления новых переменных алгебраическая суть проблемы не изменяется, тогда как трудности арифметических расчетов, связанных с решением уравнений, возрастают. По этой причине рекомендуется использовать одну из систематических процедур, описанную в предшествующей главе, или использовать по возможности ЭВМ.

Итак, мы убедились, что в случае только двух переменных рассчитывался 1 коэффициент корреляции, а при наличии трех переменных — 7, включая 1 множественный, 3 простых и 3 частных коэффициента. Для четырех переменных общее число коэффициентов возрастает до 25, а для пяти — до 81. Одной из причин их значительного увеличения является появление частных коэффициентов высокого порядка. Порядком коэффициента частной корреляции служит число фиксированных переменных. При трех переменных мы имели коэффициенты только первого порядка, такие, как $r_{YX_1 \cdot X_2}$. При четырех переменных мы имеем простые и частные первого порядка и частные коэффициенты второго порядка, такие, как $r_{YX_1 \cdot X_2 \cdot X_3}$, что выражает корреляцию Y и X_1 при фиксированных значениях X_2 и X_3 .

В математической статистике имеется общее уравнение, которое позволяет нам вычислить частный коэффициент любого порядка при наличии трех коэффициентов предшествующего порядка

$$r_{YX_1 \cdot X_2 \cdot X_3 \dots X_m}^2 = \frac{(r_{YX_1 \cdot X_2 \dots X_m} - r_{YX_2 \cdot X_3 \dots X_m} r_{X_1 X_2 \cdot X_3 \dots X_m})^2}{(1 - r_{YX_2 \cdot X_3 \dots X_m}^2)(1 - r_{X_1 X_2 \cdot X_3 \dots X_m}^2)}.$$

Ранее приведенные уравнения для определения частных коэффициентов первого порядка, включающие три переменных из трех простых

корреляций, являются просто частным случаем этого общего уравнения.

Общей формулой определения множественного коэффициента корреляции, включающей m -переменных, является

$$1 - R_{Y \cdot x_1 \dots x_m}^2 = (1 - r_{Yx_1}^2)(1 - r_{Yx_2 \cdot x_1}^2)(1 - r_{Yx_3 \cdot x_1 x_2}^2) \dots (1 - r_{Yx_m \cdot x_1 \dots x_{m-1}}^2).$$

В случае двух независимых переменных это уравнение сокращается до довольно простой формы, уже приводившейся для $R_{Y \cdot x_1 x_2}^2$.

Итак, мы увидели, что трудности арифметических расчетов возрастают по мере увеличения количества рассматриваемых переменных, однако, вероятно, самая большая трудность при анализе более чем трех переменных заключается в графическом изображении связей. Связь двух переменных может быть представлена двухмерным графиком. Связи между тремя переменными могут быть представлены трехмерной диаграммой. Но как изобразить картину связей между четырьмя или большим числом переменных? У нас нет ответа на этот вопрос. Мы не должны быть обеспокоены нашей неспособностью представить связи в четырех или большем числе измерений. Более того, наши умозаключения должны в большей мере касаться уравнений, чем диаграмм. В конце концов, у нас не вызывает затруднений сама идея связи урожайности культуры с уровнем содержания N, P и K в почве, количеством доступной воды, числом сорняков, распространением болезней и вредителей, температурой воздуха и долготой дня. При достаточном объеме данных мы могли бы составить легко решаемые уравнения, описывающие эти связи. Должно ли нас беспокоить то, что мы не можем составить такие легко решаемые уравнения? Должно ли нас беспокоить то, что мы не можем представить диаграмму, изображающую комплексное взаимодействие факторов? Ответ заключается в том, что одно уравнение может стоять тысячи диаграмм.

Необходимо упомянуть еще один момент относительно корреляции и регрессии большого числа переменных. Мы показали, что в случае трех переменных могут быть выполнены два различных анализа в зависимости от того, какая из независимых переменных рассматривается первой. При трех независимых переменных число возможных анализов возрастает до шести, а числа m независимых переменных будет $m!$ возможных перестановок переменных. (Символ « $m!$ » читается, как m -факториал и означает произведение всех чисел от единицы до m . Так, $10! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 = 3\,628\,800$). Каким будет лучший порядок группировки переменных при анализе? Или в другой формулировке: «Из большого числа независимых каким образом найти лучшую группу заданного размера?» Выбор прямого простого метода для получения лучшей группы является одной из самых нерешенных проблем статистики. Разработаны программы подхода к ее решению на ЭВМ, однако здесь лимитирующим фактором остается время.

ПОВЕРХНОСТИ ОТКЛИКА

Для анализа множественной регрессии не требуется иметь обязательно несколько независимых переменных. Они могут выражаться разной степенью одной и той же переменной, например: X , X^2 и X^3 или произведениями двух или большего числа переменных, например X_1X_2 , $X_1^2X_2$ и т. д. Так, подбор полиномиальной кривой является частным случаем анализа множественной регрессии. В главе 14 было указано, что доля вариабельности Y , приходящаяся на линейную плюс квадратичную сумму квадратов, обозначается R^2 , или коэффициентом множественной детерминации. Если мы имеем две переменных, существенно криволинейно связанных с Y , то можно найти уравнение, описывающее эту полную связь. Мы можем не только найти такое уравнение, но также показать эту связь графически с помощью одного из нескольких способов трехмерного изображения. Такой график называется поверхностью отклика.

Эксперимент с азотным удобрением и сроками уборки сахарной свеклы, описанный в главе 10, является хорошим примером. В той главе мы выявили существенные компоненты (суммы квадратов) дисперсионного анализа данных эксперимента: линейную и квадратичную для азота, линейную и квадратичную для сроков уборки, линейную и квадратичную взаимодействия азота $\times$ сроки уборки, квадратичную азота $\times$ линейную сроков уборки. Для представления действия всех этих компонентов на урожайность в одном уравнении потребуется уравнение следующей формы:

$$\hat{Y} = a + bN + cH + dN^2 + eH^2 + fNH + gN^2H.$$

Для нахождения этого уравнения необходимо решить семь систем уравнений с семью неизвестными. Исходными нормальными уравнениями будут:

$$\begin{aligned} an + b\sum H + c\sum N + d\sum H^2 + e\sum N^2 + f\sum NH + g\sum N^2H &= \sum Y; \\ a\sum H + b\sum H^2 + c\sum NH + d\sum H^3 + e\sum N^2H + f\sum NH^2 + g\sum N^2H^2 &= \sum HY; \\ a\sum N + b\sum NH + c\sum N^2 + d\sum NH^2 + e\sum N^3 + f\sum N^2H + g\sum N^3H &= \sum NY; \\ a\sum H^2 + b\sum H^3 + c\sum NH^2 + d\sum H^4 + e\sum N^2H^2 + f\sum NH^3 + g\sum N^2H^3 &= \sum H^2Y; \\ a\sum N^2 + b\sum N^2H + c\sum N^3 + d\sum N^2H^2 + e\sum N^4 + f\sum N^3H + g\sum N^4H &= \sum N^2Y; \\ a\sum NH + b\sum NH^2 + c\sum N^2H + d\sum NH^3 + e\sum N^3H + f\sum N^2H^2 + g\sum N^3H^2 &= \\ = \sum NHY; \\ a\sum N^2H + b\sum N^2H^2 + c\sum N^3H + d\sum N^2H^3 + e\sum N^4H + f\sum N^3H^2 + g\sum N^4H^2 &= \\ = \sum N^2HY. \end{aligned}$$

Решение этих семи систем уравнений покажется на первый взгляд невероятно трудной задачей, однако при надлежащем кодировании величин N и H многие из сумм получают нуль и уравнения в значительной мере упростятся. Величина H , поскольку она представляет пять сроков уборки равновеликого интервала между ними, кодируется путем использования c_1 -коэффициентов под $n=5$ в таблице А.11 (приложение). Искомые коэффициентами будут -2 , -1 , 0 , 1 и 2 .

Для кодирования уровней азота мы делим фактические дозы на 80 и получаем серии 0, 1, 2 и 4; линейными коэффициентами этих серий будут $-7, -3, -1$ и 9 (табл. А.11а).

Использование этих кодированных величин сводит к нулю следующие выражения (члены) нормальных уравнений: $\Sigma H, \Sigma N, \Sigma NH, \Sigma N^2H, \Sigma H^3, \Sigma NH^2, \Sigma N^3H, \Sigma NH^3, \Sigma N^2H^2, \Sigma N^4H$. И для подстановки в нормальные уравнения используют оставшиеся после сокращения суммы:

$$\begin{array}{ll} \Sigma H^2 = 40 & n = 20 \\ \Sigma N^2 = 700 & \Sigma Y = 1\,600 \\ \Sigma N^2 H^2 = 1\,400 & \Sigma HY = 751 \\ \Sigma N^3 = 1\,800 & \Sigma NY = 1\,430,4 \\ \Sigma H^4 = 136 & \Sigma H^2 Y = 3\,006,6 \\ \Sigma N^4 = 45\,220 & \Sigma N^2 Y = 54\,867,2 \\ \Sigma N^3 H^2 = 3\,600 & \Sigma NHY = 744,2 \\ \Sigma N^4 H^2 = 90\,440 & \Sigma N^2 HY = 25\,967,8 \end{array}$$

Нормальные уравнения примут вид:

$$\begin{array}{llll} 20a & + & 40d + 700e & = 1\,600,0; \quad (1) \\ 40b & & & 1\,400g = 751,0; \quad (2) \\ 700c & + & 1\,800e & = 1\,430,4; \quad (3) \\ 40a & + & 136d + 1\,400e & = 3\,006,6; \quad (4) \\ 700a + 1\,800c + 1\,400d + 45\,220e & & & = 54\,867,2; \quad (5) \\ & 1\,400f + & 3\,600g & = 744,2; \quad (6) \\ - 1\,400b & + & 3\,600f + 90\,440g & = 25\,967,8; \quad (7) \end{array}$$

Умножение уравнения (2) на 35 и последующее вычитание из уравнения (7) даст

$$3\,600f + 41\,440g = -317,2. \quad (8)$$

Уравнение (8), умноженное на 7, минус уравнение (6), умноженное на 18, даст

$$225\,280g = -15,616; \quad g = -0,069318.$$

Подстановка g в уравнение (6) дает $f = 0,709818$.

Подстановка g в уравнение (2) дает $b = 21,201136$.

Умножение уравнения (1) на 2 и последующее вычитание из уравнения (4) дает

$$56d = -193,4; \quad d = -3,453571.$$

Умножение уравнения (1) на 35 и последующее вычитание из уравнения (5) даст

$$1\,800c + 20\,720e = -1132,8. \quad (9)$$

Уравнение (9), умноженное на 7, минус уравнение (3), умноженное на 18, даст

$$112\,640e = -33\,676,8; \quad e = -0,298977.$$

Подстановка d и e в уравнение (1) дает

$$20a=1947,4269; \quad a=97,371345.$$

Подстановка e в уравнение (3) дает

$$700c=1968,559; \quad c=2,812227.$$

Теперь у нас есть все члены уравнения, выраженные в кодированных величинах N и H .

$$\hat{Y} = 97,371345 + 21,201136H' + 2,812227N' - \\ - 3,453571H'^2 - 0,298977N'^2 + 0,709818N'H' - 0,069318N'H'.$$

Исходные градации сроков уборки составляли 0, 3, 6, 9 и 12 недель, а дозами азота были 0; 0,8; 1,6 и 3,2 хандредвейта/акр ($\approx$ ц/га). Для применения в полученном уравнении этих единиц мы должны заменить в нем H' на $(H/3)-2$ и N' на $5N-7$. (Необходимые уравнения для перехода от H' и N' к исходным величинам H и N смотрите в заключении главы 15.) Итоговое уравнение будет иметь следующий вид:

$$\hat{Y} = 23,55 + 8,8834H + 18,1868N - 0,38373H^2 - 4,00853N^2 + \\ + 2,80045NH - 0,577652N^2H.$$

Подстановка значений N и H в это уравнение даст расчетные величины, приведенные в таблице 16.4.

Таблица 16.4. Фактические и предсказанные (расчетные) урожайности сахарной свеклы для каждого варианта, основанные на суммах четырех повторений

N	H	Y	$\hat{Y}$	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
0,0	0	22,0	23,55	-1,55	2,4025
0,0	3	47,4	46,75	0,65	0,4225
0,0	6	61,1	63,04	-1,94	3,7636
0,0	9	69,8	72,42	-2,62	6,8644
0,0	12	76,1	74,89	1,21	1,4641
0,8	0	39,4	35,53	3,87	14,9769
0,8	3	67,9	64,34	3,56	12,6736
0,8	6	85,6	86,24	-0,64	0,4096
0,8	9	105,0	101,24	3,76	14,1376
0,8	12	110,1	109,33	0,77	0,5929
1,6	0	40,7	42,39	-1,69	2,8561
1,6	3	74,4	74,59	-0,19	0,0361
1,6	6	91,9	99,88	-7,98	63,6804
1,6	9	120,1	118,27	1,83	3,3489
1,6	12	129,3	129,75	-0,45	0,2025
3,2	0	37,9	40,70	-2,80	7,8400
3,2	3	77,5	73,04	4,46	19,8916
3,2	6	96,6	98,46	-1,86	3,4596
3,2	9	122,1	116,98	5,12	26,2144
3,2	12	125,1	128,60	-3,50	12,2500
Суммы		1600,0	1599,99	0,01	197,4873

Сумму квадратов отклонений фактических от предсказанных дат необходимо разделить на 4 для перевода ее на поделяночную основу, поскольку в таблице даны суммарные урожайности по четырем повторениям.

Это дает 49,372, что точно совпадает с общей остаточной суммой квадратов дисперсионного анализа в главе 10.

Если остаточная сумма квадратов, деленная на общую сумму квадратов для вариантов, равняется выражению $(1-R^2)$, то в данном случае

$$(1-R^2) = \frac{49,372}{4969,24} = 0,0099$$

$$R^2 = 1 - 0,0099 = 0,9901.$$

Таким образом, 99% варибельности Y учитывается рассчитанным нами уравнением.

Поделив каждый член уравнения на 4 (повторность каждого варианта), получим $\bar{Y}$ в тоннах на акр (0,4 га), то есть в единицах, наи-

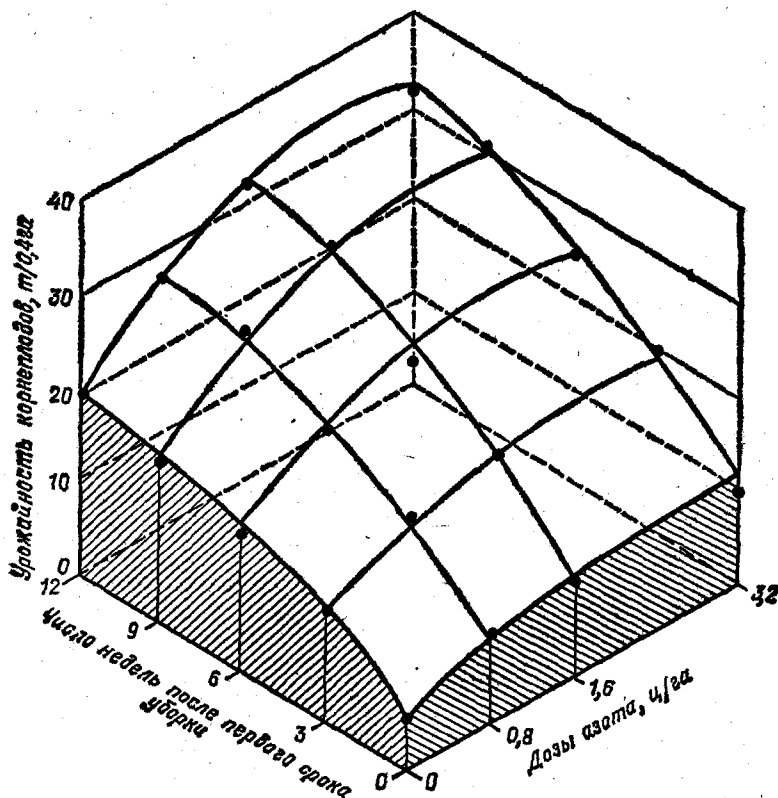


Рис. 16.2. Трехмерная поверхность отклика, соответствующая фактической отзывчивости (жирные точки) сахарной свеклы на дозы азотного удобрения и сроки уборки.

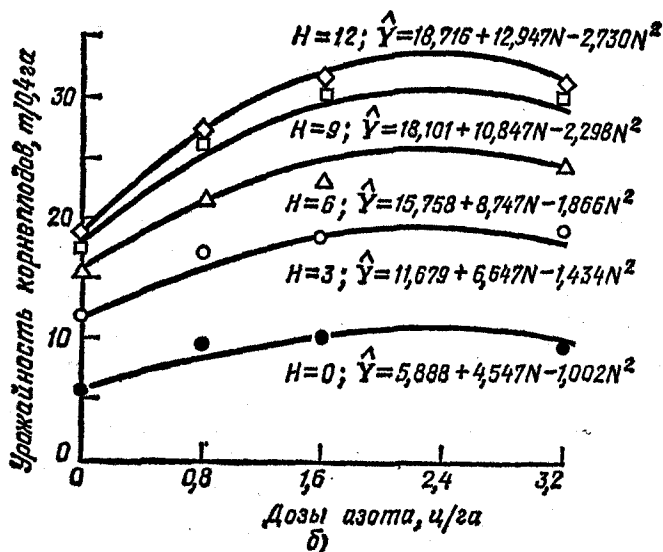
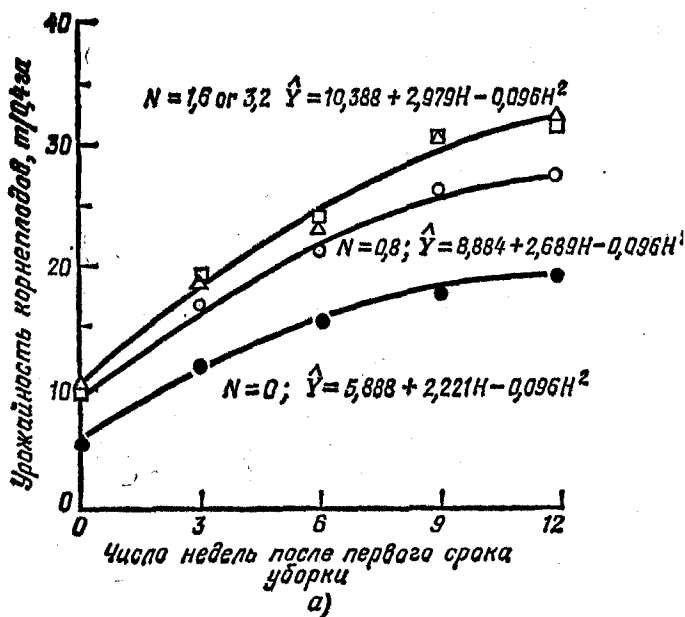


Рис. 16.3 «Срезы» поверхности отклика рисунка 16.2. Эффекты сроков уборки на каждом уровне азота (а) и эффекты уровней азота для каждого срока уборки (б). Уравнения для $N_{1,0}$ и $N_{2,1}$ были усреднены.

более подходящих для публикации. Форма уравнения при этом изменится на

$$\hat{Y} = 5,888 + 2,221H + 4,547N - 0,096H^2 - 1,002N^2 + 0,700NH - 0,144N^2H,$$

где N — хандредвейты $N/\text{акр}$ ($\approx$ ц/га) и H — число недель после даты первого срока уборки (H 1-го срока = 0). Результаты могут быть представлены несколькими способами в зависимости от целей исследования и условий, требующих особого рассмотрения.

На рисунке 16.2 представлена трехмерная поверхность отклика. На рисунке 16.3 показаны «срезы» поверхности отклика. Рисунок 16.3а дает уравнения и кривые отклика для эффектов времени уборки по каждому уровню азота, тогда как рисунок 16.3б дает то же самое для эффектов азота по каждому сроку уборки. Двухмерные уравнения рисунка 16.3а получены путем приравнивания N нулю в уравнении множественной регрессии и сокращения подобных членов, что дает $\hat{Y} = 5,888 + 2,221H - 0,096H^2$. Другие уравнения рисунка 16.3 получены подобным путем, приравнивания N по очереди к 0,8; 1,6 и 3,2. Уравнения для $N=1,6$ и 3,2 были усреднены, поскольку они почти идентичны. Уравнения рисунка 16.3б вычислены путем приравнивания H по очереди к 0, 3, 6 и 12.

Рисунок 16.3 иллюстрирует природу взаимодействия для выражений N -линейная $\times$ H -линейная и N -квадратичная $\times$ H -квадратичная. NH -выражение уравнения множественной регрессии имеет результатом различную N -линейную для каждого срока уборки и различную H -линейную для каждого уровня азота. Выражение N^2H имеет результатом различный квадратичный эффект азота для каждого срока уборки (рис. 16.3б). Для сопоставления отметьте, что в уравнении отсутствует NH^2 -выражение и как следствие квадратичный эффект H остается одинаковым по всем уровням N (рис. 16.3а).

ЗАКЛЮЧЕНИЕ

При анализе более двух переменных выделяют три вида коэффициентов корреляции.

Простая или общая корреляция — это линейная корреляция внутри любой пары переменных, независимо от значений оставшихся переменных.

Частная корреляция — это корреляция между двумя переменными, когда одна или больше из оставшихся переменных удерживаются на постоянном уровне.

Множественная корреляция — это объединенная корреляция между зависимой переменной и всеми независимыми переменными.

Формулой квадрата коэффициента простой корреляции будет

$$r_{YX_i}^2 = \frac{(\sum x_i y)^2}{\sum x_i^2 \sum y^2}.$$

квадрата коэффициента частной корреляции первого

$$r_{YX_i, X_j}^2 = \frac{(r_{YX_i} - r_{YX_j} r_{X_i X_j})^2}{(1 - r_{YX_j}^2)(1 - r_{X_i X_j}^2)}.$$

Порядком коэффициента частной корреляции служит число переменных, удерживаемых на постоянном уровне, символически изображаемом числом выражений индекса, следующего за точкой. При трех переменных мы можем иметь частные коэффициенты только первого порядка.

Коэффициент множественной корреляции при трех переменных находится из формулы

$$R_{Y \cdot X_1 X_2}^2 = \frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1}r_{YX_2}r_{X_1X_2}}{1 - r_{X_1X_2}^2}$$

Множественный коэффициент всегда положителен и, по крайней мере, равен самым большим величинам простого или частного коэффициента.

Уравнение регрессии описывает связь между зависимой переменной и всеми независимыми переменными. Оно имеет формулу

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 + \dots$$

Символы b_1 , b_2 и т. д. называются коэффициентами частной регрессии. Для нахождения уравнения регрессии, лучшим образом соответствующего фактическим данным, мы решаем следующие нормальные уравнения относительно коэффициентов частной корреляции

$$\begin{aligned} b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + \dots + b_m \Sigma x_1 x_m &= \Sigma x_1 y; \\ b_1 \Sigma x_2 x_1 + b_2 \Sigma x_2^2 + \dots + b_m \Sigma x_2 x_m &= \Sigma x_2 y; \\ &\vdots \\ b_1 \Sigma x_m x_1 + b_2 \Sigma x_m x_2 + \dots + b_m \Sigma x_m^2 &= \Sigma x_m y, \end{aligned}$$

где m — число независимых переменных.

Для решения нам требуется m уравнений с m числом членов левой стороны каждого уравнения.

Формулой для нахождения a из уравнения регрессии будет

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \dots - b_m \bar{X}_m.$$

(Символ $\bar{Y}$ означает среднее значение Y и определяется из $\Sigma Y/n$, где n — число наблюдений. Подобно $\bar{X}_i = \Sigma \bar{X}_i/n$ является средним значение X_i .) Символ $\bar{Y}$ — оценочная (расчетная) величина Y , получаемая из уравнения регрессии. Разность $Y - \bar{Y}$ представляет отклонение фактической величины от ее оценки и $\Sigma (Y - \bar{Y}) = 0$. Если эта сумма отклонений оказывается неравной нулю (без учета малой погрешности округлений), следовательно, была допущена ошибка при расчетах по формуле

$$\frac{\sum (Y - \hat{Y})^2}{\text{Общая сумма квадратов } Y} = 1 - R^2.$$

При анализе более трех переменных возникают три затруднения:

1) объем арифметических расчетов возрастает аналогично уровню подъему) экспоненциальной кривой;

2) графическое изображение связей становится труднее;

3) установление наилучшего порядка добавления или элиминирования переменных является трудной проблемой, и в статистике нет практического метода нахождения лучшей группы заданного размера из большего числа переменных.

Степени и произведения переменных могут рассматриваться в качестве дополнительных переменных. При использовании степеней и произведений двух независимых переменных для расчета уравнения множественной регрессии итоговые результаты можно представить графически в виде трехмерной поверхности отклика.

АНАЛИЗЫ ПОДСЧЕТОВ (численностей)

Большая часть материала этой книги касалась статистической обработки результатов измерений, таких, как масса, урожайность или высота. Однако мы не всегда измеряем некоторые характеристики индивидуума. Иногда мы просто классифицируем индивидуумы на две или большее число групп, такие, как: мертвые или живые; здоровые или больные; мужские или женские; красные, розовые или белые; первокурсники, второкурсники, младшие или старшие. Даже при изучении поддающихся измерению характеристик иногда более удобно классифицировать индивидуумы на группы. Например, мы хотим провести исследование, которое включало бы измерение сумм денежных доходов, приходящихся на человека. Многие из опрошенных, вероятно, откажутся назвать точную сумму, однако без колебаний назовут, к какой из трех или четырех категорий они принадлежат, и подобная классификация может быть достаточной для решения задач, поставленных нашим исследованием.

Данные, получаемые на основе подсчетов индивидуумов, принадлежащих к той или иной группе, как правило, подвергают статистической обработке, отличной от общепринятой для результатов измерений. Рассмотрим, к примеру, исследование по выявлению некоторой закономерности в отношении показателей качества яиц, откладываемых всеми несушками стада. Мы можем взвесить каждое яйцо из выборки и установить среднюю или усредненную массу яйца, равную, скажем, 21 г. Мы можем также классифицировать каждое яйцо как надтреснутое или целое и найти, что 5% яиц были надтреснутыми. Было бы бессмысленным утверждать, что в среднем яйца имели 5%-ную надтреснутость. В данном случае средняя относится к доле экземпляров выборки, имеющих это повреждение.

В главе о преобразованиях уравнений мы показали, каким образом данные подсчетов могут быть иногда преобразованы и обработаны обоснованным статистическим методом, как если бы они являлись количественными измерениями. В этой главе мы опишем метод, называемый хи-квадратом (обозначаемый символом χ^2) и предназначенный для анализа данных учетов.

Прежде чем обсуждать этот метод, мы должны рассмотреть, что нам дает классификация и подсчет индивидуумов. Сбор таких данных

обычно предназначается для решения одной или большего числа из следующих трех задач: 1) проверить одну или большее количество гипотез при внешнем отсутствии предполагаемых закономерностей в данных учетов; 2) установить, имеется ли корреляция между различными показателями и 3) проверить, взяты ли выборки из разных совокупностей.

ХИ-КВАДРАТ

Общая формула хи-квадрата, используемая для решения всех указанных проблем, имеет следующий вид:

$$\chi^2 = \sum \frac{(Ob - Ex)^2}{Ex},$$

где Ob — фактически наблюдаемая величина для каждого из двух или большего числа классов и Ex — соответствующая ожидаемая величина.

