

Ю. П. АДЛЕР, Е. В. МАРКОВА, Ю. В. ГРАНОВСКИЙ

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ ПОИСКЕ ОПТИМАЛЬНЫХ УСЛОВИЙ

*Издание второе
переработанное и дополненное*



Планирование эксперимента — новая научная дисциплина. Она применяется для решения широкого круга задач: построения интерполяционных моделей, изучения кинетики и механизма явлений, оптимизации процессов и др. Наибольшее практическое значение имеет оптимизация процессов (планирование экстремальных экспериментов). Этому направлению и посвящена монография. Она представляет собой вводный курс по планированию эксперимента при поиске оптимальных условий.

Рассчитана на научных и инженерно-технических работников.

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Эта книга вышла из печати первым изданием в 1971 г. За прошедшее время авторам пришлось выслушать и прочесть большое число мнений о книге, высказанных читателями. Одновременно накапливался опыт использования книги в качестве учебного пособия или даже учебника (в Московском институте повышения квалификации Министерства химической промышленности СССР) по курсу планирования эксперимента. Все мнения сводились к тому, что книгу следует переиздать, внося многие исправления. Читатели высказывали самые различные суждения о дополнениях, вплоть до включения в книгу планов второго порядка, отсеивающего эксперимента, априорного ранжирования и т. д. С одной стороны, включение в книгу подобного материала сделало бы ее весьма громоздкой, да и материал заслуживает, на наш взгляд, отдельного подробного рассмотрения, с другой стороны, некоторое расширение книги казалось целесообразным. В результате обсуждений, колебаний и компромиссов родилась следующая структура второго издания.

Прежде всего возникла мысль дополнить вторую главу материалом по обобщенному критерию оптимизации, получившему в последнее время широкое распространение. Попытка подробно это изложить привела к появлению новой самостоятельной главы.

Желание дополнить изложение регрессионного анализа рассмотрением более сложных случаев (неравномерное дублирование опытов, критерии оптимальности планов и т. д.) привело нас к необходимости обратиться к матричному аппарату. Тем более, что он получает все более широкое распространение не только в теоретических, но и в прикладных публикациях. Так возникла довольно большая [глава по матричному аппарату регрессии.

Хотя в тексте первого издания каждое существенное высказывание авторы стремились систематически иллюстрировать приме-

рами, последние оказались разрозненными, отрывочными и было много предложений о построении сквозного примера, чтобы можно было бы проследить все операции подряд от начала до конца. Учитывая это, мы построили такой пример.

Чтобы связать книгу со всей концепцией математической теории эксперимента, было решено написать главу по классификации планов (в ее написании принимал участие А. Н. Лисенков).

Особо отметим одно из наших решений — отказ от программированного способа изложения материала. Для этого, собственно, не было особых оснований в отзывах читателей, хотя некоторых читателей такая форма изложения раздражала и мешала чтению. Но не они определили наше решение. Авторам казалось интересным проведение своеобразного педагогического эксперимента по обучению планированию с помощью как программированных, так и обычных учебников. Для этого надо иметь один и тот же материал, изложенный двумя способами: с программированием и без него. Разумеется, такой подход не является отказом от принципов программированного обучения.

Таким образом, в новом издании осуществлены значительные изменения объема и структуры материала без существенных выходов за рамки тематики, определенной первым изданием.

При подготовке этого издания нами учтены критические замечания и пожелания многих читателей. Особенно авторы благодарны Г. В. Лейбу, А. И. Смирнову, А. А. Сухову.

Авторы будут признательны за отзывы, предложения, замечания, которые можно направлять по адресу: 117333, Москва, ул. Вавилова, д. 40. Научный совет по комплексной проблеме «Кибернетика» при Президиуме АН СССР, секция «Математическая теория эксперимента».

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Перед вами книга, посвященная одному из методов планирования эксперимента — методу Бокса — Уилсона. Ее задача — дать по возможности простое введение в предмет, рассчитанное на людей с высшим техническим образованием, а также студентов. Такая постановка задачи заставила отказаться от рассмотрения многих современных методов планирования эксперимента и ограничиться подробным изложением одного метода.

Метод Бокса — Уилсона, пожалуй, простейший из методов планирования эксперимента. Однако его успешное применение зависит от решения многих вопросов, связанных с принятием неформализованных решений при выборе параметра оптимизации, факторов, плана экспериментов и при интерпретации результатов.

В книге сделана попытка на многих примерах рассмотреть эти вопросы, обычно не освещаемые в литературе. Примеры заимствованы главным образом из собственных работ авторов. Это сделано не потому, что в литературе не было лучших. Просто для детального рассмотрения требуется слишком много информации, обычно отсутствующей в публикациях.

При изложении материала были использованы элементы программированного обучения — «метод многовариантного ответа». Это сравнительно новый метод изложения, хотя тенденция к его распространению очевидна. Его цель — активизировать усвоение материала, выработать необходимые навыки.

Для эффективной проработки материала, изложенного в такой форме, вообще говоря, требуется специальный набор текста, когда значительная часть страницы остается пустой. К сожалению, подобное расточительство в настоящее время невозможно, и в данном издании пришлось прибегнуть к паллиативу: текст набран подряд, но порции информации пронумерованы.

Несколько слов о стиле. В книге сделана попытка вести изложение в манере, близкой к разговорному языку. Это не должно ввести читателя в заблуждение: какова бы ни была форма, для овладения предметом требуется серьезный систематический труд.

Книгу в рукописи прочло много людей, чьи доброжелательные отзывы и товарищеская критика постоянно поддерживали авторов, которые выражают им свою признательность. Особенно благодарны авторы А. И. Бергу, В. Г. Горскому, И. А. Гришкану, А. Н. Захарову, Ю. Л. Клокову, Ю. П. Назаренко, В. В. Налимову, Е. Л. Мамутову, Э. М. Менчеру.

ЦЕЛИ КНИГИ

Прежде чем вы начнете знакомство с планированием эксперимента, мы хотим познакомить вас с целями, которые преследовали авторы, приступая к написанию этой книги. Некоторые термины могут показаться вам незнакомыми, но это не должно вас смущать: далее они будут определены. Нам кажется, что такое начало позволит вам уяснить общую структуру книги, а также поможет проконтролировать себя после ее изучения.

Мы стремились, чтобы читатель, проработав эту книгу, усвоил, какая информация нужна для построения факторного эксперимента при оптимизации различных процессов. Получив эту информацию, он сможет выбрать нужный план опытов, построить математическое описание процесса в области экспериментирования и провести статистический анализ, выбрать наикратчайший путь к оптимуму и осуществить движение по этому пути.

Авторы хотели бы, чтобы читатель, усвоив изложенный материал, смог:

- применить методы планирования эксперимента при оптимизации многофакторных процессов;
- дать определение параметру оптимизации;
- дать определение факторам, определяющим процесс;
- дать определение поверхности отклика;
- дать определение матрице планирования;
- дать определение условиям ортогональности и ротатабельности;
- дать определение полному факторному эксперименту 2^k и дробным репликам от него;
- дать определение уравнению регрессии;
- дать определение основным эффектам и эффектам взаимодействия;
- дать определение кратчайшему пути к оптимуму;
- выбрать параметр оптимизации;
- выбрать уровни факторов и интервалы их варьирования;
- построить полный факторный эксперимент 2^k и дробные реплики от него;
- получить уравнение регрессии;
- произвести статистический анализ уравнения регрессии;
- произвести содержательную (например, физико-химическую) интерпретацию уравнения регрессии;

найти и реализовать кратчайший путь к оптимуму.

Мы стремились, чтобы читатель получил навыки, необходимые экспериментатору для принятия решений на основе использования методов планирования эксперимента в простейших случаях.

Получив эти навыки, читатель сможет:

выбрать область, в которой имеет смысл планировать эксперимент;

использовать имеющиеся данные при составлении плана эксперимента;

принять решение о необходимых действиях после получения и статистического анализа уравнения регрессии.

ОГРАНИЧЕНИЯ

Читатель не должен питать иллюзий относительно того, с чем ему предстоит познакомиться на следующих страницах.

Мы будем предполагать, что изучаемый процесс физически осуществлен, и перед исследователем стоит задача его оптимизации.

Из многих возможных путей поиска оптимальных условий мы рассмотрим лишь один, который получил название *метода Бокса — Уилсона или метода крутого восхождения*.

Этот метод позволяет получать статистические математические модели процессов, используя факторное планирование, регрессионный анализ и движение по градиенту.

Кроме того, мы будем предполагать, что: множество определяющих факторов задано, каждый из факторов управляем, результаты опытов воспроизводятся, опыты равноценны, т. е. различием в стоимости можно пренебречь,

решается задача поиска оптимальных условий (или в некоторых случаях интерполяции), математическая модель процесса заранее не известна.

Мы будем рассматривать задачи с числом факторов от двух до пятнадцати. (Но кое-что пригодится и при большем числе факторов.)

Некоторые менее существенные ограничения будут приведены в тексте.

Конечно, столь ограниченная задача является частной, и вам может показаться, что на ее изучение не стоит тратить времени. Мы можем возразить так: это частная, но достаточно широко распространенная на практике задача; ее изучение служит основой для понимания более сложных задач.

Чтобы достигнуть вершин, надо начать движение.

ВВЕДЕНИЕ

Некоторые люди, считая ошибочными все правила и наставления, глубоко убеждены в том, что величественное вложено в нас природой, а не постигается в обучении. . . что способности чахнут и теряют свою естественность при соприкосновении с иссушающими природу учеными наставлениями.

Греческий ионим. I в н. э. О возвышенном

Мысль о том, что эксперимент можно планировать, восходит к глубокой древности. Наш далекий предок, убедившись, что острым камнем можно убить даже мамонта, несомненно выдвигал гипотезы, которые после целенаправленной экспериментальной проверки привели к созданию копья, дротика, а затем и лука со стрелами.

Он, однако, не пользовался статистическими методами, поэтому остается непонятным, как он вообще выжил и обеспечил тем самым наше существование.

Только в начале нашего века люди, наконец, поняли, что дальше дело так не пойдет, и придумали статистические методы планирования эксперимента. Честь открытия этой идеи принадлежит английскому статистiku Рональду Фишеру (конец двадцатых годов), который впервые показал целесообразность одновременного варьирования всеми факторами в противовес широко распространенному однофакторному эксперименту [1]*. Понадобилось еще несколько десятилетий, чтобы в начале пятидесятих годов появилось новое направление в планировании эксперимента, связанное с оптимизацией процессов, — планирование экстремального эксперимента. Первая работа в этой области была опубликована в 1951 г. Боксом и Уилсоном в Англии [2]. Идея метода Бокса — Уилсона крайне проста. Экспериментатору предлагается ставить последовательные небольшие серии опытов, в каждой из которых одновременно варьируются по определенным правилам все факторы. Серии организуются таким образом, чтобы после математической обработки предыдущей можно было выбрать условия проведения (т. е. спланировать) следующую серию. Так последовательно, шаг за шагом, достигается область оптимума.

Применение планирования эксперимента делает поведение экспериментатора целенаправленным и организованным, существенно способствует повышению производительности его труда и надежности полученных результатов. Важным достоинством

* Ссылки на литературу читатель найдет после каждой главы.

метода является его универсальность, пригодность в огромном большинстве областей исследования, интересующих современного человека. В нашей стране планирование эксперимента развивается с 1960 г. под руководством В. В. Налимова. Однако даже простая процедура планирования весьма коварна, в чем вы сможете неоднократно убедиться, читая эту книгу.

Планирование эксперимента еще совсем молодая область*. Она бурно развивается и вызывает все больший интерес у исследователей. Интерес вполне понятен: перспектива сократить число опытов, найти оптимум, получить количественные оценки влияния факторов и определить ошибки — крайне привлекательна.

Но, когда экспериментатор делает попытку познакомиться с планированием эксперимента, он часто сталкивается с серьезными трудностями. Больше того, иногда он просто неверно применяет методы планирования или выбирает не самый оптимальный для данной ситуации путь исследования, или допускает еще какие-нибудь досадные ошибки. При этом снижается эффективность его работы и появляется опасность дискредитации важного и полезного направления.

Эти трудности вызваны объективными причинами. Здесь и молодость данного направления, неустановившаяся терминология, недостаточность практического опыта и слишком абстрактный характер многих опубликованных работ, и недостаточная математическая подготовка экспериментаторов, и многое другое.

Наша книга представляет собой попытку сделать шаг вперед на пути преодоления этих трудностей. Она написана тремя авторами. Смысл такого объединения заключается в обобщении опыта работы каждого автора. Этот опыт возник при участии в различных исследованиях, при многочисленных консультациях экспериментаторов, а также при многократном чтении курсов лекций, самых разнообразных по структуре, длительности и составу аудитории.

Наша цель — развить у читателя навыки практического применения планирования эксперимента. Чтобы сделать задачу выполнимой, пришлось ввести жесткие ограничения, с которыми вы уже знакомы (см. стр. 10). Мы будем рассматривать один из самых простых методов планирования — метод Бокса — Уилсона.

От читателя кроме энтузиазма требуются общие представления об элементах математической статистики и о планировании эксперимента. В качестве исходного учебника статистики можно предложить, например, книгу Н. Бейли «Статистические методы в биологии» (М., «Мир», 1964). Что касается концепции планирования эксперимента, то данную работу можно рассматривать как

естественное развитие в сторону практических приложений книги одного из авторов (Ю. П. А д л е р «Введение в планирование эксперимента», М., изд-во «Металлургия», 1969). Для более подготовленного читателя можно рекомендовать монографию В. В. Налимова и Н. А. Черновой «Статистические методы планирования экстремальных экспериментов» (М., «Наука», 1965).

Мы понимаем, что во многих вузах страны в ближайшее время начнут читаться курсы или фрагменты курсов планирования эксперимента, и хотели бы помочь и тем, кого будут учить, и тем, кто будет учить.

Авторы сочтут свою задачу выполненной, если вы заинтересуетесь методами планирования эксперимента, прочитав эту книгу. Мы будем рады, если вы действительно научитесь пользоваться простейшими из этих методов и не будете сожалеть о потраченном времени.

Итак, мы начинаем.

* Очерк по истории развития планирования экспериментов вы сможете найти в гл. 17.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Мы... еще не далеко ушли от того возраста, когда, давая название чему-либо, мы полагаем, что создаем нечто новое.

М. Пруст. По направлению к Свану

Вам, конечно, хочется сразу заняться делом. Но подождите, давайте начнем со слов. Планирование эксперимента, как и всякий раздел науки, имеет свою терминологию. Вы уже убедились в этом, читая «Цели книги» и «Ограничения».

Нам предстоит рассматривать терминологию на протяжении всей книги, но некоторые наиболее общие термины собраны в этой главе, ибо без них мы не сможем понимать друг друга.

Из названия темы видно, что речь идет об экспериментальных методах. Большинство научных исследований связано с экспериментом. Он проводится в лабораториях, на производстве, на опытных полях и участках, в клиниках и т. д. Эксперимент может быть физическим, психологическим или модельным. Он может непосредственно проводиться на объекте или на его модели. Модель обычно отличается от объекта масштабом, а иногда природой.

Как вы считаете, можно ли поставить эксперимент на абстрактной математической модели?

Если модель достаточно точно описывает объект, то эксперимент на объекте может быть заменен экспериментом на модели. В последнее время наряду с физическими моделями все большее распространение получают абстрактные математические модели. Можно получать новые сведения об объекте, экспериментируя на модели, если она достаточно точно описывает объект.

Эксперимент занимает центральное место в науке. Однако возникает вопрос, насколько эффективно он используется. Джон Бернал, например, отмечал, что научные исследования организуются и проводятся настолько хаотично, что их коэффициент полезного действия может быть оценен величиной порядка 2%. Для того чтобы повысить эффективность исследований, требуется нечто совершенно новое. Одним из возможных путей является применение математических методов, построение математической теории планирования эксперимента.

Планирование эксперимента — это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью. При этом существенно следующее:

стремление к минимизации общего числа опытов;

одновременное варьирование всеми переменными, определяющими процесс, по специальным правилам — алгоритмам;

использование математического аппарата, формализующего многие действия экспериментатора;

выбор четкой стратегии, позволяющей принимать обоснованные решения после каждой серии экспериментов.

Задачи, для решения которых может использоваться планирование эксперимента, чрезвычайно разнообразны.

Поиск оптимальных условий, построение интерполяционных формул, выбор существенных факторов, оценка и уточнение констант теоретических моделей (например, кинетических), выбор наиболее приемлемых из некоторого множества гипотез о механизме явлений, исследование диаграмм состав—свойство — вот примеры задач, при решении которых применяется планирование эксперимента. Можно сказать, что там, где есть эксперимент, имеет место и наука о его проведении — планирование эксперимента.

{Поиск оптимальных условий является одной из наиболее распространенных научно-технических задач. Они возникают в тот момент, когда установлена возможность проведения процесса и необходимо найти наилучшие (оптимальные в некотором смысле) условия его реализации.} Этим задачам и посвящена наша книга.

Пусть, например, у химика возникла гипотеза о том, что при взаимодействии двух веществ должен получаться некоторый интересующий его продукт. Чтобы убедиться в правильности своей гипотезы, он начинает проводить эксперимент. Возможно, что ему повезло и он получил требуемый продукт. Однако выход продукта весьма низок, скажем, 2%. Вот тут-то и возникает задача выбора оптимальных условий. Требуется так подобрать концентрации реагирующих веществ, температуру, давление, время реакции и другие факторы, чтобы сделать выход возможно более близким к 100%. В данном примере находятся условия проведения процесса, оптимальные в смысле максимизации выхода требуемого продукта. Но это далеко не единственно возможная постановка задачи. Найденные условия оказались бы другими, если бы ставилась, например, цель минимизации себестоимости продукта или минимизации количества вредных примесей. Следует подчеркнуть, что всегда необходимо четко формулировать, в каком смысле условия должны быть оптимальными. Этим определяется выбор цели исследования. Точная формулировка цели в значительной мере определяет успех исследования, и мы посвятим этому вопросу следующую главу.

Задачи, сформулированные аналогичным образом, называются задачами оптимизации. Процесс их решения называется процессом оптимизации или просто оптимизацией. Выбор оптимального состава многокомпонентных смесей или сплавов, повышение производительности действующих установок, повышение качества продукции, снижение затрат на ее получение — вот примеры задач оптимизации.

Эксперимент, который ставится для решения задач оптимизации, называется экстремальным. Это название связано с глубокой аналогией между оптимизацией и поиском экстремума некоторой функции. Давайте рассмотрим следующие две задачи.

1. Прочность бетона в значительной степени определяется маркой цемента, количеством наполнителя и количеством воды. Требуется установить связь между прочностью бетона и названными факторами.

2. Надежность некоторого полупроводникового прибора зависит от ряда технологических факторов. Требуется так подобрать значения этих факторов, чтобы надежность прибора повысилась.

Как вы думаете, какая из этих задач является экстремальной?

Чтобы облегчить вам выбор, укажем на признак, отличающий экстремальные задачи. Задача является экстремальной, если цель ее состоит в поиске экстремума некоторой функции. Чтобы установить, какая из двух задач является экстремальной, надо обратиться к их формулировкам и выяснить, где удовлетворяются требования экстремальности. В задаче 1 требуется установить связь между прочностью бетона и тремя факторами. Здесь не определено, какая прочность является оптимальной, и не требуется ее оптимизировать. В задаче 2 необходимо повысить надежность прибора. Сама постановка задачи указывает на то, что существующая надежность не удовлетворяет экспериментатора и требуется поиск таких условий, при которых ее значения повысятся. Задачи типа 1 мы будем называть интерполяционными, а типа 2 — экстремальными.

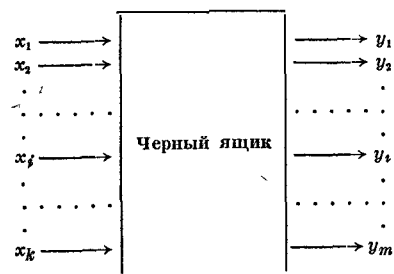


Рис. 1. Схема «черного ящика»

Чтобы продвинуться дальше, нам придется определить еще ряд важных понятий, первое из которых — «объект исследования». Для описания объекта исследования удобно пользоваться представлением о кибернетической системе, которая схематически изображена на рис. 1. Иногда такую кибернетическую систему называют «черным ящиком» [3, 4]. Стрелки справа изображают численные характеристики целей исследования. Мы

обозначаем их буквой игрек и называем параметрами оптимизации. В литературе вы можете встретить другие названия: критерий оптимизации, целевая функция, выход «черного ящика» и т. д.

Для проведения эксперимента необходимо иметь возможность воздействовать на поведение «черного ящика». Все способы такого воздействия мы обозначаем буквой икс и называем факторами. Их называют также входами «черного ящика».

При решении задачи будем использовать математические модели объекта исследования. Под математической моделью мы

понимаем уравнение, связывающее параметр оптимизации с факторами. Это уравнение в общем виде можно записать так:

$$y = \varphi(x_1, x_2, \dots, x_k),$$

где символ φ (), как обычно в математике, заменяет слова: «функция от». Такая функция называется *функцией отклика*. В четвертой главе мы рассмотрим вопрос о том, как эту функцию можно выбрать и построить. А сейчас важно понять, как получаются условия проведения опытов в том эксперименте, который мы собираемся провести.

Каждый фактор может принимать в опыте одно из нескольких значений. Такие значения будем называть уровнями. Может оказаться, что фактор способен принимать бесконечно много значений (непрерывный ряд). Однако на практике точность, с которой устанавливается некоторое значение, не беспредельна. Поэтому мы вправе считать, что всякий фактор имеет определенное число дискретных уровней. Это соглашение существенно облегчает построение «черного ящика» и эксперимента, а также упрощает оценку их сложности.

Фиксированный набор уровней факторов (т. е. установление каждого фактора на некоторый уровень) определяет одно из возможных состояний «черного ящика». Одновременно это есть условия проведения одного из возможных опытов. Если перебрать все возможные наборы состояний, то мы получим полное множество различных состояний данного «ящика». Одновременно это будет число возможных различных опытов.

Чтобы узнать число различных состояний, достаточно число уровней факторов (если оно для всех факторов одинаково) возвести в степень числа факторов k : p^k , где p — число уровней. Поупражняйтесь в подсчете числа различных состояний для разных случаев. Это вам пригодится в дальнейшем. Кроме того, вы увидите, что реальные объекты, с которыми вы сталкиваетесь ежедневно, обладают огромной сложностью. Так, на первый взгляд простая система с пятью факторами на пяти уровнях имеет 3125 состояний, а для десяти факторов на четырех уровнях их уже свыше миллиона!

В этих условиях мы просто вынуждены отказаться от таких экспериментов, которые включают все возможные опыты: перебор слишком велик. Тогда возникает вопрос: сколько и каких опытов надо включить в эксперимент, чтобы решить поставленную задачу? Здесь-то и приходит на помощь планирование эксперимента.

Однако нужно иметь в виду, что при планировании эксперимента не безразлично, какими свойствами обладает объект исследования. Укажем два основных требования, с которыми приходится считаться. Прежде всего существенно, воспроизводятся ли на объекте результаты эксперимента. Выберем некоторые уровни для всех факторов и в этих условиях проведем эксперимент. Затем повторим его несколько раз через неравные промежутки вре-

мени и сравним значения параметра оптимизации. Разброс этих значений характеризует воспроизводимость результатов. Если он не превышает некоторой заранее заданной величины (наших требований к точности эксперимента), то объект удовлетворяет требованию воспроизводимости результатов, а если превышает, то не удовлетворяет этому требованию. Мы будем рассматривать только такие объекты, для которых требование воспроизводимости выполняется.

Планирование эксперимента предполагает активное вмешательство в процесс и возможность выбора в каждом опыте тех уровней факторов, которые представляют интерес. Поэтому такой эксперимент называется активным. Объект, на котором возможен активный эксперимент, называется управляемым. Это и есть второе требование к объекту исследования.

На практике нет абсолютно управляемых объектов. На реальный объект обычно действуют как управляемые, так и неуправляемые факторы. Неуправляемые факторы влияют на воспроизводимость эксперимента и являются причиной ее нарушения. Если требования воспроизводимости не выполняются, приходится обращаться к активно-пассивному эксперименту [5].

Возможно, плохая воспроизводимость объясняется действием фактора, систематически изменяющегося (дрейфующего) во времени. Тогда нужно обращаться к специальным методам планирования [6]. Наконец, возможно, что все факторы неуправляемы. В этом случае возникает задача установления связи между параметром оптимизации и факторами по результатам наблюдений за поведением объекта, или, как говорят, по результатам пассивного эксперимента [7]. Эти случаи мы не будем рассматривать. Наша цель — изложение методов планирования экстремального эксперимента для воспроизводимых управляемых статических объектов.

Планирование экстремального эксперимента — это метод выбора количества и условий проведения опытов, минимально необходимых для отыскания оптимальных условий, т. е. для решения поставленной задачи.

Приступая к знакомству с планированием экстремального эксперимента, надо иметь в виду, что при оптимизации распространен так называемый детерминированный подход, особенно широко используемый в химии. При этом предполагается построение физической модели процесса на основании тщательного изучения механизма явлений (например, кинетики, гидродинамики), что позволяет получить математическую модель объекта в виде системы дифференциальных уравнений. Несомненно, что детерминированный и статистический (связанный с планированием эксперимента) подходы должны разумно дополнять друг друга, а не противопоставляться, как это иногда делается.

Теперь можно считать, что основные определения введены, и мы готовы перейти к детальному рассмотрению нашей задачи. Но сначала подведем итог.

* * *

В этой главе мы познакомились с основными определениями, которые используются в теории планирования экстремального эксперимента. Прежде чем приступать к эксперименту, необходимо однозначно и непротиворечиво сформулировать его цель и выбрать подходящую количественную характеристику этой цели, которую мы назвали параметром оптимизации.

Понятие «объект исследования» требует точного формального определения. Для такого определения удалось приспособить кибернетическое понятие «черный ящик» — модель объекта. Экспериментатор, вставший на путь применения методов планирования эксперимента, должен уметь формулировать свою задачу в терминах «черного ящика».

Входы «черного ящика» называются факторами. Каждый фактор может принимать некоторое определенное число различных значений, называемых уровнями. Сочетание определенных уровней всех факторов определяет возможное состояние «черного ящика» и условия одного из возможных опытов.

Совокупность всех различных возможных состояний определяет сложность «черного ящика» и общее число возможных опытов.

Результаты эксперимента используются для получения математической модели объекта исследования, которая представляет собой уравнение, связывающее параметр оптимизации и факторы. Такое уравнение называется функцией отклика.

Использование для получения модели всех возможных опытов приводит к абсурдно большим экспериментам. Задача выбора необходимых для эксперимента опытов, методов математической обработки их результатов и принятия решений — это и есть задача планирования эксперимента. Частный случай этой задачи — планирование экстремального эксперимента, т. е. эксперимента, поставленного с целью поиска оптимальных условий функционирования объекта. Планирование экстремального эксперимента — метод выбора минимального количества опытов, необходимых для отыскания оптимальных условий.

Желаем вам успеха!

Литература

1. R. A. Fisher. The Design of Experiments. 6-th ed., London, Oliver and Boyd, 1951.
2. G. E. P. Box, K. B. Wilson. On the Experimental Attainment of Optimum Conditions. — J. Roy. Statist. Soc., Ser. B, 1951, 13, N 1.
3. Н. Винер. Кибернетика. М., «Советское радио», 1968.
4. У. Р. Эшби. Введение в кибернетику. М., ИЛ, 1959.
5. Ю. П. Адлер, А. И. Ратнер, Г. Ф. Лещинская. Об активно-пассивном эксперименте. — Научные труды Гиредмета, 27. М., «Металлургия», 1969, с. 16.
6. Е. В. Маркова, А. Н. Лисенков. Планирование экспериментов в условиях неоднородностей. М., «Наука», 1973.
7. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.

ПАРАМЕТР ОПТИМИЗАЦИИ

Приплыла к нему рыбка, спросила:
Чего тебе надобно, старче?

А Пушкин Сказка о рыбаке и рыбке

При планировании экстремального эксперимента очень важно определить параметр, который нужно оптимизировать. Сделать это совсем не так просто, как кажется на первый взгляд. Цель исследования должна быть сформулирована очень четко и допускать количественную оценку. Будем называть характеристику цели, заданную количественно, параметром оптимизации. Параметр оптимизации является реакцией (откликом) на воздействие факторов, которые определяют поведение выбранной вами системы. Реакция объекта многогранна, многоаспектна. Выбор того аспекта, который представляет наибольший интерес, как раз и задается целью исследования.

При традиционном нематематическом подходе исследователь стремится как-то учесть разные аспекты, взвесить их и принять *согласованное* решение о том, какой опыт *лучше*. Однако разные экспериментаторы проведут сравнение опытов неодинаково. Различия, если хотите, одно из проявлений таланта исследователя или его бездарности.

Прежде чем сформулировать требования к параметрам оптимизации и рекомендации по их выбору, познакомимся с различными видами параметров.

2.1. Виды параметров оптимизации

В зависимости от объекта и цели исследования параметры оптимизации могут быть весьма разнообразными. Чтобы ориентироваться в этом многообразии, введем некоторую классификацию (рис. 2). Мы не стремимся к созданию полной и детальной классификации. Наша задача — построить такую условную схему, которая включала бы ряд практически важных случаев и помогала экспериментатору ориентироваться в реальных ситуациях.

Реальные ситуации, как правило, сложны. Они часто требуют одновременного учета нескольких, иногда очень многих, параметров. В принципе каждый объект может характеризоваться сразу **всей совокупностью** параметров, приведенных на рис. 2, или любым подмножеством из этой совокупности. Движение к опти-

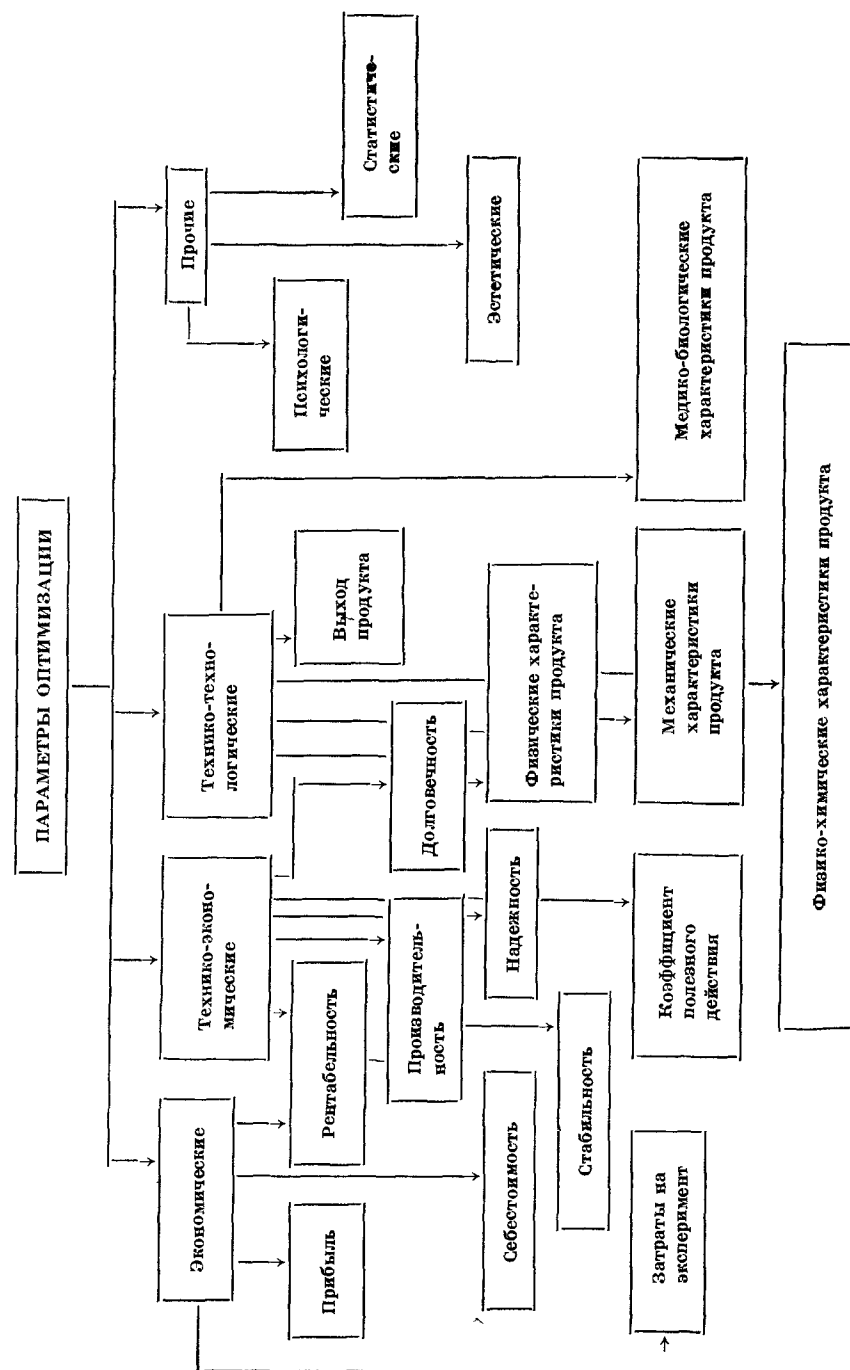


Рис. 2. Классификация параметров оптимизации

мumu возможно, если выбран один-единственный параметр оптимизации. Тогда прочие характеристики процесса уже не выступают в качестве параметров оптимизации, а служат ограничениями. Другой путь — построение обобщенного параметра оптимизации как некоторой функции от множества исходных [1—3].

Прокомментируем некоторые элементы схемы.

Экономические параметры оптимизации, такие, как прибыль, себестоимость и рентабельность, обычно используются при исследовании действующих промышленных объектов, тогда как затраты на эксперимент имеет смысл оценивать в любых исследованиях, в том числе и лабораторных. Если цена опытов одинакова (см. «Ограничения»), затраты на эксперимент пропорциональны числу опытов, которые необходимо поставить для решения данной задачи. Это в значительной мере определяет выбор плана эксперимента.

Среди технико-экономических параметров наибольшее распространение имеет производительность. Такие параметры, как долговечность, надежность и стабильность, связаны с длительными наблюдениями. Имеется некоторый опыт их использования при изучении дорогостоящих ответственных объектов, например радиоэлектронной аппаратуры.

Почти во всех исследованиях приходится учитывать количество и качество получаемого продукта. Как меру количества продукта используют выход, например, процент выхода химической реакции, выход годных изделий.

Показатели качества чрезвычайно разнообразны. В нашей схеме они сгруппированы по видам свойств. Характеристики количества и качества продукта образуют группу технико-технологических параметров.

Под рубрикой «прочие» сгруппированы различные параметры, которые реже встречаются, но не являются менее важными. Сюда попали статистические параметры, используемые для улучшения характеристик случайных величин или случайных функций. В качестве примеров назовем задачи на минимизацию дисперсии случайной величины, на уменьшение числа выбросов случайного процесса за фиксированный уровень и т. д. Последняя задача возникает, в частности, при выборе оптимальных настроек автоматических регуляторов или при улучшении свойств нитей (пряжа, искусственное волокно и др.).

С ростом сложности объекта возрастает роль психологических аспектов взаимодействия человека или животного с объектом. Так, при выборе оптимальной организации рабочего места оператора параметром оптимизации может служить число ошибочных действий в различных возможных ситуациях. Сюда относятся задачи выработки условных рефлексов типа задачи «крысы в лабиринте».

При решении задачи технической эстетики или сравнении произведений искусства возникает потребность в эстетических

параметрах. Они основаны на ранговом подходе, который будет рассмотрен ниже.

Таковы некоторые виды параметров оптимизации.

Давайте рассмотрим следующий пример [1].

Пример 1. Во время второй мировой войны несколько сот английских торговых судов на Средиземном море были вооружены зенитными орудиями для защиты от вражеских бомбардировщиков. Поскольку это мероприятие было достаточно дорогим (требовалось иметь на каждом судне боевую команду), через несколько месяцев решили оценить его эффективность. Какой из параметров оптимизации более подходит для этой цели?

Число сбитых самолетов.

Потери в судах, оснащенных орудиями, по сравнению с судами без орудий.

Если Вы считаете, что эффективность установления орудий на торговые суда можно оценить числом сбитых самолетов, то Вы вряд ли смогли бы занять пост командующего английским флотом на Средиземном море. Выбранный Вами параметр оптимизации оценивает эффективность уничтожения самолетов. В то же время ясно, что значения параметра оптимизации в этом случае будут низкими, так как существуют куда более эффективные средства для этой цели (авиация, боевой флот), чем зенитные орудия на торговых судах.

Если же Вы полагаете, что эффективность установки орудий на торговые суда можно оценить сопоставлением потерь в судах, оснащенных орудиями, с потерями в судах без орудий, то это разумный выбор параметра оптимизации, потому что основной задачей при установке орудий была защита судов. Самолеты вынуждены были теперь использовать противозенитные маневры и бомбометание с большой высоты, что уменьшало потери. Из числа атакованных самолетами торговых судов с зенитными орудиями было потоплено 10% судов, а потери в судах без орудий составили 25%. Затраты на установку орудий и содержание боевых расчетов окупались очень быстро.

2.2. Требования к параметру оптимизации

Параметр оптимизации — это признак, по которому мы хотим оптимизировать процесс. Он должен быть *количественным*, даваться числом. Мы должны уметь его измерять при любой возможной комбинации выбранных уровней факторов. Множество значений, которые может принимать параметр оптимизации, будем называть областью его определения. Области определения могут быть непрерывными и дискретными, ограниченными и неограниченными. Например, выход реакции — это параметр оптимизации с непрерывной ограниченной областью определения. Он может изменяться в интервале от 0 до 100%. Число бракованных изделий, число зерен на шлифе сплава, число кровяных телец в пробе крови — вот примеры параметров с дискретной областью определения, ограниченной снизу.

Уметь измерять параметр оптимизации — это значит располагать подходящим прибором. В ряде случаев такого прибора может не существовать или он слишком дорог. Если нет способа количественного измерения результата, то приходится воспользоваться приемом, называемым ранжированием (ранговым подходом). При этом параметрам оптимизации присваиваются оценки — ранги по заранее выбранной шкале: двухбалльной, пятибалльной и т. д. Ранговый параметр имеет дискретную ограниченную область определения. В простейшем случае область содержит два значения (да, нет; хорошо, плохо). Это может соответствовать, например, годной продукции и браку.

Ранг — это количественная оценка параметра оптимизации, но она носит условный (субъективный) характер. Мы ставим в соответствие качественному признаку некоторое число — ранг.

Для каждого физически измеряемого параметра оптимизации можно построить ранговый аналог. Потребность в построении такого аналога возникает, если имеющиеся в распоряжении исследователя численные характеристики неточны или неизвестен способ построения удовлетворительных численных оценок. При прочих равных условиях всегда нужно отдавать предпочтение физическому измерению, так как ранговый подход менее чувствителен и с его помощью трудно изучать тонкие эффекты.

Пример 2. Ваша жена решила испечь яблочный пирог по новому рецепту (аналогичный пример рассмотрен в литературе [4]). Вам, конечно, трудно остаться в стороне, и вы предлагаете ей свои услуги по оптимизации этого процесса. Цель процесса — получение вкусного пирога, но такая формулировка цели еще не дает возможности приступить к оптимизации: необходимо выбрать количественный критерий, характеризующий степень достижения цели. Можно принять следующее решение: очень вкусный пирог получает отметку 5, просто вкусный пирог — отметку 4 и т. д.

Как вы полагаете, можно ли после такого решения переходить к оптимизации процесса?

Давайте разберемся. Нам важно количественно оценить результат оптимизации. Решает ли отметка эту задачу? Конечно, потому что, как мы договорились, отметка 5 соответствует очень вкусному пирогу и т. д. Другое дело, что этот подход, называемый ранговым, часто оказывается грубым, нечувствительным. Но возможность такой количественной оценки результатов не должна вызывать сомнений.

Другие примеры рангового подхода: определение чемпиона мира по фигурному катанию или гимнастике, дегустация вин, сравнение произведений искусства и т. д. Или, если хотите, из области химии: сравнение продуктов по цвету, прозрачности, форме кристаллов.

Следующее требование: параметр оптимизации должен выражаться *одним числом*. Иногда это получается естественно, как регистрация показаний прибора. Например, скорость движения

машины определяется числом на спидометре. Чаше приходится производить некоторые вычисления. Так бывает при расчете выхода реакции. В химии часто требуется получать продукт с заданным отношением компонентов, например, $A : B = 3 : 2$. Один из возможных вариантов решения подобных задач состоит в том, чтобы выразить отношение одним числом (1,5) и в качестве параметра оптимизации пользоваться значениями отклонений (или квадратов отклонений) от этого числа.

Еще одно требование, связанное с количественной природой параметра оптимизации, — *однозначность* в статистическом смысле. Заданному набору значений факторов должно соответствовать одно с точностью до ошибки эксперимента значение параметра оптимизации. (Однако обратное неверно: одному и тому же значению параметра могут соответствовать разные наборы значений факторов.)

Для успешного достижения цели исследования необходимо, чтобы параметр оптимизации действительно оценивал эффективность функционирования системы в заранее выбранном смысле. Это требование является главным, определяющим корректность постановки задачи. «Если мы требуем победы и не знаем, что подразумеваем под этим, мы встретимся с призраком, стучащимся к нам в дверь» [5].

Представление об эффективности не остается постоянным в ходе исследования. Оно меняется по мере накопления информации и в зависимости от достигнутых результатов. Это приводит к последовательному подходу при выборе параметра оптимизации. Так, например, на первых стадиях исследования технологических процессов в качестве параметра оптимизации часто используется выход продукта. Однако в дальнейшем, когда возможность повышения выхода исчерпана, нас начинают интересовать такие параметры, как себестоимость, чистота продукта и т. д.

Говоря об оценке эффективности функционирования системы, важно помнить, что речь идет о системе в целом. Часто система состоит из ряда подсистем, каждая из которых может оцениваться своим локальным параметром оптимизации. При этом оптимальность каждой из подсистем по своему параметру оптимизации «не исключает возможности гибели системы в целом» [6].

Пример 3. При флотации сульфидной руды в лабораторных условиях изучалась эффективность применения нового реагента-пенообразователя по схеме рис. 3 [7]. В качестве параметра оптимизации выбрано извлечение (при заданном качестве) концентрата в основной флотации. После проведения эксперимента выяснилось, что реагент дает более высокий выход концентрата по сравнению с прежним пенообразователем. Как вы считаете, обоснованно ли выбран параметр оптимизации, если ставилась задача оптимизации всего процесса флотационного обогащения руды?

Параметр оптимизации — извлечение концентрата в основной флотации — при решении задачи оптимизации всего процесса

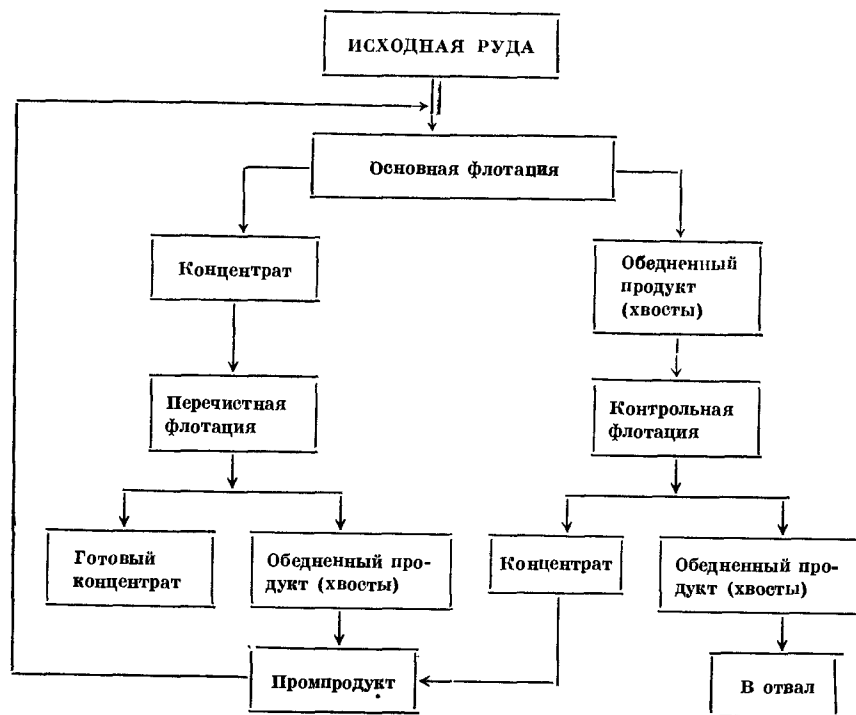


Рис. 3. Схема флотационного обогащения руды

Флотационного обогащения руды выбран не совсем обоснованно. Это правильный ответ, потому что существенно достижение конечной цели — получение готового концентрата (после перечистной флотации), а выбранный параметр оптимизации характеризует эффективность достижения промежуточной цели. Промежуточная цель — повышение выхода концентрата после основной флотации — была достигнута, но при промышленных испытаниях снизились показатели контрольной флотации, что привело к снижению извлечения и качества концентрата по всему циклу. Параметр оптимизации оказался неэффективным с точки зрения достижения конечной цели.

На эту сторону параметра оптимизации обращается внимание в книге Бира [6]: «Отличительной особенностью любой кибернетической системы можно считать полную бессмысленность рассмотрения ее иначе, как единого организма».

Мало иметь эффективный параметр оптимизации. Надо еще, чтобы он был эффективный в статистическом смысле. Понятие статистической эффективности достаточно сложное, и мы не будем здесь заниматься точными формулировками. Фактически это требование сводится к выбору параметра оптимизации, который

определяется с наибольшей возможной точностью. (Если и эта точность недостаточна, тогда приходится обращаться к увеличению числа повторных опытов.)

Пусть, например, нас интересует исследование прочностных характеристик некоторого сплава. В качестве меры прочности можно использовать как прочность на разрыв, так и макротвердость. Поскольку эти характеристики функционально связаны, то с точки зрения эффективности они эквивалентны. Однако точность измерения первой характеристики существенно выше, чем второй. Требование статистической эффективности заставляет отдать предпочтение прочности на разрыв.

Следующее требование к параметру оптимизации — требование универсальности или полноты. Под универсальностью параметра оптимизации понимается его способность всесторонне характеризовать объект. В частности, технологические параметры оптимизации недостаточно универсальны: они не учитывают экономику. Универсальностью обладают, например, обобщенные параметры оптимизации, которые строятся как функции от нескольких частных параметров [3].

Пример выбора параметра оптимизации, обладающего полнотой, рассмотрен в работе [8] для процессов зонной перекристаллизации. Обычно применяемый для этой цели коэффициент распределения, представляющий отношение концентраций примесей в твердой и жидкой фазах, излишне специфичен. Предложен более полный параметр оптимизации — энтропийная функция S

$$S = - \sum_{i=1}^m \sum_{j=1}^n c_{ij} \log c_{ij},$$

где c_{ij} — концентрация i -й примеси (при их числе m) в j -м участке слитка (при их числе n).

Желательно, чтобы параметр оптимизации имел физический смысл, был простым и легко вычисляемым.

Требование физического смысла связано с последующей интерпретацией результатов эксперимента. Не представляет труда объяснить, что значит максимум извлечения, максимум содержания ценного компонента. Эти и подобные им технологические параметры оптимизации имеют ясный физический смысл, но иногда для них может не выполняться, например, требование статистической эффективности. Тогда рекомендуется переходить к преобразованию параметра оптимизации. Преобразование, например типа $\arcsin \sqrt{y}$, может сделать параметр оптимизации статистически эффективным (например, дисперсии становятся однородными), но остается неясным: что же значит достигнуть экстремума этой величины?

Второе требование часто также оказывается весьма существенным. Для процессов разделения термодинамические параметры

оптимизации более универсальны. Однако на практике ими пользуются мало: их расчет довольно труден.

Пожалуй, из этих двух требований первое является более существенным, потому что часто удается найти идеальную характеристику системы и сравнить ее с реальной характеристикой. Иногда при этом целесообразно нормировать параметр с тем, чтобы он принимал значения от нуля до единицы.

Кроме высказанных требований и пожеланий при выборе параметра оптимизации нужно еще иметь в виду, что параметр оптимизации в некоторой степени оказывает влияние на вид математической модели исследуемого объекта. Экономические параметры, в силу их аддитивной природы, легче представляются простыми функциями, чем физико-химические показатели. Не случайно методы линейного программирования, основанные на простых моделях, получили широкое распространение именно в экономике. Температура плавления сплава является, как известно, сложной, многоэкстремальной характеристикой состава, тогда как стоимость сплава зависит от состава линейно.

Итак, вы наверное уже поняли, что найти параметр оптимизации, удовлетворяющий всем требованиям, все равно, что поймать жар-птицу.

2.3. О задачах с несколькими выходными параметрами

Какое бы качество вы ни захотели оценить, всегда найдутся по меньшей мере три противоречивых критерия его оценки.

Экон. Физики продолжают шутить

Задачи с одним выходным параметром имеют очевидные преимущества. Но на практике чаще всего приходится учитывать несколько выходных параметров. Иногда их число довольно велико. Так, например, при производстве резиновых и пластмассовых изделий приходится учитывать физико-механические, технологические, экономические, художественно-эстетические и другие параметры (прочность, эластичность, относительное удлинение, способность смеси прилипать к форме и т. д.). Математические модели можно построить для каждого из параметров, но одновременно оптимизировать несколько функций невозможно.

Обычно оптимизируется одна функция, наиболее важная с точки зрения цели исследования, при ограничениях, налагаемых другими функциями. Поэтому из многих выходных параметров выбирается один в качестве параметра оптимизации, а остальные служат ограничениями. Всегда полезно исследовать возможность уменьшения числа выходных параметров. Для этого можно воспользоваться корреляционным анализом.

При этом между всевозможными парами параметров необходимо вычислить коэффициент парной корреляции, который является общепринятой в математической статистике характеристикой связи между двумя случайными величинами. Если обозначить один параметр через y_1 , а другой — через y_2 , и число опытов, в которых они будут измеряться, — через N , так, что $u=1, 2, \dots, N$, где u — текущий номер опыта, то коэффициент парной корреляции r вычисляется по формуле

$$r_{y_1 y_2} = \frac{\sum_{u=1}^N (y_{1u} - \bar{y}_1) (y_{2u} - \bar{y}_2)}{\sqrt{\sum_{u=1}^N (y_{1u} - \bar{y}_1)^2 \sum_{u=1}^N (y_{2u} - \bar{y}_2)^2}}.$$

Здесь

$$\bar{y}_1 = \sum_{u=1}^N y_{1u} / N \text{ и } \bar{y}_2 = \sum_{u=1}^N y_{2u} / N$$

средние арифметические соответственно для y_1 и y_2 .

Значения коэффициента парной корреляции могут лежать в пределах от -1 до $+1$. Если с ростом значения одного параметра возрастает значение другого, у коэффициента будет знак плюс, а если уменьшается, то минус. Чем ближе найденное значение $r_{y_1 y_2}$ к единице, тем сильнее значение одного параметра зависит от того, какое значение принимает другой, т. е. между такими параметрами существует линейная связь, и при изучении процесса можно рассматривать только один из них. Необходимо помнить, что коэффициент парной корреляции как мера тесноты связи имеет четкий математический смысл только при линейной зависимости между параметрами и в случае нормального их распределения.

Для проверки значимости коэффициента парной корреляции нужно сравнить его значение с табличным (критическим) значением r , которое приведено в табл. 2.1. Для пользования этой таблицей нужно знать число степеней свободы $f=N-2$ и выбрать определенный уровень значимости, например, равный 0,05. Такое значение уровня значимости называют еще 5%-ным уровнем риска, что соответствует вероятности верного ответа при проверке нашей гипотезы $P=1-\alpha=0,95$, или 95%. Это значит, что в среднем только в 5% случаев возможна ошибка при проверке гипотезы.

В практических исследованиях 5%-ный уровень риска применяется наиболее часто. Но экспериментатор всегда свободен в выборе уровня значимости, и возможны ситуации, в которых, например, требуется 1%-ный уровень риска. При этом воз-

Таблица 2. 1

Критические значения коэффициента парной корреляции при $\alpha = 0,05$

Число степеней свободы f	Критическое значение r	Число степеней свободы f	Критическое значение r	Число степеней свободы f	Критическое значение r
1	0,997	9	0,602	17	0,456
2	0,950	10	0,576	18	0,444
3	0,878	11	0,553	19	0,433
4	0,811	12	0,532	20	0,423
5	0,754	13	0,514	30	0,349
6	0,707	14	0,497	50	0,273
7	0,666	15	0,482	80	0,217
8	0,632	16	0,468	100	0,195

растает надежность ответа. Проверка гипотезы сводится к сравнению абсолютной величины коэффициента парной корреляции с критическим значением. Если экспериментально найденное значение r меньше критического, то нет оснований считать, что имеется тесная линейная связь между параметрами, а если больше или равно, то гипотеза о корреляционной линейной связи не отвергается [9].

При высокой значимости коэффициента корреляции любой из двух анализируемых параметров можно исключить из рассмотрения как не содержащий дополнительной информации об объекте исследования. Исключить можно тот параметр, который технически труднее измерять, или тот, физический смысл которого менее ясен. При планировании эксперимента целесообразно измерять все параметры, затем оценить корреляцию между ними и строить модели для их минимально возможного числа или же воспользоваться обобщенным параметром. Но бывают случаи, когда приходится рассматривать и коррелированные параметры.

2.4. Резюме

Мы познакомились с некоторыми практически важными аспектами весьма сложной проблемы — выбора параметра оптимизации. Параметр оптимизации — это реакция (отклик) на воздействия факторов, которые определяют поведение изучаемой системы.

Параметры оптимизации бывают экономическими, технико-экономическими, технико-технологическими, статистическими, психологическими и т. д.

Параметр оптимизации должен быть:
 эффективным с точки зрения достижения цели;
 универсальным;
 количественным и выражаться одним числом;

статистически эффективным;

имеющим физический смысл, простым и легко вычисляемым; существующим для всех различных состояний.

В тех случаях, когда возникают трудности с количественной оценкой параметров оптимизации, приходится обращаться к ранговому подходу. В ходе исследования могут меняться априорные представления об объекте исследования, что приводит к последовательному подходу при выборе параметра оптимизации.

Из многих параметров, характеризующих объект исследования, только один, часто обобщенный, может служить параметром оптимизации. Остальные рассматриваются как ограничения.

Мы хотели бы, чтобы данная глава помогла вам при выборе параметра оптимизации в тех конкретных исследованиях, которые вы проводите.

Литература

1. Х. Гуд, Р. Э. Макол. Системотехника. Введение в проектирование больших систем. М., «Советское радио», 1962.
2. Дж. фон Нейман, О. Моргенштерн. Теория игр и экономическое поведение. М., «Наука», 1972.
3. В. М. Добкин. Выбор экономических критериев оптимизации режимных и конструктивных параметров реакторов. — Химическая промышленность, 1968, № 3.
4. H. Smith, A. Rose. Subjective responses in process investigation. — Industr. and Engng Chem. 1963, 55, N 7.
5. Н. Винер. Кибернетика. М., «Советское радио», 1968.
6. Ст. Бир. Кибернетика и управление производством. М., «Наука», 1965.
7. Л. А. Барский, И. Н. Плаксин. Критерии оптимизации разделительных процессов. М., «Наука», 1967.
8. В. Н. Виздорович, Ю. П. Адлер, А. Е. Вольян. Об оценке эффективности процессов зонной перекристаллизации. — Изв. АН СССР, Металлургия и горное дело, 1964, № 2.
9. В. В. Налимов. Применение математической статистики при анализе вещества. М., Физматгиз, 1960.

ОБОБЩЕННЫЙ ПАРАМЕТР ОПТИМИЗАЦИИ

Переключаются звук, запах, форма, цвет,
Глубокий, темный смысл обретшие в слиянье.

Ш. Бодлер. Цветы зла

Путь к единому параметру оптимизации часто лежит через обобщение. Как уже говорилось, из многих откликов, определяющих объект, очень часто трудно выбрать один, самый важный. Если же это возможно, тогда мы попадаем в ситуацию, описанную в предыдущей главе. Здесь же нам предстоит познакомиться с более сложной ситуацией, когда необходимо множество откликов обобщать (свертывать) в единый количественный признак. С таким обобщением связан ряд трудностей.

Каждый отклик имеет свой физический смысл и свою размерность. Чтобы объединить различные отклики, прежде всего приходится ввести для каждого из них некоторую безразмерную шкалу. Шкала должна быть однотипной для всех объединяемых откликов — это делает их сравнимыми. Выбор шкалы — не простая задача, зависящая от априорных сведений об откликах, а также от той точности, с которой мы хотим определить обобщенный признак.

После того как для каждого отклика построена безразмерная шкала, возникает следующая трудность — выбор правила комбинирования исходных частных откликов в обобщенный показатель. Единого правила не существует. Здесь можно идти различными путями, и выбор пути неформализован. Рассмотрим несколько различных способов построения обобщенного показателя.

3.1. Простейшие способы построения обобщенного отклика

Пусть исследуемый объект характеризуют n частных откликов y_u ($u=1, 2, \dots, n$) и каждый из этих откликов измеряется в N опытах. Тогда y_{ui} — это значение u -го отклика в i -м опыте ($i=1, 2, \dots, N$). Каждый из откликов y_u имеет свой физический смысл и, чаще всего, разную размерность. Введем простейшее преобразование: набор данных для каждого y_u поставим в соответствие с самым простым стандартным аналогом — шкалой, на которой имеется только два значения: 0 — брак, удовлетво-

рительное качество, 1 — годный продукт, удовлетворительное качество. Преобразованные значения обозначим так: y_{ui} — преобразованное значение u -го отклика в i -м опыте. Здесь мы применили шкалу, в которой использовано числовое множество из двух элементов (в данном случае 0 и 1). Стандартизовав таким образом шкалу частных откликов, мы подошли ко второму этапу — их обобщению. По какому же правилу следует комбинировать частные отклики?

Будем рассуждать следующим образом. В ситуации, когда каждый преобразованный частный отклик принимает только два значения 0 и 1, естественно желать, чтобы и обобщенный отклик принимал одно из этих двух возможных значений, причем так, чтобы значение 1 имело место, если, и только если, все частные отклики в этом опыте приняли значение 1. А если хотя бы один из откликов обратился в 0, то и обобщенный отклик будет нулем.

При таких рассуждениях для построения обобщенного отклика удобно воспользоваться формулой

$$Y_i = \sqrt[n]{\prod_{u=1}^n y_{ui}},$$

где Y_i — обобщенный отклик в i -м опыте; $\prod_{u=1}^n$ — произведение частных откликов $y_{1i} y_{2i} \dots y_{ni}$.

Корень мы ввели для того, чтобы связать эту формулу с другой, более сложной, которая будет рассмотрена далее. В данном же случае ничего не изменится, если написать

$$Y_i = \prod_{u=1}^n y_{ui}.$$

Совершенно очевидно, что такой подход слишком груб и жесток [1]. Рассмотрим пример [2].

Пример 1. При разработке оптимальной рецептуры нового пластифицированного полимерного материала качество продукции оценивалось семью выходными параметрами: y_1 — термостабильность композиции, мин; y_2 — блеск (баллы), y_3 — морозостойкость, °C; y_4 — модуль упругости при +20° C, кгс/см²; y_5 — предел прочности при растяжении, кгс/см²; y_6 — относительное удлинение при разрыве, %; y_7 — число перегибов до разрушения, шт. Для каждого частного отклика введем следующие преобразования:

$$\begin{aligned} y_{1i} &= \begin{cases} 1, & \text{если } y_{1i} > 100; \\ 0, & \text{если } y_{1i} \leq 100; \end{cases} & y_{2i} &= \begin{cases} 1, & \text{если } y_{2i} < 20; \\ 0, & \text{если } y_{2i} \geq 20. \end{cases} \\ y_{3i} &= \begin{cases} 1, & \text{если } y_{3i} < -18; \\ 0, & \text{если } y_{3i} \geq -18; \end{cases} & y_{4i} &= \begin{cases} 1, & \text{если } y_{4i} < 120; \\ 0, & \text{если } y_{4i} \geq 120; \end{cases} \end{aligned}$$

$$y_{5i} = \begin{cases} 1, & \text{если } y_{5i} > 200; \\ 0, & \text{если } y_{5i} \leq 200; \end{cases} \quad y_{6i} = \begin{cases} 1, & \text{если } y_{6i} > 200; \\ 0, & \text{если } y_{6i} \leq 200; \end{cases}$$

$$y_{7i} = \begin{cases} 1, & \text{если } y_{7i} > 25; \\ 0, & \text{если } y_{7i} \leq 25. \end{cases}$$

Данные для девяти опытов приведены в табл. 3.1.

Для комплексной оценки материалов были построены два обобщенных показателя:

$$Y_1 = \sqrt[7]{y_1 y_2 y_3 y_4 y_5 y_6 y_7},$$

учитывающий требования и разработки и заказчика, и

$$Y_2 = \sqrt[3]{y_3 y_5 y_7},$$

учитывающий только основные требования заказчика. По обобщенному показателю Y_1 могут рекомендоваться только три рецептуры (каждый опыт соответствует одной определенной рецептуре). Если же взять за основу только требования заказчика, то удовлетворительное качество имеют материалы в семи опытах.

Таблица 3.1

Натуральные, преобразованные и обобщенные отклики

Номер опыта	Натуральные частные отклики							Преобразованные частные отклики							Обобщенные отклики	
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_1	y_2	y_3	y_4	y_5	y_6	y_7	Y_1	Y_2
1	272	14	-25	103	215	299	103	1	1	1	1	1	1	1	1	1
2	187	20	-23	91	179	254	29	1	0	1	1	0	1	1	0	0
3	162	21	-24	102	216	270	99	1	0	1	1	1	1	1	0	1
4	461	14	-25	114	198	251	54	1	1	1	1	0	1	1	0	0
5	267	14	-21	105	208	268	31	1	1	1	1	1	1	1	1	1
6	250	24	-27	99	220	304	46	1	0	1	1	1	1	1	0	1
7	489	12	-25	123	201	238	33	1	1	1	0	1	1	1	0	1
8	380	14	-23	116	230	292	126	1	1	1	1	1	1	1	1	1
9	580	29	-22	100	215	304	48	1	0	1	1	1	1	1	0	1

Обратим Ваше внимание на еще один способ получения обобщенного отклика, который может применяться в тех случаях, когда для каждого из частных откликов известен «идеал», к которому нужно стремиться. Существует много способов введения метрики*, задающей «близость к идеалу». Дополним предыдущие обозначения еще одним: y_{i0} — наилучшее («идеальное») значение i -го отклика. Тогда $y_{ii} - y_{i0}$ можно рассматривать как некоторую меру близости к идеалу. Однако использовать

* Ввести метрику — значит указать правило определения расстояний между любыми парами объектов из интересующего нас множества.

разность при построении обобщенного отклика невозможно по двум причинам. Она имеет размерность соответствующего отклика, а у каждого из откликов может быть своя размерность, что препятствует, как мы уже знаем, их объединению. Отрицательный или положительный знак разности также создает неудобство. Чтобы перейти к безразмерным значениям, достаточно разность поделить на желаемое значение: $(y_{ii} - y_{i0})/y_{i0}$.

Чтобы нивелировать знаки, можно разность возводить в квадрат. Тогда обобщенный отклик получим по следующей формуле:

$$Y_1 = \sum_{i=1}^n \left(\frac{y_{ii} - y_{i0}}{y_{i0}} \right)^2.$$

Если в некотором опыте все частные отклики совпадут с идеалом, то Y станет равным нулю. Это и есть то значение, к которому нужно стремиться. Чем ближе к нулю, тем лучше. Конечно, необходимо условиться о том, что считать нижней границей, если верхняя равна нулю. Обратите внимание, нуль здесь имеет другой смысл, чем в первом случае.

Среди недостатков такой оценки выделяется нивелировка частных откликов. Все они входят в обобщенный отклик на равных правах. На практике же различные показатели бывают далеко не равноправны. Устранить этот недостаток можно введением некоторого веса a_u

$$Y_1 = \sum_{i=1}^n a_u \left(\frac{y_{ii} - y_{i0}}{y_{i0}} \right)^2,$$

причем

$$\sum_{i=1}^n a_u = 1 \text{ и } a_u > 0.$$

Чтобы проранжировать отклики по степени их важности и найти соответствующие веса, можно воспользоваться экспертными оценками [3, 4].

Мы рассмотрели простейшие способы построения обобщенного показателя. Для перехода к более сложным способам нужно научиться фиксировать более тонкие различия на шкале преобразования откликов. Здесь в основном приходится опираться на опыт экспериментатора. Но, чтобы этот опыт разумно употребить в рамках формальных процедур, его тоже нужно формализовать. Наиболее естественный путь такой формализации — введение системы предпочтений экспериментатора на множестве значений каждого частного отклика, получение стандартной шкалы и затем обобщение результатов.

Пользуясь системой предпочтений, можно получить более содержательную шкалу вместо шкалы классификаций с двумя классами. К примеру построения такой шкалы мы перейдем в следующем параграфе.

3.2. Шкала желательности

Одним из наиболее удобных способов построения обобщенного отклика является обобщенная функция желательности Харрингтона [5, 6]. В основе построения этой обобщенной функции лежит идея преобразования натуральных значений частных откликов в безразмерную шкалу желательности или предпочтительности. Шкала желательности относится к психофизическим шкалам. Ее назначение — установление соответствия между физическими и психологическими параметрами. Здесь под физическими параметрами понимаются всевозможные отклики, характеризующие функционирование исследуемого объекта. Среди них могут быть эстетические и даже статистические параметры (см. классификацию параметров во второй главе), а под психологическими параметрами понимаются чисто субъективные оценки экспериментатора желательности (предпочтительности) того или иного значения отклика. Чтобы получить шкалу желательности, удобно пользоваться готовыми разработанными таблицами соответствий между отношениями предпочтения в эмпирической и числовой (психологической) системах (табл. 3.2).

Таблица 3.2

Стандартные отметки на шкале желательности

Желательность	Отметки на шкале желательности
Очень хорошо	1,00—0,80
Хорошо	0,80—0,63
Удовлетворительно	0,63—0,37
Плохо	0,37—0,20
Очень плохо	0,20—0,00

Значение частного отклика, переведенное в безразмерную шкалу желательности, обозначается через d_u ($u=1, 2, \dots, n$) и называется частной желательностью (от desirable фр. — желательный). Шкала желательности имеет интервал от нуля до единицы. Значение $d_u=0$ соответствует абсолютно неприемлемому уровню данного свойства, а значение $d_u=1$ — самому лучшему значению свойства. Понятию «очень хорошо» соответствуют значения на шкале желательности $1 > d_u > 0,8$, а понятию

«очень плохо» — $0 < d_u < 0,2$ и т. д. Выбор отметок на шкале желательности 0,63 и 0,37 объясняется удобством вычислений: $0,63 \approx 1 - (1/e)$, $0,37 \approx 1/e$. Значение $d_u=0,37$ обычно соответствует границе допустимых значений:

Эти числа на первый взгляд напоминают черную магию. Но все объясняется достаточно просто. В табл. 3.2 представлены числа, соответствующие некоторым точкам кривой (рис. 4), которая задается уравнением $d = e^{-e^{-y}}$ или $d = \exp[-\exp(-y)]$, где \exp — принятое обозначение экспоненты. На оси ординат нанесены значения желательности, изменяющиеся от 0 до 1. По оси абсцисс указаны значения отклика, записанные в условном масштабе. За начало отсчета 0 по этой оси выбрано значение, соответствующее желательности 0,37. Выбор именно данной точки связан с тем, что она является точкой перегиба кривой, что в свою очередь создает определенные удобства при вычислениях. То же самое верно для значения желательности, соответствующего 0,63. Выбор этой кривой не является единственной возможностью. Однако она возникла в результате наблюдений за реальными решениями экспериментаторов и обладает такими полезными свойствами как непрерывность, монотонность и гладкость. Кроме того, эта кривая хорошо передает тот факт, что в областях желательностей, близких к 0 и 1, «чувствительность» ее существенно ниже, чем в средней зоне.

Симметрично относительно нуля на оси y' (y' — кодированная шкала) расположены кодированные значения отклика. Значение на кодированной шкале принято выбирать от 3 до 6. Например, на рис. 4 использовано шесть интервалов в сторону убывания и шесть — в сторону возрастания. Ниже вы познакомитесь со случаем, когда выбрано три интервала. Выбор числа интервалов определяет крутизну кривой в средней зоне.

Пример 1. Пусть среди откликов будет выход реакции, естественные границы которого заключены между 0% и 100%.

Предположим, что 100% соответствует на шкале желательности единице, а 0% — нулю. Это отображено на рис. 4 на шкале, расположенной под шкалой y' . На ней пока имеются только две точки: 0 и 100. Выбор других точек зависит от ряда обстоятельств, таких, как сложившаяся в начальный момент ситуация, требования к результату, возможности экспериментатора.

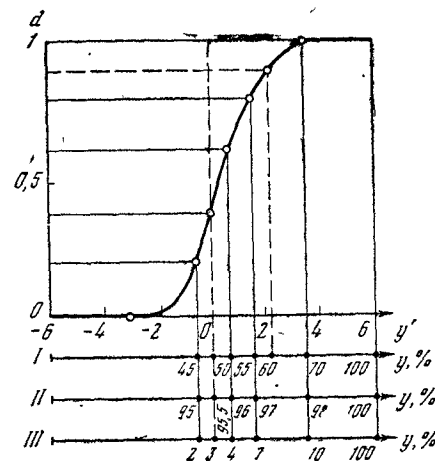


Рис. 4. Функция желательности

Можно себе представить случай, когда процесс идет с выходом около 50%. О выходе более 70% трудно мечтать, а может быть он и невозможен из-за побочных реакций. Из сказанного ясно, что в данном случае выход в 70% для экспериментатора практически эквивалентен выходу в 100%. Это и отражено при выборе второй точки справа, для которой желательность близка к единице. Третья точка ограничивает область, которой экспериментатор может дать оценку «очень хорошо». В табл. 3.2 это соответствует диапазону 1,00—0,80. Какое натуральное значение y поставить в соответствие этой точке, зависит от «азартности» или «осторожности» экспериментатора. Построение этой шкалы напоминает игру. Именно из теории игр заимствована эта терминология. Азарт азартом, но нужно еще и уметь играть. Опытный химик может сказать, что «больше 60% из этой реакции вы не выжмете». Тогда бы мы поставили в соответствие третьей точке выход 60%.

Конечно, такое рассуждение может не иметь смысла, если вследствие ошибок эксперимента мы не в состоянии различить выхода в 60 и 70%. Следовательно, разным точкам на шкале должны соответствовать разные значения выхода. «Азартный» экспериментатор может поставить в соответствие выход, скажем, в 67%. Область хороших результатов (0,80—0,63 по шкале желательности) наш экспериментатор заключил в границы 60—55%. Уже достигнутый результат — 50%, дает естественную нижнюю границу области «удовлетворительно», ну а результат ниже 45% просто никуда не годится. На рис. 4 разобранная ситуация обозначена (I).

Но все может быть и иначе. Имеется хорошо налаженный технологический процесс с выходом 95%, но возросли требования к чистоте продукта. Поэтому, наряду с другими мероприятиями, возникла задача увеличения выхода хотя бы до 98%. Эта ситуация на нашем рисунке обозначена (II). Конечно, такая шкала возможна только при прецизионных измерениях.

Совсем другая картина получается, когда речь идет о задаче синтеза нового вещества, которого до сих пор не удавалось получить в количествах, достаточных для идентификации.

При выходе менее 2% нет способа в данном случае идентифицировать продукт. Любой выход выше 10% — превосходен. Эта ситуация (III), по видимому, не нуждается в дальнейших комментариях.

Кривую желательности обычно используют как номограмму, поскольку это легко и оперативно. Так, если в ситуации (I) получен выход 63%, то ему будет соответствовать желательность 0,9, что показано пунктиром на рисунке. На практике такой простой прием часто оказывается вполне достаточным, но не всегда. Во-первых, потому что точность графического определения желательности может оказаться недостаточной, а, во-вторых, потому что эта точность зависит от положения на шкале y .

Тогда приходится прибегать к аналитическому методу определения желательности и с помощью шкалы y' и приведенного выше уравнения находить оценки желательности.

В нашем примере рассмотрен только один отклик — выход реакции, и границы его заданы естественно: от 0 до 100%. Однако, это не всегда бывает так. Стоит включить в рассмотрение отклики,

характеризующие, например, качество материала, и границы станут весьма неопределенными.

Естественно возникает вопрос, на каком основании устанавливаются границы допустимых значений для частных откликов. При этом нужно иметь в виду, что ограничения могут быть односторонними в виде $y_u \leq y_{\max}$ или $y_u \geq y_{\min}$ и двусторонними в виде $y_{\min} \leq y_u \leq y_{\max}$. Здесь возможны две ситуации. Первая, самая благоприятная, возможна, если экспериментатор располагает инструкцией, в которой четко сформулированы требования ко всем частным откликам, т. е. имеется спецификация с одним или двумя ограничивающими пределами. Тогда отметка на шкале желательности $d_u=0,37$ соответствует y_{\min} , если имеется одностороннее ограничение $y_u \geq y_{\min}$ или y_{\max} для $y_u \leq y_{\max}$. В случае двустороннего ограничения этой отметке ставится в соответствие и y_{\min} , и y_{\max} . Во второй ситуации спецификация отсутствует, тогда ограничения на шкале и другие отметки делаются весьма субъективно, на основании опыта и интуиции экспериментатора. Совершенно очевидно, что в таком случае довольно опасно руководствоваться мнением одного-единственного исследователя, а желательно учесть мнения нескольких специалистов. При обобщении ряда мнений и установлении степени согласованности между различными специалистами можно воспользоваться методом ранговой корреляции [4, 7].

3.3. Преобразование частных откликов в частные функции желательности

Представим себе такой случай, когда существует спецификация с одним или двумя ограничивающими пределами и эти пределы являются единственными значениями качества. Тогда вне пределов $d_u=0$, а внутри пределов $d_u=1$.

Пусть y_{\min} — нижний предел спецификации и $y_u \geq y_{\min}$. При таком одностороннем ограничении частная функция желательности будет иметь вид

$$d_u = \begin{cases} 0, & \text{если } y_u < y_{\min}; \\ 1, & \text{если } y_u \geq y_{\min}. \end{cases}$$

Аналогично для двустороннего ограничения получаем

$$d_u = \begin{cases} 0, & \text{если } y_u < y_{\min} \text{ и } y_u > y_{\max}; \\ 1, & \text{если } y_{\min} \leq y_u \leq y_{\max}. \end{cases}$$

Нетрудно видеть, что мы пришли к случаю, рассмотренному в § 3.1: шкала желательности выродилась в простейшую шкалу классификаций с двумя классами эквивалентности. Эти два случая показаны на рис. 5. Но такое положение, когда ограничи-

являющиеся пределы спецификации являются единственными критериями качества, встречается довольно редко, так как трудно разделить результаты твердой границей на две категории: годен, не годен. Поэтому преобразования частных откликов в их стандартные аналоги на шкале желательности осуществляются по более сложным законам. Примером такого более сложного преобразования служит таблица желательности (табл. 3.2) и соответствующая ей функция желательности.

При односторонних ограничениях $y_u \leq y_{\max}$ или $y_u \geq y_{\min}$ на рис. 6 представлена частная функция желательности для

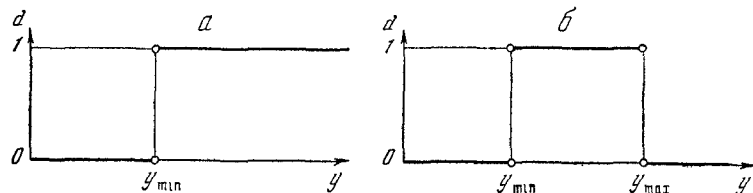


Рис. 5. Задание частной функции желательности

а — одностороннее ограничение; б — двустороннее ограничение

свойства, ограниченного с одной стороны. К числу свойств, подчиняющихся одностороннему ограничению, относятся очень многие характеристики качества материалов: теплостойкость, прочность, ударная вязкость, морозостойкость, модуль упругости, относительное удлинение при разрыве и т. д. (это можно было видеть из примера 1). Для всех этих показателей ограничение имело вид $y_u \geq y_{\min}$.

Другой вид одностороннего ограничения $y_u \leq y_{\max}$ характерен для таких показателей, как содержание различных вредных примесей, влажность, удельный вес, содержание дорогих или дефицитных компонентов и т. п.

Для двустороннего ограничения $y_{\min} \leq y_u \leq y_{\max}$ пример функции желательности показан на рис. 7. Случай двустороннего ограничения встречается не так часто, как одностороннего, и он более сложен для оценки откликов. Примерами могут служить молекулярный вес материалов, предназначенных для переработки литьем или экструзией, насыпной вес и вес утряски, индекс расплава и т. п. Индекс расплава, например, является важной характеристикой литьевых и экструзионных типов полиметилметакрилата. Слишком низкий индекс наблюдается у материалов с недостаточной текучестью, что отрицательно влияет на ход технологического процесса при переработке полимера. При высоком показателе индекса расплава возникают затруднения во время формования материала.

Построение функции желательности детально рассмотрено Т. М. Карташовой [2]. В некоторых сложных случаях к построению шкалы желательности приходится подходить итеративно.

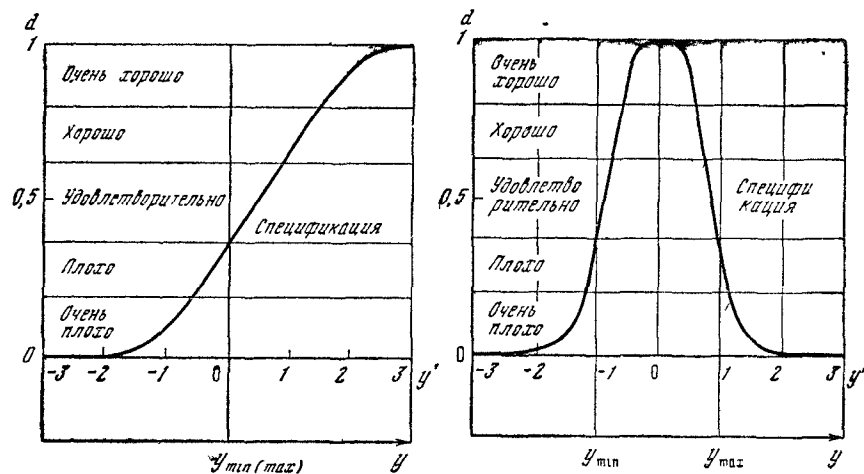


Рис. 6. Функция желательности для свойства, ограниченного с одной стороны

Рис. 7. Функция желательности для свойства, ограниченного с двух сторон

Шкала желательности есть попытка формализации представлений экспериментатора о важности тех или других значений частных откликов. Нет никакой гарантии, что такие представления можно считать правильными. Особенно, если не существует спецификации: В тех случаях, когда шкала желательности создается совместно экспериментатором и консультантом по планированию эксперимента, можно представить себе и такую ситуацию, когда не было между ними достаточно глубокого взаимопонимания и шкала получилась неудачной. Тогда нужно учесть все недостатки и построить другую шкалу, более точно отражающую предпочтительность значений откликов.

3.4. Обобщенная функция желательности

После того, как выбрана шкала желательности и частные отклики преобразованы в частные функции желательности, можно приступить к основной задаче — построению обобщенного показателя D , названного Харрингтоном *обобщенной функцией желательности*. Обобщать, то есть переходить от d_u^* к D , предлагается по формуле

$$D = \sqrt[n]{\prod_{u=1}^n d_u^*}$$

Здесь обобщенная функция желательности задается как среднее геометрическое частных желательностей. Такое представление можно рассматривать как удобную модель психологической

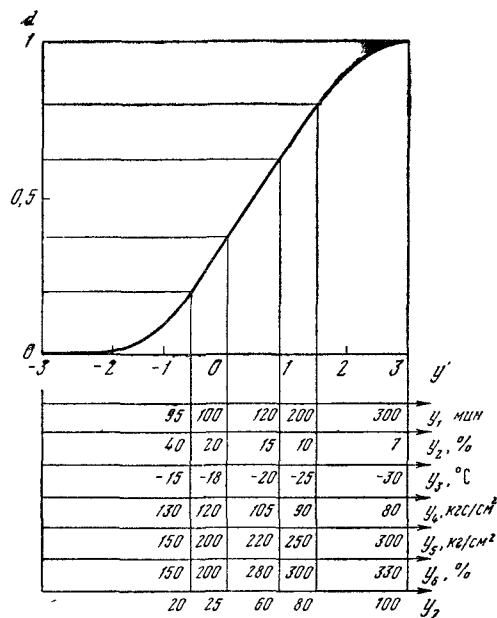


Рис. 8. Шкала желательности для нового полимерного материала

ности таков, что если хотя бы одна частная желательность $d_u=0$, то обобщенная функция тоже будет равна нулю, с другой стороны $D=1$ тогда и только тогда, когда все $d_u=1$ ($u=1, 2, \dots, n$). Обобщенная функция желательности весьма чувствительна к малым значениям частных желательностей. Можно представить себе другие задачи, когда столь жесткие свойства обобщенного критерия окажутся неприемлемыми. Тогда нужно перейти к другим способам обобщения. А пока вернемся к функции желательности.

Способ задания базовых отметок шкалы желательности, представленный в табл. 3.2, один и тот же как для частных желательностей, так и для обобщенной. Так, если $d_1, d_2, \dots, d_n=0,63$, то и $D=0,63$, если $d_1, d_2, \dots, d_n=0,37$, то и $D=0,37$ и т. п. В обобщенную функцию желательности могут входить самые разнообразные частные отклики: технологические, технико-экономические, экономические, эстетические и т. п.

В первом параграфе этой главы мы рассмотрели в примере 1 построение обобщенного отклика с использованием шкалы классификации с двумя классами эквивалентности 0 и 1. Теперь для этих же исходных данных построим обобщенную функцию желательности.

реакции исследователя, которая возникает при решении определенного класса задач. Примером может служить установление пригодности материала с данным набором свойств для использования его в заданных условиях.

Если хотя бы один частный отклик, входящий в комплекс параметров качества материала, не удовлетворяет требованиям спецификации (например, при определенной температуре материал становится хрупким и разрушается), то как бы ни были хороши прочие свойства, этот материал не может быть использован по назначению. Действительно, способ задания обобщенной функции желательности

Пример 2. Пусть полимерный материал характеризуют те же семь частных откликов, как и в примере 1. Воспользуемся более тонким способом преобразования откликов в безразмерные величины, чем мы это сделали в предыдущем примере, и построим для них шкалу желательности (рис. 8). Натуральные значения частных откликов преобразуем в частные функции желательности (см. табл. 3.3).

По обобщенному критерию D_1 четыре рецептуры получили оценку хорошо, пять — удовлетворительно, а по критерию D_2 , учитывающего только требование заказчика, три рецептуры оценены хорошими, шесть удовлетворительными. Если сравнить эти данные с примером 1, в котором шкала принимала только два значения (0 — брак и 1 — годный продукт), то становятся совершенно очевидными преимущества функции желательности как более тонкого обобщенного критерия.

Мы рассмотрели самый простой способ получения обобщенного критерия — графический. Но часто приходится прибегать к аналитической зависимости, особенно при систематических исследованиях одного и того же объекта. Тогда можно использовать полиномы невысоких степеней, обычно первой или второй степени. Проверка пригодности того или иного полинома является специальной статистической задачей. Если вы дочитаете эту книгу до конца десятой главы, то вам, вероятно, станет ясно, как ее решать.

Располагая информацией о y'_i , мы можем рассчитать d_i по уже упоминавшейся формуле

$$d_i = \exp[-\exp(-y'_i)].$$

Для получения кодированных значений y'_i в данном примере* взяты три равномерных интервала, т. е. выбран код: $-3, -2, -1, 0, +1, +2, +3$. Такой код можно поставить в соответствие абсциссе однажды построенной кривой. Если требуется регулировать ее крутизну, то это можно сделать изменением числа интервалов. Чтобы приспособить такую стандартную кривую к реальным частным откликам, достаточно дополнить табл. 3.2 следующей информацией:

d_i	y'_i	y_1	y_2	y_3	y_4	y_5	y_6	y_7
1,00—0,80	3,0	300	7	—30	80	300	330	100
0,80—0,63	1,5	200	10	—25	90	250	280	80
0,63—0,37	0,85	120	15	—20	105	220	250	60
0,37—0,20	0,00	100	20	—18	120	200	200	25
0,20—0,00	—0,50	95	40	—15	130	150	150	20

Обобщенная функция желательности является некоторым абстрактным построением и поэтому, прежде чем рекомендовать ее в качестве единого критерия оптимизации, представлялось интересным исследовать такие ее важные свойства, как адекватность,

* Кодированные значения откликов относятся к верхней границе интервала.

Таблица 3.3

Натуральные и обобщенные по функции желательности отклики

Номер опыта	Натуральные значения откликов							Частные желательности							Оцен-ка по шкале желательности	Оцен-ка по шкале желательности
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	d_1	d_2	d_3	d_4	d_5	d_6	d_7	D_1	D_2
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	d_1	d_2	d_3	d_4	d_5	d_6	d_7	D_1	D_2
1	272	14	-25	103	215	299	103	0.98	0.67	0.80	0.71	0.55	0.80	1.00	0.772	хор.
2	187	20	-23	91	179	254	29	0.77	0.36	0.73	0.79	0.32	0.53	0.29	0.500	уд.
3	162	21	-24	102	216	270	99	0.68	0.35	0.75	0.72	0.55	0.56	0.99	0.630	хор.
4	461	14	-25	114	198	251	54	1.00	0.67	0.80	0.47	0.38	0.53	0.57	0.603	уд.
5	257	14	-21	105	208	268	31	0.97	0.67	0.70	0.63	0.44	0.54	0.39	0.595	уд.
6	250	24	-27	99	220	304	46	0.95	0.32	0.90	0.70	0.63	0.81	0.48	0.646	уд.
7	489	12	-25	123	201	238	33	1.00	0.72	0.80	0.35	0.38	0.50	0.40	0.533	уд.
8	380	14	-23	116	230	292	126	1.00	0.67	0.73	0.44	0.67	0.79	1.00	0.733	хор.
9	580	29	-22	100	215	304	48	1.00	0.30	0.72	0.70	0.55	0.81	0.52	0.626	уд.

$$D_1 = \sqrt[7]{d_1 \cdot d_2 \cdot d_3 \cdot d_4 \cdot d_5 \cdot d_6 \cdot d_7}, \quad D_2 = \sqrt[9]{d_1 \cdot d_2 \cdot d_3 \cdot d_4 \cdot d_5 \cdot d_6 \cdot d_7 \cdot d_8 \cdot d_9}$$

статистическая чувствительность и эффективность. Оказалось, что для шкал желательности статистическая чувствительность и эффективность частных и обобщенной функций желательности не ниже, чем таковые для любого технологического показателя, им соответствующего [2].

Обобщенная функция желательности является количественным, однозначным, единым и универсальным показателем качества исследуемого объекта, и если добавить еще такие свойства, как адекватность, эффективность и статистическая чувствительность, то становится ясным, что ее можно использовать в качестве критерия оптимизации.

Обобщенная функция желательности нашла широкое применение для оценки качества полимерных материалов, резиновых и латексных изделий, а также при разработке различных рецептур [8—11]. Используется она в последнее время также и в промышленности [12].

3.5. Резюме

Построение обобщенного параметра оптимизации связано с созданием единого признака, количественно определяющего функционирование исследуемого объекта с многими выходными параметрами. При этом возникают некоторые трудности. Каждый выходной параметр—отклик—имеет свой физический смысл, свою размерность. Чтобы объединить различные отклики, необходимо ввести единую для всех откликов искусственную метрику. Набор данных каждого отклика нужно поставить в соответствие с некоторым стандартным аналогом, с безразмерной шкалой. Поэтому первым вопросом, который нужно решить при построении обобщенного параметра оптимизации, является вопрос о выборе шкалы. Шкала должна быть однотипной для всех объединяемых откликов. Построение шкалы во многом зависит от уровня априорных сведений о выходных параметрах, а также от той точности, с которой мы хотим определить обобщенный отклик.

Второй важный вопрос — выбор правила комбинирования исходных частных откликов в обобщенный показатель. Единого правила не существует, здесь можно идти различными путями, и выбор пути неформализован. Рассмотрено несколько различных способов построения обобщенного показателя.

Литература

1. В. Г. Горский, Ю. П. Адлер. Планирование промышленных экспериментов. М., «Металлургия», 1974.
2. Т. М. Карташова. Вопросы оптимизации при разработке рецептуры и технологии получения новых полимерных материалов. Автореферат канд. дисс., МХТИ им. Д. И. Менделеева, 1969.

3. Г. Г. Азгальдов, Э. П. Райхман. О квалиметрии. М., «Стандарты», 1973.
4. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
5. E. C. Harrington. Industr. Quality Control, 1965, 21, N 10.
6. E. C. Harrington. Chem. Engng. Progr. 1963, 42, N 59.
7. Руководящие технические материалы. Экспериментально-статистические методы получения математического описания и оптимизации сложных технологических процессов. (Ранговая корреляция), вып. 3. М., ОКБА, 1966.
8. Е. В. Маркова, Т. М. Карташова, Ю. М. Бусыгина и др. — Заводская лаборатория, 1969, 35, № 7.
9. Б. П. Штаркман, Т. М. Карташова, Э. А. Середа и др. Оптимизация рецептуры и режима желатинизации пластизолей. — Пластические массы, 1969, № 2.
10. Т. М. Карташова, Б. П. Штаркман, А. М. Шаргородский и др. Применение совмещенных планов для исследования и оптимизации процесса переработки смесей полимеров. — Пластические массы, 1969, № 9.
11. Е. В. Маркова, Т. М. Карташова, Ю. М. Бусыгина и др. Применение латинского куба второго порядка при разработке рецептуры нового полимерного материала. — Заводская лаборатория, 1969, 35, № 7.
12. C. W. Lowe. Industrial Statistics, v. 2. London, Busines Books Ltd., 1970.

Глава четвертая

ФАКТОРЫ

Что посеешь, то и пожнешь.

Поговорка

Теперь нам предстоит рассмотреть способы воздействия на оптимизируемый объект. Как вы знаете (стр. 16), способы воздействия были названы факторами.

После того как выбран объект исследования и параметр оптимизации, нужно включить в рассмотрение все существенные факторы, которые могут влиять на процесс. Если какой-либо существенный фактор окажется неучтенным, то это может привести к неприятным последствиям. Так, если неучтенный фактор произвольно флуктуировал — принимал случайные значения, которые экспериментатор не контролировал, — это значительно увеличит ошибку опыта (вы подробно познакомитесь с понятием «ошибка опыта» в гл. 8). При поддержании фактора на некотором фиксированном уровне может быть получено ложное представление об оптимуме, так как нет гарантии, что фиксированный уровень является оптимальным.

Читатель, внимательно прочитавший гл. 1 и усвоивший, что число различных состояний объекта p^k , где p — число уровней, а k — число факторов, может задать вопрос: «Ну, а как же преодолеть большое число опытов? Чем больше факторов, тем больше опытов». Действительно, число опытов растет по показательной функции. Размерность факторного пространства увеличивается, и математики в таких случаях говорят о «проклятии размерности».

Рекомендации о том, как преодолеть «проклятие размерности», вы найдете в гл. 7.

Если число факторов больше пятнадцати, нужно обратиться к методам отсеивания несущественных факторов. Здесь можно воспользоваться формализацией априорной информации [1—3], методом случайного баланса [4], планами Плаккетта-Бермана [5] и др. Иногда эти планы применяются и при меньшем числе факторов.

Мы не имеем возможности в этой книге рассказать об отсеивающих экспериментах и о формализации априорной информации. В «Ограничениях» сказано, что рассматривается случай, когда множество факторов задано и число факторов не превышает пятнадцати.

Однако обратить ваше внимание на важность выбора факторов, влияющих на процесс, на опасность пропуска существенного фактора мы сочли совершенно необходимым. От удачного выбора факторов зависит успех оптимизации.

Теперь поговорим о факторах. Начнем с определения.

4.1. Определение фактора

Фактором называется измеряемая переменная величина, принимающая в некоторый момент времени определенное значение. Факторы соответствуют способам воздействия на объект исследования.

Так же, как и параметр оптимизации, каждый фактор имеет область определения. Мы будем считать фактор заданным, если вместе с его названием указана область его определения. Под областью определения понимается совокупность всех значений, которые в принципе может принимать данный фактор. Ясно, что совокупность значений фактора, которая используется в эксперименте, является подмножеством из множества значений, образующих область определения.

Область определения может быть непрерывной и дискретной. Однако в тех задачах планирования эксперимента, которые мы собираемся рассматривать, всегда используются дискретные области определения. Так, для факторов с непрерывной областью определения, таких, как температура, время, количество вещества и т. п., всегда выбираются дискретные множества уровней. В практических задачах области определения факторов, как правило, ограничены. Ограничения могут носить принципиальный либо технический характер. (Подробно это рассмотрено в главе 6).

Произведем классификацию факторов в зависимости от того, является ли фактор переменной величиной, которую можно оценивать количественно: измерять, взвешивать, титровать и т. п., или же он — некоторая переменная, характеризующаяся качественными свойствами.

Вы уже догадались, что факторы разделяются на количественные и качественные. Качественные факторы — это разные вещества, разные технологические способы, аппараты, исполнители и т. д.

Хотя качественным факторам не соответствует числовая шкала в том смысле, как это понимается для количественных факторов, однако можно построить условную порядковую шкалу, которая ставит в соответствие уровням качественного фактора числа натурального ряда, т. е. производит кодирование. Порядок уровней может быть произволен, но после кодирования он фиксируется.

В ряде случаев граница между понятием качественного и количественного фактора весьма условна. Пусть, например,

при изучении воспроизводимости результатов химического анализа надо установить влияние положения тигля с навеской в муфельной печи. Можно разделить под печи на квадраты и считать номера квадратов уровнями качественного фактора, определяющего положение тигля. Вместо этого можно ввести два количественных фактора — ширину и длину пода печи. Качественным факторам не соответствует числовая шкала, и порядок уровней факторов не играет роли.

Время реакции, температура, концентрация реагирующих веществ, скорость подачи веществ, величина pH — это примеры наиболее часто встречающихся количественных факторов. Различные реагенты, адсорбенты, вулканизующие агенты, кислоты, металлы являются примером уровней качественных факторов.

Пример 1. Наш первый пример относится к исследованию процесса вулканизации бутадиен-стирольного каучука солями непредельных кислот. В планирование эксперимента были включены следующие факторы: \bar{x}_1 — температура вулканизации, °C; \bar{x}_2 — время вулканизации, мин; \bar{x}_3 — количество инициатора, вес. ч.; \bar{x}_4 — количество вулканизующего агента, вес. ч.; \bar{x}_5 — количество окисла, вес. ч.; \bar{x}_6 — тип окисла (окись цинка или окись магния); \bar{x}_7 — тип кислотного остатка (метакрилат, малеат); \bar{x}_8 — тип катиона соли (Na, Mg).

Довольно большое количество факторов и наличие среди них качественных объясняется тем, что планировать эксперимент приходилось на первой стадии, когда еще неясно, какие вещества нужно использовать для структурирования эластомеров, какие соли добавить для увеличения прочности, какие ускорители окажутся наиболее эффективными и т. д. [6, 7].

Пример 2. Многостадийный процесс получения ацетилацетона характерен тем, что приходилось изучать факторы, влияющие на четыре стадии процесса, так как y определялся только в конце 4-й стадии (рис. 9).

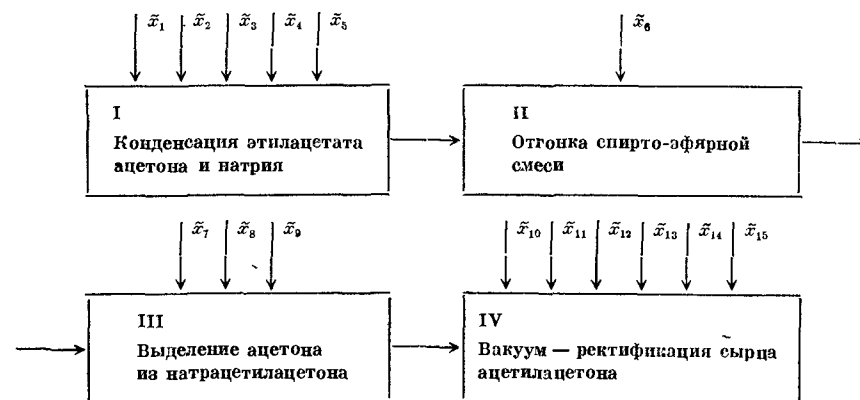


Рис. 9. Процесс получения ацетилацетона

I. Стадия конденсации. \tilde{x}_1 — температура реакции конденсации, °C; \tilde{x}_2 — время прилива ацетона, мин; \tilde{x}_3 — время выдержки, час; \tilde{x}_4 — соотношение компонентов, г/г; \tilde{x}_5 — скорость перемешивания, об/сек.

II. Стадия отгонки спирто-эфирной смеси. \tilde{x}_6 — конечная температура сухого остатка, °C.

III. Стадия выделения ацетилацетона из натрацетилацетона. \tilde{x}_7 — величина pH; \tilde{x}_8 — скорость подачи соляной кислоты, мл/сек; \tilde{x}_9 — температура при выделении, °C.

IV. Стадия вакуум-ректификации сырья ацетилацетона. \tilde{x}_{10} — температура отгонки спирто-эфирной смеси 1-й фракции, °C; \tilde{x}_{11} — температура отгонки спирто-эфирной смеси 2-й фракции, °C; \tilde{x}_{12} — температура отгонки спирто-эфирной смеси 3-й фракции, °C; \tilde{x}_{13} — время отгонки 1-й фракции, мин; \tilde{x}_{14} — время отгонки 2-й фракции, мин; \tilde{x}_{15} — время отгонки 3-й фракции, мин.

Как вы думаете, можно ли включать в планирование эксперимента факторы, относящиеся к различным стадиям?

Давайте рассуждать вместе.

В рассматриваемом примере процесс получения ацетилацетона, состоящий из четырех стадий, удобно представить как единое целое в виде одного «черного ящика» (рис. 10). На этот «черный ящик» воздействуют 15 факторов. Почему возникла необходимость рассматривать все стадии как единое целое?

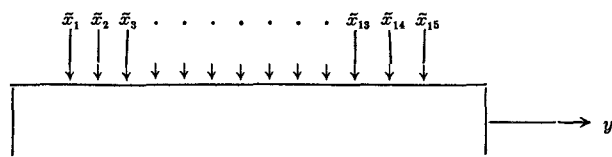


Рис. 10. Процесс получения ацетилацетона

Параметр оптимизации измеряется в конце последней стадии. На предыдущих стадиях выходной параметр не измеряется: отсутствуют нужные аналитические методики. Возможны и другие причины, например, параметр оптимизации отдельной стадии противоречит общей цели оптимизации. Но было бы неправильным считать, что во всех случаях при оптимизации многостадийных процессов нужно рассматривать все стадии как единое целое. Весьма часто оптимизация отдельных стадий вполне оправдана и очевидна.

Так, процесс получения сульфадимизина состоит из трех химических стадий: получения сульгина, получения ацетилацетона (эта стадия, как вы знаете из рассматриваемого примера, в свою очередь состоит из четырех частей) и конденсации сульгина с ацетилацетоном. Сульгин и ацетилацетон имеют самостоятельное значение. Процессы их получения выделены в отдель-

ные производства, зачастую территориально не объединенные. Оптимизировать совместно все три стадии не представляется возможным и целесообразным.

Таким образом, в планирование эксперимента можно включать факторы, относящиеся к различным стадиям, но не во всех случаях это является необходимым.

4.2. Требования, предъявляемые к факторам при планировании эксперимента

Мы дали определение понятию «фактор» и привели примеры факторов. Теперь сформулируем требования, предъявляемые к факторам.

При планировании эксперимента факторы должны быть управляемыми. Это значит, что экспериментатор, выбрав нужное значение фактора, может его поддерживать постоянным в течение всего опыта, т. е. может управлять фактором. В этом состоит особенность «активного» эксперимента. Планировать эксперимент можно только в том случае, если уровни факторов подчиняются воле экспериментатора.

Представьте себе, что вы изучаете процесс синтеза аммиака. Колонна синтеза установлена на открытой площадке. Является ли температура воздуха фактором, который можно включить в планирование эксперимента?

Температура воздуха — фактор неуправляемый. Мы еще не научились делать погоду по заказу. А в планировании могут участвовать только те факторы, которыми можно управлять, — устанавливать и поддерживать на выбранном уровне в течение опыта или менять по заданной программе. Температурой окружающей среды в данном случае управлять невозможно. Ее можно только контролировать.

Чтобы точно определить фактор, нужно указать последовательность действий (операций), с помощью которых устанавливаются его конкретные значения (уровни). Такое определение фактора будем называть *операциональным*. Так, если фактором является давление в некотором аппарате, то совершенно необходимо указать, в какой точке и с помощью какого прибора оно измеряется и как оно устанавливается. Введение операционального определения обеспечивает однозначное понимание фактора.

С операциональным определением связаны выбор размерности фактора и точность его фиксирования. Мы привыкли считать, что выбор размерности фактора не представляет особой трудности. Экспериментатор хорошо ориентируется в том, какую размерность нужно использовать. Это действительно так в тех случаях, когда существует устоявшаяся традиция, построены измерительные шкалы, приборы, созданы эталоны и т. д. Так обстоит

дело при измерении температуры, времени, давления и т. д. Но бывает, что выбор размерности превращается в весьма трудную проблему выбора измерительных шкал, сложность которой далеко выходит за рамки нашего рассмотрения.

Замена одной измерительной шкалы другой называется преобразованием шкал. Оно может быть использовано для упрощения модели объекта.

Точность замера факторов должна быть возможно более высокой. Степень точности определяется диапазоном изменения факторов. При изучении процесса, который длится десятки часов, нет необходимости учитывать доли минуты, а в быстрых процессах необходимо учитывать, быть может, доли секунды. Если факторы измеряются с большой ошибкой или особенность объекта исследования такова, что значения факторов трудно поддерживать на выбранном уровне (уровень фактора «плывет»), то экспериментатору следует обратиться к конъюнктному анализу [8, 9].

Факторы должны быть непосредственными воздействиями на объект. Факторы должны быть однозначны. Трудно управлять фактором, который является функцией других факторов. Но в планировании могут участвовать сложные факторы, такие, как соотношения между компонентами, их логарифмы и т. п.

Необходимость введения сложных факторов возникает при желании представить динамические особенности объекта в статической форме. Пусть, например, требуется найти оптимальный режим подъема температуры в реакторе. Если относительно температуры известно, что она должна нарастать линейно, то в качестве фактора вместо функции (в данном случае линейной) можно использовать тангенс угла наклона, т. е. градиент. Положение усложняется, когда исходная температура не зафиксирована. Тогда ее приходится вводить в качестве еще одного фактора. Для более сложных кривых пришлось бы ввести большее число факторов (производные высоких порядков, координаты особых точек и т. д.). Поэтому целесообразно пользоваться сложным качественным фактором — номером кривой. Различные варианты кривых рассматриваются в качестве уровней. Это могут быть разные режимы термообработки сплавов, переходные процессы в системах управления и т. д. Мы показали, как можно сложный фактор-функцию представить с помощью простых однозначных факторов.

Пример 3. При оптимизации процесса получения одного производного пиперазина изучалось влияние семи факторов, среди которых были соотношения между компонентами: \tilde{x}_1 — количество едкого натра, г/мол; \tilde{x}_2 — способ поддержания pH; \tilde{x}_3 — время прилива вещества a , час; \tilde{x}_4 — время выдержки реакционной массы, час; \tilde{x}_5 — температура реакционной среды, °C; \tilde{x}_6 — весовое соотношение вещества b и метанола, г/г; \tilde{x}_7 — мольное соотношение вещества a и вещества b , г/мол/г/мол.

Матрица планирования для этих факторов приведена в гл. 7.

А теперь ответьте, пожалуйста, на следующий вопрос. Изучается процесс растворения твердого тела в жидкости — диффузионный процесс. Может ли скорость диффузии служить фактором в планировании эксперимента?

Скорость диффузии зависит от концентрации, величины поверхности соприкосновения двух фаз — жидкости и твердого тела, от коэффициента растворения. Коэффициент растворения зависит от коэффициента диффузии и от толщины диффузионного слоя. Коэффициент диффузии, в свою очередь, является функцией нескольких переменных.

Конкретное значение скорости диффузии определяется сочетанием значений других факторов. Если бы мы могли управлять скоростью диффузии, придавая ей в каждом опыте желаемое значение, то она могла бы стать фактором. Те, кто сталкивался с диффузионными задачами, знают, как далеки мы от реализации этой возможности. Скорость диффузии не может являться фактором при планировании эксперимента.

При растворении твердого тела в жидкости образуется диффузионный слой, прилегающий к поверхности твердого тела. Состав этого слоя неодинаков в различных зонах. В пограничной части слой в большей или меньшей степени находится в состоянии равновесия, и концентрация растворенного вещества в нем приближается к концентрации насыщенного раствора, $c_{\text{нас}}$. В части слоя, прилегающей к внутреннему объему жидкости, концентрация растворенного вещества приближается к концентрации c в остальном объеме жидкости. Скорость диффузии тем больше, чем больше различие в концентрациях ($c_{\text{нас}} - c$) диффундирующего вещества

$$\frac{dc}{dt} \doteq KS (c_{\text{нас}} - c),$$

где dc/dt — скорость изменения концентрации в объеме рассматриваемой фазы, S — величина поверхности соприкосновения данных фаз, K — коэффициент растворения.

Коэффициент растворения зависит от коэффициента диффузии D растворяемого вещества и от толщины диффузионного слоя δ $K = D/\delta$. Коэффициент диффузии, в свою очередь, зависит от ряда факторов.

Теперь вам должно быть ясно, что скорость диффузии является функцией многих переменных. Скоростью диффузии весьма трудно управлять.

Такие требования, предъявляемые к факторам при планировании эксперимента, как управляемость и однозначность, не выполняются.

4.3. Требования к совокупности факторов

При согласии малое растет, при несогласии — величайшее разрушается.

Гай Саллюстий Крисп

При планировании эксперимента обычно одновременно изменяется несколько факторов. Поэтому очень важно сформулировать требования, которые предъявляются к совокупности факторов. Прежде всего выдвигается требование *совместимости*. Совместимость факторов означает, что все их комбинации осуществимы и безопасны. Это очень важное требование. Представьте себе, что вы поступили легкомысленно, не обратили внимания на требование совместимости факторов и запланировали такие условия опыта, которые могут привести к взрыву установки или осмолецию продукта. Согласитесь, что такой результат очень далек от целей оптимизации.

Несовместимость факторов может наблюдаться на границах областей их определения. Избавиться от нее можно сокращением областей. Положение усложняется, если несовместимость проявляется внутри областей определения. Одно из возможных решений — разбиение на подобласти и решение двух отдельных задач.

При планировании эксперимента важна *независимость* факторов, т. е. возможность установления фактора на любом уровне вне зависимости от уровней других факторов. Если это условие невыполнимо, то невозможно планировать эксперимент. Итак, мы подошли ко второму требованию — отсутствию корреляции между факторами. Требование некоррелированности не означает, что между значениями факторов нет никакой связи. Достаточно, чтобы связь не была линейной.

Исследуется некоторая термодинамическая система. Можно ли включить в планирование эксперимента следующие три фактора: \tilde{x}_1 — давление, атм, \tilde{x}_2 — объем, л, \tilde{x}_3 — температуру, °К?

Пусть в термодинамической системе имеет место уравнение Менделеева—Клапейрона $PV=nRT$ и заданы два фактора, например $V(x_1)$ и $T(x_2)$. Тогда $P(x_3)$ может быть вычислено. То же самое и с двумя другими парами факторов. Поэтому в планирование можно включать два (а не три) фактора. Здесь возможны три комбинации: 1) x_1 и x_2 , 2) x_1 и x_3 , 3) x_2 и x_3 .

4.4. Примеры факторов

В этой главе мы уже приводили примеры факторов, относящиеся к органической химии и полимерам. А теперь перейдем к другим областям, памятуя изречение Ньютона, что примеры так же поучительны, как и правила.

Области практических приложений планирования эксперимента чрезвычайно многообразны: химия, металлургия, биология, ме-

дицина, обогащение полезных ископаемых, пищевая и текстильная промышленность, сельское хозяйство, военное дело и др. [10].

Применяется планирование эксперимента и в несколько неожиданных областях исследования, в таких, как геронтология (наука о долголетьи), при классификации образцов древней керамики, в хлебопечении и табачном деле.

В зависимости от объектов исследования меняются и факторы. В своих примерах, к сожалению, мы не можем отразить все это многообразие. Остановимся на более типичных для нашей отечественной практики случаях, базируясь на материалах второй Всесоюзной конференции по планированию эксперимента (Москва, 1968).

В выборе примеров мы также руководствовались принципом многофакторности; приводили задачи, в которых количество факторов было бы не меньше четырех, так как придумать примеры с двумя-тремя факторами очень легко может и сам читатель.

Пример 4. При исследовании электролитического процесса получения алюминия в планирование эксперимента были включены следующие семь факторов: \tilde{x}_1 — напряжение на электролизере, в; \tilde{x}_2 — время между обработками электролизера, час; \tilde{x}_3 — концентрация фтористого магния в электролите, %; \tilde{x}_4 — концентрация фтористого кальция в электролите, %; \tilde{x}_5 — криолитовое отношение; \tilde{x}_6 — уровень электролита в ванне, см; \tilde{x}_7 — время между операциями съема угольной пены, сутки [11].

Пример 5. А вот пример факторов, влияние которых интересовало экспериментатора при оптимизации производства резисторов: \tilde{x}_1 — давление при прессовке, кг/см²; \tilde{x}_2 — температура при прессовке, °С; \tilde{x}_3 — время выдержки под давлением, мин; \tilde{x}_4 — температура в муфеле при прессовке, °С; \tilde{x}_5 — время температурной выдержки, мин; \tilde{x}_6 — дисперсность наполнителя, мк; \tilde{x}_7 — соотношение флюса и наполнителя, г/г; \tilde{x}_8 — давление при шмотировании, кг/см²; \tilde{x}_9 — дисперсность сажи, мк; \tilde{x}_{10} — время выдержки при шмотировании, мин; \tilde{x}_{11} — качество керамики оснований; \tilde{x}_{12} — дисперсность флюсов, мк [12].

Пример 6. При изучении процесса варки сульфатной целлюлозы в планирование эксперимента были включены такие пять факторов: \tilde{x}_1 — концентрация активной щелочи в варочном растворе (в единицах Na₂O), г/л; \tilde{x}_2 — сульфитность раствора, %; \tilde{x}_3 — конечная температура варки, °С; \tilde{x}_4 — продолжительность подъема температуры до конечной, мин; \tilde{x}_5 — продолжительность варки при конечной температуре, мин [13].

Пример 7. При оптимизации процесса обогащения молибденовой руды экспериментатор остановил свое внимание на следующих факторах: \tilde{x}_1 — время измельчения руды, мин; \tilde{x}_2 — расход олеата натрия, г/т; \tilde{x}_3 — расход алкилсульфата, г/т; \tilde{x}_4 — расход соды, г/т; \tilde{x}_5 — расход керосина, г/т [14].

Пример 8. При оптимизации процесса экстракции циркония и гафния из серьезнокислых растворов в качестве независимых переменных приняты: \tilde{x}_1 — концентрация металла, г/л; \tilde{x}_2 — концентрация кислоты, мол/л; \tilde{x}_3 — концентрация спирта, %; \tilde{x}_4 — соотношение объемов фаз, мл/мл [15].

Пример 9. В микробиологических исследованиях весьма важной задачей является нахождение оптимального состава питательной среды. В одной из микробиологических работ проверялось влияние следующих факторов:

\tilde{x}_1 — $MgCl_2$;	\tilde{x}_5 — H_3BO_3 ;	\tilde{x}_8 — $ZnSO_4$;
\tilde{x}_2 — $ZnCl_2$;	\tilde{x}_6 — $CaCl_2$;	\tilde{x}_9 — $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$;
\tilde{x}_3 — $FeCl_3$;	\tilde{x}_7 — $MnSO_4$;	\tilde{x}_{10} — $CoCl_2$ [16].
\tilde{x}_4 — $CuSO_4$;		

Обработка результатов эксперимента велась по пробам, полученным на четвертые сутки выращивания. В экспериментах такого рода возможно также варьировать время выращивания, температуру и т. д.

Пример 10. И, наконец, для специалистов, занимающихся животноводством, приведем пример факторов, влияющих на откорм свиней.

Определялось соотношение в рационе питательных веществ и стимуляторов, т. е. химических соединений, воздействующих на обмен веществ в организме. Изучалось влияние следующих четырнадцати факторов:

макроэлементы: \tilde{x}_1 — Ca; \tilde{x}_2 — K; \tilde{x}_3 — Mg; \tilde{x}_4 — Na;
 микроэлементы: \tilde{x}_5 — Co; \tilde{x}_6 — Zn; \tilde{x}_7 — Cu; \tilde{x}_8 — Fe; \tilde{x}_9 — J; \tilde{x}_{10} — Mn;
 витамины: \tilde{x}_{11} — A; \tilde{x}_{12} — D₂; \tilde{x}_{13} — B₁₂;
 антибиотики: \tilde{x}_{14} — биомидин.

Мы надеемся, что приведенных примеров достаточно, чтобы вы составили себе ясное представление о понятии «факторы».

4.5. Резюме

Итак, мы установили, что факторы — это переменные величины, соответствующие способам воздействия внешней среды на объект. Они определяют как сам объект, так и его состояние. Требования к факторам: управляемость и однозначность.

Управлять фактором — это значит установить нужное значение и поддерживать его постоянным в течение опыта или менять по заданной программе. В этом состоит особенность «активного» эксперимента. Планировать эксперимент можно только в том случае, если уровни факторов подчиняются воле эксперимента-тора.

Факторы должны непосредственно воздействовать на объект исследования. Трудно управлять фактором, если он является функцией других переменных, но в планировании эксперимента могут участвовать сложные факторы, такие, как логарифмы, соотношения и т. д. Факторы должны быть определены операционально.

Требования к совокупности факторов: совместимость и отсутствие линейной корреляции. Выбранное множество факторов должно быть достаточно полным. Если какой-либо существенный фактор пропущен, это приведет к неправильному определению оптимальных условий или к большой ошибке опыта. Факторы могут быть количественными и качественными.

Точность фиксации факторов должна быть высока. Степень точности определяется диапазоном изменения факторов.

Выбор факторов — очень ответственный этап при подготовке к планированию эксперимента. От удачного выбора зависит успех оптимизации.

После того как мы рассказали вам о параметре оптимизации и факторах, можно подойти к выбору модели исследуемого процесса.

Литература

1. Ю. П. Адлер, И. Ф. Александрова, Ю. В. Грановский и др. Об одном методе формализации априорной информации при планировании эксперимента. В сб. «Планирование эксперимента». М., «Наука», 1966.
2. Руководящие технические материалы. Экспериментально-статистические методы получения матем. описания и оптимизации сложных технологических процессов (Ранговая корреляция), вып. 3. М., ОКБА, 1966.
3. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.
4. F. E. Satterthwaite. Random balance experimentation. — *Technometrics*, 1959, 1, N 2.
5. R. L. Plackett, I. P. Burman. The design of optimum multifactor experiments. — *Biometrika*, 1946, 33, N 4.
6. А. А. Донцов, Е. В. Маркова, В. Э. Михлин и др. Применение математико-статистического метода для оптимизации процесса вулканизации эластомеров солями ненасыщенных кислот. — *Каучук и резина*, 1967, № 10.
7. Е. В. Маркова. Планирование эксперимента при оптимизации процессов тонкого органического синтеза. Автореферат канд. дисс., МХТИ им. Менделеева. М., 1965.
8. Н. П. Клепиков, С. Н. Соколов. Анализ и планирование экспериментов методом максимума правдоподобия. М., «Наука», 1964.
9. В. В. Федоров. Теория оптимального эксперимента. М., «Наука», 1971.
10. Ю. П. Адлер, Ю. В. Грановский. Обзор прикладных работ по планированию эксперимента. Препринт № 1. МГУ, 1967 (Изд. 2-е, препринт № 33, 1972).
11. Э. Б. Чамлик, В. М. Никитин, Э. М. Менчер. Выделение существенных факторов при электролитическом получении алюминия. В сб. «Проблемы планирования эксперимента». М., «Наука», 1969.
12. Л. Г. Власов, В. Б. Лукьянов, Б. Г. Красильников и др. Применение методов планирования экстремального эксперимента в производстве резисторов. Материалы II Всесоюзной конференции по планированию эксперимента. М., МЭИ, 1968.
13. Э. М. Менчер, Р. Э. Пен, М. Г. Малина и др. Опыт изучения варки сульфатной целлюлозы с применением статистических методов планирования эксперимента. — Материалы второй Всесоюзной конференции по планированию эксперимента. М., МЭИ, 1968.
14. Л. А. Барский, Ю. Б. Рубинштейн. Особенности планирования экстремальных экспериментов при исследовании разделительных процессов. В сб. «Проблемы планирования эксперимента». М., «Наука», 1969.
15. Н. С. Смирнова, Ю. В. Грановский, А. Л. Каплан и др. Планирование эксперимента при изучении экстракции циркония и гафния спиртами. В сб. «Проблемы планирования эксперимента». М., «Наука», 1969.
16. И. М. Чирков, Л. Е. Гурина, С. С. Рылкин. Применение методов математического планирования экспериментов в микробиологических исследованиях. В сб. «Проблемы планир. эксперим.». М., «Наука», 1969.

ВЫБОР МОДЕЛИ

Достаточно мне шум шагов их услышать,
Чтобы мог показать я, в каком направлении путь
они держат.

Г. Аполлинер. Кортес

В этой главе мы хотим дать рекомендации по выбору модели. Дело это не простое и связано со многими обстоятельствами и соображениями.

Мы говорили, что под моделью понимаем вид функции отклика (см. гл. 1)

$$y = f(x_1, x_2, \dots, x_k).$$

Выбрать модель — значит выбрать вид этой функции, записать ее уравнение. Тогда останется спланировать и провести эксперимент для оценки численных значений констант (коэффициентов) этого уравнения. Но как выбрать модель?

Чтобы постепенно продвигаться к ответу на этот вопрос, давайте сначала построим геометрический аналог функции отклика — поверхность отклика. Будем для наглядности рассматривать случай с двумя факторами.

Заметим, что в случае многих факторов геометрическая наглядность теряется. Мы попадаем в абстрактное многомерное пространство, где у нас нет навыка ориентирования. Приходится переходить на язык алгебры. Тем не менее простые примеры, которые мы сейчас рассмотрим, помогут вам, как мы думаем, при работе с многими факторами.

Мы хотим изобразить геометрически возможные состояния «черного ящика» с двумя входами. Для этого достаточно расположить плоскостью с обычной декартовой системой координат. По одной оси координат будем откладывать в некотором масштабе значения (уровни) одного фактора, а по другой оси — второго. Тогда каждому состоянию «ящика» будет соответствовать точка на плоскости.

Но, как вы помните из предыдущей главы, для факторов существуют области определения. Это значит, что у каждого фактора есть минимальное и максимальное возможные значения, между которыми он может изменяться либо непрерывно, либо дискретно. Если факторы совместимы, то границы образуют на плоскости некоторый прямоугольник, внутри которого лежат точки, соответствующие состояниям «черного ящика». Пунктирными ли-

ниями на рис. 11 обозначены границы областей определения каждого из факторов, а сплошными — границы их совместной области определения.

Чтобы указать значение параметра оптимизации, требуется еще одна ось координат. Если ее построить, то поверхность отклика будет выглядеть так, как на рис. 12. Пространство, в котором строится поверхность отклика, мы будем называть факторным пространством. Оно задается координатными осями, по которым откладываются значения факторов и параметра оптимизации*. Размерность факторного пространства зависит от числа факторов. При многих факторах поверхность отклика уже нельзя изобразить наглядно и приходится ограничиваться только алгебраическим языком.

Но для двух факторов можно даже не переходить к трехмерному пространству, а ограничиться плоскостью. Для этого достаточно произвести сечение поверхности отклика плоскостями, параллельными плоскости X_1OX_2 , и полученные в сечениях линии спроектировать на эту плоскость. Так строят, например, изображения гор и морских впадин на географических картах (рис. 13). Точка M на рисунке — это и есть та оптимальная точка, которую мы ищем. Каждая линия соответствует постоянному значению параметра оптимизации. Такая линия называется линией равного отклика. Существует соответствие между состоянием «ящика» и значением параметра оптимизации: каждому возможному состоянию «ящика» соответствует одно значение параметра оптимизации. Однако обратное неверно: одному возможному значению параметра оптимизации может соответствовать и одно, и несколько, и сколько угодно состояний «ящиков».

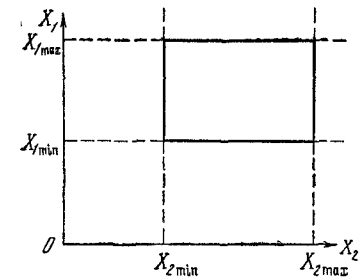


Рис. 11. Область определения факторов

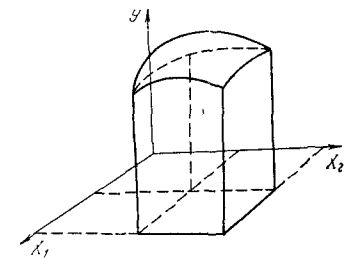


Рис. 12. Поверхность отклика

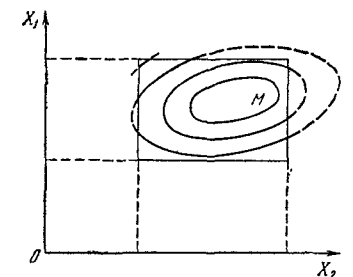


Рис. 13. Проекция сечений поверхности отклика на плоскость

* Иногда под факторным пространством понимается пространство, образованное только осями факторов.

Правда, эти утверждения справедливы, если не учитывать ошибок в определении значений параметра оптимизации. К вопросу об оценке и учете этих ошибок мы вернемся ниже, а пока не будем принимать их во внимание.

Теперь, когда мы можем представить себе поверхность отклика, пора вернуться к основному вопросу: как ставить эксперимент, чтобы найти оптимум при минимуме затрат? Это прежде всего вопрос стратегии.

Если бы мы располагали таблицей, в которой содержались бы все возможные состояния объекта и соответствующие им отклики, то особой необходимости в построении математической модели не было бы. Просто мы бы выбрали то (или те) состояние, которое соответствует наилучшему отклику. Но мы уже знаем, сколь велик перебор возможных состояний (см. гл. 1), и должны отказаться от практической реализации этой возможности.

Другая возможность — случайный выбор некоторого числа состояний и определение откликов в них, в надежде, что среди этих состояний попадутся оптимальное или по крайней мере близкие к нему состояния. Мы не будем рассматривать эту интересную возможность, так как, к сожалению, она не вписывается в нашу тему [1].

Наконец, третья возможность — строить математическую модель, чтобы с ее помощью предсказывать значения откликов в тех состояниях, которые не изучались экспериментально. Если не можем измерить отклик в каждом состоянии, то сумеем хоть предсказывать результат. Причем даже не в каждом состоянии, а только в наиболее интересных, в тех, которые приближают нас к оптимуму.

Такая стратегия приводит нас к шаговому принципу, лежащему в основе рассматриваемого метода планирования эксперимента [2, 3].

5.1. Шаговый принцип

За отказ от полного перебора состояний надо чем-то платить. Цена — это предположения, которые мы должны сделать относительно свойств неизвестной нам модели до начала эксперимента (как говорят, априори).

Некоторые из предположений мы никогда не сможем проверить. Такие предположения называются постулатами. Если в действительности они не выполняются, то весьма возможно, что мы не найдем оптимум. Точнее, мы примем за оптимум то, что на самом деле им не является (хотя, быть может, нас и удовлетворит).

Какие же предположения о свойствах поверхности отклика мы делаем? Главное — это непрерывность поверхности, ее глад-

кость и наличие единственного оптимума (быть может, и на границе области определения факторов).

Эти постулаты позволяют представить изучаемую функцию в виде степенного ряда в окрестности любой возможной точки факторного пространства (такие функции в математике называются аналитическими). Кроме того, если мы придумаем какой-то способ постепенного приближения к оптимальной точке, нужно, чтобы результат не зависел от исходной точки. Если оптимум один, то неважно, приближаемся мы к нему справа или слева, а если их несколько, да они еще неравноценны. . .

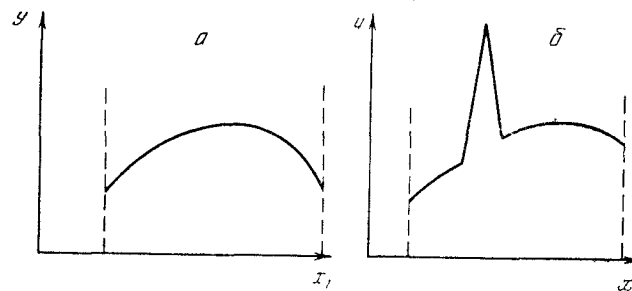


Рис. 14. Примеры функций отклика для одного фактора

На рис. 14 приведены две картинки, изображающие функции отклика для одного фактора. На рис. 14, а показан благоприятный случай. На рис. 14, б — много нарушений. Здесь и два экстремума (оптимума) и пик (нарушение гладкости и непрерывности).

Если в поисках оптимума мы начнем последовательно двигаться слева направо, то найдем наименьший из максимумов и вряд ли узнаем о существовании второго, наибольшего. Правда, он так локализован и остр, что его не мудрено пропустить и при движении с правого конца, если ставить опыты не во всех точках.

Возможно, вы обратили внимание на то, что требование непрерывности не согласуется с представлением о дискретных уровнях факторов. Однако в действительности это не страшно. Мы ведь можем считать, что фактор принимает непрерывный ряд значений (если даже некоторые значения не имеют смысла или физически нереализуемы). Важно только помнить о таком соглашении при использовании результатов. А для построения математической модели это создает значительные удобства.

Так как мы заранее считаем, что предпосылки выполняются, то надо максимально использовать возможности, которые при этом открываются.

Если, например, мы будем знать значения параметра оптимизации в нескольких соседних точках факторного пространства, мы сможем (в силу гладкости и непрерывности функции отклика)

представить себе результаты, которые можно ожидать в других соседних точках. Следовательно, можно найти такие точки, для которых ожидается наибольшее увеличение (или уменьшение, если мы ищем минимум) параметра оптимизации. Тогда ясно, что следующий эксперимент надо переносить именно в эти точки. Надо продвигаться в этом направлении, пренебрегая остальными. (Вот где экономятся опыты!) Сделав новый эксперимент, снова можно оценить направление, в котором скорее всего следует двигаться. В силу единственности оптимума мы, таким образом, рано или поздно непременно его достигнем. Это и есть шаговый принцип.

Сделаем некоторые пояснения. Мы выбираем в факторном пространстве какую-то точку и рассматриваем множество точек в ее окрестности, т. е. выбираем в области определения факторов малую подобласть. Здесь мы хотим провести эксперимент, на основании которого должна быть построена первая модель. Эту модель мы намерены использовать для предсказания результатов опытов в тех точках, которые не входили в эксперимент. Если эти точки лежат внутри нашей подобласти, то такое предсказание называется интерполяцией, а если вне — экстраполяцией. Чем дальше от области эксперимента лежит точка, для которой мы хотим предсказать результат, тем с меньшей уверенностью это можно делать. Поэтому мы вынуждены экстраполировать недалеко и использовать результаты экстраполяции для выбора усло-

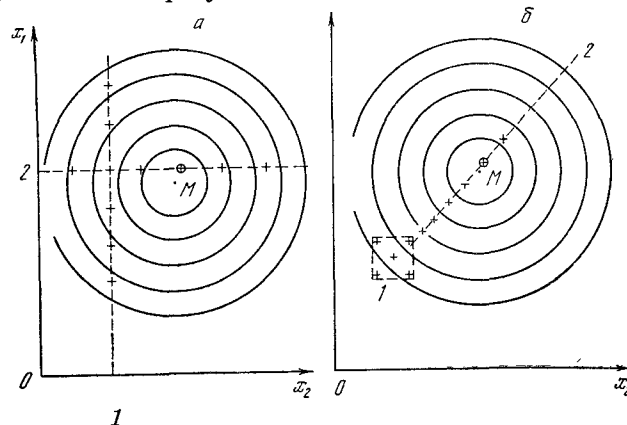


Рис. 15. Два способа поиска оптимума

вий проведения следующего эксперимента. Дальше цикл повторяется.

Попутно полученную модель можно использовать для проверки различных гипотез о механизме изучаемого явления или о его отдельных сторонах. Например, если вы предполагаете, что увеличение значения некоторого фактора должно приводить к увеличению значения параметра оптимизации, то с помощью модели можно узнать, так ли это. Такая проверка называется

интерпретацией модели. Она, конечно, имеет большое значение, и мы вернемся к ней позже (в гл. 11).

На рис. 15 изображены два варианта поиска оптимума для одной и той же поверхности. Крестиками на рисунке обозначены условия опытов. В случае а использован подход, который иногда называют классическим (метод Гаусса—Зейделя). Он состоит в том, что сначала последовательно изменяются значения одного фактора. (На рисунке этот эксперимент обозначен 1.) Затем находится и фиксируется наилучшее значение этого фактора. В этих условиях последовательно изменяются значения второго фактора (2) и т. д. (если больше факторов).

В случае б представлен простейший вариант шаговой процедуры. Сначала изучается локальная область (1), затем определяется наиболее интересное направление и в этом направлении ставятся следующие опыты (2).

Оказалось (см. рис. 15), что в обоих случаях достигнут одинаковый результат при одинаковом суммарном количестве опытов.

Как вы думаете, всегда ли эти две процедуры эквивалентны?

Что нам требуется? Выяснить, нет ли нарушений наших предпосылок. Легче всего установить, сколько оптимумов (экстремумов) имеет изображенная функция. Если экстремумов больше одного, то уже нарушена предпосылка. Кроме того, существенно, нет ли каких-нибудь нарушений гладкости и непрерывности функции (например, пиков).

Дело в том, что эффективность зависит от вида поверхности, а также от того, в какой последовательности перебираются факторы в случае а и из окрестностей какой точки начат эксперимент в случае б.

Попробуйте вместо окружностей, которые задают линии равных откликов, нарисовать эллипсы, главные оси которых составляют некоторый острый угол с осями координат. Вы увидите, что эффективность процедур окажется различной.

Вот иллюстрация, которая сразу показывает правильность вашего ответа (рис. 16). Это, разумеется, только иллюстрация. В жизни не всегда удастся за один цикл достигнуть оптимума.

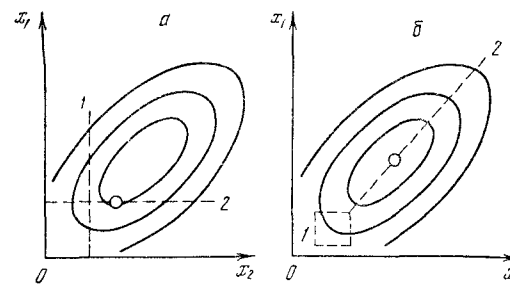


Рис. 16. Два способа поиска оптимума.

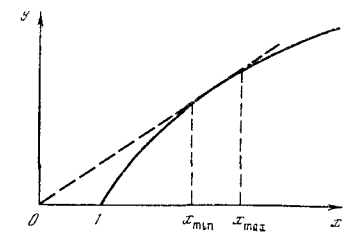


Рис. 17. График логарифмической функции

Но несомненно, что по крайней мере в отношении результата процедура б, т. е. шаговый метод, в среднем эффективнее, чем процедура а. Можно придумать и более конкурентноспособные процедуры, чем а, но они обычно требуют значительно больше опытов.

Теперь займемся выбором модели для первого эксперимента более конкретно.

5.2. Как выбрать модель?

Модели бывают разные. Моделей бывает много. Чтобы выбрать одну из них, надо понять, что мы хотим от модели, какие требования мы к ней предъявляем. Теперь мы, пожалуй, сможем сформулировать эти требования.

Исходя из выбранной стратегии, ясно, что главное требование к модели — это способность предсказывать направление дальнейших опытов, причем предсказывать с требуемой точностью. Так как до получения модели мы не знаем, какое направление нам понадобится, то естественно требовать, чтобы точность предсказания во всех возможных направлениях была одинакова.

Это значит, что в некоторой подобласти, в которую входят и координаты выполненных опытов, предсказанное с помощью модели значение отклика не должно отличаться от фактического больше чем на некоторую заранее заданную величину. Модель, которая удовлетворяет такому или какому-либо аналогичному требованию, называется адекватной. Проверка выполнимости этого требования называется проверкой адекватности модели. Разработаны специальные статистические методы, с помощью которых проверяется адекватность. Мы их рассмотрим в гл. 9.

Если несколько различных моделей отвечают нужным требованиям, то следует предпочесть ту из них, которая является самой простой.

На рис. 17 изображена логарифмическая функция. На некотором отрезке $[x_{\min}, x_{\max}]$ она с удовлетворительной точностью описывается двумя уравнениями:

$$y = \log_b x, \quad (1)$$

$$y = bx. \quad (2)$$

В уравнении (2) b — коэффициент, который мы можем оценить, например, по результатам эксперимента. Какое из уравнений, (1) или (2), по вашему мнению, проще?

Простота — вещь относительная. Если вы заранее не сформулируете точно, что называется простым, а что сложным, то невозможно произвести выбор. Вот почему на наш вопрос не было никакого другого ответа, кроме «не знаю».

На будущее мы договоримся, что при прочих равных условиях мы всегда будем предпочитать степенные ряды. Точнее, отрезки степенных рядов — алгебраические полиномы. При таком соглашении можно сказать, что уравнение (2) проще, чем уравнение (1).

Фактически мы произвели выбор класса моделей. Мы сказали, что всегда, когда это возможно, будем искать модель среди полиномов. Построение полинома возможно в окрестностях любой точки факторного пространства, поскольку мы предположили, что функция является аналитической.

Выбрать — значит сравнить. А как сравнить между собой классы моделей, если свойства объекта заранее неизвестны? Остается предполагать, что нам будут редко встречаться задачи, в которых исходные постулаты окажутся существенно неверными. Если это так, то мы действительно выбрали наиболее простой, удобный и математически разработанный класс моделей.

Возможно, что кто-то заранее выбрал для нашей задачи конкретную модель. Тогда тоже возникает необходимость в планировании эксперимента для оценки ее коэффициентов. Но мы не будем рассматривать задачи этого типа.

Давайте выпишем полиномы для случая двух факторов. Они будут различаться по максимальным степеням входящих в них переменных.

Полином нулевой степени: $y = b_0$.

Полином первой степени: $y = b_0 + b_1x_1 + b_2x_2$.

Полином второй степени: $y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$.

Полином третьей степени:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{112}x_1^2x_2 + b_{122}x_1x_2^2 + b_{111}x_1^3 + b_{222}x_2^3.$$

5.3. Полиномиальные модели

Итак, мы представили неизвестную нам функцию отклика полиномом. Операция замены одной функции другой, в каком-то смысле эквивалентной функцией называется аппроксимацией. Значит, мы аппроксимировали неизвестную функцию полиномом.

Но полиномы бывают разных степеней. Какой взять на первом шаге?

Эксперимент нужен только для того, чтобы найти численные значения коэффициентов полинома. Поэтому чем больше коэффициентов, тем больше опытов окажется необходимым. А мы стремимся сократить их число. Значит, надо найти такой полином, который содержит как можно меньше коэффициентов, но удовле-

творяет требованиям, предъявленным к модели. Чем ниже степень полинома при заданном числе факторов, тем меньше в нем коэффициентов.

Мы хотим, чтобы модель хорошо предсказывала направление наискорейшего улучшения параметра оптимизации. Такое направление называется направлением градиента. Ясно, что движение в этом направлении приведет к успеху быстрее, чем движение в любом другом направлении (это значит, что будет достигнута экономия числа опытов).

Как вы думаете, можно ли в этой связи всегда использовать полином первой степени?

С одной стороны, он содержит информацию о направлении градиента, с другой — в нем минимально возможное число коэффициентов при данном числе факторов. Единственное опасение в том, что неясно, будет ли линейная модель всегда адекватной. Ответ зависит еще и от объекта. Этим нам и предстоит сейчас заняться, чтобы завершить столь утомительную главу.

Вопрос в том, как выбрать подобласть в факторном пространстве, чтобы линейная модель оказалась адекватной. Условие аналитичности функции отклика гарантирует нам эту возможность. Всегда существует такая окрестность любой точки (точнее, почти любой точки), в которой линейная модель адекватна. Размер такой области заранее не известен, но адекватность, как вы помните, можно проверять по результатам эксперимента. Значит, выбрав сначала произвольную подобласть, мы, рано или поздно, найдем ее требуемые размеры. И как только это случится, воспользуемся движением по градиенту.

На следующем этапе мы будем искать линейную модель уже в другой подобласти. Цикл повторяется до тех пор, пока движение по градиенту не перестанет давать эффект. Это значит, что мы попали в область, близкую к оптимуму. Такая область называется «почти стационарной». Здесь линейная модель уже не нужна. Либо попаданием в почти стационарную область задача решена (случай, рассматриваемый в этой книге), либо надо переходить к полиномам более высоких степеней, например второй степени, чтобы подробнее описать область оптимума.

Удачный выбор подобласти имеет, как вы видите, большое значение для успеха всей работы. Он связан с интуитивными решениями, которые принимает экспериментатор на каждом этапе. Как это делается, мы рассмотрим ниже — в следующей главе и в гл. 11 и 13.

Кроме задачи оптимизации, иногда возникает задача построения интерполяционной модели. В этом случае нас не интересует оптимум. Просто мы хотим предсказывать результат с требуемой точностью во всех точках некоторой заранее заданной области. Тут не приходится выбирать подобласть. Необходимо последовательно увеличивать степень полинома до тех пор, пока модель не окажется адекватной. Если адекватной оказывается линейная,

или неполная квадратная модель (без членов, содержащих квадраты факторов), то ее построение аналогично тому, что требуется для оптимизации. Поэтому мы попутно будем рассматривать и эту задачу.

5.4. Резюме

Итак, мы выбрали модель, которую будем систематически использовать на первом этапе планирования эксперимента. Это алгебраический полином первой степени — линейная модель.

Чтобы произвести такой выбор, нам понадобилось научиться изображать поверхность отклика в факторном пространстве, задаваемом прямоугольными декартовыми координатами, по осям которых откладываются в некотором масштабе значения (уровни) факторов и значения параметров оптимизации. Поверхность отклика задана только в совместной области определения факторов. В этой области каждому возможному набору значений факторов (состоянию объекта) соответствует единственное значение параметра оптимизации. Для уменьшения размерности факторного пространства при геометрическом построении поверхности отклика можно использовать сечения.

Мы выяснили, что математическая модель требуется для предсказания направления градиента, т. е. направления, в котором величина параметра оптимизации улучшается быстрее, чем в любом другом направлении. Такая модель позволяет избежать полного перебора состояний объекта и тем самым уменьшить количество опытов, необходимых для отыскания оптимума.

Отказ от полного перебора требует оплаты в виде предположений о свойствах поверхности отклика, которые мы не сможем проверить. Такие предположения можно выбрать по-разному. Мы выбрали предположения об аналитичности функции отклика и об единственности оптимума. Аналитической называется такая функция, которую можно разложить в степенной ряд в окрестностях любой точки из области ее определения.

Используя эти предпосылки, можно предложить процедуру поиска оптимума, основанную на шаговом принципе. Этот принцип гласит: проводи короткие (насколько возможно) серии опытов, по их результатам строй математическую модель, используй модель для оценки градиента, ставь новые опыты только в этом направлении. Получается циклический процесс, который заканчивается при попадании в область, близкую к оптимуму («почти стационарную» область).

Чтобы выбрать теперь конкретную модель, надо сформулировать конкретные требования. К ним относятся адекватность и простота.

Под адекватностью понимается способность модели предсказывать результаты эксперимента в некоторой области с требуемой

точностью. После реализации опытов можно проверить адекватность модели.

Простота — вещь относительная. Мы просто условились считать алгебраические полиномы самыми простыми. Это соглашение базируется на накопленном разными исследователями опыте работы с такими моделями и обычно удовлетворяет экспериментатора. Кроме того, полином линеен относительно неизвестных коэффициентов, что упрощает обработку наблюдений.

Так мы выбрали класс моделей. Осталось выбрать степень полинома и подобласть, в которой надо начинать эксперимент. Эти выборы связаны между собой. Однако важно, что в принципе возможен такой выбор области, при котором линейная модель окажется адекватной. Этого достаточно, чтобы оценить градиент.

Выбор области связан с теми интуитивными решениями, которые принимает экспериментатор на каждом этапе работы.

Попутно мы упомянули о задаче построения интерполяционных моделей, которые используются для предсказания откликов во всей области. Область фиксируется заранее. Надо последовательно повышать степень полинома до тех пор, пока не найдется адекватная модель.

Так как же, наконец, выбирать условия проведения опытов в первом эксперименте, что такое на практике экспериментальный план?

Вы узнаете это, когда перейдете к следующей главе.

Литература

1. Л. А. Растригин. Статистические методы поиска. М., «Наука», 1968.
2. А. Вальд. Последовательный анализ. М., Физматгиз, 1960.
3. В. В. Налимов. Теория эксперимента. М., «Наука», 1971.

Глава шестая

ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Нет правила, которое нельзя было бы нарушить ради более прекрасного.

Л. ван Бетховен

«Наконец-то!»—воскликнет нетерпеливый читатель, справедливо полагая, что речь пойдет о планировании эксперимента. Но ... терпение. Прежде чем приступить к планированию, попытаемся дать ответы на вопросы, поставленные в четвертой главе. Прежде всего, как выбрать локальную область факторного пространства, где ее выбирать и какого размера она должна быть? Это важный этап принятия неформализованных решений, предшествующих построению плана первой серии эксперимента.