Для разъяснения этого выражения мы должны вначале определить ожидаемую величину для каждого класса индивидуумов в соответствии с нашими гипотезами. Далее ожидаемая величина вычитается из фактической величины, получаемая разность возводится в квадрат и делится на ожидаемую величину. Итоговые частные суммируют по всем классам. Далее сумма сравнивается с табличной величиной χ^2 при соответствующем числе степеней свободы.

Это сравнение дает нам приблизительную вероятность случайного получения отклонений от ожидаемых величин, равных фактически наблюдаемым или превышающим их.

Арифметические расчеты этого анализа довольно просты, кроме того, для некоторых особых случаев имеются сокращенные методы с использованием калькуляторов. Однако несколько моментов мы должны принять во внимание с целью надлежащего использования метода хи-квадрата.

1. Необходимо соблюдать осторожность при выборе рабочей гипотезы. Эта гипотеза должна быть обоснованной, базирующейся на ранее установленных фактах и принципах.

2. Следует иметь в виду тот факт, что распределение хи-квадрата является непрерывным распределением и фактически подобно нормальному распределению. С другой стороны, распределение выборок, основанных на подсчетах, является дискретным, или прерывным распределением. Если индивидуумы распределяют в один из двух классов, то мы имеем дело с так называемым биномиальным распределением. Нормальное и биномиальное распределение подобны, но не идентичны. Вот почему, как указывалось выше, таблица значений хи-квадрата дает приблизительную вероятность. Нам необходимо знать, какие ситуации приводят к слабым приближениям, так что мы можем или избежать эти ситуации, или по возможности внести коррективы, обеспечивающие более тесное приближение к истинному событию.

3. На основании заданной гипотезы надо знать, как правильно вычислить ожидаемые величины для каждого класса.

4. Число степеней свободы для входа в таблицу хи-квадратов всегда является очевидным. С этой целью необходимо ознакомиться с некоторыми правилами.

5. Интерпретацию данных, полученных методом хи-квадрата следует проводить с глубоким анализом всех факторов. Даже если фактические наблюдения не отличаются существенно от предположения — принятие гипотезы может быть неоправданным даже при соответствии полученных данных другим одинаково логичным гипотезам.

Давайте проиллюстрируем эти различные положения на примере. Предположим, что мы работаем с некоторым растением, имеющим красные и белые соцветия. Мы скрестили растения чистых линий этих двух видов соцветий и в поколении F_1 получили все соцветия красными. В поколении F_2 из восьми растений четыре имели красные и четыре белые соцветия. На основании всего, что нам уже известно, мы совершенно уверены, что красный цвет доминирует над белым, что как мы предполагаем далее, определяется единственным геном. Наши познания в генетике ведут к принятию гипотезы, которая гласит, что в поколении F_2 произойдет расщепление на красные и белые соцветия в соотношении 3 : 1 соответственно.

На основании этой гипотезы следует ожидать, что из восьми растений будет получено шесть с красными и два с белыми соцветиями. Следовательно, фактически наблюдаемое число отклоняется от ожидаемого на два. Мы спрашиваем: «Какова вероятность случайного получения отклонения от ожидаемого числа, равного фактически наблюдаемому или превышающему его?» Если эта вероятность будет очень мала, то мы отвергнем нашу гипотезу.

Признав, что хи-квадрат дает нам только приближенное значение желаемой вероятности, мы вычислим точную вероятность, основанную на биномиальном распределении. Для этого необходимо найти вероятность каждого возможного исхода и сложить вероятности для всех случаев равенства или превышения фактического отклонения при сопоставлении с ожидаемой величиной.

Вначале следует установить некоторые символы. Мы назовем гипотетическое отношение как $r_1 : r_2$. Вероятность принадлежности отдельного индивидуума к первому классу называется p и равняется $r_1/(r_1+r_2)$. Вероятность нахождения индивидуума во втором классе обозначается буквой q и равняется $r_2/(r_1+r_2)$ или $1-p$. Фактическую численность каждого класса называют n_1 и n_2 и $n_1+n_2=n$ или общим числом выборки. Символ « $n!$ » называют n -факториалом и получают путем перемножения всех целых чисел от 1 до n . Факториал нуля равен 1.

При биномиальном распределении вероятность получения выборки с численностью n_1 первого и n_2 второго класса определяют из формулы

$$\frac{p^{n_1} q^{n_2} n!}{n_1! n_2!}.$$

Таблица 17.1

Исходы (фактические соотношения)	Вероятность	Отклонение n_1 от ожида- емой величины (n_1^{-6})	Классовый интервал	Вероятность, основанная на нормаль- ной кривой
8:0	0,1001	2	$> 1,5$	0,1104
7:1	0,2670	1	От 0,5 до 1,5	0,2312
6:2	0,3115	0	От -0,5 до 0,5	0,3168
5:3	0,2076	-1	От -0,5 до -1,5	0,2312
4:4	0,0865	-2	От -1,5 до -2,5	0,0897
2:6	0,0038	-4	От -3,5 до -4,5	0,0020
1:7	0,0004	-5	От -4,5 до -5,5	0,0001
0:8	0,0000	-6	$< -5,5$	0,0000
Сумма	1,0000			

В нашем примере $r_1=3$, $r_2=1$, $p=r_1/(r_1+r_2)=3/4$, $q=r_2/(r_1+r_2)=1/4$. Вероятность получения выборки, в которой $n_1=4$ и $n_2=4$, равна

$$\frac{p^{n_1} q^{n_2} n!}{n_1! n_2!} = \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^4 \cdot \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 1 \cdot 2 \cdot 3 \cdot 4} = \frac{81}{256} \times \frac{1}{256} \times 70 = 0,0865.$$

Подобным образом мы можем вычислить вероятность любого из всех других исходов и на этом основании заполнить первые три колонки таблицы 17.1.

Последняя вероятность в действительности не равна нулю, но меньше, чем 0,00005.

Отметьте, что сумма всех вероятностей дает 1, которая и является показателем правильности проведенных расчетов.

Ожидаемая величина n_1 равна $np=8 \times 3/4=6$, что позволяет определить разность между наблюдаемыми величинами n_1 и ее ожидаемой величиной и заполнить третью колонку таблицы.

Мы можем теперь ответить на исходные вопросы. Вероятность получения отклонения от ожидаемой величины, равной двум или большему числу, равняется сумме вероятностей первого и последних пяти из девяти исходов таблицы, а именно $0,1001+0,0865+\dots+0,0000=0,2139$.

Посмотрим, как оценивают этот результат с помощью метода хи-квадрата. Используем известную формулу

$$\chi^2 = \sum \frac{(Ob - Ex)^2}{Ex} = \frac{(4-6)^2}{6} + \frac{(4-2)^2}{2} = \frac{4}{6} + \frac{4}{2} = 0,67 + 2 = 2,67.$$

Отыскивая это значение в таблице хи-квадратов (табл. А.6, приложение) при одной степени свободы мы находим, что наш фактический хи-квадрат очень близок к величине 2,706, находящейся в колонке под 10%, поэтому вероятность случайного получения отклонения, по крайней мере равного фактически наблюдаемому, равна 0,10. (Точное зна-

чение вероятности, равное 0,1025, дают более полные таблицы.) Это значительно меньше рассчитанной нами величины действительной вероятности, равной 0,2139.

Поправка Йейтса на непрерывность. В математической статистике имеется так называемая поправка Йейтса на непрерывность, которая значительно снижает расхождение между двумя методами. Предположим, что мы использовали нормальное распределение для получения оценочной вероятности каждого исхода. Для этого мы должны вначале найти дисперсию и стандартное отклонение распределения. Это достигается возведением отклонения от средней (ожидаемое значение) в квадрат для каждого фактического исхода и последующим умножением на соответствующую вероятность. Эти произведения суммируются для всех наблюдаемых исходов.

Дисперсия $= 2^2 \times 0,1001 + 1^2 \times 0,2670 + \dots + 5^2 \times 0,0004 = 1,4997$. Поскольку мы имеем дело с биномиальным распределением, то воспользуемся простой формулой для дисперсии $\sigma^2 = npq$. Таким образом, для этого примера получим

$$\sigma^2 = 8 \times 3/4 \times 1/4 = 1,5.$$

Стандартное отклонение $\sigma = \sqrt{\sigma^2} = \sqrt{1,5} = 1,225$.

Классовые интервалы могут быть теперь выражены в значениях величины z путем деления пределов каждого интервала на стандартное отклонение. Далее площадь под нормальной кривой для каждого интервала может быть найдена из таблиц вероятностных функций, которые приводятся в большинстве сборников математических таблиц.

Эти оценки приводятся в таблице 17.1 для того, чтобы показать, как различаются нормальное и биномиальное распределения. Поскольку нормальное распределение является непрерывным, мы должны объединить всю часть нормальной кривой от $n_1 - Ex = -1,5$ до $n_1 - Ex = (-2,5)$ и определить площадь этой части для того, чтобы найти вероятность равенства $n_1 - Ex = (-2)$. Подобно этому вероятность равенства $n_1 - Ex = 2$ есть часть площади под нормальной кривой от $n_1 - Ex = 1,5$ до бесконечности. Таким образом, наш вопрос в отношении вероятности получения отклонения от ожидаемого результата, равного 2 или большему числу, для случая использования нормального распределения должен быть поставлен иначе: «Какова вероятность того, что отклонение от ожидаемого для этой величины превысит 1,5?» Поправка Йейтса как раз учитывает это и заключается в простом вычитании 0,5 из абсолютной величины (независимо от знака) разности между фактическим и ожидаемым результатами.

Используя эту поправку, мы вычислим скорректированный хи-квадрат по уравнению

$$\begin{aligned} \chi^2 &= \sum \frac{(|OB - Ex| - 0,5)^2}{Ex} = \frac{(2 - 0,5)^2}{6} + \frac{(2 - 0,5)^2}{2} = \\ &= \frac{1,5^2}{6} + \frac{1,5^2}{2} = 0,375 + 1,125 = 1,50. \end{aligned}$$

(Отметьте: символ $|x|$ означает абсолютную величину x .)

Сопоставление этой величины хи-квадрата с табличными данными (табл. А.6 в приложении) показывает, что искомая вероятность находится в пределах от 0,1 до 0,5, но она значительно превышает ранее установленную величину. Более полные таблицы дают величину P , равную 0,2207, очень близкую к вычисленной вероятности, равной 0,2139. Вероятность, основанная на нормальном распределении, может быть получена таким же путем, как и для биномиального распределения, в результате сложения вероятности первой и пяти последних строк из таблицы 17.1. Это дает величину 0,2208, которая, как и следовало ожидать, равна (без учета погрешностей округления) результату, полученному с помощью метода хи-квадрат.

РЕКОМЕНДАЦИИ К ИСПОЛЬЗОВАНИЮ МЕТОДА ХИ-КВАДРАТ

Мы убедились, что даже для малых выборок (с объемом, равным восьми) различие между нормальным распределением, на котором основан хи-квадрат, и точным биномиальным распределением не очень велико. Следующие правила помогут решить, дает ли хи-квадрат достаточно близкое приближение к правильному ответу.

1. Чем больше объем выборки, тем ближе соответствие между двумя распределениями.

2. Чем больше отношение между r_1 и r_2 в наших предположениях, тем значительнее будет расхождение между двумя распределениями для заданного объема выборки. Так, если мы предполагаем отношение 1 : 1, то соответствие будет близким даже для малых выборок, но если предполагаемое отношение равно 15 : 1, то необходимо брать значительно больший объем выборки.

3. Хороший практический навык подсказывает, когда не следует использовать хи-квадрат, если наименьший ожидаемый класс численности составляет не более пяти. Если имеется более двух классов, то можно объединить классы, ожидаемые численности которых составят меньше пяти. Увеличение объема выборки можно также использовать для повышения размера ожидаемых самых наименьших классов.

4. Всегда используйте поправку Йейтса для расчета хи-квадрата, имеющего единственную степень свободы. Никогда не используйте ее для решения проблем, имеющих более чем 1 степень свободы.

В общем виде степени свободы — это число классов, которые могут принимать произвольную численность. Так, если имеется два класса, как в нашем примере, то n_1 может принимать любое значение, тогда как n_2 — фиксированная величина, потому что этот класс будут составлять оставшиеся члены (единицы) выборки, поскольку $n_2 = n - n_1$. Хи-квадрат будет опираться на одну степень свободы. При проверке любой, чуждой по отношению к нашим данным гипотезы число степеней свободы будет всегда на единицу меньше, чем число классов. Другие случаи будут рассмотрены позднее.

ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ

Интерпретация — это последний и наиболее важный этап нашего анализа данных. Мы убедились, что расхождение между наблюдаемым и ожидаемым результатом может быть легко отнесено на счет только случайного фактора. Поэтому у нас нет доводов для отвержения нашей гипотезы. Значит ли это, что мы имеем сильные доводы в пользу нашей гипотезы? Не обязательно, и как раз это положение часто понимают неправильно. Давайте представим это следующим образом. Мы можем выдвинуть много других гипотез, и наша выборка не будет представлять существенное отклонение для всех из них. Если мы имеем весомер доказательство того, что красный и белый цвет определяются единственной парой генов, тогда предполагаемое отношение 3 : 1 будет наиболее обоснованной гипотезой, и наша выборка может быть охарактеризована как обеспечившая хорошее подтверждающее доказательство. С другой стороны, доводы для утверждения о единственной паре генов можно рассматривать как очень слабые. Далее мы должны рассмотреть возможности появления под воздействием двух пар генов отношений 9 : 7 или 13 : 3. Пример нашей выборки с отношением 4 красных к 4 белым не противоречит появлению любого из этих отношений. Для выявления различий между другими возможными гипотезами необходимы дополнительные критерии или значительно большие объемы выборок в поколении F_2 .

В таблице 17.2 даны размеры выборок, необходимые для обоснования различных, часто встречающихся отношений. Например, объем выборки в 105 единиц, показанный в таблице, необходим для обосно-

Таблица 17.2. Необходимый объем выборки для отвержения по крайней мере одной из двух альтернативных гипотез (в числителе 5%-ный, в знаменателе 1%-ный уровень)

Отноше- ния	15:1	7:1	13:3	3:1	11:5	5:3	9:7
1:1	$\frac{16}{24}$	$\frac{24}{38}$	$\frac{38}{61}$	$\frac{62}{101}$	$\frac{112}{186}$	$\frac{254}{428}$	$\frac{1008}{1718}$
9:7	$\frac{20}{31}$	$\frac{33}{53}$	$\frac{56}{92}$	$\frac{105}{174}$	$\frac{243}{407}$	$\frac{977}{1664}$	
5:3	$\frac{27}{42}$	$\frac{49}{79}$	$\frac{94}{155}$	$\frac{223}{374}$	$\frac{915}{1558}$		
11:5	$\frac{39}{61}$	$\frac{80}{130}$	$\frac{195}{326}$	$\frac{823}{1398}$			
3:1	$\frac{60}{97}$	$\frac{159}{264}$	$\frac{699}{1184}$				
13:3	$\frac{114}{186}$	$\frac{543}{915}$					
7:1	$\frac{354}{589}$						

вания того, что либо отношение $3 : 1$, либо отношение $9 : 7$ будет отвергнуто на 5%-ном уровне. Табличная величина хи-квадрата, требуемая для отвержения гипотезы, равна 3,84. Если мы получим отношение $70 : 35$, то фактическая величина хи-квадрата, равная 3,46, была бы недостаточно велика для отвержения гипотезы $3 : 1$. Проверяя гипотезу $9 : 7$, мы получаем величину хи-квадрата, равную 4,22, достаточную для отвержения гипотезы на 5%-ном уровне. С другой стороны, наблюдаемое отношение $69 : 36$ имело бы значение хи-квадрата, равное 4,34 и 3,45 для $3 : 1$ и $9 : 7$ гипотез соответственно. На основании этого мы отвергли бы гипотезу $3 : 1$. Проверка этих величин хи-квадрата не осуществлялась и предлагается в качестве упражнения для читателей. Не забудьте использовать поправку на непрерывность.

ПРОВЕРКА НЕЗАВИСИМОСТИ

В отношении данных, полученных путем подсчетов, нас часто интересует: связаны ли между собой две переменные. Например, первой переменной, используемой для классификации индивидуумов, может быть уровень образования, а второй — уровень доходов. При желании мы можем проверить взаимосвязь между уровнем образования и доходами.

Мы можем произвольно выбрать два уровня одной переменной, такой, как прививка двух групп животных, инокулируя одну группу и оставляя другую без прививки. Далее, после некоторого периода времени можно выделить в каждой группе здоровых и больных животных и проверить наличие какой-либо связи между прививкой и заболеваниями. В генетических исследованиях часто возникает желание выявить, наследуются ли два признака независимо или проявляется наличие связей между ними. Решение всех этих проблем аналогично корреляционному анализу измеряемых (количественных) данных.

При анализе данных на проявление зависимости между двумя переменными наиболее удобной нулевой гипотезой будет предположение об их независимости. Если отклонение от независимости значительно превышает ее ожидаемый предел, обусловленный случайным фактором, мы отвергаем гипотезу о независимости двух переменных и принимаем альтернативную гипотезу, предлагающую их взаимозависимость.

Для того чтобы найти ожидаемую величину и применить формулу хи-квадрата, мы используем принцип теории вероятностей, который гласит: «Если два события независимы друг от друга, то вероятность одновременного наступления двух событий будет равна произведению вероятностей их отдельного наступления». Давайте проиллюстрируем этот принцип на примере и покажем, как осуществляется проверка по критерию хи-квадрата. Сто животных были подвергнуты инъекции антибиотика и после некоторого периода времени осмотрены на проявление симптомов заболевания. При проверке 88 животных оказались здоровыми, а 12 проявили симптомы заболевания. Другая группа в 200 животных не подвергалась инъекции антибиотика, и аналогичный осмотр через некоторое время позволил выявить 143 здоровых

Таблица 17.3. Заболеваемость животных (гол.) в зависимости от инъекции антибиотика (в скобках ожидаемое число)

Варианты	Здоровые	Больные	Сумма
Инъекция	88' (77)	12 (23)	100
Без инъекции	143 (154)	57 (46)	200
Суммы	231	69	300

и 57 больных животных. Эти результаты могут быть сведены в виде так называемой таблицы сопряженности 2×2 (табл. 17.3).

Мы проверим гипотезу об отсутствии связи между применением антибиотика и заболеванием. Если эти две переменные независимы, то ожидаемая доля здоровых животных из всех подвергавшихся инъекции составит произведение фактической доли здоровых на долю инъецированных. То есть $231/300 \times 100/300 = 77/300$. Поскольку в эксперименте было 300 животных, то $77/300 \times 300 = 77$ будет ожидаемое число здоровых животных из получивших антибиотик. Расчеты могут быть значительно сокращены, если принять во внимание, что общая сумма выступает знаменателем в обеих долях, которые затем перемножаются для получения объединенной вероятности. Последняя доля была затем умножена на общую сумму и получено ожидаемое число. Мы можем использовать общую сумму в наших расчетах один раз и найти ожидаемое число следующим путем: $(100 \times 231)/300 = 77$. Словами это можно выразить так: ожидаемое число здоровых животных из всех подвергавшихся инъекции равно произведению общего количества инъецированных на общее количество здоровых, деленному на общую сумму экспериментальных животных. Подобным образом может быть вычислен каждый из других ожидаемых классов. Фактически в таблице 2×2 требуется вычислить только одну ожидаемую величину. Поскольку мы ожидаем получить 77 здоровых животных из всех подвергавшихся инъекции, то оставшиеся 23 будем ожидать больными. По аналогии мы ожидаем получить 77 здоровых животных в обработанной группе и 154 из 231 здоровых животных в необработанной (контрольной) группе. Тогда из 200 контрольных животных при ожидании получить 154 головы здоровыми мы ожидаем, что оставшиеся 46 животных будут больными. Обратите внимание, что после установления численности одной из групп (классов) оставшиеся три класса становятся фиксированными. Таким образом, мы имеем лишь одну степень свободы для таблицы 2×2 . Общее правило для $r \times c$ (где r — ряды и c — колонки) таблицы сопряженности заключается в том, что степени свободы равны произведению $(r-1) \times (c-1)$.

Необходимо подчеркнуть одну особенность таблицы 2×2 : разность между наблюдаемыми и ожидаемыми величинами остается одной и той же для каждой клетки таблицы с тем лишь исключением, что две разности имеют положительный, а две другие отрицательный знак. Эта

единая разность в нашем примере равна 11 (например, $88-77=11$), и поскольку мы имеем дело с 1 степенью свободы, необходимо применить поправку Йейтса и рассматривать разность, равную 10,5.

Применив нашу формулу хи-квадрата, мы получим

$$\chi^2 = \sum \frac{(|Ob - Ex| - 0,5)^2}{Ex} = \frac{(10,5)^2}{77} + \frac{(10,5)^2}{23} + \frac{(10,5)^2}{154} + \frac{(10,5)^2}{46} = 9,34.$$

Обратившись к таблице хи-квадратов (табл. А.6, приложение), мы увидим, что при 1 степени свободы вероятность случайного получения хи-квадрата, равного 6,636, следует ожидать равной 1%, а равного 10,827, — 0,1%. Поэтому можно утверждать, что вероятность получения хи-квадрата, равного 9,34, лишь несколько выше, чем 1 из 1000; в итоге мы отвергаем гипотезу о независимости и делаем заключение о наличии связи между инъекцией антибиотика и заболеванием.

Для того чтобы показать использование хи-квадрата с целью проверки независимости между двумя парами генов, мы проанализируем некоторые данные многочисленного потомства календулы, разделенного по двум факторам: по скороспелости и интенсивности зеленой окраски (легкая хлорофилловая недостаточность). Известно, что скороспелость рецессивна по отношению к позднему созреванию и определяется в этом генетическом материале единственной парой генов. Хлорофилловая недостаточность рецессивна по отношению к нормальному состоянию и также контролируется единственной парой генов. Необходимо ответить на три вопроса. Подходит ли отношение 3 : 1 к соотношению поздне- и раннеспелых форм? Подходит ли отношение 3 : 1 к соотношению нормальных и светло-зеленых растений? Наследуются ли две пары признаков независимо или проявляется связь между ними? Организованное размещение данных в таблице сопряженности выглядит следующим образом (табл. 17.4).

Для ответа на первый вопрос, касающийся соотношения поздние : ранние, мы вычисляем хи-квадрат

$$\begin{aligned} \chi^2 &= \frac{(|4380 - 4275| - 0,5)^2}{4275} + \frac{(|1320 - 1425| - 0,5)^2}{1425} = \\ &= \frac{(104,5)^2}{4275} + \frac{(104,5)^2}{1425} = 10,22. \end{aligned}$$

Таблица 17.4. Результаты деления потомства календулы по двум признакам (в скобках дано ожидаемое число)

Сорт	Нормальные	Светло-зеленые	Суммы	Ожидаемое отношение 3 : 1
Позднеспелый	3470 (3457,9)	910 (922,1)	4380	4275
Раннеспелый	1030 (1042,1)	290 (277,9)	1320	1425
Итого	4500	1200	5700	
Ожидаемое отношение 3 : 1	4275	1425		

Полученная величина хи-квадрата почти равна требуемой (10,827) на 0,1%-ном уровне. Это означает, что в случае правильности предполагаемого отношения 3 : 1 вероятность получения отклонения, равного наблюдаемому, составила бы лишь 1 из 1000. Поэтому мы отвергаем гипотезу о том, что соотношение 3 : 1 является правильным. В действительности гипотеза о том, что позднее цветение просто доминирует над ранним цветением, не была отвергнута, потому что отмечалась некоторая ослабленность раннеспелых форм (как и для случая многих других рецессивных признаков) в сравнении с позднеспелыми. Небольшое, но существенное отклонение от отношения 3 : 1 было поэтому отнесено на счет различных уровней выживаемости (скорости развития). Следует отметить, что в данном случае было необычно большое потомство. Если бы оно составляло лишь одну десятую часть от фактической величины (570 растений) и соотношение поздние : ранние оставалось неизменным, то значение хи-квадрата составило бы только 0,94, что не достигает существенного уровня.

Вопрос о соотношении нормальные : светло-зеленые объясняют подобным образом; значение хи-квадрата, равное в данном случае 47,16, опять оценивается как очень существенное. Светло-зеленые растения, испытывающие дефицит хлорофилла, имеют еще более низкую скорость развития в сравнении с нормальными, чем раннеспелые в сравнении с позднеспелыми.

При оценке независимости мы скорее принимаем наблюдаемые отношения, чем предполагаемое отношение 3 : 1, и вычисляем ожидаемые величины на основе предположения о независимости. Таким образом, ожидаемое число позднеспелых нормальных растений будет равно

$$\frac{\text{Сумма нормальных} \times \text{сумма позднеспелых}}{\text{Общая сумма}} = \frac{4500 \times 4389}{5700} = 3457,9.$$

Ожидаемые величины для оставшихся трех клеток таблицы 17.4 можно вычислить подобным путем или путем вычитания из краевых сумм. Использование обоих способов обеспечивает проверку аккуратности расчетов. Отметим, что ($Ob-Ex$) равно 12,1 для верхней левой и нижней правой клеток таблицы и ($-12,1$) для двух других клеток. Числители слагаемых выражений для определения хи-квадрата будут одинаковыми для каждого класса. Применяв поправку Йейтса для каждой клетки таблицы, получаем $(12,1 - 0,5)^2 = (11,6)^2$. Поэтому хи-квадрат будет равен

$$\frac{11,6^2}{3456,9} + \frac{11,6^2}{922,1} + \frac{11,6^2}{1042,1} + \frac{11,6^2}{277,9} = 0,80.$$

Вероятность случайного получения такой величины находится между 10 и 50%, так что мы не имеем каких-либо доказательств, оправдывающих отвержение гипотезы о независимости.

Другой пример покажет вычисления хи-квадрата для тех случаев, когда он опирается более чем на 1 степень свободы, и каким образом может «свертываться» таблица сопряженности. Три группы (по 39 телат крупного рогатого скота в каждой) при откорме получали неоди-

Таблица 17.5. Заболеваемость телят, откармливаемых по трем рационам (ожидаемые величины — в скобках)

Число заболева- ний	Рационы			Суммы
	1	2	3	
0	9 (17,3)	16 (17,3)	17 (16,3)	52
1	1 (0,3)	0 (0,3)	0 (0,3)	1
2	0 (1,3)	3 (1,3)	1 (1,3)	4
3	7 (5,7)	9 (5,7)	1 (5,7)	17
4	3 (4,7)	5 (4,7)	6 (4,7)	14
5	4 (3,3)	1 (3,3)	5 (3,3)	10
6	2 (2,0)	1 (2,0)	3 (2,0)	6
7	0 (1,3)	2 (1,3)	2 (1,3)	4
8	1 (2,3)	2 (2,3)	4 (2,3)	7
10	2 (0,7)	0 (0,7)	0 (0,7)	2
Итого	39	39	39	117

наковые рационы. Состояние здоровья каждого животного контролировалось путем регистрации числа заболеваний. Полученные результаты приведены в таблице 17.5.

В данном случае ожидаемые величины вычисляются очень легко, поскольку точно треть животных соответствовала каждому классу рационов. Это означает, что заболевание можно ожидать у каждой трети животных, откармливаемых на рационах из каждого класса, если рационы и заболевания являются независимыми друг от друга. Мы замечаем, что многие ожидаемые величины меньше пяти, так что фактически применение формулы хи-квадрата для исходных данных будет необоснованным. Однако мы проделаем вычисления и посмотрим, каким образом сравнивают результаты с таковыми, получаемыми из свернутой таблицы.

$$\chi^2 = \sum \frac{(Ob - Ex)^2}{Ex} = \frac{(19 - 17,3)^2}{17,3} + \frac{(16 - 17,3)^2}{17,3} + \dots + \frac{(0 - 0,7)^2}{0,7} = 24,5.$$

Степени свободы = $(r-1)(c-1) = (10-1)(3-1) = 18$.