Здесь мы впервые сталкиваемся с проблемой принятия решений при планировании эксперимента. Далее мы уже не расстанемся с этой темой. Поэтому уместны несколько слов об особенностях этих этапов решения задачи [1]. Весь процесс исследования можно считать состоящим из последовательности этапов, часть из которых полностью формализована, а часть требует «интуитивных» решений. Причем, по мере развития теории формальные этапы будут играть все большую роль, но до конца не вытеснят неформализованные этапы. В силу этого, между прочим, не ожидается создание «логарифмической линейки» по планированию эксперимента и надо тратить время на его изучение.

6.1. Принятие решений перед планированием эксперимента

При выборе области эксперимента прежде всего надо оценить границы областей определения факторов. При этом должны учитываться ограничения нескольких типов. Первый тип — принципиальные ограничения для значений факторов, которые не могут быть нарушены ни при каких обстоятельствах. Например, если фактор — температура, то нижним пределом будет абсолютный нуль. Второй тип — ограничения, связанные с технико-экономическими соображениями, например, со стоимостью сырья, дефицитностью отдельных компонентов, временем ведения процесса. Третий тип ограничений, с которым чаще всего приходится иметь

дело, определяется конкретными условиями проведения процесса, например, существующей аппаратурой, технологией, организацией. В реакторе, изготовленном из некоторого материала, температуру нельзя поднять выше температуры плавления этого материала или выше рабочей температуры данного катализатора.

Оптимизация обычно начинается в условиях, когда объект уже подвергался некоторым исследованиям. Информацию, содержащуюся в результатах предыдущих исследований, будем называть априорной (т. е. полученной до начала эксперимента). Мы можем использовать априорную информацию для получения представления о параметре оптимизации, о факторах, о наилучших условиях ведения процесса и характере поверхности отклика, т. е. о том, как сильно меняется параметр оптимизации при небольших изменениях значений факторов, а также о кривизне поверхности. Для этого можно использовать графики (или таблицы) однофакторных экспериментов, осуществлявшихся в предыдущих исследованиях или описанных в литературе. Если однофакторную зависимость нельзя представить линейным уравнением (в рассматриваемой области), то в многомерном случае, несомненно, будет существенная кривизна. Обратное утверждение, к сожалению, не очевидно.

Итак, выбор экспериментальной области факторного пространства связан с тщательным анализом априорной информации. Поясним наши рассуждения примером [2].

Пример 1. Изучалось ионообменное разделение смесей группы редкоземельных элементов растворами иминодиуксусной кислоты. Параметр оптимизации — содержание неодима в выходном растворе (элюате) в процентах. Рассматривалось всего два фактора: концентрация элюанта (входного раствора), % вес (\tilde{x}_1) и pH элюанта (\tilde{x}_2). Как построить область определения факторов? Начнем с \tilde{x}_1 . Известно, что при $\tilde{x}_1 > 3$ работать нельзя, так как это предел растворимости данного вещества при нормальной температуре. Значит, верхний предел $\tilde{x}_1 = 3$. С нижним пределом дело обстоит сложнее. Здесь нельзя указать четкую границу. Известно только, что чем ниже концентрация, тем дольше идет процесс. При $\tilde{x}_1 = 0,5$ время протекания процесса находится в разумных пределах. Это и определяет нижнюю границу. Ради большой выгоды ее можно будет сдвинуть, тогда как изменить верхнюю границу практически нельзя.

Для выбора области определения \tilde{x}_2 использовались теоретические представления о процессе, из которых следует, что разделение происходит благодаря одновременному присутствию в системе двух соединений: моно- и ди-комплексов. Специальные предварительные опыты показали, что при $\text{pH} < 3$ кислота находится в недиссоциированном состоянии, а при $\text{pH} > 8$ оба соединения разрушаются. Следовательно, \tilde{x}_2 может изменяться от 3 до 8. Если факторы совместимы (а в данном случае это так), то их совместная область определения тоже задана (см. гл. 4):

Вы видите, как непросто решается этот важный вопрос. Но это только начало. Теперь в области определения надо найти локальную подобласть для планирования эксперимента. Процедура выбора этой подобласти включает два этапа: выбор основного уровня и выбор интервалов варьирования.

Выбор основного уровня. Наилучшим условиям, определенным из анализа априорной информации, соответствует комбинация (или несколько комбинаций) уровней факторов. Каждая комбинация является многомерной точкой в факторном пространстве. Ее можно рассматривать как исходную точку для построения плана эксперимента. Назовем ее основным (нулевым) уровнем. Построение плана эксперимента сводится к выбору экспериментальных точек, симметричных относительно нулевого уровня.

В разных случаях мы располагаем различными сведениями об области наилучших условий. Если имеются сведения о координатах одной наилучшей точки и нет информации о границах определения факторов, то остается рассматривать эту точку в качестве основного уровня. Аналогичное решение принимается, если границы известны и наилучшие условия лежат внутри области.

Положение усложняется, если эта точка лежит на границе (или весьма близко к границе) области. Тогда приходится основную уровень выбирать с некоторым сдвигом от наилучших условий.

Может случиться, что координаты наилучшей точки неизвестны, но есть сведения о некоторой подобласти, в которой процесс идет достаточно хорошо. Тогда основной уровень выбирается либо в центре, либо в случайной точке этой подобласти. Сведения о подобласти можно получить, анализируя изученные ранее подобные процессы, из теоретических соображений или из предыдущего эксперимента.

Наконец, возможен случай с несколькими эквивалентными точками, координаты которых различны. Когда отсутствуют дополнительные данные (технологического, экономического характера и т. д.), выбор произволен. Конечно, если эксперимент недорог и требует немного времени, можно приступить к построению планов экспериментов вокруг нескольких точек.

Следующий пример иллюстрирует одну из возможных ситуаций.

Пример 2. На рис. 18 изображена область определения для двух факторов. Крестиком отмечены наилучшие условия, известные из априорной информации. Известно также, что имеется возможность дальнейшего улучшения параметра оптимизации, а данное значение нас не удовлетворяет. Эту точку нельзя рассматривать в качестве основного уровня. Дело в том, что она расположена на границе области определения. Требование симметрии экспериментальных точек относительно нулевого уровня привело бы в этом случае к выходу за границы области определения, чего делать также нельзя.

Резюмируем наши рассуждения о принятии решений при выборе основного уровня в виде блок-схемы (см. стр. 73).

После того как нулевой уровень выбран, переходим к следующему шагу — выбору интервалов варьирования.

Выбор интервалов варьирования. Теперь наша цель состоит в том, чтобы для каждого фактора выбрать два уровня, на которых он будет варьироваться в эксперименте.

Представьте себе координатную ось, на которой откладываются значения данного фактора, для определенности — температуры. Пусть основной уровень уже выбран и равен 100°C . Это значение изображается точкой. Тогда два интересующих нас уровня можно изобразить двумя точками, симметричными относительно первой. Будем называть один из этих уровней верхним, а второй — нижним. Обычно за верхний уровень принимается тот, который соответствует большему значению фактора, хотя это не обязательно, а для качественных факторов вообще безразлично.

Интервалом варьирования факторов называется некоторое число (свое для каждого фактора), прибавление которого к основному уровню дает верхний, а вычитание — нижний уровни фактора. Другими словами, интервал варьирования — это расстояние на координатной оси между основным и верхним (или нижним) уровнем. Таким образом, задача выбора уровней сводится к более простой задаче выбора интервала варьирования.

Заметим еще, что для упрощения записи условий эксперимента и обработки экспериментальных данных масштабы по осям выбираются так, чтобы верхний уровень соответствовал $+1$, нижний -1 , а основной — нулю. Для факторов с непрерывной областью определения это всегда можно сделать с помощью преобразования

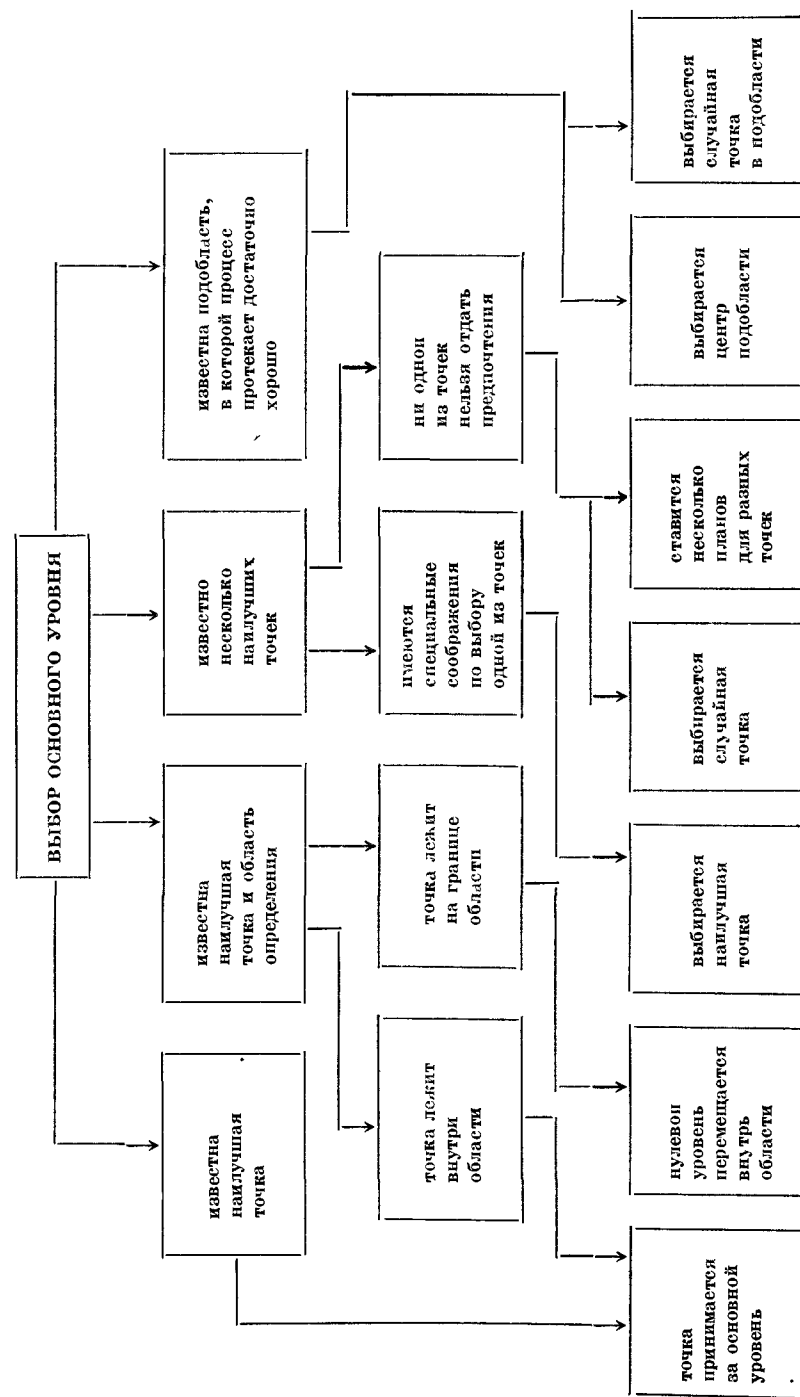
$$x_j = \frac{\tilde{x}_j - \tilde{x}_{j0}}{I_j},$$

где x_j — кодированное значение фактора, \tilde{x}_j — натуральное значение фактора, \tilde{x}_{j0} — натуральное значение основного уровня, I_j — интервал варьирования, j — номер фактора.

Для качественных факторов, имеющих два уровня, один уровень обозначается $+1$, а другой -1 ; порядок уровней не имеет значения.

Пусть процесс определяется четырьмя факторами. Основной уровень и интервалы варьирования выбраны следующим образом.

	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4
Основной уровень	3	30	1,5	15
Интервал варьирования	2	10	1	10



Блок-схема принятия решений при выборе основного уровня

Остановимся на первом факторе. Отметим на координатной оси три уровня: нижний, основной и верхний.

Натуральные значения	1	2	3	5	\bar{x}_1
Кодированные значения	-1	\times	0	+1	x_1

Нужно найти кодированное значение для $\bar{x}_1 = 2,0$. Это значение лежит между 1,0 и 3,0, т. е. между -1 и 0 в кодированном масштабе. Так как в натуральном масштабе 2,0 лежит посередине между 1,0 и 3,0, то ему соответствует -0,5 в кодированном масштабе. (Для $\bar{x}_1 = 2,5$ будет $x_1 = -0,25$, для $\bar{x}_1 = 1,5$ будет $x_1 = -0,75$ и т. д.)

На выбор интервалов варьирования накладываются естественные ограничения сверху и снизу. Интервал варьирования не может быть меньше той ошибки, с которой экспериментатор фиксирует уровень фактора. Иначе верхний и нижний уровни окажутся неразличимыми. С другой стороны, интервал не может быть настолько большим, чтобы верхний или нижний уровни оказались за пределами области определения. Внутри этих ограничений обычно еще остается значительная неопределенность выбора, которая устраняется с помощью интуитивных решений.

Обратите внимание, что при решении задачи оптимизации мы стремимся выбрать для первой серии экспериментов такую под-область, которая давала бы возможность для шагового движения к оптимуму. В задачах же интерполяции интервал варьирования охватывает всю описываемую область.

Выбор интервалов варьирования — задача трудная, так как она связана с неформализованным этапом планирования эксперимента. Возникает вопрос, какая априорная информация может быть полезна на данном этапе? Это — сведения о точности, с которой экспериментатор фиксирует значения факторов, о кривизне поверхности отклика и о диапазоне изменения параметра оптимизации. Обычно эта информация является ориентировочной (в некоторых случаях она может оказаться просто ошибочной), но это единственная разумная основа, на которой можно начинать планировать эксперимент. В ходе эксперимента ее часто приходится корректировать.

Точность фиксирования факторов определяется точностью приборов и стабильностью уровня в ходе опыта. Для упрощения схемы принятия решений мы введем приближенную классификацию, полагая, что есть низкая, средняя и высокая точности. Можно, например, считать, что поддержание температуры в реакторе с погрешностью не более 1% соответствует высокой, не более 5% — средней, а более 10% — низкой точности.

Источником сведений о кривизне поверхности отклика могут служить уже упоминавшиеся графики однофакторных зависимо-

стей, а также теоретические соображения. Из графиков сведения о кривизне можно получить визуально. Некоторое представление о кривизне дает анализ табличных данных, так как наличие кривизны соответствует непропорциональное изменение параметра оптимизации при равномерном изменении фактора. Мы будем различать три случая: функция отклика линейна, функция отклика существенно нелинейна и информация о кривизне отсутствует.

Наконец, полезно знать, в каких диапазонах меняются значения параметра оптимизации в разных точках факторного пространства. Если имеются результаты некоторого множества опытов, то всегда можно найти наибольшее или наименьшее значения параметра оптимизации. Разность между этими значениями будем называть диапазоном изменения параметра оптимизации для данного множества опытов. Условимся различать широкий и узкий диапазоны. Диапазон будет узким, если он несущественно отличается от разброса значений параметра оптимизации в повторных опытах. (Этот разброс, как вы узнаете в гл. 8, определяет ошибку опыта.) В противном случае будем считать диапазон широким. Учтем также случай, когда информация отсутствует. Итак, для принятия решений используется априорная информация о точности фиксирования факторов, кривизне поверхности отклика и диапазоне изменения параметра оптимизации. Каждое сочетание градаций перечисленных признаков определяет ситуацию, в которой нужно принимать решение.

При принятых нами градациях возможно $3^3 = 27$ различных ситуаций. Они представлены на рис. 19—21 в виде кружочков, цифры в которых соответствуют порядковым номерам ситуаций.

Теперь мы приблизились к принятию решения о выборе интервалов варьирования. Для интервалов также введем градацию. Будем рассматривать широкий, средний и узкий интервалы варьирования, а также случай, когда трудно принять однозначное решение. Размер интервала варьирования составляет некоторую долю от области определения фактора. Можно, например, условиться о следующем: если интервал составляет не более 10% от области определения, считать его узким, не более 30% — средним и в остальных случаях — широким. Это, конечно, весьма условно, и в каждой конкретной задаче приходится специально определять эти понятия, которые зависят не только от размера области определения, но и от характера поверхности отклика, и от точности фиксирования факторов.

Перейдем к рассмотрению блок-схем принятия решений. На первой схеме (рис. 19) представлены девять ситуаций, имеющих место при низкой точности фиксирования факторов. При выборе решений учитываются информация о кривизне поверхности отклика и о диапазоне изменения параметра оптимизации. Типичное решение — широкий интервал варьирования. Узкий интервал

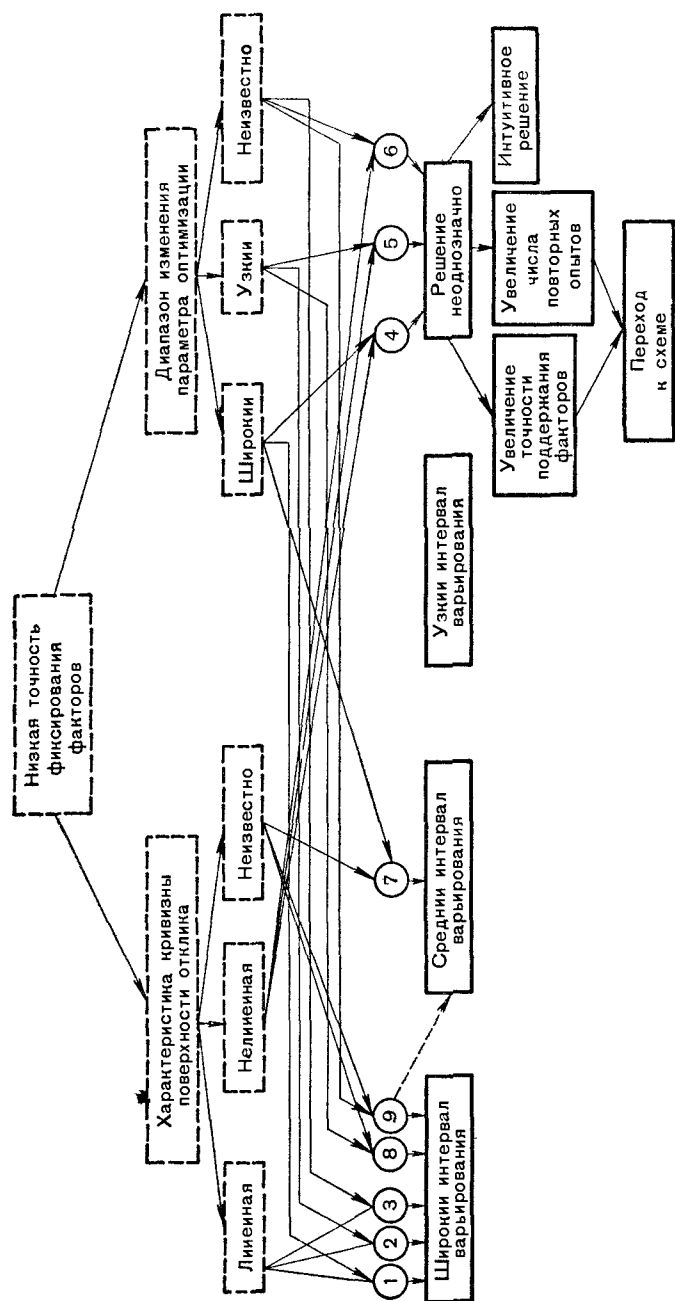


Рис. 19. Принятие решений при низкой точности фиксации факторов

варьирования совершенно не используется, что вполне понятно при низкой точности.

Пусть ситуация определяется следующими признаками: поверхность отклика линейна, а диапазон изменения параметра оптимизации узок. Какое решение вы бы предпочли? Эта ситуация обозначена на нашей схеме номером 2. Признаки ситуации определяются стрелками, направленными к данному кружочку. Стрелка, выходящая из кружочка, указывает решение. Низкая точность фиксирования факторов приводит к отказу от выбора узкого интервала варьирования, иначе результаты могут оказаться неразличимыми. Нам известно, что поверхность линейна. Это не налагает ограничений на расширение интервалов. Кроме того, надо учитывать сведения о диапазоне изменения параметра оптимизации. Он узок, а мы стремимся получить в эксперименте различающиеся значения параметра оптимизации. Поэтому интервал следует увеличивать.

Вернемся снова к блок-схеме. Вы видите, что средний интервал варьирования в этой схеме выбирается дважды, причем в девятой ситуации как редко применяемая альтернатива. Здесь отсутствует информация об обоих признаках, и выбор широкого интервала представляется более естественным.

Наибольшие трудности возникают, когда поверхность отклика нелинейна. Появляется противоречие между низкой точностью фиксирования факторов и кривизной. Первая требует расширения интервала, а вторая — сужения. Решение оказывается неоднозначным. Как поступить? Приходится рассматривать дополнительные рекомендации (см. блок-схему). Прежде всего нужно выяснить, нельзя ли увеличить точность эксперимента либо за счет инженерных решений, либо за счет увеличения числа повторных опытов. Если это возможно, то решения принимаются на основе блок-схемы (рис. 20) для средней точности фиксирования факторов. Если это невозможно, то для принятия решения нет достаточных оснований и оно становится интуитивным.

Это блок-схема, как и последующие, служит весьма грубым приближением к действительности. На практике учитывается еще масса обстоятельств. Например, решения, принимаемые по каждому фактору в отдельности, корректируются при рассмотрении совокупности факторов.

На рис. 20 изображена блок-схема для случая средней точности фиксирования факторов.

Характерен выбор среднего интервала варьирования. Лишь в случае нелинейной поверхности и широкого диапазона рекомендуется узкий интервал варьирования. При сочетаниях линейной поверхности с узким диапазоном и отсутствием информации о диапазоне выбирается широкий интервал варьирования. Пунктиром, как и выше, показаны редко применяемые альтернативы.

Наконец, на рис. 21 построена блок-схема для случая высокой точности фиксирования фактора. Сочетание высокой точ-

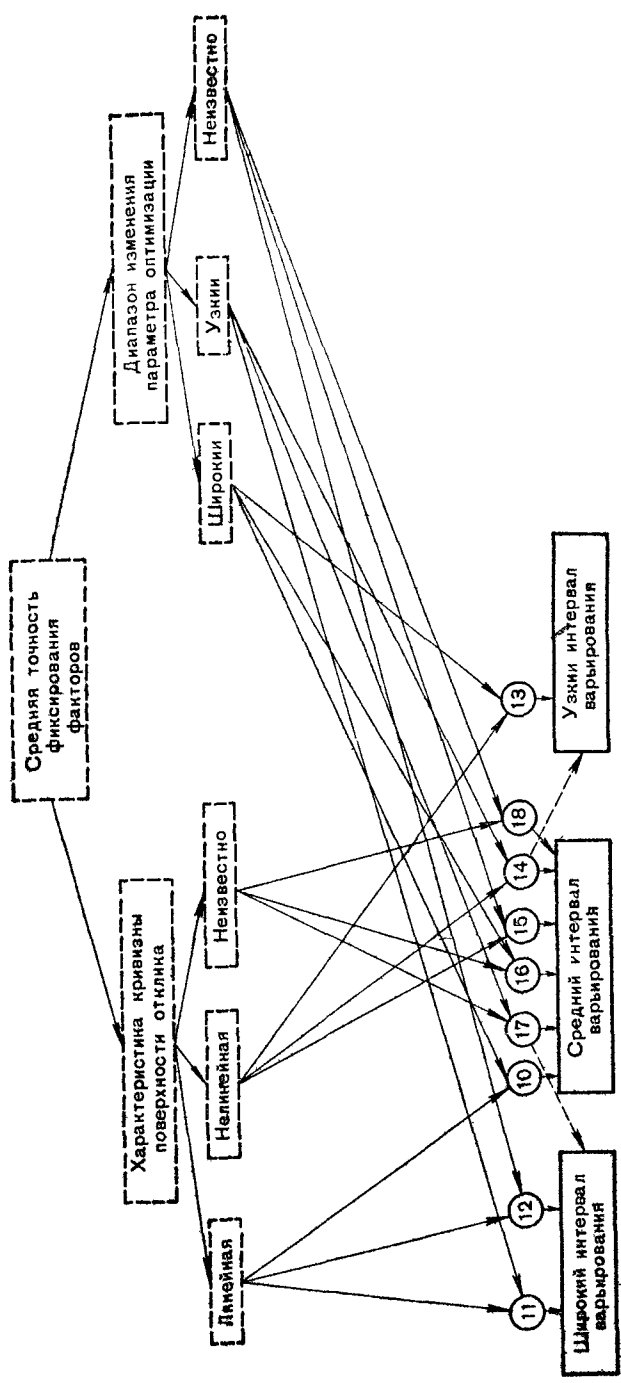


Рис. 20. Принятие решений при средней точности фиксирования факторов

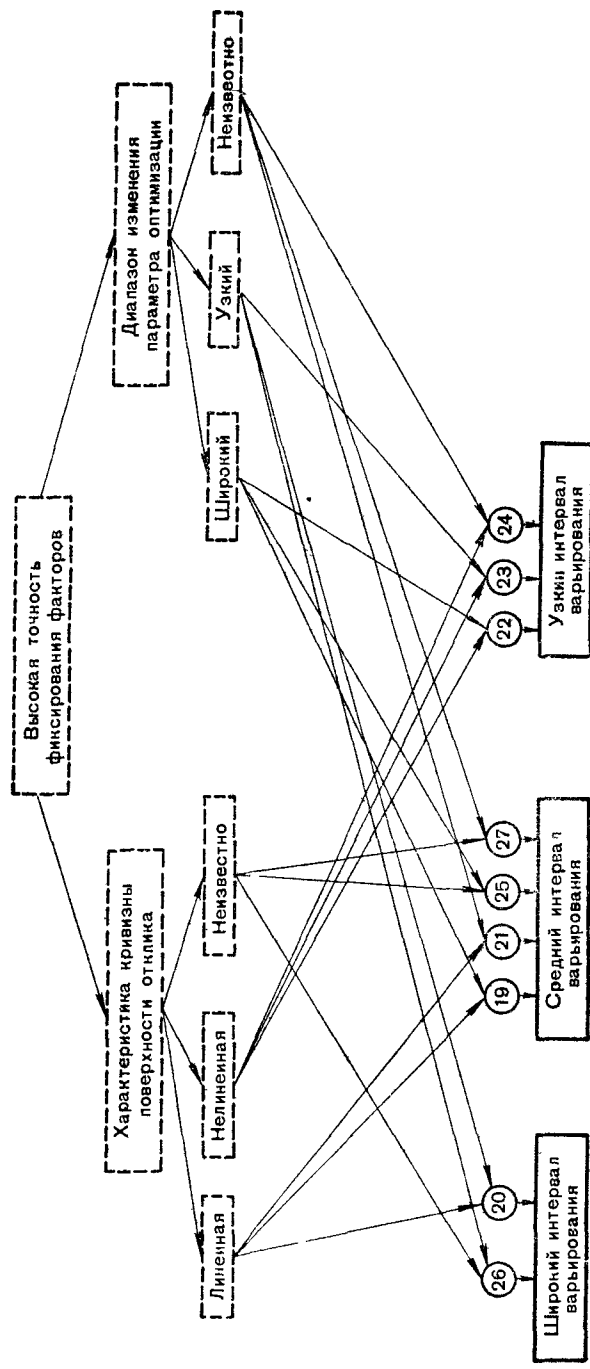


Рис. 21. Принятие решений при высокой точности фиксирования факторов

ности с нелинейностью поверхности всегда приводит к выбору узкого интервала. Довольно часто выбирается средний интервал и лишь в двух случаях широкий. В обоих последних блок-схемах отсутствуют неоднозначные решения.

Пример 3. Давайте продолжим рассмотрение примера 1. Вы помните, что область определения факторов была выбрана следующим образом: для \tilde{x}_1 от 0,5 до 3, для \tilde{x}_2 от 3 до 8. Основной уровень: $\tilde{x}_1=1,5$, $\tilde{x}_2=7,0$.

Экспериментатор имел такую априорную информацию: точность фиксирования факторов средняя, поверхность отклика линейна, диапазон изменения параметра оптимизации довольно узок. Этот случай соответствует ситуации 11 блок-схемы (рис. 20). Принимаемое решение — широкий интервал варьирования. Экспериментатор выбрал такие интервалы: $I_1=0,5$, $I_2=1,0$, что составляет 20% от области определения факторов. Это несомненно широкие интервалы варьирования. Отметим еще, что для \tilde{x}_2 основной уровень выбран вблизи границы области определения. Поэтому рекомендация о выборе широкого интервала варьирования приводит к совпадению верхнего уровня с этой границей. Так на практике осуществляется выбор интервалов варьирования. Ниже, в гл. 11, мы продолжим рассмотрение этого примера и посмотрим, как оправдываются принятые решения.

Итак, вооружившись умением выбирать основной уровень и интервалы варьирования факторов, мы готовы приступить к построению плана проведения эксперимента.

6.2. Полный факторный эксперимент типа 2^k

Первый этап планирования эксперимента для получения линейной модели основан, как мы договорились, на варьировании факторов на двух уровнях. В этом случае, если число факторов известно, можно сразу найти число опытов, необходимое для реализации всех возможных сочетаний уровней факторов. Простая формула, которая для этого используется, уже приводилась в гл. 1, и мы ее напомним: $N = 2^k$, где N — число опытов, k — число факторов, 2 — число уровней. В общем случае эксперимент, в котором реализуются все возможные сочетания уровней факторов, называется полным факторным экспериментом. Если число уровней каждого фактора равно двум, то имеем полный факторный эксперимент типа 2^k [3, 4].

Таблица 6.1

Матрица планирования эксперимента 2^2

Номер опыта	x_1	x_2	y	Номер опыта	x_1	x_2	y
1	-1	-1	y_1	3	-1	+1	y_3
2	+1	-1	y_2	4	+1	+1	y_4

Нетрудно написать все сочетания уровней в эксперименте с двумя факторами. Напомним, что в планировании эксперимента используются кодированные значения факторов: +1 и -1 (часто для простоты записи единицы опускают). Условия эксперимента можно записать в виде таблицы, где строки соответствуют различным опытам, а столбцы — значениям факторов. Будем называть такие таблицы матрицами планирования эксперимента.

Матрица планирования для двух факторов приведена в табл. 6.1.

Каждый столбец в матрице планирования называют вектор-столбцом, а каждую строку — вектор-строкой. Таким образом, в табл. 6.1 мы имеем два вектора-столбца независимых переменных и один вектор-столбец параметра оптимизации. То, что записано в этой таблице в алгебраической форме, можно изобразить геометрически. Найдем в области определения факторов точку, соответствующую основному уровню, и проведем через нее новые оси координат, параллельные осям натуральных значений факторов. Далее, выберем масштабы по новым осям так, чтобы интервал варьирования для каждого фактора равнялся единице. Тогда условия проведения опытов будут соответствовать вершинам квадрата, центром которого является основной уровень, а каждая сторона параллельна одной из осей координат и равна двум интервалам (рис. 22). Номера вершин квадрата соответствуют номерам опытов в матрице планирования. Площадь, ограниченная квадратом, называется областью эксперимента. Иногда удобнее считать областью эксперимента площадь, ограниченную окружностью, описывающей квадрат. В задачах интерполяции область эксперимента есть область предсказываемых значений y .

Запись матрицы планирования, особенно для многих факторов, громоздка. Для ее сокращения удобно ввести условные буквенные обозначения строк.

Это делается следующим образом. Порядковый номер фактора ставится в соответствие строчной букве латинского алфавита: $x_1 - a$, $x_2 - b$, ... и т. д. Если теперь для строки матрицы планирования выписать латинские буквы только для факторов, находящихся на верхних уровнях, то условия опыта будут заданы однозначно. Опыт со всеми факторами на нижних уровнях условимся обозначать (1). Матрица планирования вместе с принятыми буквенными обозначениями приведена в табл. 6.2.

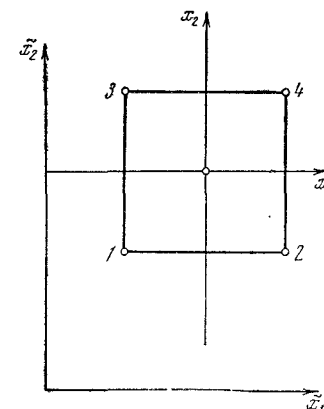


Рис. 22. Геометрическая интерпретация полного факторного эксперимента 2^2

Таблица 6.2

Матрица планирования эксперимента 2^3

Номер опыта	x_1	x_2	Буквенные обозначения строк	y
1	-1	-1	(I)	y_1
2	+1	-1	a	y_2
3	-1	+1	b	y_3
4	+1	+1	ab	y_4

Теперь вместо полной записи матрицы планирования можно пользоваться только буквенными обозначениями. Ниже приведена буквенная запись еще одного плана: c , b , a , abc , (I) , bc , ac , ab . Матрица планирования приведена в табл. 6.3.

Таблица 6.3

Матрица планирования эксперимента 2^3

Номер опыта	x_1	x_2	x_3	Буквенные обозначения строк	y
1	-1	-1	+1	c	y_1
2	-1	+1	-1	b	y_2
3	+1	-1	-1	a	y_3
4	+1	+1	+1	abc	y_4
5	-1	-1	-1	(I)	y_5
6	-1	+1	+1	bc	y_6
7	+1	-1	+1	ac	y_7
8	+1	+1	-1	ab	y_8

Таким образом вы построили полный факторный эксперимент 2^3 . Он имеет восемь опытов и включает все возможные комбинации уровней трех факторов.

Если для двух факторов все возможные комбинации уровней легко найти прямым перебором (или просто запомнить), то с ростом числа факторов возникает необходимость в некотором приеме построения матриц. Из многих возможных обычно используется три приема, основанных на переходе от матриц меньшей размерности к матрицам большей размерности. Рассмотрим первый. При добавлении нового фактора каждая комбинация уровней исходного плана встречается дважды: в сочетании с нижним и верхним уровнями нового фактора. Отсюда естественно появляется прием: записать исходный план для одного уровня нового фактора, а затем повторить его для другого уровня. Вот как это выглядит при переходе от эксперимента 2^2 к 2^3 (табл. 6.4):

Таблица 6.4

Построение матрицы планирования эксперимента 2^3

Номер опыта	x_1	x_2	x_3	y	Номер опыта	x_1	x_2	x_3	y
1	-	-	+	y_1	5	-	-	-	y_5
2	-	+	+	y_2	6	-	+	-	y_6
3	+	-	+	y_3	7	+	-	-	y_7
4	+	+	+	y_4	8	+	+	-	y_8

Этот прием распространяется на построение матриц любой размерности.

Рассмотрим второй прием. Для этого введем правило перемножения столбцов матрицы. При построчном перемножении двух столбцов матрицы произведение единиц с одноименными знаками дает +1, а с разноименными —1. Воспользовавшись этим правилом, получим для случая, который мы рассматриваем, вектор-столбец произведения x_1x_2 в исходном плане. Далее повторим еще раз исходный план, а у столбца произведений знаки поменяем на обратные

Этот прием тоже можно перенести на построение матриц любой размерности, однако он сложнее, чем первый.

Третий прием основан на правиле чередования знаков. В первом столбце знаки меняются поочередно, во втором столбце они чередуются через два, в третьем — через 4, в четвертом — через 8 и т. д. по степеням двойки. Если в табл. 6.4 поменять местами столбцы для x_1 и x_2 , то получится нужная матрица.

По аналогии с полным факторным экспериментом 2^2 можно дать геометрическую интерпретацию полного факторного эксперимента 2^3 . Геометрической интерпретацией полного факторного эксперимента 2^3 служит куб, координаты вершин которого задают условия опытов.

Если поместить центр куба в точку основного уровня факторов, а масштабы по осям выбрать так, чтобы интервал варьирования равнялся единице, то получится куб, изображенный на рис. 23. Куб задает область эксперимента, а центр куба является ее центром.

К сожалению, мы не умеем рисовать картинки для числа факторов $k > 3$. Но фигура, задающая область эксперимента в многомерном пространстве, является некоторым аналогом куба. Будем называть эту фигуру гиперкубом.

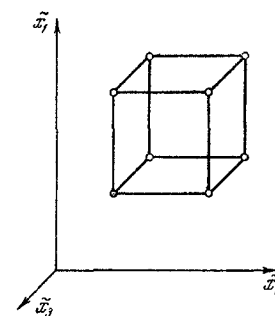


Рис. 23. Геометрическая интерпретация полного факторного эксперимента 2^3

6.3. Свойства полного факторного эксперимента типа 2^k

Мы научились строить матрицы планирования полных факторных экспериментов с факторами на двух уровнях. Теперь выясним, какими общими свойствами эти матрицы обладают независимо от числа факторов. Говоря о свойствах матриц, мы имеем в виду те из них, которые определяют качество модели. Ведь эксперимент и планируется для того, чтобы получить модель, обладающую некоторыми оптимальными свойствами. Это значит, что оценки коэффициентов модели должны быть наилучшими и что точность предсказания параметра оптимизации не должна зависеть от направления в факторном пространстве, ибо заранее неясно, куда предстоит двигаться в поисках оптимума.

Два свойства следуют непосредственно из построения матрицы. Первое из них — симметричность относительно центра эксперимента — формулируется следующим образом: алгебраическая сумма элементов вектор-столбца каждого фактора равна нулю, или $\sum_{i=1}^N x_{ji} = 0$, где j — номер фактора, N — число опытов, $j = 1, 2, \dots, k$.

Второе свойство — так называемое условие нормировки — формулируется следующим образом: сумма квадратов элементов каждого столбца равна числу опытов, или $\sum_{i=1}^N x_{ji}^2 = N$. Это следствие того, что значения факторов в матрице задаются $+1$ и -1 .

Мы рассмотрели свойства отдельных столбцов матрицы планирования. Теперь остановимся на свойстве совокупности столбцов.

Сумма почленных произведений любых двух вектор-столбцов матрицы равна нулю, или $\sum_{i=1}^N x_{ji}x_{ui} = 0$, $j \neq u$, $j, u = 0, 1, 2, \dots, k$. Это важное свойство называется ортогональностью матрицы планирования.

Последнее, четвертое свойство называется ротатабельностью, т. е. точки в матрице планирования подбираются так, что точность предсказания значений параметра оптимизации одинакова на равных расстояниях от центра эксперимента и не зависит от направления. Тех, кто интересуется доказательством этого утверждения, мы отсылаем к работе [3].

Даны две матрицы планирования:

	x_1	x_2		x_1	x_2
	—	—		—	+
a)	+	—	b)	+	—
	—	+		—	+
	+	+		+	—

Давайте проверим, как выполняются все три свойства для каждой из матриц. Первое свойство $\sum_{i=1}^N x_{ji} = 0$ выполняется для всех столбцов обеих матриц. Действительно, для первого столбца матрицы а) имеем

$$(-1) + (+1) + (-1) + (+1) = 0.$$

Аналогичный результат получается для всех других столбцов.

Второе свойство $\sum_{i=1}^N x_{ji}^2 = N$ — также выполняется для обеих матриц. Например, для того же столбца имеем

$$(-1)^2 + (+1)^2 + (-1)^2 + (+1)^2 = 4.$$

С третьим свойством, однако, дело обстоит иначе. Если для матрицы а) формула $\sum_{i=1}^N x_{ji}x_{ui} = 0$, $j \neq u$ выполняется, то в случае б) это не так. Действительно

$$\begin{aligned} &(-1)(+1) + (+1)(-1) + (-1)(+1) + \\ &+ (+1)(-1) = -4 \neq 0. \end{aligned}$$

6.4. Полный факторный эксперимент и математическая модель

Давайте еще раз вернемся к матрице 2^2 (табл. 6.2). Для движения к точке оптимума нам нужна линейная модель $y = b_0 + b_1x_1 + b_2x_2$. Наша цель — найти по результатам эксперимента значения неизвестных коэффициентов модели. До сих пор, говоря о линейной модели, мы не останавливались на важном вопросе о статистической оценке ее коэффициентов. Теперь необходимо сделать ряд замечаний по этому поводу. Можно утверждать, что эксперимент проводится для проверки гипотезы о том, что линейная модель $\eta = \beta_0 + \beta_1x_1 + \beta_2x_2$ адекватна. Греческие буквы использованы для обозначения «истинных» генеральных значений соответствующих неизвестных. Эксперимент, содержащий конечное число опытов, позволяет только получить выборочные оценки для коэффициентов уравнения $y = b_0 + b_1x_1 + \dots + b_kx_k$. Их точность и надежность зависят от свойств выборки и нуждаются в статистической проверке. Как производится такая проверка, вы узнаете в гл. 9. А пока займемся вычислением оценок коэффициентов. Их можно вычислить по простой формуле

$$b_j = \frac{\sum_{i=1}^N x_{ji}y_i}{N}, \quad j = 0, 1, \dots, k,$$

обоснование которой дается в гл. 9. Воспользуемся этой формулой для подсчета коэффициентов b_1 и b_2

$$b_1 = \frac{(-1)y_1 + (+1)y_2 + (-1)y_3 + (+1)y_4}{4},$$

$$b_2 = \frac{(-1)y_1 + (-1)y_2 + (+1)y_3 + (+1)y_4}{4}$$

Вы видите, что благодаря кодированию факторов расчет коэффициентов превратился в простую арифметическую процедуру. Для подсчета коэффициента b_1 используется вектор-столбец x_1 , а для b_2 — столбец x_2 . Остается неясным, как найти b_0 . Если наше уравнение $y = b_0 + b_1x_1 + b_2x_2$ справедливо, то оно верно и для средних арифметических значений переменных: $\bar{y} = b_0 + b_1\bar{x}_1 + b_2\bar{x}_2$. Но в силу свойства симметрии $\bar{x}_1 = \bar{x}_2 = 0$. Следовательно, $\bar{y} = b_0$. Мы показали, что b_0 есть среднее арифметическое значение параметра оптимизации. Чтобы его получить, необходимо сложить все y и разделить на число опытов. Чтобы привести эту процедуру в соответствие с формулой для вычисления коэффициентов, в матрицу планирования удобно ввести вектор-столбец фиктивной переменной x_0 , которая принимает во всех опытах значение $+1$. Это было уже учтено в записи формулы, где j принимало значения от 0 до k .

Теперь у нас есть все необходимое, чтобы найти неизвестные коэффициенты линейной модели

$$y = b_0x_0 + b_1x_1 + b_2x_2.$$

Коэффициенты при независимых переменных указывают на силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает фактор. Если коэффициент имеет знак плюс, то с увеличением значения фактора параметр оптимизации увеличивается, а если минус, то уменьшается. Величина коэффициента соответствует вкладу данного фактора в величину параметра оптимизации при переходе фактора с нулевого уровня на верхний или нижний.

Иногда удобно оценивать вклад фактора при переходе от нижнего к верхнему уровню. Вклад, определенный таким образом, называется эффектом фактора (иногда его называют основным или главным эффектом). Он численно равен удвоенному коэффициенту. Для качественных факторов, варьируемых на двух уровнях, основной уровень не имеет физического смысла. Поэтому понятие «эффект фактора» является здесь естественным.

Планируя эксперимент, на первом этапе мы стремимся получить линейную модель. Однако у нас нет гарантии, что в выбранных интервалах варьирования процесс описывается линейной моделью. Существуют способы проверки пригодности линейной модели (проверка адекватности — см. гл. 9). А если модель не-

линейна, как количественно оценить нелинейность, пользуясь полным факторным экспериментом?

Один из часто встречающихся видов нелинейности связан с тем, что эффект одного фактора зависит от уровня, на котором находится другой фактор. В этом случае говорят, что имеет место эффект взаимодействия двух факторов. Полный факторный эксперимент позволяет количественно оценивать эффекты взаимодействия. Для этого надо, пользуясь правилом перемножения столбцов, получить столбец произведения двух факторов. При вычислении коэффициента, соответствующего эффекту взаимодействия, с новым вектор-столбцом можно обращаться так же, как с вектор-столбцом любого фактора. Для полного факторного эксперимента 2^2 матрица планирования с учетом эффекта взаимодействия представлена в табл. 6.5. Очень важно, что при добавлении столбцов эффектов взаимодействий все рассмотренные свойства матриц планирования сохраняются.

Таблица 6.5

Матрица планирования эксперимента 2^2 с эффектом взаимодействия

Номер опыта	x_0	x_1	x_2	x_1x_2	y	Номер опыта	x_0	x_1	x_2	x_1x_2	y
1	+1	+1	+1	+1	y_1	3	+1	-1	-1	+1	y_3
2	+1	-1	+1	-1	y_2	4	+1	+1	-1	-1	y_4

Теперь модель выглядит следующим образом:

$$y = b_0x_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2.$$

Коэффициент b_{12} вычисляется обычным путем

$$b_{12} = \frac{(+1)y_1 + (-1)y_2 + (+1)y_3 + (-1)y_4}{4}.$$

Столбцы x_1 и x_2 задают планирование — по ним непосредственно определяются условия опытов, а столбцы x_0 и x_1x_2 служат только для расчета.

Обращаем ваше внимание на то, что при оптимизации мы стремимся сделать эффекты взаимодействия возможно меньшими. В задачах интерполяции, напротив, их выявление часто важно и интересно.

Покажем на примере еще один способ расчета коэффициентов, известный под названием метода Йетса [5]. Все операции по расчету приведены в табл. 6.6.

Слева в этой таблице выписан вектор-столбец значений параметра оптимизации. Первая операция (2-й столбец) состоит в попарном сложении и вычитании этих значений, причем верхнее число вычитается из нижнего. Вторая операция (3-й столбец) состоит в том же действии, но уже с числами второго столбца.

Таблица 6.6

Расчет коэффициентов регрессии по методу Йетса

1	2	3	1	2	3
y_1	$y_1 + y_2$	$y_1 + y_2 + y_3 + y_4$	y_3	$y_2 - y_1$	$y_3 + y_4 - y_1 - y_2$
y_2	$y_3 + y_4$	$y_2 - y_1 + y_4 - y_3$	y_4	$y_4 - y_3$	$y_1 - y_3 - y_2 + y_4$

Если теперь числа, оказавшиеся в третьем столбце, разделить на число опытов, то получим значения коэффициентов. Операции сложения и вычитания повторяются столько раз, сколько имеется факторов.

Рассмотрим пример.

Пример 4. В табл. 6.7 приводятся результаты эксперимента, а под ней расчет коэффициентов.

Таблица 6.7

Расчетная матрица и результаты *

Номер опыта	x_0	x_1	x_2	$x_1 x_2$	Буквенные обозначения строк	y
1	+1	-1	-1	+1	(1)	95
2	+1	+1	-1	-1	a	90
3	+1	-1	+1	-1	b	85
4	+1	+1	+1	+1	ab	82

* См. пример 3.

95	185	352	$b_0 = 88,0$
90	167	-8	$b_1 = -2,0$
85	-5	-18	$b_2 = -4,5$
82	-3	2	$b_{12} = 0,5$

Обратите внимание на то, что порядок коэффициентов в последнем столбце соответствует порядку буквенных обозначений матрицы планирования. Так, строке (1) соответствует b_0 , строке (a) — b_1 и т. д. Порядок буквенных обозначений зависит от порядка опытов, который должен быть фиксированным.

С ростом числа факторов число возможных взаимодействий быстро растет. Мы рассмотрели самый простой случай, когда имелось одно взаимодействие. Обратимся теперь к полному факторному эксперименту 2^3 .

Матрица планирования 2^3 с учетом всех возможных взаимодействий приведена в табл. 6.8.

Таблица 6.8

Полный факторный эксперимент 2^3

Номер опыта	x_0	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$	y
1	+1	-1	-1	-1	+1	-1	-1	+1	y_1
2	+1	+1	-1	-1	-1	-1	+1	+1	y_2
3	+1	-1	+1	-1	-1	+1	-1	+1	y_3
4	+1	+1	+1	-1	+1	+1	+1	+1	y_4
5	+1	-1	-1	+1	+1	-1	+1	-1	y_5
6	+1	+1	-1	+1	-1	+1	-1	-1	y_6
7	+1	-1	+1	+1	-1	-1	+1	-1	y_7
8	+1	+1	+1	+1	+1	-1	-1	-1	y_8

Вы, по-видимому, испытывали затруднения при построении столбца эффекта взаимодействия $x_1 x_2 x_3$. Он получается перемножением всех трех столбцов и называется эффектом взаимодействия второго порядка. Эффект взаимодействия двух факторов называется эффектом взаимодействия первого порядка. Вообще эффект взаимодействия максимального порядка в полном факторном эксперименте имеет порядок, на единицу меньший числа факторов. Довольно часто применяются синонимы: парные эффекты взаимодействия ($x_1 x_2$, $x_2 x_3$, ...), тройные ($x_1 x_2 x_3$, $x_2 x_3 x_4$, ...) и т. д.

Полное число всех возможных эффектов, включая b_0 , линейные эффекты и взаимодействия всех порядков, равно числу опытов полного факторного эксперимента. Чтобы найти число возможных взаимодействий некоторого порядка, можно воспользоваться обычной формулой числа сочетаний

$$C_k^m = \frac{n!}{m! (k - m)!},$$

где k — число факторов, m — число элементов во взаимодействии. Так, для плана 2^4 число парных взаимодействий равно шести

$$C_4^2 = \frac{4!}{2!2!} = 6.$$

Поясним физический смысл эффекта взаимодействия следующим примером. Пусть на некоторый химический процесс влияют два фактора: температура и время реакции.

В области низких температур увеличение времени увеличивает выход продукта. При переходе в область высоких температур эта закономерность нарушается. Здесь, напротив, необходимо уменьшать время реакции. Это и есть проявление эффекта взаимодействия.

Ортогональность матрицы планирования позволяет получить независимые друг от друга оценки коэффициентов. Это означает, что величина любого коэффициента не зависит от того, какие величины имеют другие коэффициенты.

Однако сформулированные выше утверждения справедливы лишь в том случае, если модель включает только линейные эффекты и эффекты взаимодействия. Между тем существенными могут оказаться коэффициенты при квадратах факторов, их кубах и т. п. Так, для случая существенных квадратичных членов в двухфакторном эксперименте модель можно записать так:

$$y = b_0x_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2.$$

Какую информацию о квадратичных членах можно извлечь из полного факторного эксперимента?

Попытка построения вектор-столбцов для x_1^2 и x_2^2 приводит к получению единичных столбцов, совпадающих друг с другом и со столбцом x_0 . Так как эти столбцы неразличимы, то нельзя сказать, за счет чего получилась величина b_0 . Она включает значение свободного члена и вклады квадратичных членов. В этом случае говорят, что имеет место смешанная оценка. Это символически записывается следующим образом:

$$b_0 \rightarrow \beta_0 + \sum_{j=1}^k \beta_{jj},$$

где b_0 — вычисленный нами коэффициент, а греческими буквами, как принято в статистике, обозначены неизвестные истинные значения свободного члена (β_0) и квадратичных коэффициентов (β_{jj}). Если бы мы сделали сколь угодно много опытов, то в пределе получили бы истинные значения коэффициентов. На практике реализуются лишь малые выборки, по которым вычисляются оценки истинных коэффициентов.

По отношению к квадратичной модели для двух факторов получается такая система смешивания:

$$b_0 \rightarrow \beta_0 + \sum_{j=1}^k \beta_{jj}, \quad b_2 \rightarrow \beta_2, \quad b_1 \rightarrow \beta_1, \quad b_{12} \rightarrow \beta_{12}.$$

Следовательно, оценки всех коэффициентов, кроме b_0 , не смешаны.

Число опытов в полном факторном эксперименте превышает число коэффициентов линейной модели, причем тем больше, чем больше факторов. Разность между числом опытов и числом коэффициентов во многих случаях оказывается очень велика, и возникает естественное желание сократить число необходимых опытов. Этим мы и займемся в следующей главе. Но прежде подведем итог сказанному.

6.5. Резюме

Первой серией опытов предшествует этап неформализованных решений, направленных на выбор локальной области факторного пространства. При этом оцениваются границы областей определения факторов, задаваемые либо принципиальными ограничениями, либо технико-экономическими соображениями, либо конкретными условиями проведения процесса. Установление области связано с тщательным анализом априорной информации об изменении параметра оптимизации и о кривизне поверхности отклика.

Локальная область проведения эксперимента выбирается в два этапа: определение основного уровня и интервалов варьирования. Основной (нулевой) уровень — многомерная точка в факторном пространстве, задаваемая комбинацией уровней факторов. Построение плана эксперимента сводится к выбору экспериментальных точек, симметричных относительно основного уровня. При установлении основного уровня приходится рассматривать различные ситуации. Ситуации задаются информацией о наилучших точках и определяют решения.

Следующий этап — выбор интервалов варьирования факторов. Для каждого фактора определяются два уровня, на которых он варьируется в эксперименте. Уровни факторов изображаются двумя точками на координатной оси, симметричными относительно основного уровня. Один из уровней — верхний, другой — нижний. Интервалом варьирования факторов называется некоторое число (свое для каждого фактора), прибавление которого к основному уровню дает верхний, а вычитание — нижний уровень.

Для упрощения записи условий эксперимента и обработки экспериментальных данных масштабы по осям задают так, чтобы верхний уровень соответствовал $+1$, нижний -1 , основной — нулю.

На выбор интервалов варьирования накладываются ограничения снизу (он не может быть меньше ошибки фиксирования уровня фактора) и сверху (верхний или нижний уровни не должны выходить за область определения).

В задачах оптимизации выбирают подобласть, которая давала бы возможность реализовать шаговую процедуру движения к оптимуму. В задачах интерполяции интервал варьирования охватывает всю описываемую область.

При определении интервала варьирования используется информация о точности, с которой фиксируются значения факторов, о кривизне поверхности отклика и о диапазоне изменения параметра оптимизации. Для принятых градаций этих признаков существует 27 различных ситуаций. Низкая точность фиксирования факторов определяет типичное решение — широкий интервал варьирования. Для средней точности характерен выбор среднего

ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

интервала. Высокая точность обычно приводит либо к узкому, либо к среднему интервалам.

Эксперимент, в котором реализуются все возможные сочетания уровней, называется полным факторным экспериментом. Если число уровней равно двум, то это полный факторный эксперимент типа 2^k . Условия эксперимента представляют в виде таблицы — матрицы планирования, где строки соответствуют различным опытам, а столбцы — значениям факторов. Геометрическая интерпретация полных факторных планов: план 2^2 задается координатами вершин квадрата, план 2^3 — координатами вершин куба, при $k > 3$ — координатами вершин гиперкуба.

Полный факторный эксперимент типа 2^k обладает свойствами симметричности, нормировки, ортогональности, ротатабельности (для линейной модели).

Коэффициенты, вычисленные по результатам эксперимента, указывают на силу влияния факторов. Эффект фактора численно равен удвоенному коэффициенту. В тех случаях, когда эффект одного фактора зависит от уровня, на котором находится другой фактор, говорят о наличии эффекта взаимодействия двух факторов. Для его количественной оценки получают столбец произведений этих факторов и обращаются с ним как с вектор-столбцом любого фактора.

Из полного факторного эксперимента нельзя извлечь информацию о квадратичных членах. Вектор-столбцы для квадратичных членов совпадают друг с другом и со столбцом x_0 . Величина свободного члена b_0 включает вклады квадратичных членов, получается смешанная оценка. Оценки остальных коэффициентов не смешаны.

В полном факторном эксперименте разность между числом опытов и числом коэффициентов велика. Возникает проблема уменьшения числа опытов. Этому вопросу посвящена следующая глава.

Литература

1. Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. О принятии решений в неформализованных ситуациях. В сб. «Методологические проблемы кибернетики», Материалы Всесоюзного совещания, т. 2. М., «Наука», 1970.
2. Н. М. Прутков, Ю. В. Грановский, Л. И. Мартыненко и др. Применение статистического метода Бокса — Уилсона для нахождения оптимальных условий разделения смесей р. з. э. при элюировании растворами иминодиуксусной кислоты. — Ж. неорганич. хим., 1970, 15, № 2.
3. W. Cochran, G. M. Cox. Experimental Designs, 2 ed. N. Y. — J. Wiley, 1960.
4. Д. Финни. Введение в теорию планирования эксперимента. М., «Наука», 1970.
5. Ч. Хикс. Основные принципы планирования эксперимента. М., «Мир», 1967.

Не числом, а умением.

Поговорка

Количество опытов в полном факторном эксперименте значительно превосходит число определяемых коэффициентов линейной модели. Другими словами, полный факторный эксперимент обладает большой избыточностью опытов. Было бы заманчивым сократить их число за счет той информации, которая не очень существенна при построении линейных моделей. При этом нужно стремиться к тому, чтобы матрица планирования не лишилась своих оптимальных свойств. Сделать это не так просто, но все же возможно. Итак начнем поиск путей минимизации числа опытов [1—3].

7.1. Минимизация числа опытов

Начнем с самого простого — полного факторного эксперимента 2^2 . Напишем еще раз эту хорошо нам известную матрицу (табл. 7.1).

Таблица 7.1

Полный факторный эксперимент 2^2

Номер опыта	x_0	x_1	x_2	(x_3) x_1x_2	y	Номер опыта	x_0	x_1	x_2	(x_3) x_1x_2	y
1	+	—	—	+	y_1	3	+	—	+	—	y_3
2	+	+	—	—	y_2	4	+	+	+	+	y_4

Пользуясь таким планированием, можно вычислить четыре коэффициента и представить результаты эксперимента в виде неполного квадратного уравнения

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2.$$

Если имеются основания считать, что в выбранных интервалах варьирования процесс может быть описан линейной моделью, то достаточно определить три коэффициента: b_0 , b_1 и b_2 . Остается

одна степень свободы. Употребим ее для минимизации числа опытов. При линейном приближении $b_{12} \rightarrow 0$ и вектор-столбец $x_1 x_2$ можно использовать для нового фактора x_3 . Поставим этот фактор в скобках над взаимодействием $x_1 x_2$ и посмотрим, каковы будут оценки коэффициентов. Здесь уже не будет тех отдельных оценок, которые мы имели в полном факторном эксперименте 2^k . Оценки смешаются следующим образом:

$$b_1 \rightarrow \beta_1 + \beta_{23}; \quad b_2 \rightarrow \beta_2 + \beta_{13}; \quad b_3 \rightarrow \beta_3 + \beta_{12}.$$

Но нас это не должно огорчать. Ведь мы постулируем линейную модель, и, следовательно, все парные взаимодействия незначимы. Главное, мы нашли средство минимизировать число опытов: вместо восьми опытов для изучения трех факторов оказывается можно поставить четыре! При этом матрица планирования не теряет своих оптимальных свойств (ортогональность, ротатабельность и т. п.), в чем вы можете самостоятельно убедиться. Найденное правило можно сформулировать так: чтобы сократить число опытов, нужно новому фактору присвоить вектор-столбец матрицы, принадлежащий взаимодействию, которым можно пренебречь. Тогда значение нового фактора в условиях опытов определяется знаками этого столбца.

Посмотрите, пожалуйста, на три матрицы, приведенные ниже. Эти матрицы предлагаются взамен полного факторного эксперимента 2^3 , требующего, как вы знаете, восьми опытов. Каким бы из них вы воспользовались?

Матрица № 1						Матрица № 2					
№	x_0	x_1	x_2	x_3	y	№	x_0	x_1	x_2	x_3	y
1	+	—	—	+	y_1	1	+	—	—	+	y_1
2	+	+	+	—	y_2	2	+	+	+	+	y_2
3	+	—	+	—	y_3	3	+	—	+	—	y_3
4	+	+	—	+	y_4						

Матрица № 3					
№	x_0	x_1	x_2	x_3	y
1	+	—	—	+	y_1
2	+	+	+	+	y_2
3	+	—	+	—	y_3
4	+	+	—	—	y_4

Проверим свойства матрицы № 1. Каждый вектор-столбец матрицы, кроме первого, содержит равное число $+1$ и -1 . Это означает, что выполняется условие:

$$\sum_{i=1}^4 x_{ji} = 0.$$

Теперь перемножим каждую пару вектор-столбцов и посмотрим, будет ли сумма произведений равна 0. К сожалению,

$$\sum_{i=1}^4 x_{2i} x_{3i} = -4,$$

т. е. совершена какая-то ошибка в выборе матрицы. Постараемся ее найти. Вектор-столбцы для x_1 и x_2 не вызывают сомнения. Ведь эта часть матрицы — полный факторный эксперимент 2^2 . А как построен вектор-столбец для x_3 ? Элементы этого столбца обратны по знаку элементам соседнего столбца x_2 . Два этих столбца оказались взаимосвязанными: $x_3 = -x_2$. При этом $b_3 \rightarrow \beta_3 - \beta_2$ и $b_2 \rightarrow \beta_2 - \beta_3$. В таком планировании не могут быть отдельно оценены основные эффекты. Значит, мы потеряли информацию о двух линейных коэффициентах нашей модели. Таким планированием воспользоваться невозможно.

Матрица № 2 содержит всего три опыта. Три опыта недостаточны для оценки четырех коэффициентов: b_0 , b_1 , b_2 и b_3 . Кроме того, ни одно из свойств, присущих полному факторному эксперименту, здесь не выполняется, за исключением нормировки. Матрица № 3 сохраняет все свойства полного факторного эксперимента. Она дает возможность оценить свободный член b_0 и три коэффициента при линейных членах, потому что для x_3 использован вектор-столбец $x_1 x_2$ полного факторного эксперимента 2^2 .

Если мы в дополнение к столбцам матрицы № 3 вычислим еще столбцы для произведений $x_1 x_3$ и $x_2 x_3$, то увидим, что элементы столбца $x_1 x_3$ совпадут с элементами столбца x_2 , а элементы столбца $x_2 x_3$ — с элементами столбца x_1 . Найденные нами коэффициенты будут оценками для совместных эффектов

$$b_1 \rightarrow \beta_1 + \beta_{23}; \quad b_2 \rightarrow \beta_2 + \beta_{13}; \quad b_3 \rightarrow \beta_3 + \beta_{12}.$$

Такое планирование нас вполне устраивает. Мы смешали эффекты взаимодействия с основными эффектами. (Но все основные эффекты оцениваются раздельно друг от друга!) Так как постулируется линейная модель, то предполагается, что эффекты взаимодействия близки к нулю, и поэтому $b_1 \approx \beta_1$, $b_2 \approx \beta_2$, $b_3 \approx \beta_3$.

Мы рассмотрели самый простой случай: матрицу из четырех опытов для трехфакторного планирования. С увеличением числа факторов вопрос о минимизации числа опытов превращается в довольно сложную задачу. Рассмотрим ее детально. При этом нам не обойтись без новых определений и понятий.

7.2. Дробная реплика

Поставив четыре опыта для оценки влияния трех факторов, мы воспользовались половиной полного факторного эксперимента 2^3 , или «полуреplikой». Если бы мы x_3 приравняли к $-x_1 x_2$, то получили бы вторую половину матрицы 2^3 . В этом случае:

$b_1 \rightarrow \beta_1 - \beta_{23}$; $b_2 \rightarrow \beta_2 - \beta_{13}$, $b_3 \rightarrow \beta_3 - \beta_{12}$. При реализации обеих полуреплик можно получить отдельные оценки для линейных эффектов и эффектов взаимодействия, как и в полном факторном эксперименте 2^3 .

Объединение этих двух полуреплик и есть полный факторный эксперимент 2^3 .

Матрица из восьми опытов для четырехфакторного планирования будет полурепликой от полного факторного эксперимента 2^4 , а для пятифакторного планирования — четверть-репликой от 2^5 . В последнем случае два линейных эффекта приравниваются к эффектам взаимодействия. Для обозначения дробных реплик, в которых p линейных эффектов приравнены к эффектам взаимодействия, удобно пользоваться условным обозначением 2^{k-p} . Так, полуреплика от 2^6 запишется в виде 2^{6-1} , а четверть-реплика от 2^5 — в виде 2^{5-2} .

Таблица 7.2

Условия обозначения дробных реплик и число опытов

Число факторов	Дробная реплика	Условное обозначение	Число опытов	
			для дробной реплики	для полного факторного эксперимента
3	1/2-реплика от 2^3	2^{3-1}	4	8
4	1/2-реплика от 2^4	2^{4-1}	8	16
5	1/4-реплика от 2^5	2^{5-2}	8	32
6	1/8-реплика от 2^6	2^{6-3}	8	64
7	1/16-реплика от 2^7	2^{7-4}	8	128
5	1/2-реплика от 2^5	2^{5-1}	16	32
6	1/4-реплика от 2^6	2^{6-2}	16	64
7	1/8-реплика от 2^7	2^{7-3}	16	128
8	1/16-реплика от 2^8	2^{8-4}	16	256
9	1/32-реплика от 2^9	2^{9-5}	16	512
10	1/64-реплика от 2^{10}	2^{10-6}	16	1024
11	1/128-реплика от 2^{11}	2^{11-7}	16	2048
12	1/256-реплика от 2^{12}	2^{12-8}	16	4096
13	1/512-реплика от 2^{13}	2^{13-9}	16	8192
14	1/1024-реплика от 2^{14}	2^{14-10}	16	16384
15	1/2048-реплика от 2^{15}	2^{15-11}	16	32768

Условные обозначения дробных реплик и количество опытов приведены в табл. 7.2.

7.3. Выбор полуреплик. Генерирующие соотношения и определяющие контрасты

При построении полуреплики 2^{3-1} существует всего две возможности: приравнять x_3 к $+x_1x_2$ или к $-x_1x_2$. Поэтому есть только две полуреплики 2^{3-1} (табл. 7.3).

Таблица 7.3

Две полуреплики 2^{3-1}

Номер опыта	I. $x_3 = x_1x_2$				Номер опыта	II. $x_3 = -x_1x_2$			
	x_1	x_2	x_3	$x_1x_2x_3$		x_1	x_2	x_3	$x_1x_2x_3$
1	+	+	+	+	1	+	+	-	-
2	-	-	+	+	2	-	-	-	-
3	+	-	-	+	3	+	-	+	-
4	-	+	-	+	4	-	+	+	-

Для произведения трех столбцов матрицы I выполняется соотношение: $+1 = x_1x_2x_3$, а матрицы II: $-1 = x_1x_2x_3$. Вы видите, что все знаки столбцов произведений одинаковы и в первом случае равны плюс единице, а во втором — минус единице.

Символическое обозначение произведения столбцов, равного $+1$ или -1 , называется *определяющим контрастом*. Контраст помогает определять смешанные эффекты. Для того чтобы определить, какой эффект смешан с данным, нужно помножить обе части определяющего контраста на столбец, соответствующий данному эффекту. Так, если $1 = x_1x_2x_3$, то для x_1 имеем

$$x_1 = x_1^2x_2x_3 = x_2x_3,$$

так как всегда $x_i^2 = 1$.

Для x_2 находим

$$x_2 = x_1x_2^2x_3 = x_1x_3,$$

для x_3

$$x_3 = x_1x_2x_3^2 = x_1x_2.$$

Это значит, что коэффициенты линейного уравнения будут оценками

$$b_1 \rightarrow \beta_1 + \beta_{23}, \quad b_2 \rightarrow \beta_2 + \beta_{13}, \quad b_3 \rightarrow \beta_3 + \beta_{12}.$$

Соотношение, показывающее, с каким из эффектов смешан данный эффект, называется *генерирующим соотношением*.

Полуреплики, в которых основные эффекты смешаны с двухфакторными взаимодействиями, носят название планов с разрешающей способностью III (по наибольшему числу факторов в определяющем контрасте). Такие планы принято обозначать: 2^{k-1}_{III} .

При выборе полуреплики 2^{4-1} возможно восемь решений:

- | | | |
|----------------------|----------------------|-------------------------|
| 1) $x_4 = x_1x_2$, | 4) $x_4 = -x_2x_3$, | 7) $x_4 = x_1x_2x_3$, |
| 2) $x_4 = -x_1x_2$, | 5) $x_4 = x_1x_3$, | 8) $x_4 = -x_1x_2x_3$. |
| 3) $x_4 = x_2x_3$, | 6) $x_4 = -x_1x_3$, | |

Разрешающая способность этих полуреplik различна. Так, реплики 1—6 имеют по три фактора в определяющем контрасте, а 7—8 по четыре. Реплики 7 и 8 имеют максимальную разрешающую способность и называются главными. Разрешающая способность задается системой смешивания данной реплики. Она будет максимальной, если линейные эффекты смешаны с эффектами взаимодействия наибольшего возможного порядка.

При отсутствии априорной информации об эффектах взаимодействия экспериментатор стремится выбрать реплику с наибольшей разрешающей способностью, так как тройные взаимодействия обычно менее важны, чем парные. Если существует информация об эффектах взаимодействия, то она должна использоваться при выборе реплики.

Реплики, в которых нет ни одного главного эффекта, смешанного с другим главным эффектом или парным взаимодействием, а все парные взаимодействия смешаны друг с другом, носят название планов с разрешающей способностью IV (по наибольшему числу факторов в определяющем контрасте). Они имеют обозначение $2\frac{1}{4}$. Полуреплика, заданная определяющим контрастом $1 = +x_1x_2x_3x_4$, имеет только четные комбинации букв в каждой строке. Ее можно записать следующим образом, считая строку (1) четной:

(1), ad , bd , ab , ac , cd , bc , $abcd$.

А полуреплика, заданная $1 = -x_1x_2x_3x_4$, имеет только нечетные комбинации

a , b , c , d , abd , acd , abc , bcd .

Такие полуреплики называют главными полурепликами, так как они обладают наибольшей разрешающей способностью.

Пусть выбраны полуреплики, заданные определяющими контрастами $1 = x_1x_2x_3x_4$ и $1 = -x_1x_2x_3x_4$. Совместные оценки здесь определяются соотношениями:

$$\begin{array}{lll} x_1 = x_2x_3x_4, & x_1x_3 = x_2x_4, & x_3 = -x_1x_2x_4, \\ x_2 = x_1x_3x_4, & x_1x_4 = x_2x_3, & x_4 = -x_1x_2x_3, \\ x_3 = x_1x_2x_4, & & x_1x_2 = -x_3x_4, \\ x_4 = x_1x_2x_3, & x_1 = -x_2x_3x_4, & x_1x_3 = -x_2x_4, \\ x_1x_2 = x_3x_4, & x_2 = -x_1x_3x_4, & x_1x_4 = -x_2x_3. \end{array}$$

Такой тип смешивания даст возможность оценивать линейные эффекты совместно с эффектами взаимодействий второго порядка, а взаимодействия первого порядка — совместно друг с другом.

Если полуреплики заданы генерирующими соотношениями $x_4 = x_1x_2$ и $x_4 = -x_1x_2$, то в этом случае определяющими контра-

тами являются $1 = x_1x_2x_4$ и $1 = -x_1x_2x_4$, следовательно, мы получаем планы с разрешающей способностью III и некоторые основные эффекты смешиваем с парными взаимодействиями:

$$\begin{array}{lll} x_1 = x_2x_4, & x_2x_3 = x_1x_3x_4, & x_3 = -x_1x_2x_3x_4, \\ x_2 = x_1x_4, & x_3x_4 = x_1x_2x_3, & x_4 = -x_1x_2, \\ x_3 = x_1x_2x_3x_4, & & x_1x_3 = -x_2x_3x_4, \\ x_4 = x_1x_2, & x_1 = -x_2x_4, & x_2x_3 = -x_1x_3x_4, \\ x_1x_3 = x_2x_3x_4, & x_2 = -x_1x_4, & x_3x_4 = -x_1x_2x_3. \end{array}$$

Разрешающая способность этих полуреplik ниже, чем у планов с разрешающей способностью IV, с помощью которых линейные эффекты определяются независимо от парных взаимодействий.

Эти полуреплики имеют в каждой строке как четные, так и нечетные комбинации букв. Такие полуреплики не являются главными. Разумен выбор такой полуреплики, если имеется априорная информация о большей значимости тройных взаимодействий по сравнению с парными или о незначимости трех парных взаимодействий x_2x_4 , x_1x_4 , x_1x_2 .

Как видите, выбор дробной реплики требует много терпения и труда. Но другого пути нет. Применяя дробное планирование, нужно точно знать систему смешивания, четко представлять, какую информацию приходится терять.

Теперь рассмотрим пример полуреплики $2\frac{1}{4}$.

Пример 1. Этот пример относится к планированию эксперимента для отыскания оптимальных условий получения нового полимерного серусодержащего антиоксиданта, синтезированного превращением высокомолекулярного полистирола с серой [4]. Задача состояла в получении стабилизатора, введение которого в изотактический полипропилен увеличивало бы период индукции, не ухудшая физико-механических свойств полимера. В качестве факторов рассматривались переменные, показанные в табл. 7.4.

Матрица планирования представляла собой полуреплику от 2^4 , заданную генерирующим соотношением $x_4 = x_1x_2x_3$. Определяющим контрастом является $1 = x_1x_2x_3x_4$. Умножая определяющий контраст последовательно на x_1 , x_2 , x_3 и x_4 , определяем совместно оценки линейных эффектов и взаимодействий

$$\begin{array}{lll} b_1 \rightarrow \beta_1 + \beta_{234}, & b_3 \rightarrow \beta_3 + \beta_{124}, & b_{12} \rightarrow \beta_{12} + \beta_{34}, \\ b_2 \rightarrow \beta_2 + \beta_{134}, & b_4 \rightarrow \beta_4 + \beta_{123}, & b_{13} \rightarrow \beta_{13} + \beta_{24}, \\ & & b_{23} \rightarrow \beta_{23} + \beta_{14}. \end{array}$$

Матрица планирования, результаты эксперимента и коэффициенты регрессии показаны в табл. 7.5.

Анализ результатов и поиск оптимальных условий приводятся в последующих главах.

Таблица 7.4

Уровни факторов и интервалы варьирования

Факторы	Уровни факторов			Интервал варьирования
	-1	0	+1	
x_1 — температура реакционной среды, °C	200	220	240	20
x_2 — дозировка серы в исходной смеси, вес. ч.	3	6	9	3
x_3 — время реакции, мин.	40	100	160	60
x_4 — дозировка антиоксиданта в полипропилене, вес. ч.	1	2	3	1

Таблица 7.5

Матрица планирования 2^{4-1}

Номер опыта	x_0	x_1	x_2	x_3	x_4	$x_1x_2 = x_3x_4$	$x_1x_3 = x_2x_4$	$x_2x_3 = x_1x_4$	ν
1	+	+	+	-	-	+	-	-	10
2	+	-	-	-	-	+	+	+	9
3	+	+	-	-	+	-	+	+	15
4	+	-	+	-	+	-	+	-	25
5	+	+	+	+	+	+	+	+	26
6	+	-	-	+	+	+	-	-	14
7	+	+	-	+	-	-	+	-	5
8	+	-	+	+	-	-	-	+	20
Коэффициенты b_i	15,50	-1,50	4,75	0,75	4,50	-0,75	0,75	2,0	

Поговорим теперь о полуреплике 2^{5-1} .

При выборе полуреплики 2^{5-1} в распоряжении экспериментатора имеется множество вариантов. Так, x_5 можно приравнять к одному из шести парных взаимодействий. В этом случае получим полуреплику с разрешающей способностью III. Очевидно, это будет не лучший выбор полуреплики. Далее, x_5 можно приравнять к одному из четырех тройных взаимодействий. Тогда получим план с разрешающей способностью IV, и все линейные эффекты будут смешаны с тройными взаимодействиями. И, наконец, полуреплика может быть задана генерирующими соотношениями $x_5 = x_1x_2x_3x_4$ или $x_5 = -x_1x_2x_3x_4$. Определяющими контрастами в этом случае будут $1 = x_1x_2x_3x_4x_5$ и $1 = -x_1x_2x_3x_4x_5$. Такие реплики носят название планов с разрешающей способностью V и обозначаются 2^{5-1}_V .

Пусть имеется пять факторов и для них нужно выбрать полуреплику с наибольшей разрешающей способностью.

Для полуреплики, заданной генерирующим соотношением $x_5 = x_1x_2x_3x_4$, $1 = x_1x_2x_3x_4x_5$, следовательно,

$$\begin{aligned} b_1 &\rightarrow \beta_1 + \beta_{345}, & b_{14} &\rightarrow \beta_{14} + \beta_{35}, & b_{35} &\rightarrow \beta_{35} + \beta_{14}, \\ b_2 &\rightarrow \beta_2 + \beta_{12345}, & b_{15} &\rightarrow \beta_{15} + \beta_{34}, & b_{45} &\rightarrow \beta_{45} + \beta_{13}, \\ b_3 &\rightarrow \beta_3 + \beta_{145}, & b_{23} &\rightarrow \beta_{23} + \beta_{1245}, & b_{123} &\rightarrow \beta_{123} + \beta_{245}, \\ b_4 &\rightarrow \beta_4 + \beta_{135}, & b_{24} &\rightarrow \beta_{24} + \beta_{1235}, & b_{124} &\rightarrow \beta_{124} + \beta_{235}, \\ b_5 &\rightarrow \beta_5 + \beta_{134}, & b_{25} &\rightarrow \beta_{25} + \beta_{1234}, & b_{125} &\rightarrow \beta_{125} + \beta_{234}, \\ b_{12} &\rightarrow \beta_{12} + \beta_{2345}, & b_{34} &\rightarrow \beta_{34} + \beta_{15}, & b_{134} &\rightarrow \beta_{134} + \beta_5 \text{ и т. д.} \\ b_{13} &\rightarrow \beta_{13} + \beta_{45}, \end{aligned}$$

Смешивание основных эффектов с тройными взаимодействиями, когда существуют эффекты взаимодействия более высокого порядка, нельзя признать наилучшим, если нет специальных соображений.

Если выбрана полуреплика, заданная генерирующим соотношением $x_5 = x_1x_2x_3x_4$ и, следовательно, определяющим контрастом $1 = x_1x_2x_3x_4x_5$, то коэффициенты определяют такие смешанные оценки:

$$\begin{aligned} b_1 &\rightarrow \beta_1 + \beta_{2345}, & b_5 &\rightarrow \beta_5 + \beta_{1234}, & b_{14} &\rightarrow \beta_{14} + \beta_{235}, \\ b_2 &\rightarrow \beta_2 + \beta_{1345}, & b_{12} &\rightarrow \beta_{12} + \beta_{345}, & b_{15} &\rightarrow \beta_{15} + \beta_{234}, \\ b_3 &\rightarrow \beta_3 + \beta_{1245}, & b_{13} &\rightarrow \beta_{13} + \beta_{245}, & b_{23} &\rightarrow \beta_{23} + \beta_{145} \text{ и т. д.} \\ b_4 &\rightarrow \beta_4 + \beta_{1235}, \end{aligned}$$

Получили полуреплику с разрешающей способностью V. В таких планах линейные эффекты смешаны со взаимодействиями третьего порядка, а взаимодействия первого порядка — с взаимодействиями второго порядка. Эта полуреплика имеет преимущества по сравнению с ранее рассмотренной репликой.

Возможны двадцать два решения при выборе плана 2^{5-1} :

- 1) $x_5 = x_1x_2$, 9) $x_5 = x_3x_4$, 16) $x_5 = -x_1x_3x_4$,
- 2) $x_5 = -x_1x_2$, 10) $x_5 = -x_2x_4$, 17) $x_5 = x_1x_2x_4$,
- 3) $x_5 = x_1x_3$, 11) $x_5 = x_3x_4$, 18) $x_5 = -x_1x_2x_4$,
- 4) $x_5 = -x_1x_3$, 12) $x_5 = -x_3x_4$, 19) $x_5 = x_2x_3x_4$,
- 5) $x_5 = x_1x_4$, 13) $x_5 = x_1x_2x_3$, 20) $x_5 = -x_2x_3x_4$,
- 6) $x_5 = -x_1x_4$, 14) $x_5 = -x_1x_2x_3$, 21) $x_5 = x_1x_2x_3x_4$,
- 7) $x_5 = x_2x_3$, 15) $x_5 = x_1x_3x_4$, 22) $x_5 = -x_1x_2x_3x_4$,
- 8) $x_5 = -x_2x_3$,

Мы не станем рассматривать выбор полуреплик 2^{6-1} , 2^{7-1} и т. д. Такими полурепликами редко пользуются на практике. Ведь полуреплика 2^{6-1} требует 32 опыта, а для эксперимента

тора выгодны планы 2^{6-2} или 2^{6-3} , требующие соответственно 16 и 8 опытов. Поэтому с ростом числа факторов возрастает дробность применяемых реплик.

Заметим, что при построении главных полуреplik в определяющий контраст надо включать наибольшее число факторов.

Построение 1/4-реplik мы рассмотрим в следующем разделе.

7.4. Выбор 1/4-реплик.

Обобщающий определяющий контраст

Дорогой читатель, приступая к этому параграфу, вам придется запастись еще большим терпением. Но вы будете вознаграждены, ибо с увеличением дробности не только возрастают ваши усилия, но и уменьшается число опытов. А ради этого стоит потрудиться и тщательно разобраться в выборе 1/4-реplik.

При исследовании влияния пяти факторов можно поставить не 16 опытов, как в предыдущем примере, а только 8, т. е. воспользоваться репликой 2^{5-2} . Здесь возможны двенадцать решений, если x_4 приравнять парному взаимодействию, а x_5 — тройному:

- 1) $x_4 = x_1x_2$, $x_5 = x_1x_2x_3$,
- 2) $x_4 = x_1x_2$, $x_5 = -x_1x_2x_3$,
- 3) $x_4 = -x_1x_2$, $x_5 = x_1x_2x_3$,
- 4) $x_4 = -x_1x_2$, $x_5 = -x_1x_2x_3$,
- 5) $x_4 = x_1x_3$, $x_5 = x_1x_2x_3$,
- 6) $x_4 = x_1x_3$, $x_5 = -x_1x_2x_3$,
- 7) $x_4 = -x_1x_3$, $x_5 = x_1x_2x_3$,
- 8) $x_4 = -x_1x_3$, $x_5 = -x_1x_2x_3$,
- 9) $x_4 = x_2x_3$, $x_5 = x_1x_2x_3$,
- 10) $x_4 = x_2x_3$, $x_5 = -x_1x_2x_3$,
- 11) $x_4 = -x_2x_3$, $x_5 = x_1x_2x_3$,
- 12) $x_4 = -x_2x_3$, $x_5 = -x_1x_2x_3$.

Допустим, выбран пятый вариант: $x_4 = x_1x_3$ и $x_5 = x_1x_2x_3$. Тогда определяющими контрастами являются: $1 = x_1x_3x_4$ и $1 = x_1x_2x_3x_5$.

Если перемножить эти определяющие контрасты, то получится третье соотношение, задающее элементы столбца $1 = x_2x_4x_5$. Чтобы полностью охарактеризовать разрешающую способность реплики, необходимо записать обобщающий определяющий контраст $1 = x_1x_3x_4 = x_2x_4x_5 = x_1x_2x_3x_5$.

Система смешивания определяется умножением обобщающего определяющего контраста последовательно на x_1 , x_2 , x_3 и т. д.

$$\begin{aligned} x_1 &= x_3x_4 = x_1x_2x_4x_5 = x_2x_3x_5, \\ x_2 &= x_1x_2x_3x_4 = x_4x_5 = x_1x_3x_5, \end{aligned}$$

$$\begin{aligned} x_3 &= x_1x_4 = x_2x_3x_4x_5 = x_1x_2x_5, \\ x_4 &= x_1x_3 = x_2x_5 = x_1x_2x_3x_4x_5, \\ x_5 &= x_1x_3x_4x_5 = x_2x_4 = x_1x_2x_3, \\ x_1x_2 &= x_2x_3x_4 = x_1x_4x_5 = x_3x_5, \\ x_1x_3 &= x_3x_4x_5 = x_1x_2x_4 = x_2x_3. \end{aligned}$$

Получается довольно сложная система смешивания линейных эффектов с эффектами взаимодействия первого, второго, третьего и четвертого порядков. Если, например, коэффициенты $b_{12} \rightarrow \beta_{12} + \beta_{234} + \beta_{145} + \beta_{35}$ и $b_{15} \rightarrow \beta_{15} + \beta_{345} + \beta_{124} + \beta_{23}$ отличаются от нуля, то возникают сомнения, можно ли пренебрегать другими парными взаимодействиями, с которыми смешаны линейные эффекты. Тогда следует поставить вторую серию опытов, выбрав нужным образом другую 1/4-реплику.

При этом можно воспользоваться методом «перевала». Смысл этого метода заключается в том, что вторая четверть-реплика получается из первой путем изменения всех знаков матрицы на обратные. Тогда в обобщающем определяющем контрасте тройные произведения имеют знак, противоположный их знаку в первой четверть-реплике. Тройные произведения определяют парные взаимодействия в совместных оценках для линейных эффектов. Усредняя результаты обеих четверть-реplik, можно получить линейные эффекты, не смешанные с парными взаимодействиями.

Пример 2. Для дополнения 1/4-реплики, заданной определяющими контрастами $1 = x_1x_3x_4$ и $1 = x_1x_2x_3x_5$, можно использовать реплику, заданную генерирующими соотношениями $x_4 = -x_1x_2$ и $x_5 = -x_1x_2x_3$; реплику, заданную генерирующими соотношениями $x_4 = -x_1x_3$ и $x_5 = x_1x_2x_3$; реплику, заданную генерирующими соотношениями $x_4 = x_1x_3$ и $x_5 = -x_1x_2x_3$.

В случае $x_4 = -x_1x_2$ и $x_5 = -x_1x_2x_3$ обобщающий определяющий контраст запишется в виде $1 = -x_1x_2x_4 = -x_1x_2x_3x_5 = +x_3x_4x_5$.

Посмотрим, какой будет система смешивания

$$\begin{aligned} x_1 &= -x_2x_4 = -x_2x_3x_5 = +x_1x_3x_4x_5, \\ x_2 &= -x_1x_4 = -x_1x_3x_5 = +x_2x_3x_4x_5, \\ x_3 &= -x_1x_2x_3x_4 = -x_1x_2x_5 = +x_4x_5, \\ x_4 &= -x_1x_2 = -x_1x_2x_3x_4x_5 = +x_3x_5, \\ x_5 &= -x_1x_2x_4x_5 = -x_1x_2x_3 = +x_3x_4. \end{aligned}$$

Сложим две четверть-реплики

$$\begin{aligned} x_1 &= -x_2x_4 = x_3x_4 = +x_1x_3x_4x_5 = x_1x_2x_4x_5, \\ x_2 &= -x_1x_4 = x_4x_5 = +x_2x_3x_4x_5 = x_1x_2x_3x_4 \text{ и т. д.} \end{aligned}$$

Ясно, что дополнение первой 1/4-реплики второй не привело к улучшению ситуации, а, напротив, осложнило ее, так как линейные эффекты смешиваются с двумя парными взаимодействиями и уничтожаются тройные взаимодействия.

Если исследовать 1/4-реплику, заданную генерирующими соотношениями $x_4 = -x_1x_3$ и $x_5 = x_1x_2x_3$, то обобщающим контрастом в этом случае явятся $1 = -x_1x_3x_4 = x_1x_2x_3x_5 = -x_2x_4x_5$.

Система смешивания будет

$$\begin{aligned}x_1 &= -x_3x_4 = x_2x_3x_5 = -x_1x_2x_4x_5, \\x_2 &= -x_1x_2x_3x_4 = x_1x_3x_5 = -x_4x_5, \\x_3 &= -x_1x_4 = x_1x_2x_5 = -x_2x_3x_4x_5, \\x_4 &= -x_1x_3 = x_1x_2x_3x_4x_5 = -x_2x_5, \\x_5 &= -x_1x_3x_4x_5 = x_1x_2x_3 = -x_3x_4, \\x_1x_2 &= -x_2x_3x_4 = x_3x_5 = -x_1x_4x_5, \\x_1x_5 &= -x_3x_4x_5 = x_2x_3 = -x_1x_2x_4.\end{aligned}$$

При сложении двух 1/4-реплик получается следующая картина:

$$\begin{aligned}b_1 &\rightarrow \beta_1 + \beta_{235}, & b_4 &\rightarrow \beta_4 + \beta_{12345}, & b_{12} &\rightarrow \beta_{12} + \beta_{35}, \\b_2 &\rightarrow \beta_2 + \beta_{135}, & b_5 &\rightarrow \beta_5 + \beta_{123}, & b_{15} &\rightarrow \beta_{15} + \beta_{23} \text{ и т. д.} \\b_3 &\rightarrow \beta_3 + \beta_{125},\end{aligned}$$

Мы получили план с разрешающей способностью IV и освободили линейные эффекты от смешивания с парными взаимодействиями.

В случае 1/4-реплики, заданной генерирующими соотношениями $x_4 = -x_1x_3$ и $x_5 = x_1x_2x_3$ и обобщающим определяющим контрастом $1 = x_1x_3x_4 = -x_1x_2x_3x_5 = -x_2x_4x_5$, получается следующая система смешивания:

$$\begin{aligned}x_1 &= x_3x_4 = -x_2x_3x_5 = -x_1x_2x_4x_5, \\x_2 &= x_1x_2x_3x_4 = -x_1x_3x_5 = -x_4x_5, \\x_3 &= x_1x_4 = -x_1x_2x_5 = -x_2x_3x_4x_5, \\x_4 &= x_1x_3 = -x_1x_2x_3x_4x_5 = -x_2x_5, \\x_5 &= x_1x_3x_4x_5 = -x_1x_2x_3 = -x_2x_4, \\x_1x_2 &= x_2x_3x_4 = -x_3x_5 = -x_1x_4x_5, \\x_1x_5 &= x_3x_4x_5 = -x_2x_3 = -x_1x_2x_4.\end{aligned}$$

При сложении двух 1/4-реплик получаем

$$\begin{aligned}b_1 &\rightarrow \beta_1 + \beta_{34}, & b_4 &\rightarrow \beta_4 + \beta_{13}, & b_{12} &\rightarrow \beta_{12} + \beta_{234}, \\b_2 &\rightarrow \beta_2 + \beta_{1234}, & b_5 &\rightarrow \beta_5 + \beta_{1345}, & b_{15} &\rightarrow \beta_{15} + \beta_{345} \text{ и т. д.} \\b_3 &\rightarrow \beta_3 + \beta_{14},\end{aligned}$$

Итак, линейные эффекты освобождаются от совместных оценок с тройными эффектами взаимодействий и в некоторых случаях — с парными. Такое планирование целесообразно применить в том случае, если ставится задача освободить все линейные эффекты от взаимодействий второго порядка и два линейных эффекта (в данном случае b_2 и b_5) — от взаимодействий первого порядка.

Таким образом, для дополнения 1/4-реплики до 1/2-реплики, если есть подозрения, что эффекты взаимодействия первого порядка отличаются от нуля, нужно взять вторую четверть-реплику с обобщающим контрастом, в котором два тройных произведения имеют отрицательный знак, так как тройные произведения опре-

деляют парные взаимодействия в совместных оценках для линейных эффектов.

Однако можно представить себе и такой случай, когда целесообразно освободить линейные эффекты от эффектов взаимодействий второго порядка и только часть из линейных эффектов (например, b_2 и b_5) от парных взаимодействий. Тогда нужно выбрать 1/4-реплику таким образом, чтобы в обобщающем определяющем контрасте произведение четырех членов имело отрицательный знак, так как это произведение определяет тройные взаимодействия в совместных оценках для линейных эффектов.

Теперь рассмотрим пример реплики 2^{5-2} .

Пример 3. Применение плана 2^{5-2} относится к оптимизации процесса получения одного производного пиперазина по схеме В. Было решено при планировании эксперимента варьировать пятью факторами, представленными в табл. 7.6.

Параметром оптимизации служил выход реакции в процентах.

Таблица 7.6

Уровни факторов и интервалы варьирования

Факторы	Уровни факторов			Интервал варьирования
	-1	0	+1	
\tilde{x}_1 — соотношение между NaOH и веществом a	1 : 1	1,25 : 1	1,5 : 1	0,25*
\tilde{x}_2 — соотношение между веществом c и веществом a	1 : 1	1,25 : 1	1,5 : 1	0,25*
\tilde{x}_3 — время выдержки, час	3	4	5	1
\tilde{x}_4 — температура, °C	20	25	30	5
\tilde{x}_5 — время прилива вещества, мин	20	40	60	20

* Этот интервал относится к первому члену пропорции

В табл. 7.7 приведена матрица планирования эксперимента.

В данном случае при планировании использована 1/4-реплика от полного факторного эксперимента 2^5 . При этом ставится 8 опытов вместо 32. Матрица задана генерирующими соотношениями

$$x_4 = x_1x_2x_3, \quad x_5 = -x_1x_2 \text{ и имеет обобщающий контраст}$$

$$1 = x_1x_2x_3x_4 = -x_1x_2x_5 = -x_3x_4x_5.$$

Выбрано генерирующее соотношение $x_5 = -x_1x_2$, поскольку взаимодействия x_1x_3 и x_2x_3 предполагались существенными.

Совместные оценки такой 1/4-реплики

$$\begin{aligned}b_1 &\rightarrow \beta_1 - \beta_{25} + \beta_{234} - \beta_{1345}, & b_5 &\rightarrow \beta_5 - \beta_{12} - \beta_{34} + \beta_{12345}, \\b_2 &\rightarrow \beta_2 - \beta_{15} + \beta_{134} - \beta_{2345}, & b_{13} &\rightarrow \beta_{13} + \beta_{24} - \beta_{235} - \beta_{145}, \\b_3 &\rightarrow \beta_3 - \beta_{45} + \beta_{124} - \beta_{1235}, & b_{14} &\rightarrow \beta_{14} + \beta_{23} - \beta_{245} - \beta_{135}, \\b_4 &\rightarrow \beta_4 - \beta_{35} + \beta_{125} - \beta_{1245},\end{aligned}$$

Таблица 7.7

Матрица планирования 2^{6-2}

Номер опыта	x_0	x_1	x_2	x_3	x_4	x_5	y
1	+	—	—	—	—	—	50,0
2	+	+	+	—	—	—	57,2
3	+	—	—	+	+	—	48,1
4	+	+	—	+	—	+	46,0
5	+	—	+	+	—	+	64,8
6	+	+	—	—	+	+	45,3
7	+	—	+	—	+	+	54,8
8	+	+	+	+	+	—	53,0
b_j	52,400	—2,025	5,050	0,575	—2,100	0,325	

Была намечена вторая серия опытов для случая, если поиск оптимальных условий окажется неэффективным. Обобщающий определяющий контраст второй $1/4$ -реплики имеет вид $1 = x_1 x_2 x_3 x_4 = x_1 x_2 x_5 = x_3 x_4 x_5$. Он был выбран так, чтобы знаки тройных произведений оказались противоположными знакам тех же произведений в первой четверть-реплике.

7.5. Реплики большой дробности

Если вы, наш терпеливый читатель, осилили два предыдущих параграфа, вы сможете выбрать реплику любой дробности. Процедура выбора совершенно аналогична.

При выборе $1/8$ -реплики 2^{6-3} можно воспользоваться вектор-столбцами трех взаимодействий, например, так:

- 1) $x_4 = x_1 x_2$, $x_5 = x_1 x_3$, $x_6 = x_2 x_3$;
- 2) $x_4 = x_1 x_3$, $x_5 = x_2 x_3$, $x_6 = x_1 x_2 x_3$;
- 3) $x_4 = x_1 x_2$, $x_5 = x_2 x_3$, $x_6 = x_1 x_2 x_3$;
- 4) $x_4 = x_1 x_2$, $x_5 = x_1 x_3$, $x_6 = x_1 x_2 x_3$.

Для каждого из этих решений можно сделать шесть перестановок.

Итого получается 24 возможности выбора $1/8$ -реплики. Это при условии, что мы всюду выбираем положительные генерирующие соотношения.

Из четырех приведенных выше решений наименее удачно первое, поскольку все линейные эффекты смешиваются с парными взаимодействиями.

Если априорно известно, что из всех взаимодействий наиболее существенно $x_1 x_2$, то нужно выбрать второе решение, если $x_1 x_3$ — третье, а если $x_2 x_3$ — четвертое.

Допустим, мы выбрали четвертое решение, предполагая, что из факторов x_4 , x_5 , x_6 наиболее существенным является x_4 . Приравняем x_4 тройному взаимодействию и запишем генерирующие соотношения

$$x_4 = x_1 x_2 x_3, \quad x_5 = x_1 x_2, \quad x_6 = x_1 x_3.$$

Ограничимся парными и тройными взаимодействиями.

Для $1/8$ -реплики с генерирующими соотношениями

$$x_4 = x_1 x_2 x_3, \quad x_5 = x_1 x_2, \quad x_6 = x_1 x_3$$

имеем следующие определяющие контрасты:

$$1 = x_1 x_2 x_3 x_4, \quad 1 = x_1 x_2 x_5, \quad 1 = x_1 x_3 x_6.$$

Если попарно перемножить эти определяющие контрасты, то получим

$$1 = x_3 x_4 x_5, \quad 1 = x_2 x_4 x_6, \quad 1 = x_2 x_3 x_5 x_6.$$

Произведение трех определяющих контрастов равно

$$1 = x_1 x_4 x_5 x_6.$$

Чтобы полностью охарактеризовать разрешающую способность данной $1/8$ -реплики, запишем обобщающий определяющий контраст

$$1 = x_1 x_2 x_3 x_4 = x_1 x_2 x_5 = x_1 x_3 x_6 = x_3 x_4 x_5 = x_2 x_4 x_6 = x_2 x_3 x_5 x_6 = x_1 x_4 x_5 x_6.$$

Получается следующая система смешивания:

$$\begin{aligned} b_1 &\rightarrow \beta_1 + \beta_{25} + \beta_{36} + \beta_{234} + \beta_{456}, & b_4 &\rightarrow \beta_4 + \beta_{35} + \beta_{26} + \beta_{123} + \beta_{156}, \\ b_2 &\rightarrow \beta_2 + \beta_{15} + \beta_{46} + \beta_{134} + \beta_{356}, & b_5 &\rightarrow \beta_5 + \beta_{12} + \beta_{34} + \beta_{236} + \beta_{146}, \\ b_3 &\rightarrow \beta_3 + \beta_{16} + \beta_{45} + \beta_{124} + \beta_{256}, & b_6 &\rightarrow \beta_6 + \beta_{13} + \beta_{24} + \beta_{235} + \beta_{145}. \end{aligned}$$

Рассмотрим пример $1/16$ -реплики от 2^7 .

Пример 4. Этот пример относится к планированию эксперимента при оптимизации процесса получения производного пиперазина по способу А. Изучалось влияние на выход продукта семи факторов, приведенных в табл. 7.8. Использовалась $1/16$ часть от полного факторного эксперимента 2^7 . Это дает возможность сократить число опытов до 8 вместо 128.

В табл. 7.9 приведена матрица планирования и соответствующие коэффициенты линейного уравнения.

Реплика задана генерирующими соотношениями

$$x_4 = x_1 x_2; \quad x_5 = x_1 x_3; \quad x_6 = x_2 x_3; \quad x_7 = x_1 x_2 x_3.$$

Для них имеем следующие определяющие контрасты:

$$1 = x_1 x_2 x_4; \quad 1 = x_1 x_3 x_5; \quad 1 = x_2 x_3 x_6; \quad 1 = x_1 x_2 x_3 x_7.$$

Таблица 7.8

Уровни факторов и интервалы варьирования

Факторы	Уровни факторов			Интервал варьирования
	-1	0	+1	
x_1 — количество NaOH, внесенного в реакционную массу до прилива вещества a , г/мол	0,0075	0,0180	0,0285	0,0105
x_2 — способ поддержания pH (вид раствора)	18%-ный раствор NaOH в метаноле	—	40%-ный водный раствор NaOH	—
x_3 — время прилива вещества a и раствора NaOH, час	3	4,5	6	1,5
x_4 — время выдержки, час	1	2	3	1,0
x_5 — температура, °C	20	25	30	5,0
x_6 — весовое соотношение вещества b и метанола, г/г	1 : 3	1 : 3,5	1 : 4	0,5 *
x_7 — молярное соотношение веществ a и b	1 : 1	1 : 1,1	1 : 1,2	0,1 *

* Этот интервал относится ко второму члену пропорции.

Таблица 7.9

Матрица планирования 2^{7-4} и коэффициенты

Номер опыта	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
1	+	+	+	+	+	+	+	+	19,3
2	+	+	—	—	—	—	+	+	23,8
3	+	—	—	+	+	—	—	+	31,3
4	+	—	+	—	—	+	—	+	12,8
5	+	—	—	—	+	+	+	—	32,0
6	+	—	+	+	—	—	+	—	14,0
7	+	+	+	—	+	—	—	—	25,0
8	+	+	—	+	—	+	—	—	30,5
b_j	+23,5875	+1,0625	-5,8125	+0,1875	+3,3125	+0,0625	-1,3125	-1,7875	

Обобщающий определяющий контраст

$$1 = x_1x_2x_4 = x_1x_3x_5 = x_2x_3x_6 = x_1x_2x_5x_7 = x_2x_3x_4x_5 = x_1x_3x_4x_6 = x_3x_4x_7 = \\ = x_1x_2x_5x_6 = x_2x_5x_7 = x_1x_6x_7 = x_4x_5x_6 = x_1x_4x_5x_7 = x_2x_4x_6x_7 = \\ = x_3x_5x_6x_7 = x_1x_2x_3x_4x_5x_6x_7.$$

Такой обобщающий определяющий контраст получен в результате парного перемножения исходных контрастов, затем — умножения по три и по четыре.

Если всеми коэффициентами взаимодействия, начиная с тройных, можно пренебречь, то коэффициенты будут совместными оценками:

$$\begin{aligned} b_1 &\rightarrow \beta_1 + \beta_{24} + \beta_{35} + \beta_{57}, & b_5 &\rightarrow \beta_5 + \beta_{13} + \beta_{46} + \beta_{27}, \\ b_2 &\rightarrow \beta_2 + \beta_{14} + \beta_{36} + \beta_{57}, & \beta_6 &\rightarrow \beta_6 + \beta_{23} + \beta_{45} + \beta_{17}, \\ b_3 &\rightarrow \beta_3 + \beta_{15} + \beta_{26} + \beta_{47}, & b_7 &\rightarrow \beta_7 + \beta_{24} + \beta_{25} + \beta_{16}, \\ b_4 &\rightarrow \beta_4 + \beta_{12} + \beta_{56} + \beta_{37}. \end{aligned}$$

Разрешающая способность такой реплики чрезвычайно мала, так как каждый линейный эффект определяется совместно с тремя парными взаимодействиями. Такой репликой можно пользоваться только в том случае, если все парные взаимодействия равны нулю. В большинстве случаев, начиная исследование процесса, трудно априорно предсказать, будут эффекты взаимодействия или нет. Поэтому экспериментатор должен наметить план дальнейших опытов для случая, если парные эффекты значимы и поиск оптимальных условий будет неэффективным. В нашем примере нужно реализовать специальных образом выбранную вторую реплику 2^{7-4} (метод перевала).

Матрицу планирования для этой реплики можно получить из первой реплики, изменив в ней все знаки на обратные. Такая реплика задается генерирующими соотношениями

$$x_4 = -x_1x_2; \quad x_5 = -x_1x_3; \quad x_6 = -x_2x_3; \quad x_7 = +x_1x_2x_3.$$

В обобщающем определяющем контрасте все тройные произведения оказываются со знаком минус, и поэтому в совместных оценках для линейных эффектов не будет парных взаимодействий со знаком плюс. Усредняя результаты вычислений для таких двух реплик, можно получить отдельные оценки для всех линейных эффектов.

Мы последовательно рассмотрели реплики различной дробности: $1/2$ -реплику от 2^8 , $1/2$ -реплику от 2^4 , $1/4$ -реплику от 2^5 , $1/8$ -реплику от 2^6 и $1/16$ -реплику от 2^7 . В первом случае необходимы четыре опыта. Во всех прочих случаях экспериментатор ставит восемь опытов. С ростом числа факторов увеличивается дробность реплик и усложняется система смешивания. Предельное число факторов для восьми опытов — семь. В этом случае оценивается восемь коэффициентов линейного уравнения $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7$; число степеней свободы равно нулю. При числе факторов от 8 до 15 приходится ставить 16 опытов. С ростом числа факторов дробность реплик растет следующим образом: $1/32$ -реплика от 2^9 , $1/64$ -реплика от 2^{10} , $1/128$ -реплика от 2^{11} , $1/256$ -реплика от 2^{12} , $1/512$ -реплика от 2^{13} , $1/1024$ -реплика от 2^{14} и $1/2048$ -реплика от 2^{15} . Предельное число факторов для 16 опытов — пятнадцать. План с предельным числом факторов для данного числа опытов и заданной модели называется насыщенным. В этом случае число опытов равно числу оцениваемых коэффициентов. Все рекомендации для выбора системы смешивания аналогичны приведенным выше. Можно, далее, рассматривать построение дробных планов для числа факторов от 16 до 31 (при этом необходимо ставить 32 опыта), для числа

Таблица 7.10

Планирование опытов при изучении процесса хлорирования титановых шлаков

Основной уровень Интервал варьирования Верхний уровень Нижний уровень	16,6 5,2	750 50	3,5 1,5	2 1	5,65 1,35	4,0 0,5	8,3 1,7	1 1	6,65 1,35	5,7 1,3	4,7 1,3	7 3	7 3	8 2	1:1 0,5*		
	24,8 11,4	800 700	5 2	3 1	7,0 4,3	4,5 0,5	10,0 6,6	2 0	8,0 5,3	7,0 4,4	6,0 3,4	10 4	10 4	10 6	1:0,5 1:1,5		
Номер опыта	Кодовые обозначения переменных																\bar{y}
	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	39,2
2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	48,5
3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	48,0
4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	44,7
5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	47,2
6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	41,7
7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	33,4
8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	40,5
9	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	48,2
10	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	49,4
11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	98,5
12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	95,0
13	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	92,5
14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	103,75
15	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	110,0
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	102,5
b_j	75,25	8,69	3,75	1,20	28,79	-5,09	8,38	7,40	1,29	1,66	1,65	-5,12	-3,71	7,78	-0,68	-4,99	151,0
																	48,7
																	151,05
																	151,1
																	93,5
																	103,7
																	127,0
																	123,5
																	125,25
																	85,0
																	103,0

* Интервал относится ко второму члену пропорции.

факторов от 32 до 63 (здесь необходимы 64 опыта) и т. д. Однако для решения столь сложных задач рекомендуется применять методы отбора факторов, например метод случайного баланса [5].

Поэтому в «Ограничениях» мы и отметили, что будем рассматривать задачи с числом факторов от 2 до 15.

В качестве последнего примера рассмотрим применение плана 2^{15-11} .

Пример 5. Изучалось влияние следующих 15 факторов на скорость хлорирования титановых шлаков в расплаве [6]. Пусть \bar{x}_1 — расход хлора; \bar{x}_2 — температура; \bar{x}_3 — концентрация углерода в расплаве; \bar{x}_4 — концентрация TiO_2 в расплаве; \bar{x}_5 — \bar{x}_{11} — компоненты, определяющие состав шлака; \bar{x}_{12} — \bar{x}_{15} — компоненты расплава.

Столь большое число факторов было включено в программу исследования потому, что процесс хлорирования плохо изучен. Планы такого типа применяются в планировании эксперимента довольно редко, так как число степеней свободы равно нулю и невозможно провести последовательный статистический анализ. Здесь даже нет смысла выписывать совместные оценки, линейные эффекты смешаны со 105 парными взаимодействиями. Освободиться от них можно с помощью метода «перевала»: нужно дополнить первую реплику второй, изменив все знаки в матрице планирования на обратные. В табл. 7.10 представлены уровни факторов и интервал варьирования, матрица планирования и результаты экспериментов. В последней строке помещены коэффициенты.

7.6. Резюме

Дробные реплики находят широкое применение при получении линейных моделей. Целесообразность их применения возрастает с ростом количества факторов. В табл. 7.10 показано, что при исследовании влияния 15 факторов можно в 2048 раз сократить число опытов, применяя реплику большой дробности (16 опытов вместо 32768). Эффективность применения дробных реплик зависит от удачного выбора системы смешивания линейных эффектов с эффектами взаимодействия, а также от умелой стратегии экспериментирования в случае значимости некоторых взаимодействий. Априорные сведения о взаимодействиях могут оказать большую услугу экспериментатору.

При построении дробных реплик используют следующее правило: для того чтобы сократить число опытов при введении в планирование нового фактора, нужно поместить этот фактор в вектор-столбец матрицы, принадлежащий взаимодействию, которым можно пренебречь.

Реплики, которые используются для сокращения опытов в 2^m раз, где $m=1, 2, 3, 4, \dots$, называются регулярными. Они позволяют производить расчет коэффициентов уравнения так же просто, как и в случае полного факторного эксперимента.

При применении дробных реплик линейные эффекты смешиваются с эффектами взаимодействий. Чтобы определить систему смешивания, нужно знать определяющие контрасты и генерирующие соотношения. Определяющим контрастом называется символическое обозначение произведения любых столбцов, равное ± 1 .

Чтобы определить, какие взаимодействия смешаны с данным линейным эффектом, нужно умножить определяющий контраст на этот линейный эффект и получить генерирующие соотношения. Например, если имеются следующие генерирующие соотношения: $x_1 = x_2 x_3$, $x_2 = x_1 x_3$ и $x_3 = x_1 x_2$, то определяющий контраст будет $1 = x_1 x_2 x_3$.

Эффективность реплики зависит от системы смешивания. Реплики, у которых линейные эффекты смешаны с взаимодействиями наивысшего порядка, являются наиболее эффективными, так как обладают наибольшей разрешающей способностью.

Для освобождения линейных эффектов от взаимодействий первого порядка можно использовать метод «перевала». Смысл метода в добавлении новой реплики, все знаки которой противоположны исходной реплике.

С ростом числа факторов быстро увеличивается число реплик различной дробности. Эти реплики характеризуются обобщающими определяющими контрастами, которые получают перемножением по два, по три и т. д. исходных определяющих контрастов.

Мы научились строить полные и дробные факторные эксперименты. Давайте теперь посмотрим, как их реализовать.

Литература

1. W. S. Connor, S. Young. Fractional Factorial Designs for Experiments with Factors of two and three Levels. National Bureau of Standards, Applied Mathematics Series, 58, 1961.
2. O. L. Davies, W. A. Hay. The construction and uses of fractional factorial designs in industrial research. Biometrics, 6, 1950.
3. В. З. Бродский. Многофакторные регулярные планы. М., изд-во МГУ, 1972.
4. Е. В. Маркова, П. Фарка, И. П. Борисова и др. Метод планирования эксперимента при поиске оптимальных условий получения полимерного серусодержащего антиоксиданта. — Пластические массы, 1969, № 1.
5. F. E. Satterthwaite. Random Balance Experimentation. — Technometrics, 1959, 1, N 2.
6. П. П. Хомяков, Ю. П. Адлер, В. В. Налимов. Выявление факторов, влияющих на скорость хлорирования титановых шлаков в расплаве. — Заводская лаборатория, 1963, 29, № 1.

Глава восьмая

ПРОВЕДЕНИЕ ЭКСПЕРИМЕНТА

Перед ошибками захлопываем дверь.
В смятенье Истина: «Как я войду теперь?»
Р. Тагор. Искры

Эта глава посвящена конфликту между математикой и реальной действительностью. Тому, кто строит изящное здание в виде матрицы планирования, нужно помнить о тех темных силах, под действием которых здание может рухнуть, не принеся никому пользы. Нужно помнить об ошибках опыта! Как бы остроумно вы ни планировали эксперимент, какие бы системы смешивания ни избирали, все труды ваши будут напрасны, если вы не спуститесь на землю, к реальной действительности и не продумаете все детали постановки опыта.

В этой главе рассказывается, как нужно готовиться к опыту, как реализовать матрицу планирования, как подсчитывать ошибки и классифицировать их, как бороться с некоторыми систематическими ошибками в виде неоднородности сырья, различий в аппаратуре и т. д. Но прежде всего надо как-то формализовать сведения, имеющиеся об объекте. Для этого можно воспользоваться анкетой, приведенной в следующем параграфе.

8.1. Анкета для сбора априорной информации

Постановка задачи, выбор параметров оптимизации

1. Краткое описание процесса, объекта.
 2. Формулировка цели исследования (если задач несколько — проранжировать их по степени важности).
 3. Выбор параметров оптимизации (откликов).
- Заполните следующую таблицу, включив в нее все возможные отклики

Номер отклика	Название	Размерность	Область определения	Точность	Примечание

4. Желаемый результат. Число и точность.

5. Какой результат будет считаться отличным, хорошим, удовлетворительным, неудовлетворительным.

Выбор факторов

1. Список всех «подозреваемых»: факторов, которые могут влиять на процесс.

2. Список факторов, включаемых в реальный эксперимент.

Номер фактора	Название	Размерность	Область определения	Область интереса	Точность	Примечание

3. Существуют ли возможности установления значения фактора на любом заданном уровне?

4. Сохраняются ли заданные значения уровней в течение опыта?

5. Могут ли некоторые комбинации уровней факторов привести к остановке процесса (например, взрыв, нетехнологичность и т. д.)?

Число опытов

1. Желаемое число опытов, ограничения на число опытов.

2. Желаемый срок проведения исследования.

3. Примерная длительность одного опыта.

4. Стоимость и затраты труда при проведении одного опыта серии.

5. Желаемое число уровней для одного фактора.

6. Возможность выполнения параллельных опытов и их желаемое число.

7. Возможность проведения параллельных измерений.

8. Желаемая стратегия проведения опытов (например, по одному в день и т. д.).

Учет априорной информации

1. Условия и результаты, достигнутые при изучении аналогичных процессов.

2. Результаты предварительного эксперимента и данные (литературные и собственные) о величине ошибки эксперимента.

3. Взаимодействия факторов.

В следующем параграфе приведен конкретный пример постановки задачи, в котором использованы некоторые части этой анкеты.

8.2. Реализация плана эксперимента

К проведению опытов необходимо тщательно подготовиться, собрать опытную установку, проверить и прокалибровать приборы, подготовить исходное сырье, составить специальный журнал. Журнал заранее оформляют в соответствии с методикой и планом опытов так, чтобы была ясна последовательность действий. Первую страницу можно посвятить выбору цели исследования и параметрам оптимизации, с указанием их размерностей. Желательно перечислить все параметры, которые могут служить характеристиками процесса и указать, какая между ними существует корреляция. Если же сведения о корреляции отсутствуют, целесообразно подсчитать коэффициенты парной корреляции, проверить их значимость и выделить группу некоррелированных параметров. На второй странице перечислить факторы и поместить таблицу уровней факторов и интервалов варьирования. Не забудьте указать единицы измерения факторов! Для матрицы планирования удобно отвести разворот журнала, чтобы имела возможность дополнить ее до расчетной матрицы, записать повторные опыты и примечания.

Чтобы облегчить работу лаборанта и исключить ошибки при выборе условий опыта, в рабочей матрице планирования целесообразно проставлять не только кодовые значения факторов, но и натуральные.

При составлении рабочей матрицы планирования необходимо оставить место для столбцов, в которых отмечаются даты постановки опытов и фамилии лаборантов, если опыты проводят несколько человек. Имея перед собой план опытов, необходимо подсчитать количество исходного сырья и заранее его подготовить. Желательно, чтобы сырье было однородное. Если требование однородности выполнить невозможно, нужно заблаговременно определить количество различных партий сырья и соответствующим образом разбить матрицу планирования на блоки. На этом вопросе мы далее остановимся подробно. Отдельные страницы нужно отвести для расчетов, которые необходимы для определения количеств всех компонентов реакции и т. п., а также для анализа результатов эксперимента. Все расчеты должны сохраняться до окончания работы.

Пример 1. В качестве примера приведем оформление журнала при оптимизации процесса получения сульфадимизина.

Страница 1

Планирование эксперимента при оптимизации процесса получения сульфадимизина

Цель исследования: определение оптимальных условий процесса конденсации сульгина с ацетилацетоном в присутствии уксусной кислоты.

Параметры, характеризующие процесс: y_1 — выход сульфадимизина по сульгину, %, y_2 — качество сульфадимизина.

Формулировка задачи оптимизации. Достижение максимального выхода сульфадимизина, качество которого удовлетворяет требованиям фармакопей: y_1 — параметр оптимизации ($y_1 \rightarrow 100\%$), y_2 — ограничение.

Качество продукта определяется по процентному содержанию сульфадимизина в получаемом продукте и по его температуре плавления. Согласно требованиям фармакопей содержание основного вещества в получаемом продукте должно быть не менее 99%, а температура плавления должна находиться в пределах 196—200° С, т. е. $y_2' \geq 99\%$, $196^\circ \text{С} < y_2'' < 200^\circ \text{С}$.

Страница 2

Факторы, определяющие процесс: x_1 — время реакции, час; x_2 — содержание ацетилацетона в реакционной массе, %; x_3 — содержание уксусной кислоты в реакционной массе, %; x_4 — температура реакционной массы, °С; x_5 — качество ацетилацетона, %; x_6 — качество сульфина, %.

Выбор варьируемых факторов. Принято решение изменять в опытах первые три фактора. Качество ацетилацетона и сульфина решено поддерживать постоянным (таким, как на действующем производстве) $x_5 = 90\%$; $x_6 = 98\%$.

Температура реакционной среды является производной состава и давления. Поэтому она, если не применять специальных способов воздействия на температуру, не является независимой величиной и не может служить в качестве фактора. Однако температуру необходимо контролировать в течение всех опытов.

Страница 3

Выбор технологии. Сульфин загружается одновременно с ацетилацетоном и уксусной кислотой. Реакция проводится при перемешивании реакционной смеси и непрерывном отгоне воды.

Необходимые анализы. Анализ исходного сырья: ацетилацетона, сульфина, уксусной кислоты (следует описание методик). Анализы получаемых продуктов: сульфадимизина в осадке, сульфадимизина в фильтрате, сульфина в фильтрате (следует описание методик).

Описание экспериментальной установки. Опыты проводятся на лабораторной установке, состоящей из стеклянной конической колбы емкостью 250 мл, снабженной металлической якорной мешалкой и обратным холодильником. Температура реакционной массы измеряется термопарой, подключенной к электрическому потенциометру, и непрерывно записывается на картограмму. Колба обогревается электрической баней, наполненной вазелиновым маслом. Температура в бане автоматически регулируется с помощью реле и контактного термометра и поддерживается около 160° С.

Страница 4

Выбор основного уровня и интервалов варьирования. Для того чтобы выбрать уровни факторов, следует собрать и проанализировать литературные и заводские данные. По заводскому регламенту процесс проводится при следующих условиях: $x_1 = 27$ час, $x_2 = 4\%$, $x_3 = 16\%$. При этом $y_1 = 84\%$.

Опубликованные данные и сведения из отчетов (априорная информация)

Влияние времени реакции (x_1). Данные об оптимальном времени реакции в лабораторных условиях противоречивы. Так, в отчете № 1 указано, что опыты проводились при $x_1 = 21$ час, затем время уменьшили до 12 час. Уменьшение времени не снизило существенно выход реакции. В отчете № 2 описы-

ваются опыты с различным временем: 18, 24 и 30 час. Наилучший выход получен при $x_1 = 24$ час.

Влияние избытка ацетилацетона (x_2). Данные о влиянии избытка ацетилацетона от стехиометрического соотношения также противоречивы. В одном отчете указано, что содержание ацетилацетона сверх 10% является нецелесообразным, в другом оптимальным считается 40% избытка ацетилацетона.

Влияние уксусной кислоты (x_3). Вопрос о влиянии процентного содержания уксусной кислоты специально не исследован. Считается, что x_3 целесообразно поддерживать около 16—17%. Предполагается, что это растворитель и его концентрация может изменяться в широких пределах.

На основании анализа имеющихся сведений решено выбрать следующие уровни и интервалы варьирования факторов (табл. 8.1)

Таблица 8.1

Уровни факторов и интервалы варьирования

Факторы	Уровни			I	Размерность
	—1	0	+1		
x_1	16	18	20	2	час
x_2	20	24	28	4	%
x_3	12	15	18	3	%

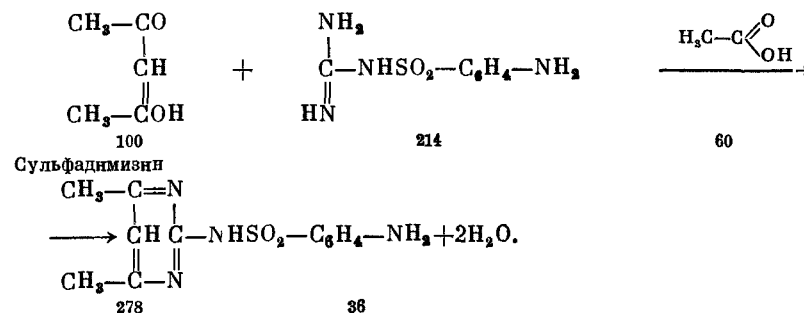
Страница 5

Расчет компонентов реакции

Ацетилацетон

Сульфин

Уксусная кислота



Расчет по стехиометрическим соотношениям на 12,5 г ацетилацетона:

$$\begin{array}{l}
 100-214 \\
 12,5 - z
 \end{array}
 \quad z = \frac{214 \cdot 12,5}{100} = 26,7 \text{ г;}$$

ацетилацетона — 12,5 г, сульфина — 26,7 г.

Расчет ацетилацетона с избытком 20%: $12,5 + 12,5 \cdot 0,2 = 15$ г и т. д.

Таблица 8.2

Матрица планирования и результаты эксперимента

Номер опыта в матрице	Случайный порядок реализации опытов	Дата проведения опытов *	x_1		x_2		x_3		$y, \%$	
			код	ч	код	°/о	код	°/о	повторные опыты	средний результат
1	2 9	10/III 5/IV	—	16	—	20	—	12	80,23 81,93	81,08
2	6 13	21/III 14/IV	+	20	—	20	—	12	86,50 84,80	85,65
3	1 15	8/III 21/IV	—	16	+	28	—	12	82,45 82,10	82,27
4	7 10	27/III 6/IV	+	20	+	28	—	12	89,50 91,30	90,40
5	3 16	12/III 27/IV	—	16	—	20	+	18	85,10 84,80	84,95
6	8 14	1/IV 16/IV	+	20	—	20	+	18	90,30 —	89,95
	14'	17/IV							89,60	
7	4 12	15/III 12/IV	—	16	+	28	+	18	85,60 84,90	85,25
8	5 11	20/III 9/IV	+	20	+	28	+	18	88,02 88,48	88,25

* Начало проведения всех опытов в 8 ч. В таблице необходимо указать фамилии лаборантов по сменам. Для каждого опыта дать картограмму температуры.

Имея такую таблицу, можно приступить к обработке экспериментальных данных.

Каждая горизонтальная строка матрицы — это условия опыта. Условия опытов чрезвычайно разнообразны. Ведь мы занимаемся планированием многофакторного эксперимента, когда все факторы изменяются одновременно. Приступая к планированию эксперимента, мы должны отказаться от привычного однофакторного эксперимента, который проводится по принципу «изменяй один фактор, а прочие держи постоянными».

Мы хотим заниматься исследованием сложных многофакторных систем и понимаем, что однофакторный эксперимент нам не поможет. И это не только наше мнение. У. Р. Эшби во «Введении в кибернетику» писал: «Тот факт, что в течение столетий могли принимать такую догму, как «изменяйте факторы по одному», показывает, что ученые занимались в основном исследова-

$y, \%$		$y, ^\circ\text{C}$		Примечания		
повторные опыты	средний результат	повторные опыты	средний результат	начало кипения	начало помутнения	выпадение осадка
94,2 99,9	97,05	196 197	196,5	8 ч 45 мин 8 ч 38 мин	20 ч 20 мин 20 ч 18 мин	20 ч 40 мин 20 ч 35 мин
99,6 100,0	99,80	197 196	196,5	8 ч 38 мин 8 ч 40 мин	20 ч 20 мин 20 ч 23 мин	20 ч 40 мин 20 ч 43 мин
99,2 100,0	99,60	195 196	195,5	8 ч 32 мин 8 ч 30 мин	20 ч 40 мин 20 ч 45 мин	20 ч 55 мин 21 ч
99,8 99,9	99,85	196 197	196,5	8 ч 20 мин 8 ч 25 мин	20 ч 20 ч 15 мин	20 ч 32 мин 21 ч
99,5 99,8	99,65	197 198	197,5	8 ч 25 мин 8 ч 25 мин	19 ч 19 ч 05 мин	19 ч 15 мин 19 ч 18 мин
99,6 —	99,65	196 —	196,0	8 ч 30 мин При кипении	20 ч 10 мин из колбы сильно жидкость **	20 ч 35 мин выбрасывало
99,7		196		8 ч 20 мин	20 ч	20 ч 20 мин
99,5 99,6	99,55	197 196	196,5	8 ч 25 мин 8 ч 28 мин	18 ч 15 мин 18 ч 45 мин	18 ч 35 мин 19 ч
99,5 99,7	99,60	197 197	197,0	8 ч 35 мин 8 ч 32 мин	20 ч 45 мин 21 ч	21 ч 21 ч 45 мин

** Опыт повторить.

нием систем, допускающих этот метод, ибо в сложных системах он часто неприменим по существу».

Проанализируем, в чем состоит недостаток однофакторного эксперимента и почему им нецелесообразно пользоваться при исследовании многофакторных систем. При однофакторном эксперименте, варьируя одним фактором и стабилизируя все прочие на произвольно выбранных уровнях, экспериментатор получает зависимость параметра оптимизации только от одного фактора и определяет локальный оптимум. Далее он повторяет аналогичную процедуру для второго, третьего и k -го фактора. В результате длительной и кропотливой работы, требующей много средств и времени, опытные данные представляются десятками графиков, которые в сущности имеют иллюстративный характер.

За время эксперимента могут происходить изменения в аппаратуре, сырье и т. д. Все это вносит изменения в результаты

эксперимента, вследствие чего данные многих опытов являются несопоставимыми. В планировании эксперимента разработана четкая стратегия экспериментирования. Экспериментатор может минимизировать число опытов, пользуясь шаговым принципом и дробным планированием.

Имеются способы борьбы с неконтролируемым дрейфом, вызванным изменением аппаратуры, сырья и т. п. (об этом мы расскажем вам в этой главе). Все это не предусмотрено в однофакторном эксперименте.

Давайте посмотрим, что значит однофакторный эксперимент при исследовании semifакторной системы. Итак, экспериментатор хочет исследовать влияние семи факторов на некоторый параметр оптимизации и решил проводить однофакторный эксперимент. Для того чтобы построить кривую, обычно берут четыре-пять экспериментальных точек. Возьмем четыре точки. Необходимое количество опытов при реализации всевозможных комбинаций равно $N=4^7=16384$. Совершенно ясно, что такое количество опытов реализовать невозможно. Значит, экспериментатор произвольно отбросит очень многие комбинации и реализует небольшую часть опытов, изменяя факторы по одному при постоянных значениях прочих факторов. Тогда естественно возникнет вопрос, как будут выглядеть кривые, если прочие факторы заставить варьировать на другом уровне? Несомненно, кривые изменятся. Перебрать все комбинации — значит поставить 16384 опыта.

На предыдущих страницах мы рассказывали вам о шаговом принципе исследования поверхности отклика и о дробном факторном эксперименте. Все эти приемы и методы предлагаются экспериментатору для разумной минимизации числа опытов.

Нам пришлось сделать довольно большое отступление. Возвратимся к матрице планирования.

Одновременное изменение всех факторов вносит в условия опытов большое разнообразие. Разнообразие увеличивается с ростом числа факторов.

Обратите, пожалуйста, внимание на матрицу на стр. 106 (матрица предыдущей главы). В опыте № 1 все факторы находятся на нижних уровнях. Это значит, что опыт проводится при следующих условиях: \bar{x}_1 — соотношение между NaOH и диэтилкарбаминхлоридом 1 : 1; \bar{x}_2 — соотношение между пиперазином и диэтилкарбаминхлоридом 1 : 1; \bar{x}_3 — время выдержки, 3 час; \bar{x}_4 — температура, 20° С; \bar{x}_5 — время прилива диэтилкарбаминхлорида, 20 мин.

В опыте № 2 первые два фактора поддерживаются на верхних уровнях, а все остальные — на нижних. В опыте № 3 на верхних уровнях находятся \bar{x}_3 и \bar{x}_4 и т. д. В последнем опыте все факторы, кроме \bar{x}_5 , находятся на верхних уровнях.

В результате такого многообразия условий получаются различные значения параметра оптимизации. Так, наименьшее зна-

чение в матрице — 45,3%, а наибольшее — 64,8%. Эта разница весьма ощутима и можно сделать вывод, что условия опыта № 5 лучше, чем условия опыта № 6.

Но всегда ли легко определить, что один опыт лучше другого? Неправильно утверждать, что условия опыта № 3 более выгодны с точки зрения величины параметра оптимизации, чем опыта № 4 ($y_3=48,1$, $y_4=46,0$). Экспериментаторы, вооруженные статистическими методами, поступают осторожно. Они должны проверить, значимо ли отличаются результаты опытов № 3 и № 4 друг от друга. Интуитивное мнение является чрезвычайно слабым указанием на возможное превосходство опыта № 3. Нужны объективные оценки. Прежде всего вам должна быть известна ошибка опыта, которая может быть столь большой, что разница в выходе потонет на ее фоне. И, напротив, ошибка может иметь столь малую величину, что разница в 2,1% окажется значимой.

Если известна ошибка опыта, то значимость различий двух средних можно проверить с помощью t -критерия (критерия Стьюдента) по формуле

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

где \bar{y}_1 — среднее значение выхода в одном опыте, \bar{y}_2 — среднее значение выхода в другом опыте, s — ошибка опыта (рассматривается случай, когда ошибки для первого и второго опыта близки одна к другой), n_1 — количество наблюдений в первом опыте, n_2 — количество наблюдений во втором опыте.

Эта формула предназначена для сравнения средних значений двух малых выборок с равными дисперсиями. Проверка значимости ведется по табулированным значениям t -критерия.

Так, если в нашем случае $s=1$, количество параллельных опытов одинаково и равно двум, то

$$t = \frac{2,1}{1 \sqrt{\frac{1}{2} + \frac{1}{2}}} = 2,1.$$

Число степеней свободы равно $n_1+n_2-2=2$. Табличное значение t -критерия для $f=2$ и 5%-ного уровня значимости равно 4,3. Это означает, что вероятность того, что при 2 степенях свободы значение величины t будет больше по абсолютной величине чем 4,3, равна 0,05. Поскольку экспериментальное значение t меньше табличного, то с вероятностью $P=1-\alpha=0,95$ можно считать, что различия между результатами двух опытов нет.

А теперь познакомимся с вычислением ошибки опыта, или, как ее часто называют, ошибки воспроизводимости.

8.3. Ошибки параллельных опытов

Каждый эксперимент содержит элемент неопределенности вследствие ограниченности экспериментального материала. Постановка повторных (или параллельных) опытов не дает полностью совпадающих результатов, потому что всегда существует ошибка опыта (ошибка воспроизводимости). Эту ошибку и нужно оценить по параллельным опытам. Для этого опыт воспроизводится по возможности в одинаковых условиях несколько раз и затем берется среднее арифметическое всех результатов. Среднее арифметическое \bar{y} равно сумме всех n отдельных результатов, деленной на количество параллельных опытов n

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{\sum_{i=1}^n y_i}{n}.$$

Отклонение результата любого опыта от среднего арифметического можно представить как разность $y_i - \bar{y}$, где y_i — результат отдельного опыта. Наличие отклонения свидетельствует об изменчивости, вариации значений повторных опытов. Для измерения этой изменчивости чаще всего используют *дисперсию*. Дисперсией называется среднее значение квадрата отклонений величины от ее среднего значения. Дисперсия обозначается s^2 и выражается формулой

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1},$$

где $(n-1)$ — число степеней свободы, равное количеству опытов минус единица. Одна степень свободы использована для вычисления среднего.

Корень квадратный из дисперсии, взятый с положительным знаком, называется средним квадратическим отклонением, стандартом или квадратичной ошибкой

$$s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}.$$

Стандарт имеет размерность той величины, для которой он вычислен. Дисперсия и стандарт — это меры рассеяния, изменчивости. Чем больше дисперсия и стандарт, тем больше рассеяны значения параллельных опытов около среднего значения

Рассказывая о расчете среднего значения, мы употребили такую формулировку: среднее арифметическое равно сумме всех n отдельных результатов, деленной на количество параллельных опытов n .

Такая формулировка может вызвать у вас некоторые размышления. Действительно, читатель, имеющий математико-статистическую подготовку, может дополнить эту формулировку, и она будет выглядеть следующим образом: среднее арифметическое равно сумме всех n отдельных результатов, деленной на n , если они имеют нормальное распределение.

Например, наличие резко отклоняющихся результатов (так называемых грубых наблюдений) свидетельствует о нарушении закона нормального распределения. При наличии грубых наблюдений нужно сначала их исключить, а затем подсчитывать среднее арифметическое и дисперсию.

Далее мы остановимся на этом вопросе подробнее.

Рамки этой книги не позволяют сделать экскурс в область математической статистики и подробно остановиться на этом вопросе. Мы можем рекомендовать вам книгу Н. Бейли «Статистические методы в биологии», где этот вопрос изложен на уровне, доступном для понимания экспериментаторов [1].

Надо всегда следить, чтобы не нарушались необходимые условия выполнения той или иной операции. Иначе вы рискуете принять абсурд за истину.

Ошибка опыта является суммарной величиной, результатом многих ошибок: ошибок измерений факторов, ошибок измерений параметра оптимизации и др. Каждую из этих ошибок можно, в свою очередь, разделить на составляющие.

Вопрос о классификации ошибок довольно сложный и вызывает много дискуссий. В качестве примера одной из возможных схем классификации мы приведем схему из книги Ю. В. Кельнича «Теория ошибок измерений» (М., изд-во «Недра», 1967) (рис. 24).

Все ошибки принято разделять на два класса: систематические и случайные. Систематические ошибки порождаются причинами, действующими регулярно, в определенном направлении. Чаще всего эти ошибки можно изучить и определить количественно. Систематические ошибки находят, калибруя измерительные приборы и сопоставляя опытные данные с изменяющимися внешними условиями (например, при градуировке термометра по реперным точкам, при сравнении с эталонным прибором). Если систематические ошибки вызываются внешними условиями (переменной температуры, сырья и т. д.), следует компенсировать их влияние. Как это делать, мы покажем в следующих параграфах. Случайными ошибками называются те, которые появляются нерегулярно, причины возникновения которых неизвестны и которые невозможно учесть заранее.

Систематические и случайные ошибки состоят из множества элементарных ошибок. Для того чтобы исключить инструментальные ошибки, следует проверять приборы перед опытом, иногда в течение опыта и обязательно после опыта. Ошибки при проведении самого опыта возникают вследствие неравномерного нагрева реакционной среды, разного способа перемешивания и т. п.

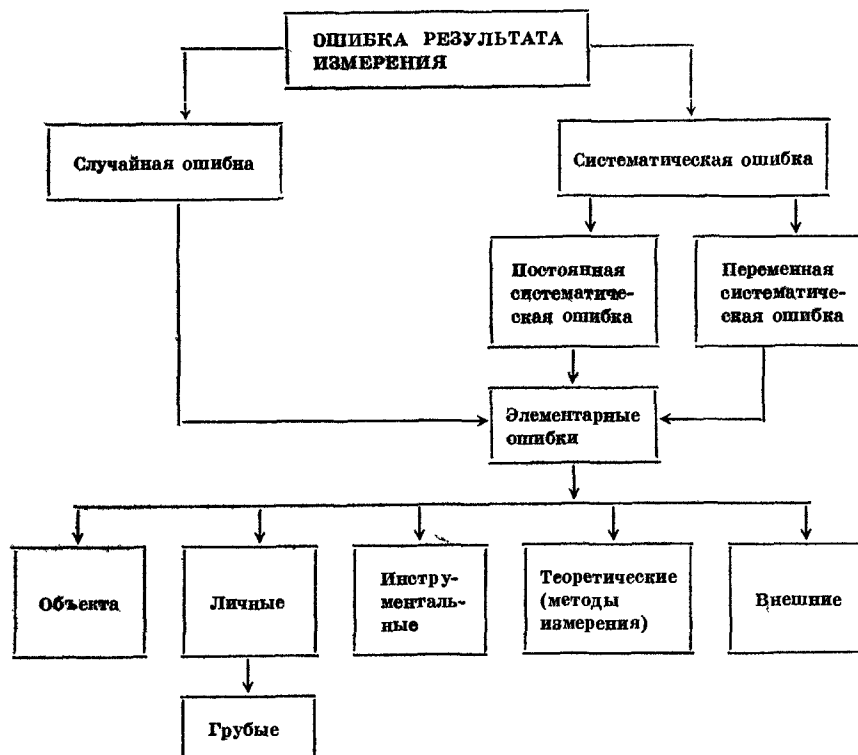


Рис. 24. Компоненты ошибки измерений

При повторении опытов такие ошибки могут вызвать большой разброс экспериментальных результатов.

Как мы уже знаем, очень важно исключить из экспериментальных данных грубые ошибки, так называемый брак при повторных опытах. Ни в коем случае, конечно, нельзя вносить поправки самовольно. Для отброса ошибочных опытов существуют правила.

Для определения брака используют, например, критерий Стьюдента

$$\frac{y - \bar{y}}{s} \geq t.$$

Значение t берут из таблицы t -распределения Стьюдента. Опыт считается бракованным, если экспериментальное значение критерия t по модулю больше табличного значения.

Пример 2. Обратимся к конкретному примеру. При исследовании процесса коррозии четыре повторных опыта показали следующие значения скорости коррозии: 3,580, 2,370, 2,710 и 2,761 $\text{мг/см}^2 \cdot \text{час}$.

Результат первого опыта поставлен под сомнение, так как он выделяется на фоне остальных трех опытов.

Исключим первый опыт из расчета и по остальным произведем вычисление среднего арифметического и стандарта

$$\bar{y} = \frac{2,370 + 2,710 + 2,761}{3} = 2,613;$$

$$s = \sqrt{\frac{(-0,243)^2 + 0,097^2 + 0,148^2}{3-1}} = \sqrt{\frac{0,090}{2}} \approx 0,21.$$

Если произведем проверку по критерию Стьюдента, то получим

$$\frac{3,580 - 2,613}{0,21} = \frac{0,967}{0,21} = 4,6.$$

При числе степеней свободы $f=2$ и уровне значимости 0,05 $t=4,303$. Экспериментальное значение больше табличного, поэтому сомнительный результат можно считать браком.

Можно воспользоваться еще и другими критериями.

Пусть имеется n повторных наблюдений y_1, y_2, \dots, y_n и возникает подозрение, что i -е наблюдение несовместимо с остальными. Подсчитаем среднее значение \bar{y} и ошибку $s_{(y)}$:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}; \quad s_{(y)} = \sqrt{\frac{\sum_{i=1}^n (\bar{y} - y_i)^2}{n-1}}.$$

Найдем относительное отклонение для i -го определения

$$r_i = \frac{y_i - \bar{y}}{s_{(y)} \sqrt{\frac{n-1}{n}}}.$$

Если найденное значение r_i для любого i -го измерения не превосходит по абсолютной величине табличного значения r для выбранного уровня значимости и числа степеней свободы $f=n-2$, то можно принять гипотезу об однородности результатов наблюдений.

При больших значениях числа степеней свободы r -распределение весьма близко к нормальному распределению. В этих случаях для проверки однородности наблюдений можно пользоваться трехсигмовым критерием 3σ , полагая, что выборочная дисперсия хорошо характеризует генеральную дисперсию. Если ни одно из отклонений при большом числе измерений не превосходит по абсолютной величине 3σ , то допустимо считать все наблюдения совместимыми. Критерий r_i применим для оценки любого i -го наблюдения. Для оценки специально выбранных наблюдений, минималь-

ных или максимальных, используется распределение максимального отклонения:

$$r_{\max} = \frac{y_{\max} - \bar{y}}{s_{\{y\}} \sqrt{\frac{n-1}{n}}}, \quad r_{\min} = \frac{\bar{y} - y_{\min}}{s_{\{y\}} \sqrt{\frac{n-1}{n}}}.$$

В табл. 8.3 приведены значения r_{\max} (r_{\min}) для уровня значимости 0,01 и 0,05 и числа степеней свободы от 1 до 23.

Таблица 8.3

Значения r_{\max} (r_{\min}) для уровней значимости 0,01 и 0,05 [2]

Число степеней свободы	Уровень значимости		Число степеней свободы	Уровень значимости		Число степеней свободы	Уровень значимости	
	0,01	0,05		0,01	0,05		0,01	0,05
1	1,414	1,412	9	2,606	2,343	17	2,932	2,600
2	1,723	1,689	10	2,663	2,387	18	2,959	2,623
3	1,955	1,869	11	2,714	2,426	19	2,984	2,644
4	2,130	1,996	12	2,759	2,461	20	3,008	2,644
5	2,265	2,093	13	2,800	2,493	21	3,030	2,683
6	2,374	2,172	14	2,837	2,523	22	3,051	2,701
7	2,464	2,237	15	2,871	2,551	23	3,071	2,717
8	2,540	2,294	16	2,903	2,557			

Если рассчитанное значение r_{\max} (r_{\min}) превышает табличное, то оцениваемый результат может быть отнесен к грубым и не включаться в расчет.

Предположим, что в первом опыте по экстракционному разделению циркония и гафния получены значения параллельных опытов 12,15; 10,86; 16,00. Оценим последнее значение:

$$\bar{y} = 13,00; \quad s_{\{y\}}^2 = 6,172; \quad s_{\{y\}} = 2,48;$$

$$r_{\max} = \frac{16,0 - 13,0}{2,02} = 1,48.$$

Табличное значение r_{\max} для уровня значимости 0,05 и одной степени свободы равно 1,412. Наблюдение 16,00 может быть признано грубым.

Здесь показаны самые простые приемы, которыми можно пользоваться при исключении ошибочных результатов [1, 2]. Рекомендуем вам познакомиться с другими приемами [2].

Отметим еще, что повторные опыты нельзя путать с повторными измерениями в одном опыте. Такие измерения часто делаются и являются полезными, но не могут заменить повторных опытов.

8.4. Дисперсия параметра оптимизации

Мы рассмотрели, как подсчитывается дисперсия в каждом опыте, т. е. в каждой горизонтальной строке матрицы планирования.

Матрица планирования состоит из серии опытов, и дисперсия всего эксперимента получается в результате усреднения дисперсий всех опытов. По терминологии, принятой в планировании эксперимента, речь идет о подсчете дисперсии параметра оптимизации $s_{\{y\}}^2$ или, что то же самое, дисперсии воспроизводимости эксперимента $s_{\text{воспр}}^2$.