Сопоставляя наш вычисленный хи-квадрат, равный 24,5 (таблица А.6, приложение), против 18 степеней свободы, находим, что вероятность случайного получения наблюдаемых результатов составляет несколько выше 10%. И все-таки у нас нет достаточных оснований для отвержения гипотезы об отсутствии связи между состоянием здоровья животных и видом рациона.

В целях обеспечения требований правила о том, что численность каждого ожидаемого (нерасчетного) класса не должна быть ниже 5, мы можем свернуть таблицу путем объединения частоты классов: 1, 2 и 3; 4 и 5; 6, 7, 8 и 10. Это дает новую таблицу (табл. 17.6).

Расчеты хи-квадрата дадут нам величину 10,61, которую мы находим в таблице приложений против 6-ти степеней свободы. Мы определяем, что она почти точно равняется табличной величине при 10%

Таблица 17.6. Свернутый вид таблицы 17.5

Классы числа заболеваний	Рационы			Суммы
	1	2	3	
0	19 (17,3)	16 (17,3)	17 (17,3)	52
1—3	8 (7,3)	12 (7,3)	2 (7,3)	22
4—5	7 (8,0)	6 (8,0)	11 (8,0)	24
6—10	5 (6,3)	5 (6,3)	9 (6,3)	19
Итого	39	39	39	117

вероятности. Поэтому наши выводы будут аналогичными полученным из исходной таблицы, хотя подобное встречается не во всех случаях. Свертывание таблицы — всегда надежный способ для исключения в расчетах слишком небольших ожидаемых классов. Более того, это сокращает расчеты хи-квадрата. Отметим, что в этом примере не использовалась поправка на непрерывность, потому что мы имели дело с более чем одной степенью свободы.

НЕОДНОРОДНОСТЬ

Третье и последнее использование хи-квадрата будет рассмотрено для проверки гипотезы о принадлежности группы выборок к одной совокупности. Рассмотрим потомство календулы от восьми растений при расчленении каждого на нормальные и светло-зеленые формы, как показано в таблице 17.7.

Мы выполним два вида анализов. Во-первых, проверим каждое потомство и итоговые данные со всех потомств на отклонение от гипотетического отношения 3 : 1.

Таблица 17.7. Нормальные и светло-зеленые экземпляры потомства восьми растений календулы

Потомство	Нормальные	Светло-зеленые	χ^2 (3:1)	χ^2 (3106:854)
1	315	85	3,00	0,023
2	602	170	3,65	0,094
3	868	252	3,73	0,578
4	174	42	3,56	0,575
5	192	48	3,20	0,348
6	165	39	3,76	0,723
7	161	43	1,67	0,028
8	629	175	4,48	0,019
Итого			27,05	2,388
Сумма	3106	854	24,91	0,000
Неоднородность			2,14	2,388

Хи-квадраты, вычисленные для каждого потомства, показаны в четвертой колонке. Эти величины были рассчитаны без учета поправки на непрерывность, потому что мы хотим сложить их, а слагаемы лишь неприведенные хи-квадраты. Отметим, что только один из них превышает величину 3,84 требуемого порога существенности для 5%-ного уровня. Поэтому индивидуальные потомства дают нам очень мало доводов для отвержения нашей гипотезы. Однако подобное заключение все еще не обосновано, поскольку семь из восьми потомств дали хорошее соответствие (т. е. несущественно отклоняются от отношения 3 : 1); в опыте получено более чем достаточно доказательств верности нашей гипотезы. Поэтому мы должны продолжить анализ. Сложение восьми индивидуальных хи-квадратов, каждый с 1 степенью свободы, дает итоговый (общий) хи-квадрат, равный 27,05 при 8 степенях свободы. Эта величина превышает табличное значение, равное 26,125, на 0,001%-ном уровне. Другими словами, вероятность того, что такая величина может быть получена в результате случайности, составляет менее 1 случая из 1000. Другую проверку можно сделать по отношению к общим суммам, равным 3106 для нормальных и 854 для светло-зеленых растений. Ожидаемые числа составят $3960 \times 3/4 = 2970$ и $3960 \times 1/4 = 990$, тогда

$$\chi^2 = \frac{(3106 - 2970)^2}{2970} + \frac{(854 - 990)^2}{990} = 24,91.$$

Это значительно превышает табличную величину хи-квадрата для 1 степени свободы на 0,001%-ном уровне, так что теперь мы совершенно определенно отвергаем гипотезу о том, что все потомства являются выборками из совокупности с отношением 3 : 1. Нас еще интересует, могут ли все эти потомства представлять выборки из одной совокупности. Для проверки этой гипотезы мы вычисляем так называемый хи-квадрат неоднородности.

Хи-квадрат неоднородности равен общему хи-квадрату минус хи-квадрат суммы.

Поскольку общий хи-квадрат был равен 27,05 и хи-квадрат суммы 24,91, то хи-квадрат неоднородности будет равен 2,14 при 7 степенях свободы. Как видно из таблицы приложений, этот хи-квадрат даже меньше величины 2,167, требуемой для 0,95%-ного уровня. Вероятность того, что хи-квадрат подобной величины или большей может быть случайно получен на основе группы однородных выборок, составляет около 95%. Все эти оценки можно обобщить в виде таблицы, подобной таблице дисперсионного анализа (табл. 17.8).

Таблица 17.8. Итоговые данные анализа восьми потомств календулы, основанного на отношении 3:1

Источники	df	Хи-квадрат
Общий	8	27,05 ***.
Сумма	1	24,91 ***
Неоднородность	7	2,14 несущественно

Таблица 17.9. Результаты анализа данных по календуле, основанного на фактических суммах

Источник	df	Хи-квадрат
Общий	8	2,388
Сумма	1	0,000
Неоднородность	7	2,388

Вместо проверки каждого потомства на соответствие гипотетическому отношению мы можем оценить наблюдаемое отношение сумм растений. Это сделано в последней колонке таблицы 17.7. Объединенный хи-квадрат безусловно имеет нулевое значение, поскольку наблюдаемым отношением является то отношение, которое мы оцениваем. Таблица 17.9 аналогична вышеприведенной.

Мы все еще не имеем доказательств неоднородности и потому делаем заключение о том, что мы имеем дело с однородной группой потомств и что нашей лучшей оценкой действительного отношения будет отношение 3106 : 854.

Отметим, что расчеты в этой последней проверке были точно такими же, как при проверке независимости. Другими словами, при проверке каждой выборки на соответствие наблюдаемому общему отношению мы получим равенство, в котором хи-квадрат неоднородности будет равен хи-квадрату независимости. Только в случае сопоставления отношений выборок и сумм с гипотетическим отношением возникает необходимость расчленения общего хи-квадрата на два компонента.

В таблице 17.10 показано, как изменился бы статистический анализ, если бы первые четыре потомства имели такое же абсолютное отклонение от отношения 3 : 1, но в обратном направлении.

Таблица 17.10. Гипотетическая группа данных по календуле, отображающая неоднородность

Потомство	Нормальные растения	Светло-зеленые растения	χ^2 (3:1)	χ^2 (2950:1010)
1	285	115	3,00	2,05
2	556	216	3,65	2,49
3	812	308	3,73	2,35
4	150	66	3,56	2,90
5	192	48	3,20	3,82
6	165	39	3,76	4,38
7	161	43	1,67	2,10
8	629	175	4,48	5,92
Итого			27,05	26,01
Сумма	2950	1010	0,54	0,00
Неоднородность			26,51	26,01

Отметим, что данные величины сумм теперь значительно приблизились к отношению 3 : 1, но хи-квадрат неоднородности остался очень существенным. Опять мы отвергаем гипотезу о том, что все потомства являются выборками из совокупности, которая подчиняется отношению 3 нормальные к 1 светло-зеленому. В этом случае отклонение гипотезы опирается на сильное доказательство неоднородности группы выборок, так что сложение данных не оправдало себя.

В течение всего обсуждения материала этой главы мы использовали единственную формулу

$$\chi^2 = \sum \frac{(Ob - Ex)^2}{Ex}.$$

Допускалось лишь небольшое изменение в тех случаях, когда возникла необходимость введения поправки на непрерывность. В особых случаях для обеспечения сокращенных методов расчета применяют многие модификации этой формулы. При необходимости вычисления большого количества хи-квадратов рекомендуем обращаться к более сложным пособиям по статистике, в которых указаны сокращенные формулы. Для читателя, испытывающего лишь периодически необходимость применения метода хи-квадрат, следует ограничиться знакомством с этой основной формулой.

ЗАКЛЮЧЕНИЕ

Общая формула расчета хи-квадрата имеет вид

$$\chi^2 = \sum \frac{(Ob - Ex)^2}{Ex}.$$

Индивидуумы, односторонне сгруппированные в два или большее число классов, могут сравниваться с гипотетическим отношением. Число степеней свободы будет на единицу меньше числа классов.

Сравнивая вычисленный хи-квадрат с табличным, мы можем найти вероятность случайного появления отклонения, по крайней мере, равного наблюдаемому.

Индивидуумы, двухсторонне группируемые на r и c классы, могут подвергаться проверке на независимость между двумя критериями классификации. Число степеней свободы будет равно произведению $(r-1) \times (c-1)$. Если два или большее число образцов по отдельности проверяют, сопоставляя с общим гипотетическим соотношением, то сумма итоговых хи-квадратов может быть расчленена на две компоненты, как показано ниже.

Источник	df
Общий	$r (c-1)$
Сумма	$(c-1)$
Неоднородность	$(r-1) (c-1)$

Число классов, по которым распределяются индивидуумы каждой выборки, равно c , а r представляет число выборок.

ПУТИ ПОВЫШЕНИЯ ТОЧНОСТИ ЭКСПЕРИМЕНТОВ

Точность эксперимента определяется его способностью выявить действительные эффекты вариантов. Обычно чем более точен эксперимент, тем меньшее различие по вариантам он способен выявить. Чем больше вариабельность между экспериментальными единицами одного варианта, тем выше будет ошибка, связанная с оценкой разности между двумя средними, и тем не менее эксперимент будет точен в отношении выявления эффектов вариантов. Стандартная ошибка разности двух средних уменьшается с уменьшением s и увеличением n и $s_{\bar{d}} = \sqrt{2s^2/n}$, где n — число повторностей. Таким образом, сущность методов увеличения точности эксперимента заключается в снижении неконтролируемой вариабельности по делянкам или увеличении эффективного числа повторений.

Точность может быть улучшена посредством: 1) увеличения повторности; 2) тщательного выбора вариантов; 3) совершенствования техники; 4) отбора экспериментального материала; 5) выбора экспериментальной единицы; 6) введения дополнительных приемов и 7) плановой группировки экспериментальных единиц.

УВЕЛИЧЕНИЕ ПОВТОРНОСТИ

Точность эксперимента можно всегда повысить за счет дополнительных повторностей, однако степень улучшения быстро падает по мере увеличения числа повторностей. Например, для того чтобы удвоить точность сравнения двух средних эксперимента с четырьмя повторностями, необходимо повысить повторность до 16. Это следует из влияния числа повторностей (n) на величину наименьшей существенной разности между двумя средними на заданном уровне значимости $HSP = t\sqrt{2s^2/n}$. Вышеприведенное сравнение не отличается совершенной точностью, поскольку с увеличением n t становится несколько меньше, однако оно достаточно верно для использования в качестве правила в практических расчетах.

Обычно в опытах с полевыми и овощными культурами для достижения обоснованной точности требуется от четырех до восьми повторений. Планируя эксперимент, вы должны иметь обоснованную уверенность в том, что он будет способен выявить действительную разность

желаемого размера. Если вероятность того, что вы достигнете поставленной цели с заданным числом повторений, составляет малую величину, а другие обоснованные средства улучшения точности отсутствуют, то мы настоятельно советуем не проводить эксперимент или по крайней мере отложить его до тех пор, когда будет достаточно ресурсов для его проведения и хорошая уверенность в достижении поставленной цели.

Таблица 2.1 в книге Кохрана и Кокса [3] является удобным источником определения числа повторностей, необходимых для выявления конкретной разности. Эта таблица основана на формуле $r \geq 2[(CV)^2/D^2](t_1 + t_2)^2$, где CV — коэффициент вариации [$CV = s(100)/\bar{Y}$]; D — разность, которую вы желаете выявить, выраженная в процентах от средней по опыту; t_1 — табличная величина для конкретного уровня значимости (например, 5%) и числа степеней свободы для экспериментальной ошибки и t_2 — табличная величина для числа степеней свободы ошибки и вероятности $(1-P) \cdot 2$, где P — вероятность получения существенного результата в данном цикле эксперимента. Если $P = 0,80$, тогда $(1-P)2 = 0,40$ — двухсторонняя площадь кривой t -распределения, основанной на степенях свободы для экспериментальной ошибки.

Перед использованием этого уравнения следует определить по приведенной формуле требуемое, по вашему мнению, число повторностей, а затем взять его приблизительное значение (r). На основе этого r снова решите уравнение, а затем возьмите следующее по величине значение r в качестве требуемого числа повторностей.

Например, мы хотим провести эксперимент, включающий шесть вариантов по схеме рендомизированных блоков. Мы желаем иметь 80%-ную вероятность выявления различий между средними величинами с достоверностью, равной 10% от экспериментальной средней на 5%-ном уровне значимости. Другие эксперименты, проведенные с такими же экспериментальными единицами, показывают, что хорошо поставленный опыт должен иметь коэффициент вариации около 5%. Мы полагаем, что будет достаточно шести повторностей. Таким образом, для первого уравнения имеем: $r=6$, число вариантов $= 6 (n=6)$, df ошибки $= (r-1)(n-1)=25$, $t_1=2,060$, $t_2=0,856$ (см. в приложении таблицу А.2) и $r \geq 2(5/10)^2(2,060+0,856)^2=4,25$.

Теперь возьмем $r=5$, тогда df ошибки $= (4-1)(5-1)=20$, $t_1=2,086$ и $t_2=0,860$. Решив уравнение, получим $r \geq 2(5/10)^2(2,086+0,860)^2=4,34$; поэтому мы выбираем 5 как нашу оценку необходимого числа повторностей. После проведения эксперимента мы будем иметь 80%-ную вероятность выявления 10%-ной разности на 5%-ном уровне при пяти повторностях до тех пор, пока коэффициент вариации не превысит его ожидаемую величину.

ВЫБОР ВАРИАНТОВ

Тщательный выбор вариантов — это не только важное условие достижения целей исследования, но и увеличения точности эксперимента. Например, при изучении эффекта гербицида, фунгицида, удобрений

или инсектицида более важно определить реакцию экспериментальных единиц на увеличение доз воздействующего материала, чем решить, существенно или несущественно различается воздействие двух соседних доз. Таким образом, надлежащий ряд доз делает возможным планирование более чувствительных критериев существенности, чем только сравнение соседних средних ранжированного ряда. Как указывалось ранее, дозы с равновеликими интервалами, покрывающие размах ожидаемого отклика, являются наиболее эффективными для установления кривой доза — отзывчивость и упрощения расчетов сумм квадратов и уравнений откликов. Так же, как указывалось в главе 3, факториальный эксперимент, где одновременно оцениваются два или больше видов факторов, может значительно улучшить точность сравнений главного фактора.

СОВЕРШЕНСТВОВАНИЕ ТЕХНИКИ

Неправильная техника ведения эксперимента может увеличить экспериментальную ошибку и исказить эффекты вариантов. Хорошая техника должна обеспечить: 1) однородное воздействие изучаемого фактора (вариантов); 2) подходящую и несмещенную меру оценки действия вариантов; 3) отсутствие грубых ошибок и 4) контроль за внешними факторами воздействия, так чтобы поставить варианты в одинаковые условия для сравнения.

ОТБОР ЭКСПЕРИМЕНТАЛЬНОГО МАТЕРИАЛА

Для определенных исследований желательно иметь тщательно отобранный и однородный материал. При выборе экспериментального материала, однако, должны иметь в виду ту совокупность, в отношении которой надо сделать заключение. Так, для большинства сельскохозяйственных исследований важно использовать такие виды экспериментального материала, которые находят практическое применение в производстве.

ВЫБОР ЭКСПЕРИМЕНТАЛЬНОЙ ЕДИНИЦЫ

Размер и форма делянки воздействуют на точность полевого опыта. Как правило, вариабельность уменьшается с увеличением размера делянки, однако при достижении определенного размера степень увеличения точности стремительно падает по мере дальнейшего увеличения размеров. Для определения урожайности в полевом опыте точность обычно мало возрастает при использовании делянок площадью более 400 м². Для большинства культур учетная площадь делянок в 40—80 м² обеспечивает хорошую точность. Леклерг и др. [5] рассматривают в своей книге многие примеры полевых делянок различных размеров и формы для различных культур и приводят обширный и заслуживающий внимания список источников. Четырехугольные делянки с длинной стороной, направленной вдоль наибольшего изменения почвенного плодородия, отличаются наибольшей эффективностью преодо-

ления почвенной неоднородности. Увеличение числа животных или деревьев, составляющих экспериментальную единицу, также повышает точность эксперимента. Однако при возможности индивидуального воздействия вариантов на животных или деревья увеличение точности будет больше в случае использования в качестве экспериментальных единиц отдельных индивидуумов и повышенной повторности вариантов, чем в случае использования такого же общего числа животных или деревьев и составления экспериментальных единиц из нескольких индивидуумов.

ВВЕДЕНИЕ ДОПОЛНИТЕЛЬНЫХ ПРИЕМОВ — КОВАРИАЦИЯ

Один из приемов снижения ошибки в эксперименте — это устранение вариабельности Y , связанной с некоторой независимой переменной X . Этот прием называется ковариацией.

Предположим, что в полевом опыте имелась значительная вариация густоты стеблестоя культуры по делянкам. Если сделать обоснованную оценку поделаночных урожаев, которые были бы получены в случае одинаковой густоты стеблестоя, то мы можем улучшить точность измерения эффектов вариантов. Оценка, основанная на предположении прямой пропорциональности урожаев и густоты стеблестоя, не будет обоснованной для данного примера, поскольку она почти всегда дает завышенную урожайность делянок с меньшей густотой стеблестоя.

Другой пример полезного применения ковариационного анализа относится к экспериментам по откорму животных, где начальные массы тела животных различаются. Если установлена зависимость между приростом массы тела и исходной массой, то можно внести коррективы, повышающие точность оценки эффектов вариантов.

В целом ковариация представляет довольно сложный статистический метод как в отношении необходимых вычислений, так и интерпретации результатов. Во многих источниках списка литературы к данной главе ковариация рассматривается детально. На основании

Таблица 18.1. Гипотетические даты, представляющие значения двух переменных X и Y в опыте с четырьмя повторностями и пятью вариантами, размещенными по схеме рендомизированных блоков

Блоки	X					Y				
	1	2	3	4	суммы	1	2	3	4	суммы
Варианты:										
1	8	6	7	7	28	7	5	6	6	26
2	8	4	12	12	36	9	5	9	9	32
3	4	10	10	8	32	6	12	10	12	40
4	1	7	4	12	24	9	11	10	18	48
5	9	8	12	11	40	4	7	15	20	56
Суммы	30	35	45	50	160	45	40	50	65	200

Таблица 18.2. Раздельный дисперсионный анализ данных X и Y из таблицы 18.1

Источник вариации	df	SSX	MSX	F	SSY	MSY	F
Общий	19	186			334		
Блоки	3	50	16,67		70	23,33	
Варианты	4	40	10,00	1,25	160	40,00	4,62 *
Ошибка	12	96	8,00		104	8,67	

нашего опыта мы знаем, что лишь немногие сельскохозяйственные исследователи отчасти используют ковариационный анализ, так что подробный обзор техники анализа вряд ли будет целесообразным. Поэтому мы опишем лишь в общем виде технику ковариационного анализа и некоторые простые аспекты интерпретации.

Таблица 18.1 составлена из некоторых гипотетических дат, обеспечивающих легкие расчеты с целью иллюстрации техники ковариационного анализа. Вы можете считать, что X и Y представляют густоту стеблестоя и урожайность, исходную массу тела и прирост массы или любую другую пару переменных, с которыми вы можете иметь дело.

Дисперсионный анализ значений обоих переменных X и Y может выполняться обычным путем; конечные результаты показаны в таблице 18.2.

Мы отмечаем, что варианты не оказывают существенного влияния на переменную X , тогда как их влияние на Y существенно на 5%-ном уровне.

Для ковариационного анализа необходимо в дополнение к сумме квадратов X и Y вычислить суммы их произведений, которые мы будем обозначать символом SXY .

Во-первых, необходимо определить корректирующее выражение

$$C = \frac{(\Sigma X)(\Sigma Y)}{rn} = \frac{(160)(200)}{20} = 1600.$$

Сумма произведений для блоков равна

$$SXYB = \frac{\Sigma T_{bx}T_{by}}{n} - C = \frac{(30)(45) + \dots + (50)(65)}{5} - 1600 = 50.$$

Для вариантов

$$SXYT = \frac{\Sigma t_{tx}t_{ty}}{b} - C = \frac{(28)(24) + \dots + (40)(56)}{4} - 1600 = 24.$$

Общая сумма произведений равна

$$SXY = \Sigma XY - C = (8)(7) + \dots + (11)(20) - 1600 = 142.$$

Сумму произведений для ошибки можно определить путем вычитания

$$SXYE = SXY - SXYB - SXYT = 142 - 50 - 24 = 68.$$

Таблица 18.3. Компоненты ошибок X и Y после устранения эффектов блоков и вариантов и общей средней

Блоки	X					Y				
	1	2	3	4	суммы	1	2	3	4	суммы
Варианты:										
1	3	0	-1	-2	0	2	1	0	-3	0
2	1	-4	2	1	0	2	-1	1	-2	0
3	-2	3	1	-2	0	-3	4	0	-1	0
4	-3	2	-3	4	0	-2	1	-2	3	0
5	1	-1	1	-1	0	1	-5	1	3	0
Суммы	0	0	0	0	0	0	0	0	0	0

Для того чтобы показать, откуда происходят суммы квадратов и произведений для ошибки и окончательно выяснить, как получается уравнение регрессии, мы устраним из каждой даты эффекты блоков и вариантов и значение общей средней, оставляя лишь остаточные компоненты ошибки по аналогии с расчетами в главе 5 (табл. 18.3).

Легко установить, что суммы квадратов этих компонентов равны суммам квадратов ошибки таблицы 18.2, полученным в дисперсионном анализе. Также сумма произведений соответствующих компонентов X и Y равна величине $SXYE$, полученной косвенно при вычитании. Как раз эти 20 пар компонентов ошибки используются для вычисления регрессии Y и X , свободной от влияния вариантов и блоков.

Из главы, посвященной линейной корреляции регрессии, мы узнали, что сумма квадратов отклонения от регрессии может быть найдена из выражения $(1-r^2)SSY$.

Это можно записать по-иному:

$$(1-r^2)SSY = \left[1 - \frac{(SXY)^2}{(SSX)(SSY)} \right] SSY = SSY - \frac{SXY^2}{SSX}.$$

Эту сумму квадратов отклонения от регрессии можно считать суммой квадратов Y после устранения влияния X на Y . Поэтому это равенство называют « Y , приведенный к X ».

Теперь мы имеем всю информацию, необходимую для проведения ковариационного анализа (табл. 18.4).

Сумма квадратов ошибки для Y , приведенного к X , равна

$$SSY - \frac{SXY^2}{SSX} = 104 - \frac{68^2}{96} = 55,833.$$

Эта величина имеет 11 степеней свободы в отличие от 12 для неприведенной ошибки.

Степени свободы, суммы квадратов и произведений в ряду, называемом «варианты + ошибка», получают просто, путем сложения их по рядам «варианты» и «ошибка». Затем мы получаем сумму квадратов

Таблица 18.4. Ковариационный анализ данных таблицы 18.2

Источники вариации	df	Суммы квадратов и произведений			df	Y, приведенный к X		
		SSX	SXY	SSY		SS	MS	F
Общий	19	186	142	334				
Блоки	3	50	50	70				
Варианты	4	40	24	160				
Ошибка	12	96	68	104	11	55,833	5,076	
Варианты + ошибка	16	136	92	265	15	201,765		
Варианты приведенные					4	145,932	36,483	7,19 **

для Y , приведенного к X , таким же путем в этом же ряду, как и для ошибки:

$$(\text{варианты} + \text{ошибка}) \text{ приведенная } SS = 264 - \frac{92^2}{136} = 201,765.$$

Сумму квадратов вариантов для Y , приведенного к X , теперь получают путем вычитания $201,765 - 55,833 = 145,932$. Важно отметить, что приведенная сумма квадратов для вариантов не может быть получена прямым путем применения формулы $SSY - SXY^2/SSX$ для строки вариантов. В этом случае мы будем иметь $160 - 24^2/40 = 145,6$. Тот факт, что эта величина довольно близка правильному ответу, является лишь совпадением. Эти две величины обычно не бывают столь близкими.

Коэффициент регрессии находят по строке «ошибка» путем использования обычной зависимости $b = SXY/SSX = 68/96 = 0,70833$. В качестве информации интересно посмотреть, что произойдет в случае приведения выражений ошибки для Y в таблице 18.3 к соответствующим выражениям ошибки для X . Это можно осуществить с помощью решения уравнения Y_{ij} приведенный $= Y_{ij} - bX_{ij}$ для каждого значения Y в таблице, как показано в таблице 18.5.

✓

Таблица 18.5. Значения ошибки для Y , приведенного к X

Блоки	1	2	3	4
Варианты:				
1	-0,12500	1,00000	0,70833	-1,58333
2	1,29167	1,83333	-0,41667	-2,70833
3	-1,58333	1,87500	-0,70833	0,41667
4	0,12500	-0,41667	0,12500	0,16667
5	0,29167	-4,29167	0,29167	3,70833

Не только суммы по блокам и вариантам остались равны нулю, как и должно быть, но также сумма квадратов этих скорректированных значений ошибки равна 55,833 — точно такой же величине, как и при ковариационном анализе.

Корректирование более чем одного источника вариации. Независимо от схемы эксперимента или числа исследуемых факторов можно придерживаться общей формы таблицы ковариационного анализа (табл. 18.4). Следует помнить, что величину каждого корректируемого источника вариации суммы квадратов и произведений для этого источника надо прибавлять к соответствующим суммам квадратов и произведений для ошибки. Итоговая строчка таблицы «источник + ошибка» используется для вычисления суммы квадратов Y , приведенного к X , и из этого результата вычитают исправленную сумму квадратов для ошибки, чтобы найти исправленную сумму квадратов для исследуемого источника вариации. Мы проиллюстрируем эту процедуру путем расчленения суммы квадратов вариантов нашего примера на четыре источника вариации или компоненты, каждый с одной степенью свободы.

Компоненты	Коэффициенты					$\Sigma (c_i T_i)_X$	$\Sigma (c_i T_i)_Y$	Σc_i^2
I	4	-1	-1	-1	-1	-20	-80	20
II	0	3	-1	-1	-1	12	-48	12
III	0	0	2	-1	-1	0	-24	6
IV	0	0	0	1	-1	-16	-8	2

Суммы квадратов для каждого компонента получают из обычной формулы $[\Sigma (c_i T_i)]^2 / r (\Sigma c_i^2)$. Для суммы произведений требуется некоторая модификация этой формулы

$$SXY = \Sigma (c_i T_i)_X \Sigma (c_i T_i)_Y / r (\Sigma c_i^2).$$

Результаты ковариационного анализа расчлененных эффектов вариантов приводятся в таблице 18.6. Необходимо отметить очень важную особенность этой таблицы. Неприведенные суммы квадратов и произведений отвечают слагаемости, то есть суммы четырех компонентов равны общим суммам квадратов и произведений для вариантов. С другой стороны, приведенные суммы квадратов не отвечают слагаемости. Сумма четырех компонентов равна 141,026, тогда как общая приведенная сумма квадратов для вариантов равна 145,932. Это означает, что мы не можем найти приведенный компонент суммы квадратов путем вычитания всех оставшихся компонентов из общей приведенной суммы квадратов для вариантов.

Корректировка средних по вариантам. Часто желательно оценить, какими стали бы средние вариантов зависимой переменной, если средние независимой переменной были бы одинаковыми для всех вариан-

Таблица 18.6. Итоги ковариационного анализа расчлененных эффектов вариантов

Источник вариации	df	Сумма квадратов и произведений			df	Y, приведенный к X		
		SSX	SXY	SSY		SS	MS	F
Общий	19	186	142	334				
Блоки	3	50	50	70				
Варианты	4	40	24	160				
Компонент I (KI)	1	5	20	80				
Компонент II (KII)	1	3	-12	48				
Компонент III (KIII)	1	0	0	24				
Компонент IV (KIV)	1	32	16	8				
Ошибка	12	96	68	104	11	55,833	5,076	
KI + ошибка	13	101	88	184	12	107,327		
KI приведенный					1	51,494	51,494	10,14 **
KII + ошибка	13	99	56	152	12	120,323		
KII приведенный					1	64,490	64,490	12,70 **
KIII + ошибка	13	96	68	128	12	79,83		
KIII приведенный					1	24,000	24,000	4,73 не
KIV + ошибка	13	128	84	112	12	56,875		
KIV приведенный					1	1,042	1,042	0,21 не
Сумма четырех компонентов						141,026		

тов. Эти приведенные средние находят из уравнения $\hat{Y}_i = \bar{Y}_i - b(X_i - \bar{X})$, где b это ошибка SXY / ошибку SSX . В нашем примере $b = 68/96 = 0,7083$ и приведенные средние будут равны.

$\bar{Y}_i$	$(\bar{X}_i - \bar{X})$	$b(\bar{X}_i - \bar{X})$	$\hat{Y}$
6	-1	-0,7083	6,7083
8	1	0,7083	7,2917
10	0	0,0000	10,0000
12	-2	-1,4166	13,4166
14	2	1,4166	12,5834

Можно ожидать, что приведенная сумма квадратов для вариантов может быть найдена непосредственно из приведенных средних по вариантам. И действительно, это иногда предлагается в качестве приближительного метода ковариационного анализа, если нет существенного воздействия вариантов на X , т. е. на независимую переменную. Однако имеется одно положение, которое в книгах по статистике редко полностью разъясняется: «Сумма квадратов приведенных средних по вариантам всегда больше, чем приведенная сумма квадратов для вариантов».

Разницу находят из уравнения

$$\frac{[SSXE(SXYT) - SXYE(SXXT)]^2}{(SSXE)^2(SXXT + SXXE)}.$$

В нашем примере она будет равна

$$\frac{96(24) - 68(40)^2}{(96)^2(40 + 96)} = 0,1381.$$

Сумма квадратов приведенных вариантов равна

$$4(6,7083^2 + \dots + 12,5834^2) - \frac{200^2}{20} = 146,0694.$$

(Отметьте, что поскольку мы работаем со средними, то производим умножение, а не деление на число повторений до вычитания корректирующего выражения.) Приведенная сумма квадратов для вариантов, полученная в ковариационном анализе, равнялась 145,932, и разница между этими двумя суммами квадратов составляет 0,1374, т. е. такую же величину, как и вычисленную по формуле, но без округлений.

Поскольку сумма квадратов для вариантов, получаемая на основе приведенных средних, всегда превышает истинную сумму квадратов, то и значения F также слишком высоки. Поэтому если при использовании приблизительного метода находят F -значения, лишь слегка превышающие уровень существенности, то надо применять точную технику анализа. С другой стороны, если F -значения приблизительного метода несущественны, то можно быть уверенным, что и точный метод покажет их несущественность.

Сравнение двух приведенных средних по вариантам. Поскольку дисперсия приведенных средних по вариантам больше, чем правильно приведенный средний квадрат вариантов, то обычная HSP не обеспечит соответствующую оценку сравнения приведенных средних. В отношении технической стороны расчетов возникает необходимость проведения оценки стандартной ошибки разности для каждой пары средних. Формула имеет вид

$$s_d^2 = \text{приведенная } EMS \left[\frac{2}{r} + \frac{(\bar{x}_p - \bar{x}_q)^2}{SSXE} \right].$$

В случае если число степеней свободы для ошибки равно 20 или больше и при этом отсутствует существенное воздействие вариантов на X , то приблизительная формула, используемая для всех пар средних, будет иметь такой вид

$$s_d^2 = \text{приведенная } EMS \left[\frac{2}{r} + \frac{2SSXT}{r(t-1)SSXE} \right].$$

Интерпретация результатов ковариационного анализа. В результате ковариационного анализа почти всегда значительно снижается средний квадрат ошибки; также снижается обычно и приведенный средний квадрат вариантов. По этой причине величины F для вариантов после корректирования могут быть больше или меньше, чем до корректирования. Интерпретация результатов зависит от того, име-

лось ли существенное воздействие на X , т. е. на независимую переменную, или нет.

В случае отсутствия существенного воздействия вариантов на X и на Y до, но не после корректирования, это будет означать, что отмеченный эффект вариантов на Y был преувеличен за счет случайной вариации внутри X , и поэтому вывод следует делать со значительной осторожностью.

Если воздействие вариантов на X не было существенно, а на Y существенно после, а не до корректирования, то действительное воздействие вариантов, вероятно, было скрыто за счет вариации внутри X .

Если варианты существенно воздействовали на X , тогда значение F после корректирования обычно бывает меньше, чем до корректирования. Если оно все еще существенно, тогда мы делаем вывод о существенном воздействии вариантов на Y сверх и кроме воздействия, связанного с вариацией внутри X .

Мы убедились, что техника ковариационного анализа значительно более запутанная, чем обычного дисперсионного анализа, а интерпретация результатов часто бывает трудной. Избегайте по возможности случайного распределения известной независимой переменной — это лучшее, что можно посоветовать. Последнее осуществляется путем тщательной группировки экспериментальных единиц в блоки, что делает возможным устранение большей части вариабельности, связанной с X , наряду с эффектом блоков.

СПЛАНИРОВАННАЯ ГРУППИРОВКА ЭКСПЕРИМЕНТАЛЬНЫХ ЕДИНИЦ — СХЕМА

Мы посвятили большую часть этой книги обсуждению экспериментальных схем и их роли в улучшении точности экспериментов. В опытно-м деле имеется много других схем, которые мы не обсуждали. На основании нашего опыта, однако, можно утверждать, что схемы, представленные здесь, используются в преобладающем большинстве сельскохозяйственных исследований. Читатель, заинтересованный в других схемах, должен познакомиться с более сложными пособиями, такими, как книга Кохрана и Кокса [3].

ЗАКЛЮЧЕНИЕ

Точность — это способность эксперимента выявить действительный эффект варианта. Она может быть улучшена посредством повышенной повторности, выбора вариантов, совершенствования техники, снижения вариабельности экспериментальных единиц одного варианта, увеличения размера экспериментальных единиц (до определенных пределов), использования ковариации и применения более эффективной экспериментальной схемы.

СПИСОК ЛИТЕРАТУРЫ

1. Alder H. L., Roessler E. V. 1968, 5th Ed. *Introduction to Probability and Statistics* W. H. Freeman and Co., San Francisco, 333 p.
2. Bancroft T. A. 1968. *Topics in Intermediate Statistical Methods*, Volume One. The Iowa State University Press, Ames, Iowa, 129 p.
3. Cochran W. G., Cox G. M. 1964, 2nd Ed. *Experimental Design*. John Wiley and Sons, Inc., New York, 617 p.
4. Finney D. J. 1962, 2nd Ed. *An Introduction to Statistical Science in Agriculture*. Munksgaard, Copenhagen, Denmark, and Oliver and Boyd Ltd., Edinburg, Scotland, 216 p.
5. LeClerc E. L., Leonard W. H., Clark A. G. 1962, 2nd Ed. *Field Plot Technique*. Burgess Publishing Co., Minneapolis, Minnesota, 373 p.
6. Snedecor G. W. Cochran W. G. 1967, 6th Ed. *Statistical Methods*. The Iowa State University Press, Ames, Iowa 593 p.
7. Sokal R. R., Rohlf R. J. 1969. *Biometry, The Principles and Practice of Statistics in Biological Research*. W. H. Freeman and Co., San Francisco, 776 p.
8. Steel R. G. D., Torrie J. H. 1960. *Principles and Procedures of Statistics with Special Reference to the Biological Sciences* McGraw-Hill Book Co., Inc., New York, 481 p.

ПРИЛОЖЕНИЕ

Таблица А.1. Случайные числа

При рендомизации любой группы из 10 или меньшего числа единиц произвольно возьмите начальную цифру и продолжайте отбор по рядам, колонкам или диагоналям в любом направлении. Записывайте числа по мере их появления, пропуская те, которые превосходят по величине общее количество рендомизируемых единиц и уже записанные числа. Если вы желаете провести рендомизацию для числа единиц, превышающего 10, можно взять двойные колонки или ряды, что даст двухзначные цифры и продолжить процесс рендомизации аналогично вышеописанному

8	2	0	3	1	4	5	8	2	1	7	2	7	3	8	5	2	9	0	6	3	1	6	4
0	8	7	3	3	1	9	7	5	2	5	7	6	9	8	0	5	2	5	3	4	0	1	0
2	3	3	8	6	1	4	2	4	0	2	6	1	8	9	0	5	2	6	9	8	7	7	2
4	7	5	5	6	3	0	7	7	1	9	1	6	1	7	4	1	7	1	3	7	9	3	7
1	9	3	9	5	3	4	9	5	5	2	7	5	8	0	3	4	8	8	1	2	7	5	3
2	8	7	8	1	4	1	4	6	9	4	2	4	1	5	2	9	4	6	2	1	5	2	8
8	4	8	5	1	3	9	6	6	0	4	7	2	1	9	0	2	0	6	7	0	6	0	1
0	3	8	8	4	7	5	1	5	1	7	3	4	5	2	0	7	4	7	9	6	6	7	7
3	5	3	1	9	3	7	4	9	5	0	2	0	1	4	6	2	5	4	5	8	5	0	9
3	4	5	9	5	2	7	9	8	9	0	5	5	8	5	1	3	7	7	3	5	5	4	7
4	1	5	3	0	9	1	3	7	2	5	8	7	7	1	3	6	3	9	7	8	7	9	1
7	2	9	5	6	7	8	5	4	5	3	4	5	4	1	9	8	6	7	5	7	9	3	1
5	9	2	8	9	8	6	4	4	1	5	3	7	7	0	8	0	2	5	6	0	6	1	2
1	3	3	3	9	0	5	2	8	7	4	0	9	0	3	7	3	1	7	9	4	5	5	2
4	6	0	1	0	8	6	2	1	0	0	5	0	3	1	5	4	9	0	3	7	4	7	0
7	7	0	6	6	3	2	8	8	5	8	9	5	6	4	0	5	9	1	8	0	5	4	9
3	3	8	5	7	5	7	4	3	4	5	7	9	6	9	5	0	7	7	6	6	8	8	5
9	1	7	1	3	6	9	2	9	1	9	4	2	3	3	0	8	1	8	7	7	6	4	7
6	2	2	8	0	9	4	5	3	7	2	5	4	6	6	5	3	7	4	0	4	6	5	8
1	7	5	9	0	0	2	0	5	6	5	8	5	1	9	5	6	6	5	0	4	6	5	6
0	3	9	6	9	4	7	3	5	7	0	6	5	4	7	1	1	8	5	3	2	8	0	9
3	0	8	2	8	1	4	4	1	6	7	6	6	9	9	9	7	5	8	9	6	4	5	9
9	4	9	1	2	2	0	1	3	2	4	6	7	9	1	0	8	2	9	8	3	2	6	2
7	2	5	1	4	4	9	6	5	2	8	5	5	1	0	8	2	6	2	0	6	9	2	2
9	9	2	5	7	4	3	1	2	3	6	4	1	5	2	4	0	4	2	2	8	7	1	8
2	0	9	1	8	9	4	4	6	1	4	8	6	7	9	2	5	0	6	9	3	3	0	1
6	5	2	6	1	2	1	7	7	1	4	7	8	1	4	3	2	7	3	7	4	0	0	1
1	2	9	9	6	4	2	5	3	2	7	4	3	2	3	3	8	5	3	3	6	5	5	3
3	2	8	3	7	9	6	0	4	8	6	0	5	4	1	1	4	9	0	5	0	9	4	4
0	9	3	4	1	1	9	5	8	3	2	4	6	7	3	4	4	9	2	3	7	2	5	7
6	7	5	3	4	2	1	5	5	0	1	2	4	7	5	5	2	6	8	7	8	2	8	0
9	6	0	1	3	0	5	3	6	6	2	9	6	0	3	4	7	6	1	1	9	1	6	5
4	6	9	9	6	7	8	5	8	1	2	9	2	6	2	4	4	9	0	5	5	4	6	2
9	7	7	1	9	2	6	5	6	3	3	6	3	6	8	3	9	9	8	7	7	2	7	9
7	5	3	3	3	3	7	3	7	6	7	3	9	1	1	2	3	9	0	9	5	9	6	5
2	8	1	3	1	3	4	2	1	0	3	1	2	3	2	0	2	3	9	7	7	5	0	6
6	0	9	4	8	8	5	5	3	7	9	0	0	0	0	1	9	2	0	6	1	5	8	4
3	5	9	0	7	7	0	1	8	1	2	9	3	4	6	9	2	8	9	8	9	8	6	5
4	4	8	1	1	7	4	4	7	4	4	4	1	6	5	9	3	6	5	9	8	3	2	4
6	3	9	7	0	6	2	5	3	3	2	6	0	5	1	2	4	3	7	1	0	7	8	2

Таблица А.2. Распределение t^*

Степени свободы	Вероятность получения такой или большей величины (уровни значимости)					
	0,400	0,200	0,100	0,050	0,010	0,001
1	1,376	3,078	6,314	12,706	63,657	
2	1,061	1,886	2,920	4,303	9,925	31,598
3	0,978	1,638	2,353	3,182	5,841	12,941
4	0,941	1,533	2,132	2,776	4,604	8,610
5	0,920	1,476	2,015	2,571	4,032	6,859
6	0,906	1,440	1,943	2,447	3,707	5,959
7	0,895	1,415	1,895	2,365	3,499	5,405
8	0,889	1,397	1,860	2,306	3,355	5,041
9	0,883	1,383	1,833	2,262	3,250	4,781
10	0,879	1,372	1,812	2,228	3,169	4,587
11	0,876	1,363	1,796	2,201	3,106	4,437
12	0,873	1,356	1,782	2,179	3,055	4,318
13	0,870	1,350	1,771	2,160	3,012	4,221
14	0,868	1,345	1,761	2,145	2,977	4,140
15	0,866	1,341	1,753	2,131	2,947	4,073
16	0,865	1,337	1,746	2,120	2,921	4,015
17	0,863	1,333	1,740	2,110	2,898	3,965
18	0,862	1,330	1,734	2,101	2,878	3,922
19	0,861	1,328	1,729	2,093	2,861	3,883
20	0,860	1,325	1,725	2,086	2,845	3,850
21	0,859	1,323	1,721	2,080	2,831	3,819
22	0,858	1,321	1,717	2,074	2,819	3,792
23	0,858	1,319	1,714	2,069	2,807	3,767
24	0,857	1,318	1,711	2,064	2,797	3,745
25	0,856	1,316	1,708	2,060	2,787	3,725
26	0,856	1,315	1,706	2,056	2,779	3,707
27	0,855	1,314	1,703	2,052	2,771	3,690
28	0,855	1,313	1,701	2,048	2,763	3,674
29	0,854	1,311	1,699	2,045	2,756	3,659
30	0,854	1,310	1,697	2,042	2,750	3,646
35	0,852	1,306	1,690	2,030	2,724	3,591
40	0,851	1,303	1,684	2,021	2,704	3,551
45	0,850	1,301	1,680	2,014	2,690	3,520
50	0,849	1,299	1,676	2,008	2,678	3,496
55	0,849	1,297	1,673	2,004	2,669	3,476
60	0,848	1,296	1,671	2,000	2,660	3,460
70	0,847	1,294	1,667	1,994	2,648	3,435
80	0,847	1,293	1,665	1,989	2,638	3,416
90	0,846	1,291	1,662	1,986	2,631	3,402
100	0,846	1,290	1,661	1,982	2,625	3,390
120	0,845	1,289	1,658	1,980	2,617	3,373
∞	0,8416	1,2816	1,6448	1,9600	2,5758	2,2905

* Определенные значения этой таблицы взяты из таблицы III книги Фишера и Нейтса: Statistical tables for Biological, Agricultural, and Medical Research, опубликованной by Oliver & Boyd, Edinburgh (ранее издававшейся by Longman Group Ltd., London). С разрешения авторов и издателей другая часть была вычислена по способу, предлагаемому Ченом и Маковски (см. подпись под таблицей А.3).

Таблица А.3. 10%, 5% и 1% уровни значимости для F-распределения *

df числителя	p	Степени свободы знаменателя								
		1	2	3	4	5	6	7	8	9
1	0,10	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86
	0,05	161	200	216	225	230	234	237	239	241
	0,01	4052	4999	5403	5625	5764	5859	5928	5981	6022
2	0,10	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38
	0,05	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
	0,01	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	0,10	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24
	0,05	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
	0,01	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	0,10	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94
	0,05	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
	0,01	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	0,10	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32
	0,05	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
	0,01	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	0,10	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96
	0,05	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
	0,01	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	0,10	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72
	0,05	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
	0,01	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	0,10	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56
	0,05	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
	0,01	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	0,10	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44
	0,05	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
	0,01	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	0,10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35
	0,05	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
	0,01	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	0,10	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27
	0,05	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
	0,01	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	0,10	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21
	0,05	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
	0,01	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	0,10	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16
	0,05	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
	0,01	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	0,10	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12
	0,05	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
	0,01	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	0,10	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09
	0,05	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
	0,01	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89

(больший средний квадрат)

10	11	12	13	14	15	16	17	18	19	20
60,19 242 6056	60,47 243 6083	60,71 244 6106	60,90 245 6126	61,07 245 6143	61,22 246 6157	61,35 246 6170	61,46 247 6181	61,57 247 6191	61,66 248 6201	61,74 248 6209
9,39	9,40	9,41	9,41	9,42	9,42	9,43	9,43	9,44	9,44	9,44
19,40	19,40	19,41	19,42	19,42	19,43	19,43	19,44	19,44	19,44	19,45
99,40	99,41	99,42	99,42	99,43	99,43	99,44	99,44	99,44	99,45	99,45
5,23	5,22	5,22	5,21	5,20	5,20	5,20	5,19	5,19	5,19	5,18
8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66
27,23	27,13	27,05	26,98	26,92	26,87	26,83	26,79	26,75	26,72	26,69
3,92	3,91	3,90	3,89	3,88	3,87	3,86	3,86	3,85	3,85	3,84
5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80
14,55	14,45	14,37	14,31	14,25	14,20	14,15	14,11	14,08	14,05	14,02
3,30	3,28	3,27	3,26	3,25	3,24	3,23	3,22	3,22	3,21	3,21
4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56
10,05	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55
2,94	2,92	2,90	2,89	2,88	2,87	2,86	2,85	2,85	2,84	2,84
4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87
7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40
2,70	2,68	2,67	2,65	2,64	2,63	2,62	2,61	2,61	2,60	2,59
3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44
6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24	9,21	6,18	6,16
2,54	2,52	2,50	2,49	2,48	2,46	2,45	2,45	2,44	2,43	2,42
3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15
5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36
2,42	2,40	2,38	2,36	2,35	2,34	2,33	2,32	2,31	2,30	2,30
3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94
5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81
2,32	2,30	2,28	2,27	2,26	2,24	2,23	2,22	2,22	2,21	2,20
2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
4,85	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41
2,25	2,23	2,21	2,19	2,18	2,17	2,16	2,15	2,14	2,13	2,12
2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65
4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10
2,19	2,17	2,15	2,13	2,12	2,10	2,09	2,08	2,08	2,07	2,06
2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54
4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86
2,14	2,12	2,10	2,08	2,07	2,05	2,04	2,03	2,02	2,01	2,01
2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46
4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66
2,10	2,07	2,05	2,04	2,02	2,01	2,00	1,99	1,98	1,97	1,96
2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39
3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51
2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,95	1,94	1,93	1,92
2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33
3,80	3,73	3,67	3,61	3,57	3,56	3,49	3,45	3,42	3,40	3,37

DF зна- мена- теля	P	Степени свободы числителя									
		1	2	3	4	5	6	7	8	9	
16	0,10	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	
	0,05	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	
	0,01	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	
17	0,10	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	
	0,05	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	
	0,01	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	
18	0,10	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	
	0,05	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	
	0,01	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	
19	0,10	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	
	0,05	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	
	0,01	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	
20	0,10	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	
	0,05	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	
	0,01	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	
21	0,10	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	
	0,05	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	
	0,01	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	
22	0,10	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	
	0,05	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	
	0,01	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	
23	0,10	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	
	0,05	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	
	0,01	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	
24	0,10	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	
	0,05	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	
	0,01	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	
25	0,10	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	
	0,05	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	
	0,01	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	
26	0,10	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	
	0,05	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	
	0,01	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	
27	0,10	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	
	0,05	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	
	0,01	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	
28	0,10	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	
	0,05	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	
	0,01	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	
29	0,10	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	
	0,05	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	
	0,01	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	
30	0,10	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	
	0,05	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	
	0,01	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	

(большой средний квадрат)

10	11	12	13	14	15	16	17	18	19	20
2,03	2,01	1,99	1,97	1,95	1,94	1,93	1,92	1,91	1,90	1,89
2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28
3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26
2,00	1,98	1,96	1,94	1,93	1,91	1,90	1,89	1,88	1,87	1,86
2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23
3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,27	3,21	3,19	3,16
1,98	1,95	1,93	1,92	1,90	1,89	1,87	1,86	1,85	1,84	1,84
2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19
3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08
1,96	1,93	1,91	1,89	1,88	1,86	1,85	1,84	1,83	1,82	1,81
2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16
3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00
1,94	1,91	1,89	1,87	1,86	1,84	1,83	1,82	1,81	1,80	1,79
2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12
3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94
1,92	1,90	1,87	1,86	1,84	1,83	1,81	1,80	1,79	1,78	1,78
2,32	2,28	2,25	2,22	2,20	2,18	2,16	2,14	2,12	2,11	2,10
3,31	3,24	3,17	3,12	3,07	3,03	2,99	2,96	2,93	2,90	2,88
1,90	1,88	1,86	1,84	1,83	1,81	1,80	1,79	1,78	1,77	1,76
2,30	2,26	2,23	2,20	2,17	2,15	2,13	2,11	2,10	2,08	2,07
3,26	3,18	3,12	3,07	3,02	2,98	2,94	2,91	2,88	2,85	2,83
1,89	1,87	1,84	1,83	1,81	1,80	1,78	1,77	1,76	1,75	1,74
2,27	2,24	2,20	2,18	2,15	2,13	2,11	2,09	2,08	2,06	2,05
3,21	3,14	3,07	3,02	2,97	2,93	2,89	2,86	2,83	2,80	2,78
1,88	1,85	1,83	1,81	1,80	1,78	1,77	1,76	1,75	1,74	1,73
2,25	2,22	2,18	2,15	2,13	2,11	2,09	2,07	2,05	2,04	2,03
3,17	3,09	3,03	2,98	2,93	2,89	2,85	2,82	2,79	2,76	2,74
1,87	1,84	1,82	1,80	1,79	1,77	1,76	1,75	1,74	1,73	1,72
2,24	2,20	2,16	2,14	2,11	2,09	2,07	2,05	2,04	2,02	2,01
3,13	3,06	2,99	2,94	2,89	2,85	2,81	2,78	2,75	2,72	2,70
1,86	1,83	1,81	1,79	1,77	1,76	1,75	1,73	1,72	1,71	1,71
2,22	2,18	2,15	2,12	2,09	2,07	2,05	2,03	2,08	2,00	1,99
3,09	3,02	2,96	2,90	2,86	2,81	2,78	2,75	2,72	2,69	2,66
1,85	1,82	1,80	1,78	1,76	1,75	1,74	1,76	1,71	1,70	1,70
2,20	2,17	2,13	2,10	2,08	2,06	2,04	2,02	2,00	1,99	1,97
3,06	2,99	2,93	2,87	2,82	2,78	2,75	2,71	2,68	2,66	2,63
1,84	1,81	1,79	1,77	1,75	1,74	1,73	1,71	1,70	1,69	1,69
2,19	2,15	2,12	2,09	2,06	2,04	2,02	2,00	1,99	1,97	1,96
3,03	2,96	2,90	2,84	2,79	2,75	2,72	2,68	2,65	2,63	2,60
1,83	1,80	1,78	1,76	1,75	1,73	1,72	1,71	1,69	1,68	1,68
2,18	2,14	2,10	2,08	2,05	2,03	2,01	1,99	1,97	1,96	1,94
3,00	2,93	2,87	2,81	2,77	2,73	2,69	2,66	2,63	2,60	2,57
1,82	1,79	1,77	1,75	1,74	1,72	1,71	1,70	1,69	1,68	1,67
2,16	2,13	2,09	2,06	2,04	2,01	1,99	1,98	1,96	1,95	1,93
2,98	2,91	2,84	2,79	2,74	2,70	2,66	2,63	2,60	2,57	2,55

Др. зна- мена- теля	P	Степени свободы числителя									
		1	2	3	4	5	6	7	8	9	10
32	0,10	2,87	2,48	2,26	2,13	2,04	1,97	1,91	1,87	1,83	
	0,05	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	
	0,01	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	
34	0,10	2,86	2,47	2,25	2,12	2,02	1,96	1,90	1,86	1,82	
	0,05	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	
	0,01	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	
36	0,10	2,85	2,46	2,24	2,11	2,01	1,94	1,89	1,85	1,81	
	0,05	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	
	0,01	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	
38	0,10	2,84	2,45	2,23	2,10	2,01	1,94	1,88	1,84	1,80	
	0,05	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	
	0,01	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	
40	0,10	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	
	0,05	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	
	0,01	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	
42	0,10	2,83	2,43	2,22	2,08	1,99	1,92	1,86	1,82	1,78	
	0,05	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	
	0,01	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	
44	0,10	2,82	2,43	2,21	2,08	1,98	1,91	1,86	1,81	1,78	
	0,05	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	
	0,01	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	
46	0,10	2,82	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77	
	0,05	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	
	0,01	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	
48	0,10	2,81	2,42	2,20	2,07	1,97	1,90	1,85	1,80	1,77	
	0,05	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	
	0,01	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	
50	0,10	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	
	0,05	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	
	0,01	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	
55	0,10	2,80	2,40	2,19	2,05	1,95	1,88	1,83	1,78	1,75	
	0,05	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	
	0,01	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	
60	0,10	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	
	0,05	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	
	0,01	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	