Вы помните, что дисперсия в каждом опыте, состоящем из n повторных наблюдений, подсчитывается по формуле

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}.$$

При подсчете дисперсии параметра оптимизации квадрат разности между значением y_q в каждом опыте и средним значением из n повторных наблюдений \bar{y} нужно просуммировать по числу опытов в матрице N , а затем разделить на $N(n-1)$.

Так мы приходим к формуле

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{q=1}^n (y_{iq} - \bar{y})^2}{N(n-1)},$$

где $i=1, 2, \dots, N$; $q=1, 2, \dots, n$.

Такой формулой можно пользоваться в случаях, когда число повторных опытов одинаково во всей матрице.

Для двух повторных опытов формула принимает совсем простой вид

$$s_{\{y\}}^2 = \frac{2 \sum_{i=1}^N (y_{i1} - y_{i2})^2}{N}.$$

Дисперсию воспроизводимости проще всего рассчитывать, когда соблюдается равенство числа повторных опытов во всех экспериментальных точках. На практике часто приходится сталкиваться со случаями, когда число повторных опытов различно. Это происходит вследствие отброса грубых наблюдений, неуверенности экспериментатора в правильности некоторых результатов (в таких случаях возникает желание еще и еще раз повторить опыт) и т. п.

Тогда при усреднении дисперсий приходится пользоваться средним взвешенным значением дисперсий, взятым с учетом числа

$$s^2_{(y)} = \frac{s^2_1 f_1 + s^2_2 f_2 + \dots + s^2_N f_N}{f_1 + f_2 + \dots + f_N} = \frac{\sum_{i=1}^N f_i s^2_i}{\sum_{i=1}^N f_i},$$

где s^2_1 — дисперсия первого опыта, s^2_2 — дисперсия второго опыта и т. д., f_1 — число степеней свободы в первом опыте, равное числу параллельных опытов n_1 минус 1, т. е. $f_1 = n_1 - 1$; f_2 — число степеней свободы во втором опыте и т. д.

Число степеней свободы средней дисперсии принимается равным сумме чисел степеней свободы дисперсий, из которых она вычислена.

Обращаем ваше внимание на то, что вы совершите ошибку, если возьмете среднее значение дисперсий без учета числа степеней свободы, а также если возьмете среднее значение стандартных отклонений. Стандартные отклонения нужно возвести в квадрат и затем взять взвешенное среднее, как указано выше.

Случай с неравным числом наблюдений, который мы рассмотрели выше, связан с нарушением ортогональности матрицы. Поэтому здесь нельзя использовать расчетные формулы для коэффициентов, приведенные в гл. 6. и 7.

Этот вопрос мы рассмотрим в гл. 9, когда будем рассказывать о расчете дисперсии адекватности.

Итак, вы имеете формулы для расчета дисперсий, воспроизводимости эксперимента. Казалось бы, все обстоит хорошо. И все же... вы уже, наверное, чувствуете, что речь пойдет о «некоторых» ограничениях. Действительно, это так. Формулами

$$s^2_{(y)} = \frac{\sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{N(n-1)} \quad \text{и} \quad s^2_{(y)} = \frac{\sum_{i=1}^N f_i s^2_i}{\sum_{i=1}^N f_i}$$

можно пользоваться только в том случае, если дисперсии однородны. Последнее означает, что среди всех суммируемых дисперсий нет таких, которые бы значительно превышали все остальные.

Одним из требований регрессионного анализа, с которым вы познакомитесь в следующей главе, является однородность дисперсий.

Вы, конечно, понимаете, что для проверки неоднородности дисперсий нужны количественные критерии. Для того чтобы познакомиться с ними, нужно перейти к следующему параграфу.

8.5. Проверка однородности дисперсий

Проверка однородности дисперсий производится с помощью различных статистических критериев. Простейшим из них является критерий Фишера, предназначенный для сравнения двух дисперсий. Критерий Фишера (F -критерий) представляет собою отношение большей дисперсии к меньшей. Полученная величина сравнивается с табличной величиной F -критерия (см. стр. 152).

Если полученное значение дисперсионного отношения больше приведенного в таблице для соответствующих степеней свободы и выбранного уровня значимости, это означает, что дисперсии значимо отличаются друг от друга, т. е. что они неоднородны.

Пример 3. Пусть $s^2_1 = 5,14$, $n_1 = 7$ и $f_1 = 6$; $s^2_2 = 0,324$ для $n_2 = 6$ и $f_2 = 5$. В данном примере отношение дисперсий равно $5,14/0,324 = 15,9$ при $f_1 = 6$ и $f_2 = 5$. Почти в каждом пособии по математической статистике помещена таблица отношений дисперсий для различных степеней свободы и различного уровня значимости. Имеется она и в нашей книге. Выбираем наиболее популярный уровень значимости 0,05. В таблице по горизонтали отложены числа степеней свободы для большей дисперсии f_1 , а по вертикали — числа степеней свободы для меньшей дисперсии f_2 . Для $f_1 = 6$ и $f_2 = 5$ $F_{\text{таб}} = 4,40$. Это значит: вероятность того, что экспериментальное значение F будет больше чем 4,40, равна 0,05 или 5%. Наше $F_{\text{пол}} = 15,90$. Оно значительно превышает табличное значение.

Мы проверяли гипотезу об однородности дисперсий. Наша гипотеза состояла в том, что обе группы экспериментальных данных получены из одной и той же совокупности и дают одинаковое рассеяние. Установили, что одна дисперсия значимо отличается от другой (для выбранного уровня значимости).

Если сравниваемое количество дисперсий больше двух и одна дисперсия значительно превышает остальные, можно воспользоваться критерием Кохрена. Этот критерий пригоден для случаев, когда во всех точках имеется одинаковое число повторных опытов. При этом подсчитывается дисперсия в каждой горизонтальной строке матрицы

$$s^2 = \frac{\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{n-1},$$

а затем из всех дисперсий находится наибольшая s^2_{max} , которая делится на сумму всех дисперсий. Критерий Кохрена — это отношение максимальной дисперсии к сумме всех дисперсий

$$G = \frac{s^2_{\text{max}}}{\sum_{i=1}^N s^2_i}.$$

С этим критерием связаны числа степеней свободы $f_1 = n - 1$ и $f_2 = N$. Гипотеза об однородности дисперсий подтверждается,

если экспериментальное значение критерия Кохрена не превышает табличного значения. Тогда можно усреднять дисперсии и пользоваться формулой

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{N(n-1)}.$$

Пример 4. В начале главы, показывая, как нужно оформлять журнал, мы привели матрицу 2^3 с двумя повторными опытами. Мы сказали: вот с такой таблицей можно приступать к обработке экспериментальных данных. Воспользуемся этой таблицей для расчета дисперсии воспроизводимости. Перепишем ее так, чтобы было удобно производить расчет (табл. 8.4).

Таблица 8.4

Расчет дисперсии воспроизводимости

Номер опыта	Матрица планирования	y'	y''	\bar{y}	Δy	$(\Delta y)^2$	s_i^2
1	(1)	80,23	81,93	81,08	-0,85	0,722	1,444
2	a	86,50	84,80	85,65	0,85	0,722	1,444
3	b	82,45	82,10	82,27	0,18	0,031	0,062
4	ab	89,50	91,30	90,40	-0,90	0,810	1,620
5	c	85,10	84,80	84,95	0,15	0,023	0,046
6	ac	90,30	89,60	89,95	0,35	0,123	0,246
7	bc	85,60	84,90	85,25	0,35	0,123	0,246
8	abc	88,02	88,48	88,25	-0,23	0,053	0,106
$\sum_{i=1}^N (\Delta y_i)^2$						2,607	

Дисперсия в каждом опыте равна

$$s^2 = \frac{\sum_{i=1}^2 (y_i - \bar{y})^2}{2-1} = 2(\Delta y^2).$$

Максимальная дисперсия оказалась в опыте № 4.

Экспериментальный критерий Кохрена равен $G=1,620/5,214=0,31$. Табличный критерий Кохрена равен: $G=0,68$. Экспериментальный критерий Кохрена не превышает значения табличного. Гипотеза об однородности дисперсий подтверждается.

Дисперсия воспроизводимости равна

$$s_{\{y\}}^2 = \frac{2 \sum_{i=1}^8 (y_{iq} - \bar{y}_i)^2}{8} = \frac{2 \cdot 2,607}{8} = 0,652.$$

Пример 5. Теперь обратимся к примеру с различным числом повторных опытов (табл. 8.5)

Проведем подсчет дисперсии в каждом опыте и дисперсию воспроизводимости (если не возникнет предположение, что дисперсии неоднородны).

Таблица 8.5

Матрица планирования 2^{3-1} с различным числом повторных опытов *

Номер опыта	Матрица планирования	y^I	y^{II}	y^{III}	y^{IV}	y	Δy^I
1	c	87,31	86,01	—	—	86,66	0,65
2	abc	92,3	91,8	—	—	92,05	0,25
3	b	87,2	88,7	87,5	88,0	87,85	-0,65
4	a	84,0	84,9	84,2	—	84,37	-0,37

* y — выход реакции, %.

Таблица 8.5 (окончание)

Номер опыта	Δy^{II}	Δy^{III}	Δy^{IV}	$(\Delta y^I)^2$	$(\Delta y^{II})^2$	$(\Delta y^{III})^2$	$(\Delta y^{IV})^2$	f
1	-0,65			0,422	0,422			1
2	-0,25			0,062	0,062			1
3	0,85	-0,35	0,15	0,422	0,723	0,122	0,022	3
4	0,53	-0,17		0,137	0,281	0,029		2

$$s_1^2 = (0,422 + 0,422)/(2-1) = 0,844;$$

$$s_2^2 = (0,062 + 0,062)/(2-1) = 0,124;$$

$$s_3^2 = (0,422 + 0,723 + 0,122 + 0,022)/(4-1) = 0,429;$$

$$s_4^2 = (0,137 + 0,281 + 0,029)/(3-1) = 0,223;$$

$$s_{\{y\}}^2 = \frac{0,844 \cdot 1 + 0,124 \cdot 1 + 0,429 \cdot 3 + 0,223 \cdot 2}{1+1+3+2} = \frac{2,701}{7} = 0,386.$$

В данном примере не возникает предположение о неоднородности дисперсий, поскольку все они имеют одинаковый порядок.

Если возникает предположение о наличии неоднородности, следует попытаться его проверить. Для этой цели можно воспользоваться критерием Бартлетта. По уже знакомой вам формуле подсчитывается дисперсия воспроизводимости

$$s_{\{y\}}^2 = \sum_{i=1}^N f_i s_i^2 / \sum_{i=1}^N f_i.$$

Далее находится величина

$$\frac{1}{c} \left(f \lg s_{\{y\}}^2 - \sum_{i=1}^N f_i \lg s_i^2 \right),$$

$$\text{где } c = 0,4343 \left[1 + \frac{1}{3(N-1)} \left(\sum_{i=1}^N \frac{1}{f_i} - \frac{1}{f} \right) \right].$$

Здесь число степеней свободы равно $N-1$, где N — число сравниваемых дисперсий. При планировании эксперимента типа 2^k это число равно числу опытов в матрице.

Бартлет показал, что величина $\frac{1}{c} \left(f \lg s_{\{y\}}^2 - \sum_{i=1}^N f_i \lg s_i^2 \right)$ приближенно подчиняется χ^2 -распределению с $(N-1)$ степенями свободы. Значимость критерия Бартлета проверяется обычным способом.

Критерий Бартлета базируется на нормальном распределении. Если имеются отклонения от нормального распределения, то проверка неоднородности дисперсий может привести к ошибочным результатам.

Пример 6. Предлагаем вашему вниманию следующую задачу. В четырех опытах с разным числом повторных наблюдений получены результаты, приведенные в табл. 8.6.

Таблица 8.6

Исходные данные для расчета критерия Бартлета

Номер опыта	s_i^2	f_i	Номер опыта	s_i^2	f_i
1	3,50	4	3	5,88	3
2	4,22	5	4	11,36	3

Рассчитаем $s_{\{y\}}^2$ и воспользуемся критерием Бартлета, а затем ответим на вопрос, верна ли гипотеза об однородности дисперсий. По данным табл. 8.6 мы получаем: $\sum_{i=1}^N f_i = 15$ и $s_{\{y\}}^2 = 5,79$. Находим величину c

$$c = 0,4343 \left[1 + \frac{1}{3(4-1)} \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{3} + \frac{1}{3} - \frac{1}{15} \right) \right] = 0,4850.$$

Теперь мы можем определить χ^2

$$\chi^2 = \frac{1}{0,4850} (15 \lg 5,79 - 4 \lg 3,50 - 5 \lg 4,22 - 3 \lg 5,88 - 3 \lg 11,36) = 1,567.$$

Экспериментальное значение χ^2 -критерия равно 1,567. Табличное значение для трех степеней свободы и уровня значимости 0,05 равно 7,815, и мы приходим к выводу, что дисперсии однородны.

Приступать к расчету ошибки воспроизводимости, к регрессионному анализу (а также к дисперсионному анализу, который мы не рассматриваем в этой книге) можно только после того, как дисперсии выдержали проверку на однородность. Экспериментаторы часто пренебрегают такой проверкой, объясняя это трудоемкостью расчетов и сложностью критерия Бартлета.

Экспериментаторам, которым претит кропотливая работа при экспериментальных расчетах, можно предложить использование F -критерия даже в тех случаях, когда число дисперсий больше двух. Делается это следующим образом. Из всех дисперсий выделяются наибольшая и наименьшая. По F -критерию производится проверка, значимо ли они различаются между собой. Ясно, что если наибольшая и наименьшая дисперсии не отличаются значимо, то дисперсии, имеющие промежуточные значения, также не могут значимо отличаться друг от друга. Тогда всю группу дисперсий можно считать принадлежащей к единой совокупности. В таких случаях нет надобности применять критерий Бартлета.

Мы показали вам, как нужно проверять гипотезу об однородности дисперсий. Вы теперь знаете, какими формулами нужно пользоваться, если гипотеза об однородности дисперсий верна. А что же делать экспериментатору, если дисперсии все-таки оказались неоднородными? В таких случаях часто оказывается полезным изменение масштаба для параметра оптимизации. При этом вводится некоторая математическая функция от параметра оптимизации, например квадратный корень или логарифм.

Использование таких методов выходит за рамки элементарного анализа, и в случае необходимости экспериментатору целесообразно обращаться за советом к специалисту по планированию эксперимента.

8.6. Рандомизация

Сотри случайные черты —
И ты увидишь: мир прекрасен.
А. Блок

Чтобы исключить влияние систематических ошибок, вызванных внешними условиями (переменной температуры, сырья, лаборанта и т. д.), рекомендуется случайная последовательность при постановке опытов, запланированных матрицей. Опыты необходимо рандомизировать во времени. Термин «рандомизация» происходит от английского слова random — случайный. Почему рандомизация опытов важна, мы попытаемся показать на следующем примере.

Пример 7. В табл. 8.7 приведена матрица 2^3 , полученная из матрицы 2^2 обычным способом: два раза повторен план 2^2 , причем в первых четырех опытах x_3 имеет верхнее значение, а в последних четырех опытах — нижнее значение. Допустим, что экспериментатор может поставить в первый день четыре опыта и во второй день также четыре опыта.

Можно ли опыты ставить подряд и в первый день реализовать опыты № 1, 2, 3 и 4, а во второй — 5, 6, 7 и 8? Ставя опыты подряд, вы разбиваете матрицу на две части или на два блока: в первый блок входят опыты № 1, 2, 3 и 4, во второй — № 5, 6, 7 и 8. Если внешние условия первого дня каким-то образом отличались от внешних условий второго дня, то это способствовало возникновению некоторой систематической ошибки. Обозначим

Матрица 2^3 , нерандомизированная во времени

Номер опыта	x_1	x_2	x_3	y	Номер опыта	x_1	x_2	x_3	y
1	+	+	+	y_1	5	+	+	—	y_5
2	—	—	+	y_2	6	—	—	—	y_6
3	+	—	+	y_3	7	+	—	—	y_7
4	—	+	+	y_4	8	—	+	—	y_8

$$b_3 = \frac{1}{8} [(y_1 + \varepsilon) + (y_2 + \varepsilon) + (y_3 + \varepsilon) +$$

$$+(y_4 + \varepsilon) - y_5 - y_6 - y_7 - y_8] \rightarrow \beta_3 + \frac{\varepsilon}{2},$$

где β_3 — истинное значение коэффициента при x_3 . Таким образом, возможное различие во внешних условиях смешалось с величиной линейного коэффициента b_3 и исказило это значение. В такой последовательности опыты ставить нельзя. Опыты нужно рандомизировать во времени, т. е. придать последовательности опытов случайный характер.

- Приведем простой пример рандомизации условий эксперимента. В полном факторном эксперименте 2^3 предполагается каждое значение параметра оптимизации определять по двум параллельным опытам. Нужно случайно расположить всего 16 опытов. Присвоим параллельным опытам номера с 9 по 16, и тогда опыт № 9 будет повторным по отношению к первому опыту, десятый — ко второму и т. д. Следующий этап рандомизации — использование таблицы случайных чисел. Обычно таблица случайных чисел приводится в руководствах по математической статистике. Фрагмент таблицы помещен на стр. 135. В случайном месте таблицы выписываются числа с 1 по 16 с отбрасыванием чисел больше 16 и уже выписанных. В нашем случае, начиная с четвертого столбца, можно получить такую последовательность:

2; 15; 9; 5; 12; 14; 8; 13; 16; 1; 3; 7; 4; 6; 11; 10.

Это значит, что первым реализуется опыт № 2, вторым — опыт № 7 и т. д.

Выборную случайным образом последовательность опытов не рекомендуется нарушать,

Фрагмент таблицы случайных чисел [2]

56	66	25	32	38	64	70	26	27	67	77	40	04	34	63	98	99	89	31	46	12	90	50	28	96
88	40	52	02	29	82	69	34	50	21	74	00	91	27	52	98	72	03	45	65	30	89	71	45	91
87	63	88	23	62	51	07	69	59	02	89	49	14	98	53	41	92	36	07	76	85	37	84	37	47
32	25	21	15	08	82	34	57	57	35	22	03	33	48	84	37	37	29	38	37	89	76	25	09	69
44	61	88	23	13	01	59	47	64	04	99	59	96	20	30	87	31	33	69	45	58	48	00	83	48
94	44	08	67	79	41	61	41	15	60	11	88	83	24	82	24	07	78	61	89	42	58	88	22	16
13	24	09	00	65	46	38	61	12	90	62	41	11	59	85	18	42	61	29	88	76	04	21	80	
78	27	84	05	99	85	75	67	80	05	57	05	71	70	31	31	99	99	06	96	53	99	25	13	
42	39	30	02	34	99	46	68	45	15	19	74	15	50	17	44	80	13	86	38	40	45	82	13	
04	52	43	96	38	13	83	80	72	34	20	84	56	19	49	59	14	85	42	99	71	16	34	33	
82	85	77	30	16	69	32	46	46	30	84	20	68	72	98	94	62	63	59	44	00	89	06	15	
38	48	84	88	24	55	46	48	60	06	90	08	83	83	93	40	90	88	25	26	85	74	55	80	
91	19	05	68	22	58	04	63	21	16	23	38	25	43	32	98	94	65	35	35	16	91	07	12	
54	81	87	21	31	40	46	17	62	63	99	71	14	12	64	51	68	50	60	78	22	69	51	98	
65	43	75	12	91	20	36	25	57	92	33	65	95	48	75	00	06	65	25	90	16	29	34	14	
49	98	71	31	80	59	57	32	43	07	85	06	64	75	27	29	17	06	11	30	68	70	97	87	
03	98	68	89	39	71	87	32	14	99	42	10	25	37	30	08	27	75	43	97	54	20	69	93	
56	04	21	34	92	89	81	52	15	12	84	11	12	66	87	48	21	06	86	08	35	39	52	28	
48	09	36	95	36	20	82	53	32	89	92	68	50	88	17	37	92	02	23	43	63	24	69	80	
23	97	10	96	57	74	07	95	26	44	93	08	43	30	41	86	45	74	33	78	84	33	38	76	
43	97	55	45	98	35	69	45	96	80	46	26	39	96	33	60	20	73	30	79	17	19	03	47	
40	05	08	50	79	89	58	19	86	48	27	98	99	24	08	94	19	15	81	29	82	14	35	88	
66	97	10	69	02	25	36	43	71	76	00	67	56	12	69	07	89	55	63	31	50	72	20	33	
15	62	38	72	92	03	76	09	30	75	77	80	04	24	54	67	60	10	79	26	21	60	03	48	
77	81	15	14	67	55	24	22	20	55	36	93	07	69	37	72	22	43	46	32	56	15	75	25	
18	87	05	09	96	45	14	72	41	46	12	67	46	72	02	59	06	17	49	12	73	28	23	52	
08	58	53	63	66	13	07	04	48	71	39	07	46	96	40	20	86	79	11	81	74	11	15	23	
16	07	79	57	61	42	19	68	15	12	60	21	53	12	07	04	99	88	22	39	75	16	69	13	
84																								

8.7. Разбиение матрицы типа 2^k на блоки

Если экспериментатор располагает сведениями о предстоящих изменениях внешней среды, сырья, аппаратуры и т. п., то целесообразно планировать эксперимент таким образом, чтобы эффект влияния внешних условий был смешан с определенным взаимодействием, которое не жалко потерять. Так, при наличии двух партий сырья матрицу 2^3 можно разбить на два блока таким образом, чтобы эффект сырья сказался на величине трехфакторного взаимодействия. Тогда все линейные коэффициенты и парные взаимодействия будут освобождены от влияния неоднородности сырья (табл. 8.8).

В этой матрице при составлении блока 1 отображены все строки, для которых $x_1x_2x_3=+1$, а при составлении блока 2 — все строки, для которых $x_1x_2x_3=-1$. Различие в сырье можно рассматривать как новый фактор x_4 . Тогда матрица 2^3 , разбитая на два блока, представляет собой полуреплику 2^{4-1} с определяющим контрастом $1=x_1x_2x_3x_4$.

Таблица 8.8

Разбиение матрицы 2^3 на два блока

Блок	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	+	—	—	+	+	—	—	+	$y_1+\varepsilon$
	+	+	—	—	—	—	+	+	$y_2+\varepsilon$
	+	—	+	—	—	+	—	+	$y_3+\varepsilon$
	+	+	+	+	+	+	+	+	$y_4+\varepsilon$
2	+	—	—	—	+	+	+	—	y_5
	+	+	—	+	—	+	—	—	y_7
	+	—	+	+	—	—	+	—	y_7
	+	+	+	—	+	—	—	—	y_8

Мы предлагаем вам для данной матрицы (табл. 8.8) рассчитать все коэффициенты и посмотреть, какие коэффициенты смешаны с эффектом сырья:

$$b_0 = \frac{1}{8} [(y_1 + \varepsilon) + (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) + y_5 + y_6 + y_7 + y_8],$$

$$b_0 \rightarrow \beta_0 + \frac{\varepsilon}{2};$$

$$b_1 = \frac{1}{8} [-(y_1 + \varepsilon) + (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) -$$

$$- y_5 + y_6 - y_7 + y_8],$$

$$b_1 \rightarrow \beta_1;$$

$$b_2 = \frac{1}{8} [-(y_1 + \varepsilon) - (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) -$$

$$- y_5 - y_6 + y_7 + y_8],$$

$$b_2 \rightarrow \beta_2;$$

$$b_3 = \frac{1}{8} [(y_1 + \varepsilon) - (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) -$$

$$- y_5 + y_6 + y_7 - y_8],$$

$$b_3 \rightarrow \beta_3;$$

$$b_{12} = \frac{1}{8} [(y_1 + \varepsilon) - (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) +$$

$$+ y_5 - y_6 - y_7 + y_8],$$

$$b_{12} \rightarrow \beta_{12};$$

$$b_{13} = \frac{1}{8} [-(y_1 + \varepsilon) - (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) +$$

$$+ y_5 + y_6 - y_7 - y_8],$$

$$b_{13} \rightarrow \beta_{13};$$

$$b_{23} = \frac{1}{8} [-(y_1 + \varepsilon) + (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) +$$

$$+ y_5 - y_6 + y_7 - y_8],$$

$$b_{23} \rightarrow \beta_{23};$$

$$b_{123} = \frac{1}{8} [(y_1 + \varepsilon) + (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) -$$

$$- y_5 - y_6 - y_7 - y_8],$$

$$b_{123} \rightarrow \beta_{123} + \frac{\varepsilon}{2}.$$

Эффект сырья отразился на подсчете свободного члена b_0 и эффекта взаимодействия второго порядка b_{123} .

Аналогично можно разбить на два блока любой эксперимент типа 2^k . Главное — правильно выбрать взаимодействие, которым можно безболезненно пожертвовать. При отсутствии априорных сведений выбирают взаимодействие самого высокого порядка: $x_1x_2x_3$ для 2^3 , $x_1x_2x_3x_4$ для 2^4 , $x_1x_2x_3x_4x_5$ для 2^5 и т. д. Но если экспериментатору известно, что одно из парных взаимодействий лишено, например, физико-химического смысла, то можно пожертвовать парным взаимодействием.

В нашей практике встречалось много задач, в которых взаимодействия высокого порядка оказывались более значимыми, чем парные взаимодействия. Когда взаимодействие выбрано, в первый блок группируются все опыты, в которых это взаимодействие равно $+1$, а во второй, где оно равно -1 .

Теперь посмотрим, как можно разбить матрицу на четыре блока. Пусть нужно поставить эксперимент 2^4 . Заведомо известно, что

имеется четыре источника неоднородности, которые могут значительно исказить результаты эксперимента. При наличии четырех источников неоднородности нужно матрицу 2^4 разбить на четыре блока так, чтобы линейные эффекты были освобождены от влияния межблокового эффекта. Чтобы произвести разбиение матрицы 2^4 на четыре блока по четыре опыта в каждом, нужно выбрать три взаимодействия, которыми можно пожертвовать (число взаимодействий определяется числом степеней свободы, смешивающимися с различием между блоками: $f=4-1=3$). Два таких взаимодействия можно выбрать произвольно, а третье оказывается однозначно определенным по следующему правилу: нужно взять алгебраическое произведение первых двух выбранных взаимодействий и заменить единицей каждый множитель, стоящий в квадрате. Так, если двумя произвольно выбранными взаимодействиями являются парные x_1x_2 и x_3x_4 , то третьим будет $x_1x_2x_3x_4$. Если выбранными являются тройные $x_1x_2x_3$ и $x_2x_3x_4$, то третьим будет x_1x_4 . При разбиении матрицы 2^4 на четыре блока одно из парных взаимодействий окажется неизбежно смешанным с межблоковым эффектом

Пусть мы выбрали для смешивания три взаимодействия: $x_1x_2x_3$, $x_2x_3x_4$ и x_1x_4 . Включаем в первый блок те опыты, которые имеют четное количество букв, одинаковых с буквами, входящими в символы трех выбранных взаимодействий (при этом удобно пользоваться кодовым обозначением матрицы с помощью латинских букв).

При разбиении на блоки принято обозначать факторы заглавными латинскими буквами. Мы будем пользоваться этими обозначениями наряду с нашими x_j .

Опыт (1), где все факторы на нижних уровнях, удовлетворяет этому условию, так как имеется 0 общих букв со всеми взаимодействиями. Опыт bc также удовлетворяет этому условию, так как его символ имеет две общие буквы с $x_1x_2x_3$ (ABC) и $x_2x_3x_4$ (BCD) и ни одной с x_1x_4 (AD). Двумя другими, удовлетворяющими условию опытами, будут acd и abd , имеющие по две буквы со всеми взаимодействиями.

В результате получается блок 1 (табл. 8.9). Для определения состава следующего блока выбираем какое-либо неиспользованное испытание, например a , и умножаем на этот символ каждый член блока 1, получаем блок 2.

Блок 1	Блок 2	Блок 3	Блок 4
(1)	a	b	d
bc	abc	c	bcd
acd	cd	$abcd$	ac
abd	bd	ad	ab

Аналогичная операция проводится для определения состава блоков 3 и 4. Путем выбора неиспользованного испытания b получаем блок 3, используя d — блок 4. Запишем эту матрицу

Таблица 8.9
Разбиение матрицы 2^4 на четыре блока

ν	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
$x_1x_2x_3x_4$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$x_2x_3x_4$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$x_1x_2x_4$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$x_1x_3x_4$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$x_1x_2x_3$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_3x_4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_2x_4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_2x_3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_1x_4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_1x_3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_1x_2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
x_0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Буквенное обозначение матрицы	(1)	bc	acd	abd	a	abc	cd	bd	b	c	$abcd$	ad	d	bcd	ac	ab
Блок	1	2				3				4						

в кодовом обозначении $+1$ и -1 и проверим, какие взаимодействия смешаны с межблоковым эффектом.

В матрице табл. 8.9 можно видеть, что в каждом блоке для всех эффектов, за исключением смешанных, соблюдается равенство числа $+1$ и -1 . Следовательно, межблоковый эффект отразится на подсчете b_0 , b_{14} , b_{123} и b_{234} . Остальные коэффициенты регрессии освобождены от влияния источников неоднородности.

Матрицу типа 2^k можно разбить на количество блоков 2^n (n — степень двойки) при $n < k$. Так, матрица 2^3 разбивается на два блока по четыре опыта в каждом и на четыре блока по два опыта в каждом. Матрица 2^4 — на два блока по восемь опытов в каждом, на четыре блока по четыре опыта и на восемь блоков по два опыта и т. д. Мы не имеем возможности подробно останавливаться на этом вопросе. С разбиением матриц на блоки вы можете познакомиться в работе [3, 4]

8.8. Резюме

В этой главе мы обратили ваше внимание на то, что к опыту нужно тщательно готовиться: собрать и наладить опытную установку, проверить приборы, подготовить исходное сырье, разработать журнал. Тщательная подготовка к опыту будет способствовать уменьшению ошибки опыта. Ошибка опыта является суммарной величиной, состоящей из ряда ошибок: ошибок при измерении факторов, параметра оптимизации и ошибок при проведении опыта. Ошибки подразделяются на случайные и систематические. Для того чтобы компенсировать влияние систематических ошибок, опыты нужно рандомизировать во времени. Если экспериментатору заранее известны источники систематических ошибок, например известно количество различных партий сырья, следует разбивать матрицу планирования на блоки. При этом межблоковый эффект заведомо смешивается с взаимодействиями, которыми экспериментатор может пренебречь.

Особое внимание следует уделять проверке однородности дисперсий, так как это — одна из предпосылок, лежащих в основе регрессионного анализа. Для проверки однородности дисперсий можно использовать критерии Фишера, Кохрена или Бартлетта. Очень важно отбросить грубые наблюдения — брак при постановке повторных опытов.

Воспроизводимость эксперимента является одним из важнейших требований планирования эксперимента.

Литература

1. Н. Бейли. Статистические методы в биологии. М., ИЛ, 1962.
2. В. В. Налимов. Применение математической статистики при анализе вещества. М., Физматгиз, 1960.
3. Е. В. Маркова, А. Н. Лисенков. Планирование эксперимента в условиях неоднородностей. М., «Наука», 1973.
4. К. А. Браунли. Статистические исследования в производстве. М., ИЛ, 1949.

Глава девятая

ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

Когда все сущее, сменяясь каждый час,
В нестройный хор сливается вокруг нас,
Кто звуки мерные в порядке размещает,
Чьей речи верный ритм явился и тверд,
Кто единичное искусно обобщает
Объединяя все в торжественный аккорд?

Гете. Фауст

Тщательное, скрупулезное выполнение эксперимента, несомненно, является главным условием успеха исследования. Это общее правило, и планирование эксперимента не относится к исключениям.

Однако нам не безразлично, как обработать полученные данные. Мы хотим извлечь из них всю информацию и сделать соответствующие выводы. Как всегда, мы находимся между Сциллой и Харибдой. С одной стороны, не извлечь из эксперимента все, что из него следует, — значит пренебречь нелегким трудом экспериментатора. С другой стороны, сделать утверждения, не следующие из экспериментальных данных, — значит создавать иллюзии, заниматься самообманом (и обманом тоже, хотя и невольным).

Статистические методы обработки результатов позволяют нам не перейти разумной меры риска. Поэтому мы отводим эту главу для их рассмотрения [1—8].

9.1. Метод наименьших квадратов

Статистики разработали много разнообразных методов обработки результатов эксперимента. Но, пожалуй, ни один из них не может конкурировать по популярности, по широте приложений с методом наименьших квадратов (МНК), который был развит усилиями Лежандра и Гаусса более 150 лет назад.

Давайте попробуем разобраться в этом методе. Начнем с простого случая: один фактор, линейная модель. Интересующая нас функция отклика (которую мы будем также называть уравнением регрессии) имеет вид

$$y = b_0 + b_1 x_1.$$

Это хорошо известное вам уравнение прямой линии. Наша цель — вычисление неизвестных коэффициентов b_0 и b_1 . Мы провели экс-

перимент, чтобы использовать при вычислениях его результаты. Как это сделать наилучшим образом?

Если бы все экспериментальные точки лежали строго на прямой линии, то для каждой из них было бы справедливо равенство

$$y_i - b_0 - b_1 x_{1i} = 0,$$

где $i=1, 2, \dots, N$ — номер опыта. Тогда не было бы никакой проблемы. На практике это равенство нарушается и вместо него приходится писать

$$y_i - b_0 - b_1 x_{1i} = \xi_i,$$

где ξ_i — разность между экспериментальным и вычисленным по уравнению регрессии значениями y в i -й экспериментальной точке. Эту величину иногда называют невязкой.

Действительно, невязка возникает по двум причинам: из-за ошибки эксперимента и из-за непригодности модели. Причем эти причины смешаны и мы не можем, не получив дополнительной информации, сказать, какая из них преобладает.

Можно постулировать, что модель пригодна. Тогда невязка будет порождаться только ошибкой опыта. (Еще можно, конечно, постулировать, что ошибка опыта равна нулю. Тогда невязка будет связана только с пригодностью модели, и пригодной будет такая модель, для которой все невязки равны нулю.)

Обычно оценивают независимо ошибку опыта (помните предыдущую главу?) и проверяют пригодность модели.

Конечно, мы хотим найти такие коэффициенты регрессии, при которых невязки будут минимальны. Это требование можно записать по-разному. В зависимости от этого мы будем получать разные оценки коэффициентов. Вот одна из возможных записей

$$U = \sum_{i=1}^N \xi_i^2 = \min,$$

которая приводит к методу наименьших квадратов.

Возможен и метод наименьших кубов $\sum_{i=1}^N |\xi_i|^3 = \min$, так как

условие, которое мы выбираем, произвольно.

Беда заключается в том, что он хуже МНК с другой точки зрения: мы будем получать оценки коэффициентов со значительно меньшей точностью. Да и в вычислительном отношении этот путь сложнее.

Существует и метод, в котором минимизируется сумма модулей (абсолютных величин) невязок. Но этот путь связан с дополнительными вычислительными трудностями. Условие МНК — это удачный компромисс.

В последнее время были предложены другие подходы. Можно, например, минимизировать модуль максимальной невязки. Это записывается так:

$$\min_i \max_i |\xi_i|.$$

Предложений можно сделать сколько угодно, но мы не будем более на них останавливаться и перейдем непосредственно к МНК.

Когда мы ставим эксперимент, то обычно стремимся провести больше (во всяком случае не меньше) опытов, чем число неизвестных коэффициентов. Поэтому система линейных уравнений

$$\xi_i = y_i - b_0 - b_1 x_{1i}$$

оказывается переопределенной и часто противоречивой (т. е. она может иметь бесконечно много решений или может не иметь решений). Переопределенность возникает, когда число уравнений больше числа неизвестных; противоречивость — когда некоторые из уравнений несовместимы друг с другом.

Только если все экспериментальные точки лежат на прямой, то система становится определенной и имеет единственное решение.

МНК обладает тем замечательным свойством, что он делает определенной любую произвольную систему уравнений. Он делает число уравнений равным числу неизвестных коэффициентов.

Наше уравнение регрессии имеет вид

$$y = b_0 + b_1 x_1.$$

В нем два неизвестных коэффициента. Значит, применяя МНК, мы получим два уравнения.

Давайте попробуем их получить.

Мы писали

$$U = \sum_{i=1}^N \xi_i^2 = \min.$$

Это соотношение можно записать иначе

$$U = \sum_{i=1}^N \xi_i^2 = \sum_{i=1}^N (y_i - b_0 - b_1 x_{1i})^2 = \min.$$

Вы, конечно, помните из курса математики, что минимум некоторой функции, если он существует, достигается при одновременном равенстве нулю частных производных по всем неизвестным, т. е.

$$\frac{\partial U}{\partial b_0} = 0, \quad \frac{\partial U}{\partial b_1} = 0.$$

Вот откуда берутся наши уравнения для определения коэффициентов. Теперь, как говорится, дело техники:

$$-2 \sum_{i=1}^N (y_i - b_0 - b_1 x_{1i}) = 0, \quad -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_{1i}) x_{1i} = 0.$$

Для вычислений удобно раскрыть скобки и провести простые преобразования, которые дают

$$Nb_0 + \sum_{i=1}^N x_{1i} b_1 = \sum_{i=1}^N y_i, \quad \sum_{i=1}^N x_{1i} b_0 + \sum_{i=1}^N x_{1i}^2 b_1 = \sum_{i=1}^N y_i x_{1i}.$$

Окончательные формулы для вычисления коэффициентов регрессии, которые удобно находить с помощью определителей, имеют вид

$$b_0 = \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_{1i}^2 - \sum_{i=1}^N y_i x_{1i} \sum_{i=1}^N x_{1i}}{N \sum_{i=1}^N x_{1i}^2 - \left(\sum_{i=1}^N x_{1i} \right)^2};$$

$$b_1 = \frac{N \sum_{i=1}^N y_i x_{1i} - \sum_{i=1}^N y_i \sum_{i=1}^N x_{1i}}{N \sum_{i=1}^N x_{1i}^2 - \left(\sum_{i=1}^N x_{1i} \right)^2}.$$

Посмотрим теперь, как вычисляются суммы, входящие в эти формулы.

Результаты эксперимента представляются следующей матрицей (табл. 9.1). Для выполнения вычислений ее расширяют так, как представлено в табл. 9.2.

Вы, конечно, заметили, что в этой таблице сделано больше вычислений, чем требуется для расчета b_0 и b_1 . Эти «лишние» данные нужны для проверки правильности расчетов.

Возможны два способа проверки. Первый из условия

$$\sum_{i=1}^N (x_{1i} + y_i)^2 = \sum_{i=1}^N x_{1i}^2 + 2 \sum_{i=1}^N y_i x_{1i} + \sum_{i=1}^N y_i^2,$$

которое хорошо вам известно еще из школьной математики. (Оно должно выполняться не только для сумм, но и в каждой строчке таблицы.) Второй способ использует условие $\bar{y} = b_0 + b_1 \bar{x}_1$. Подставляя в это соотношение \bar{y} и \bar{x}_1 из последней строки таблицы и один из коэффициентов, можно найти другой коэффициент и сравнить с расчетным.

Вторая из проверок является наиболее полной, наиболее жесткой. Она проверяет не только вычисления сумм, но и вычисления коэффициентов.

Таблица 9.2

Расчетная таблица для вычисления коэффициентов регрессии

Номер опыта	x_1	y	x_1^2	$y x_1$	y^2	$x_1 + y$	$(x_1 + y)^2$
1	x_{11}	y_1	x_{11}^2	$y_1 x_{11}$	y_1^2	$x_{11} + y_1$	$(x_{11} + y_1)^2$
2	x_{12}	y_2	x_{12}^2	$y_2 x_{12}$	y_2^2	$x_{12} + y_2$	$(x_{12} + y_2)^2$
...
i	x_{1i}	y_i	x_{1i}^2	$y_i x_{1i}$	y_i^2	$x_{1i} + y_i$	$(x_{1i} + y_i)^2$
...
N	x_{1N}	y_N	x_{1N}^2	$y_N x_{1N}$	y_N^2	$x_{1N} + y_N$	$(x_{1N} + y_N)^2$
Σ	$\sum_{i=1}^N x_{1i}$	$\sum_{i=1}^N y_i$	$\sum_{i=1}^N x_{1i}^2$	$\sum_{i=1}^N y_i x_{1i}$	$\sum_{i=1}^N y_i^2$	—	$\sum_{i=1}^N (x_{1i} + y_i)^2$
Среднее значение	\bar{x}_1	\bar{y}					

Таблица 9.1

Условия и результаты опытов

Номер опыта	x_1	y
1	x_{11}	y_1
2	x_{12}	y_2
...
i	x_{1i}	y_i
...
N	x_{1N}	y_N

На практике используют обе проверки, чтобы в случае ошибок в таблице не считать зря коэффициенты.

Имейте в виду: никакая проверка не гарантирует вас от ошибок в записи исходных данных. Будьте внимательны!

Имейте в виду: никакие результаты вычислений нельзя ни использовать, ни даже обсуждать, пока они не проверены. Иначе вы рискуете впасть в заблуждение и, в лучшем случае, потерять время.

Ну вот мы и научились вычислять коэффициенты. Давайте нанесем исходные данные и полученное уравнение на график (рис. 25).

Выделим для удобства рассмотрения несколько экспериментальных точек и отрезков нашего уравнения в большем масштабе (рис. 26).

Мы выбрали пять экспериментальных точек, которые пронумеровали цифрами 1, 2, 3, 4, 5. Четвертая точка оказалась лежащей на линии. МНК состоит в том, чтобы минимизировать сумму квадратов отрезков, характеризующих расхождение между экспериментальными точками и полученным уравнением. Мы минимизировали сумму квадратов пунктирных отрезков.

Если бы наше уравнение регрессии имело вид

$$x_1 = b_0 + b_1 y,$$

то мы минимизировали бы сумму сплошных отрезков.

Во всех формулах тогда пришлось бы x_1 и y поменять местами и коэффициенты получились бы другими (если, конечно, не все невязки равны нулю).

Мы находим невязки по оси y , поэтому и минимизируется сумма квадратов вертикальных отрезков. Обе линии совпадут только в том случае, если все невязки равны нулю, т. е. если все экспериментальные точки лежат точно на прямой линии.

Теперь мы можем узнать, какая же получилась сумма квадратов невязок. Будем называть ее *остаточной суммой квадратов*.

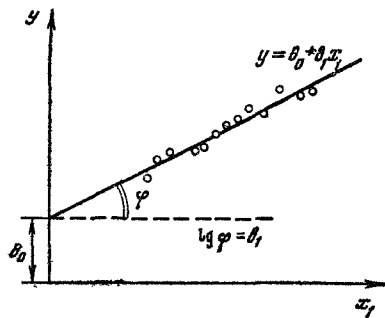


Рис. 25. Линейное уравнение регрессии

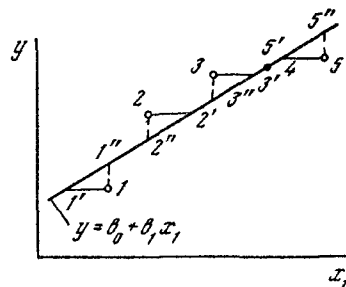


Рис. 26. Линейное уравнение регрессии (фрагмент)

Из рисунков видно, что для этого надо вычислить по уравнению значения y в условиях каждого опыта. Будем называть такое значение предсказанным и обозначать \hat{y} . Затем надо найти все невязки (отрезки), возвести их в квадрат и сложить (табл. 9.3).

Таблица 9.3

Расчет остаточной суммы квадратов

Номер опыта	y	\hat{y}	$\Delta y = y - \hat{y}$	Δy^2
1	y_1	\hat{y}_1	Δy_1	Δy_1^2
2	y_2	\hat{y}_2	Δy_2	Δy_2^2
...
i	y_i	\hat{y}_i	Δy_i	Δy_i^2
...
N	y_N	\hat{y}_N	Δy_N	Δy_N^2
				$\sum_{i=1}^N \Delta y_i^2$

Величина $\sum_{i=1}^N \Delta y_i^2$ и есть остаточная сумма квадратов, которую мы раньше обозначили $\sum_{i=1}^N \xi_i^2$. МНК гарантирует, что эта величина минимально возможная.

Итак, мы научились находить наилучшие в смысле МНК оценки коэффициентов линейного уравнения для одного фактора. Это, конечно, полезно, но нас интересуют многофакторные задачи.

Обобщение на многофакторный случай не связано с какими-либо принципиальными трудностями. Правда, вычисления значительно усложняются и требуют привлечения аппарата алгебры матриц. Рассмотрим его в следующей главе. А пока мы воспользуемся тем, что наши матрицы планирования ортогональны. Если вы забыли это понятие, то обратитесь к стр. 84 и повторите его. Далее будем рассматривать только этот случай, который позволяет резко упростить вычисления, что составляет одно из преимуществ планирования эксперимента.

Можно показать, что для любого числа факторов коэффициенты будут вычисляться по формуле

$$b_j = \frac{\sum_{i=1}^N y_i x_{ij}}{N}.$$

В этой формуле $j=0, 1, 2, \dots, k$ — номер фактора. Ноль записан для вычисления b_0 . Действительно, посмотрите на формулы для вычисления коэффициентов регрессии на стр. 144. В первой формуле

$$\sum_{i=1}^N x_{i1} = 0$$

в силу симметричности плана. Поэтому после сокращения формулы приобретают вид

$$b_0 = \frac{\sum_{i=1}^N y_i}{N}, \quad b_1 = \frac{\sum_{i=1}^N y_i x_{i1}}{\sum_{i=1}^N x_{i1}^2},$$

где

$$\sum_{i=1}^N x_{i1}^2 = N,$$

что совпадает с написанным выше.

Так как каждый фактор (кроме x_0) варьируется на двух уровнях $+1$ и -1 , то вычисления сводятся к приписыванию столбцу y знаков соответствующего фактору столбца и алгебраическому сложению полученных значений. Деление результата на число опытов в матрице планирования дает искомый коэффициент. Это очень простая формула, но вам необходимо научиться пользоваться ею безошибочно.

При вычислениях линейных моделей по дробным репликам никаких особенностей не появляется. Все точно так же. Допол-

нительные трудности возникают, если мы хотим найти коэффициенты неполного квадратного уравнения (если нас интересуют эффекты взаимодействия). Тогда уравнение регрессии будет иметь вид

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + b_{12}x_1x_2 + \dots + b_{k-1,k}x_{k-1}x_k.$$

Конечно, можно интересоваться не всеми эффектами взаимодействия, а только определенными. В полном факторном эксперименте можно оценить все взаимодействия. Для дробных реплик это не так.

Если вы построили полуреплику 2^{4-1} с определяющим контрастом $1 = x_1x_2x_3x_4$, то раздельных оценок b_{12} и b_{34} получить нельзя, так как имеет место соотношение $x_1x_2 = x_3x_4$. В этом можно легко убедиться, если выписать столбцы интересующих нас эффектов. А это необходимо для вычисления коэффициентов. В нашем случае столбец x_1x_2 совпадает со столбцом x_3x_4 .

Формулу для вычислений коэффициентов можно записать так:

$$b_{uj} = \frac{\sum_{i=1}^N y_i x_{ui} x_{ji}}{N}.$$

Здесь $u, j = 1, 2, \dots, k$ — номера факторов, $u \neq j$.

Обратите внимание, что в силу ортогональности эффекты взаимодействия оцениваются независимо от линейных эффектов.

9.2. Регрессионный анализ

До сих пор мы пользовались МНК как *вычислительным приемом*. Нам нигде не приходилось вспоминать о статистике. Но, как только мы начинаем проверять какие-либо гипотезы о пригодности модели или о значимости коэффициентов, приходится вспоминать о статистике. И с этого момента МНК превращается в регрессионный анализ.

А регрессионный анализ, как всякий статистический метод, применим при определенных предположениях, постулатах.

Первый постулат. Параметр оптимизации y есть *случайная величина* с нормальным законом распределения. Дисперсия воспроизводимости, которую мы научились находить в седьмой главе, — одна из характеристик этого закона распределения.

В данном случае, как и по отношению к любым другим постулатам, нас интересуют два вопроса: как проверить его выполнимость и к чему приводят его нарушения?

При наличии большого экспериментального материала (десятки параллельных опытов) гипотезу о нормальном распределении можно проверить стандартными статистическими тестами

(например, χ^2 -критерием). К сожалению, экспериментатор редко располагает такими данными, поэтому приходится принимать этот постулат на веру. (Кроме тех случаев, когда заведомо известно, что это не так и требуется специальное рассмотрение. Мы не будем на них останавливаться.)

В том, что y — случайная величина, обычно сомневаться не приходится.

Какие последствия связаны, по вашему мнению, с нарушением первого постулата?

При нарушении нормальности мы лишаемся возможности установления вероятностей, с которыми справедливы те или иные высказывания. В этом таится большая опасность. Мы рискуем загноготизировать себя численными оценками и вероятностями, за которыми ничего не стоит. Это даже хуже волюнтаризма. Вот почему надо очень внимательно относиться к возможным нарушениям предпосылок.

Второй постулат. Дисперсия y не зависит от абсолютной величины y . С этим требованием мы уже встречались в восьмой главе.

Выполнимость этого постулата проверяется с помощью критерия однородности дисперсий в разных точках факторного пространства. Нарушение этого постулата недопустимо. Если однородность дисперсий все же отсутствует, то необходимо такое преобразование y , которое делает дисперсии однородными. Увы, его не всегда легко найти. Довольно часто помогает логарифмическое преобразование, с которого обычно начинают поиски.

Третий постулат. Значения факторов суть *неслучайные величины*. Это несколько неожиданное утверждение практически означает, что установление каждого фактора на заданный уровень и его поддержание существенно точнее, чем ошибка воспроизводимости.

Нарушение этого постулата приводит к трудностям при реализации матрицы планирования. Поэтому оно обычно легко обнаруживается экспериментатором.

Существует еще четвертый постулат, налагающий ограничения на взаимосвязь между значениями факторов. У нас он выполняется автоматически в силу ортогональности матрицы планирования.

Если с постулатами все в порядке, то можно проверять статистические гипотезы.

9.3. Проверка адекватности модели

Первый вопрос, который нас интересует после вычисления коэффициентов модели, это проверка ее пригодности. Мы будем называть такую проверку *проверкой адекватности модели*.

Ниже (рис. 27, а, б) приведены два рисунка с одинаковым расположением экспериментальных точек и, следовательно, одинаковым разбросом относительно линии регрессии, но с различ-

ным средним разбросом в точках (с различной дисперсией воспроизводимости). Разброс в точках показан, как это иногда делается, отрезками прямых, составляющих доверительный интервал, равный $\pm 2s_{\{y\}}$. Модель можно считать адекватной только в первом случае.

В данном случае разброс в точках такого же порядка, что и разброс относительно линии. Поэтому можно предполагать, что построенная модель пригодна. (Дальше мы выясним, как проверить это количественно.) Во втором случае опыты «слишком» точны. Требуется более сложная модель, чтобы точность ее предсказания была сравнима с точностью эксперимента.

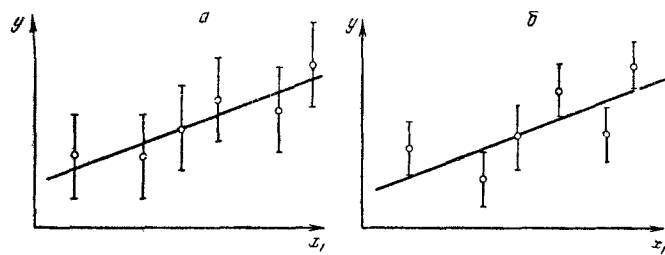


Рис. 27. Проверка адекватности

Это качественные соображения, а нам нужна количественная мера.

Для характеристики среднего разброса относительно линии регрессии вполне подходит остаточная сумма квадратов. Неудобство состоит в том, что она зависит от числа коэффициентов в уравнении: введите столько коэффициентов, сколько вы провели независимых опытов, и получите остаточную сумму, равную нулю. Поэтому предпочитают относить ее на один «свободный» опыт. Число таких опытов называется числом степеней свободы (f).

Числом степеней свободы в статистике называется разность между числом опытов и числом коэффициентов (констант), которые уже вычислены по результатам этих опытов независимо друг от друга.

Если, например, вы провели полный факторный эксперимент 2^3 и нашли линейное уравнение регрессии, то число степеней свободы

$$f = N - (k + 1) = 8 - (3 + 1) = 4.$$

Остаточная сумма квадратов, деленная на число степеней свободы, называется *остаточной дисперсией, или дисперсией адекватности* ($s_{\text{ак}}^2$).

$$s_{\text{ак}}^2 = \frac{\sum_{i=1}^N \Delta y_i^2}{f}.$$

Рассмотрим пример. Требуется найти число степеней свободы для $s_{\text{ак}}^2$ в следующем случае: план 2^{4-1} и четыре параллельных опыта в нулевой точке для вычисления ошибки опыта; модель линейная.

Один из возможных ответов — семь степеней свободы — основан на следующих предположениях. Вероятно, вы рассуждали так. Прделано 12 опытов: $2^{4-1} = 8$ плюс 4 нулевых. В уравнение входит 5 коэффициентов. Следовательно, $f = 12 - 5 = 7$. Здесь не учтено, что параллельные опыты нельзя считать самостоятельными, так как они дублируют друг друга. Поэтому они все дают одну степень свободы. Другой неправильный ответ — 4 степени свободы. Этот неправильный ответ получился, вероятно, из следующего рассуждения. Прделано 12 опытов: восемь по матрице планирования и четыре нулевых. Так как все нулевые опыты тождественны, то они дают одну степень свободы. Число коэффициентов в модели равно пяти. Следовательно, $f = 9 - 5 = 4$. Вы не обратили внимание на то, что опыты в нулевой точке не используются при вычислении коэффициентов и не могут поэтому войти в число степеней свободы.

Правильный ответ — три степени свободы. Действительно, мы провели 12 опытов, но четыре опыта в нулевой точке были проведены для других целей и в вычислении коэффициентов не участвовали, поэтому они не входят в число степеней свободы. (А если бы входили — такие случаи возможны, то давали бы не четыре, а только одну степень свободы.) Число коэффициентов модели — пять. Следовательно,

$$f = 8 - 5 = 3.$$

Запомните правило: в планировании эксперимента число степеней свободы для дисперсии адекватности равно числу различных опытов, результаты которых используются при подсчете коэффициентов регрессии, минус число определяемых коэффициентов. Вам еще представится случай поупражняться в определении числа степеней свободы для дисперсии адекватности.

Для проверки гипотезы об адекватности можно использовать F -критерий (этот критерий уже использовался для сравнения двух дисперсий)

$$F = \frac{s_{\text{ак}}^2}{s_{\{y\}}^2}.$$

Величину, стоящую в числителе этой формулы, мы только что научились считать, а знаменатель — старый знакомый (см. гл. 8) — это дисперсия воспроизводимости со своим числом степеней свободы.

Удобство использования критерия Фишера состоит в том, что проверку гипотезы можно свести к сравнению с табличным значением. Фрагмент соответствующей таблицы, который может удовлетворить ваши нужды не только в упражнениях этой книги, но и в большинстве случаев практики, приведен ниже (табл. 9.4).

Таблица 9.4

Значения F -критерия Фишера при 5%-ном уровне значимости [2]

f_1	$f_2 = 1$	2	3	4	5	6	12	24	∞
1	164,4	199,5	215,7	224,6	230,2	234,0	244,9	249,0	254,3
2	18,5	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,5	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,9
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,9
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,5	2,2	2,0	1,7
28	4,2	3,3	3,0	2,7	2,6	2,4	2,1	1,9	1,7
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

Таблица построена следующим образом. Столбцы связаны с определенным числом степеней свободы для числителя f_1 , строки — для знаменателя f_2 . На пересечении соответствующих строки и столбца стоят критические значения F -критерия. Как правило, в технических задачах используется уровень значимости 0,05.

Если рассчитанное значение F -критерия не превышает табличного, то с соответствующей доверительной вероятностью модель можно считать адекватной. При превышении табличного значения эту приятную гипотезу приходится отвергать.

Если модель адекватна, то мы можем перейти к крутому восхождению. Если нет — приходится преодолевать дополнительные трудности. Это мы обсудим ниже. Но во всех случаях интересно проверять еще значимость отдельных коэффициентов регрессии.

9.4. Проверка значимости коэффициентов

Проверка значимости каждого коэффициента проводится независимо.

Ее можно осуществлять двумя равноценными способами: проверкой по t -критерию Стьюдента или построением доверительного интервала. При использовании полного факторного эксперимента или регулярных дробных реплик доверительные интервалы для всех коэффициентов (в том числе и эффектов взаимодействия) равны друг другу.

Прежде всего надо, конечно, найти дисперсию коэффициента регрессии $s_{\{b_j\}}^2$. Она определяется в нашем случае по формуле

$$s_{\{b_j\}}^2 = \frac{s_{\{y\}}^2}{N}, \text{ если параллельные опыты отсутствуют.}$$

Из формулы видно, что дисперсии всех коэффициентов равны друг другу, так как они зависят только от ошибки опыта и числа опытов.

Теперь легко построить доверительный интервал (Δb_j)

$$\Delta b_j = \pm t s_{\{b_j\}}.$$

Здесь t — табличное значение критерия Стьюдента при числе степеней свободы, с которыми определялась $s_{\{y\}}^2$, и выбранном уровне значимости (обычно 0,05); $s_{\{b_j\}}$ — квадратичная ошибка коэффициента регрессии

$$s_{\{b_j\}} = \pm \sqrt{s_{\{b_j\}}^2}.$$

Формулу для доверительного интервала можно записать в следующей эквивалентной форме:

$$\Delta b_j = \pm \frac{ts_{\{y\}}}{\sqrt{N}}.$$

Коэффициент значим, если его абсолютная величина больше доверительного интервала. Доверительный интервал задается верхней и нижней границами $b_j + \Delta b_j$ и $b_j - \Delta b_j$.

Для отыскания значений t -критерия можно воспользоваться таблицей, фрагмент из которой приведен в табл. 9.5.

Таблица построена следующим образом. Столбцы соответствуют различным степеням свободы и значениям критерия.

Пусть в двух разных задачах случайно оказались два численно равных коэффициента регрессии. Доверительные интервалы для них оказались различными. Из них значим только второй

Задача	b_j	Δb_j
1	5,3	$\pm 5,5$
2	5,3	$\pm 2,6$

Таблица 9.5

Значения t -критерия Стьюдента при 50%-ном уровне значимости [2]

Число степеней свободы	Значения t -критерия	Число степеней свободы	Значения t -критерия	Число степеней свободы	Значения t -критерия
1	12,71	11	2,201	21	2,080
2	4,303	12	2,179	22	2,074
3	3,182	13	2,160	23	2,069
4	2,776	14	2,145	24	2,064
5	2,571	15	2,131	25	2,060
6	2,447	16	2,120	26	2,056
7	2,365	17	2,110	27	2,052
8	2,306	18	2,101	28	2,048
9	2,262	19	2,093	29	2,045
10	2,228	20	2,086	30	2,042
				∞	1,960

В действительности чем уже доверительный интервал (при заданном α), тем с большей уверенностью можно говорить о значимости коэффициента.

Помните рабочее правило: если абсолютная величина коэффициента больше, чем доверительный интервал, то коэффициент значим.

Если вам больше нравится проверять значимость коэффициентов по t -критерию, то воспользуйтесь формулой

$$t = \frac{|b_j|}{s(b_j)}.$$

Вычисленное значение t -критерия сравнивается с табличным при заданном α и соответствующем числе степеней свободы. Полученные выводы о значимости коэффициентов, конечно, должны совпадать с предыдущими.

Так производится проверка значимости коэффициентов.

9.5. Резюме

Итак, в этой главе вы освоили основные методы обработки экспериментальных данных, полученных при планировании эксперимента. Вы научились не только вычислять коэффициенты регрессии, но и проводить статистические оценки адекватности и значимости.

Мы подробно рассмотрели метод наименьших квадратов — эффективный и простой способ получения оценок коэффициентов регрессии. Эти оценки приводят к минимально возможной остаточной сумме квадратов и в этом смысле являются оптимальными.

Одновременно мы установили важное требование обязательной проверки правильности вычислений и научились выполнять это требование при применении метода наименьших квадратов.

Вы узнали, что МНК становится частью регрессионного анализа при проверке статистических гипотез. При этом должны выполняться следующие постулаты: 1) параметр оптимизации — случайная величина с нормальным законом распределения; 2) дисперсия параметра оптимизации не зависит от значений параметра оптимизации; 3) значения факторов — неслучайные величины; 4) факторы не коррелированы.

Мы выяснили, как можно проверить выполнимость этих постулатов и к чему приводит их нарушение.

Всякая модель ценна постольку, поскольку она верно отражает описываемое явление. Мы выбрали подходящий статистический метод проверки адекватности модели, основанный на критерии Фишера, и научились им пользоваться.

Кроме проверки адекватности следует проводить проверку значимости коэффициентов. Эта проверка осуществляется с помощью критерия Стьюдента. Мы рассмотрели два варианта такой проверки: с помощью построения доверительных интервалов и непосредственно сравнением с табличным значением критерия.

Литература

1. Е. С. Венцель. Теория вероятностей. М., «Наука», 1969.
2. К. Доерфель. Статистика в аналитической химии. М., «Мир», 1969.
3. В. П. Спиридонов, А. А. Лопаткин. Математическая обработка физико-химических данных. М., изд-во МГУ, 1970.
4. Д. Худсон. Статистика для физиков. Изд. 2-е. М., «Мир», 1970.
5. Н. Дрейпер, Г. Смит. Прикладной регрессионный анализ. М., «Статистика», 1973.
6. С. А. Айвазян. Статистическое исследование зависимостей. М., «Металлургия», 1966.
7. Ю. Нейман. Вводный курс теории вероятностей и математической статистики. М., «Наука», 1968.
8. Л. Яноши. Теория и практика обработки результатов измерений. М., «Мир», 1968.

МАТРИЧНЫЙ ПОДХОД К РЕГРЕССИОННОМУ АНАЛИЗУ

Что, в сущности, буква и цифра? Не глаза ли два,
Которым открыта вся суть естества?

Валюсар. Тирукурал

При изложении предыдущего материала мы не использовали матричную алгебру. Однако существует множество причин, по которым целесообразно ввести этот язык. Вот некоторые из них: многие сложные случаи, возникающие в планировании эксперимента, такие как планы с неравномерным дублированием опытов, планы с разным числом уровней и т. п., требуют матричной формы описания; рассмотрение некоторых критериев оптимальности планов естественно вести на матричном языке; многие работы, к которым вы захотите обратиться, прочтя эту книгу, используют матричные обозначения. Да и вообще алгебра матриц становится все более универсальным языком современной науки и техники. Данная глава не претендует на полное изложение матричного аппарата. Мы будем использовать только те определения и операции, которые необходимы для рассмотрения регрессионного анализа. Более полное изложение матричной алгебры можно найти в [1—4].

10.1. Метод наименьших квадратов для одного фактора

Будем действовать в такой последовательности. Сначала рассмотрим численный пример линейного уравнения для одного фактора, который в общем виде рассмотрен в предыдущей главе. Затем опишем его на матричном языке. Это позволит нам в дальнейшем сделать некоторые обобщения.

Таблица 10.1

Условия и результаты опытов

Номер опыта	x_1	y	Номер опыта	x_1	y
1	-2	0	4	+1	3
2	-1	1	5	+2	4
3	0	2			

Пусть известно, что y связан с x_1 линейным уравнением $y = b_0 + b_1 x_1$.

В результате проведенного эксперимента была получена следующая таблица, аналогичная табл. 9.1, в которой x_1 принимает пять кодированных значений (табл. 10.1). Давайте произведем вычисления в соответствии с предписанием табл. 9.2 (табл. 10.2).

Таблица 10.2

Расчетная таблица вычисления коэффициентов регрессии

Номер опыта	x_1	y	x_1^2	$y x_1$	y^2	$x_1 + y$	$(x_1 + y)^2$
1	-2	0	4	0	0	-2	4
2	-1	1	1	-1	1	0	0
3	0	2	0	0	4	2	4
4	+1	3	1	3	9	4	16
5	+2	4	4	8	16	6	36
Σ	0	10	10	10	30	—	60
Среднее значение	0	2					

Воспользовавшись полученными ранее формулами, вычислим коэффициенты:

$$b_0 = \frac{\sum_{i=1}^5 y_i \sum_{i=1}^5 x_{1i}^2 - \sum_{i=1}^5 y_i x_{1i} \sum_{i=1}^5 x_{1i}}{5 \sum_{i=1}^5 x_{1i}^2 - \left(\sum_{i=1}^5 x_{1i} \right)^2} = \frac{10 \cdot 10 - 10 \cdot 0}{5 \cdot 10 - 0} = 2;$$

$$b_1 = \frac{5 \sum_{i=1}^5 y_i x_{1i} - \sum_{i=1}^5 y_i \sum_{i=1}^5 x_{1i}}{5 \sum_{i=1}^5 x_{1i}^2 - \left(\sum_{i=1}^5 x_{1i} \right)^2} = \frac{5 \cdot 10 - 10 \cdot 0}{5 \cdot 10 - 0} = 1.$$

Теперь можно написать уравнение $y = 2 + 1 \cdot x_1$. С помощью этого примера мы воспроизвели процедуру МНК, изложенную в предыдущей главе. Давайте воспользуемся теперь другим языком.

В нашем примере участвуют три множества элементов: элементы, задающие условия проведения опытов, элементы, характеризующие их результаты, и неизвестные коэффициенты, которые нужно определить.

Так, элементы, характеризующие результаты опытов, можно представить в виде столбца

$$Y = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix},$$

неизвестным коэффициентам соответствует столбец

$$B = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix},$$

а элементы, задающие условия опытов, удобно представить в виде таблицы

$$X = \begin{array}{cc} x_0 & x_1 \\ \begin{bmatrix} +1 & -2 \\ +1 & -1 \\ +1 & 0 \\ +1 & +1 \\ +1 & +2 \end{bmatrix} \end{array}.$$

В этой таблице появился столбец x_0 , состоящий из $+1$. Он введен для удобства вычислений всех коэффициентов, включая свободный член b_0 . Фактически это значит, что мы переписали исходное уравнение так: $y = b_0 x_0 + b_1 x_1$. Об этом уже говорилось в шестой главе.

Таблицы, в которых собраны упорядоченные некоторым образом элементы, называются матрицами. Следовательно, в нашем случае Y , B и X являются матрицами. Элементы Y упорядочены по номерам опытов, элементы X , кроме того, — по номерам переменных (x_0 , x_1), а элементы B — по номерам коэффициентов, которые соответствуют номерам переменных.

Для того чтобы воспользоваться матрицами при описании регрессионного анализа, необходимо ввести некоторые операции над ними, что мы и сделаем в следующем параграфе.

10.2. Некоторые операции над матрицами

Матрицы, описанные в предыдущем параграфе, различаются по числу элементов, числу строк и числу столбцов. Так, в матрице Y один столбец, пять строк и пять элементов, а в матрице X тоже пять строк, но два столбца и десять элементов. Если число строк и число столбцов различны, то матрицы называются *прямоугольными*, а при равном числе строк и столбцов — *квадратными*. Все матрицы из этого примера —

прямоугольные. Если матрица имеет всего один столбец, то ее называют матрицей-столбцом или вектор-столбцом. Примерами служат матрицы Y и B . Аналогично можно определить и матрицы-строки (векторы-строки).

Вернемся к нашему примеру. На основании исходных данных можно записать систему из пяти уравнений по одному уравнению для каждого опыта: $y_i = b_0 x_{0i} + b_1 x_{1i}$ ($i=1, 2, \dots, 5$) или в развернутой форме:

$$0 = b_0 \cdot 1 + b_1 \cdot (-2);$$

$$1 = b_0 \cdot 1 + b_1 \cdot (-1);$$

$$2 = b_0 \cdot 1 + b_1 \cdot (0);$$

$$3 = b_0 \cdot 1 + b_1 \cdot (+1);$$

$$4 = b_0 \cdot 1 + b_1 \cdot (+2).$$

На матричном языке эта система уравнений выглядит следующим образом:

$$Y = XB, \text{ т. е. } \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} +1 & -2 \\ +1 & -1 \\ +1 & 0 \\ +1 & +1 \\ +1 & +2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.$$

Чтобы эти две записи стали эквивалентными, необходимо ввести определенные правила перемножения матриц. Будем в произведении различать матрицу, стоящую слева, и матрицу, стоящую справа. Перемножить две матрицы это значит получить матрицу произведений, элементы которой находятся по следующим правилам.

Элементы первой строки матрицы, стоящей слева, умножаются на соответствующие элементы матрицы, стоящей справа, и полученные произведения складываются. В нашем случае имеем: $(+1)b_0 + (-2)b_1$. Для получения элемента, стоящего на пересечении первого столбца и второй строки, аналогичная операция продлевается со второй строкой матрицы, стоящей слева, и тем же самым первым столбцом матрицы, стоящей справа, т. е. $(+1)b_0 + (-1)b_1$. Продолжая таким образом до последней строки матрицы, стоящей слева, получаем все элементы первого столбца матрицы произведений. Эта процедура повторяется столько раз, сколько вектор-столбцов содержит матрица, стоящая справа. В нашем случае эта матрица имеет только один столбец. Из определения видно, что матрица произведений имеет столько столбцов, сколько матрица, стоящая справа, и столько строк, сколько матрица, стоящая слева. В рассматриваемом примере матрица-

произведение имеет один столбец и пять строк, что соответствует размерности матрицы Y . И тогда матрица-произведение имеет вид

$$\begin{bmatrix} 1 & b_0 + b_1(-2) \\ 1 & b_0 + b_1(-1) \\ 1 & b_0 + b_1(0) \\ 1 & b_0 + b_1(+1) \\ 1 & b_0 + b_1(+2) \end{bmatrix}.$$

Сопоставление матрицы-произведения с системой уравнений убеждает нас в тождественности матричной и нематричной форм записей. Вектор Y , оказывается, и есть матрица произведений в данном случае. Элементы матрицы-произведения называются скалярными произведениями вектор-строки матрицы, стоящей слева, и соответствующего вектор-столбца матрицы, стоящей справа. В правилах перемножения матриц существуют особенности, не имеющие аналога в числах. Так, безразлично, в каком порядке записаны матрицы в произведении. Вы, паверное, заметили, что левая и правая матрицы неравноправны. Если вы захотите умножить матрицу B на матрицу X (BX), то убедитесь, что этого сделать невозможно, ибо длины векторов, входящих в скалярное произведение, должны быть согласованы.

Таким образом, для двух произвольных матриц произведение существует, если число столбцов матрицы, стоящей слева, равно числу строк матрицы, стоящей справа. Ясно, что для двух квадратных матриц одинакового размера существуют оба произведения (справа и слева), однако они могут быть различными. Матрицы, произведение которых не зависит от порядка сомножителей, называются *коммутирующими*. В общем же случае для произведения матриц коммутативный закон не выполняется.

Перейдем теперь к системе нормальных уравнений МНК, которая в нашем случае выглядит следующим образом (см. стр. 144):

$$5b_0 + 0b_1 = 10; \quad 0b_0 + 10b_1 = 10.$$

Можно показать, что в матричном виде она запишется следующим образом: $X^T X B = X^T Y$. Здесь X^T обозначает матрицу, *транспонированную* по отношению к матрице X . Протранспонировать матрицу — это значит столбцы исходной матрицы сделать строками транспонированной матрицы, сохранив их последовательность. Так, в нашем случае транспонированная матрица

$$X^T = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 \\ -2 & -1 & 0 & +1 & +2 \end{bmatrix}.$$

Для получения системы нормальных уравнений нам пришлось умножить обе части исходной системы уравнений слева на X^T . Давайте выполним эти операции

$$X^T Y = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 \\ -2 & -1 & 0 & +1 & +2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & +1 & +2 & +3 & +4 \\ 0 & +(-1) & +0 & +3 & +8 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \end{bmatrix};$$

$$X^T X = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 \\ -2 & -1 & 0 & +1 & +2 \end{bmatrix} \cdot \begin{bmatrix} +1 & -2 \\ +1 & -1 \\ +1 & 0 \\ +1 & +1 \\ +1 & +2 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & +1 & +1 & +1 & +1 & -2 & -1 & 0 & +1 & +2 \\ -2 & -1 & 0 & +1 & +2 & +4 & +1 & 0 & +1 & +4 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}.$$

Теперь можно записать систему уравнений:

$$\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}.$$

Читателю предлагается возможность убедиться в том, что полученная матричная запись в точности соответствует исходной системе нормальных уравнений.

Матрица $X^T X$ называется *матрицей системы нормальных уравнений*. Она обладает рядом важных для нас свойств. Прежде всего заметим, что в этой матрице два элемента, расположенных симметрично относительно диагонали, идущей с левого верхнего угла в правый нижний (так называемой *главной диагонали*), равны между собой. В нашем случае это нули. Такое свойство характерно для матриц систем нормальных уравнений МНК, так как векторы, входящие в скалярные произведения, коммутативны.

Матрица, элементы которой симметричны относительно главной диагонали, называется *симметричной*. Если все элементы вне главной диагонали равны нулю, то такая матрица называется *диагональной*. В дальнейшем нам понадобится еще одна разновидность диагональных матриц. Диагональная матрица, все элементы главной диагонали которой являются единицами, называется *единичной* матрицей. Единичная матрица играет в алгебре матриц такую же роль, какую единица — в алгебре чисел.

Решить систему нормальных уравнений это значит записать в явном виде элементы вектора B (b_0 и b_1). Если бы мы имели дело с числами, то для этого нужно было бы поделить обе части на коэффициент при неизвестном и получить ответ. Но для матриц вместо деления (которое не определено) используется специальная опе-

рация умножения на *обратную* матрицу. Задача состоит в том, чтобы превратить матрицу, стоящую перед матрицей неизвестных коэффициентов, в единичную. Тогда умножение вектора \mathbf{B} на единичную матрицу его не изменит, а чтобы равенство не нарушилось, и правую часть придется домножить на соответствующую матрицу. Если условиться обозначать обратную матрицу степенью -1 , то предыдущие рассуждения приведут к следующей записи: $(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) \mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. Здесь система нормальных уравнений МНК умножена слева на матрицу, обратную к матрице системы нормальных уравнений.

Произведение обратной матрицы на прямую справа равно единичной матрице, которую условимся обозначать \mathbf{E} : $\mathbf{E} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X})$. В этом равенстве участвуют три матрицы. Матрица системы нормальных уравнений $\mathbf{X}^T \mathbf{X}$, которую называют *прямой* матрицей, $(\mathbf{X}^T \mathbf{X})^{-1}$ — обратная матрица. Перепишем это равенство для нашей задачи

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}.$$

Неизвестные элементы обратной матрицы обозначены a_{ij} , где $i=1, 2$ соответствует строке, а $j=1, 2$ — столбцу. Найдем эти элементы

$$\begin{aligned} 1 &= a_{11} \cdot 5 + a_{12} \cdot 0; & 0 &= a_{21} \cdot 5 + a_{22} \cdot 0; \\ 0 &= a_{11} \cdot 0 + a_{12} \cdot 10; & 1 &= a_{21} \cdot 0 + a_{22} \cdot 10. \end{aligned}$$

Отсюда следует, что $a_{11}=1/5$, $a_{12}=0$, $a_{21}=0$, $a_{22}=1/10$.

Запишем обратную матрицу

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}.$$

Лишь благодаря простоте примера можно было воспользоваться столь элементарной процедурой. В общем случае приходится прибегать к более сложным алгоритмам и вычислительной технике [5, 6].

Отметим некоторые существенные свойства обратной матрицы. Произведение прямой и обратной матрицы коммутативно. Если прямую матрицу обозначить \mathbf{A} , то $\mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{E}$.

Матрица, обратная к симметричной, тоже будет симметрична. На главной диагонали матрицы, обратной к диагональной, будут стоять числа, обратные соответствующим числам, стоящим на диагонали прямой матрицы. Зная это свойство, мы могли не продвигать предыдущие вычисления, а сразу записать обратную матрицу для нашего примера.

Продолжим вычисление для примера. Подставим известные матрицы в уравнение для вектора коэффициентов $\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.

Имеем

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix} \cdot \begin{bmatrix} 10 \\ 10 \end{bmatrix}.$$

Осталось перемножить матрицы и получить

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1/5 \cdot 10 & + 0 \cdot 10 \\ 0 \cdot 10 & + 1/10 \cdot 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Две матрицы равны, когда равны их соответствующие элементы, поэтому $b_0=2$, $b_1=1$. Таким образом, мы получили результат, совпадающий с полученным ранее без использования матриц.

Введем еще одно важное понятие: каждой квадратной матрице можно поставить в соответствие некоторое число, называемое *определителем* этой матрицы. Определитель представляет собой алгебраическую сумму всех возможных произведений, в каждое из которых входит по одному элементу от каждой строки и от каждого столбца. Причем знак произведения (+ или —) зависит от положения элементов данного произведения матрицы.

Посчитаем определитель для матрицы системы нормальных уравнений. Здесь возможны два произведения, каждое из которых содержит два сомножителя $5 \cdot 10 = 50$ и $0 \cdot 0 = 0$.

Определитель принято обозначать

$$\begin{vmatrix} 5 & 0 \\ 0 & 10 \end{vmatrix} = 50 - 0 = 50$$

или

$$\det \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} = 50$$

(\det — определитель от латинского determinant).

Вычислим определитель обратной матрицы

$$\begin{vmatrix} 1/5 & 0 \\ 0 & 1/10 \end{vmatrix} = \frac{1}{50} - 0 = \frac{1}{50}.$$

Можно заметить, что между определителями прямой и обратной матрицы существует в данном случае простое соотношение: они являются взаимнообратными числами. Такое соотношение выполняется и в общем случае.

Определитель может быть любым действительным числом, как положительным, так и отрицательным. Он может оказаться и равным нулю. С последним случаем связаны некоторые особенности свойств матрицы. Матрица, определитель которой равен нулю, не имеет обратной. Такую матрицу называют *особенной*, *вырожденной* или *сингулярной*. Если же определитель матрицы не равен нулю, то матрица называется *неособенной*, *невырожденной* или *несингулярной*.

Например, у следующей матрицы определитель равен нулю и она является вырожденной

$$\det \begin{bmatrix} -1 & +2 \\ +1 & -2 \end{bmatrix} = (-1) \cdot (-2) - (+1) \cdot (+2) = 0.$$

Значит, решение системы нормальных уравнений возможно только тогда, когда матрица невырождена, т. е. $\det (X^T X) \neq 0$. Это предполагалось и имело место в нашем примере.

10.3. Обобщение метода наименьших квадратов на многофакторный линейный случай

Пусть имеется k факторов и известно, что отклик и факторы связаны линейно: $y = b_0 x_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$. Выпишем для этого случая матрицы X , Y и B

$$X = \begin{bmatrix} x_{01} & x_{11} & \dots & x_{k1} \\ x_{02} & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{0N} & x_{1N} & \dots & x_{kN} \end{bmatrix}_{(N \times k)}; \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{(N \times 1)}; \quad B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}_{[(k+1) \times 1]}.$$

Запишем исходную систему линейных уравнений:

$$Y = XB;$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{01} & x_{11} & \dots & x_{k1} \\ x_{02} & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{0N} & x_{1N} & \dots & x_{kN} \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}.$$

После преобразований, аналогичных рассмотренным в предыдущем параграфе, придем к следующей формуле:

$$B = (X^T X)^{-1} X^T Y;$$

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \left(\begin{bmatrix} x_{01} & x_{02} & \dots & x_{0N} \\ x_{11} & x_{12} & \dots & x_{1N} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \dots & x_{kN} \end{bmatrix} \cdot \begin{bmatrix} x_{01} & x_{11} & \dots & x_{k1} \\ x_{02} & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{0N} & x_{1N} & \dots & x_{kN} \end{bmatrix} \right)^{-1} \times \\ \times \begin{bmatrix} x_{01} & x_{02} & \dots & x_{0N} \\ x_{11} & x_{12} & \dots & x_{1N} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \dots & x_{kN} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Скалярные произведения удобно представлять в виде сумм, т. е. матрицу системы нормальных уравнений можно записать в следующем виде:

$$X^T X = \begin{bmatrix} \sum x_0^2 & \sum x_0 x_1 & \sum x_0 x_2 & \dots & \sum x_0 x_k \\ \sum x_0 x_1 & \sum x_1^2 & \sum x_1 x_2 & \dots & \sum x_1 x_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_0 x_k & \sum x_1 x_k & \sum x_2 x_k & \dots & \sum x_k^2 \end{bmatrix}.$$

Так как суммирование ведется от 1 до N по всему множеству опытов, индекс суммирования мы опустили.

Аналогично $X^T Y$ есть вектор сумм произведений:

$$X^T Y = \begin{bmatrix} \sum y x_0 \\ \sum y x_1 \\ \sum y x_2 \\ \vdots \\ \sum y x_k \end{bmatrix}.$$

Чтобы получить ответ, т. е. вектор B , остается обратить матрицу $X^T X$ и умножить обратную матрицу на $X^T Y$.

Проведем обработку данных табл. 7.9 матричным способом

$$X = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \end{bmatrix};$$

$$X^T = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \end{bmatrix};$$

$$X^T X = \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix};$$

10.4. Статистический анализ

В предыдущей главе вначале была продемонстрирована процедура МНК, а затем рассмотрены статистические аспекты. Так же построена и эта глава

Перейдем к статистическому анализу в матричной форме. Будем предполагать, что постулаты регрессионного анализа выполняются. Что значит провести статистический анализ? Это значит проверить ряд статистических гипотез гипотезу об адекватности заданной модели, гипотезу о значимости отдельных коэффициентов регрессии и др. Дальнейшее изложение направлено на прояснение некоторых обстоятельств, определивших вид формул статистического анализа регрессионной модели. Знакомство с элементами алгебры матриц поможет сделать это.

Фундаментальную роль в анализе уравнения регрессии играет матрица

$$M^{-1} = (X^T X s_{\{y\}}^2)^{-1},$$

которая называется матрицей дисперсий ковариаций. Прячая матрица M называется информационной матрицей Фишера.

В структуре матрицы дисперсий-ковариаций содержится вся информация о статистических свойствах модели. Провести статистический анализ значит извлечь эту информацию. Для этого прежде всего перейдем от матрицы, обратной к матрице системы нормальных уравнений, к матрице M^{-1} . Оценка дисперсии воспроизводимости $s_{\{y\}}^2$ — скаляр; $X^T X$ — квадратная матрица. Умножить матрицу на скаляр слева или справа — значит умножить на этот скаляр каждый элемент матрицы.

Полученные таким образом произведения имеют определенный статистический смысл. Так, на главной диагонали матрицы-произведения стоят оценки дисперсий коэффициентов регрессии, вне главной диагонали расположены оценки ковариаций.

Чтобы познакомиться с понятием *ковариация*, рассмотрим два произвольных вектор-столбца матрицы X . Во многих случаях важно знать, сколь сильна линейная связь между этими векторами. Ковариация является одной из мер такой связи. Чтобы найти ковариацию, сначала центрируют оба вектора, а затем вычисляют их скалярное произведение. Центрирование используется для устранения неопределенности, связанной с выбором начала координат. Пусть, например, изучается ковариация между температурой и каким-нибудь другим фактором. Если значения температуры записываются в шкале Цельсия, то без центрирования значение ковариации получится иное, чем для шкалы Кельвина. При центрировании же это не произойдет.

Ковариация определяется по формуле

$$\text{cov} \{x_1 x_2\} = \sum_{i=1}^N (x_{1i} - \bar{x}_1) (x_{2i} - \bar{x}_2).$$

$$(X^T X)^{-1} = \begin{bmatrix} 1/8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/8 \end{bmatrix};$$

$$b_0 = \frac{\sum y_i x_{0i}}{8} = \frac{188,7}{8} = 23,5875;$$

$$b_1 = \frac{\sum y_i x_{1i}}{8} = \frac{8,5}{8} = 1,0625,$$

$$b_2 = \frac{\sum y_i x_{2i}}{8} = \frac{-46,5}{8} = -5,8125;$$

$$b_3 = \frac{\sum y_i x_{3i}}{8} = \frac{1,5}{8} = 0,1875;$$

$$b_4 = \frac{\sum y_i x_{4i}}{8} = \frac{26,5}{8} = 3,3125;$$

$$b_5 = \frac{\sum y_i x_{5i}}{8} = \frac{0,5}{8} = 0,0625,$$

$$b_6 = \frac{\sum y_i x_{6i}}{8} = \frac{-10,5}{8} = -1,3125;$$

$$b_7 = \frac{\sum y_i x_{7i}}{8} = \frac{-14,3}{8} = -1,7875.$$

$$X^T Y = \begin{bmatrix} \sum y_i x_{0i} \\ \sum y_i x_{1i} \\ \sum y_i x_{2i} \\ \sum y_i x_{3i} \\ \sum y_i x_{4i} \\ \sum y_i x_{5i} \\ \sum y_i x_{6i} \\ \sum y_i x_{7i} \end{bmatrix};$$

Мы получили те же результаты, что и в седьмой главе. Детально рассмотренная процедура МНК в матричной форме показывает, как получают формулы для коэффициентов регрессии, которые использовались в предыдущих главах.

Аналогичным путем можно оценить эффекты взаимодействия, входящие в модель. Для этого надо расширить матрицу X , включив в нее столбцы взаимодействий. Все остальные операции производятся совершенно аналогично. В векторе Y появляются при этом элементы, соответствующие эффектам взаимодействий. Расширение матрицы X подобным образом называют *линеаризацией*. Это эквивалентно замене эффектов взаимодействия новыми линейными членами. Подобная процедура возможна только тогда, когда все коэффициенты входят в уравнение линейно. Такие уравнения называются *линейными по параметрам* и только они рассматриваются в нашей книге. В некоторых случаях приходится использовать уравнения, *нелинейные по параметрам*. Примером может служить уравнение Аррениуса $y = b_0 e^{b_1/x_1}$, которое используется в химической кинетике [7, 8].

Заметим, что это выражение совпадает с числителем коэффициента парной корреляции, с которым мы уже сталкивались (§ 2.3).

Давайте построим матрицу M^{-1} для однофакторной линейной модели. Информационная матрица M равна:

$$M = (X^T X \cdot s_{\{y\}}^2) = \begin{bmatrix} N & \sum x_{1i} \\ \sum x_{1i} & \sum x_{1i}^2 \end{bmatrix} \cdot s_{\{y\}}^2.$$

Матрица 'дисперсий-ковариаций M^{-1} равна:

$$M^{-1} = \begin{bmatrix} \frac{\sum x_{1i}^2}{N \sum x_{1i}^2 - (\sum x_{1i})^2} s_{\{y\}}^2 & -\frac{\sum x_{1i}}{N \sum x_{1i}^2 - (\sum x_{1i})^2} s_{\{y\}}^2 \\ -\frac{\sum x_{1i}}{N \sum x_{1i}^2 - (\sum x_{1i})^2} s_{\{y\}}^2 & \frac{N}{N \sum x_{1i}^2 - (\sum x_{1i})^2} s_{\{y\}}^2 \end{bmatrix} =$$

$$= \begin{bmatrix} s_{\{b_0\}}^2 & \text{cov}\{b_0, b_1\} \\ \text{cov}\{b_0, b_1\} & s_{\{b_1\}}^2 \end{bmatrix}.$$

Ортогональные планы, рассматриваемые в этой книге, обладают тем свойством, что ковариации между всеми парами коэффициентов регрессии равны нулю. Это можно проиллюстрировать примером из предыдущего параграфа:

$$M^{-1} = \begin{bmatrix} s_{\{y\}}^2/8 & & & & & \\ & s_{\{y\}}^2/8 & & & & \\ & & s_{\{y\}}^2/8 & & & 0 \\ & & & s_{\{y\}}^2/8 & & \\ & & & & s_{\{y\}}^2/8 & \\ 0 & & & & & s_{\{y\}}^2/8 \\ & & & & & & s_{\{y\}}^2/8 \\ & & & & & & & s_{\{y\}}^2/8 \end{bmatrix}$$

(так принято сокращенно записывать квадратные матрицы, когда все внедиагональные элементы равны нулю). В этом примере $s_{\{y\}}^2 = 1$ и поэтому дисперсии b -коэффициентов $s_{\{b_j\}}^2 = 0,125$. Таким образом, мы пришли к формуле

$$s_{\{b_j\}}^2 = s_{\{y\}}^2 / N,$$

которая уже фигурировала раньше. Она справедлива для ортогональных планов.

Рассмотрим теперь проверку адекватности линейного уравнения регрессии. Дисперсия адекватности равна

$$s_{\text{а.к}}^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - (k + 1)}.$$

Числитель этого выражения — остаточная сумма квадратов — в матричной форме имеет вид

$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 = (Y - \hat{Y})^T (Y - \hat{Y}) = Y^T Y - B^T X^T Y.$$

Приведем конкретный пример (§ 6.4, пример № 4) расчета остаточной суммы квадратов разными способами (см. табл. 10.3).

Таблица 10.3

Расчет остаточной суммы квадратов

Номер опыта	x_0	x_1	x_2	y	\hat{y}	$y - \hat{y}$	$(y - \hat{y})^2$
1	+	—	—	95,0	94,5	+0,5	0,25
2	+	+	—	90,0	90,5	—0,5	0,25
3	+	—	+	85,0	85,5	—0,5	0,25
4	+	+	+	82,0	81,5	+0,5	0,25

Линейное уравнение регрессии имеет вид $\hat{y} = 88,0 - 2,0x_1 - 4,5x_2$. Обычный способ вычисления остаточной суммы дает

$$\sum_{i=1}^4 (y_i - \hat{y}_i)^2 = (0,25 + 0,25 + 0,25 + 0,25) = 1,00.$$

Воспользуемся теперь матричной записью. Для этого придется ввести еще одну матричную операцию — операцию *вычитания*. Разностью двух матриц одинакового размера называется матрица, элементы которой являются разностями соответствующих элементов матриц уменьшаемого и вычитаемого, взятых в том же порядке:

$$Y - \hat{Y} = \begin{bmatrix} 95,0 \\ 90,0 \\ 85,0 \\ 82,0 \end{bmatrix} - \begin{bmatrix} 94,5 \\ 90,5 \\ 85,5 \\ 81,5 \end{bmatrix} = \begin{bmatrix} +0,5 \\ -0,5 \\ -0,5 \\ +0,5 \end{bmatrix}.$$

Транспонирование и перемножение дают

$$(Y - \hat{Y})^T (Y - \hat{Y}) = [+0,5 \ -0,5 \ -0,5 \ +0,5] \begin{bmatrix} +0,5 \\ -0,5 \\ -0,5 \\ +0,5 \end{bmatrix} =$$

$$= 0,25 + 0,25 + 0,25 + 0,25 = 1,00.$$

Если воспользоваться другим выражением для остаточной суммы квадратов, то

$$\mathbf{Y}^T \mathbf{Y} - \mathbf{B}^T \mathbf{X}^T \mathbf{Y} = [95,0 \ 90,0 \ 85,0 \ 82,0] \cdot \begin{bmatrix} 95,0 \\ 90,0 \\ 85,0 \\ 82,0 \end{bmatrix} -$$

$$- [88,0 \ -2,0 \ -4,5] \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 95 \\ 90 \\ 85 \\ 82 \end{bmatrix} = 1,00.$$

Здесь мы встретились с необходимостью перемножить три матрицы. Эту операцию можно выполнять либо справа налево, либо слева направо. Сначала перемножается первая пара матриц. А затем матрица-произведение рассматривается как новая матрица и перемножается с третьей. Конечно, размеры всех матриц должны удовлетворять условиям существования произведения, что и выполняется в нашем случае.

Кроме статистических оценок, уже знакомых по предыдущей главе, введем еще одну оценку — оценку *дисперсии предсказанного значения отклика*. Если имеется адекватное уравнение регрессии, то его можно использовать для предсказания результата какого-нибудь нового опыта в некоторой точке факторного пространства. Для этого достаточно подставить в уравнение координаты этой точки и произвести алгебраические операции. Очевидно точность такого предсказания будет неодинакова в разных точках факторного пространства. Чтобы учесть это различие и вводится дисперсия предсказанного значения отклика.

Пусть известное уравнение имеет вид $y = b_0 + b_1 x_1$. Координаты предсказываемой точки задаются вектором $\mathbf{X}_i^T = [1 \ x_i]$. Отсюда следует, что

$$\hat{y}_i = [1 \ x_i] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}.$$

В рамках предпосылок регрессионного анализа x — случайная величина, а b_0 и b_1 — случайные величины, так как они являются функциями результатов эксперимента. Следовательно, \hat{y} — тоже случайная величина, связанная с некоторой суммой двух величин b_0 и b_1 . Дисперсии и ковариация b_0 и b_1 уже известны. Они являются элементами матрицы дисперсий-ковариаций. Для определения дисперсии предсказанного значения отклика $s_{\{y\}}^2$ можно воспользоваться законом сложения ошибок [9]. Если $y = f(z_1, z_2, \dots, z_k)$, где z_j — независимые случайные величины, входящие в известное уравнение, то

$$s_{\{y\}}^2 \approx \sum_{j=1}^k \left(\frac{\partial y}{\partial z_j} \right)_i^2 s_{\{z_j\}}^2 + \sum_{j=1}^k \sum_{l=1, l \neq j}^k \left(\frac{\partial^2 y}{\partial z_j \partial z_l} \right)_i \text{cov} \{z_j, z_l\}.$$

В нашем случае

$$s_{\{y\}}^2 \approx \left(\frac{\partial y}{\partial b_0} \right)_i^2 s_{\{b_0\}}^2 + \left(\frac{\partial y}{\partial b_1} \right)_i^2 s_{\{b_1\}}^2 + \left(\frac{\partial^2 y}{\partial b_0 \partial b_1} \right)_i \text{cov} \{b_0, b_1\}$$

или

$$s_{\{y\}}^2 \approx s_{\{b_0\}}^2 + 2x_i \text{cov} \{b_0, b_1\} + x_i^2 s_{\{b_1\}}^2.$$

Последнее соотношение можно записать также в матричной форме:

$$s_{\{y\}}^2 = \mathbf{X}_i^T (\mathbf{X}^T \mathbf{X})^{-1} s_{\{y\}}^2 \mathbf{X}_i = [1 \ x_i] \begin{bmatrix} s_{\{b_0\}}^2 & \text{cov} \{b_0, b_1\} \\ \text{cov} \{b_0, b_1\} & s_{\{b_1\}}^2 \end{bmatrix} \begin{bmatrix} 1 \\ x_i \end{bmatrix}.$$

Давайте произведем вычисления для нескольких опытов из примера предыдущего параграфа. Возьмем в качестве первой точки начало координат — нулевой опыт. В этом случае вектор \mathbf{X}_i имеет вид $\mathbf{X}_0^T = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$. Матрица

$$\mathbf{M}^{-1} = (\mathbf{X}^T \mathbf{X} s_{\{y\}}^2)^{-1} = \begin{bmatrix} 1/8 & & & & & & & \\ & 1/8 & & & & & & \\ & & 1/8 & & & & & \\ & & & 0 & & & & \\ & & & & 1/8 & & & \\ & & & & & 1/8 & & \\ & & & 0 & & & 1/8 & \\ & & & & & & & 1/8 \end{bmatrix},$$

так как $s_{\{y\}}^2 = 1$.

Отсюда получим дисперсию предсказанного значения y , который в этой точке равен 23,5875

$$s_{\{y_0\}}^2 = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \begin{bmatrix} 1/8 & & & & & & & \\ & 1/8 & & & & & & \\ & & 1/8 & & & & & \\ & & & 0 & & & & \\ & & & & 1/8 & & & \\ & & & & & 1/8 & & \\ & & & 0 & & & 1/8 & \\ & & & & & & & 1/8 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} =$$

$$= [1/8 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 1/8 = 0,125.$$

Таким образом, точность предсказания отклика в центре эксперимента оказалась равной дисперсии коэффициентов. Посмотрим, какое значение примет $s^2_{\{g\}}$ в точке с координатами, равными 2: $X_g^T = [1 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2]$. Тогда

$$s^2_{\{g\}} = [1 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2] \begin{bmatrix} 1/8 & & & & & & \\ & 1/8 & & & & & \\ & & 1/8 & & & & \\ & & & 1/8 & & & \\ & & & & 1/8 & & \\ & & & & & 1/8 & \\ & & & & & & 1/8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} =$$

$$= 3,625.$$

В этой точке $\hat{y}_g = 20,0125$, а дисперсия $s^2_{\{g\}} = 3,625$. Обратите внимание, как существенно увеличилась дисперсия предсказания при удалении от центра.

Описывая статистический анализ, мы до сих пор не принимали во внимание повторных наблюдений. Перейдем теперь к рассмотрению этого вопроса.

10.5. Взвешенный метод наименьших квадратов и статистический анализ

Каждый опыт несет некоторую информацию об объекте. Опыты, различающиеся условиями проведения, несут информацию об эффектах факторов, а параллельные опыты позволяют оценить дисперсию воспроизводимости. С ростом числа параллельных опытов растет точность эксперимента и оцениваемые эффекты можно определить с большей надежностью. На практике встречаются различные случаи дублирования опытов. Может оказаться, что к моменту начала эксперимента воспроизводимость опытов известна по предыдущим исследованиям. Так бывает иногда в задачах анализа вещества, когда используется методика с заранее известной ошибкой воспроизводимости. Если предполагать, что в намечаемой серии опытов ошибка не изменится и нет опасности появления грубых наблюдений, то параллельные опыты можно не ставить. Если же мы не располагаем такой информацией по предыдущим исследованиям или считаем наше предположение слишком жестким, тогда приходится дублировать опыты. Сделать это можно по-разному: в одной точке, в нескольких точках и во всех. В качестве одной точки выбирается центр плана или некоторая строка матрицы. В других случаях бывает равное число параллельных (равномерное дублирование) или различное (неравномерное дублирование). Последнее часто имеет место потому,

что часть опытов может оказаться потерянной: не удался анализ, сломалась установка и т. д. Различные варианты дублирования опытов приводят к различным вариантам обработки данных.

Начнем с наиболее распространенного случая — *равномерное дублирование*. Если при записи матрицы X не делать различия между параллельными и различными опытами, то число строк в матрице будет равно Nn , где N — число различных опытов; n — число параллельных опытов. Это приведет к некоторым изменениям в системе нормальных уравнений.

Пример 1. Пусть реализован эксперимент 2^2 с двумя параллельными опытами в каждой точке, который дал следующие результаты:

$$X = \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix}; \quad Y = \begin{bmatrix} 0,8 \\ 1,3 \\ 1,7 \\ 2,6 \\ 0,6 \\ 1,5 \\ 1,7 \\ 2,7 \end{bmatrix}; \quad B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}.$$

Выпишем матрицу системы нормальных уравнений и найдем оценки коэффициентов

$$X^T X = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{bmatrix};$$

$$X^T Y = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 \end{bmatrix} \begin{bmatrix} 0,8 \\ 1,3 \\ 1,7 \\ 2,6 \\ 0,6 \\ 1,5 \\ 1,7 \\ 2,7 \end{bmatrix} = \begin{bmatrix} 12,9 \\ 4,5 \\ 3,3 \end{bmatrix}.$$

Вычислим значения b -коэффициентов

$$B = (X^T X)^{-1} X^T Y = \begin{bmatrix} 1/8 & 0 & 0 \\ 0 & 1/8 & 0 \\ 0 & 0 & 1/8 \end{bmatrix} \begin{bmatrix} 12,9 \\ 4,5 \\ 3,3 \end{bmatrix} = \begin{bmatrix} 1,6125 \\ 0,5625 \\ 0,4125 \end{bmatrix}.$$

Таким образом, $\hat{y} = 1,6125 + 0,5625x_1 + 0,4125x_2$.

В этом варианте расчета различные и параллельные опыты не дифференцировались. Можно поступить иначе, рассматривая матрицу X как матрицу различных опытов. Тогда для учета информации о параллельных опытах будем использовать так называемую матрицу весов. Она представляет собой квадратную диагональную матрицу P размера $N \times N$. Элементы главной диагонали равны числу повторных опытов соответствующих строк матрицы X . Нумерация строк матрицы X должна совпадать с нумерацией строк матрицы P .

В нашем примере реализовано четыре различных опыта с двумя параллельными. Поэтому матрица X и матрица P имеют вид

$$X = \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix}; \quad P = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Тогда система нормальных уравнений МНК имеет вид $(X^T P X)B = X^T P \bar{Y}$, где \bar{Y} — вектор-столбец средних значений по соответствующему числу параллельных опытов. Это усреднение необходимо, чтобы привести в соответствие размеры матриц, входящих в систему нормальных уравнений. Коэффициенты регрессии определяются по формуле

$$\begin{aligned} B &= (X^T P X)^{-1} X^T P \bar{Y} = \\ &= \left(\begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix} \right)^{-1} \times \\ &\times \begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0,70 \\ 1,40 \\ 1,70 \\ 2,65 \end{bmatrix} = \\ &= \left(\begin{bmatrix} 2 & 2 & 2 & 2 \\ -2 & -2 & 2 & 2 \\ -2 & 2 & -2 & 2 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 2 & 2 & 2 & 2 \\ -2 & -2 & 2 & 2 \\ -2 & 2 & -2 & 2 \end{bmatrix} \begin{bmatrix} 0,70 \\ 1,40 \\ 1,70 \\ 2,65 \end{bmatrix} = \\ &= \begin{bmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{bmatrix}^{-1} \begin{bmatrix} 0,70 \cdot 2 & +1,40 \cdot 2 & +1,70 \cdot 2 & +2,65 \cdot 2 \\ -0,70 \cdot 2 & -1,40 \cdot 2 & +1,70 \cdot 2 & +2,65 \cdot 2 \\ -0,70 \cdot 2 & +1,40 \cdot 2 & -1,70 \cdot 2 & +2,65 \cdot 2 \end{bmatrix} = \end{aligned}$$

$$= \begin{bmatrix} \frac{(0,70 + 1,40 + 1,70 + 2,65) \cdot 2}{8} \\ \frac{(-0,70 - 1,40 + 1,70 + 2,65) \cdot 2}{8} \\ \frac{(-0,70 + 1,40 - 1,70 + 2,65) \cdot 2}{8} \end{bmatrix} = \begin{bmatrix} 1,6125 \\ 0,5625 \\ 0,4125 \end{bmatrix}.$$

Как и следовало ожидать, результаты совпадают.

Хотелось бы обратить ваше внимание на то, что при равномерном дублировании сохраняется ортогональность плана, и матрица нормальных уравнений остается диагональной. При отсутствии параллельных опытов матрица весов становится единичной.

А как теперь будет выглядеть статистический анализ результатов такого эксперимента? Рассмотрим проверку адекватности модели. При наличии числа повторных опытов n , равного для всех строк плана, дисперсия адекватности равна

$$s_{ад}^2 = \frac{n \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2}{N - (k + 1)}.$$

Числитель этого выражения в матричной форме имеет вид

$$n \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2 = \bar{Y}^T P \bar{Y} - B^T X^T P \bar{Y}.$$

Повторные опыты накладывают более жесткие условия на проверку адекватности, так как рассчитанный F -критерий увеличивается в n раз и для принятия гипотезы адекватности требуется большее соответствие экспериментальных и расчетных точек. В рассматриваемом примере $s_{ад}^2 = 0,0312$. Дисперсия воспроизводимости для одинакового числа повторных опытов подсчитывалась, как уже говорилось в гл. 8, по формуле

$$s_{воспр}^2 = \frac{\sum_{i=1}^N \sum_{q=1}^n (y_{iq} - \bar{y}_i)^2}{N(n-1)}.$$

Эта формула справедлива для однородных дисперсий.

Составим таблицу для расчета дисперсии воспроизводимости (табл. 10.4). Проверка показывает, что выборочные дисперсии

Таблица 10.4

Расчет дисперсии воспроизводимости

Номер опыта	Матрица планирования	y'	y''	y	Δy	Δy^2	s_i^2
1	(1)	0,8	0,6	0,70	0,10	0,0100	0,020
2	b	1,3	1,5	1,40	0,10	0,0100	0,020
3	a	1,7	1,7	1,70	0,00	0	0
4	ab	2,6	2,7	2,65	0,05	0,0025	0,005

однородны $G_{\text{эсп}}=0,44$, $G_{\text{табл}}=0,91$, $\alpha=0,05$, $s_{\text{воспр}}^2=0,0112$. Для проверки адекватности линейной модели найдем F -критерий: $F=s_{\text{ад}}^2/s_{\text{воспр}}^2=0,0312/0,0112 \approx 2,8$. Табличное значение критерия Фишера для числа степеней свободы 1; 4 и 5% уровня значимости (табл. 9.4) равно 7,7. Гипотеза адекватности линейной модели может быть принята.

Осталось проверить значимость b -коэффициентов. Дисперсия оценки b коэффициентов равна $s_{\{b_j\}}^2 = s_{\text{воспр}}^2/Nn = 0,0014$. Дисперсия воспроизводимости $s_{\text{воспр}}^2$, деленная на число параллельных опытов n , называется *дисперсией среднего* и обозначается $s_{\{g\}}^2$.

Отсюда имеем $s_{\{b_j\}}^2 = s_{\{g\}}^2/N = 0,0056/4 = 0,0014$. Этот же результат получается из матрицы дисперсий-ковариаций $(X^T P X s_{\{y\}}^2)^{-1}$ или в нашем примере

$$\begin{bmatrix} 1/8 & 0 & 0 \\ 0 & 1/8 & 0 \\ 0 & 0 & 1/8 \end{bmatrix} 0,0112 = \begin{bmatrix} 0,0014 & 0 & 0 \\ 0 & 0,0014 & 0 \\ 0 & 0 & 0,0014 \end{bmatrix}.$$

Тогда $s_{\{b_0\}}^2 = s_{\{b_1\}}^2 = s_{\{b_2\}}^2 = 0,0014$. Если при проверке адекватности используется $s_{\{g\}}^2$, то числитель F -критерия не нужно умножать на n , поскольку на это число уже поделен знаменатель. Для варианта с равномерным дублированием опытов на практике можно использовать следующую эквивалентную схему обработки результатов, учитывающую усреднение непосредственно.

1. Определим коэффициенты регрессии

$$b_j = \frac{\sum_{i=1}^N \bar{y}_i x_{ij}}{N}.$$

Матрица X в этом случае содержит только отличающиеся вектор-строки, а матрица $P=nE$.

2. Найдем дисперсию адекватности

$$s_{\text{ад}}^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2}{N - (k+1)}.$$

3. Оценим дисперсию среднего по строкам

$$s_i^2 = \frac{\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{n(n-1)}.$$

4. Проверим гипотезу об однородности дисперсий и после ее принятия найдем общую дисперсию среднего

$$s_{\{y\}}^2 = \sum_{i=1}^N s_i^2 / N.$$

5. Затем вычислим дисперсии оценок коэффициентов регрессии

$$s_{\{b_j\}}^2 = s_{\{g\}}^2 / N.$$

6. Наконец, проверим гипотезу адекватности модели

$$F = s_{\text{ад}}^2 / s_{\{y\}}^2.$$

Применим эту схему обработки результатов к нашему примеру. В табл. 10.5 повторены матрица планирования и средние значения откликов, а также приведены данные, необходимые при проверке адекватности модели. Коэффициенты регрессии $b_0 = 1,6125$;

Таблица 10.5

Матрица планирования и результаты опытов

Номер опыта	x_0	x_1	x_2	g	\bar{g}	$\bar{g} - \bar{g}$	$(\bar{g} - \bar{g})^2 \cdot 10^4$
1	+1	-1	-1	0,70	0,6375	0,0625	39
2	+1	-1	+1	1,40	1,4625	-0,0625	39
3	+1	+1	-1	1,70	1,7625	-0,0625	39
4	+1	+1	+1	2,65	2,5875	0,0625	39

$b_1 = 0,5625$, $b_2 = 0,4125$. Дисперсия адекватности $s_{\text{ад}}^2 = 156 \cdot 10^{-4}$. Дисперсии среднего по строкам (табл. 10.4) $s_1^2 = 0,01$, $s_2^2 = 0,01$, $s_3^2 = 0$, $s_4^2 = 0,0025$. Критерий Кохрена $G = 0,44$, гипотеза об однородности дисперсий принимается. Общая дисперсия среднего $s_{\{y\}}^2 = 0,0225/4 = 0,0056$. Дисперсия оценок коэффициентов регрессии $s_{\{b_j\}}^2 = 0,0056/4 = 0,0014$. Проверка гипотезы адекватности модели $F = 0,0156/0,0056 \approx 2,8$.

Пример 2. Пусть реализован план с неравномерным дублированием опытов [10], в котором первый опыт дублирован дважды. Матрица X имеет вид

$$X = \begin{bmatrix} +1 & -1 & -1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & +1 \end{bmatrix}.$$

Запишем решение системы нормальных уравнений с учетом весов

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \left(\begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & +1 \end{bmatrix} \right)^{-1} \times \\ \times \begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0,5(y_{11} + y_{12}) \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}.$$

Выполним указанные действия

$$\mathbf{X}^T \mathbf{P} = \begin{bmatrix} 2 & 1 & 1 & 1 \\ -2 & 1 & -1 & 1 \\ -2 & -1 & 1 & 1 \end{bmatrix}, \quad \mathbf{X}^T \mathbf{P} \mathbf{X} = \begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & 1 \\ -1 & 1 & 5 \end{bmatrix}.$$

Подчеркнем, что дублирование одного опыта нарушило ортогональность плана. Применение стандартных формул для подсчета коэффициентов регрессии, используемых в случае ортогональных планов, стало неправомерным.

Найдем обратную матрицу (см. [1])

$$(\mathbf{X}^T \mathbf{P} \mathbf{X})^{-1} = \frac{1}{28} \begin{bmatrix} 6 & 1 & 1 \\ 1 & 6 & -1 \\ 1 & -1 & 6 \end{bmatrix}.$$

Наконец,

$$\mathbf{X}^T \mathbf{P} \bar{\mathbf{Y}} = \begin{bmatrix} \frac{2(y_{11} + y_{12})}{2} + y_2 + y_3 + y_4 \\ \frac{-2(y_{11} + y_{12})}{2} + y_2 - y_3 + y_4 \\ \frac{-2(y_{11} + y_{12})}{2} - y_2 + y_3 + y_4 \end{bmatrix}.$$

В результате имеем

$$\mathbf{B} = \frac{1}{28} \begin{bmatrix} 6 & 1 & 1 \\ 1 & 6 & -1 \\ 1 & -1 & 6 \end{bmatrix} \cdot \begin{bmatrix} y_{11} + y_{12} + y_2 + y_3 + y_4 \\ -y_{11} - y_{12} + y_2 - y_3 + y_4 \\ -y_{11} - y_{12} - y_2 + y_3 + y_4 \end{bmatrix}.$$

Заметим, что к тому же результату можно придти, используя обычные формулы для нахождения оценок коэффициентов регрессии. Для этого в рассматриваемом примере достаточно сравнить соответствующие матрицы

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & -1 & +1 \\ -1 & -1 & -1 & +1 & +1 \end{bmatrix} \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & -1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & +1 \end{bmatrix} = \begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & 1 \\ -1 & 1 & 5 \end{bmatrix},$$

$$\mathbf{X}^T \mathbf{Y} = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & -1 & +1 \\ -1 & -1 & -1 & +1 & +1 \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \\ = \begin{bmatrix} y_{11} + y_{12} + y_2 + y_3 + y_4 \\ -y_{11} - y_{12} + y_2 - y_3 + y_4 \\ -y_{11} - y_{12} - y_2 + y_3 + y_4 \end{bmatrix}.$$

Положим, что в ходе экспериментирования производилось дублирование точек в соответствии с матрицей \mathbf{X} и получены результаты, изображаемые вектором \mathbf{Y}

$$\mathbf{X} = \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & -1 \\ \vdots & \vdots & \vdots \\ +1 & +1 & -1 \\ +1 & +1 & -1 \\ \vdots & \vdots & \vdots \\ +1 & +1 & +1 \\ +1 & +1 & +1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{matrix} n_1 \text{ раз} \\ n_2 \text{ раз} \\ n_4 \text{ раз} \end{matrix}; \quad \mathbf{Y} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{41} \\ y_{42} \\ \vdots \\ y_{4n_4} \end{bmatrix},$$

где n_i — число параллельных (дублированных) опытов в i -х условиях $i=1, 2, \dots, N$. Общее число всех опытов будет равно

$$\sum_{i=1}^N n_i = \sum_{i=1}^N (n_i - 1) + N.$$

Перейдем к более лаконичной форме записи условий и результатов эксперимента. Для этого введем матрицу весов \mathbf{P} . Это квадратная диагональная матрица с элементами $p_{ii} = n_i$

$$\mathbf{P} = \begin{bmatrix} n_1 & & 0 \\ & n_2 & \\ 0 & & n_3 \\ & & & n_4 \end{bmatrix}.$$

Эта матрица в совокупности с матрицей \mathbf{X} , содержащей только неповторяющиеся строки, задает условия эксперимента. А его ре-

зультаты тогда можно представить в виде вектора \bar{Y} , состоящего из средних наблюдений по дублированным опытам. Для учета различных вариантов дублирования составим таблицу, в которой приведем формулы для различных случаев (табл. 10.6).

Таблица 10.6

Суммы квадратов и числа степеней свободы при разных вариантах дублирования опытов

Источник рассеяния	Характер дублирования опытов	Суммы квадратов	Число степеней свободы
Ошибка эксперимента	Неравномерный	$\sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - y_i)^2$	$\sum_{i=1}^N (n_i - 1)$
	Равномерный	$\sum_{i=1}^N \sum_{q=1}^n (y_{iq} - y_i)^2$	$N(n - 1)$
	Дублирование в одной точке $i = 1$	$\sum_{q=1}^{n_1} (y_{1q} - y_1)^2$	$n_1 - 1$
	Дублирование в отдельной серии из L опытов	$\sum_{i=1}^L (y_i - \bar{y})^2$	$L - 1$
Неадекватность модели	Неравномерный	$\sum_{i=1}^N n_i (\bar{y}_i - \hat{y}_i)^2$	$N - (k + 1)$
	Равномерный	$n \sum_{i=1}^N (y_i - \hat{y}_i)^2$	$N - (k + 1)$
	Дублирование в одной точке $i = 1$	$n_1 (y_1 - \hat{y}_1)^2 + \sum_{i=2}^N (y_i - \hat{y}_i)^2$	$N - (k + 1)$
	Дублирование в отдельной серии из L опытов	$\sum_{i=1}^N (y_i - \hat{y}_i)^2$	$N - (k + 1)$

Для удобства обозначений условимся в случае равенства всех n_i писать $n_i = \text{const} = n$, а номер единственного опыта, который дублируется, не нарушая общности, будем считать первым. В табл. 10.6 также указаны формулы для случая, когда дисперсия, характеризующая ошибку опыта, находится из независимой серии опытов, результаты которой не используются при вычислении коэффициентов регрессии.

Таким образом, неравномерность дублирования должна учитываться и при регрессионном и при дисперсионном анализе. На практике это делается не всегда.

Приведенная таблица нуждается в комментариях. Начнем с пояснения содержащихся в ней формул. Для рассмотрения частных случаев, возможных при различных стратегиях дублирования, выразим соответствующие суммы квадратов с помощью обычной формы записи. Прежде всего запишем выражение для остаточной суммы квадратов

$$S_{\text{ост}} = \sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \hat{y}_i)^2.$$

Разложим эту сумму на составляющие с помощью следующего преобразования:

$$S_{\text{ост}} = \sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i + \bar{y}_i - \hat{y}_i)^2,$$

где \bar{y}_i — среднее значение отклика по опытам i -й серии

$$\bar{y}_i = \frac{1}{n_i} \sum_{q=1}^{n_i} y_{iq}.$$

Раскрывая выражение для $S_{\text{ост}}$, будем иметь

$$S_{\text{ост}} = \sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)^2 + \sum_{i=1}^N \sum_{q=1}^{n_i} (\bar{y}_i - \hat{y}_i)^2 + + 2 \sum_{i=1}^N (\bar{y}_i - \hat{y}_i) \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i).$$

Сумма $\sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)$, входящая в последнее слагаемое, равна нулю

$$\sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i) = \sum_{q=1}^{n_i} y_{iq} - n_i \bar{y}_i = 0.$$

Поэтому

$$S_{\text{ост}} = \sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)^2 + \sum_{i=1}^N n_i (\bar{y}_i - \hat{y}_i)^2$$

дает разложение $S_{\text{ост}}$ на два слагаемых, первое из которых есть $S_{\text{востр}}$, а второе — $S_{\text{ах}}$.

С величиной $S_{\text{ост}}$ связано число степеней свободы, равное

$$f_{\text{ост}} = \sum_{i=1}^N n_i - p,$$

где p — число коэффициентов регрессии, входящих в регрессионную зависимость (в линейном случае $p=k+1$).

Величина

$$S_{\text{воспр}} = \sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)^2$$

связана с числом степеней свободы

$$f_{\text{воспр}} = \sum_{i=1}^N (n_i - 1).$$

Следовательно, дисперсия, характеризующая ошибку эксперимента, есть

$$s_{\{y\}}^2 = \frac{S_{\text{воспр}}}{f_{\text{воспр}}} = \frac{\sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)^2}{\sum_{i=1}^N (n_i - 1)}.$$

Тот же результат будет, если использовать выражения для частных дисперсий в каждой серии дублированных опытов. Действительно, для i -й серии параллельных опытов можно записать

$$s_i^2 = \frac{S_{\text{воспр } i}}{f_{\text{воспр } i}} = \frac{\sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)^2}{n_i - 1}.$$

Тогда объединенная оценка дисперсии, характеризующей ошибку опыта, может быть найдена из соотношения

$$s_{\{y\}}^2 = \frac{f_{\text{воспр } 1}}{\sum_{i=1}^N f_{\text{воспр } i}} s_1^2 + \frac{f_{\text{воспр } 2}}{\sum_{i=1}^N f_{\text{воспр } i}} s_2^2 + \dots + \frac{f_{\text{воспр } N}}{\sum_{i=1}^N f_{\text{воспр } i}} s_N^2.$$

Применение этой формулы оправданно лишь при условии однородности частных дисперсий. Однородность дисперсий может быть проверена в общем случае с помощью критерия Бартлета. Запишем нашу формулу в виде

$$s_{\{y\}}^2 = \sum_{i=1}^N s_{\text{воспр } i}^2 f_i / \sum_{i=1}^N f_{\text{воспр } i}.$$

Отсюда следует, что она идентична с предыдущей.

Теперь вернемся к сумме квадратов

$$S_{\text{ак}} = \sum_{i=1}^N n_i (\bar{y}_i - \hat{y}_i)^2.$$

Она связана с числом степеней свободы $f_{\text{ак}} = N - (k+1)$ или в общем случае $f_{\text{ак}} = N - p$. Следовательно, дисперсия, связанная с неадекватностью уравнения регрессии, есть

$$s_{\text{ак}}^2 = \frac{S_{\text{ак}}}{f_{\text{ак}}} = \frac{\sum_{i=1}^N n_i (\bar{y}_i - \hat{y}_i)^2}{N - (k+1)}.$$

Это позволяет проверить гипотезу о том, что экспериментальные данные не противоречат полученному описанию. Проверка гипотезы производится с помощью дисперсионного отношения Фишера

$$F = s_{\text{ак}}^2 / s_{\{y\}}^2.$$

Таким образом, содержащиеся в таблице формулы являются частными случаями одного и того же разложения остаточной суммы квадратов $S_{\text{ост}}$. Зная процедуру разложения остаточной суммы квадратов, можно всякий раз получить нужную формулу. Однако на практике более удобно иметь эти формулы в готовом виде.

Наиболее общий случай представляет собой неравномерное дублирование. Чтобы получить здесь значение F -критерия, нужно сумму квадратов, связанную с дисперсией воспроизводимости, разделить на соответствующее ей число степеней свободы и получить знаменатель F -критерия (если построчные дисперсии однородны по критерию Бартлета)

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{q=1}^{n_i} (y_{iq} - \bar{y}_i)^2}{\sum_{i=1}^N (n_i - 1)}.$$

Затем надо взять сумму квадратов, связанную с дисперсией адекватности, поделить на соответствующее ей число степеней свободы и получить числитель

$$s_{\text{ак}}^2 = \frac{\sum_{i=1}^N n_i (\bar{y}_i - \hat{y}_i)^2}{N - p}.$$

Аналогично поступают во всех других случаях. Вариант равномерного дублирования уже неоднократно обсуждался. При дублировании в одной строке матрицы (в нашей таблице эта строка имеет индекс 1, хотя ясно, что выбор строк произволен) симметрия нарушается, это находит отражение в формулах. Если дела экспериментатора так плохи, что из-за экономии времени приходится ставить параллельные опыты только в одной точке, то часто лучше всего выбирать центр плана, ибо при этом не нарушается ортогональность.

Последний из приведенных в таблице вариантов — дублирование в отдельной серии из L опытов, не входящих в план. Это означает, что параллельные опыты не ставятся, а информация об ошибке опыта черпается из какого-то независимого источника, например, из публикаций предшественников.

Перейдем теперь к рассмотрению требований, которые обычно предъявляются к планам и называются критериями оптимальности.

10.6. Критерии оптимальности планов

Построение плана эксперимента можно интерпретировать как выбор строк матрицы X , их числа и последовательности проведения. Этот выбор осуществляется разными способами и соответственно приводит к разным результатам. Это значит, что b -коэффициенты могут быть оценены с разной точностью, что они будут иметь разные ковариации, что предсказанное значение отклика получится с разными дисперсиями и т. д. В зависимости от того, какие требования экспериментатор предъявляет к модели, он может придти к той или иной формулировке требований к матрице X . Формализация этих требований связана с критериями оптимальности. Критерии оптимальности удобно формулировать в терминах свойств матрицы $M = X^T X$ или матрицы M^{-1} . Именно эти матрицы непосредственно связаны с оценками модели и функционально зависят от матрицы X . Так, например, при диагональной матрице M план оказывается ортогональным, т. е. все столбцы матрицы X взаимноортогональны и коэффициенты модели независимы: $\text{cov}\{b_i, b_j\} = 0$.

Таким образом, мы пришли к критерию, который уже был рассмотрен в шестой главе и который является одним из самых существенных для планов, обсуждаемых в нашей книге. Он относится к группе критериев, связанных с оценками свойств коэффициентов. Кроме этой группы критериев будем различать критерии, определяющие предсказательные свойства модели, и критерии, сформулированные без использования матрицы M , такие, как композиционность, возможность разбиения плана на ортогональные блоки, насыщенность и т. д. [11,12].

Начнем рассмотрение с критериев первой группы. В их основе лежит концепция совместных эффективных оценок, которая берет свое начало с работ Р. Фишера, рассматривавшего проблему получения наилучших оценок при обработке экспериментальных данных.

Американский математик Дж. Кифер распространил такое требование на задачи построения планов. Различные критерии этой группы приведены в табл. 10.7. В таблице каждый критерий определяется тремя различными, но эквивалентными способами. Таблица начинается с критерия D -оптимальности,

Таблица 10.7

Критерии оптимальности планов для оценок коэффициентов модели

Критерий	Интерпретация		
	алгебраическая	геометрическая	статистическая
D -оптимальность	Минимум определителя матрицы дисперсий-ковариаций $\min \det M^{-1}$	Минимум объема эллипсоида рассеяния оценок параметров	Минимум обобщенной дисперсии
A -оптимальность	Минимум следа матрицы дисперсий-ковариаций $\min \text{tr} M^{-1}$	Минимум суммы квадратов длин осей эллипсоида рассеяния и минимум длин диагоналей прямоугольника, описанного около эллипсоида	Минимальная средняя дисперсия оценок коэффициентов
E -оптимальность	Минимум собственного значения матрицы дисперсий-ковариаций $\min \max \lambda (M^{-1})$	Минимум максимальной проекции оси эллипсоида рассеяния на координатные оси пространства параметров	Отдельные оценки параметров не обладают слишком большими дисперсиями и ковариациями
Минимум дисперсии оценок коэффициентов	Минимум диагонального элемента матрицы дисперсий-ковариаций $\min \max s\{b_j\}$	Минимум максимальной проекции оси эллипсоида рассеяния на координатные оси пространства параметров	Минимум дисперсии оценок коэффициентов
Минимум суммы относительных ошибок оценок	$\min \sum_{j=0}^k \frac{s\{b_j\}}{b_j}$		Минимум суммы относительных ошибок оценок
Ортогональность	Диагональность матрицы дисперсий-ковариаций $\sum_{i=1}^N x_i x_{ij} = \begin{cases} 0 & j \neq u \\ 1 & j = u \end{cases}$	Направление главных осей эллипсоида рассеяния совпадает с направлением координатных осей пространства параметров	Оценки параметров независимы $\text{cov}\{b_j, b_u\} = 0 \quad j \neq u$

с названием которого часто связывается вся концепция оптимального построения планов. Смысл его — минимизация дисперсии всех коэффициентов регрессии, рассматриваемых как нечто единое, как вектор. Дисперсию вектора коэффициентов принято называть *обобщенной дисперсией*, которая задается известной нам функцией от матрицы дисперсий-ковариаций — *определителем*. Чем меньше определитель, тем меньше обобщенная дисперсия. Как всегда в математике наряду с алгебраическим представлением можно использовать и геометрическое. Однако в этом случае вместо уже знакомого нам факторного пространства приходится вводить *пространство параметров*, в котором координатные оси задаются значениями коэффициентов регрессии.

Между размерностями факторного пространства и пространства параметров не существует однозначной связи. В случае одного фактора и линейного уравнения регрессии пространство параметров будет двумерным (так как в уравнение входит два коэффициента b_0 и b_1). Но можно представить себе вариант перехода в одномерное пространство, если, например, незначим один из коэффициентов, или в трехмерное пространство, когда от уравнения прямой приходится переходить к параболе. Аналогичное рассуждение имеет место и для большего числа факторов.

Если число факторов и вид уравнения заданы, тогда размерность пространства параметров определяется однозначно. Так, для линейного уравнения с k факторами пространство параметров имеет размерность $k+1$. Каждой точке в таком пространстве соответствует вектор оценок коэффициентов, а определитель матрицы дисперсий-ковариаций пропорционален объему *эллипсоида рассеяния* оценок параметров. Причем центр эллипсоида совмещен с МНК оценкой. Эллипсоид является многомерным аналогом обычного доверительного интервала. Для одномерного пространства параметров он вырождается в отрезок, который и есть доверительный интервал. На плоскости (когда имеется два параметра) получится эллипс. Эллиптическая форма доверительного интервала связана с предпосылкой регрессионного анализа о нормальном распределении.

Геометрическая интерпретация в пространстве параметров характерна для критериев этой группы.

Для критерия A -оптимальности матрица X выбирается так, чтобы достигнуть минимума суммы квадратов длин главных осей эллипсоида рассеяния. Алгебраически это соответствует минимуму еще одной функции матрицы дисперсии-ковариации, которая называется *следом* и обозначается $\text{tr } M^{-1}$ (от trase фр.) или $\text{sp } M^{-1}$ (от Spur нем.). Следом квадратной матрицы называется сумма ее диагональных элементов. Вспомним, что на диагонали матрицы M^{-1} находятся дисперсии b -коэффициентов. Значит A -оптимальность обеспечивает минимум суммы дисперсий b -коэффициентов без учета их ковариаций и, следовательно, минимум средней дисперсии.

Кроме таких функций от матрицы M^{-1} как определитель и след для критериев оптимальности можно использовать и другие ее функции. Примером может служить критерий E -оптимальности, минимизирующий максимальное *собственное значение* этой матрицы. Собственное значение принадлежит к характеристикам структуры матрицы. Этому вопросу посвящена обширная литература, к которой мы и отсылаем читателя [13, 14]. Всегда существует опасность, что эллипсоид рассеяния может получить слишком вытянутую, бананоподобную форму. При этом некоторые b -коэффициенты попадут в неблагоприятные условия. Критерий E -оптимальности позволяет уменьшить эту опасность, поскольку он минимизирует самую длинную ось эллипсоида рассеяния.

Подобным образом можно интерпретировать и остальные критерии, приведенные в табл. 10.7, что предоставляется читателю в качестве самостоятельного упражнения.

Перейдем к рассмотрению критериев второй группы. Описание этих критериев проводится не на языке пространства параметров, а в более привычных терминах факторного пространства и функции отклика (см. табл. 10.8). G - и Q -критерии связаны

Таблица 10.8

Критерии оптимальности планов для предсказательных свойств модели

Критерий	Интерпретация	
	алгебраическая	статистическая
G -оптимальность	$\min \max X_i^T M^{-1} X_i$	Минимум максимального значения дисперсии оценки поверхности отклика
Q -оптимальность	$\min \sum_{i=1}^N X_i^T M^{-1} X_i$	Минимум средней дисперсии оценки поверхности отклика
Ротатабельность	$X_i^T M^{-1} X_i = f(\rho)$	Постоянство дисперсии предсказания на равных расстояниях от центра эксперимента
Униформность	$X_i^T M^{-1} X_i \approx \text{const}$ при $0 \leq \rho \leq 1$	Дисперсия предсказания постоянна в некоторой области вокруг центра эксперимента (например, с единичным радиусом)

с дисперсией предсказания значений отклика. G -критерий минимизирует максимальную дисперсию предсказания, Q -критерий — среднюю дисперсию. Если план G -оптимален, то экспериментатор имеет гарантию, что в области планирования не окажется точек, в которых точность оценки поверхности отклика будет слишком *низкая*.

Критерии ротатабельности и униформности связаны с требованием постоянства дисперсии предсказания на некоторых фиксированных расстояниях от центра эксперимента. Ротатабельность плана означает, что оценки дисперсии предсказания инвариантны (независимы) относительно вращения координатных осей факторного пространства. Иными словами, дисперсия предсказания не будет зависеть от того, в каком направлении осуществляется движение из начала координат, а зависит только от расстояния между интересующей нас точкой и началом. Униформность в дополнение к этому требует, чтобы в некоторой окрестности начала координат, обычно внутри сферы единичного радиуса, дисперсия предсказания оставалась приблизительно постоянной.

Кроме этих двух групп критериев, при упоминании которых мы не претендовали на полноту перечисления, существует еще большое число требований, принимаемых во внимание. Укажем некоторые из них.

Наиболее естественное желание экспериментатора — уменьшение числа опытов. Минимальное число опытов задается числом коэффициентов модели, а приближение к нему служит мерой *насыщенности* плана. Таким образом, насыщенность плана оказывается одним из возможных критериев оптимальности. Именно стремление удовлетворить этому критерию привело к созданию дробных реплик. Заметим, что дробные реплики одновременно удовлетворяют по крайней мере двум критериям: они ортогональны и насыщены.

Большое значение имеет требование *композиционности*, позволяющее разделить эксперимент на части и в случае необходимости последовательно реализовывать одну часть за другой без потери информации. Примером может служить переход от $1/4$ -реплики к полуреплике, а затем к полному факторному эксперименту, если система смешивания оказалась сложной и не позволила выделить интересующие эффекты. Это соответствует последовательному переходу от простой линейной модели к модели с взаимодействиями. То же самое возможно и при переходе от линейных моделей к модели второго порядка и т. д.

Близким к требованию композиционности является требование *разбиения плана на ортогональные блоки*. Композиционность требуется из-за того, что мы не знаем заранее, какой окажется адекватная модель, а разбивать на блоки приходится из-за того, что возможно изменение внешних условий и важно защититься от их влияния. С примерами разбиения на блоки вы могли ознакомиться в гл. 8.

Список возможных критериев отнюдь не исчерпывается описанным [11]. Но даже если ограничиться перечисленным, то естественно может возникнуть вопрос, зачем нужны простому экспериментатору D -, G - и прочие оптимальности?

Мы уже говорили выше о таких свойствах полного факторного эксперимента и дробных реплик, как ортогональность и ротатабельность. Ортогональность обеспечивает независимость оценок b -коэффициентов, что очень существенно при интерпретации силы влияния факторов и их взаимодействий, а также, как мы увидим далее, для оценки направления градиента при движении к оптимуму. Поскольку заранее трудно предполагать, в какую сторону будет направлен градиент, то полезно стремиться и к ротатабельности, обеспечивающей одинаковую точность предсказания в разных направлениях.

Рассмотрим D -оптимальность. Этот критерий связан с оценкой уравнения регрессии в целом и смысл его становится более понятным, когда речь идет об интерполяционных (описательных) задачах. Здесь уже экспериментатор не стремится изучать влияние каждого фактора в отдельности. Для него важно получить минимальную обобщенную дисперсию коэффициентов, что обеспечивает достаточные предсказательные свойства модели внутри области эксперимента. Такие же свойства обеспечиваются и G -критерием. Это происходит не случайно, так как D - и G -критерии во многих случаях эквивалентны [12].

Критерии первой группы отличаются от D -критерия тем, что используют иные свойства вектора оценок коэффициентов регрессии, отличные от обобщенной дисперсии. Могут встретиться постановки задач, требующие того или иного критерия. Однако экспериментатор обычно не может осуществить этот выбор самостоятельно [12].

Планы, рассматриваемые в нашей книге, обладают тем свойством, что одновременно являются оптимальными во многих смыслах. Они не только ортогональны и ротатабельны, но еще и D -, G -, A - и E -оптимальны. Такое совпадение имеет место всегда, когда $\mathbf{X}^T\mathbf{X} = N\mathbf{E}$, где \mathbf{E} — единичная матрица. Матрица \mathbf{X} , состоящая из $+1$ и -1 и удовлетворяющая этому условию, называется матрицей Адамара. Очевидно, матрицы полного факторного эксперимента 2^k и некоторых регулярных дробных реплик принадлежат к этому классу.

Таким образом, рассматриваемые в данной книге планы попутно с теми критериями, на основе которых они строились, удовлетворяют еще и ряду других критериев. Иногда критерии могут оказываться противоречивыми. Тогда выбор наиболее подходящего из них требует большого искусства. Именно искусства, потому что общей теории выбора критериев оптимальности пока не существует [11].

На этом мы заканчиваем главу, в которой вы познакомились с матричным аппаратом регрессионного анализа.

Литература

1. С. Сирл, У. Госман. Матричная алгебра в экономике. М., «Статистика», 1974.
2. Дж. Хедли. Линейная алгебра. М., «Высшая школа», 1966.
3. Р. Беллман. Введение в теорию матриц. М., «Наука», 1969.
4. Ф. Р. Гантмахер. Теория матриц. Изд. 3-е. М., «Наука», 1967.
5. Д. К. Фадеев, В. Н. Фадеева. Вычислительные методы линейной алгебры. М., Физматгиз, 1960.
6. Дж. Форсайт, К. Молер. Численное решение систем линейных алгебраических уравнений. М., «Мир», 1969.
7. Н. Дрейпер, Г. Смит. Прикладной регрессионный анализ. М., «Статистика», 1973.
8. Д. Химмельблау. Анализ процессов статистическими методами. М., «Мир», 1973.
9. В. В. Налимов. Применение математической статистики при анализе вещества. М., Физматгиз, 1960.
10. В. Г. Горский, Ю. П. Адлер. О регрессионном и дисперсионном анализе при планировании эксперимента с неравномерным дублированием опытов. — Заводская лаборатория, 1971, 37, № 3, с. 319—325.
11. В. В. Налимов, Т. И. Голикова. Логические основания планирования эксперимента. — Препринт № 20 Лаборат. статистич. методов, изд-во МГУ. М., 1971.
12. В. В. Федоров. Теория оптимального эксперимента. М., «Наука», 1971.
13. К. Ланцош. Практические методы прикладного анализа. М., Физматгиз, 1962.
14. Д. Х. Уилкинсон. Алгебраическая проблема собственных значений. М., «Наука», 1971.

Глава одиннадцатая

ПРИНЯТИЕ РЕШЕНИЙ ПОСЛЕ ПОСТРОЕНИЯ МОДЕЛИ

Не нам, господа, подражать Плинию.
Наше дело выравнивать линию.

К. Прутков

11.1. Интерпретация результатов

Адекватная линейная модель, которой мы теперь располагаем, имеет вид полинома первой степени. Коэффициенты полинома являются частными производными функции отклика по соответствующим переменным. Их геометрический смысл — тангенсы углов наклона гиперплоскости к соответствующей оси. Большой по абсолютной величине коэффициент соответствует большему углу наклона и, следовательно, более существенному изменению параметра оптимизации при изменении данного фактора.

До сих пор мы употребляли абстрактный математический язык. Перевод модели на язык экспериментатора называется интерпретацией модели.

Задача интерпретации весьма сложна. Ее решают в несколько этапов. Первый этап состоит в следующем. Устанавливается, в какой мере каждый из факторов влияет на параметр оптимизации. Величина коэффициента регрессии — количественная мера этого влияния. Чем больше коэффициент, тем сильнее влияет фактор. О характере влияния факторов говорят знаки коэффициентов. Знак плюс свидетельствует о том, что с увеличением значения фактора растет величина параметра оптимизации, а при знаке минус — убывает. Интерпретация знаков при оптимизации зависит от того, ищем ли мы максимум или минимум функции отклика. Если $y \rightarrow \max$, то увеличение значений всех факторов, коэффициенты которых имеют знак плюс, благоприятно, а имеющих знак минус — неблагоприятно. Если же $y \rightarrow \min$, то, наоборот, благоприятно увеличение значений тех факторов, знаки коэффициентов которых отрицательны.

Далее выясняется, как расположить совокупность факторов в ряд по силе их влияния на параметр оптимизации. Факторы, коэффициенты которых незначимы, конечно не интерпретируются. Можно сказать только, что при данных интервалах варьирования и ошибке воспроизводимости они не оказывают существенного влияния на параметр оптимизации.

Изменение интервалов варьирования приводит к изменению коэффициентов регрессии. Абсолютные величины коэффициентов

регрессии увеличиваются с увеличением интервалов. Инвариантами к изменению интервалов остаются знаки линейных коэффициентов регрессии. Однако и они изменятся на обратные, если при движении по градиенту (гл. 12) мы «проскочим» экстремум.

В некоторых задачах представляет интерес построение уравнения регрессии для натуральных значений факторов. Уравнение для натуральных переменных можно получить, используя формулу перехода (стр. 72). Коэффициенты регрессии изменятся. При этом пропадает возможность интерпретации влияния факторов по величинам и знакам коэффициентов регрессии. Вектор-столбцы натуральных значений переменных в матрице планирования уже не будут ортогональными, коэффициенты определяются зависимо друг от друга. Если же поставлена задача получения интерполяционной формулы для натуральных переменных, такой прием допустим.

Пример 1. Определение оптимальных условий ионнообменного разделения неодима и празеодима (стр. 70) $y \rightarrow \max$.

Вспомним, что \bar{x}_1 — концентрация промывающего раствора (элюанта), \bar{x}_2 — pH этого же раствора, y — процентное содержание неодима в выходящем растворе — элюате.

После обработки экспериментальных данных получено уравнение регрессии

$$\hat{y} = 88,0 - 2,0x_1 - 4,5x_2, \quad s_{\{b_j\}} = 0,30.$$

К увеличению параметра оптимизации приводит уменьшение значений факторов. Подставим в уравнение регрессии разные кодированные значения факторов и посмотрим, при каких значениях факторов увеличивается параметр оптимизации.

Если подставить в уравнение значения $x_1 = +1$ и $x_2 = +1$, то получится

$$y = 88,0 - 2,0(+1) - 4,5(+1) = 81,5.$$

Теперь подставим значения $x_1 = -1$ и $x_2 = -1$

$$y = 88,0 - 2,0(-1) - 4,5(-1) = 94,5.$$

Уменьшение значений факторов действует благоприятно. Для увеличения параметра оптимизации нужно уменьшать значения факторов.

Запомните правило: если коэффициент регрессии отрицателен, то для увеличения параметра оптимизации надо уменьшать значение фактора, а если положителен, то увеличивать.

При минимизации параметра оптимизации можно изменить знаки коэффициентов (кроме b_0) на обратные и поступать, как в первом случае.

Теперь мы получили основу для перехода к следующему этапу. Априорные сведения дают некоторые представления о характере действия факторов. Источниками таких сведений могут служить теория изучаемого процесса, опыт работы с аналогичными процессами или предварительные опыты и т. д.

Если, например, ожидается, что с ростом температуры должно происходить увеличение параметра оптимизации, а коэффициент регрессии имеет знак минус, то возникает противоречие. Возможны две причины возникновения такой ситуации: либо в эксперименте допущена ошибка и он должен быть подвергнут ревизии, либо неверны априорные представления. Нужно иметь в виду, что эксперимент проводится в локальной области факторного пространства и коэффициент отражает влияние фактора только в этой области. Заранее не известно, в какой мере можно распространить результат на другие области. Теоретические же представления имеют обычно более общий характер. Кроме того, априорная информация часто основывается на однофакторных зависимостях. При переходе к многофакторному пространству ситуация может изменяться. Поэтому мы должны быть уверены, что эксперимент проведен корректно. Тогда для преодоления противоречия можно выдвигать различные гипотезы и проверять их экспериментально. Эксперименты по проверке гипотез тоже можно планировать, но эти задачи здесь мы не рассматриваем.

В тех, довольно редких, случаях, когда имеется большая априорная информация, позволяющая выдвигать гипотезы о механизме явлений, можно перейти к следующему этапу интерпретации. Он сводится к проверке гипотез о механизме явлений и выдвижению новых гипотез.

Получение информации о механизме явлений не является обязательным в задачах оптимизации, но возможность такого рода следует использовать. Здесь особое внимание приходится уделять эффектам взаимодействия факторов. Как их интерпретировать?

Пусть в некоторой задаче взаимодействие двух факторов значимо и имеет положительный знак. Это свидетельствует о том, что одновременное увеличение, как и одновременное уменьшение, значений двух факторов приводит к увеличению параметра оптимизации (без учета линейных эффектов). А если эффект взаимодействия факторов x_1 и x_2 имеет отрицательный знак? Любая комбинация разных знаков x_1 и x_2 приводит к росту параметра оптимизации.

Запомните правило: если эффект взаимодействия имеет положительный знак, то для увеличения параметра оптимизации требуется одновременное увеличение или уменьшение значений факторов, например сочетания: $x_1 = +1$ и $x_2 = +1$ или $x_1 = -1$ и $x_2 = -1$. Для уменьшения параметра оптимизации факторы должны одновременно изменяться в разных направлениях, например $x_1 = +1$ и $x_2 = -1$ или $x_1 = -1$ и $x_2 = +1$.