(большой средний квадрат)

10	11	12	13	14	15	16	17	18	19	20
1,81	1,78	1,76	1,74	1,72	1,71	1,69	1,68	1,67	1,66	1,65
2,14	2,10	2,07	2,04	2,01	1,99	1,97	1,95	1,94	1,92	1,91
2,93	2,86	2,80	2,74	2,70	2,65	2,62	2,58	2,55	2,53	2,50
1,79	1,77	1,75	1,73	1,71	1,69	1,68	1,67	1,66	1,65	1,64
2,12	2,08	2,05	2,02	1,99	1,97	1,95	1,93	1,92	1,90	1,89
2,89	2,82	2,76	2,70	2,66	2,61	2,58	2,54	2,51	2,49	2,46
1,78	1,76	1,73	1,71	1,70	1,68	1,67	1,66	1,65	1,64	1,63
2,11	2,07	2,03	2,00	1,98	1,95	1,93	1,92	1,90	1,88	1,87
2,86	2,79	2,72	2,67	2,62	2,58	2,54	2,51	2,48	2,45	2,43
1,77	1,75	1,72	1,70	1,69	1,67	1,66	1,65	1,63	1,62	1,61
2,09	2,05	2,02	1,99	1,96	1,94	1,92	1,90	1,88	1,87	1,85
2,83	2,75	2,69	2,64	2,59	2,55	2,51	2,48	2,45	2,42	2,40
1,76	1,74	1,71	1,70	1,68	1,66	1,65	1,64	1,62	1,61	1,61
2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,89	1,87	1,85	1,84
2,80	2,73	2,66	2,61	2,56	2,52	2,48	2,45	2,42	2,39	2,37
1,75	1,73	1,71	1,69	1,67	1,65	1,64	1,63	1,62	1,61	1,60
2,06	2,03	1,99	1,96	1,94	1,91	1,89	1,87	1,86	1,84	1,83
2,78	2,70	2,64	2,59	2,54	2,50	2,46	2,43	2,40	2,37	2,34
1,75	1,72	1,70	1,68	1,66	1,65	1,63	1,62	1,61	1,60	1,59
2,05	2,01	1,98	1,95	1,92	1,90	1,88	1,86	1,84	1,83	1,81
2,75	2,68	2,62	2,56	2,52	2,47	2,44	2,40	2,37	2,35	2,32
1,74	1,71	1,69	1,67	1,65	1,64	1,63	1,61	1,60	1,59	1,58
2,04	2,00	1,97	1,94	1,91	1,89	1,87	1,85	1,83	1,82	1,80
2,78	2,66	2,60	2,54	2,50	2,45	2,42	2,38	2,35	2,33	2,30
1,73	1,71	1,69	1,67	1,65	1,63	1,62	1,61	1,59	1,58	1,57
2,03	1,99	1,96	1,93	1,90	1,88	1,86	1,84	1,82	1,81	1,79
2,71	2,64	2,58	2,53	2,48	2,44	2,40	2,37	2,33	2,31	2,28
1,73	1,70	1,68	1,66	1,64	1,63	1,61	1,60	1,59	1,58	1,57
2,03	1,99	1,95	1,92	1,89	1,87	1,85	1,83	1,81	1,80	1,78
2,70	2,63	2,56	2,51	2,46	2,42	2,38	2,35	2,32	2,29	2,27
1,72	1,69	1,67	1,65	1,63	1,61	1,60	1,59	1,58	1,56	1,55
2,01	1,97	1,93	1,90	1,88	1,85	1,83	1,81	1,79	1,78	1,76
2,66	2,59	2,53	2,47	2,42	2,38	2,34	2,31	2,28	2,25	2,23
1,71	1,68	1,66	1,64	1,62	1,60	1,59	1,58	1,56	1,55	1,54
1,99	1,95	1,92	1,89	1,86	1,84	1,82	1,80	1,78	1,76	1,75
2,63	2,56	2,50	2,44	2,39	2,35	2,31	2,28	2,25	2,22	2,20

DF зна- мена- теля	P	Степени свободы числителя								
		1	2	3	4	5	6	7	8	9
65	0,10	2,78	2,39	2,17	2,03	1,94	1,87	1,81	1,77	1,73
	0,05	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03
	0,01	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69
70	0,10	2,78	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72
	0,05	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02
	0,01	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67
80	0,10	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71
	0,05	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00
	0,01	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64
100	0,10	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69
	0,05	3,91	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97
	0,01	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59
120	0,10	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68
	0,05	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96
	0,01	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
150	0,10	2,74	2,34	2,12	1,98	1,89	1,81	1,76	1,71	1,67
	0,05	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94
	0,01	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53
200	0,10	2,73	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66
	0,05	3,89	3,01	2,65	2,42	2,26	2,14	2,06	1,98	1,93
	0,01	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50
400	0,10	2,72	2,32	2,10	1,96	1,86	1,79	1,73	1,69	1,65
	0,05	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90
	0,01	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	2,45
1000	0,10	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64
	0,05	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89
	0,01	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43
∞	0,10	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63
	0,05	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88
	0,01	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

* Данные этой таблицы рассчитаны по способу, предлагаемому в статье Hubert J. Chen 76—3, Memphis State University, Department of Mathematical Sciences (1976).

(большой средний квадрат)

10	11	12	13	14	15	16	17	18	19	20
1,70	1,67	1,65	1,63	1,61	1,59	1,58	1,57	1,55	1,54	1,53
1,98	1,94	1,90	1,87	1,85	1,82	1,80	1,78	1,76	1,75	1,73
2,61	2,53	2,47	2,42	2,37	2,33	2,29	2,26	2,23	2,20	2,17
1,69	1,66	1,64	1,62	1,60	1,59	1,57	1,56	1,55	1,54	1,53
1,97	1,93	1,89	1,86	1,84	1,81	1,79	1,77	1,75	1,74	1,72
2,59	2,51	2,45	2,40	2,35	2,31	2,27	2,23	2,20	2,18	2,15
1,68	1,65	1,63	1,61	1,59	1,57	1,56	1,55	1,53	1,52	1,51
1,95	1,91	1,88	1,84	1,82	1,79	1,77	1,75	1,73	1,72	1,70
2,55	2,48	2,42	2,36	2,31	2,27	2,23	2,20	2,17	2,14	2,12
1,66	1,64	1,61	1,59	1,57	1,56	1,54	1,53	1,52	1,50	1,49
1,93	1,89	1,85	1,82	1,79	1,77	1,75	1,73	1,71	7,69	1,68
2,50	2,43	2,37	2,31	2,27	2,22	2,19	2,15	2,12	2,09	2,07
1,65	1,63	1,60	1,58	1,56	1,55	1,53	1,52	1,50	1,49	1,48
1,91	1,87	1,83	1,80	1,78	1,75	1,73	1,71	1,69	1,67	1,66
2,47	2,40	2,34	2,28	2,23	2,19	2,15	2,12	2,09	2,06	2,03
1,64	1,61	1,59	1,57	1,55	1,53	1,52	1,50	1,49	1,48	1,47
1,89	1,85	1,82	1,79	1,76	1,73	1,71	1,69	1,67	1,66	1,64
2,44	2,37	2,31	2,25	2,20	2,16	2,12	2,08	2,06	2,03	2,00
1,63	1,60	1,58	1,56	1,54	1,52	1,51	1,49	1,48	1,47	1,46
1,88	1,84	1,80	1,77	1,74	1,72	1,69	1,67	1,66	1,64	1,62
2,41	2,34	2,27	2,22	2,17	2,13	2,09	2,06	2,03	2,00	1,97
1,61	1,59	1,56	1,54	1,52	1,50	1,49	1,47	1,46	1,45	1,44
1,85	1,81	1,78	1,74	1,72	1,69	1,67	1,65	1,63	1,61	1,60
2,37	2,29	2,23	2,17	2,13	2,08	2,05	2,01	1,98	1,95	1,92
1,61	1,58	1,55	1,53	1,51	1,49	1,48	1,46	1,45	1,44	1,43
1,84	1,80	1,76	1,73	1,70	1,68	1,65	1,63	1,61	1,60	1,58
2,34	2,27	2,20	2,15	2,10	2,06	2,02	1,98	1,95	1,92	1,90
1,60	1,57	1,55	1,52	1,50	1,49	1,47	1,45	1,44	1,43	1,42
1,83	1,79	1,75	1,72	1,69	1,67	1,64	1,62	1,60	1,58	1,57
2,32	2,25	2,18	2,13	2,08	2,04	1,99	1,96	1,93	1,90	1,88

and A. B. Makowsky, "On Approximations to the F-Distribution and Its Inverse", Report

Таблица А.А. Существенные значения фактора Стиюдента (R) для умножения на НСР при сравнении разноразмерных средних ранжированного ряда на 5%-ном уровне; p — интервалы между средними и n — степени свободы для ошибок

n	p:																			
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100				
4	1,00	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02				
5	1,00	1,03	1,04	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05	1,05				
6	1,00	1,03	1,05	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06	1,06				
7	1,00	1,04	1,06	1,07	1,07	1,07	1,08	1,08	1,08	1,08	1,08	1,08	1,08	1,08	1,08	1,08				
8	1,00	1,04	1,06	1,08	1,09	1,09	1,09	1,09	1,09	1,09	1,09	1,09	1,09	1,09	1,09	1,09				
9	1,00	1,04	1,07	1,08	1,09	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10				
10	1,00	1,05	1,07	1,09	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10	1,10				
11	1,00	1,05	1,08	1,09	1,10	1,11	1,11	1,11	1,11	1,11	1,11	1,11	1,11	1,12	1,12	1,12				
12	1,00	1,05	1,08	1,09	1,10	1,11	1,12	1,12	1,12	1,12	1,12	1,12	1,13	1,13	1,13	1,13				
13	1,00	1,05	1,08	1,09	1,10	1,11	1,12	1,12	1,13	1,13	1,13	1,13	1,13	1,13	1,13	1,13				
14	1,00	1,05	1,08	1,10	1,11	1,12	1,13	1,13	1,14	1,14	1,14	1,14	1,15	1,15	1,15	1,15				
15	1,00	1,05	1,08	1,10	1,12	1,12	1,13	1,14	1,14	1,14	1,15	1,15	1,15	1,15	1,15	1,15				
16	1,00	1,05	1,08	1,10	1,12	1,12	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,16	1,16	1,16				
17	1,00	1,05	1,08	1,10	1,12	1,13	1,13	1,14	1,15	1,15	1,16	1,16	1,16	1,16	1,16	1,16				
18	1,00	1,05	1,08	1,10	1,12	1,13	1,13	1,14	1,15	1,15	1,16	1,16	1,17	1,17	1,17	1,17				
19	1,00	1,05	1,08	1,10	1,12	1,13	1,14	1,15	1,15	1,16	1,16	1,17	1,17	1,17	1,17	1,17				
20	1,00	1,05	1,08	1,10	1,12	1,13	1,14	1,15	1,15	1,16	1,17	1,17	1,18	1,18	1,18	1,18				
22	1,00	1,05	1,08	1,10	1,12	1,13	1,14	1,15	1,16	1,17	1,18	1,18	1,18	1,18	1,19	1,19				
24	1,00	1,05	1,08	1,10	1,12	1,13	1,14	1,15	1,16	1,17	1,18	1,18	1,18	1,19	1,19	1,19				
26	1,00	1,05	1,08	1,10	1,12	1,13	1,15	1,15	1,16	1,17	1,18	1,19	1,19	1,19	1,20	1,20				
28	1,00	1,05	1,08	1,10	1,12	1,14	1,15	1,16	1,16	1,17	1,18	1,19	1,20	1,20	1,20	1,20				
30	1,00	1,05	1,08	1,11	1,12	1,14	1,15	1,16	1,17	1,18	1,20	1,20	1,21	1,21	1,21	1,21				
40	1,00	1,05	1,08	1,11	1,13	1,14	1,15	1,16	1,17	1,19	1,20	1,21	1,22	1,23	1,23	1,23				
60	1,00	1,05	1,09	1,11	1,13	1,14	1,16	1,17	1,18	1,19	1,21	1,21	1,22	1,24	1,24	1,24				
100	1,00	1,05	1,09	1,11	1,14	1,15	1,16	1,18	1,19	1,20	1,21	1,22	1,23	1,24	1,26	1,26				
∞	1,00	1,05	1,09	1,12	1,14	1,15	1,17	1,18	1,19	1,21	1,22	1,23	1,24	1,25	1,30	1,32				

Таблица А.5. Существенные значения фактора Стьюдента (R) для умножения на HCp при сравнении разнородных средних ратифицированного ряда на 1%-ном уровне; p — интервалы между средними и n — степени свободы для «ошибки»

n	p															
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
3	1,00	1,03	1,04	1,05	1,07	1,08	1,08	1,09	1,09	1,09	1,10	1,11	1,13	1,13	1,13	1,13
4	1,00	1,04	1,06	1,08	1,09	1,09	1,11	1,12	1,12	1,12	1,14	1,14	1,15	1,15	1,15	1,15
5	1,00	1,05	1,07	1,08	1,10	1,11	1,12	1,11	1,13	1,13	1,16	1,18	1,18	1,19	1,19	1,19
6	1,00	1,05	1,08	1,10	1,12	1,13	1,15	1,15	1,16	1,16	1,18	1,18	1,20	1,20	1,20	1,20
7	1,00	1,05	1,08	1,10	1,12	1,14	1,16	1,16	1,17	1,17	1,19	1,19	1,21	1,21	1,21	1,21
8	1,00	1,06	1,08	1,10	1,12	1,14	1,16	1,17	1,17	1,18	1,20	1,20	1,22	1,22	1,22	1,23
9	1,00	1,06	1,09	1,11	1,13	1,15	1,16	1,17	1,18	1,20	1,21	1,22	1,24	1,24	1,24	1,24
10	1,00	1,05	1,09	1,11	1,13	1,14	1,15	1,17	1,17	1,19	1,20	1,22	1,23	1,23	1,23	1,23
11	1,00	1,05	1,08	1,10	1,12	1,14	1,15	1,16	1,17	1,19	1,20	1,21	1,21	1,22	1,22	1,22
12	1,00	1,05	1,08	1,10	1,12	1,14	1,15	1,16	1,17	1,18	1,19	1,20	1,21	1,21	1,21	1,21
13	1,00	1,05	1,08	1,10	1,12	1,14	1,15	1,16	1,17	1,18	1,19	1,20	1,20	1,20	1,20	1,20
14	1,00	1,05	1,08	1,10	1,12	1,14	1,15	1,16	1,17	1,18	1,19	1,20	1,20	1,20	1,20	1,20
15	1,00	1,05	1,08	1,10	1,11	1,13	1,14	1,15	1,16	1,18	1,18	1,19	1,20	1,20	1,20	1,20
16	1,00	1,05	1,08	1,10	1,11	1,13	1,14	1,15	1,16	1,17	1,18	1,19	1,20	1,20	1,20	1,20
17	1,00	1,05	1,08	1,10	1,11	1,13	1,14	1,15	1,16	1,17	1,18	1,19	1,19	1,19	1,19	1,19
18	1,00	1,05	1,08	1,10	1,11	1,13	1,14	1,15	1,16	1,17	1,18	1,18	1,19	1,19	1,19	1,19
19	1,00	1,05	1,07	1,09	1,11	1,13	1,14	1,15	1,16	1,17	1,18	1,18	1,19	1,19	1,19	1,19
20	1,00	1,05	1,07	1,09	1,11	1,13	1,14	1,15	1,16	1,17	1,18	1,18	1,19	1,19	1,19	1,19
22	1,00	1,05	1,07	1,09	1,11	1,12	1,14	1,15	1,15	1,17	1,17	1,18	1,19	1,19	1,19	1,19
24	1,00	1,05	1,07	1,09	1,11	1,12	1,13	1,15	1,15	1,17	1,17	1,18	1,19	1,19	1,20	1,20
26	1,00	1,05	1,07	1,09	1,11	1,12	1,13	1,15	1,15	1,17	1,18	1,18	1,19	1,19	1,20	1,20
28	1,00	1,04	1,07	1,09	1,11	1,12	1,13	1,14	1,15	1,17	1,18	1,18	1,19	1,19	1,21	1,21
30	1,00	1,04	1,07	1,09	1,11	1,12	1,13	1,14	1,15	1,17	1,18	1,19	1,20	1,20	1,21	1,21
40	1,00	1,04	1,07	1,09	1,11	1,12	1,14	1,15	1,15	1,17	1,18	1,19	1,20	1,20	1,23	1,23
60	1,00	1,04	1,07	1,09	1,11	1,12	1,14	1,15	1,15	1,17	1,18	1,19	1,20	1,20	1,24	1,24
100	1,00	1,04	1,07	1,09	1,11	1,12	1,14	1,15	1,15	1,17	1,18	1,19	1,20	1,21	1,25	1,25
∞	1,00	1,04	1,07	1,09	1,11	1,12	1,14	1,15	1,15	1,17	1,18	1,19	1,20	1,21	1,26	1,26

Таблица А.6. Распределение χ^2 (хи-квадрат) *

Степени свободы	Вероятность получения такого или большего значения							
	0,99	0,95	0,90	0,50	0,10	0,05	0,01	0,001
1	0,0002	0,00393	0,0158	0,455	2,706	3,841	6,635	10,827
2	0,0201	0,103	0,211	1,386	4,605	5,991	8,210	13,815
3	0,115	0,352	0,584	2,366	6,251	7,815	11,345	16,268
4	0,297	0,711	1,064	3,357	7,779	9,488	13,277	18,465
5	0,554	1,145	1,610	4,351	9,236	11,070	15,086	20,517
6	0,872	1,635	2,204	5,348	10,645	12,592	16,812	22,457
7	1,239	2,167	2,833	6,346	12,017	14,067	18,475	24,322
8	1,645	2,733	3,490	7,344	13,362	15,507	20,090	26,125
9	2,088	3,325	4,168	8,343	14,684	16,919	21,666	27,877
10	2,558	3,940	4,865	9,342	15,987	18,307	23,209	29,588
11	3,053	4,575	5,578	10,341	17,275	19,675	24,725	31,264
12	3,571	5,226	6,304	11,340	18,549	21,026	26,217	32,909
13	4,107	5,892	7,042	12,340	19,812	22,362	27,688	34,528
14	4,660	6,571	7,790	13,339	21,064	23,685	29,141	36,123
15	5,229	7,261	8,547	14,339	22,307	24,996	30,578	37,697
16	5,812	7,962	9,312	15,338	23,542	26,296	32,000	39,252
17	6,408	8,672	10,085	16,338	24,769	27,587	33,409	40,790
18	7,015	9,390	10,865	17,338	25,989	28,869	34,805	42,312
19	7,633	10,117	11,651	18,338	27,204	30,144	36,191	43,820
20	8,260	10,851	12,443	19,337	28,412	31,410	37,566	45,315
21	8,897	11,591	13,240	20,337	29,615	32,671	38,932	46,797
22	9,542	12,338	14,041	21,337	30,813	33,924	40,289	48,268
23	10,196	13,091	14,848	22,337	32,007	35,172	41,638	49,728
24	10,856	13,848	15,659	23,337	33,196	36,415	42,980	51,179
25	11,524	14,611	16,473	24,337	34,382	37,652	44,314	52,620
26	12,198	15,379	17,292	25,336	35,563	38,885	45,642	54,052
27	12,879	16,151	18,114	26,336	36,741	40,113	46,963	55,476
28	13,565	16,928	18,939	27,336	37,916	41,337	48,278	56,893
29	14,256	17,708	19,768	28,336	39,087	42,557	49,588	58,302
30	14,953	18,493	20,599	29,339	40,256	43,773	50,892	59,703

* Сокращенный вид таблицы IV из книги Фишера и Иейтса (см. подпись под таблицей А.2).

Таблица А.7. Значения коэффициента корреляции (r) для соответствующих уровней значимости *

Степени свободы	Вероятность получения такого или большего значения (уровни значимости)			
	0,1	0,05	0,01	0,001
1	0,9879	0,9969	0,9999	1,0000
2	0,9000	0,9500	0,9900	0,9990
3	0,8054	0,8783	0,9587	0,9912
4	0,7293	0,8114	0,9172	0,9741
5	0,6694	0,7545	0,8745	0,9507

Степени свободы	Вероятность получения такого или большего значения (уровни значимости)			
	0,1	0,05	0,01	0,001
6	0,6215	0,7067	0,8343	0,9249
7	0,5822	0,6664	0,7977	0,8982
8	0,5494	0,6319	0,7646	0,8721
9	0,5214	0,6021	0,7348	0,8471
10	0,4973	0,5760	0,7079	0,8233
11	0,4762	0,5529	0,6835	0,8010
12	0,4575	0,5324	0,6614	0,7800
13	0,4409	0,5139	0,6411	0,7603
14	0,4259	0,4973	0,6226	0,7420
15	0,4124	0,4821	0,6055	0,7246
16	0,4000	0,4683	0,5897	0,7084
17	0,3887	0,4555	0,5751	0,6932
18	0,3783	0,4438	0,5614	0,6787
19	0,3687	0,4329	0,5487	0,6652
20	0,3598	0,4227	0,5368	0,6524
25	0,3233	0,3809	0,4869	0,5974
30	0,2960	0,3494	0,4487	0,5541
35	0,2746	0,3246	0,4182	0,5189
40	0,2573	0,3044	0,3932	0,4896
45	0,2428	0,2875	0,3721	0,4648
50	0,2306	0,2732	0,3541	0,4433
60	0,2108	0,2500	0,3248	0,4078
70	0,1954	0,2319	0,3017	0,3799
80	0,1829	0,2172	0,2830	0,3568
90	0,1726	0,2050	0,2673	0,3375
100	0,1638	0,1946	0,2540	0,3211

* Сокращенный вид таблицы VI из книги Фишера и Иейтса (см. подпись под таблицей А.2.).

Таблица А.8. Преобразование процентов и градусы углов *

%	0	1	2	3	4	5	6	7	8	9
0	0	5,7	8,1	10,0	11,5	12,9	14,2	15,3	16,4	17,5
10	18,4	19,4	20,3	21,1	22,0	22,8	23,6	24,4	25,1	25,8
20	26,6	27,3	28,0	28,7	29,3	30,0	30,7	31,3	31,9	32,6
30	33,2	33,8	34,4	35,1	35,7	36,3	36,9	37,5	38,1	38,6
40	39,2	39,8	40,4	41,0	41,6	42,1	42,7	43,3	43,9	44,4
50	45,0	45,6	46,1	46,7	47,3	47,9	48,4	49,0	49,6	50,2
60	50,8	51,4	51,9	52,5	53,1	53,7	54,3	54,9	55,6	56,2
70	56,8	57,4	58,1	58,7	59,3	60,0	60,7	61,3	62,0	62,7
80	63,4	64,2	64,9	65,6	66,4	67,2	68,0	68,9	69,7	70,6
90	71,6	72,5	73,6	74,7	75,8	77,1	78,5	80,0	81,9	84,3
100	90,0	—	—	—	—	—	—	—	—	—

* Сокращенный вид таблицы X из книги Фишера и Иейтса (см. подпись под таблицей А.2.).

Таблица А.9. Логарифмы

Нату- ральные числа	0	1	2	3	4	5	6	7	8	9	Поправка								
											1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10

Натуральные числа	Поправка									
	0	1	2	3	4	5	6	7	8	9
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445

Нату- ральные числа	Поправка										
	0	1	2	3	4	5	6	7	8	9	
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	4
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9026	4
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	4
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	4
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	4
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	4
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	4
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	4
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	4
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	4

Таблица А.10. Квадраты и квадратные корни

N	N²	√N	√10N	N	N²	√N	√10N
1,00	1,0000	1,00000	3,16228	1,45	2,1025	1,20416	3,80789
1,01	1,0201	1,00499	3,17805	1,46	2,1316	1,20830	3,82099
1,02	1,0404	1,00995	3,19374	1,47	2,1609	1,21244	3,83406
1,03	1,0609	1,01489	3,20936	1,48	2,1904	1,21655	3,84708
1,04	1,0816	1,01980	3,22490	1,49	2,2201	1,22066	3,86005
1,05	1,1025	1,02470	3,24037	1,50	2,2500	1,22474	3,87298
1,06	1,1236	1,02956	3,25576	1,51	2,2801	1,22882	3,88587
1,07	1,1449	1,03441	3,27109	1,52	2,3104	1,23288	3,89872
1,08	1,1664	1,03923	3,28634	1,53	2,3409	1,23693	3,91152
1,09	1,1881	1,04403	3,30151	1,54	2,3716	1,24097	3,92428
1,10	1,2100	1,04881	3,31662	1,55	2,4025	1,24499	3,93700
1,11	1,2321	1,05357	3,33167	1,56	2,4336	1,24900	3,94968
1,12	1,2544	1,05830	3,34664	1,57	2,4649	1,25300	3,96232
1,13	1,2769	1,06301	3,36155	1,58	2,4964	1,25698	3,97492
1,14	1,2996	1,06771	3,37639	1,59	2,5281	1,26095	3,98748
1,15	1,3225	1,07238	3,39116	1,60	2,5600	1,26491	4,00000
1,16	1,3456	1,07703	3,40588	1,61	2,5921	1,26886	4,01248
1,17	1,3689	1,08167	3,42053	1,62	2,6244	1,27279	4,02492
1,18	1,3924	1,08628	3,43511	1,63	2,6569	1,27671	4,03733
1,19	1,4161	1,09087	3,44964	1,64	2,6896	1,28062	4,04969
1,20	1,4400	1,09545	3,46410	1,65	2,7225	1,28452	4,06202
1,21	1,4641	1,10000	3,47851	1,66	2,7556	1,28841	4,07431
1,22	1,4884	1,10454	3,49285	1,67	2,7889	1,29228	4,08656
1,23	1,5129	1,10905	3,50714	1,68	2,8224	1,29615	4,09878
1,24	1,5376	1,11355	3,52136	1,69	2,8561	1,30000	4,11096
1,25	1,5625	1,11803	3,53553	1,70	2,8900	1,30384	4,12311
1,26	1,5876	1,12250	3,54965	1,71	2,9241	1,30767	4,13521
1,27	1,6129	1,12694	3,56371	1,72	2,9584	1,31149	4,14729
1,28	1,6384	1,13137	3,57771	1,73	2,9929	1,31529	4,15933
1,29	1,6641	1,13578	3,59166	1,74	3,0276	1,31909	4,17133
1,30	1,6900	1,14018	3,60555	1,75	3,0625	1,32288	4,18330
1,31	1,7161	1,14455	3,61939	1,76	3,0976	1,32665	4,19524
1,32	1,7424	1,14891	3,63318	1,77	3,1329	1,33041	4,20714
1,33	1,7689	1,15326	3,64692	1,78	3,1684	1,33417	4,21900
1,34	1,7956	1,15758	3,66060	1,79	3,2041	1,33791	4,23084
1,35	1,8225	1,16190	3,67423	1,80	3,2400	1,34164	4,24264
1,36	1,8496	1,16619	3,68782	1,81	3,2761	1,34536	4,25441
1,37	1,8769	1,17047	3,70135	1,82	3,3124	1,34907	4,26615
1,38	1,9044	1,17473	3,71484	1,83	3,3489	1,35277	4,27785
1,39	1,9321	1,17898	3,72827	1,84	3,3856	1,35647	4,28952
1,40	1,9600	1,18322	3,74166	1,85	3,4225	1,36015	4,30116
1,41	1,9881	1,18743	3,75500	1,86	3,4596	1,36382	4,31277
1,42	2,0164	1,19164	3,76829	1,87	3,4969	1,36748	4,32435
1,43	2,0449	1,19583	3,78153	1,88	3,5344	1,37113	4,33590
1,44	2,0736	1,20000	3,79473	1,89	3,5721	1,37477	4,34741