Если эффект взаимодействия имеет отрицательный знак, то для увеличения параметра оптимизации факторы должны одновременно изменяться в разных направлениях, например

$$x_1 = +1 \text{ и } x_2 = -1 \text{ или } x_1 = -1 \text{ и } x_2 = +1.$$

Для уменьшения параметра оптимизации требуется одновременное увеличение или уменьшение факторов, т. е.

$$x_1 = +1 \text{ и } x_2 = +1, \text{ или } x_1 = -1 \text{ и } x_2 = -1.$$

Вы видите, что интерпретация эффектов взаимодействия не так однозначна, как линейных эффектов. В каждом случае имеется два варианта. Какому из вариантов отдавать предпочтение? Прежде всего нужно учесть знаки линейных эффектов соответствующих факторов. Если эффект взаимодействия имеет знак плюс и соответствующие линейные эффекты отрицательны, то выбор однозначен: сочетание -1 и -1 . Однако возможен случай, когда знаки линейных эффектов различны. Тогда приходится учитывать численные значения коэффициентов и жертвовать самым малым эффектом.

Иногда приходится учитывать технологические соображения: например, эксперимент в одной области факторного пространства дороже (или труднее), чем в другой.

Пример 2. Рассмотрим один из простейших примеров интерпретации, связанной с гипотезами о механизме действия факторов (см. стр. 118). Изучалось влияние трех факторов на выход сульфадимизина. По поводу влияния концентрации уксусной кислоты x_3 априори выдвигалась следующая гипотеза. Предполагалось, что уксусная кислота является растворителем, не участвующим в процессе. Из уравнения регрессии

$$\hat{y} = 85,975 + 2,588x_1 + 0,568x_2 + 1,125x_3 - 0,588x_1x_3 - 0,918x_2x_3, \quad (s_{\{b\}} = 0,28)$$

видно, что существенным оказался не только b_3 , но также b_{13} и b_{23} . Этот факт ставит под сомнение первоначальную гипотезу, и можно предположить, что уксусная кислота активно участвует в процессе.

Заканчивая этот параграф, упомянем еще об интерпретации эффектов взаимодействия высоких порядков. Если значимым оказался эффект взаимодействия трех факторов, например $x_1x_2x_3$, то его можно интерпретировать следующим образом. Этот эффект может иметь знак плюс, если отрицательные знаки будут у четного числа факторов (ноль или любые два). Знак минус будет, если нечетное число факторов имеет знак минус (все три или любой один). Это правило распространяется на взаимодействия любых порядков. Пользуются еще таким приемом. Произведение двух факторов условно считают одним фактором и сводят трехфакторное взаимодействие к парному и т. д.

Мы сказали, что интерпретация результатов — это перевод с одного языка на другой. Такой перевод обеспечивает взаимопонимание между статистиком и экспериментатором, работающими совместно над задачами оптимизации. Интерпретация уравнения регрессии важна не только для понимания процесса, но и для принятия решений при оптимизации.

11.2. Принятие решений после построения модели процесса

Нам придется принимать решения в сложных ситуациях. Решения зависят от числа факторов, дробности плана, цели исследования (достижение оптимума, построение интерполяционной формулы) и т. д. Количество возможных решений по примерной оценке достигает нескольких десятков тысяч. Поэтому мы будем рассматривать только наиболее часто встречающиеся нам случаи и выделим «типичные» решения. Положение здесь сложнее, чем в случае принятия решений о выборе основного уровня и интервалов варьирования факторов (гл. 6), где удалось рассмотреть все варианты. Ситуации будем различать по адекватности и неадекватности модели, значимости и незначимости коэффициентов регрессии в модели, информации о положении оптимума.

Обсудим сначала принятие решения для адекватного линейного уравнения регрессии.

Линейная модель адекватна. Здесь возможны три варианта ситуации: 1) все коэффициенты регрессии значимы; 2) часть коэффициентов регрессии значима, часть незначима; 3) все коэффициенты регрессии незначимы.

В каждом варианте оптимум может быть близко, далеко или о его положении нет информации (неопределенная ситуация). Рассмотрим первый вариант.

Если область оптимума близка, возможны три решения: окончание исследования, переход к планам второго порядка и движение по градиенту.

Переход к планированию второго порядка дает возможность получить математическое описание области оптимума и найти экстремум. Хотя мы и не рассматриваем вопросы построения планов второго порядка, эту возможность надо также учитывать. Подробные рекомендации по применению планирования второго порядка вы найдете, например, в руководстве [1].

Движение по градиенту используется при малой ошибке опыта, поскольку на фоне большой ошибки трудно установить приращение параметра оптимизации.

Решение при неопределенной ситуации или удаленной области оптимума одно и то же: движение по градиенту.

Второй вариант — часть коэффициентов регрессии значима, часть незначима. Вам пока придется поверить, что движение по градиенту наиболее эффективно, если коэффициенты значимы. Поэтому выбираются решения, реализация которых приводит к получению значимых коэффициентов. На этом этапе важно выдвинуть гипотезы, объясняющие незначимость эффектов. Это может быть и неудачный выбор интервалов варьирования, и включение (из осторожности) факторов, не влияющих на параметр оптимизации, и большая ошибка опыта, и т. д. Решение зависит от того, какую гипотезу мы предпочитаем.

Если, например, выдвинута первая гипотеза, то возможно такое решение: расширение интервалов варьирования по незначимым факторам и постановка новой серии опытов. Изменение интервалов варьирования иногда сочетают с переносом центра эксперимента в точку, соответствующую условиям наилучшего опыта. Невлияющие факторы стабилизируются и исключаются из дальнейшего рассмотрения. Другие возможные решения для получения значимых коэффициентов: увеличение числа параллельных опытов и доработка плана. Увеличение числа параллельных опытов приводит к уменьшению дисперсии воспроизводимости и соответственно дисперсии коэффициентов регрессии. Опыты могут быть повторены либо во всех точках плана, либо в некоторых.

Доработка плана осуществляется несколькими способами: методом «перевала» — у исходной реплики изменяют знаки на обратные (в этом случае основные эффекты оказываются не смешанными с парными эффектами взаимодействия); переходом к полному факторному эксперименту; переходом к реплике меньшей дробности; переходом к плану второго порядка (если область оптимума близка).

Реализация любого из этих решений требует значительных экспериментальных усилий. Поэтому иногда можно и не следовать строго правилу «двигайтесь по всем факторам», а пойти на некоторый риск и двигаться только по значимым факторам.

Наконец, если область оптимума близка, то возможно принятие таких же решений, как и в случае значимости всех коэффициентов регрессии.

Рассмотрим последний вариант: линейная модель адекватна, все коэффициенты регрессии незначимы (кроме b_0). Чаще всего это происходит вследствие большой ошибки эксперимента или узких интервалов варьирования. Поэтому возможные решения направлены прежде всего на увеличение точности эксперимента и расширение интервалов варьирования. Увеличение точности, как вы уже знаете, может достигаться двумя путями: благодаря улучшению методики проведения опытов или вследствие постановки параллельных опытов.

Если область оптимума близка, то возможно также окончание исследования.

В заключение приведем блок-схему принятия решения в задаче определения оптимальных условий, линейная модель адекватна (рис. 28). В блок-схеме пунктирными линиями обведены ситуации, сплошными линиями — принимаемые решения.

Пример 3. При определении оптимальных условий технологического процесса получения волокна из полипропилена в качестве независимых переменных были выбраны: \bar{x}_1 — температура расплава, °С; \bar{x}_2 — давление расплава, кг/см²; \bar{x}_3 — скорость намотки на бобину, м/мин; \bar{x}_4 — температура нагревателей, °С; \bar{x}_5 — скорость вытягивания, м/мин; \bar{x}_6 — кратность вытягивания.

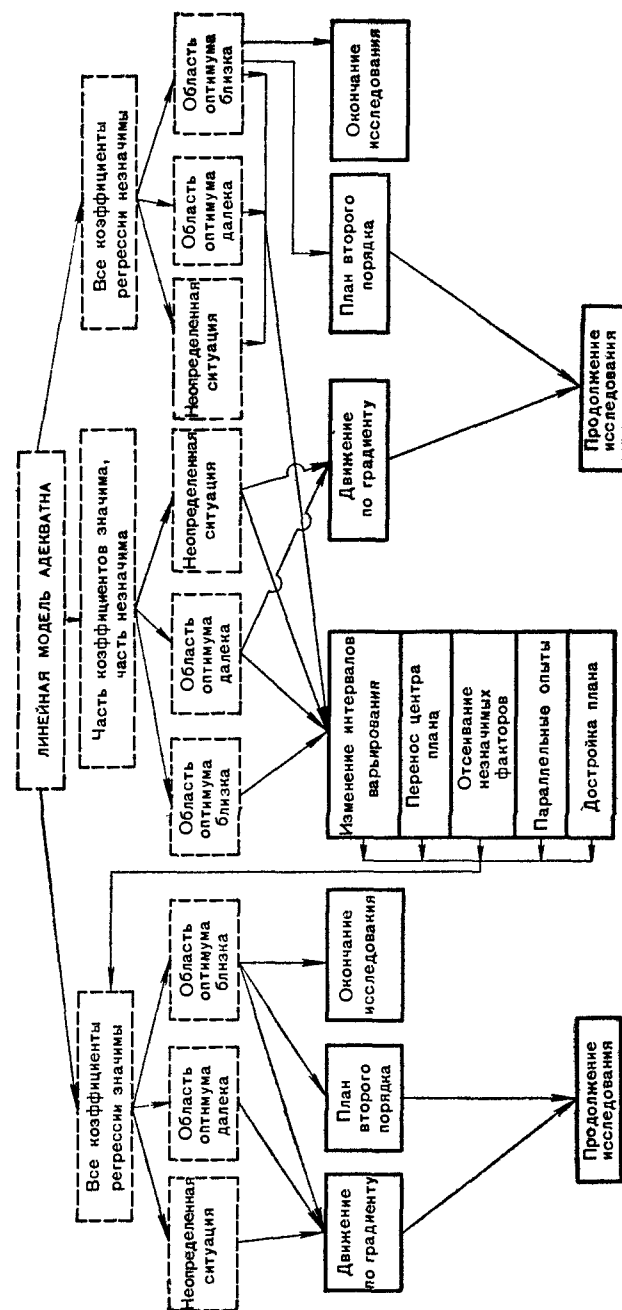


Рис. 28. Принятие решений в задаче определения оптимальных условий; линейная модель адекватна

Параметр оптимизации — прочность, волокна. Условия, матрица планирования и результаты этих дорогостоящих и трудоемких опытов приведены в табл. 11.1 [2].

Таблица 11.1

Матрица планирования и результаты опытов

Уровень	Фактор						
	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	\tilde{x}_5	\tilde{x}_6	
Основной интервал варьирования Верхний Нижний	300	50	2,40	150	0,35	7,2	
	10	7	0,47	5	0,12	0,3	
	310	57	2,87	155	0,47	7,5	
	290	43	1,93	145	0,23	6,9	
Опыты	Кодированные значения факторов						Отклик y
	x_1	x_2	x_3	x_4	x_5	x_6	
1	-1	-1	-1	-1	-1	-1	53,4
2	+1	-1	+1	-1	-1	+1	65,3
3	+1	-1	-1	+1	+1	-1	54,2
4	-1	+1	+1	-1	+1	-1	56,2
5	-1	+1	-1	+1	-1	+1	52,8
6	+1	+1	+1	+1	-1	-1	52,2
7	+1	+1	-1	-1	+1	+1	65,1
8	-1	-1	+1	+1	+1	+1	52,8

Здесь использована 1/8-реплика от полного факторного эксперимента 2^6 с генерирующими соотношениями $x_4 = x_1 x_2 x_3$, $x_5 = -x_1 x_3$, $x_6 = -x_2 x_3$.

Получены следующие оценки коэффициентов регрессии и ошибки в их определении:

$$\begin{aligned} b_0 &= 56,500, & b_3 &= 0,125, & b_6 &= 2,500, \\ b_1 &= 2,700, & b_4 &= -3,500, & s\{b\} &= 1,060. \\ b_2 &= 0,0749, & b_5 &= 0,575. \end{aligned}$$

Линейное уравнение регрессии адекватно. Из шести коэффициентов регрессии три коэффициента (b_1 , b_4 , b_6) значимы. Информации о положении области оптимума нет.

Рассмотрим два варианта принятия решения: 1) движение по градиенту; 2) расширение интервалов варьирования.

Оценим первый вариант. Из шести коэффициентов регрессии только три оказались значимыми, так что движение может быть неэффективным. Далее применена 1/8-реплика от полного факторного эксперимента; смешанность эффектов высока, и не исключено, что оценки коэффициентов регрессии являются суммарными оценками нескольких значимых эффектов. С другой

стороны, устранение незначимости линейных эффектов требует постановки новых опытов, а они длительны и дороги. В крутом восхождении мы рискуем напрасно поставить только 2—3 опыта. Поэтому решение о движении по градиенту кажется нам разумным.

В данном примере было принято решение о движении по градиенту.

Второй вариант — с помощью дополнительных опытов устранить незначимость эффектов. Действительно, только три коэффициента из шести оказались значимыми, эффекты смешаны довольно сильно: движение по градиенту может быть неэффективным. Поэтому решение об изменении интервалов варьирования кажется правильным. Единственное, что не учтено этим решением, — длительность и трудоемкость опытов. Изменение интервалов варьирования требует не менее восьми дополнительных опытов. Это трудно осуществить на практике.

Пример 4. В задаче ионообменного разделения неодима и празеодима (стр. 70) получено адекватное уравнение регрессии: $\hat{y} = 88,0 - 2,0x_1 - 4,5x_2$; $s\{b\} = 0,30$. Все коэффициенты регрессии значимы, область оптимума близка (наилучший опыт серии $y_1 = 95\%$). Цель исследования — получение выхода 99—100%, число опытов лимитировано. Варианты решения: 1) движение по градиенту; 2) окончание исследования; 3) переход к плану второго порядка.

Первый вариант — движение по градиенту. Это наиболее приемлемое решение. Несмотря на близость области оптимума, целесообразно увеличить выход на несколько процентов за счет реализации небольшого (2—3) числа опытов. Этой цели отвечает решение о движении по градиенту, тем более что постановка плана второго порядка потребовала бы проведения еще не менее 5 опытов.

Второй вариант — исследование можно закончить. Закончить или продолжить исследование — решает экспериментатор, исходя из тех задач, которые перед ним стоят. Здесь представлялось важным увеличить выход на несколько процентов по сравнению с лучшим опытом серии ($y_1 = 95\%$).

Третий вариант — следует переходить к планированию второго порядка. По условию задачи важно было увеличить выход за счет двух-трех опытов. Этому отвечает движение по градиенту.

В данной задаче было использовано движение по градиенту, расчет которого приведен в гл. 12.

Остается только упомянуть о задаче построения интерполяционной формулы: цель исследования достигнута, если получена адекватная модель.

Перейдем к следующему разделу — принятие решения в случае неадекватной линейной модели.

Линейная модель неадекватна. Если линейная модель неадекватна, значит не удастся аппроксимировать поверхность отклика плоскостью. Формальные признаки (кроме величины F-критерия), по которым можно установить неадекватность линейной модели, следующие.

1. Значимость хотя бы одного из эффектов взаимодействия.
2. Значимость суммы коэффициентов регрессии при квадратичных членах $\sum \beta_{ii}$. Оценкой этой суммы служит разность между b_0

и значением зависимой переменной в центре плана y_0 . Если разность превосходит ошибку опыта, то гипотеза о незначимости коэффициентов при квадратичных членах не может быть принята. Однако надо учесть, что сумма может быть незначима и при значимых квадратичных эффектах, если они имеют разные знаки.

Для неадекватной модели мы не будем делать различия между случаями значимых и незначимых линейных коэффициентов регрессии, поскольку решения для них обычно совпадают.

Решения, принимаемые для получения адекватной модели: изменение интервалов варьирования факторов, перенос центра плана, достройка плана.

Наиболее распространенный прием — изменение интервалов варьирования. Он, конечно, требует постановки новой серии опытов. Иногда отказываются от построения адекватной модели, чтобы ценой нескольких опытов проверить возможность движения по градиенту. Это решение нельзя считать достаточно корректным. Движению по градиенту обычно предшествует оценка кривизны поверхности отклика (по сумме коэффициентов при квадратичных членах) и сопоставление величин линейных эффектов и эффектов взаимодействия. Если вклад квадратичных членов и эффектов взаимодействия невелик, то решение о движении по градиенту представляется возможным.

Еще одно решение: включение в модель эффектов взаимодействия и движение с помощью неполного полинома второго порядка. Этот прием связан с получением и анализом уравнений второго порядка. Направление градиента будет меняться от точки к точке.

Если область оптимума близка, то, как и в блок-схеме рис. 28, возможны варианты окончания исследования и перехода к построению плана второго порядка.

На рис. 29 приведена блок-схема принятия решений в задаче оптимизации для случая, когда линейная модель неадекватна.

Пример 5. Оптимизировался процесс получения фармацевтического препарата (карбометоксисульфанилгуанидина). В качестве факторов были выбраны: \tilde{x}_1 — отношение растворителя к основному веществу, г/л; \tilde{x}_2 — температура реакционной массы, °С; \tilde{x}_3 — время реакции, мин.

Параметр оптимизации — выход продукта в процентах. Условия, матрица планирования и результаты опытов приведены в табл. 11.2.

Получены следующие результаты:

$$\begin{aligned} b_0 &= 23,15, & b_3 &= 9,47, & b_{23} &= 3,64, & s^2_{\{b\}} &= 0,12, \\ b_1 &= 1,92, & b_{12} &= 0,04, & b_{123} &= -0,87, & s^2_{\{y\}} &= 0,97. \\ b_2 &= 10,35, & b_{13} &= -0,91, \end{aligned}$$

Линейное уравнение регрессии оказалось неадекватным: $F_{\text{экс}} = 32,74$ при табличном значении 4,12. Область оптимума далека.

Варианты решения: 1) постановка новой серии опытов, связанная с изменением интервалов варьирования и переносом центра; 2) движение по градиенту.

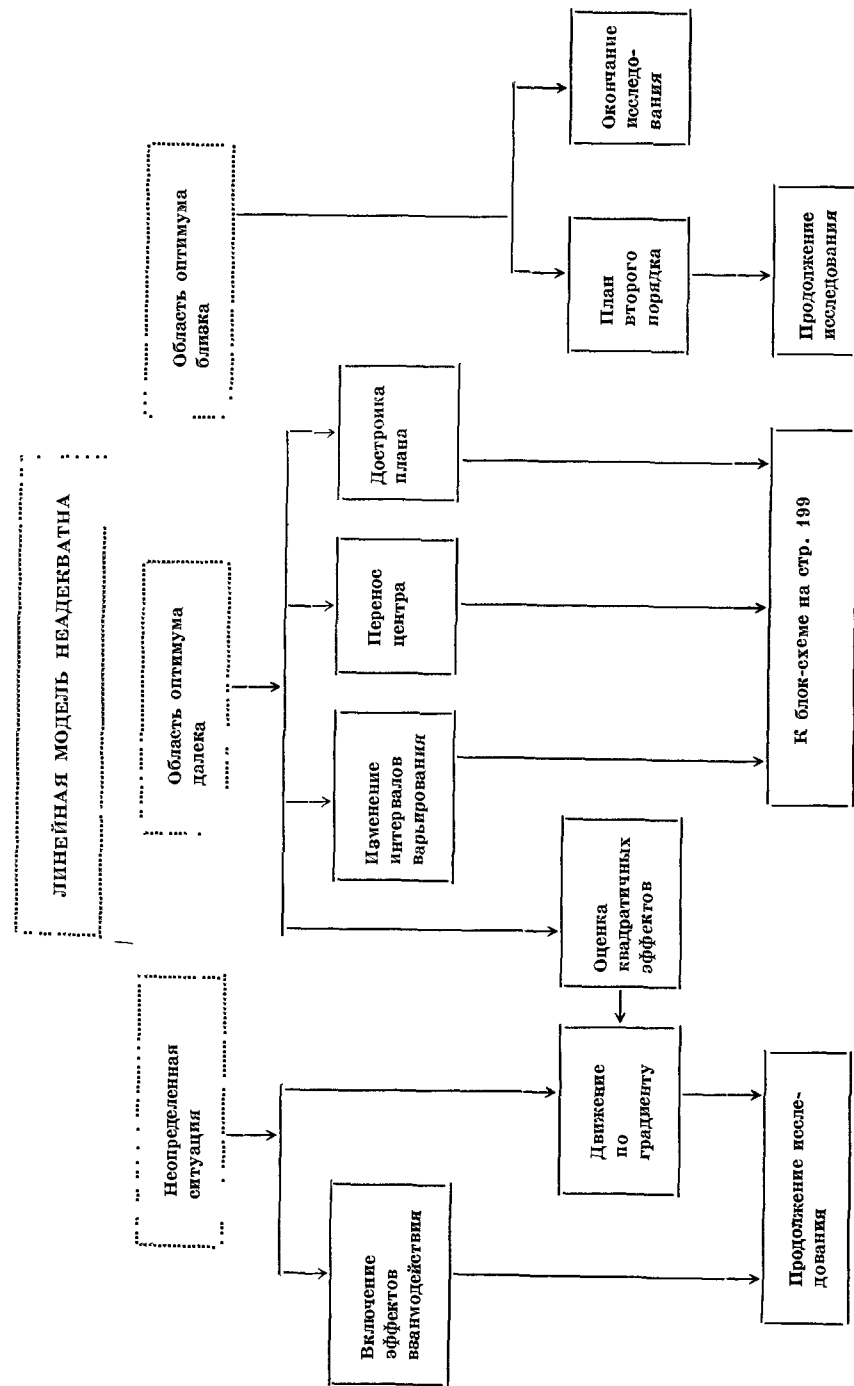


Рис. 29. Принятие решений в задаче определения оптимальных условий; линейная модель неадекватна

Таблица 11.2

Матрица планирования и результаты опытов

Уровень	\bar{x}_1	\bar{x}_2	\bar{x}_3	
Основной интервал варьирования	0,7 0,2	135 5	30 15	
Верхний уровень	0,9	140	45	
Нижний уровень	0,5	130	15	
Опыты	Кодированные значения факторов			
	x_1	x_2	x_3	y
1	+1	+1	+1	46,80
2	+1	-1	+1	20,47
3	-1	-1	+1	16,80
4	-1	-1	-1	5,08
5	+1	+1	-1	24,15
6	+1	-1	-1	8,89
7	-1	+1	-1	16,63
8	-1	+1	+1	46,45

Изменение интервалов варьирования факторов и попытка получения адекватной модели — в данной ситуации вполне приемлемое решение. Интервалы варьирования нужно изменить по факторам x_2 и x_3 . Изменение интервалов можно дополнить перенесением центра эксперимента в условия опытов 1 или 8, давших лучшие результаты.

Таким образом, это решение требует реализации еще восьми опытов.

Проанализируем второе вполне возможное решение. Три эффекта взаимодействия (b_{13} , b_{23} , b_{123}) оказались значимыми, так что постановка новой серии опытов с уменьшением интервалов варьирования представляется разумным решением. Но в то же время линейные эффекты не смешаны с эффектами взаимодействия, и их вклад в уравнение регрессии значительно превышает вклад взаимодействий. Напомним, что опыты дороги. Поэтому решение о проведении двух-трех опытов крутого восхождения более всего в данной ситуации соответствует цели достижения максимального выхода с минимальными затратами, хотя и существует риск не получить улучшения результатов.

При выполнении этой работы исследователи выбрали движение по градиенту и улучшили результаты в два раза (см. гл. 12).

Особый случай возникает при использовании насыщенных планов. При значимости всех коэффициентов регрессии ничего нельзя сказать об адекватности или неадекватности модели. Движение

по градиенту в такой ситуации показывает правильность предположения, что коэффициенты регрессии являются оценками для линейных эффектов.

Остановимся теперь на задаче построения интерполяционной формулы.

11.3. Построение интерполяционной формулы.

Линейная модель неадекватна

Первое, что следует сделать при решении этой задачи, — включить в уравнение эффекты взаимодействия. Конечно, такое решение возможно, если был применен ненасыщенный план. После введения эффектов взаимодействия может не хватать степеней свободы и потребуется реализация еще двух-трех опытов внутри области эксперимента для проверки гипотезы адекватности.

Все остальные способы построения интерполяционной формулы связаны с необходимостью проведения новых опытов. Один из них — достройка плана. Используются все те же приемы, что и при устранении незначимости коэффициентов регрессии (стр. 196): метод «перевала», достройка до полного факторного эксперимента, до дробной реплики, для которой ранее смешанные эффекты становятся «чистыми», достройка до плана второго порядка.

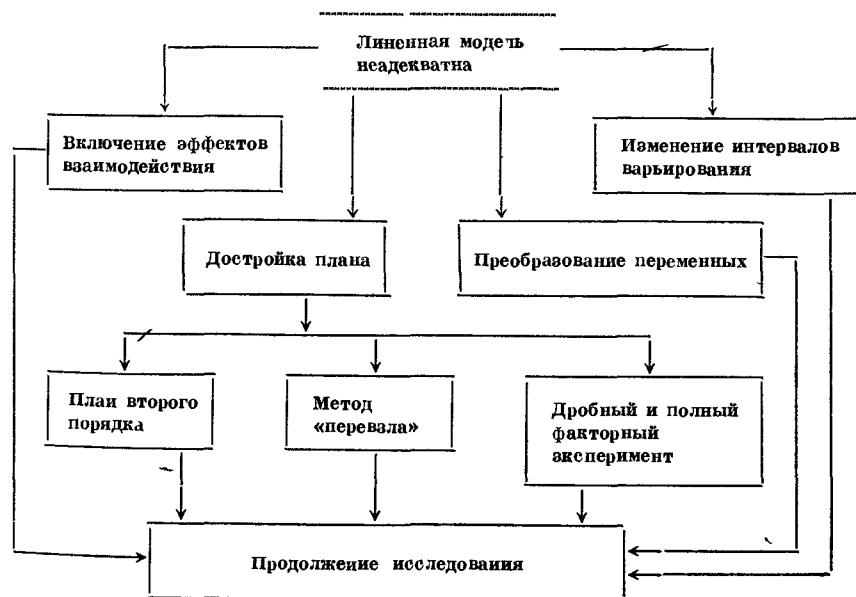


Рис. 30. Принятие решений в задаче построения интерполяционной формулы; линейная модель неадекватна

Еще один, хотя и не очень распространенный прием, — преобразование зависимых и независимых переменных, о котором упоминалось в гл. 2. Однако его подробное рассмотрение выходит за рамки нашей книги.

Наконец, если не удалось все-таки получить адекватную модель, то остается разбить область эксперимента на несколько подобластей и описать отдельно каждую из них. Это требует уменьшения интервалов варьирования факторов.

Приведем блок-схему принятия решений в задаче построения интерполяционной формулы для случая, когда линейная модель неадекватна (рис. 30). Если линейная модель адекватна, то задача решена.

Пример 6. В качестве факторов при построении математической модели яичного экстрактора были выбраны: \tilde{x}_1 — диаметр турбинки, мм; \tilde{x}_2 — скорость вращения турбинки, об/мин; \tilde{x}_3 — температура, °С; \tilde{x}_4 — концентрация свободной кислоты в водном растворе, г/мл; \tilde{x}_5 — высота слоя жидкости в ячейке, мм; \tilde{x}_6 — соотношение фаз в эмульсии.

Параметр оптимизации — продолжительность полного расслаивания в мин. Условия, матрица планирования и результаты опытов приведены в табл. 11.3. Использована 1/4-реплика от полного факторного эксперимента 2^6 . Линейное уравнение регрессии оказалось неадекватным. Затем были введены три несмешанных между собой эффекта взаимодействия факторов, имеющих наибольшую абсолютную величину

$$\hat{y} = 12,16 + 0,53x_1 + 0,53x_2 - 1,38x_3 - 3,22x_4 + 1,44x_5 - 0,62x_6 - 0,84x_1x_4 - 0,50x_1x_6 - 0,78x_2x_4.$$

Это уравнение адекватно описывает процесс $s_{\{y\}} = 0,39$. Рассчитанное значение $F_{\text{эксп}} = 2,4$ при табличном значении $F = 2,7$. Уравнение было использовано при проектировании промышленного аппарата.

Вот один из возможных приемов построения интерполяционной модели.

11.4. Резюме

Перевод модели с абстрактного математического языка на язык экспериментатора мы называли интерпретацией модели. Интерпретация — сложный процесс, который проводится в несколько этапов. Он включает оценку величины и направления влияния отдельных факторов и их взаимодействий, сопоставление влияния совокупности факторов, проверку правильности априорных представлений и в некоторых случаях проверку и выдвижение гипотез о механизме процесса.

Сочетание возможных действий с различными экспериментальными ситуациями приводит к десяткам тысяч возможных решений. Поэтому обсуждаются только «типичные» решения. Ситуации различаются по адекватности и неадекватности модели, значимости и незначимости коэффициентов регрессии, положению оптимума.

Для линейной адекватной модели со значимыми коэффициен-

Таблица 11.3

Матрица планирования и результаты опытов

Уровень	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	\tilde{x}_5	\tilde{x}_6	
Основной	90	600	26	0,40	195	0,8115	
интервал варьирования	10	100	4	0,29	25	0,0975	
Верхний	100	700	30	0,69	220	0,909	
Нижний	80	500	22	0,11	170	0,714	
Опыты	Кодированное значение факторов						
	x_1	x_2	x_3	x_4	x_5	x_6	y
1	−1	+1	+1	+1	−1	−1	7,00
2	−1	−1	−1	−1	+1	−1	16,50
3	−1	−1	−1	+1	−1	−1	9,50
4	−1	+1	+1	+1	+1	+1	9,00
5	+1	+1	+1	+1	+1	+1	7,75
6	+1	−1	−1	+1	+1	+1	10,75
7	−1	+1	−1	+1	+1	+1	11,50
8	+1	−1	−1	−1	−1	+1	13,25
9	+1	+1	−1	+1	−1	−1	8,50
10	−1	+1	+1	−1	+1	−1	14,00
11	−1	−1	+1	−1	−1	+1	9,25
12	+1	−1	+1	−1	+1	−1	17,25
13	+1	+1	+1	−1	−1	+1	14,50
14	+1	+1	−1	−1	+1	−1	22,00
15	−1	+1	−1	−1	−1	+1	16,25
16	+1	−1	+1	+1	−1	−1	7,50

тами регрессии возможны: движение по градиенту, план второго порядка, окончание исследования. Если часть коэффициентов регрессии незначима, то возможен выбор одного из решений, позволяющих получать коэффициенты регрессии значимыми: изменение интервалов варьирования, перенос центра плана, отсеивание незначимых факторов, параллельные опыты, достройка плана. Кроме того, остается движение по градиенту, а если область оптимума близка, то реализация плана второго порядка или окончание исследования.

Наконец, если все коэффициенты незначимы, то выбираются решения по реализации плана второго порядка или окончанию исследования (область оптимума близка) либо решения, позволяющие получать значимые коэффициенты регрессии (область оптимума далека и неопределенная ситуация).

Линейная модель неадекватна. Если область оптимума близка, то исследование либо заканчивается, либо реализуется план второго порядка. Такие решения, как изменение интервалов варьиро-

вания, перенос центра плана, достройка плапа, движение по градиенту, применяются при любом положении оптимума. Возможно включение в модель эффектов взаимодействия факторов и движение с помощью неполного полинома второго порядка, а также оценка квадратичных эффектов для получения информации о кривизне поверхности отклика перед движением по градиенту.

Наконец, если поставлена задача построения интерполяционной формулы, то на получении адекватной модели исследование заканчивается, а в случае неадекватной модели принимается одно из следующих решений: включение в модель эффектов взаимодействия, достройка плана, преобразование переменных, изменение интервалов варьирования.

Л и т е р а т у р а

1. В. В. Налимов, И. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.
2. Н. С. Иванов, Е. Н. Марина, Д. Ф. Фильберт и др. Применение математической статистики при исследовании процесса формирования и вытягивания полипропиленового волокна. В сб. «Карбоцепные волокна». М., Химия, 1966.

Глава двенадцатая

КРУТОЕ ВОСХОЖДЕНИЕ ПО ПОВЕРХНОСТИ ОТКЛИКА

Куда ведешь, тропинка милая?

Из песни

Решения, которые обсуждались в предыдущей главе, направлены на то, чтобы обеспечить эффективное движение по градиенту. Давайте посмотрим, как на практике осуществить это движение.

12.1. Движение по градиенту

Посмотрите на рис. 31. На нем изображены кривые равного выхода поверхности отклика для двух независимых переменных. Они подобны линиям равной высоты на географических картах. Поверхность отклика имеет вид холма с вершиной в точке «0». Если попытаться попасть в окрестность этой точки из точки *A* с помощью одного из вариантов однофакторного эксперимента, то мы сначала должны стабилизировать первый фактор, например x_1 , и изменять в направлении *AC* второй фактор до тех пор, пока увеличивается выход. За точкой *C* выход падает, и поэтому в ней стабилизируем x_2 и изменяем x_1 в направлении *CD* по такому же правилу и т. д.

Не кажется ли вам, что путь к вершине довольно извилист? Он становится еще более трудоемким при возрастании числа независимых переменных. Наиболее короткий путь к вершине — направление градиента функции отклика. На рис. 31 это направление *AB*, перпендикулярное линиям уровня. Градиент непрерывной однопозначной функции φ есть вектор

$$\nabla \varphi = \frac{\partial \varphi}{\partial x_1} i + \frac{\partial \varphi}{\partial x_2} j + \dots + \frac{\partial \varphi}{\partial x_k} k,$$

где $\nabla \varphi$ — обозначение градиента, $\partial \varphi / \partial x_i$ — частная производная функции по *i*-му фактору, *i*, *j*, *k* — единичные векторы в направлении координатных осей.

Следовательно, составляющие градиента суть частные производные функции отклика, оценками которых являются, как мы уже говорили, коэффициенты регрессии.

Изменяя независимые переменные пропорционально величинам коэффициентов регрессии, мы будем двигаться в направлении

градиента функции отклика по самому крутому пути. Поэтому процедура движения к почти стационарной области называется *крутым восхождением*.

Величины составляющих градиента определяются формой поверхности отклика и теми решениями, которые были приняты при выборе параметра оптимизации, нулевой точки и интервалов варьирования. Знак составляющих градиента зависит только от формы поверхности отклика и положения нулевой точки (рис. 32).

Пусть интервал варьирования некоторого значимого фактора равен 10 единицам. Как вы считаете, изменится ли составляющая

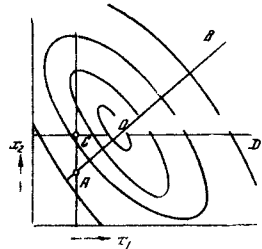


Рис. 31. Движение по поверхности отклика методами однофакторного эксперимента и градиента

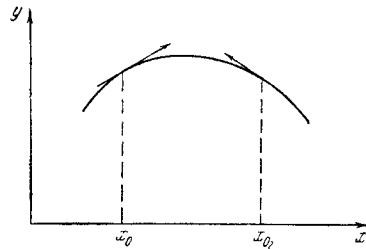


Рис. 32. Зависимость знака градиента от формы поверхности и положения нулевой точки

градиента, если в качестве единицы измерения воспользоваться вначале миллиметром, а затем дюймом? Переход от миллиметров к дюймам эквивалентен значительному увеличению интервала варьирования: 1 дюйм, как известно, равен 25,4 мм.

В первом случае интервал варьирования равнялся 10 мм, а во втором — 254 мм. Такое изменение интервала варьирования не может не иметь последствий для значимого фактора. Сильно увеличится коэффициент регрессии и вместе с ним — составляющая градиента.

В большинстве задач выбор размерности не является проблемой. Этот выбор определяется характером задач, традициями и существующей системой мер и измерительных приборов. Когда размерность фиксирована, то все ясно. Однако важно помнить, что размерность влияет на величины составляющих градиента, а их знаки инвариантны относительно изменения масштабов.

Итак, для данной поверхности отклика выбраны нулевая точка и интервалы варьирования, проведен эксперимент и найдены оценки коэффициентов регрессии. После этого направление градиента задается однозначно и является единственным. При этом предполагается, что имеется только один оптимум.

Теперь займемся расчетом направления градиента.

12.2. Расчет крутого восхождения

Технику расчета крутого восхождения удобно рассмотреть на простейшем примере в случае одного фактора (рис. 33).

Значение коэффициента регрессии равно тангенсу угла между линией регрессии и осью данного фактора. Если его умножить на интервал варьирования, который является прилежащим катетом в прямоугольном треугольнике OAB , то получится противолежащий катет AB , который и дает координаты точки, лежащей на градиенте.

Обобщение на случай k факторов делается механически, так как все эффекты независимы друг от друга. Существенно только соотношение произведений коэффициентов на соответствующие

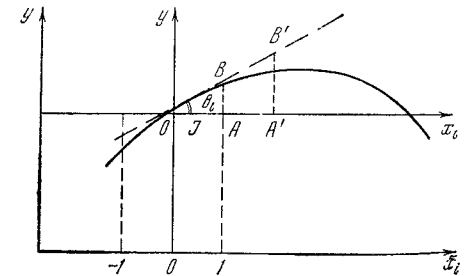


Рис. 33. Расчет координат точек в направлении градиента

интервалы. Их абсолютные величины могут все одновременно умножаться или делиться на любое положительное число. При этом получаются точки, лежащие на том же градиенте, но с другим шагом. Эта процедура заключается в том, чтобы к нулевому уровню последовательно алгебраически прибавлять величины, пропорциональные составляющим градиента.

Сразу возникает вопрос: а как выбрать шаг движения по градиенту? Это еще один этап, для которого не существует формализованного решения. Небольшой шаг потребует значительного числа опытов при движении к оптимуму, большой шаг увеличивает вероятность проскока области оптимума. Во всяком случае, аналогично выбору интервалов варьирования (гл. 6), нижняя граница задается возможностью фиксирования двух соседних опытов, а верхняя — областью определения фактора. Для облегчения работы шаги обычно округляют.

На расчет градиента не оказывает влияние b_0 . Для качественных факторов на двух уровнях либо фиксируется лучший уровень, либо градиент реализуется дважды для каждого уровня в отдельности. Незначимые факторы стабилизируются на любом уровне в интервале ± 1 . Если нет специальных соображений, то выбирают нулевой уровень. Если же по экономическим соображениям, например, выгодно поддерживать нижний уровень, то выбирают его. В движении по градиенту эти факторы не участвуют [1, 2].

Рассмотрим пример расчета крутого восхождения для процесса ионообменного разделения смеси редкоземельных элементов (см. стр. 70, 199).

Пример 1. В табл. 12.1 приведены условия, матрица планирования и результаты серии опытов, а также расчет крутого восхождения.

Этапы расчета крутого восхождения.

1. Расчет составляющих градиента.

$$b_1 \times I_1 = -1, 0, \quad b_2 \times I_2 = -4, 5.$$

Теперь мы должны прибавлять составляющие градиента к основному уровню факторов. Берем условия опыта № 5: $\bar{x}_1=0,5$, $\bar{x}_2=2,5$. В опыте № 6 факторы имеют уже нереальные значения, следовательно, можно сделать вывод, что шаг движения велик.

2. Воспользуемся условием: умножение составляющих градиента на любое положительное число дает точки, также лежащие на градиенте. В данной задаче удобно изменять рН (\bar{x}_2) на 0,5, т. е. уменьшить составляющую градиента в 9 раз. Во столько же раз уменьшается и составляющая градиента по первому фактору (—0,11). Изменению составляющих градиента соответствует в табл. 12.1 строка: шаг при изменении \bar{x}_2 на 0,5. Наконец, методы анализа позволяют задавать значение \bar{x}_1 с точностью до одного знака после запятой, шаг по этому фактору округляется.

3. Последний этап расчета: последовательное прибавление составляющих градиента к основному уровню. Получаем серию опытов крутого восхождения (в табл. 12.1 — опыты 5—9). Эти опыты часто называют мысленными.

А если бы мы признали разумным изменить фактор \bar{x}_1 на —0,20, то правильно ли рассчитана серия опытов крутого восхождения?

Опыты	\bar{x}_1	\bar{x}_2
5	1,3	6,0
6	1,1	5,0
7	0,9	4,0
8	0,7	3,0
9	0,5	2,0

Составляющая градиента уменьшилась по первому фактору в $\frac{1,0}{0,2}=5$ раз. Во столько же раз уменьшилась составляющая градиента по второму фактору ($\frac{-4,5}{5}=-0,9$), или, округляя эту величину, имеем шаг по второму фактору —1,0. Опыты 5—9 (см. выше) как раз и получены прибавлением $\bar{x}_1=-0,2$ и $\bar{x}_2=-1,0$ к основным уровням факторов.

Таким образом, расчет сводится к тому, чтобы выбрать шаг движения по одному из факторов и пропорционально произведениям коэффициентов регрессии на интервалы варьирования рассчитать шаги по другим факторам.

Иногда имеет смысл оценить ожидаемые значения параметра оптимизации в мысленных опытах. Проведем расчет для опыта № 7 крутого восхождения (табл. 12.1). Мы собираемся для оценки параметра оптимизации использовать уравнение регрессии: $\hat{y}=88,0-2,0 x_1-4,5 x_2$. Однако в табл. 12.1 приведены натураль-

Таблица 12.1

Матрица планирования, результаты и расчет крутого восхождения

Уровень	\bar{x}_1	\bar{x}_2	
Основной	1,5	7,0	
Интервал варьирования	0,5	1,0	
Верхний	2,0	8,0	
Нижний	1,0	6,0	
Опыты	Кодированные значения факторов		
	x_1	x_2	y^*
1	—1	—1	95,0
2	+1	—1	90,0
3	—1	+1	85,0
4	+1	+1	82,0
b_j	—2,0	—4,5	
$b_j \times I_j$	—1,0	—4,5	
Шаг при изменении \bar{x}_2 на 0,5	—0,11	—0,5	
Округление	—0,1	—0,5	
Опыты			
5	1,4	6,5	
6	1,3	6,0	
7	1,2	5,5	
8	1,1	5,0	
9	1,0	4,5	

* Среднее значение из двух параллельных опытов

ные значения факторов, а в уравнении применяются кодированные значения. Поэтому необходимо перевести натуральные значения в кодированные. Кодированные величины получаются с помощью уже известной вам формулы (стр. 72): $x_j=(\bar{x}_j-\bar{x}_{j0})/I_j$, $x_1=-6,0$, $x_2=-1,5$. Подставляя эти значения в уравнение регрессии, получим $\hat{y}_7=95,95$ (\hat{y} —так обозначают предсказанную по уравнению регрессии величину зависимой переменной). Аналогично для опыта № 8 $x_1=-0,8$; $x_2=-2,0$ и $\hat{y}_8=98,6$. Экспериментально полученные значения могут не совпадать с расчетными: величины независимых переменных выходят за область эксперимента [3].

Рассмотрим еще один пример по расчету крутого восхождения для процесса экстракции гафния трибутилфосфатом.

Таблица 12.2

Матрица планирования, результаты опытов и крутое восхождение

Уровень	Факторы				
	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	
Основной	3,0	30	1,5 : 1	15	
Интервал варьирования	2,0	10	1,0	10	
Верхний	5,0	40	2,5 : 1	25	
Нижний	1,0	20	0,5 : 1	5	
Опыты	Кодированные значения факторов				Отклик $y \cdot 10^{2*}$
	x_1	x_2	x_3	x_4	
1	-1	-1	-1	-1	2,060
2	+1	-1	+1	-1	3,400
3	-1	-1	+1	+1	3,095
4	-1	+1	-1	+1	4,995
5	+1	+1	-1	-1	15,150
6	+1	-1	-1	+1	9,225
7	-1	+1	+1	-1	8,250
8	+1	+1	+1	+1	5,995
b_j	0,0195	0,02030	-0,0137	-0,0066	
$b_j \times I_j$	0,0390	0,203	-0,0137	-0,066	
Шаг при изменении \bar{x}_2 на 5	0,9605	5	-0,3399	-1,6256	
Округление	1,0	5,0	-0,3	-2,0	
9	4,0	35,0	1,2 : 1	13	
10	5,0	40,0	0,9 : 1	11	
11	6,0	45,0	0,6 : 1	9	
12	7,0	50,0	0,3 : 1	7	
13	8,0	55,0	0 : 1	5	
14	9,0	60,0	—	3	

* Среднее значение из двух параллельных опытов, рандомизированных во времени

Пример 2. В табл. 12.2 так же, как и в табл. 12.1, приведена стандартная форма представления условий, матрицы планирования и результатов опытов вместе с крутым восхождением [4].

Остались нерассмотренными два момента: как влияют на крутое восхождение соотношения численных значений коэффициентов регрессии и почему движение по градиенту начинается из нулевой точки.

Представим себе, что в адекватном линейном уравнении значим только один коэффициент. Тогда в движении по градиенту будет участвовать только один фактор. Многофакторная задача вырождается в однофакторную. А это менее эффективно. Рассмотренный случай является крайним, но в практике довольно часто b -коэффи-

циенты существенно различаются между собой, оставаясь значимыми.

Функция, величины коэффициентов которой различаются не существенно, называется симметричной относительно коэффициентов. Движение по градиенту для симметричной функции наиболее эффективно. Удачным выбором интервалов варьирования можно сделать симметричной любую линейную функцию для значимых факторов.

На первом этапе планирования не всегда удастся получить симметричную функцию. Если функция резко асимметрична (коэффициенты различаются на порядок), то выгоднее вновь поставить

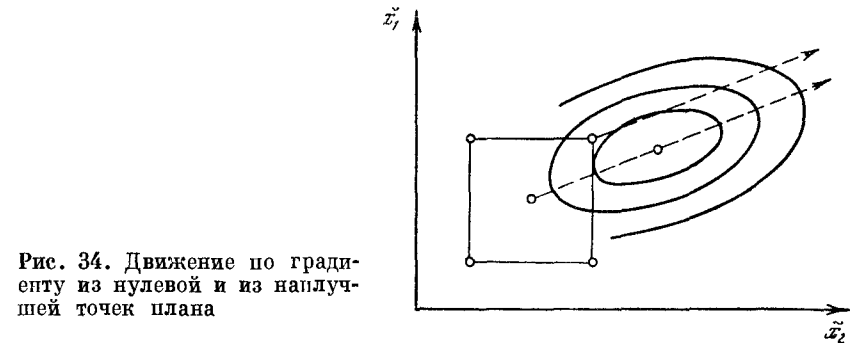


Рис. 34. Движение по градиенту из нулевой и из наилучшей точек плана

эксперимент, изменив интервалы варьирования, а не двигаться по градиенту. Так, в работе [4] после первой серии опытов получено следующее уравнение регрессии:

$$\hat{y} \cdot 10^3 = 22,48 - 1,01x_1 + 19,46x_2 + 1,46x_3 - 1,05x_4.$$

Здесь b_2 на порядок превышает остальные коэффициенты, которые статистически незначимы. Это связано скорее с неудачным выбором интервалов варьирования, чем с отсутствием соответствующих эффектов. Движение по градиенту нецелесообразно. Решение увеличить вдвое интервал варьирования незначимых факторов привело к такому результату:

$$\hat{y} \cdot 10^3 = 65,5 + 19,5x_1 + 20,3x_2 - 13,7x_3 - 6,6x_4.$$

На этот раз функция оказалась симметричной, и было реализовано крутое восхождение, показанное на стр. 212.

Направление градиента определяется единственным способом, и движение должно начинаться из нулевой точки. На рис. 34 приведена простая геометрическая иллюстрация этого факта. Хорошо видно, что движение из наилучшей точки плана проходит в стороне от оптимальных условий.

Можно рассуждать иначе. Функция отклика, вид которой нам неизвестен, разлагалась в ряд Тейлора в окрестности нулевой точки. Именно к этой точке и относится оценка градиента.

Вы узнали, как производится расчет градиента. Займемся теперь практической реализацией опытов в направлении градиента.

12.3. Реализация мысленных опытов

Рассчитав составляющие градиента, мы получили условия мысленных опытов. Число мысленных опытов зависит от задачи. Ограничением сверху служит граница области определения хотя бы по одному из факторов. Иногда по технологическим соображениям нет смысла определять условия многих опытов. Обычно рассчитывается 5–10 мысленных опытов.

Как реализовать мысленные опыты? Нужно ли ставить все опыты подряд или только некоторые из них? С какого опыта начинать? Если модель адекватна, то начинают реализацию с тех опытов, условия которых выходят за область эксперимента хотя бы по одному из факторов. Для неадекватной модели часто один-два опыта выполняют в области эксперимента. Для мысленных опытов процесса ионообменного разделения (табл. 12.1) имеет смысл ставить эксперимент начиная с опыта № 7 ($\bar{x}_1=1,2$, $\bar{x}_2=5,5$). Этот опыт по фактору x_2 уже выходит за пределы области эксперимента.

Если начинать реализацию мысленных опытов с опыта № 5, условия которого $\bar{x}_1=1,4$, $\bar{x}_2=6,5$, то значения факторов для этого опыта находятся внутри области эксперимента (основной уровень $\bar{x}_1=1,5$; $\bar{x}_2=7,0$; $I_1=0,5$; $I_2=1,0$). Так как модель адекватна, то значение параметра оптимизации для этого опыта можно получить просто расчетом $\hat{y}_1=90,65$.

Условия мысленных опытов следует тщательно обдумать и убедиться, что нет затруднений в их реализации. Если что-то не ладится, можно изменить шаг и рассчитать мысленные опыты заново.

Существует две принципиально различные стратегии реализации мысленных опытов. Все намеченные к реализации опыты ставятся одновременно либо последовательно по некоторой программе. Одновременно могут ставиться все мысленные опыты, через один, через два и т. д. Последовательный принцип заключается в том, что вначале ставятся два-три опыта, анализируются результаты и принимается решение о постановке новых опытов. Выбор стратегий определяется стоимостью опытов, их длительностью и условиями экспериментирования.

Представьте себе задачу, в которой опыт длится несколько месяцев, но одновременно можно поставить довольно большое число опытов. При последовательной стратегии реализация мысленных опытов надолго затягивается. Выгоднее реализовать сразу все намеченные опыты. Это характерно для сельскохозяйственных, биологических, металлургических задач и т. д.

Преимущество одновременной реализации опытов в том, что эта стратегия исключает временной дрейф.

Когда опыты быстры и дешевы, эта стратегия вполне пригодна. А если опыты дороги, приходится пользоваться последо-

вательной стратегией, так как минимизация числа опытов приобретает большую актуальность.

Имеется несколько вариантов последовательной стратегии. Можно реализовать опыты по одному и после каждого анализировать результаты. Другой путь — ставятся одновременно два-три опыта и затем принимаются решения.

При незначительном изменении параметра оптимизации (поверхность пологая) следующим реализуется далеко отстоящий опыт, при сильном изменении (поверхность крутая) — близлежащий.

Иногда пользуются методом «ножниц»: реализуются два крайних мысленных опыта, а затем прощупывается пространство внутри этого интервала. Минимальное число опытов — три, так как оптимум необходимо захватить «в вилку». Два опыта могут оказаться достаточными, когда координаты оптимума близки к координатам опытов исходного плана или же когда попытка продвинуться по неадекватной модели оказывается неудачной.

Крутое восхождение может считаться эффективным, если хотя бы один из реализованных опытов даст лучший результат по сравнению с наилучшим опытом серии. Например, наилучший опыт № 5 в серии (стр. 212) $y_5=15,1$, а реализованный опыт № 10 $y_{10}=30,4$, следовательно, крутое восхождение эффективно. Затем из всех реализованных опытов выбирается тот, который дал лучший результат, и его условия принимаются за основной уровень факторов в следующей серии опытов. Если в одном из реализованных опытов достигнуты оптимальные условия, то эксперимент заканчивается.

В процессе получения волокна из полипропилена (стр. 196) крутое восхождение привело к следующим результатам: $y_9=60,1$; $y_{10}=59,7$; $y_{11}=57,8$; $y_{12}=55,3$. Средний выход в серии опытов $b_0=56,5$, максимальный выход — $y=65,3$.

При последовательном поиске можно было бы сократить число опытов. Действительно, уже первый опыт не дал улучшения результатов. Однако возражений против проведения второго или даже третьего опытов нет. Четвертый опыт, видимо, уже не нужен.

Таким образом, вместо четырех опытов достаточно было провести два-три.

Рассмотрим два примера движения по градиенту.

Пример 3. Оптимизация процесса получения карбометоксисульфанилгуанидина (табл. 12.3).

Крутое восхождение оказалось весьма эффективным: средний выход в предыдущей серии опытов $b_0=23,28$, максимальный выход $y=46,80$ (стр. 202) [5].

Пример 4. Оптимизация процесса получения производного пиперазина (табл. 12.4).

Таблица 12.3

Крутое восхождение при оптимизации процесса получения карбометоксисульфанилуанидина

Условия движения по градиенту	Факторы			Отклик y
	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	
$b_j \times I_j$	0,356	51,51	140,40	
Шаг при изменении \tilde{x}_2 на 5	0,0368	5	14,30	
Округление	0,03	5	14,00	
Опыты				
9	0,73	140	44	—
10	0,76	145	58	—
11	0,79	150	72	66,70
12	0,82	155	86	—
13	0,85	160	100	72,50
14	0,88	165	114	68,40

Таблица 12.4

Крутое восхождение

Условия движения по градиенту	Факторы					
	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	\tilde{x}_5	$y, \%$
$I_j \times b_j$	$2,025 \cdot 0,25 = -0,506$	$5,05 \cdot 0,25 = 1,26$		$-2,10 \cdot 5 = -10,5$		
Шаг при $\tilde{x}_4 = -2$	-0,0963	0,24		-2		
Округление	0,09	0,25	—	-2		
Нулевой уровень	1,25	1,25	4,00	25	40	55,0
Опыты						
9	1,16	1,50	4,00	23	40	—
10	1,07	1,75	4,00	21	40	—
11	0,98	2,00	4,00	19	40	68,8
12	0,89	2,25	4,00	17	40	74,3
13	0,80	2,50	4,00	15	40	66,8

Это крутое восхождение оказалось также очень эффективным. В опыте № 12 выход целевого продукта равен 74,3%, что на 20% выше выхода в нулевой точке. К сожалению, крутое восхождение бывает столь эффективным не всегда [6].

Эти примеры иллюстрируют захват оптимума «в вилку».

Остановимся на некоторых особенностях реализации опытов крутого восхождения.

Рассмотрим следующую ситуацию. При эффективном крутом восхождении достигается граница области определения одного

из факторов. По этому фактору дальше двигаться нельзя. Возможны два решения: зафиксировать значение этого фактора и дальше двигаться по остальным или остановиться и поставить новую серию опытов линейного приближения. На практике чаще предпочитают первое решение. В этом случае нужно продолжить расчет мысленных опытов и выбрать стратегию их реализации.

Особого рассмотрения заслуживает постановка повторных опытов. Чаще всего повторные опыты не ставятся, а дублируется только наилучший результат. Будет, конечно, не хуже, если ставить параллельные опыты во всех точках.

Иногда приходится считаться с возможностью временного дрейфа. Ведь между исходной серией опытов и движением по гра-

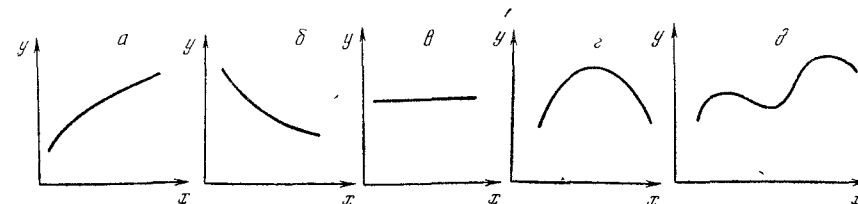


Рис. 35. Ситуации при движении по градиенту

диенту может пройти значительное время. Здесь можно рекомендовать систематическое повторение нулевых точек исходного плана, рандомизированных с точками крутого восхождения. Это дает возможность проверить гипотезу о наличии дрейфа.

При движении по градиенту возможны пять ситуаций, схематично представленных на рис. 35.

Наиболее благоприятные случаи даны на рис. 35, а, г и д, где движение по градиенту оказалось эффективным. В случае а параметр оптимизации все время возрастает, в случае г он проходит через максимум, что было проиллюстрировано примерами 3 и 4. Более сложным является случай д, где нарушена предпосылка одноэкстремальности.

Случай б иллюстрирует неэффективное крутое восхождение. Вместо ожидаемого увеличения параметра оптимизации наблюдается уменьшение. Здесь либо план эксперимента расположен в области оптимума, либо есть грубые ошибки.

Наконец, в случае в все опыты на градиенте дают одно и то же значение. Поверхность отклика имеет вид постоянного гребня.

В соответствии с шаговым принципом «ползания» по поверхности отклика крутое восхождение может осуществляться многократно, пока не будет достигнута почти стационарная область.

Принятие решений после крутого восхождения, которое мы рассмотрим в следующей главе, зависит от рассмотренных выше ситуаций.

Вы познакомились с крутым восхождением по поверхности отклика. Крутое восхождение — это движение в направлении градиента функции отклика. Градиент задается частными производными, а частные производные функции отклика оцениваются коэффициентами регрессии. В крутом восхождении независимые переменные изменяют пропорционально величинам коэффициентов регрессии и с учетом их знаков. Составляющие градиента однозначно получаются умножением коэффициентов регрессии на интервалы варьирования по каждому фактору. Серия опытов в направлении градиента рассчитывается последовательным прибавлением к основному уровню факторов величин, пропорциональных составляющим градиента.

Реализацию мысленных опытов для адекватной модели начинают с опыта, условия которого выходят за область эксперимента хотя бы по одному из факторов. Для неадекватной модели один-два опыта выполняют в области эксперимента. Возможно проведение сразу всех мысленных опытов. Более экономная процедура состоит в проведении двух-трех опытов, оценке результатов и принятии решений о прекращении или дальнейшем проведении экспериментов (последовательный поиск). При движении по градиенту возникают различные ситуации, определяющие принятие дальнейших решений.

Литература

1. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.
2. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
3. Н. М. Пруткова, Ю. В. Грановский, Л. И. Мартыненко и др. Применение статистического метода Бокса—Уилсона для нахождения оптимальных условий разделения р. з. э. при элюировании растворами иминодиуксусной кислоты. — Ж. неорганич. хим., 1970, 15, № 2.
4. Л. Н. Комиссарова, Ю. В. Грановский, Н. М. Пруткова и др. Определение оптимальных условий экстракции микроколичеств гафния трибутилфосфатом. — Заводская лаборатория, 1963, 29, № 1.
5. Е. В. Маркова. Планирование эксперимента при оптимизации процессов тонкого органического синтеза. Автореферат канд. дисс. МХТИ им. Менделеева, М., 1965.
6. И. П. Рутман, Я. С. Карпман, Е. В. Маркова и др. Оптимизация процесса получения одного производного пиперазина. — В сб. трудов ОКБА, № 3, 1969.

ПРИНЯТИЕ РЕШЕНИЙ ПОСЛЕ КРУТОГО ВОСХОЖДЕНИЯ

Свобода воли означает... не что иное, как способность принимать решения со знанием дела.

Ф. Энгельс. Анти-Дюринг

После завершения крутого восхождения вас ожидают довольно разнообразные ситуации, требующие принятия решений о дальнейших действиях.

Ситуации различаются по признаку: оказалось крутое восхождение эффективным или нет. Положение оптимума (близко, далеко, неопределенно) также имеет значение в принятии решений. В некоторых случаях нужно учитывать адекватность (или неадекватность) линейной модели [1—3].

13.1. Крутое восхождение эффективно

Об эффективности движения по градиенту можно судить по величине параметра оптимизации. Движение по градиенту считается эффективным, если реализация мысленных опытов, рассчитанных на стадии крутого восхождения, приводит к улучшению значения параметра оптимизации по сравнению с самым хорошим результатом в матрице.

При эффективном крутом восхождении возможны два исхода: область оптимума достигнута или область оптимума не достигнута.

Область оптимума достигнута. Этот случай является самым легким в смысле принятия решений. Экспериментатор может окончить исследование, если задача заключалась в достижении области оптимума, или продолжить исследование, если задача заключалась не только в достижении области оптимума, но и в детальном ее изучении. При этом необходимо достроить линейный план до плана второго порядка и результаты эксперимента представить в виде полинома второй степени. Перечисленные два варианта принятия решений следуют из концепции Бокса—Уилсона, согласно которой задача оптимизации условно разбивается на два этапа. Первый этап — крутое восхождение с целью скорейшего достижения области оптимума. При этом используется линейное планирование. Линейный план может использоваться один или несколько раз в зависимости от интенсивности

продвижения. Второй этап — описание области оптимума методами нелинейного планирования. При эффективном крутом восхождении весьма часто удается быстро приблизиться к области оптимума (совершить крутое восхождение один раз). Исследователь попадает в область оптимума, которая не может быть описана линейным приближением, и движение по методу крутого восхождения заканчивается. Завершается первый этап оптимизации. Метод крутого восхождения не решает вопроса о самой лучшей точке поверхности отклика, об экстремуме. Чтобы изучить область оптимума, необходимо перейти ко второй стадии планирования — к исследованию почти стационарной области. В принятии решений мы должны рассматривать и этот вариант, хотя изложение нелинейного планирования выходит за рамки нашей книги.

Область оптимума не достигнута. В этом случае ставится линейный план следующего цикла и исследование продолжается.

В предыдущей главе на стр. 216 приведено крутое восхождение (первый цикл) для процесса получения карбометоксисульфанилгуанидина. Оно привело к увеличению выхода реакции до 72,5%. Предполагается, что при удачном подборе условий реакции выход может быть повышен до 95—97%.

Возможные варианты решений: 1) окончить исследование; 2) построить план второго порядка для исследования области оптимума; 3) построить линейный план следующего цикла крутого восхождения.

Первые два варианта кажутся нецелесообразными, выход реакции может быть повышен, при удачном подборе условий реакции, до 95—97%. На стадии крутого восхождения получено только 72,5%. Можно предположить, что имеется резерв более чем в 20%. Поставленная цель не достигнута. Область оптимума достаточно далека.

Решение построить линейный план следующего цикла представляется наиболее целесообразным. Нужно попытаться второй раз совершить крутое восхождение и приблизиться к области оптимума. Такое решение и было принято экспериментатором.

При построении линейного плана второго цикла прежде всего возникает вопрос о выборе центра эксперимента. Самая простая рекомендация — расположить центр нового плана в той части факторного пространства, которая соответствует условиям наилучшего опыта при крутом восхождении (см. гл. 6).

Пример 1. Оптимизируем процесс получения карбометоксисульфанилгуанидина (матрица планирования первого цикла приведена на стр. 202, крутое восхождение — на стр. 216).

Крутое восхождение первого цикла привело к увеличению выхода реакции до 72,5%. Увеличение выхода явилось следствием повышения температуры и возрастания времени реакции. Это учитывалось при выборе локальной области факторного пространства во втором цикле планирования.

Что же касается первого фактора (отношения количества растворителя к количеству основного вещества), то этот фактор изменялся на стадии крутого восхождения наиболее медленно. И действительно, коэффициент b_1 значительно меньше b_2 и b_3 ($b_1=1,78$; $b_2=10,28$; $b_3=9,36$).

С технологической точки зрения увеличение x_1 нежелательно. Поэтому при выборе условий второй серии опытов значение x_1 было уменьшено. Уровни факторов и интервалы варьирования второй серии опытов, а также матрица планирования и результаты эксперимента приведены в табл. 13.1.

Уравнение регрессии получается в виде полинома первой степени

$$\hat{y}=75,901-3,524x_1+3,251x_2+2,436x_3.$$

Таблица 13.1

Вторая серия опытов на стадии крутого восхождения

Уровень	Факторы			
	\bar{x}_1	\bar{x}_2	\bar{x}_3	
Основной	0,5	160	55	
Интервал варьирования	0,2	6	15	
Верхний	0,7	166	70	
Нижний	0,3	154	40	

Опыты	Кодированные значения факторов			
	x_1	x_2	x_3	y
1	+	+	+	76,00
2	+	—	+	74,05
3	—	—	+	80,90
4	—	—	—	73,00
5	+	+	—	76,81
6	+	—	—	62,65
7	—	+	—	81,40
8	—	+	+	82,40

Таблица 13.2

Расчет дисперсии линейной модели

Номер опыта	y	\hat{y}	$ y - \hat{y} $	$(y - \hat{y})^2$	Номер опыта	y	\hat{y}	$ y - \hat{y} $	$(y - \hat{y})^2$
1	76,00	78,06	2,06	4,24	5	76,81	73,19	3,62	13,10
2	74,05	71,56	2,49	6,25	6	62,65	66,69	4,04	16,32
3	80,90	78,61	2,29	5,29	7	81,40	80,24	1,16	1,34
4	73,00	73,74	0,74	0,55	8	82,40	85,11	2,71	7,34

Расчет дисперсии приведен в табл. 13.2.

$$s_{\text{дх}}^2 = 54,35/4 = 13,59; \quad F_{\text{эксп}} = 13,59/0,97 = 14,01; \quad F_{\text{табл}} = 3,8.$$

Линейное приближение неадекватно. Ниже приведены два возможных решения: 1) движение по градиенту; 2) переход к нелинейному планированию.

При движении по градиенту (линейное приближение неадекватно в области оптимума близка) вероятность успеха мала. Однако, если опыт очень дороги (а нелинейное планирование требует опытов значительно больше, чем линейное), можно рассчитать опыты крутого восхождения и некоторые из них реализовать. Это требует немного экспериментальных усилий.

Кроме того, реализация опытов по крутому восхождению может заинтересовать экспериментатора по технологическим соображениям. Обратите внимание на то, что знак b_1 отрицательный. Значит соотношение между растворителем и основным веществом будет уменьшаться. Растворителем является этиленгликоль. Чем меньше его будет в реакционной массе, тем лучше.

В других случаях следует переходить к нелинейному планированию.

Таблица 13.3

Крутое восхождение (вторая серия)

Условия движения по градиенту	Факторы			y
	\bar{x}_1	\bar{x}_2	\bar{x}_3	
$b_j \times I_j$	$-3,524 \times 0,2 = -0,705$	$3,251 \times 6 = 19,506$	$2,436 \times 15 = 36,540$	
Шаг при изменении \bar{x}_1 на 0,1	-0,1	2,76	5,17	
Округление	-0,1	3	5	
Опыты				
9 (нулевой уровень)	0,5	160	55	77,3
10	0,4	163	60	82,2
11	0,3	166	65	—
12	0,2	169	70	83,6
13	0,1	172	75	84,8
14	0	175	80	86,9

Результаты крутого восхождения приведены в табл. 13.3. Выход реакции повышен примерно до 85%. По сравнению с выходом в центре эксперимента это — увеличение на 7,5%, а по сравнению с лучшим опытом в матрице — всего лишь на 2,4% ($s_y^2 = 0,97$). Тем не менее крутое восхождение оказалось весьма примечательным. В опыте № 13 максимальный выход получен при значении $\bar{x}_1 = 0,1$. До планирования эксперимента считалось, что процесс получения карбометоксиульфанилгуанидина может успешно протекать при $\bar{x}_1 > 0,7$, а выход реакции более 70% был неизвестен.

В данном случае крутое восхождение в некорректном применении (в условиях нелинейности и близости к области оптимума), привело в область факторного пространства, где при $\bar{x}_1 = 0,1$ получен довольно высокий выход реакции.

Неопределенная ситуация. Когда y не имеет ограничения и экспериментатор не может определить степень близости оптимума, возможны два решения: построение линейного плана следующего цикла или, если достигнут требуемый результат, окончание работы.

Общая картина принятия решений для случая, когда крутое восхождение оказалось эффективным, показана на рис. 36.

13.2. Крутое восхождение неэффективно

В каждом положении отыщется что-нибудь утешительное, если хорошо поискать.

Д. Дефо

Принимать решения при неэффективном движении по градиенту гораздо сложнее. Принятие решений во многом зависит от определенности ситуации (далеко от оптимума, близко, неопределенно) и от адекватности линейной модели. Наиболее типичные случаи показаны на блок-схеме рис. 37.

Рассмотрим каждую ситуацию отдельно.

Область оптимума близка. Если при реализации матрицы планирования удалось получить достаточно высокие значения параметра оптимизации и при крутом восхождении улучшить их не удалось, то наиболее типичными являются решения: 1) окончание исследования (выбирается лучший опыт); 2) построение плана второго порядка для описания области оптимума.

Если линейная модель была неадекватна, то возможно третье решение — возврат к блок-схеме стр. 201 для выяснения причины неадекватности линейной модели.

Пример 2. Имеется следующая ситуация: исходный план — полуреплика, линейная модель неадекватна, крутое восхождение оказалось неэффективным, область оптимума близка. Параметром оптимизации является выход полезного продукта. Максимально возможный выход — 100%. При реализации полуреплики получен наибольший выход — 80%. Ошибка опыта — 1%.

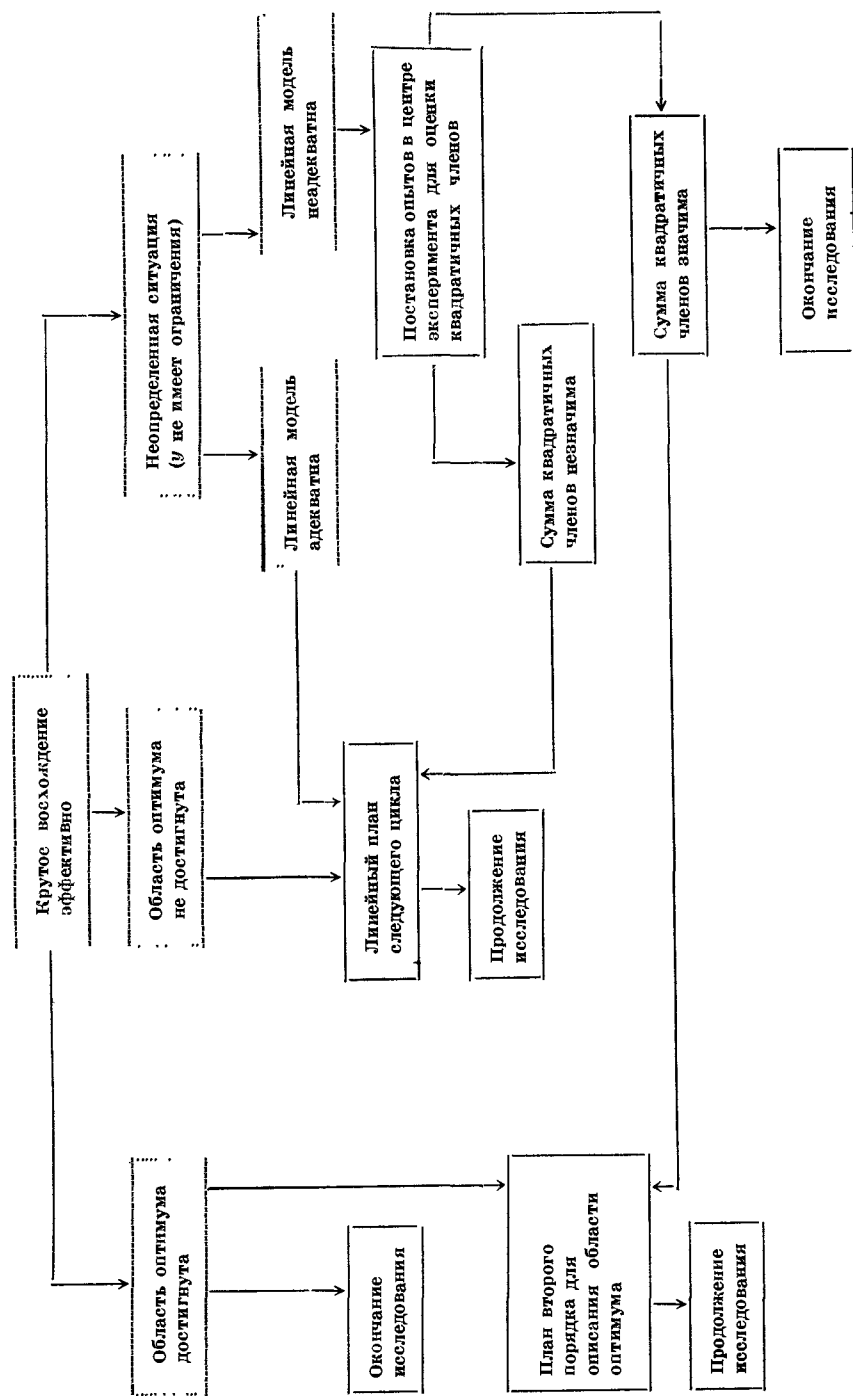
Какому из трех решений можно отдать предпочтение?