N	N*	$\sqrt{N}$	$\sqrt{10N}$	N	N*	$\sqrt{N}$	$\sqrt{10N}$
1,90	3,6100	1,37840	4,35890	2,35	5,5225	1,53297	4,84768
1,91	3,6481	1,38203	4,37035	2,36	5,5696	1,53623	4,85798
1,92	3,6864	1,38564	4,38178	2,37	5,6169	1,53948	4,86826
1,93	3,7249	1,38924	4,39318	2,38	5,6644	1,54272	4,87852
1,94	3,7636	1,39284	4,40454	2,39	5,7121	1,54596	4,88876
1,95	3,8025	1,39642	4,41588	2,40	5,7600	1,54919	4,89898
1,96	3,8416	1,40000	4,42719	2,41	5,8081	1,55242	4,90918
1,97	3,8809	1,40357	4,43847	2,42	5,8564	1,55563	4,91935
1,98	3,9204	1,40712	4,44972	2,43	5,9049	1,55885	4,92950
1,99	3,9601	1,41067	4,46094	2,44	5,9536	1,56205	4,93964
2,00	4,0000	1,41421	4,47214	2,45	6,0025	1,56525	4,94975
2,01	4,0401	1,41774	4,48330	2,46	6,0516	1,56844	4,95984
2,02	4,0804	1,42127	4,49444	2,47	6,1009	1,57162	4,96991
2,03	4,1209	1,42478	4,50555	2,48	6,1504	1,57480	4,97996
2,04	4,1616	1,42829	4,51664	2,49	6,2001	1,57797	4,98999
2,05	4,2025	1,43178	4,52769	2,50	6,2500	1,58114	5,00000
2,06	4,2436	1,43527	4,53872	2,51	6,3001	1,58430	5,00999
2,07	4,2849	1,43875	4,54973	2,52	6,3504	1,58745	5,01996
2,08	4,3264	1,44222	4,56070	2,53	6,4009	1,59060	5,02991
2,09	4,3681	1,44568	4,57165	2,54	6,4516	1,59374	5,03984
2,10	4,4100	1,44914	4,58258	2,55	6,5025	1,59687	5,04975
2,11	4,4521	1,45258	4,59347	2,56	6,5536	1,60000	5,05964
2,12	4,4944	1,45602	4,60435	2,57	6,6049	1,60312	5,06952
2,13	4,5369	1,45945	4,61519	2,58	6,6564	1,60624	5,07937
2,14	4,5796	1,46287	4,62601	2,59	6,7081	1,60935	5,08920
2,15	4,6225	1,46629	4,63681	2,60	6,7600	1,61245	5,09902
2,16	4,6656	1,46969	4,64758	2,61	6,8121	1,61555	5,10882
2,17	4,7089	1,47309	4,65833	2,62	6,8644	1,61864	5,11859
2,18	4,7524	1,47648	4,66905	2,63	6,9169	1,62173	5,12835
2,19	4,7961	1,47986	4,67974	2,64	6,9696	1,62481	5,13809
2,20	4,8400	1,48324	4,69042	2,65	7,0225	1,62778	5,14782
2,21	4,8841	1,48661	4,70106	2,66	7,0756	1,63095	5,15752
2,22	4,9284	1,48997	4,71169	2,67	7,1289	1,63401	5,16720
2,23	4,9729	1,49332	4,72229	2,68	7,1824	1,63707	5,17687
2,24	5,0176	1,49666	4,73286	2,69	7,2361	1,64012	5,18652
2,25	5,0625	1,50000	4,74342	2,70	7,2900	1,64317	5,19615
2,26	5,1076	1,50333	4,75395	2,71	7,3441	1,64621	5,20577
2,27	5,1529	1,50665	4,76445	2,72	7,3984	1,64924	5,21536
2,28	5,1984	1,50997	4,77493	2,73	7,4529	1,65227	5,22494
2,29	5,2441	1,51327	4,78539	2,74	7,5076	1,65529	5,23450
2,30	5,2900	1,51658	4,79583	2,75	7,5625	1,65831	5,24404
2,31	5,3361	1,51987	4,80625	2,76	7,6176	1,66132	5,25357
2,32	5,3824	1,52315	4,81664	2,77	7,6729	1,66433	5,26308
2,33	5,4289	1,52643	4,82701	2,78	7,7284	1,66733	5,27257
2,34	5,4756	1,52971	4,83735	2,79	7,7841	1,67033	5,28205

N	N [*]	$\sqrt{N}$	$\sqrt{10N}$	N	N [*]	$\sqrt{N}$	$\sqrt{10N}$
2,80	7,8400	1,67332	5,29150	3,25	10,5625	1,80278	5,70088
2,81	7,8961	1,67631	5,30094	3,26	10,6276	1,80555	5,70964
2,82	7,9524	1,67929	5,31037	3,27	10,6929	1,80831	5,71839
2,83	8,0089	1,68226	5,31977	3,28	10,7584	1,81108	5,72713
2,84	8,0656	1,68523	5,32917	3,29	10,8241	1,81384	5,73585
2,85	8,1225	1,68819	5,33854	3,30	10,8900	1,81659	5,74456
2,86	8,1796	1,69115	5,34790	3,31	10,9561	1,81934	5,75326
2,87	8,2369	1,69411	5,35724	3,32	11,0224	1,82209	5,76194
2,88	8,2944	1,69706	5,36656	3,33	11,0889	1,82483	5,77062
2,89	8,3521	1,70000	5,37587	3,34	11,1556	1,82757	5,77927
2,90	8,4100	1,70294	5,38516	3,35	11,2225	1,83030	5,78792
2,91	8,4681	1,70587	5,39444	3,36	11,2896	1,83303	5,79655
2,92	8,5264	1,70880	5,40370	3,37	11,3569	1,83576	5,80517
2,93	8,5849	1,71172	5,41295	3,38	11,4244	1,83848	5,81378
2,94	8,6436	1,71464	5,42218	3,39	11,4921	1,84120	5,82237
2,95	8,7025	1,71756	5,43139	3,40	11,5600	1,84391	5,83095
2,96	8,7616	1,72047	5,44059	3,41	11,6281	1,84662	5,83952
2,97	8,8209	1,72337	5,44977	3,42	11,6964	1,84932	5,84808
2,98	8,8804	1,72627	5,45894	3,43	11,7649	1,85203	5,85662
2,99	8,9401	1,72916	5,46809	3,44	11,8336	1,85472	5,86515
3,00	9,0000	1,73205	5,47723	3,45	11,9025	1,85742	5,87367
3,01	9,0601	1,73494	5,48635	3,46	11,9716	1,86011	5,88218
3,02	9,1204	1,73781	5,49545	3,47	12,0409	1,86279	5,89067
3,03	9,1809	1,74069	5,50454	3,48	12,1104	1,86548	5,89915
3,04	9,2416	1,74356	5,51362	3,49	12,1801	1,86815	5,90762
3,05	9,3025	1,74642	5,52268	3,50	12,2500	1,87083	5,91608
3,06	9,3636	1,74929	5,53173	3,51	12,3201	1,87350	5,92453
3,07	9,4249	1,75214	5,54076	3,52	12,3904	1,87617	5,93296
3,08	9,4864	1,75499	5,54977	3,53	12,4609	1,87883	5,94138
3,09	9,5481	1,75784	5,55878	3,54	12,5316	1,88149	5,94979
3,10	9,6100	1,76068	5,56776	3,55	12,6025	1,88414	5,95819
3,11	9,6721	1,76352	5,57674	3,56	12,6736	1,88680	5,96657
3,12	9,7344	1,76635	5,58570	3,57	12,7449	1,88944	5,97495
3,13	9,7969	1,76918	5,59464	3,58	12,8164	1,89209	5,98331
3,14	9,8596	1,77200	5,60357	3,59	12,8881	1,89473	5,99166
3,15	9,9225	1,77482	5,61249	3,60	12,9600	1,89737	6,00000
3,16	9,9856	1,77764	5,62139	3,61	13,0321	1,90000	6,00833
3,17	10,0489	1,78045	5,63028	3,62	13,1044	1,90263	6,01664
3,18	10,1124	1,78326	5,63915	3,63	13,1769	1,90526	6,02495
3,19	10,1761	1,78606	5,64801	3,64	13,2496	1,90788	6,03324
3,20	10,2400	1,78885	5,65685	3,65	13,3225	1,91050	6,04152
3,21	10,3041	1,79165	5,66569	3,66	13,3956	1,91311	6,04979
3,22	10,3684	1,79444	5,67450	3,67	13,4689	1,91573	6,05805
3,23	10,4329	1,79722	5,68331	3,68	13,5424	1,91833	6,06630
3,24	10,4976	1,80000	5,69210	3,69	13,6161	1,92094	6,07454

N	N [*]	$\sqrt{N}$	$\sqrt{10N}$	N	N [*]	$\sqrt{N}$	$\sqrt{10N}$
3,70	13,6900	1,92354	6,08276	4,15	17,2225	2,03715	6,44205
3,71	13,7641	1,92614	6,09098	4,16	17,3056	2,03961	6,44981
3,72	13,8384	1,92873	6,09918	4,17	17,3889	2,04206	6,45755
3,73	13,9129	1,93132	6,10737	4,18	17,4724	2,04450	6,46529
3,74	13,9876	1,93391	6,11555	4,19	17,5561	2,04695	6,47302
3,75	14,0625	1,93649	6,12372	4,20	17,6400	2,04939	6,48074
3,76	14,1376	1,93907	6,13188	4,21	17,7241	2,05183	6,48845
3,77	14,2129	1,94165	6,14003	4,22	17,8084	2,05426	6,49615
3,78	14,2884	1,94422	6,14817	4,23	17,8929	2,05670	6,50385
3,79	14,3641	1,94679	6,15630	4,24	17,9776	2,05913	6,51153
3,80	14,4400	1,94936	6,16441	4,25	18,0625	2,06155	6,51920
3,81	14,5161	1,95192	6,17252	4,26	18,1476	2,06398	6,52687
3,82	14,5924	1,95448	6,18061	4,27	18,2329	2,06640	6,53452
3,83	14,6689	1,95704	6,18870	4,28	18,3184	2,06882	6,54217
3,84	14,7456	1,95959	6,19677	4,29	18,4041	2,07123	6,54981
3,85	14,8225	1,96214	6,20484	4,30	18,4900	2,07364	6,55744
3,86	14,8996	1,96469	6,21289	4,31	18,5761	2,07605	6,56506
3,87	14,9769	1,96723	6,22093	4,32	18,6624	2,07846	6,57267
3,88	15,0544	1,96977	6,22896	4,33	18,7489	2,08087	6,58027
3,89	15,1321	1,97231	6,23699	4,34	18,8356	2,08327	6,58787
3,90	15,2100	1,97484	6,24500	4,35	18,9225	2,08567	6,59545
3,91	15,2881	1,97737	6,25300	4,36	19,0096	2,08806	6,60303
3,92	15,3664	1,97990	6,26099	4,37	19,0969	2,09045	6,61060
3,93	15,4449	1,98242	6,26897	4,38	19,1844	2,09284	6,61816
3,94	15,5236	1,98494	6,27694	4,39	19,2721	2,09523	6,62571
3,95	15,6025	1,98746	6,28490	4,40	19,3600	2,09762	6,63325
3,96	15,6816	1,98997	6,29285	4,41	19,4481	2,10000	6,64078
3,97	15,7609	1,99249	6,30079	4,42	19,5364	2,10238	6,64831
3,98	15,8404	1,99499	6,30872	4,43	19,6249	2,10476	6,65582
3,99	15,9201	1,99750	6,31664	4,44	19,7136	2,10713	6,66333
4,00	16,0000	2,00000	6,32456	4,45	19,8025	2,10950	6,67083
4,01	16,0801	2,00250	6,33246	4,46	19,8916	2,11187	6,67832
4,02	16,1604	2,00499	6,34035	4,47	19,9809	2,11424	6,68581
4,03	16,2409	2,00749	6,34823	4,48	20,0704	2,11660	6,69328
4,04	16,3216	2,00998	6,35610	4,49	20,1601	2,11896	6,70075
4,05	16,4025	2,01246	6,36396	4,50	20,2500	2,12132	6,70820
4,06	16,4836	2,01494	6,37181	4,51	20,3401	2,12368	6,71565
4,07	16,5649	2,01742	6,37966	4,52	20,4304	2,12603	6,72309
4,08	16,6464	2,01990	6,38749	4,53	20,5209	2,12838	6,73053
4,09	16,7281	2,02237	6,39531	4,54	20,6116	2,13073	6,73795
4,10	16,8100	2,02485	6,40312	4,55	20,7025	2,13307	6,74537
4,11	16,8921	2,02731	6,41093	4,56	20,7936	2,13542	6,75278
4,12	16,9744	2,02978	6,41872	4,57	20,8849	2,13776	6,76018
4,13	17,0569	2,03224	6,42651	4,58	20,9764	2,14009	6,76757
4,14	17,1396	2,03470	6,43428	4,59	21,0681	2,14243	6,77495

N	N²	√N	√10N	N	N²	√N	√10N
4,60	21,1600	2,14476	6,78233	5,05	25,5025	2,24722	7,10634
4,61	21,2521	2,14709	6,78970	5,06	25,6036	2,24944	7,11337
4,62	21,3444	2,14942	6,79706	5,07	25,7049	2,25167	7,12039
4,63	21,4369	2,15174	6,80441	5,08	25,8064	2,25389	7,12741
4,64	21,5296	2,15407	6,81175	5,09	25,9081	2,25610	7,13442
4,65	21,6225	2,15639	6,81909	5,10	26,0100	2,25832	7,14143
4,66	21,7156	2,15870	6,82642	5,11	26,1121	2,26053	7,14843
4,67	21,8089	2,16102	6,83374	5,12	26,2144	2,26274	7,15542
4,68	21,9024	2,16333	6,84105	5,13	26,3169	2,26495	7,16240
4,69	21,9961	2,16564	6,84836	5,14	26,4196	2,26716	7,16938
4,70	22,0900	2,16795	6,85565	5,15	26,5225	2,26936	7,17635
4,71	22,1841	2,17025	6,86294	5,16	26,6256	2,27156	7,18331
4,72	22,2784	2,17256	6,87023	5,17	26,7289	2,27376	7,19027
4,73	22,3729	2,17486	6,87750	5,18	26,8324	2,27596	7,19722
4,74	22,4676	2,17715	6,88477	5,19	26,9361	2,27816	7,20417
4,75	22,5625	2,17945	6,89202	5,20	27,0400	2,28035	7,21110
4,76	22,6576	2,18174	6,89928	5,21	27,1441	2,28254	7,21803
4,77	22,7529	2,18403	6,90652	5,22	27,2484	2,28473	7,22496
4,78	22,8484	2,18632	6,91375	5,23	27,3529	2,28692	7,23187
4,79	22,9441	2,18861	6,92098	5,24	27,4576	2,28910	7,23878
4,80	23,0400	2,19089	6,92820	5,25	27,5625	2,29129	7,24569
4,81	23,1361	2,19317	6,93542	5,26	27,6676	2,29347	7,25259
4,82	23,2324	2,19545	6,94262	5,27	27,7729	2,29565	7,25948
4,83	23,3289	2,19773	6,94982	5,28	27,8784	2,29783	7,26636
4,84	23,4256	2,20000	6,95701	5,29	27,9841	2,30000	7,27324
4,85	23,5225	2,20227	6,96419	5,30	28,0900	2,30217	7,28011
4,86	23,6196	2,20454	6,97137	5,31	28,1961	2,30434	7,28697
4,87	23,7169	2,20681	6,97854	5,32	28,3024	2,30651	7,29383
4,88	23,8144	2,20907	6,98570	5,33	28,4089	2,30868	7,30068
4,89	23,9121	2,21133	6,99285	5,34	28,5156	2,31084	7,30753
4,90	24,0100	2,21359	7,00000	5,35	28,6225	2,31301	7,31437
4,91	24,1081	2,21585	7,00714	5,36	28,7296	2,31517	7,32120
4,92	24,2064	2,21811	7,01427	5,37	28,8369	2,31733	7,32803
4,93	24,3049	2,22036	7,02140	5,38	28,9444	2,31948	7,33485
4,94	24,4036	2,22261	7,02851	5,39	29,0521	2,32164	7,34166
4,95	24,5025	2,22486	7,03562	5,40	29,1600	2,32379	7,34847
4,96	24,6016	2,22711	7,04273	5,41	29,2681	2,32594	7,35527
4,97	24,7009	2,22935	7,04982	5,42	29,3764	2,32809	7,36206
4,98	24,8004	2,23159	7,05691	5,43	29,4849	2,33024	7,36885
4,99	24,9001	2,23383	7,06399	5,44	29,5936	2,33238	7,37564
5,00	25,0000	2,23607	7,07107	5,45	29,7025	2,33452	7,38241
5,01	25,1001	2,23830	7,07814	5,46	29,8116	2,33666	7,38918
5,02	25,2004	2,24054	7,08520	5,47	29,9209	2,33880	7,39594
5,03	25,3009	2,24277	7,09225	5,48	30,0304	2,34094	7,40270
5,04	25,4016	2,24499	7,09930	5,49	30,1401	2,34307	7,40945

N	N°	VN	V $\sqrt{10N}$	N	N°	VN	V $\sqrt{10N}$
5,50	30,2500	2,34521	7,41620	5,95	35,4025	2,43926	7,71362
5,51	30,3601	2,34734	7,42294	5,96	35,5216	2,44131	7,72010
5,52	30,4704	2,34947	7,42967	5,97	35,6409	2,44336	7,72658
5,53	30,5809	2,35160	7,43640	5,98	35,7604	2,44540	7,73305
5,54	30,6916	2,35372	7,44312	5,99	35,8801	2,44745	7,73951
5,55	30,8025	2,35584	7,44983	6,00	36,0000	2,44949	7,74597
5,56	30,9136	2,35797	7,45654	6,01	36,1201	2,45153	7,75242
5,57	31,0249	2,36008	7,46324	6,02	36,2404	2,45357	7,75887
5,58	31,1364	2,36220	7,46994	6,03	36,3609	2,45561	7,76531
5,59	31,2481	2,36432	7,47663	6,04	36,4816	2,45764	7,77174
5,60	31,3600	2,36643	7,48331	6,05	36,6025	2,45967	7,77817
5,61	31,4721	2,36854	7,48999	6,06	36,7236	2,46171	7,78460
5,62	31,5844	2,37065	7,49667	6,07	36,8449	2,46374	7,79102
5,63	31,6969	2,37276	7,50333	6,08	36,9664	2,46577	7,79744
5,64	31,8096	2,37487	7,50999	6,09	37,0881	2,46779	7,80385
5,65	31,9225	2,37697	7,51665	6,10	37,2100	2,46982	7,81025
5,66	32,0356	2,37908	7,52330	6,11	37,3321	2,47184	7,81665
5,67	32,1489	2,38118	7,52994	6,12	37,4544	2,47386	7,82304
5,68	32,2624	2,38328	7,53658	6,13	37,5769	2,47588	7,82943
5,69	32,3761	2,38537	7,54321	6,14	37,6996	2,47790	7,83582
5,70	32,4900	2,38747	7,54983	6,15	37,8225	2,47992	7,84219
5,71	32,6041	2,38956	7,55645	6,16	37,9456	2,48193	7,84857
5,72	32,7184	2,39165	7,56307	6,17	38,0689	2,48395	7,85493
5,73	32,8329	2,39374	7,56968	6,18	38,1924	2,48596	7,86130
5,74	32,9476	2,39583	7,57628	6,19	38,3161	2,48797	7,86766
5,75	33,0625	2,39792	7,58288	6,20	38,4400	2,48998	7,87401
5,76	33,1776	2,40000	7,58947	6,21	38,5641	2,49199	7,88036
5,77	33,2929	2,40208	7,59605	6,22	38,6884	2,49399	7,88670
5,78	33,4084	2,40416	7,60263	6,23	38,8129	2,49600	7,89303
5,79	33,5241	2,40624	7,60920	6,24	38,9376	2,49800	7,89937
5,80	33,6400	2,40832	7,61577	6,25	39,0625	2,50000	7,90569
5,81	33,7561	2,41039	7,62234	6,26	39,1876	2,50200	7,91202
5,82	33,8724	2,41247	7,62889	6,27	39,3129	2,50400	7,91833
5,83	33,9889	2,41454	7,63544	6,28	39,4384	2,50599	7,92465
5,84	34,1056	2,41661	7,64199	6,29	39,5641	2,50799	7,93095
5,85	34,2225	2,41868	7,64853	6,30	39,6900	2,50998	7,93725
5,86	34,3396	2,42074	7,65506	6,31	39,8161	2,51197	7,94355
5,87	34,4569	2,42281	7,66159	6,32	39,9424	2,51396	7,94984
5,88	34,5744	2,42487	7,66812	6,33	40,0689	2,51595	7,95613
5,89	34,6921	2,42693	7,67463	6,34	40,1956	2,51794	7,96241
5,90	34,8100	2,42899	7,68115	6,35	40,3225	2,51992	7,96869
5,91	34,9281	2,43105	7,68765	6,36	40,4496	2,52190	7,97496
5,92	35,0464	2,43311	7,69415	6,37	40,5769	2,52389	7,98123
5,93	35,1649	2,43516	7,70065	6,38	40,7044	2,52587	7,98749
5,94	35,2836	2,43721	7,70714	6,39	40,8321	2,52784	7,99375

N	N ²	$\sqrt{N}$	$\sqrt{10N}$	N	N ²	$\sqrt{N}$	$\sqrt{10N}$
6,40	40,9600	2,52982	8,00000	6,85	46,9225	2,61725	8,27647
6,41	41,0881	2,53180	8,00625	6,86	47,0596	2,61916	8,28251
6,42	41,2164	2,53377	8,01249	6,87	47,1969	2,62107	8,28855
6,43	41,3449	2,53574	8,01873	6,88	47,3344	2,62298	8,29458
6,44	41,4736	2,53772	8,02496	6,89	47,4721	2,62488	8,30060
6,45	41,6025	2,53969	8,03119	6,90	47,6100	2,62679	8,30662
6,46	41,7316	2,54165	8,03741	6,91	47,7481	2,62869	8,31264
6,47	41,8609	2,54362	8,04363	6,92	47,8864	2,63059	8,31865
6,48	41,9904	2,54558	8,04984	6,93	48,0249	2,63249	8,32466
6,49	42,1201	2,54755	8,05605	6,94	48,1636	2,63439	8,33067
6,50	42,2500	2,54951	8,06226	6,95	48,3025	2,63629	8,33667
6,51	42,3801	2,55147	8,06846	6,96	48,4416	2,63818	8,34266
6,52	42,5104	2,55343	8,07465	6,97	48,5809	2,64008	8,34865
6,53	42,6409	2,55539	8,08084	6,98	48,7204	2,64197	8,35464
6,54	42,7716	2,55734	8,08703	6,99	48,8601	2,64386	8,36062
6,55	42,9025	2,55930	8,09321	7,00	49,0000	2,64575	8,36660
6,56	43,0336	2,56125	8,09938	7,01	49,1401	2,64764	8,37257
6,57	43,1649	2,56320	8,10555	7,02	49,2804	2,64953	8,37854
6,58	43,2964	2,56515	8,11172	7,03	49,4209	2,65141	8,38451
6,59	43,4281	2,56710	8,11788	7,04	49,5616	2,65330	8,39047
6,60	43,5600	2,56905	8,12404	7,05	49,7025	2,65518	8,39643
6,61	43,6921	2,57099	8,13019	7,06	49,8436	2,65707	8,40238
6,62	43,8244	2,57294	8,13634	7,07	49,9849	2,65895	8,40833
6,63	43,9569	2,57488	8,14248	7,08	50,1264	2,66083	8,41427
6,64	44,0896	2,57682	8,14862	7,09	50,2681	2,66271	8,42021
6,65	44,2225	2,57876	8,15475	7,10	50,4100	2,66458	8,42615
6,66	44,3556	2,58070	8,16088	7,11	50,5521	2,66646	8,43208
6,67	44,4889	2,58263	8,16701	7,12	50,6944	2,66833	8,43801
6,68	44,6224	2,58457	8,17313	7,13	50,8369	2,67021	8,44393
6,69	44,7561	2,58650	8,17924	7,14	50,9796	2,67208	8,44985
6,70	44,8900	2,58844	8,18535	7,15	51,1225	2,67395	8,45577
6,71	45,0241	2,59037	8,19146	7,16	51,2656	2,67582	8,46168
6,72	45,1584	2,59230	8,19756	7,17	51,4089	2,67769	8,46759
6,73	45,2929	2,59422	8,20366	7,18	51,5524	2,67955	8,47349
6,74	45,4276	2,59615	8,20975	7,19	51,6961	2,68142	8,47939
6,75	45,5625	2,59808	8,21584	7,20	51,8400	2,68328	8,48528
6,76	45,6976	2,60000	8,22192	7,21	51,9841	2,68514	8,49117
6,77	45,8329	2,60192	8,22800	7,22	52,1284	2,68701	8,49706
6,78	45,9684	2,60384	8,23408	7,23	52,2729	2,68887	8,50294
6,79	46,1041	2,60576	8,24015	7,24	52,4176	2,69072	8,50882
6,80	46,2400	2,60768	8,24621	7,25	52,5625	2,69258	8,51469
6,81	46,3761	2,60960	8,25227	7,26	52,7076	2,69444	8,52056
6,82	46,5124	2,61151	8,25833	7,27	52,8529	2,69629	8,52643
6,83	46,6489	2,61343	8,26438	7,28	52,9984	2,69815	8,53229
6,84	46,7856	2,61534	8,27043	7,29	53,1441	2,70000	8,53815

N	N ^a	$\sqrt{N}$	$\sqrt{10N}$	N	N ^a	$\sqrt{N}$	$\sqrt{10N}$
7,30	53,2900	2,70185	8,54400	7,75	60,0625	2,78388	8,80341
7,31	53,4361	2,70370	8,54985	7,76	60,2176	2,78568	8,80909
7,32	53,5824	2,70555	8,55570	7,77	60,3729	2,78747	8,81476
7,33	53,7289	2,70740	8,56154	7,78	60,5284	2,78927	8,82043
7,34	53,8756	2,70924	8,56738	7,79	60,6841	2,79106	8,82610
7,35	54,0225	2,71109	8,57321	7,80	60,8400	2,79285	8,83176
7,36	54,1696	2,71293	8,57904	7,81	60,9961	2,79464	8,83742
7,37	54,3169	2,71477	8,58487	7,82	61,1524	2,79643	8,84308
7,38	54,4644	2,71662	8,59069	7,83	61,3089	2,79821	8,84873
7,39	54,6121	2,71846	8,59651	7,84	61,4656	2,80000	8,85438
7,40	54,7600	2,72029	8,60233	7,85	61,6225	2,80179	8,86002
7,41	54,9081	2,72213	8,60814	7,86	61,7796	2,80357	8,86566
7,42	55,0564	2,72397	8,61394	7,87	61,9369	2,80535	8,87130
7,43	55,2049	2,72580	8,61974	7,88	62,0944	2,80713	8,87694
7,44	55,3536	2,72764	8,62554	7,89	62,2521	2,80891	8,88257
7,45	55,5025	2,72947	8,63134	7,90	62,4100	2,81069	8,88819
7,46	55,6516	2,73130	8,63713	7,91	62,5681	2,81247	8,89382
7,47	55,8009	2,73313	8,64292	7,92	62,7264	2,81425	8,89944
7,48	55,9504	2,73496	8,64870	7,93	62,8849	2,81603	8,90505
7,49	56,1001	2,73679	8,65448	7,94	63,0436	2,81780	8,91067
7,50	56,2500	2,73861	8,66025	7,95	63,2025	2,81957	8,91628
7,51	56,4001	2,74044	8,66603	7,96	63,3616	2,82135	8,92188
7,52	56,5504	2,74226	8,67179	7,97	63,5209	2,82312	8,92749
7,53	56,7009	2,74408	8,67756	7,98	63,6804	2,82489	8,93308
7,54	56,8516	2,74591	8,68332	7,99	63,8401	2,82666	8,93868
7,55	57,0025	2,74773	8,68907	8,00	64,0000	2,82843	8,94427
7,56	57,1536	2,74955	8,69483	8,01	64,1601	2,83019	8,94986
7,57	57,3049	2,75136	8,70057	8,02	64,3204	2,83196	8,95545
7,58	57,4564	2,75318	8,70632	8,03	64,4809	2,83373	8,96103
7,59	57,6081	2,75500	8,71206	8,04	64,6416	2,83549	8,96660
7,60	57,7600	2,75681	8,71780	8,05	64,8025	2,83725	8,97218
7,61	57,9121	2,75862	8,72353	8,06	64,9636	2,83901	8,97775
7,62	58,0644	2,76043	8,72926	8,07	65,1249	2,84077	8,98332
7,63	58,2169	2,76225	8,73499	8,08	65,2864	2,84253	8,98888
7,64	58,3696	2,76405	8,74071	8,09	65,4481	2,84429	8,99444
7,65	58,5225	2,76586	8,74643	8,10	65,6100	2,84605	9,00000
7,66	58,6756	2,76767	8,75214	8,11	65,7721	2,84781	9,00555
7,67	58,8289	2,76948	8,75785	8,12	65,9344	2,84956	9,01110
7,68	58,9824	2,77128	8,76356	8,13	66,0969	2,85132	9,01665
7,69	59,1361	2,77308	8,76926	8,14	66,2596	2,85307	9,02219
7,70	59,2900	2,77489	8,77496	8,15	66,4225	2,85482	9,02774
7,71	59,4441	2,77669	8,78066	8,16	66,5856	2,85657	9,03327
7,72	59,5984	2,77849	8,78635	8,17	66,7489	2,85832	9,03881
7,73	59,7529	2,78029	8,79204	8,18	66,9124	2,86007	9,04434
7,74	59,9076	2,78209	8,79773	8,19	67,0761	2,86182	9,04986

N	N ²	$\sqrt{N}$	$\sqrt{10N}$	N	N ²	$\sqrt{N}$	$\sqrt{10N}$
8,20	67,2400	2,86356	9,05539	8,65	74,8225	2,94109	9,30054
8,21	67,4041	2,86531	9,06091	8,66	74,9956	2,94279	9,30591
8,22	67,5684	2,86705	9,06642	8,67	75,1680	2,94449	9,31128
8,23	67,7329	2,86880	9,07193	8,68	75,3424	2,94618	9,31665
8,24	67,8976	2,87054	9,07744	8,69	75,5161	2,94788	9,32202
8,25	68,0625	2,87228	9,08295	8,70	75,6900	2,94958	9,32738
8,26	68,2276	2,87402	9,08845	8,71	75,8641	2,95127	9,33274
8,27	68,3929	2,87576	9,09395	8,72	76,0384	2,95296	9,33809
8,28	68,5584	2,87750	9,09945	8,73	76,2129	2,95466	9,34345
8,29	68,7241	2,87924	9,10494	8,74	76,3876	2,95635	9,34880
8,30	68,8900	2,88097	9,11043	8,75	76,5625	2,95804	9,35414
8,31	69,0561	2,88271	9,11592	8,76	76,7376	2,95973	9,35949
8,32	69,2224	2,84444	9,12140	8,77	76,9129	2,96142	9,36483
8,33	69,3889	2,88617	9,12688	8,78	77,0884	2,96311	9,37017
8,34	69,5556	2,88791	9,13236	8,79	77,2641	2,96479	9,37550
8,35	69,7225	2,88964	9,13783	8,80	77,4400	2,96648	9,38083
8,36	69,8896	2,89137	9,14330	8,81	77,6161	2,96816	9,38616
8,37	70,0569	2,89310	9,14877	8,82	77,7924	2,96985	9,39149
8,38	70,2244	2,89482	9,15423	8,83	77,9689	2,97153	9,39681
8,39	70,3921	2,89655	9,15969	8,84	78,1456	2,97321	9,40213
8,40	70,5600	2,89828	9,16515	8,85	78,3225	2,97489	9,40744
8,41	70,7281	2,90000	9,17061	8,86	78,4996	2,97658	9,41276
8,42	70,8964	2,90172	9,17606	8,87	78,6769	2,97825	9,41807
8,43	71,0649	2,90345	9,18150	8,88	78,8544	2,97993	9,42338
8,44	71,2336	2,90517	9,18695	8,89	79,0321	2,98161	9,42868
8,45	71,4025	2,90639	9,19239	8,90	79,2100	2,98329	9,43398
8,46	71,5716	2,90861	9,19783	8,91	79,3881	2,98496	9,43928
8,47	71,7409	2,91033	9,20326	8,92	79,5664	2,98664	9,44458
8,48	71,9104	2,91204	9,20869	8,93	79,7449	2,98831	9,44987
8,49	72,0801	2,91376	9,21412	8,94	79,9236	2,98998	9,45516
8,50	72,2500	2,91548	9,21954	8,95	80,1025	2,99166	9,46044
8,51	72,4201	2,91719	9,22497	8,96	80,2816	2,99333	9,46573
8,52	72,5904	2,91890	9,23038	8,97	80,4609	2,99500	9,47101
8,53	72,7609	2,92062	9,23580	8,98	80,6404	2,99666	9,47629
8,54	72,9316	2,92233	9,24121	8,99	80,8201	2,99833	9,48156
8,55	73,1025	2,92404	9,24662	9,00	81,0000	3,00000	9,48683
8,56	73,2736	2,92575	9,25203	9,01	81,1801	3,00167	9,49210
8,57	73,4449	2,92746	9,25743	9,02	81,3604	3,00333	9,49737
8,58	73,6164	2,92916	9,26283	9,03	81,5409	3,00500	9,50263
8,59	73,7881	2,93087	9,26823	9,04	81,7216	3,00666	9,50789
8,60	73,9600	2,93258	9,27362	9,05	81,9025	3,00832	9,51315
8,61	74,1321	2,93428	9,27901	9,06	82,0836	3,00998	9,51840
8,62	74,3044	2,93598	9,28440	9,07	82,2649	3,01164	9,52365
8,63	74,4769	2,93769	9,28978	9,08	82,4464	3,01330	9,52890
8,64	74,6496	2,93939	9,29516	9,09	82,6281	3,01496	9,53415

N	N [*]	$\sqrt{N}$	$\sqrt{10N}$	N	N [*]	$\sqrt{N}$	$\sqrt{10N}$
9,10	82,8100	3,01662	9,53939	9,55	91,2025	3,09031	9,77241
9,11	82,9921	3,01828	9,54463	9,56	91,3936	3,09192	9,77753
9,12	83,1744	3,01993	9,54987	9,57	91,5849	3,09354	9,78264
9,13	83,3569	3,02159	9,55510	9,58	91,7764	3,09516	9,78775
9,14	83,5396	3,02324	9,56033	9,59	91,9681	3,09677	9,79285
9,15	83,7225	3,02490	9,56556	9,60	92,1600	3,09839	9,79796
9,16	83,9056	3,02655	9,57079	9,61	92,3521	3,10000	9,80306
9,17	84,0889	3,02820	9,57601	9,62	92,5444	3,10161	9,80816
9,18	84,2724	3,02985	9,58123	9,63	92,7369	3,10322	9,81326
9,19	84,4561	3,03150	9,58645	9,64	92,9296	3,10483	9,81835
9,20	84,6400	3,03315	9,59166	9,65	93,1225	3,10644	9,82344
9,21	84,8241	3,03480	9,59687	9,66	93,3156	3,10805	9,82853
9,22	85,0084	3,03645	9,60208	9,67	93,5089	3,10966	9,83362
9,23	85,1929	3,03809	9,60729	9,68	93,7024	3,11127	9,83870
9,24	85,3776	3,03974	9,61249	9,69	93,8961	3,11288	9,84378
9,25	85,5625	3,04138	9,61769	9,70	94,0900	3,11448	9,84886
9,26	85,7476	3,04302	9,62289	9,71	94,2841	3,11609	9,85393
9,27	85,9329	3,04467	9,62808	9,72	94,4784	3,11769	9,85901
9,28	86,1184	3,04631	9,63328	9,73	94,6729	3,11929	9,86408
9,29	86,3041	3,04795	9,63846	9,74	94,8676	3,12090	9,86914
9,30	86,4900	3,04959	9,64365	9,75	95,0625	3,12250	9,87421
9,31	86,6761	3,05123	9,64883	9,76	95,2576	3,12410	9,87927
9,32	86,8624	3,05287	9,65401	9,77	95,4529	3,12570	9,88433
9,33	87,0489	3,05450	9,65919	9,78	95,6484	3,12730	9,88939
9,34	87,2356	3,05614	9,66437	9,79	95,8441	3,12890	9,89444
9,35	87,4225	3,05778	9,66954	9,80	96,0400	3,13050	9,89949
9,36	87,6096	3,05941	9,67471	9,81	96,2361	3,13209	9,90454
9,37	87,7969	3,06105	9,67988	9,82	96,4324	3,13369	9,90959
9,38	87,9844	3,06268	9,68504	9,83	96,6289	3,13528	9,91464
9,39	88,1721	3,06431	9,69020	9,84	96,8256	3,13688	9,91968
9,40	88,3600	3,06594	9,69536	9,85	97,0225	3,13847	9,92472
9,41	88,5481	3,06757	9,70052	9,86	97,2196	3,14006	9,92975
9,42	88,7364	3,06920	9,70567	9,87	97,4169	3,14166	9,93479
9,43	88,9249	3,07083	9,71082	9,88	97,6144	3,14325	9,93982
9,44	89,1136	3,07246	9,71597	9,89	97,8121	3,14484	9,94485
9,45	89,3025	3,07409	9,72111	9,90	98,0100	3,14643	9,94987
9,46	89,4916	3,07571	9,72625	9,91	98,2081	3,14802	9,95490
9,47	89,6809	3,07734	9,73139	9,92	98,4064	3,14960	9,95992
9,48	89,8704	3,07896	9,73653	9,93	98,6049	3,15119	9,96494
9,49	90,0601	3,08058	9,74166	9,94	98,8036	3,15278	9,96995
9,50	90,2500	3,08221	9,74679	9,95	99,0025	3,15436	9,97497
9,51	90,4401	3,08383	9,75192	9,96	99,2016	3,15595	9,97998
9,52	90,6304	3,08545	9,75705	9,97	99,4009	3,15753	9,98499
9,53	90,8209	3,08707	9,76217	9,98	99,6004	3,15911	9,98999
9,54	91,0116	3,08869	9,76729	9,99	99,8001	3,16070	9,99500

Таблица А.11. Коэффициенты, делители и К-величины для расчленения суммы квадратов и построения кривой четвертой степени на основе значений переменных с равновеликими интервалами

п: 3		4			5				6			
c_1	c_2	c_1	c_2	c_3	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
-1	1	-3	1	-1	-2	2	-1	1	-5	5	-5	1
0	-2	-1	-1	3	-1	-1	2	-4	-3	-1	7	-3
1	1	1	-1	-3	0	-2	0	6	-1	-4	4	2
		3	1	1	1	-1	-2	-4	1	-4	-4	2
					2	2	1	1	3	-1	-7	-3
									5	5	5	1
Делители												
2	6	20	20		10	14	10	70	70	84	180	28
K_1	1/3		5/16					1/7				5/96
K_2	1/2		1/20					1/10				1/70
K_3			41/240					17/60				101/4320
K_4	1/2		1/16					1/14				1/224
K_5			1/48					1/12				1/864
K_6								1/24				1/768
K_7								31/168				95/2688
K_8								3/35				27/256

п: 7				8				9			
c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
-3	5	-1	3	-7	7	-7	7	-4	28	-14	14
-2	0	1	-7	-5	1	5	-13	-3	7	7	-21
-1	-3	1	1	-3	-3	7	-3	-2	-8	13	-11
0	-4	0	6	-1	-5	3	9	-1	-17	9	9
1	-3	-1	1	1	-5	-3	9	0	-20	0	18
2	0	-1	-7	3	-3	-7	-3	1	-17	-9	9
3	5	1	3	5	1	-5	-13	2	-8	-13	-11
				7	7	7	7	3	7	-7	-21
								4	28	14	14
Делители											
28	84	6	154	168	168	264	616	60	2772	990	2002
K_1			1/21				1/32				5/693
K_2			1/28				1/68				1.60
K_3			7/36				37/3168				59/5940
K_4			1/84				1/672				1/924
K_5			1/36				1/3168				1/1188
K_6			1/264				1/16896				1/3432
K_7			67/1848				179/59136				115/24024
K_8			3/77				9/512				9/1001

n: 10				11			
c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
-9	6	-42	18	-5	15	-30	6
-7	2	14	-22	-4	6	6	-6
-5	-1	35	-17	-3	-1	22	-6
-3	-3	31	3	-2	-6	23	-1
-1	-4	12	18	-1	-9	14	4
1	-4	-12	18	0	-10	0	6
3	-3	-31	3	1	-9	-14	4
5	-1	-35	-17	2	-6	-23	-1
7	2	-14	-22	3	-1	-22	-6
9	6	42	18	4	6	-6	-6
				5	15	30	6
Делители							
330	132	8580	2860	110	858	4290	286
K_1			1/32			5/429	
K_2			1/330			1/110	
K_3			293/205920			89/25740	
K_4			1/1056			1/858	
K_5			1/41184			1/5148	
K_6			1/109824			1/3432	
K_7			41/54912			25/3432	
K_8			9/1280			3/143	

n: 12				13			
c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
-11	55	-33	33	-6	22	-11	99
-9	25	3	-27	-5	11	0	-66
-7	1	21	-33	-4	2	6	-96
-5	-17	25	-13	-3	-5	8	-54
-3	-29	19	12	-2	-10	7	11
-1	-35	7	28	-1	-13	4	64
1	-35	-7	28	0	-14	0	84
3	-29	-19	12	1	-13	-4	64
5	-17	-25	-13	2	-10	-7	11
7	1	-21	-33	3	-5	-8	-54
9	25	-3	-27	4	2	-6	-96
11	55	33	33	5	11	0	-66
				6	22	11	99
Делители							
572	12012	5148	8008	182	2002	572	68068
K_1			1/336				1/143
K_2			1/572				1/182
K_3			85/61776				25/3432
K_4			1/16016				1/2002
K_5			1/61776				1/3432
K_6			1/439296				1/116688
K_7			419/1537536				19/62832
K_8			27/7168				3/2431

n:	14				15			
	c ₁	c ₂	c ₃	c ₄	c ₁	c ₂	c ₃	c ₄
	-13	13	-143	143	-7	91	-91	1001
	-11	7	-11	-77	-6	52	-13	-429
	-9	2	66	-132	-5	19	35	-869
	-7	-2	98	-92	-4	-8	58	-704
	-5	-5	95	-13	-3	-29	61	-249
	-3	-7	63	63	-2	-44	49	251
	-1	-8	24	108	-1	-53	27	621
	1	-8	-24	108	0	-56	0	756
	3	-7	-67	63	1	-53	-27	621
	5	-5	-95	-13	2	-44	-49	251
	7	-2	-98	-92	3	-29	-61	-249
	9	2	-66	-132	4	-8	-58	-704
	11	7	11	-77	5	19	-35	-869
	13	13	143	143	6	52	13	-429
					7	91	91	1001
Делители	910	728	97240	136136	280	37128	39780	6466460
K ₁				5/448				1/663
K ₂				1/910				1/280
K ₃				581/2333760				167/238680
K ₄				1/5824				1/12376
K ₅				1/466752				1/47736
K ₆				1/3734016				1/2217072
K ₇				575/13069056				331/15519504
K ₈				3/3584				27/230945

n:	16				17			
	c ₁	c ₂	c ₃	c ₄	c ₁	c ₂	c ₃	c ₄
	-15	35	-455	273	-8	40	-28	52
	-13	21	-91	-91	-7	25	-7	-13
	-11	9	143	-221	-6	12	7	-39
	-9	-1	267	-201	-5	1	15	-39
	-7	-9	301	-101	-4	-8	18	-24
	-5	-15	265	23	-3	-15	17	-3
	-3	-19	179	129	-2	-20	13	17
	-1	-21	63	189	-1	-23	7	31
	1	-21	-63	189	0	-24	0	36
	3	-19	-179	129	1	-23	-7	31
	5	-15	-265	23	2	-20	-13	17
	7	-9	-301	-101	3	-15	-17	-3
	9	-1	-267	-201	4	-8	-18	-24
	11	9	-143	-221	5	1	-15	-39
	13	21	91	-91	6	12	-7	-39
	15	35	455	273	7	25	7	-13
					8	40	28	52
Делители	1360	5712	1007760	470288	408	7752	3876	16796
K ₁				5/1344				1/323
K ₂				1/1360				1/408
K ₃				761/12093120				43/23256
K ₄				1/22848				1/7752
K ₅				1/2418624				1/23256
K ₆				1/12899328				1/201552
K ₇				755/45147648				61/201552
K ₈				3/7168				9/4199

п: 18				19			
c ₁	c ₂	c ₃	c ₄	c ₁	c ₂	c ₃	c ₄
-17	68	-68	68	-9	51	-204	612
-15	44	-20	-12	-8	34	-68	-68
-13	23	13	-47	-7	19	28	-388
-11	5	33	-51	-6	6	89	-453
-9	-10	42	-36	-5	-5	120	-354
-7	-22	42	-12	-4	-14	126	-168
-5	-31	35	13	-3	-21	112	42
-3	-37	23	33	-2	-26	83	227
-1	-40	8	44	-1	-29	44	352
1	-40	-8	44	0	-30	0	396
3	-37	-23	33	1	-29	-44	352
5	-31	-35	13	2	-26	-83	227
7	-22	-42	-12	3	-21	-112	42
9	-10	-42	-36	4	-14	-126	-168
11	5	-33	-51	5	-5	-120	-354
13	23	-13	-47	6	6	-89	-453
15	44	20	-12	7	19	-28	-388
17	68	68	68	8	34	68	-69
				9	51	204	612
Делители							
1938	23256	23256	28424	570	13466	213180	2288132
K ₁			1/576				5/2261
K ₂			1/1938				1/570
K ₃			193/558144				269/1279080
K ₄			1/62016				1/13566
K ₅			1/558144				1/255816
K ₆			1/5457408				1/3922512
K ₇			137/2728704				535/27457584
K ₈			9/5632				9/52003

n: 20				21			
c ₁	c ₂	c ₃	c ₄	c ₁	c ₂	c ₃	c ₄
-19	57	-969	1938	-10	190	-285	969
-17	39	-357	-102	-9	133	-114	0
-15	23	85	-1122	-8	82	12	-510
-13	9	377	-1402	-7	37	98	-630
-11	-3	539	-1187	-6	-2	149	-615
-9	-13	591	-687	-5	-35	170	-406
-7	-21	553	-77	-4	-62	166	-130
-5	-27	445	503	-3	-83	142	150
-3	-31	287	948	-2	-98	103	385
-1	-33	99	1188	-1	-107	54	540
1	-33	-99	1188	0	-110	0	594
3	-31	-287	948	1	-107	-54	540
5	-27	-445	503	2	-98	-103	385
7	-21	-553	-77	3	-83	-142	150
9	-13	-591	-687	4	-62	-166	-130
11	-3	-539	-1187	5	-35	-170	-406
13	9	-377	-1402	6	-2	-149	-615
15	23	-85	-1122	7	37	-98	-680
17	39	357	-102	8	82	-12	-510
19	57	969	1938	9	133	114	0
				10	190	285	969
Делители							
2660	17556	4903140	22881320	770	201894	432630	5720330
K ₁			1/528				5/9177
K ₂			1/2660				1/770
K ₃		1193/58837680					329/2595780
K ₄		1/70224					1/67298
K ₅		1/11767536					1/519156
K ₆		1/251040768					1/9806280
K ₇		1187/878642688					131/13728792
K ₈		3/56320					27/260015

n: 22				23			
c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
-21	35	-133	1197	-11	77	-77	1463
-19	25	-57	57	-10	56	-35	133
-17	16	0	-570	-9	37	-3	-627
-15	8	40	-810	-8	20	20	-950
-13	1	65	-775	-7	5	35	-955
-11	-5	77	-563	-6	-8	43	-747
-9	-10	78	-258	-5	-19	45	-417
-7	-14	70	70	-4	-28	42	-42
-5	-17	-55	365	-3	-35	35	315
-3	-19	35	585	-2	-40	25	605
-1	-20	12	702	-1	-43	13	793
1	-20	-12	702	0	-44	0	858
3	-19	-35	585	1	-43	-13	793
5	-17	-55	365	2	-40	-25	605
7	-14	-70	70	3	-35	-35	315
9	-10	-78	-258	4	-28	-42	-42
11	-5	-77	-563	5	-19	-45	-417
13	1	-65	-775	6	-8	-43	-747
15	8	-40	-810	7	5	-35	-955
17	16	0	-570	8	20	-20	-950
19	25	57	57	9	37	3	-627
21	35	133	1197	10	56	35	133
				11	77	77	1463
Делители							
3542	7084	96140	8748740	1012	35420	32890	131231100
K_1			1/352				1/805
K_2			1/3542				1/1012
K_3		289/2307360					79/197340
K_4		1/56672					1/35520
K_5		1/2307360					1/197340
K_6		1/239965440					1/22496760
K_7		1439/839879040					787/157477320
K_8		3/36608					1/15295

п: 24				25			
c ₁	c ₂	c ₃	c ₄	c ₁	c ₂	c ₃	c ₄
-23	252	-1771	253	-12	92	-506	1518
-21	187	-847	33	-11	69	-253	253
-19	127	-133	-97	-10	48	-55	-517
-17	73	391	-157	-9	29	93	-897
-15	25	745	-165	-8	12	196	-982
-13	-17	949	-137	-7	-3	259	-857
-11	-53	1023	-87	-6	-16	287	-597
-9	-83	987	-27	-5	-27	285	-267
-7	-107	861	33	-4	-36	258	78
-5	-125	665	85	-3	-43	211	393
-3	-137	419	123	-2	-48	149	643
-1	-143	143	143	-1	-51	77	803
1	-143	-143	143	0	-52	0	858
3	-137	-419	123	1	-51	-77	803
5	-125	-665	85	2	-48	-149	643
7	-107	-861	33	3	-43	-211	393
9	-83	-987	-27	4	-36	-258	78
11	-53	-1023	-87	5	-27	-285	-267
13	-17	-949	-137	6	-16	-287	-597
15	25	-745	-165	7	-3	-259	-857
17	73	-391	-157	8	12	-196	-982
19	127	133	-97	9	29	-93	-897
21	187	847	33	10	48	55	-517
23	253	1771	253	11	69	253	253
				12	92	506	1518
Делители							
4600	394680	17760600	394680	1300	53820	1480050	14307150
K ₁			5/13728				1/1035
K ₂			1/4600				1/1300
K ₃			1721/213127200				467/8880300
K ₄			1/526240				1/53820
K ₅			1/42625440				1/1776060
K ₆			1/75778560				1/34337160
K ₇			49/7577856				133/34337160
K ₈			27/73216				1/16675

Таблица А.11.А. Коэффициенты и делители для некоторых отобранных групп вариаций с разновеликими интервалами

X	c ₁	c ₂	X	c ₁	c ₂	X	c ₁	c ₂			
1	-13	5	0	-7	3	1	-5	3			
5	-1	-9	2	-1	-5	2	-2	-4			
10	14	4	5	8	2	5	7	1			
	366	122		114	38		78	26			
X	c ₁	c ₂	X	c ₁	c ₂	X	c ₁	c ₂			
2	-11	5	0	-2	4	1	-4	2			
5	-2	-8	1	-1	-5	2	-1	-3			
10	13	3	5	3	1	4	5	1			
	294	98		14	42		42	14			
X	c ₁	c ₂	c ₃	X	c ₁	c ₂	c ₃	X	c ₁	c ₂	c ₃
1	-11	20	-8	0	-7	7	-3	1	-2	1	-1
2	-7	-4	14	1	-3	-4	8	2	-1	-1	2
4	1	-29	-7	2	1	-8	-6	4	1	-1	-2
8	17	13	1	4	9	5	1	5	2	1	1
	460	1426	310		140	154	110		10	4	10
X	c ₁	c ₂	c ₃	X	c ₁	c ₂	c ₃	X	c ₁	c ₂	c ₃
0	-2	43	-6	0	-5	9	-5	1	-3	3	-9
1	-1	-17	15	1	-3	-3	9	2	-2	-1	14
2	0	-49	-10	3	1	-13	-5	5	1	-5	-7
5	3	23	1	6	7	7	1	8	4	3	2
	14	5068	362		84	308	132		30	44	330
X	c ₁	c ₂	c ₃	X	c ₁	c ₂	c ₃				
0	-4	107	-18	1	-7	63	-10				
1	-3	-5	25	2	-5	-4	15				
5	1	-205	-9	5	1	-107	-6				
10	6	103	2	10	11	48	1				
	62	64108	1034		196	17738	362				
X	c ₁	c ₂	c ₃	c ₄	X	c ₁	c ₂	c ₃	c ₄		
0	-3	31	-837	21	1	-26	30	-176	-31454		
1	-2	2	752	-64	2	-21	11	76	63213		
2	-1	-19	916	56	4	-11	-19	252	-37667		
4	1	-37	-1016	-14	8	9	-47	-181	4854		
8	5	23	185	1	16	49	25	29	1054		
	40	3224	3171610	4870		3720	4216	133858	6428823890		
X	c ₁	c ₂	c ₃								
0	-31	93	-35433								
1	-25	40	4600								
2	-19	-6	28078								
4	-7	-77	35289								
8	17	-135	-39829								
16	65	85	7295								
	6510	41664	4949911360								

Таблица А.12. Коэффициенты для построения кривой, отображающей процесс во времени и расчленения сумм квадратов данных, полученных через одинаковые промежутки времени полного цикла

Значения X для n*					U ₁	V ₁	U ₂	V ₂	U ₃	V ₃	U ₄	V ₄	U ₅	V ₅	
4	6	8	12	24											
0	0	0	0	0	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000	0,000
				1	0,966	0,259	0,866	0,500	0,707	0,500	0,707	0,500	0,707	0,500	0,866
			1	2	0,866	0,500	0,500	0,866	0,000	0,866	0,000	-0,500	1,000	-0,500	0,863
				3	0,707	0,707	0,000	1,000	-0,707	1,000	-0,707	-0,707	0,000	-1,000	0,000
		1	2	4	0,500	0,866	-0,500	0,866	-1,000	0,866	-1,000	-0,500	0,000	-0,500	-0,866
				5	0,259	0,966	-0,866	0,500	-0,707	0,500	-0,707	-0,500	0,000	-0,500	-0,866
			2	6	0,000	1,000	-1,000	0,000	0,000	0,000	0,000	-1,000	1,000	0,000	0,000
1				7	-0,259	0,966	-0,866	-0,500	0,707	-0,866	0,707	-0,500	0,500	0,866	0,866
	2		4	8	-0,500	0,866	-0,500	0,000	1,000	-1,000	1,000	-0,500	0,500	0,866	0,866
		3		9	-0,707	0,707	0,000	0,500	0,866	-0,866	0,707	-0,500	0,000	0,000	0,000
			5	10	-0,866	0,500	0,500	0,866	0,000	-0,866	0,000	-0,500	0,000	0,866	0,866
				11	-0,966	0,259	0,866	0,259	0,866	-0,500	-0,707	-0,500	0,500	0,866	0,866
			6	12	-1,000	0,000	1,000	1,000	0,866	0,000	-1,000	-0,500	1,000	0,866	0,866
2	3	4		13	-0,966	-0,259	0,866	0,500	0,866	0,000	0,707	-0,500	0,000	0,866	0,866
			7	14	-0,866	-0,500	0,500	0,000	0,866	1,000	0,707	-0,500	0,000	0,866	0,866
				15	-0,707	-0,707	0,000	0,000	1,000	0,866	0,000	-0,500	0,000	0,866	0,866
		5		16	-0,500	-0,866	-0,500	-0,500	0,866	0,000	1,000	-0,500	0,000	0,866	0,866
			8	17	0,000	-1,000	-1,000	-1,000	0,000	0,866	0,000	-0,500	0,000	0,866	0,866
		6		18	0,259	-0,966	-0,866	-0,866	-0,500	-0,500	-0,707	0,707	0,707	0,866	0,866
			9	19	0,500	-0,866	-0,500	-0,500	-0,866	-0,500	-1,000	0,000	0,000	0,866	0,866
				20	0,707	-0,707	0,000	-0,500	-0,866	-1,000	-0,707	0,000	0,000	0,866	0,866
		7		21	0,866	-0,500	0,500	0,500	-1,000	-1,000	-0,707	-0,707	-0,707	0,000	0,000
			10	22	0,966	-0,259	0,866	0,866	-0,500	-0,866	0,000	0,000	-0,707	-0,707	-0,866
				23											

Таблица А.12. (продолжение)
Значения X для $n=7$

	U_1	V_1	U_2	V_2	U_3	V_3
0	1,000	0,000	1,000	0,000	1,000	0,000
1	0,623	0,782	-0,223	0,975	-0,901	0,434
2	-0,223	0,975	-0,901	-0,434	0,623	-0,782
3	-0,901	0,434	0,623	-0,782	-0,223	0,975
4	-0,901	-0,434	0,623	0,782	-0,223	-0,975
5	-0,223	-0,975	-0,901	0,434	0,623	0,782
6	0,623	-0,782	-0,223	-0,975	-0,901	-0,434

Значение X для $n=52$

	U_1	V_1	U_2	V_2	U_3	V_3	U_4	V_4
0	1,000	0,000	1,000	0,000	1,000	0,000	1,000	0,000
1	0,993	0,121	0,971	0,239	0,935	0,355	0,885	0,465
2	0,971	0,239	0,885	0,465	0,749	0,663	0,568	0,823
3	0,935	0,355	0,749	0,663	0,465	0,885	0,121	0,993
4	0,885	0,465	0,568	0,823	0,121	0,993	-0,355	0,935
5	0,823	0,568	0,355	0,935	-0,239	0,971	-0,749	0,663
6	0,749	0,663	0,121	0,993	-0,598	0,823	-0,971	0,239
7	0,663	0,749	-0,121	0,993	-0,823	0,568	-0,971	-0,239
8	0,568	0,823	-0,355	0,935	-0,971	0,239	-0,749	-0,663
9	0,465	0,885	-0,568	0,823	-0,993	-0,121	-0,355	-0,935
10	0,355	0,935	-0,749	0,885	-0,885	-0,465	0,121	-0,993
11	0,239	0,971	-0,885	0,465	-0,663	-0,749	0,568	-0,823
12	0,121	0,993	-0,971	0,239	-0,355	-0,935	0,885	-0,465
13	0,000	1,000	-1,000	0,000	0,000	-1,000	1,000	0,000
14	-0,121	0,993	-0,971	-0,239	0,355	-0,935	0,885	0,465
15	-0,239	0,971	-0,885	-0,465	0,663	-0,749	0,568	0,823
16	-0,355	0,935	-0,749	-0,663	0,885	-0,465	0,121	0,993
17	-0,465	0,885	-0,568	-0,823	0,993	-0,121	-0,355	0,935

	U_1	V_1	U_2	V_2	U_3	V_3	U_4	V_4
18	-0,568	0,823	-0,355	-0,935	0,971	0,239	-0,749	0,663
19	-0,663	0,749	-0,121	-0,993	0,823	0,568	-0,971	0,239
20	-0,749	0,663	0,121	-0,993	0,568	0,823	-0,971	-0,239
21	-0,823	0,568	0,355	-0,935	0,239	0,971	-0,749	-0,663
22	-0,885	0,465	0,568	-0,823	-0,121	0,993	-0,355	-0,935
23	-0,935	0,355	0,749	-0,663	-0,465	0,885	0,121	-0,993
24	-0,971	0,239	0,885	-0,465	-0,749	0,663	0,568	-0,823
25	-0,993	0,121	0,971	-0,239	-0,935	0,355	0,885	-0,465
26	-1,000	0,000	1,000	0,000	-1,000	0,000	1,000	0,000
27	-0,993	-0,121	0,971	0,239	-0,935	-0,355	0,885	0,465
28	-0,971	-0,239	0,885	0,465	-0,749	-0,663	0,568	0,823
29	-0,935	-0,355	0,749	0,663	-0,465	-0,885	0,121	0,993
30	-0,885	-0,465	0,568	0,823	-0,121	-0,971	-0,355	0,935
31	-0,823	-0,568	0,355	0,935	0,239	-0,823	-0,749	0,663
32	-0,749	-0,663	0,121	0,993	0,568	-0,823	-0,971	0,239
33	-0,663	-0,749	-0,121	0,993	0,823	-0,568	-0,971	-0,239
34	-0,568	-0,823	-0,355	0,935	0,971	-0,239	-0,749	-0,663
35	-0,465	-0,885	-0,568	0,823	0,993	0,121	-0,355	-0,935
36	-0,355	-0,935	-0,749	0,663	0,885	0,465	0,121	-0,993
37	-0,239	-0,971	-0,885	0,465	0,663	0,749	0,568	-0,823
38	-0,121	-0,993	-0,971	0,239	0,355	0,935	0,885	-0,465
39	0,000	-1,000	-1,000	0,000	0,000	1,000	1,000	0,000
40	0,121	-0,993	-0,971	-0,239	-0,355	0,935	0,885	0,465
41	0,239	-0,971	-0,885	-0,465	-0,663	0,749	0,568	0,823
42	0,355	-0,935	-0,749	-0,663	-0,885	0,465	0,121	0,993
43	0,465	-0,885	-0,568	-0,823	-0,993	0,121	-0,355	0,935
44	0,568	-0,823	-0,355	-0,935	-0,971	-0,239	-0,749	0,663
45	0,663	-0,749	-0,121	-0,993	-0,823	-0,568	-0,971	0,239
46	0,749	-0,663	0,121	-0,993	-0,568	-0,823	-0,971	-0,239
47	0,823	-0,568	0,355	-0,935	-0,239	-0,971	-0,749	-0,663
48	0,885	-0,465	0,568	-0,823	0,121	-0,993	-0,355	-0,935
49	0,935	-0,355	0,749	-0,663	0,465	-0,885	0,121	-0,993
50	0,971	-0,239	0,885	-0,465	0,749	-0,663	0,568	-0,823
51	0,993	-0,121	0,971	-0,239	0,935	-0,355	0,885	-0,465

- μ — средняя генеральной совокупности.
- μ_y — средняя совокупности выборочных средних.
- σ^2 — дисперсия совокупности.
- σ_y^2 — дисперсия совокупности средних.
- σ — стандартное отклонение совокупности.
- Y — значение переменной, значение варьирующего признака.
- $\bar{Y}$ — средняя выборки.
- S^2 — дисперсия выборок.
- S — стандартное отклонение выборок.
- σ_y — стандартная ошибка совокупности.
- S_y — стандартная ошибка выборки.
- μ_d — разность между средними.
- d — разность между выборочными средними.
- s_d^2 — дисперсия разности.
- s_d — ошибка разности между выборочными средними.
- s_w^2 — дисперсия внутри выборок.
- n — число вариантов.
- r — количество дат в выборке, повторность в полевом опыте.
- N — число объектов совокупности.
- f — частота встречаемости любого значения переменной.
- i — показатель подразделения.
- j — показатель подразделения.
- $x = X - \bar{X}$.
- $y = Y - \bar{Y}$.
- ANOVA — дисперсионный анализ.
- df — число степеней свободы.
- t — критерий Стьюдента, $t = \frac{(\bar{Y} - \mu)}{s_{\bar{y}}}$.
- Σ — суммирование.
- Z — критерий нормального распределения, $Z = \frac{(\bar{Y} - \mu)}{\sigma_{\bar{y}}}$.
- CL — доверительный предел.
- t_{05}, t_{01} — табличные значения критерия t Стьюдента для 5%-ного и 1%-ного уровня значимости.
- F — критерий Фишера, $F = \frac{s_1^2}{s_2^2}$.
- F_{05}, F_{01} — табличные значения критерия F Фишера для 5%-ного и 1%-ного уровня значимости.

- SS — общая сумма квадратов.
- SSV — сумма квадратов для вариантов.
- SST — сумма квадратов для повторений.
- SSB — сумма квадратов для блоков.
- SSR — сумма квадратов для рядов.
- SSC — сумма квадратов для столбцов.
- SSE — сумма квадратов для ошибки.
- MSV — средний квадрат для вариантов.
- MSR — средний квадрат для рядов.
- MSC — средний квадрат для столбцов.
- MSE — средний квадрат для ошибки.
- $df(T)$ — число степеней свободы для вариантов.
- $df(E)$ — число степеней свободы для ошибки.
- C — корректирующий фактор (поправка) в дисперсионном анализе.
- CV — коэффициент вариации.
- s_b^2 — дисперсия между выборками.
- χ^2 — фактическое значение критерия хи-квадрат Пирсона.
- $\chi_{0.05}^2, \chi_{0.01}^2$ — табличные значения хи-квадрат для 5%-ного и 1%-ного уровня значимости.
- H_0 — нулевая гипотеза.
- $HCP_{0.05}, HCP_{0.01}$ — наименьшие существенные разности для 5%-ного и 1%-ного уровня значимости.
- D — критерий Дункана.
- r — коэффициент линейной корреляции.
- s_r — ошибка коэффициента корреляции.
- b — коэффициент регрессии.
- b_{yx} — коэффициент регрессии Y на X .
- b_{xy} — коэффициент регрессии X на Y .
- $\bar{Y}$ — оценочная величина.
- n — число пар наблюдений.
- s_b — ошибка коэффициента регрессии.
- s_{yx} — ошибка отклонения от регрессии.
- $r_{xy \cdot z}$ — частные линейные коэффициенты корреляции.
- $r_{xz \cdot y}, r_{zy \cdot x}$ —
- $R_{x \cdot yz}, R_{y \cdot xz}, R_{z \cdot xy}$ — множественные линейные коэффициенты корреляции.
- R^2 — множественный коэффициент детерминации.
- P — вероятность.
- P_1 — уровень значимости.

- Абсцисса 155
Аддитивность 133
Адекватный анализ логической оценки 66
Анализ дисперсионный 38
— ковариационный 264
— корреляционный 154, 158
— регрессионный 154, 171
— статистический 17
— Стьюдента — Ньюмена — Кьюлаза 66
Анализ подсчетов (численности) 244
Асимптота 190
Асимптотическая кривая 189
- Биномальное распределение 246
Блоки 14, 57
— расщепленные 86
- Вариант 10
— опыта 10, 19
Варианта 19
Вариабельность 13
Варианса 24
Вариация внутриблочная 57
Варьирование блоков 59
Взаимодействие и его расчленение 72, 87, 91, 104
Взвешенная средняя 202
Выборка 20
— репрезентативная 20
— случайная 20
Выборка из совокупности 28
Выделение существенных средних 64
- Гармоник 221
Гипотеза 14
— нулевая 33
— статистическая 32
Гипотетические данные 134
Гистограмма частот 21
Главные делянки 86
— эффекты 100
Группировка 20
Групповые сравнения 71
- Двухфакторный опыт (сплит-плот) 86
Дедуктивная проблема 9
Делянка 19, 35
Делянки первого порядка 86
— второго порядка 86
Десятичный логарифм 136
Дисперсионный анализ 38
— — основы метода 38
— — схемы 49
— — примеры 53, 58, 80, 101
— — двух выборок 38
— — при полной рендомизации 52, 55
— — опыта с рендомизированными блоками 57, 60
— — опыта с латинским квадратом 80
— — двухфакторного опыта 88, 95
— — трехфакторного опыта 99, 104
— — данных за ряд лет 125
Дисперсия 24
— выборочных средних (вариантов) 39
— коэффициента регрессии 39
— отдельных вариантов 178
— совокупности 36
— средних 36
Доверительный интервал 28, 30, 169
— для генеральной средней 37
— для средней разности 45
Доверительные пояса 171
— пределы 32, 169

Достоверность опыта 15
Дулиттл-метод 198

Заблуждения 173
Защитная полоса 16
Звездочка (*, **, ***) 131, 135, 139,
141, 145

Изменчивость 13, 18
Индуктивная проблема 10, 12
Интервал доверительный 29, 30
— для средней разности 48
— классовый 20
Исследование 9, 14

Квадраты и квадратные корни 289
Ковариация 263
Ковариационный анализ 264
— интерпретация 269
— корректирование источников вариации 268
— корректирование средних по вариантам 267
— сравнение приведенных средних 269
Корректирующий фактор 25, 59, 81,
90, 103, 114
Корреляция 154, 157
— измерения 156
— криволинейная 179
— линейная 154
— множественная 226
— обратная (отрицательная) 154
— прямая (положительная) 154
— частная 227, 241
— для трех переменных 229
— для трех и более переменных 234
Корреляционный анализ 154, 157
Коэффициент вариации 26
— взаимодействия 76
— детерминации 157
— линейной корреляции 146, 155
— множественной корреляции 227
— ортогональности 69, 83, 92, 116
— ранговой корреляции 160, 162
— частной корреляции 228
Кривая асимптотическая 189
— квадратичная 198
— кубическая 198
— нормального распределения 21

— распределения частот 21
— полиномиального типа 190
— периодическая (Фурье) 202
— сочетания кривых 200
— степенная 180
— экспоненциальная 186
Криволинейная зависимость 179
Критерий Дункана 66
— значимости 14
— однородности дисперсий Бартлетта 136, 142
— существенности 13, 32
Критерий t для оценки существенности различий 44
— Тьюки 138
— Фишера F 16, 34
— хи-квадрат (χ^2) 137, 245
— Пуассона 137, 245

Латинский квадрат 78
— двойной 78
— дисперсионного анализа 80
— источников варьирования 80
— рендомизации 80
— степеней свободы 80
— суммы квадратов 81
— средних квадратов 82
Линейная регрессия 154
Ловушки 173
Логарифмические преобразования 180
Логика 9
Локальный контроль 14, 18

Мантиса 136
Математическая статистика 17
Медиана 24
Метод наименьших квадратов 164
— ортогональных полиномов 210
— ранговых различий 177
— хи-квадрат (χ^2) 246
Многоранговые критерии 246
Многоранговый метод оценки разностей 66
Множественный коэффициент детерминации 196
Мультипликативный эффект варианта 134, 153
Мышление дедуктивное 9
— индуктивное 9

Наблюдения 10

Наблюдаемая величина 245

НСР для вариантов с одинаковой повторностью 65

— разной повторностью 65

Наименьшая существенная разность, НСР 45, 64, 97, 108, 109, 118, 129

Научный метод 14

Независимость 251

— дисперсий 132

— проверки 251, 252

— средних 132

Неоднородность 256

Непрерывность 132

Неслагаемость 133

Неучтенная вариация 59

Нормальное распределение 22, 24

Нормальное уравнение 237

Нормальность распределения компонентов ошибки 130

Нулевая гипотеза 33

Нулевой разряд 152

— уровень 152

Объем выборки 24

— совокупности 22

Однородность дисперсий 131

Одноярусное расположение делянок 78

Ожидаемая величина 245

Округление опытных данных 46

Ордината 155

Ортогональность 68

Ортогональные коэффициенты 92

— полиномы 83

Остаток 139, 143

Отзывчивость (тренд) 73

Ошибка эксперимента 13

— стандартная 95, 106, 124

Оценка p 155

Оценка статистическая 13

Оценка существенности по НСР 66

— по F -критерию 66

Парабола 191, 198

— кубическая 208

— пятой степени 208

— четвертой степени 208

Параметр 20, 35

Переменная 19

— дискретная 19

— зависимая 155

— непрерывная 19

— независимая 155

Планируемые F -критерии 68

Поверхность отклика 236

— трехмерная 239

— «срезы» 240

Повторение 19

— организованное 19

Повторность 14, 250

Повторные учеты, как судбделянки 120

Подбор периодической кривой 219

— полиномальной кривой 211

— для вариантов с равновеликими интервалами 211

— для вариантов с неравновеликими интервалами 218

Полином 83, 191

— в экспериментах с повторностью 198

— коэффициентов 115

— ортогональный 210

— типа кривых 190

Полоса 110

Поправка Иейтса 248

Полуамплитуда 202

Преобразование 130

— логарифмическое 140

— через квадратный корень 143

— через арксинус, угол 147

Преобразование шкалы 150

Приведенный хи-квадрат 142

Принципы планирования схем опытов 14

Проверка независимости 251

Равновеликий интервал 210

Распределение 30

— биномиальное 147, 246

— выборочных средних 29

— дискретное 245

— нормальное 144, 245

— непрерывное 245

— Фишера F 33

Пуассона 143, 144

Распределение t 30, 273

Распределение χ^2 137, 245

Расчеты дисперсии 24

— стандартного отклонения 26

- с помощью калькулятора 26
- Регрессия 157
- анализ 171
- коэффициент 157
- линейная 154
- множественная 226
- Рендомизация 14, 17, 52
- в латинском квадрате 79
- в двухфакторном опыте 88
- в трехфакторном опыте 100
- организованная 57
- полная 52
- Ряды 78
- разрядные 153
- ранжированные 24, 67
- — средних 67
- Свойства средней 24
- Слагаемость 133
- главных эффектов 130
- Случайность 14
- Случайная величина 11
- Случайные числа 14, 272
- Совокупность 13, 19
- индивидуальных вариантов 36
- средних 36
- Сокращенный Дулиттл-метод 198
- метод определения трендов 210
- Сплит—плот (двухфакторный опыт) 86
- дисперсионный анализ 87
- источники вариации 90
- рендомизация 87
- Сплит-сплит-плот (трехфакторный опыт)
- дисперсионный анализ 99
- оценка существенных средних 104
- принципы рендомизации 99
- стандартные ошибки и НСР 106
- Сравнения групповые 71
- ортогональные 68
- Средние отклонения 22
- Средний квадрат (дисперсия) 40, 61, 103
- Средняя арифметическая 23
- средняя генеральная 23
- совокупность 22
- выборочных средних 36
- индивидуальных вариантов 36
- Стандартное отклонение 27, 36
- Стандартная ошибка средней 29, 117

- Стандартный анализ серии наблюдений 120
- метод 161
- Статистические характеристики 20
- Статистическая существенность 162
- Степенная кривая 180
- Степень риска 15
- Столбец 27, 78
- Субделянки 87
- Суб-субделянки 99
- Субсуммы 140
- Сумма квадратов 28, 40, 59, 90, 103, 114
- Существенность 32
- Схема опыта, эксперимента 13, 49
- полной рендомизации 52, 53
- Таблица дисперсионного анализа 89
- Табулирование 16, 99
- Теснота связи 177
- Тесты Тьюки 66
- Шеффи 66
- Точечная диаграмма 159
- Точность эксперимента, повышение 15, 260, 269
- — — за счет выбора вариантов 261
- — — — — выбора экспериментальной единицы 262
- — — — — ковариации 263
- — — — — отбора экспериментального материала 262
- — — — — совершенствования техники 262
- — — — — увеличения повторности 260
- Тренд 73, 210
- Трехмерная диаграмма 235
- поверхности отклика 239
- Трехфакторный опыт (сплит-сплит-плот) 99
- Увеличение повторности 260
- Умножающий эффект варианта 134
- Уравнение квадратное 191
- кватерное 191
- квинтичное 191
- кубическое 191
- линейное 191
- пятой степени 203
- регрессии 157

- четвертой степени 208
- Уровень существенности
 - 1-ный уровень существенности 33
 - 5-ный уровень существенности 33

Фаза периодической кривой 202
 Факториальность 15
 Факториальный эксперимент 48, 51
 Фиксированные переменные 234
 Функциональная зависимость 168
 Фунт 53, 138

Хандредвейт 241
 Хи-квадрат, χ^2 интерпретация результатов 137, 245, 250
 — рекомендации к использованию 249

Частота 21
 Частная корреляция 226

Шкалы 150
 — преобразованные 150
 — разрядные 151

Эксперимент 9, 14
 — планирование 14
 — повышение точности 260
 — схемы 49
 Экспериментальная единица 13, 34
 — ошибка 59
 Экспоненциальная кривая 186
 Экстраполяция 176
 Элемент случайности 11
 Эллипс 227
 Эффект блоков 14
 — вариантов 74
 — взаимодействия 74
 Эффекты главные 74

Явление взаимодействия 49, 72

<i>Предисловие к русскому изданию</i>	3
<i>Предисловие к английскому изданию</i>	7
Глава 1. Логика, исследование и эксперимент	9
Дедуктивное и индуктивное мышление	9
Исследование, научный метод и эксперимент	14
Этапы эксперимента	15
Заключение	17
Глава 2. Некоторые основные понятия	19
Нормальное распределение	22
Статистические системы оформления расчетов, средние и стандартные отклонения	22
Выборка из нормального распределения (совокупности)	28
Статистические гипотезы и критерии существенности	32
Заключение	35
Глава 3. Дисперсионный анализ и t-критерий	38
ANOVA (дисперсионный анализ) двух выборок	38
Совокупность разностей между средними	42
Округление опытных данных	46
Факториальные эксперименты	48
Схема эксперимента и дисперсионный анализ	49
Заключение	50
Глава 4. Схема полной рендомизации	52
Рендомизация	52
Дисперсионный анализ	53
Суть и назначение анализа	55
Заключение	56
Глава 5. Схема рендомизированных (организованных) блоков (повторений)	57
Рендомизация	58
Дисперсионный анализ	58
Суть и назначение анализа	60
Величина F	62
Заключение	63
Глава 6. Выделение существенных средних (оценка частных различий)	64
Наименьшая существенная разность (НСР)	64
Многоранговые критерии	66
Планируемые F -критерии	68
Заключение	76

Глава 7. Схема латинского квадрата	78
Рандомизация	79
Дисперсионный анализ	80
Выделение существенных средних	82
Заключение	85
Глава 8. Схема расщепленных делянок — сплит-плот (двухфакторный опыт)	86
Рандомизация	88
Дисперсионный анализ	88
Выявление существенных средних. Оценка частных различий	92
Заключение	98
Глава 9. Схема расщепленных делянок. Сплит-сплит-плот (трехфакторный опыт)	99
Табулирование данных	99
Дисперсионный анализ	99
Оценка существенных средних	104
Заключение	109
Глава 10. Схема расщепленных блоков	110
Дисперсионный анализ	112
Стандартные ошибки	117
Заключение	119
Глава 11. Использование повторных учетов в качестве субделянок	120
Статистический анализ каждой серии наблюдений	120
Объединение данных за два или большее число лет	125
Заключение	129
Глава 12. Преобразования исходных данных (что делать, если данные нарушают установленные закономерности)	130
Предпосылки, составляющие основу дисперсионного анализа	130
Проверка соответствия данных предположениям дисперсионного анализа	134
Логарифмическое преобразование	140
Преобразования через квадратный корень	143
Преобразования через арксинус или угол	147
Предварительно преобразованные шкалы	150
Заключение	152
Глава 13. Линейная корреляция и регрессия	154
Идея	154
Измерение корреляции	155
Регрессия	156
Сравнение корреляции и регрессии	157
Расчеты линейной корреляции	158
Статистическая существенность	162
Линия регрессии	164
Доверительные пределы	169
Регрессионный анализ данных эксперимента, имеющего повторность	171
Заблуждения (ловушки)	173
Заключение	177
Глава 14. Криволинейные зависимости	179
Выбор кривой для описания зависимости	179
Заключение	207
Глава 15. Сокращенные методы регрессии	210
Подбор полиномиальной кривой	211
Подбор периодической кривой	219
Заключение	224

Глава 16. Матричные методы корреляции	226
Кoeffициенты корреляции	226
Кoeffициенты регрессии	228
Пример корреляции с тремя переменными	229
Корреляция для трех переменных	234
Поверхности отклика	236
Заключение	241
Глава 17. Анализ дисперсии	244
Хи-квадрат	245
Рекомендации к использованию метода	249
Интерпретация результатов	250
Проверка независимости	251
Неоднородность	256
Заключение	259
Глава 18. Пути повышения точности экспериментов	260
Увеличение повторности	260
Выбор вариантов	261
Совершенствование техники	262
Отбор экспериментального материала	262
Выбор экспериментальной единицы	262
Введение дополнительных приемов — ковариация	263
Спланированная группировка экспериментальных единиц — схема	270
Заключение	270
Список литературы	270
Приложение	272
Краткий указатель символов	310
Предметный указатель	312

Т. ЛИТТЛ, Ф. ХИЛЛЗ

**СЕЛЬСКОХОЗЯЙСТВЕННОЕ ОПЫТНОЕ
ДЕЛО**

Планирование и анализ

Заведующий редакцией А. Т. Докторов

Редактор В. Н. Брусов

Художник В. А. Миттельман

**Художественный редактор Б. К. Дормидон-
тов**

Технический редактор Н. В. Суржева

**Корректоры С. В. Вишнякова и Е. В. Та-
расенко**

ИБ № 2613

Сдано в набор 16.02.81. Подписано к печати
04.09.81. Формат 60×90^{1/16}. Бумага офсет. № 2.
Гарнитура литературная. Печать высокая. Усл.
печ. л. 20. Усл. кр.-отт. 20,32. Уч.-изд. л. 22,51.
Изд. № 219. Тираж 5000 экз. Заказ № 1016. Цена
1 р. 70 к.

Ордена Трудового Красного Знамени издательство
«Колос», 107807, ГСП, Москва, В-53, ул. Садовая-
Спасская, 18

Набрано в ордена Октябрьской Революции и орде-
на Трудового Красного Знамени Первой Образцо-
вой типографии имени А. А. Жданова Союзполи-
графпрома при Государственном комитете СССР по
делам издательств, полиграфии и книжной торговли.
Москва, М-54, Валовая, 28

Отпечатано с матриц на Белоцерковской книжной
фабрике, 256400, г. Белая Церковь, ул. Карла Мар-
кса, 4.