Предлагается три варианта: 1) окончить исследование; 2) перейти к нелинейному планированию второго порядка; 3) построить полуреплику до полного факторного эксперимента.

Первое решение — окончить исследование. Давайте проанализируем ситуацию. Разница в 20% между максимальным и наилучшим выходом весьма ощутима. Видимо, целесообразно продолжить исследование и постараться улучшить значение параметра оптимизации.

Окончить исследование можно в том случае, если бы ставилась цель только приблизиться к области высокого выхода.

Второе решение — построить линейный план до плана второго порядка. Это одно из возможных решений. Если бы исходным планом был полный факторный эксперимент, то такое решение было бы наиболее целесообраз-



36. Принятие решений после крутого восхождения; крутое восхождение эффективно

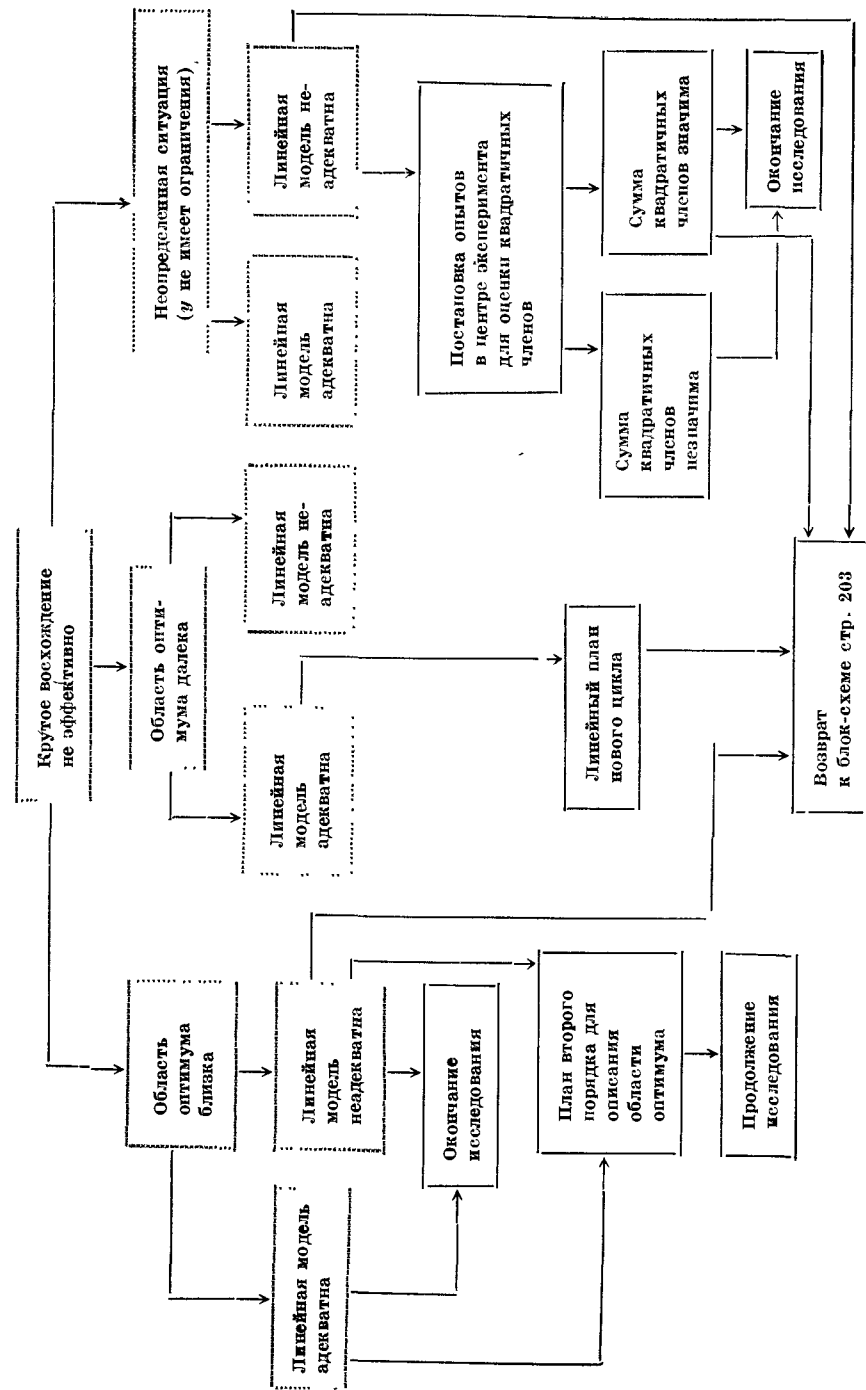


Рис. 37. Принятие решений после крутого восхождения; крутое восхождение неэффективно

ным. Но вы имели дело с дробным факторным экспериментом. В этом случае линейные оценки смешаны с эффектами взаимодействий. Поэтому имеет смысл подумать также и о другом решении.

Третье решение — достроить полуреплику до полного факторного эксперимента. Это решение представляется разумным. Наряду с этим можно также рассматривать переход к нелинейному планированию.

Область оптимума далека. Линейная модель адекватна. Если область оптимума далека и линейная модель адекватна, казалось бы, имеются все предпосылки, чтобы крутое восхождение оказалось эффективным. Тем не менее на практике крутое восхождение нередко оказывается неэффективным. Возможное объяснение — в характере поверхности отклика. Мы исходим из предпосылки, что поверхность

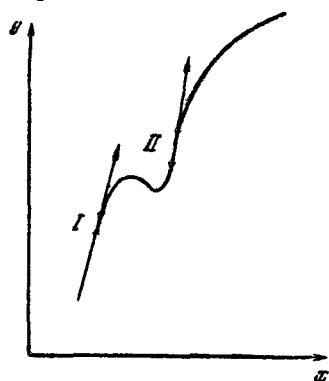


Рис. 38. Пример крутого восхождения; область оптимума далека, линейная модель адекватна, крутое восхождение неэффективно

I — исследованная область факторного пространства в первом цикле крутого восхождения, II — исследованная область факторного пространства во втором цикле крутого восхождения

отклика гладкая и одноэкстремальная. В действительности она может иметь, например, вид, показанный на рис. 38. В таких случаях целесообразно передвинуться в другую область факторного пространства и построить линейный план второго цикла крутого восхождения.

Область оптимума далека. Линейная модель неадекватна. Здесь возможно единственное решение: возвратиться к блок-схеме гл. 11 и выяснить причины неадекватности линейной модели. Напомним некоторые причины, вследствие которых крутое восхождение могло оказаться неэффективным.

1. Интервалы варьирования выбраны неудачно.
2. Исходная модель строилась по полуреплике. Нужно достроить полуреплику до полного факторного эксперимента, получить отдельные оценки для всех коэффициентов регрессии и совершить новое крутое восхождение.

3. Исходная модель строилась по дробной реплике 2^{k-p} , где $p > 1$. Целесообразно использовать метод «перевала», т. е. построить матрицу второй серии опытов, изменив все знаки на обратные. Это даст возможность освободить линейные эффекты от совместных оценок с парными взаимодействиями. Положение не улучшится, если значимыми являются взаимодействия более высокого порядка.

В случае нелинейности исходной модели можно попытаться преобразовать параметр оптимизации. Это обычный прием для снижения степени полинома.

Пример 3. Перед вами (табл. 13.4) план 2^{4-1} с генерирующим соотношением $x_4 = x_1 x_2 x_3$. Этот план применялся при оптимизации процесса получения новокаина. Здесь \tilde{x}_1 — время реакции, мин; \tilde{x}_2 — температура реакционной среды, °C; \tilde{x}_3 — избыток натриевой соли парааминобензойной кислоты, %; \tilde{x}_4 — концентрация натриевой соли парааминобензойной кислоты, %. Параметром оптимизации является выход реакции, %

Таблица 13.4

Матрица планирования 2^{4-1} и результаты эксперимента

Уровень	Факторы				
	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	
Нижний	45	50	0	3,5	
Основной	60	60	4	6,5	
Верхний	75	70	8	9,5	
I_j	15	10	4	3	

Опыты	x_0	x_1	x_2	x_3	x_4	$x_1 x_2 + x_3 x_4$	$x_1 x_3 + x_2 x_4$	$x_2 x_3 + x_1 x_4$	y
1	+	—	—	—	—	+	+	+	61,8
2	+	+	+	+	+	+	+	+	55,4
3	+	+	+	—	—	+	—	—	61,5
4	+	+	—	+	—	—	+	—	61,5
5	+	—	+	+	—	—	—	+	62,0
6	+	+	—	—	+	—	—	+	58,0
7	+	—	+	—	+	—	+	—	56,3
8	+	—	—	+	+	+	—	—	52,7
b_j	58,65	0,45	0,15	—0,75	—3,05	—0,80	0,40	0,65	

$s_{\text{воспр}}^2 = 1;$
 $s_{\{b_j\}}^2 = 0,125;$
 $2,3s_{\{b_j\}} = 0,81.$

В этом примере линейное приближение оказалось неадекватным и крутое восхождение неэффективным.

Какое решение целесообразно принять? Возможны три варианта: 1) построить новый план, уменьшив интервалы варьирования (это позволяет избавиться от эффектов взаимодействия и, возможно, сделать линейное приближение адекватным); 2) достроить линейный план до плана второго порядка, 3) достроить полуреплику до полного факторного эксперимента с тем, чтобы освободить линейные эффекты от смешивания с взаимодействиями второго порядка.

Прежде чем построить новый план, надо обратить внимание на величину $t_{s(b_j)}$. Для $t=2,3$ она равна $\pm 0,81$. Только один коэффициент регрессии b_4 оказался значимым.

Если принять решение построить новый план с уменьшением интервалов варьирования, то может оказаться, что ни один линейный эффект не выделится на фоне ошибок.

Приемлемое решение — достроить полуреплику до полного факторного эксперимента и совершить новое крутое восхождение при неискаженных линейных оценках.

Вот как выглядит достроенный план (табл. 13.5).

Таблица 13.5

Матрица планирования 2⁴

Номер опыта	x_0	x_1	x_2	x_3	x_4	y , %	Номер опыта	x_0	x_1	x_2	x_3	x_4	y , %
1	+	—	—	—	—	61,8	9	+	—	—	—	+	55,0
2	+	+	—	—	—	64,5	10	+	+	—	—	+	58,3
3	+	—	+	—	—	65,3	11	+	—	+	—	+	56,3
4	+	+	+	—	—	61,5	12	+	+	+	—	+	53,3
5	+	—	—	+	—	62,8	13	+	—	—	+	+	52,7
6	+	+	—	+	—	61,6	14	+	+	—	+	+	65,9
7	+	—	+	+	—	62,0	15	+	—	+	+	+	66,0
8	+	+	+	+	—	71,5	16	+	+	+	+	+	55,4

Полученное уравнение регрессии имеет вид

$$\hat{y} = 60,85 + 0,62x_1 + 0,54x_2 + 1,36x_3 - 2,99x_4 - 1,61x_1x_2 + 0,72x_1x_3 -$$

$$- 0,25x_1x_4 + 0,93x_2x_3 - 0,65x_2x_4 + 0,78x_3x_4 - 0,018x_1x_2x_3 -$$

$$- 2,13x_1x_2x_4 - 0,43x_1x_3x_4 - 0,12x_2x_3x_4 - 2,15x_1x_2x_3x_4,$$

$$s_{\text{воипр}}^2 = 1,67; \quad 2s_{\{b_j\}} = 0,65; \quad s_{\{b_j\}} = 0,325.$$

Уравнение регрессии существенно изменилось. Все линейные коэффициенты оказались значимыми, за исключением b_1 . Большой вклад в движение по градиенту вносит x_3 ($b_3 = 1,36$), который в полуреплике оказался незначимым. Произошло это потому, что $b_{124} = -2,13$. А мы предполагали, что оценка $b_3 \rightarrow \beta_3 + \beta_{124}$ будет неискаженной. На этом примере мы еще раз убеждаемся, что эффекты взаимодействий второго порядка не всегда менее значимы, чем эффекты взаимодействий первого порядка. Кстати, в этом примере даже $b_{1234} = -2,15$.

Совершим теперь новое крутое восхождение. Оно оказалось более удачным, чем первое (табл. 13.6).

Крутое восхождение неэффективно. Положение оптимума неопределенное. Если нет информации о положении оптимума и на стадии крутого восхождения не удалось улучшить значение параметра оптимизации, то можно рекомендовать поставить опыты в центре эксперимента с тем, чтобы оценить вклад квадратичных членов. При значимой сумме можно приступить к достройке линейного плана до плана второго порядка, так как наличие квадратичных членов свидетельствует о близости к почти стационарной области.

Обратим еще раз ваше внимание на то, что при незначимой сумме обратного вывода делать нельзя, ибо возможен, например, такой случай: $b_{11} = 5,7$, $b_{22} = -5,3$, $b_{11} + b_{22} = +0,4$. Сумма незначима,

Таблица 13.6

Крутое восхождение ($\tilde{x}_1 = 60$ мин)

Условия движения по градиенту	Факторы			y , %
	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	
$b_j \times I_j$	$0,54 \times 10 = 5,4$	$1,36 \times 4 = 5,44$	$-2,99 \times 3 = -8,97$	
Шаг	0,600	0,606	—1	
Округление	0,600	0,6	—1	
Опыты				
17	60	4	6,5	
18	60,6	4,6	5,5	
19	61,2	5,2	4,5	
20	61,8	5,8	3,5	65,8
21	62,4	6,4	2,5	79,3
22	63,0	7,0	1,5	63,3

так как коэффициенты имеют разные знаки. Это случай, когда имеется два оптимума. Если же есть основание полагать, что оптимум один, то при незначимой сумме квадратичных членов можно приступить ко второму циклу крутого восхождения.

13.3. Резюме

Рассмотрены наиболее типичные решения после крутого восхождения. Как принимать решение, зависит от эффективности крутого восхождения, а также от определенности ситуации (далеко от оптимума, близко, неопределенно) и от адекватности линейной модели.

Если крутое восхождение эффективно и область оптимума близка, возможны два решения: окончание исследования и достройка линейного плана до плана второго порядка в целях описания области оптимума. Какое решение выбрать — это зависит от того, как сформулирована задача оптимизации.

Если область оптимума далека, решение одно: построение линейного плана нового цикла.

В неопределенной ситуации, когда экспериментатор не может определить степень близости оптимума, можно переходить к новому линейному плану.

При неэффективном крутом восхождении приходится возвращаться к блок-схеме гл. 11. Если линейная модель неадекватна, следует поставить опыты в центре эксперимента для грубой оценки квадратичных членов уравнения регрессии. Если сумма квадратичных членов значима, это может свидетельствовать о близости к почти стационарной области. Тогда следует приступить к построению плана второго порядка или кончать исследо-

вания. При незначимой сумме квадратичных членов нужное решение выбрать довольно трудно. Наиболее типичным является построение линейного плана нового цикла.

Литература

1. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
2. Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. О принятии решений в неформализованных ситуациях. Методологические проблемы кибернетики. Материалы к Всесоюзной конференции, т. 2, стр. 40—52. Научный Совет по комплексной проблеме «Кибернетика» АН СССР. М., 1970.
3. Е. В. Маркова, Ю. П. Адлер. О принятии решений в неформализованных ситуациях при планировании экстремального эксперимента. Информационные материалы, № 8(45), с. 63—81. Научный совет по комплексной проблеме «Кибернетика» АН СССР. М., 1970.

Глава четырнадцатая

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Дурной конец заранее отброшен,
Он должен, должен, должен быть хорошим!

В. Врежт. Добрый человек из Сезуана

Еще совсем немного усилий и вы сможете, наконец, сами спланировать эксперимент. Но прежде нам предстоит рассмотреть вопрос о том, как заканчивается исследование. Это, если хотите, последний этап принятия решений. Кроме того, пора подвести итог всему сказанному. Надо оценить трудности, с которыми мы сталкивались, и перспективы.

14.1. Чем кончается эксперимент?

Главный признак, по которому мы судим об окончании исследования, — это значение параметра оптимизации. Хорошо, если оно достигло возможного предела. Тогда все ясно. Что делать, если предел не достигнут? Здесь прежде всего важно, достигнута ли цель. Совсем не всегда мы стремимся к теоретическому пределу. Давайте последовательно рассмотрим возможные ситуации. Начнем с более приятной, когда цель достигнута.

В этом случае прежде всего мы должны провести интерпретацию результата. Важно понять, соответствует ли полученный результат нашим исходным теоретическим представлениям о процессе. Если да, то мы получили подтверждение правильности теории. Если нет, есть повод для ее пересмотра. Во всех случаях мы получили практический результат, к которому стремились, а может быть и что-то еще сверх ожиданий.

Интерпретацией, однако, дело не кончается. Теперь перед нами главный враг — масштаб. (Если, конечно, речь идет о переносе полученных результатов на промышленную установку.) Вы помните из «Ограничений», что речь шла о лабораторном эксперименте. Даже попытка повторить результат в другой лаборатории на аналогичной установке не всегда приводит к успеху, а что будет, если провести эксперимент в полупромышленном или промышленном масштабе?

Результат, будь то оптимальный режим или адекватная интерполяционная модель, безусловно применим к данному конкретному объекту, на котором проводилось исследование. Иногда этого достаточно. Но обычно мы стремимся распространить его

на другие аналогичные объекты — в другие лаборатории или на производство. Вот тут-то нас и поджидает опасность.

Известно, например, что выполнение анализа одной и той же пробы вещества одним и тем же методом в разных лабораториях связано обычно с большей ошибкой, чем ошибка воспроизводимости в одной лаборатории. Аналогичная или даже худшая ситуация возникает при перенесении технологического процесса на другую установку. Причины вполне естественны. Это, прежде всего, различия в свойствах исходного сырья, оборудования и навыках обслуживающего персонала.

До экспериментальной проверки мы не можем сказать, воспроизводится ли, например, оптимальный режим на другой установке. Но у нас нет лучшей рекомендации, чем считать этот режим исходным для исследования процесса на новой установке.

Если экспериментальная проверка показала, что результат воспроизводится с требуемой точностью (хотя, быть может, и хуже, чем на исходной), то задачу можно считать решенной. Можно, конечно, продолжать исследование, переходя к планам второго порядка или к эволюционному планированию на промышленной установке.

Во всяком случае, такой результат благоприятен. После интерпретации модели можно отдохнуть перед новой задачей.

Если же воспроизвести результат не удастся, то приходится реализовать на новой установке новый план и снова заниматься поиском оптимума.

Давайте рассмотрим пример (см. гл. 14, пример 6). Исследовался аппарат для экстракционного разделения редкоземельных элементов. Задача ставилась так: найти интерполяционную формулу, пригодную для расчета смесительно-отстойных экстракторов ящичного типа (разных размеров). В эксперименте использовался аппарат размером (сечением) 300×200 мм. Адекватное интерполяционное уравнение в виде неполного полинома второго порядка для шести факторов оказалось следующим (на основании плана 2^6-2):

$$\bar{y} = 12,16 + 0,53x_1 + 0,53x_2 - 1,38x_3 - 3,22x_4 + 1,44x_5 - 0,62x_6 - 0,84x_1x_2 - 0,50x_1x_3 - 0,78x_2x_4.$$

Проверка полученных рекомендаций в аппарате размером 325×325 мм, что представляло практический интерес, дала отклонения от уравнения в разных точках 5—7%.

В данном случае для расчета аппаратов требуется точность не ниже 10%. Эксперимент привел к лучшим результатам, поэтому можно признать работу законченной, что и было сделано.

Достаточно часто результат бывает не так хорош и новый эксперимент неизбежен. Правда, вероятность того, что мы начинаем новый эксперимент в почти стационарной области, если центром плана будет лучший результат предыдущей серии, довольно велика. Это дает надежду, что небольшой уточняющий эксперимент будет достаточен.

Как вы помните, мы рассматривали случай, когда цель достигнута. Теперь нам предстоит менее приятное занятие — рассмотреть случаи, когда цель достигнуть не удалось. Здесь не до масштабного перехода. Надо разобраться в причинах неудачи.

Причин может быть много, и они весьма разнообразны. Их общей классификации еще нет. Поэтому мы ограничимся отдельными примерами. Примеры будут искусственными, так как в литературе трудно найти описание подобных ситуаций (хотя последние встречаются достаточно часто). Об этом писать не принято (вернее — не принято публиковать).

Представим себе, что экспериментатору пришлось несколько раз совершить движение по градиенту. В первой серии восхождение было эффективным, но цель достигнута не была. Вторая серия неожиданно привела к резкому ухудшению результатов. Пришлось ставить третью серию опытов. Вернувшись в область наилучших значений, экспериментатор снова провел опыты и в третий раз совершил крутое восхождение. Оно оказалось эффективным. Координаты наилучшей точки не совпали с наилучшим результатом первой серии. Впору опустить руки. Попытаемся понять, в чем дело.

Ниже перечислены некоторые из возможных причин возникновения данной ситуации: имеет место регулярный временной дрейф; не включен в рассмотрение важный фактор, произвольно изменяющийся от серии к серии; оба эти обстоятельства действуют одновременно.

Мы не можем предположить, что причиной всему просто дисперсия воспроизводимости, так как крутое восхождение оказывалось эффективным. Если имеет место дрейф, то картина может оказаться похожей на ту, которую мы нарисовали. К сожалению, интерпретация не является однозначной, ибо та же картина возможна и при неучтенном факторе.

Фактически приходится одновременно выдвигать обе гипотезы о дрейфе и о неучтенном факторе. А какую из них проверять сначала — зависит от конкретной задачи и интуитивных решений экспериментатора. Дрейф часто связан с изменениями свойств сырья, старением установки (например, старение катализатора) или изменениями внешних условий.

Где искать пропущенный фактор, сказать трудно. Все зависит от конкретной задачи.

Надо подчеркнуть, что временной дрейф можно интерпретировать как неучтенный фактор, и наоборот. Однако средства, которые используются для преодоления этих трудностей, различны, о чем вы уже, вероятно, догадались.

Кроме того, возможно, что обе причины имеют место одновременно.

Выявление и устранение причин необходимо для успешного решения задачи. Готовых рецептов нет. В каждом случае надо рассматривать конкретную ситуацию.

Многokrатно реализовав шаговую процедуру, вы получили некоторый наилучший результат, который, однако, хуже требуемого. Опыты хорошо воспроизводятся. Все дальнейшие попытки улучшения ни к чему не привели.

Возможные объяснения: достигнут предел возможностей данного объекта, предъявляемым требованиям он не может удовлетворить; на поверхности отклика существует несколько экстремумов и найден не наилучший.

Давайте посмотрим, откуда взялись предложенные варианты.

Так как все делалось «по правилам», а цели достигнуть не удалось, то либо этого нельзя сделать на данном объекте, либо нарушены какие-то предпосылки. Первое положение часто имеет место при ошибочной оценке возможностей объекта. Предполагается, например, что химическая реакция может идти с выходом в 100%. Но при экспериментальной проверке оказывается, что из-за побочных реакций выход более 80% невозможен. Тогда приходится пересматривать представления о возможностях объекта, особенно если ни в теории, ни у аналогичных объектов не известны требуемые результаты. (Вас не должно смущать известное высказывание Эйнштейна о том, как делаются открытия: все знают, что чего-то сделать нельзя, а один человек случайно не знает, вот он-то и делает открытие. Мы ведь учимся сейчас не делать открытия, а доводить до оптимума уже идущие процессы.)

Можно также предполагать, что существуют другие экстремумы, т. е. нарушена предпосылка об унимодальности. Если такое подозрение есть, то его стоит проверить экспериментально, например, повторив крутое восхождение из другой начальной точки. Получение другого оптимального значения подтвердит гипотезу (хотя возвращение в ту же точку ее, к сожалению, не отвергает).

— Прежде чем приступить к экспериментальной проверке гипотезы о многих экстремумах, что требует дополнительных усилий, важно понять, не является ли полученный результат теоретическим пределом. Не стоит проводить эксперимент, если улучшение невозможно.

Как уже отмечалось выше (см. гл. 5), рассмотрение этого случая выходит за рамки нашей книги. Но указание на его возможность совершенно необходимо.

Конкретные решения зависят, конечно, от постановки задачи и интуиции экспериментатора.

Список трудностей можно легко продолжить. Некоторые из них, например случай незначимых коэффициентов линейного уравнения регрессии, мы рассматривали в связи с принятием решений (см. гл. 11 и 13).

Вы видите, сколь разнообразны и сложные ситуации могут возникать при планировании эксперимента. А мы указали далеко не все ситуации. Вам понадобятся непредубежденность и терпение. Но разве существуют легкие пути познания?

14.2. Перспективы

Применение методов планирования эксперимента требует, конечно, творческого подхода. Положение усугубляется тем, что планирование требует отказа от привычного для экспериментатора языка. При этом пропадают привычные надежные ориентиры.

С другой стороны, планирование эксперимента связано с коллективным методом исследования. В этом есть известное облегчение. Если вы работаете в контакте со специалистом по планированию эксперимента, то нет необходимости разбираться в тонкостях всех методов. Но зато надо выработать общий язык, что вовсе не так просто, как кажется. Одна из задач нашей книги как раз и состоит в том, чтобы помочь вам в выработке такого языка.

Наконец, вы должны ясно понимать, что вовсе не все на свете экспериментальные задачи надо обязательно решать с помощью планирования эксперимента. Так, в ряде случаев объект можно наблюдать, но им нельзя управлять. (Ситуации биолог — естественная популяция, астроном — галактика и т. д.) Здесь математический аппарат используется для оптимальной обработки экспериментальных данных, но не для планирования экспериментов. Планирование — не панацея от всех бед, но мощный инструмент исследования, если применять его в подходящих случаях.

Решив изучать методы планирования, вы, конечно, не сможете ограничиться этой книжкой. Ведь мы рассмотрели только один из многих методов планирования эксперимента. Впереди вас ждет много интересного. Вы научитесь работать в многофакторных ситуациях (отсеивать несущественные факторы, исследовать качественные факторы и т. д.), исследовать область оптимума, планировать промышленный эксперимент, использовать планирование при проверке теоретических гипотез и при построении диаграмм состав — свойство. . . Словом, перед вами откроется новый мир. Чтобы помочь войти в него, мы советуем вам познакомиться со следующей литературой [1—14].

14.3. Резюме

Мы полагаем, что, прочтя все предыдущие страницы, вы усвоили материал и теперь знаете, что в некоторых практически важных случаях имеет смысл планировать многофакторный эксперимент.

Чтобы его реализовать, необходимо точно сформулировать цель исследования и выбрать подходящий единственный параметр оптимизации. Очень важно, чтобы множество факторов было полным и чтобы объект удовлетворял требованию управляемости.

Тогда, если известны условия, в которых процесс, в принципе, протекает, пусть даже сколь угодно плохо, можно с помощью интуитивных решений выбрать нулевой уровень и интервалы варьирования факторов. Это открывает возможность выбора под-

ходящего плана — дробной решетки или полного факторного эксперимента.

Наиболее ответственная часть работы — реализация плана. Здесь требуется высокая культура эксперимента, тщательное соблюдение условий, рандомизация опытов.

Обработка результатов с помощью метода наименьших квадратов приводит к получению математической модели объекта в области экспериментирования в виде полинома первой степени. После получения полинома проводится проверка его адекватности и статистической значимости коэффициентов. В этот момент снова приходится вступать на путь принятия интуитивных решений. Это сложное и тонкое дело, требующее навыка.

При благоприятном исходе, если это необходимо, реализуется движение по градиенту — крутое восхождение.

Процедура крутого восхождения в сочетании с принятием решений циклически продолжается, пока не достигается оптимум или пока не найден наилучший возможный результат.

Если оптимум достигнут, то обычно возникает задача масштабирования. Если нет, то приходится разбираться в ситуации и проверять новые гипотезы. Завершение отдельных этапов и работы в целом сопровождается интерпретацией результатов. Все вместе и составляет планирование эксперимента методом Бокса—Уилсона при поиске оптимальных условий проведения процессов.

Литература

1. Ю. П. Адлер, Ю. В. Грановский. Обзор прикладных работ по планированию эксперимента. Изд. 2-е., препринт № 33, МГУ (1972).
2. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.
3. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
4. Ч. Хикс. Основные принципы планирования эксперимента. М., «Мир», 1967.
5. Д. Дж. Уайлд. Методы поиска экстремума. М., «Наука», 1967.
6. Планирование эксперимента. Сб. статей под ред. Г. К. Круга. М., «Наука», 1966.
7. Проблемы планирования эксперимента. Сб. статей под ред. Г. К. Круга. М., «Наука», 1969.
8. Планирование и автоматизация эксперимента в научных исследованиях. Сб. статей под ред. Г. К. Круга. М., «Советское радио», 1974.
9. Е. В. Маркова, А. Н. Лисенков. Планирование эксперимента в условиях неоднородностей. М., «Наука», 1973.
10. В. Г. Горский; Ю. П. Адлер. Планирование промышленных экспериментов. М., «Металлургия», 1974.
11. В. В. Налимов. Теория эксперимента. М., «Наука», 1971.
12. В. В. Федоров. Теория оптимального эксперимента. М., «Наука», 1974.
13. Новые идеи в планировании эксперимента. Под ред. В. В. Налимова. М., «Наука», 1969.
14. Планирование оптимальных экспериментов. Под ред. М. Б. Малютова. МГУ, 1975.

Глава пятнадцатая

КАК РЕШАТЬ ЗАДАЧУ ОПТИМИЗАЦИИ. КОНКРЕТНЫЙ ПРИМЕР

Единственная практическая проблема —
«Что делать дальше?»

Энон. Физики продолжают шутить

Час расставания близок. Чувство сожаления скрашивается надеждой, что вы в какой-то мере уже подготовлены к применению методов планирования эксперимента для оптимизации многофакторных процессов. Однако после чтения четырнадцати глав от всех рекомендаций и советов, часто носящих неоднозначный характер, от обилия разнородных примеров, появилось смутное чувство, что на практике планирование эксперимента использовать трудно. Конечно, все понимают, что уверенность появляется не сразу, она приходит с жизненным опытом после решения нескольких конкретных задач. В этой главе мы ограничимся разбором одной из таких задач — задачи экстракционного разделения циркония и гафния трибутилфосфатом [1]. На этой задаче мы последовательно рассмотрим все логические и вычислительные операции, характерные для процедуры оптимизации. Как театр начинается с вешалки, так и всякая задача начинается с формулировки.

15.1. Предпланирование эксперимента

Задача получения циркония с низким содержанием гафния возникла в связи с использованием циркония и его сплавов как конструкционного материала для ядерных реакторов. Конструкционные материалы должны иметь хорошие механические свойства и быть коррозионно стойкими к действию теплоносителей. Так как мощность реакторов пропорциональна количеству имеющихся в них нейтронов, поглощение нейтронов конструкционными материалами должно быть небольшим. Цирконий и его сплавы являются чуть ли не единственными материалами, удовлетворяющими этим требованиям. Гафний, обладая близкими с цирконием химическими свойствами, довольно сильно поглощает нейтроны.

Циркониевые руды содержат 1—3% гафния, однако содержание гафния в цирконии, используемого как конструкционный материал, не должно превышать сотых долей процента [2].

Экстракционные методы разделения циркония и гафния нашли широкое применение в технологии получения этих металлов. По одному из наиболее распространенных вариантов технологии экстракция проводится раствором трибутилфосфата (ТБФ) в инертном разбавителе из азотнокислых растворов. Однако условия процесса, предлагаемые различными авторами для статической экстракции и одностадийного разделения, в большой степени различаются. Не удивительно поэтому, что величина параметра оптимизации, характеризующего процесс разделения — коэффициент разделения металлов, лежит в диапазоне от 4 до 21 [3].

Выбор параметра оптимизации. Рассмотрим теперь вопрос о параметре оптимизации для данной задачи. В этом случае, как и всегда, сначала обсуждается все множество откликов, известное по литературным данным и из опыта экспериментатора. Затем делается попытка построить единственный параметр оптимизации, а если это не удастся, то приходится использовать некоторое обобщение.

Экстракция относится к классу разделительных процессов, а для них предложено несколько десятков параметров — технологических, термодинамических, экономических, статистических и т. д. Общими и существенными характеристиками разделительных процессов являются полнота извлечения ценного компонента, производительность по готовому продукту и качество продукта. Рост любой из них приводит к падению других характеристик [4].

В настоящем исследовании мы также имеем дело с количеством (полнота извлечения) и качеством (низкое содержание гафния) продукта. В схеме классификации параметров оптимизации (гл. 2) характеристики количества и качества продукта отнесены к группе технико-технологических параметров. Если только ограничиться параметрами этой группы, то следует учитывать, что они недостаточно универсальны, не учитывают экономику процесса. Но их использование вполне допустимо при решении конкретных задач типа оптимизации отдельных стадий, основных операций и т. д. Параметры оптимизации этой группы просты, легко вычисляемы и имеют ясный физический смысл.

Из группы технико-технологических параметров оптимизации для процессов экстракции, реализуемых в лабораторном масштабе, широко распространены коэффициент распределения, фактор извлечения и коэффициент разделения [5]. Коэффициент распределения D равен отношению общей концентрации вещества в органической фазе к его концентрации в водной фазе. Фактор извлечения R (процент экстракции) — доля или процент от общего количества вещества, экстрагируемого в органическую фазу при данных условиях. Коэффициент разделения двух металлов равен отношению коэффициентов распределения. Как следует из этих определений, между коэффициентом распределения и фактором извлечения при равенстве объемов фаз существует связь:

$R = 100D / (D + 1)$. Еще предлагалось характеризовать разделение двух металлов фактором обогащения, равным отношению факторов извлечения [6]. Для систем с высокой разделяющей способностью использовался параметр оптимизации, равный сумме извлечений одного из металлов в органическую фазу, а другого — в водную фазу [7].

В качестве параметра оптимизации выбираем коэффициент разделения. При этом учтем, что в нескольких предварительных опытах было достигнуто значение коэффициента разделения циркония и гафния, близкое к десяти, а извлечение суммы металлов в органическую фазу составило 30—40%. Легко установить, что извлечение гафния в органическую фазу невелико, всего несколько процентов. Для сокращения сроков исследования на первых этапах решения задачи, по-видимому, можно ограничиться одним параметром оптимизации — коэффициентом разделения. Но на последующих этапах очевидно следует вводить характеристики процесса, связанные с полнотой извлечения, например извлечение суммы металлов в органическую фазу. Тогда же целесообразно будет обратиться к обобщенному параметру оптимизации.

Последовательный подход к выбору параметров оптимизации не должен вызывать возражений. Некоторым примером может служить шахматная игра, когда конечная цель достигается последовательным выполнением иных задач: максимального развития своих фигур, достижения материального перевеса и т. д. [8].

Здесь естественно ставить задачу получения максимальных значений коэффициента разделения, так как при эффективном разделении металлов сокращаются расходы экстрагента и упрощается аппаратурная схема процесса.

Еще одно обстоятельство в пользу выбора коэффициента разделения. Представляет интерес сравнение результатов планируемых опытов с данными классических однофакторных экспериментов, известных из литературы.

Таким образом, в этой задаче решено использовать коэффициент разделения. Нужно, однако, иметь в виду, что коэффициент разделения может иметь большую величину и при плохом разделении (например, $D_1 = 10^3$, $D_2 = 10$) [6]. Следовательно, когда извлечение сразу двух металлов в органическую фазу велико, применение этого критерия нецелесообразно.

К моменту начала эксперимента в предварительных опытах наилучшие значения коэффициента разделения равнялись 10, а в литературе встречались значения около 20. Если бы в результате оптимизации удалось получить значения, превосходящие 20, то работу можно было бы считать успешной.

Из априорных данных известна примерная оценка дисперсии воспроизводимости: $s_{\{y\}}^2 = 1,0$.

Выбор факторов. В гл. 4 мы отмечали, что следует включать в рассмотрение все существенные факторы, которые могут влиять

на процесс. Необходимо пояснить, что значит существенные факторы.

Принципиально на любой объект исследования влияет неограниченно большое число факторов. Какие из них являются существенными, а какие нет, определяется постановкой задачи, условиями проведения эксперимента, состоянием теоретических представлений в этой области. Например, достижение в производственных условиях требуемого по количеству и качеству продукта варьированием небольшого числа факторов удешевляет процесс, в то время как во многих лабораторных исследованиях экономика может отступать на задний план, а первостепенное значение приобретает содержательная информация о механизме процесса, получаемая при варьировании многих факторов. Имеет смысл на первых этапах исследования включать в рассмотрение большее число факторов (хотя бы по сравнению с аналогичными задачами, выполняемыми с помощью однофакторных экспериментов). Экспериментальная оценка «силы влияния» позволит выделить наиболее существенные факторы.

Обсудим вопрос о выборе факторов. Для данной экстракционной системы с целью отбора факторов были использованы методы априорного ранжирования и случайного баланса [9, 10]. Может быть, когда-нибудь мы вам и расскажем об этих весьма интересных методах выделения наиболее влияющих факторов. Они интересны еще и потому, что здесь возможно предложить последовательную схему оптимизации многофакторных процессов методами планирования эксперимента: метод априорного ранжирования факторов → метод случайного баланса → метод крутого восхождения → планирование второго порядка [3]. Возвращаясь к нашей задаче, ограничимся кратким замечанием: из двенадцати «подозреваемых» факторов было выделено четыре: концентрации суммы металлов и азотной кислоты в исходном водном растворе, концентрация реагента ТБФ в разбавителе и соотношение объемов фаз. Эти результаты не противоречат существующим теоретическим представлениям о роли различных факторов в подобных экстракционных системах [6].

Итак, варьируемые факторы: \bar{x}_1 — концентрация металлов на сумму двуокисей циркония и гафния $\{\Sigma Zr(Hf)O_2\}$ в исходном водном растворе, г/л; \bar{x}_2 — концентрация азотной кислоты в исходном водном растворе, мол/л; \bar{x}_3 — концентрация ТБФ в *о*-ксилоле, объемные %; \bar{x}_4 — соотношение объемов фаз, органической к водной.

Методика эксперимента. Для приготовления исходных растворов применяли оксинитрат циркония, содержащий 1,5–2,1 % гафния по отношению к сумме двуокисей. Исходные азотнокислые растворы оксинитрата с различной концентрацией соли и кислоты получались разбавлением концентрированного водного раствора соли. Экстрагентом служил ТБФ, разбавленный *о*-ксилолом. Опыты проводили в статических условиях. Экстрагент предва-

рительно насыщался азотной кислотой. Раствор и экстрагент (по 20 мл) перемешивали в делительных воронках в течение 15 мин на вибрационном аппарате. Реэкстракция осуществлялась двукратным встряхиванием органической фазы с равными объемами дистиллированной воды. Водную фазу после экстракции и реэкстракции анализировали на содержание суммы окислов циркония и гафния гравиметрическим методом. Содержание гафния определялось спектральным методом. Концентрация азотной кислоты устанавливалась титрованием по метиловому красному.

15.2. Выбор условий проведения опытов

Выбор области определения факторов. Вначале рассмотрим выбор области для первого фактора — концентрации металла. Здесь, как это часто бывает, приходится сталкиваться с противодействующими тенденциями. С одной стороны, выгодно повышать концентрацию суммы металлов, так как при этом усиливается процесс подавления экстракции гафния цирконием, и коэффициент разделения растёт. С другой стороны, повышению концентрации металлов препятствуют трудности приготовления исходных растворов и образование еще одной, третьей фазы, что снижает разделение.

По данным разных авторов [3], концентрация суммы металлов (это и есть первый фактор) изменяется от 35 до 120 г/л. Если «классический» эксперимент, с помощью которого получены эти цифры, можно сравнить с выстрелом по цели из ружья, то метод крутого восхождения — это уже выстрел из пушки. Имеет смысл выбрать «площадь поражения» побольше, а потому — область определения для концентрации суммы металлов установлена от 20 до 150 г/л.

Второй фактор — концентрация азотной кислоты. При концентрации менее 3 мол/л цирконий и гафний экстрагируются слабо. С ростом концентрации увеличивается и извлечение металлов в органическую фазу. При концентрации кислоты выше 5 мол/л гафний хорошо извлекается в органическую фазу, разделение падает. Здесь «рассогласование» литературных данных гораздо меньше, чем по концентрации металла: диапазон всего от 5 до 6 мол/л. Мы выбрали область определения от 3 до 8 мол/л.

Перейдем к концентрации реагента в разбавителе и соотношению объемов фаз. Литературные рекомендации (полученные, естественно, однофакторным экспериментом) сводятся к таким цифрам: $\bar{x}_3=50-60$ об. %, $\bar{x}_4=1:1-2:1$. Целесообразно уменьшать концентрацию ТБФ в разбавителе в надежде улучшить экономические показатели процесса. Повышению концентрации реагента препятствует высокая вязкость растворов. Поэтому растворы концентрации выше 70% обычно не используются. Для ТБФ выбрана область 20–70 об. %.

Предварительными экспериментами установлено слабое извлечение металлов при соотношении объемов фаз менее 1 : 1. Верхняя граница определяется экономическими соображениями. Нам задана область определения для соотношения объемов фаз от 1 : 1 до 4 : 1. Сведения об областях определения факторов, точности поддержания уровней и области, в которой по литературным данным и результатам предварительных экспериментов оптимизация наиболее целесообразна (область интереса), представлены в табл. 15.1.

Таблица 15.1

Характеристики факторов*

Фактор	Область определения	Область интереса
Концентрация металлов на сумму двуокисей в исходном водном растворе, г/л	20—150	20—60
Концентрация азотной кислоты в исходном водном растворе, мол/л	3—8	3—7
Концентрация трибутилфосфата в разбавителе, об. %	20—70	20—50
Соотношение фаз (органическая к водной)	1 : 1—4 : 1	1 : 1—2 : 1

* Точность (коэффициент вариации) для всех факторов не более 3%.

Теперь перейдем к выбору основного уровня факторов.

Принятие решений при выборе основного уровня факторов. Уже отмечалось ранее, что оптимизации новых многофакторных процессов методами планирования эксперимента обычно предшествуют некоторые (мы их называем предварительными) опыты, из которых мы узнаем, что процесс «идет», опыты воспроизводятся и т. д. Несколько подобных опытов в подобласти $\bar{x}_1=20-40$ г/л; $\bar{x}_2=3-5$ мол/л; $\bar{x}_3=20-40$ об. %; $\bar{x}_4=1:1-2:1$ были проделаны и в нашем случае. Здесь процесс экстракционного разделения протекал без трудностей: расслаивание водной и органической фаз достигалось достаточно быстро, не образовывалась третья фаза и т. д. Правда, значения коэффициента разделения в предварительных опытах не превышали десяти. Будем считать, что в отмеченной подобласти процесс разделения протекал достаточно хорошо. Нашей ситуации соответствуют два эквивалентных решения: выбор центра подобласти и выбор случайной точки в подобласти. Мы остановились на первом решении. Таким образом, основной уровень факторов равен: $\bar{x}_1=30$ г/л, $\bar{x}_2=4$ мол/л, $\bar{x}_3=30$ об. %, $\bar{x}_4=1,5:1$.

Пока вам кажется, что процедуры выбора параметров оптимизации, факторов и их основного уровня проходят «гладко». И часто бывает именно так. Можно сказать, что пока «полет совершается в тихую погоду». Но впереди «грозовой» фронт — принятие реше-

ний при выборе интервалов варьирования факторов. К этому этапу мы сейчас и обратимся.

Принятие решений при выборе интервалов варьирования факторов. Согласно сделанным ранее рекомендациям (гл. 6), при выборе интервалов варьирования факторов следует учитывать точность фиксирования значений факторов, информацию о кривизне поверхности отклика и диапазоне изменения параметра оптимизации. Погрешности в фиксировании факторов не превышают 3%, что было подтверждено несколькими специально поставленными опытами. По нашей приближенной классификации — это средняя точность. Диапазон изменения параметра оптимизации может быть признан широким, так как в предварительных опытах разброс значений параметра оптимизации существенно превышал ошибку воспроизводимости. Осталось рассмотреть вопрос о кривизне поверхности отклика. По факторам \bar{x}_2 , \bar{x}_3 и \bar{x}_4 кажется разумным предположение об отсутствии информации по кривизне поверхности отклика. Только по фактору \bar{x}_1 можно полагать наличие нелинейной зависимости, в связи со значительным расхождением приводимых рекомендаций (35—123 г/л).

Для выбора интервалов обращаемся к блок-схеме принятия решений при средней точности фиксирования факторов (гл. 6). Напомним, что ситуации обозначены номерами в кружочках, признаки ситуации определяются стрелками, направленными к данному кружочку, стрелка, выходящая из него, указывает решение. Выбранные признаки в рассматриваемой задаче для факторов \bar{x}_2 , \bar{x}_3 , \bar{x}_4 соответствуют ситуации 16 и принимаемое решение — средний интервал варьирования. Что же касается фактора \bar{x}_1 , то здесь имеет место ситуация 13 блок-схемы и другое решение — узкий интервал варьирования.

Таблица 15.2

Уровни и интервалы варьирования факторов (первая серия)

Уровни	Факторы			
	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4
Основной	30	4	30	1,5
Интервал варьирования	5	1	10	0,5
Верхний	35	5	40	2
Нижний	25	3	20	1

Условия первой серии опытов приведены в табл. 15.2. Интервал варьирования по первому фактору составляет примерно 4% от области определения, по второму и третьему факторам — 20%, по четвертому фактору — 17%. Если интервал составляет не

более 10% от области определения, то его можно отнести к узким, если не более 30% — к средним. В нашем примере эти условия соблюдены, поэтому можно считать выбор решения соответствующим блок-схеме. В этой экспериментальной области существует возможность установления факторов на любых уровнях и заданные значения уровней остаются постоянными в течение опыта.

15.3. Выбор и реализация плана (первая серия)

В гл. 5 был изложен шаговый принцип поиска оптимума, связанный с реализацией коротких серий опытов, позволяющих строить модели для оценки градиента. Мы предполагаем, что в выбранной области первой серии опытов линейная модель окажется адекватной.

Выбор плана. При выборе плана прежде всего приходится учитывать критерии оптимальности и число опытов. В нашем случае ясно, что искомый план должен быть двухуровневым, так как нас интересует линейная модель, ортогональным и ротатабельным. Ортогональность позволяет двигаться по градиенту пропорционально коэффициентам линейной модели и независимо интерпретировать эффекты. Ротатабельность обеспечивает гарантированное равенство дисперсий предсказания при движении в любом направлении от центра эксперимента. Всем этим требованиям удовлетворяет факторный эксперимент 2^4 . Однако если принять во внимание стремление к минимизации опытов, то целесообразно применить полуреплику, которая относительно линейной модели сохраняет оптимальные свойства полного факторного эксперимента.

Существуют альтернативные планы, такие, как симплекс-планирование Плакетта—Бермана и др. (см. гл. 16), однако они не удовлетворяют одновременно всем сформулированным требованиям. Эффективность дробной реплики определяется системой смешивания (гл. 7). Здесь нет каких-либо специальных соображений о большей значимости тройных взаимодействий по сравнению с парными взаимодействиями, поэтому нет смысла выбирать полуреплику 2^{4-1} с разрешающей способностью III. Коль скоро мы намечали движение по градиенту, целесообразно получить линейные эффекты, свободные от парных эффектов взаимодействия. Такие оценки позволяют получать главные полуреплики от полного факторного эксперимента 2^4 , планы с разрешающей способностью IV. Мы выбрали полуреплику с генерирующим соотношением $x_4 = x_1 x_2 x_3$.

Матрица планирования приведена в табл. 15.3. При выборе такого типа планирования получаем следующую систему смешивания оценок:

Таблица 15.3

Матрица планирования опытов (первая серия)

Номер опыта	x_1	x_2	x_3	x_4	Номер опыта	x_1	x_2	x_3	x_4
1	-1	-1	-1	-1	5	+1	+1	-1	-1
2	+1	-1	+1	-1	6	+1	-1	-1	+1
3	-1	-1	+1	+1	7	-1	+1	+1	-1
4	-1	+1	-1	+1	8	+1	+1	+1	+1

$$\begin{aligned} b_0 &\rightarrow \beta_0, & b_3 &\rightarrow \beta_3, & b_{13} &\rightarrow \beta_{13} + \beta_{24}, \\ b_1 &\rightarrow \beta_1, & b_4 &\rightarrow \beta_4, & b_{14} &\rightarrow \beta_{14} + \beta_{23}, \\ b_2 &\rightarrow \beta_2, & b_{12} &\rightarrow \beta_{12} + \beta_{34}, \end{aligned}$$

Таблица 15.4

Условия проведения опытов (первая серия)

Номер опыта	Номер параллельного опыта	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	Номер опыта	Номер параллельного опыта	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4
1	9	25	3	20	1	5	13	35	5	20	1
2	10	35	3	40	1	6	14	35	3	20	2
3	11	25	3	40	2	7	15	25	5	40	1
4	12	25	5	20	2	8	16	35	5	40	2

Условия проведения первой серии опытов в натуральных переменных представлены в табл. 15.4. Переход от натуральных значений факторов к кодированным значениям задается формулами:

$$x_1 = \frac{\tilde{x}_1 - 30}{5}, \quad x_2 = \frac{\tilde{x}_2 - 4}{1}, \quad x_3 = \frac{\tilde{x}_3 - 30}{10}, \quad x_4 = \frac{\tilde{x}_4 - 1,5}{0,5}.$$

Еще один важный вопрос для обсуждения — выбор числа параллельных опытов. Они нужны для исключения грубых наблюдений, оценки дисперсии воспроизводимости. Заметим, что если не проводятся параллельные опыты, то возможное неверное установление значения параметра оптимизации хотя бы в одном опыте изменит оценки всех коэффициентов регрессии. Реализация хотя бы двух параллельных опытов позволяет избежать этой ошибки. В нашем случае предварительная оценка дисперсии воспроизводимости ($s_{\langle y \rangle}^2 = 1,0$) казалось установленной достаточно надежно, поэтому было выбрано минимальное число параллельных опытов — два. Номера параллельных опытов приведены в табл. 15.4.

Последняя операция перед проведением эксперимента — рандомизация опытов, заключающаяся в выборе случайной последо-

вательности при постановке опытов, для исключения влияния систематических ошибок, вызванных внешними условиями (гл. 8).

Общее число проводимых опытов — шестнадцать. С помощью таблицы случайных чисел получена последовательность: 15, 13, 10, 5, 14, 4, 6, 1, 7, 8, 3, 2, 9, 12, 11, 16. В табл. 15.5 приведены порядок выполнения, матрица планирования и результаты первой серии опытов (здесь y' и y'' — результаты параллельных опытов, \bar{y} — их среднее значение).

Таблица 15.5

Порядок проведения и результаты опытов, матрица планирования (первая серия)

Номер опыта	Порядок проведения двух повторных опытов	x_0	x_1	x_2	x_3	x_4	y'	y''	\bar{y}
1	8;13	+1	-1	-1	-1	-1	2,40	1,46	1,93
2	3;12	-1	+1	-1	+1	-1	3,58	2,78	3,18
3	11;15	-1	-1	-1	+1	+1	6,10	4,50	5,30
4	6;14	+1	-1	+1	-1	+1	4,76	3,12	3,94
5	2;4	+1	+1	+1	-1	-1	5,35	4,15	4,75
6	5;7	+1	+1	-1	-1	+1	4,10	2,70	3,40
7	1;9	+1	-1	+1	+1	-1	5,42	4,00	4,71
8	10;16	-1	+1	+1	+1	+1	15,41	12,99	14,20

15.4. Обработка результатов эксперимента

Обработку результатов проводим по схеме с равномерным дублированием опытов (гл. 10) в следующей последовательности.

1. Оценка дисперсий среднего арифметического в каждой строке матрицы.

2. Проверка однородности дисперсий с помощью критерия Кохрена.

3. Если дисперсии однородны, то проводится расчет оценки дисперсии воспроизводимости.

4. Определение коэффициентов регрессии.

5. Проверка адекватности модели.

6. Проверка значимости коэффициентов регрессии.

Запишем формулы, по которым будем проводить расчеты (по дисперсиям средних, а не по индивидуальным, как в гл. 8):

$$1) s_i^2 = \frac{\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{n(n-1)};$$

$$2) G = \frac{s_{\max}^2}{\sum_{i=1}^n s_i^2};$$

$$3) s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{Nn(n-1)};$$

$$4) b_j = \frac{\sum_{i=1}^N \bar{y}_i x_{ji}}{N}, \quad b_{uj} = \frac{\sum_{i=1}^N \bar{y}_i x_{ui} x_{ji}}{N};$$

$$5) s_{ax}^2 = \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2 / f, \quad f = N - (k+1), \quad F = s_{ax}^2 / s_{\{y\}}^2;$$

$$6) s_{\{b_j\}}^2 = s_{\{y\}}^2 / N, \quad \Delta b_j = \pm t s_{\{b_j\}}.$$

Значения дисперсий среднего арифметического каждого опыта приведены в табл. 15.6. Критерий Кохрена равен: $G = 1,464/4,510 = 0,324$. Табличное значение критерия для восьми разных опытов и числа степеней свободы $n-1=1$ равно 0,679 (уровень значимости 0,05). Экспериментальная величина G -критерия не превышает табличного значения, гипотеза об однородности дисперсий принимается. Дисперсия воспроизводимости равна $s_{\{y\}}^2 = \sum_{i=1}^N s_i^2 / 8 = 0,564$. Число степеней свободы этой дисперсии $N(n-1) = 8$.

Таблица 15.6

Дисперсии среднего арифметического (первая серия опытов)

Номер опыта	s_i^2	Номер опыта	s_i^2	Номер опыта	s_i^2	Номер опыта	s_i^2
1	0,220	3	0,640	5	0,360	7	0,504
2	0,160	4	0,672	6	0,490	8	1,464

Составим расчетную матрицу (табл. 15.7). Оценки коэффициентов регрессии и дисперсий в их определении:

$$\begin{aligned} b_0 &= 5,1762; & b_8 &= 1,6712; & b_{18} &= 0,6362; \\ b_1 &= 1,2062; & b_4 &= 1,5337; & b_{14} &= 0,8837; \\ b_2 &= 1,7237; & b_{12} &= 1,3687; & s_{\{b_j\}}^2 &= 0,564/8 = 0,0705. \end{aligned}$$

Результаты расчета остаточной суммы квадратов при проверке адекватности линейной модели приведены в табл. 15.8. В этой таблице \bar{y} — средние из двух параллельных значений экспери-

Таблица 15.7

Расчетная матрица и результаты опытов (первая серия)

Номер опыта	x_0	x_1	x_2	x_3	x_4	$x_1x_2 = x_0x_4$	$x_1x_3 = x_2x_4$	$x_1x_4 = x_2x_3$	\hat{y}
1	+1	-1	-1	-1	-1	+1	+1	+1	1,93
2	+1	+1	-1	+1	-1	-1	+1	-1	3,18
3	+1	-1	+1	+1	+1	+1	-1	-1	5,30
4	+1	+1	+1	-1	+1	-1	+1	-1	3,94
5	+1	+1	-1	-1	-1	+1	-1	-1	4,75
6	+1	+1	-1	+1	+1	-1	-1	+1	3,40
7	+1	-1	+1	+1	-1	-1	-1	+1	4,71
8	+1	+1	+1	+1	+1	+1	+1	+1	14,20

Таблица 15.8

Расчет остаточной суммы квадратов (первая серия опытов)

Номер опыта	y	\hat{y}	$\Delta y = \hat{y} - y$	$(\Delta y)^2$	Номер опыта	y	\hat{y}	$\Delta y = \hat{y} - y$	$(\Delta y)^2$
1	1,93	-0,9587	-2,8887	8,344	5	4,75	4,9012	0,1512	0,023
2	3,18	4,7962	1,6162	2,612	6	3,40	4,5212	1,1212	1,257
3	5,30	5,4512	0,1512	0,023	7	4,71	5,8312	1,1212	1,257
4	3,94	5,5562	1,6162	2,612	8	14,20	11,3112	-2,8887	8,344

ментальных величин параметра оптимизации, \hat{y} — рассчитанные по уравнению регрессии. Дисперсия адекватности $s_{ад}^2 = 24,46/3 = 8,15$. Число степеней свободы дисперсии адекватности $f = 8 - (4 + 1) = 3$. Критерий для проверки гипотезы адекватности модели $F = 8,15/0,564 = 14,4$. Табличное значение критерия Фишера (гл. 9) для числа степеней свободы числителя 3 и знаменателя 8 равно 4,1. Экспериментальная величина F -критерия превышает табличное значение, гипотеза об адекватности модели отвергается. Этот вывод мы могли сделать и принимая во внимание значимость эффектов взаимодействия факторов, что собственно является другим критерием неадекватности линейной модели.

Осталось оценить значимость коэффициентов регрессии. Величина t -критерия для уровня значимости 0,05 и числа степеней свободы, с которыми определялась $s_{\hat{y}}^2$, равна 2,306 (гл. 9). Доверительный интервал $\Delta b_j = \pm 2,306 \cdot 0,266 = 0,613$. Абсолютные величины коэффициентов регрессии больше доверительного интервала, гипотеза о незначимости коэффициентов регрессии отвергается.

15.5. Интерпретация результатов.

Принятие решений

после построения модели (первая серия)

Начнем этот раздел с интерпретации результатов, — перевода модели на язык экспериментатора. Так как линейная модель неадекватна, при интерпретации необходимо учитывать парные эффекты взаимодействия факторов. Здесь, конечно, следует помнить, что эффекты взаимодействия первого порядка попарно смешаны. Линейные коэффициенты регрессии примерно одинаково влияют на параметр оптимизации. Характер их влияния также одинаков: с увеличением значений факторов растет и величина параметра оптимизации. Если же еще учесть, что коэффициенты эффектов взаимодействия положительны, то можно считать, что увеличение концентраций металла, кислоты и реагента, а также соотношения фаз приводят к росту коэффициента разделения. Это не противоречит литературным данным, предполагавшим существование наилучших значений коэффициента разделения циркония и гафния в области более высоких значений факторов. Величина коэффициента разделения, полученная в последнем опыте (см. табл. 15.7), близка к некоторым литературным данным.

Теперь займемся принятием решений. Условия наилучшего опыта (опыт № 8, табл. 15.7), по-видимому, лежат в области, близкой к оптимуму. В блок-схеме принятия решений в задаче определения оптимальных условий, линейная модель неадекватна, этой ситуации соответствуют два возможных решения: реализация плана второго порядка и окончание исследования. Принимать последнее решение у нас нет оснований — мы еще не можем считать задачу решенной. Поэтому останавливаемся на другом решении — реализации плана второго порядка.

Здесь несколько слов следует сказать о различных возможных вариантах перехода к планированию второго порядка. Наиболее распространенные на практике планы второго порядка для четырех факторов, ортогональные, ротатабельные, D -оптимальные и некоторые другие, содержат от 24 до 31 опытов. «Ядро» таких планов составляет полный факторный эксперимент 2^4 . Возможно выполнение сразу всех опытов или другой путь — последовательное композиционное «дистраивание» плана. Сначала реализуется полуреплика от полного факторного эксперимента, проверяется гипотеза адекватности линейной модели. Если модель адекватна, то можно переходить к движению по градиенту. При неадекватной модели следует «дистраивать» полуреплику либо до плана второго порядка, либо до полного факторного эксперимента. В последнем случае снова проверяется гипотеза адекватности линейной модели и т. д.

Второй путь — последовательная реализация плана, кажется предпочтительным с точки зрения экономии опытов. И он действительно весьма распространен на практике.

Таким образом, мы пришли к решению: перенос центра плана в условия наилучшего опыта первой серии и последовательное построение плана второго порядка. Теперь остается принять решение о выборе интервалов варьирования факторов во второй серии опытов. Снова обращаемся к блок-схеме принятия решений при средней точности фиксирования факторов (гл. 6). В отличие от процедуры принятия решений в первой серии опытов, сейчас появилась информация о нелинейности поверхности отклика. Широкий диапазон изменения параметра оптимизации вместе с установленной характеристикой поверхности отклика приводит к единственному решению — узкому интервалу варьирования факторов.

15.6. Реализация плана (вторая серия)

Условия, матрица планирования и результаты второй серии опытов представлены в табл. 15.9. Интервалы варьирования факторов уменьшены в два раза по сравнению с первой серией опытов

Таблица 15.9

Условия, матрица планирования и результаты опытов (вторая серия)

Уровни		Факторы				Уровни	Факторы			
		\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4		\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4
Основной интервал варьи- рования		35	5,0	40	2	Верхний Нижний	37,5	5,5	45	2,25
		2,5	0,5	5,0	0,25		32,5	4,5	35	1,75
Опыты	Порядок про- ведения двух повторных опытов	Кодированные значения факторов								
		x_0	x_1	x_2	x_3	x_4	y'	y''	g	
1	4; 11	+1	-1	-1	-1	-1	12,15	10,85	11,50	
2	8; 9	+1	+1	-1	+1	-1	8,70	7,70	8,20	
3	3; 14	+1	-1	-1	+1	+1	9,84	7,16	8,50	
4	10; 15	+1	-1	+1	-1	+1	15,21	12,79	14,00	
5	2; 12	+1	+1	+1	-1	-1	12,10	13,30	12,70	
6	1; 13	+1	+1	-1	-1	+1	9,25	10,75	10,00	
7	5; 7	+1	-1	+1	+1	-1	13,20	11,80	12,50	
8	6; 16	+1	+1	+1	+1	+1	12,85	14,15	13,50	

и составляют не более 10% от области определения факторов. В этой серии снова реализована полуреплика от полного факторного эксперимента 2^4 с генерирующим соотношением $x_4 = x_1 x_2 x_3$

и двумя параллельными опытами. Кроме того, был добавлен еще один опыт в центре плана для оценки значимости суммы коэффициентов регрессии при квадратичных членах ($\sum \beta_{jj}$). В этом опыте получены следующие значения параметра оптимизации: $y' = 13,37$; $y'' = 14,83$; $\bar{y} = 14,10$.

Обработку результатов проводим по той же схеме. Значения дисперсий среднего арифметического каждого опыта второй серии приведены в табл. 15.10. В этой серии учтена информация о параллельных опытах в центре плана. Критерий Кохрена $G = 1,795/6,300 = 0,28$. Табличное значение критерия для девяти опытов и одной степени свободы равно 0,638; гипотеза об однородности дисперсий не отвергается. Дисперсия воспроизводимости равна 0,700, для нее число степеней свободы девять.

Таблица 15.10

Дисперсии среднего арифметического (вторая серия опытов)

Номер опыта	s_i^2	Номер опыта	s_i^2	Номер опыта	s_i^2
1	0,422	4	1,464	7	0,490
2	0,250	5	0,360	8	0,422
3	1,795	6	0,562	9	0,533

Оценки коэффициентов регрессии и дисперсии в их определении:

$$\begin{aligned}
 b_0 &= 11,3625; & b_4 &= 0,1375; \\
 b_1 &= -0,2625; & b_{12} &= 0,1875; & s_{\{b_j\}}^2 &= 0,0875; \\
 b_2 &= 1,8125; & b_{13} &= 0,4375; \\
 b_3 &= -0,6875; & b_{14} &= 0,5125; & s_{\{b_j\}} &= 0,296.
 \end{aligned}$$

Результаты расчета остаточной суммы квадратов при проверке адекватности линейной модели во второй серии опытов представлены в табл. 15.11. Дисперсия адекватности и критерий Фишера равны $s_{ад}^2 = 3,912/3 = 1,304$; $F = 1,304/0,700 = 1,86$. Табличное значение F -критерия для трех и девяти степеней свободы равно 3,9 (гл. 8); нет оснований отбрасывать гипотезу адекватности линейной модели.

Величина доверительного интервала для коэффициентов регрессии $\Delta b_j = 0,296 \cdot 2,26 = 0,669$. Оказалось, что два линейных коэффициента регрессии (b_1 и b_4), а также все эффекты взаимодействия первого порядка незначимы, что могло быть результатом сужения интервалов варьирования.

Ранее отмечалось (гл. 9), что кроме величины F -критерия, формальными признаками, по которым можно установить неадекватность линейной модели, являются: 1) значимость хотя бы

Таблица 15.11

Расчет остаточной суммы квадратов (вторая серия опытов)

Номер опыта	y	\hat{y}	$\Delta y = \hat{y} - y$	$(\Delta y)^2$
1	11,50	10,3625	-1,1375	1,2939
2	8,20	8,4625	0,2625	0,0689
3	8,50	9,2625	0,7625	0,5814
4	14,00	14,2625	0,2625	0,0689
5	12,70	13,4625	0,7625	0,5814
6	10,00	10,1125	0,1125	0,0126
7	12,50	12,6125	0,1125	0,0126
8	13,50	12,3625	-1,1375	1,2937

$$\sum_{i=1}^N \Delta y_i^2 = 3,913$$

одного из эффектов взаимодействий; 2) значимость суммы коэффициентов регрессии при квадратичных членах $\Sigma \beta_{jj}$. Первый признак подтвердил гипотезу адекватности. Оценкой суммы коэффициентов регрессии служит разность между b_0 и значением зависимой переменной в центре плана y_0 . В нашем случае $y_0 - b_0 = 14,10 - 11,36 = 2,74$; эта величина значительно превосходит ошибку опыта $s_{(y)} = 0,84$ и поэтому гипотеза о незначимости коэффициентов при квадратичных членах не может быть принята.

Еще во введении мы предупреждали, что даже простая процедура планирования эксперимента может оказаться весьма коварной. Этот пример — подтверждение тому. Наличие квадратичных эффектов указывает на кривизну поверхности отклика, что приводит к плану второго порядка. Сужение же интервалов варьирования факторов привело к тому, что гипотеза адекватности линейной модели не была отвергнута.

15.7. Интерпретация результатов.

Принятие решений после построения модели
(вторая серия)

Среднее значение b_0 во второй серии опытов оказалось примерно в два раза выше, чем в первой серии. Это еще один аргумент в пользу гипотезы о близости области оптимума. В этой области факторного пространства влияние концентрации кислоты выше, чем влияние концентрации экстрагента. Обратим внимание на изменение знака коэффициента b_3 по сравнению с первой серией опытов. Изменение знака указывает, что область оптимальных значений по концентрации экстрагента «где-то рядом».

Теперь придется снова принимать решение. Перенос центра эксперимента в лучшую точку можно рассматривать как некоторый способ поиска оптимальных условий. И поскольку этот прием оказался эффективным, обратимся к блок-схеме принятия решений после крутого восхождения, крутое восхождение эффективно (гл. 12). Здесь, как и при принятии решений после первой серии опытов, те же два варианта при близости области оптимума: окончание исследования и план второго порядка для описания области оптимума. Будем выполнять намеченный ранее план — дотраивание полурешетки до плана второго порядка.

Это решение потребовало выполнения еще 16 опытов. Затем было рассчитано уравнение регрессии второго порядка. С его помощью найдены условия опытов, для которых величина параметра оптимизации оказалась близкой к 30. Однако описание этих опытов лежит за той чертой, которая отделяет нашу книгу от грочего мира.

Литература

1. Ю. В. Грановский, И. А. Чернова, Ю. П. Адлер и др. Математическая модель для процесса экстракционного разделения гафния и циркония трибутилфосфатом. — Заводская лаборатория, 1963, 29, № 1.
2. Металлургия циркония. Под ред. Г. А. Меерсона и Ю. В. Гагаринского. М., ИЛ, 1959.
3. В. И. Спицын, Ю. В. Грановский, Л. Н. Комиссарова. Планирование экспериментов при изучении экстракции циркония и гафния. — В сб. «Планирование эксперимента». М., «Наука», 1966.
4. Л. А. Барский, И. Н. Пляксин. Критерии оптимизации разделительных процессов. М., «Наука», 1967.
5. И. П. Алимарин, Ю. А. Золотов. Терминология экстракции. — Ж. анал. хим., 1971, 26, № 5.
6. Ю. А. Золотов, Б. З. Иофа, Л. К. Чукалин. Экстракция галогенидных комплексов металлов. М., «Наука», 1973.
7. М. Ю. Медведев, В. Г. Майоров, А. Г. Бабкин и др. Изучение оптимальных условий экстракции и разделения ниобия и тантала. — В сб. «Планирование эксперимента». М., «Наука», 1966.
8. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
9. Ю. П. Адлер, И. Ф. Александрова, Ю. В. Грановский и др. Об одном методе формализации априорной информации при планировании эксперимента. В сб. «Планирование эксперимента». М., «Наука», 1966.
10. Ю. В. Грановский, Ю. П. Адлер, В. В. Налимов и др. Отсеивающие эксперименты при изучении разделения циркония и гафния экстракцией трибутилфосфатом. — Заводская лаборатория, 1963, 29, № 10.

О КЛАССИФИКАЦИИ ЭКСПЕРИМЕНТАЛЬНЫХ ПЛАНОВ

На правилах покоится игра.

Г. Гессе. Игра в бисер

Итак, мы рассмотрели подробно одну из задач планирования эксперимента. В действительности экспериментатор сталкивается с огромным разнообразием постановок задач и, соответственно, экспериментальных планов. Поэтому мы решили написать эту главу, цель которой связать наше изложение со всем множеством задач, объединенных общим названием «математическая теория эксперимента».

В зависимости от задачи исследования, свойств объекта, выполнения математических предпосылок, наличия априорной информации и т. д. можно выбрать тот или иной класс планов для получения необходимой информации [1, 2].

Создание единой системы классификации экспериментальных планов представляет собой сложную задачу. Оно связано с выявлением и отбором признаков, позволяющих проводить однозначную классификацию всего множества известных планов. На данном этапе в качестве предварительной классификации можно предложить систему, которая включает в себя следующие классы планов: 1) планы дисперсионного анализа; 2) планы отсеивающего эксперимента; 3) планы многофакторного анализа; 4) планы для изучения поверхности отклика; 5) планы для динамических задач планирования; 6) планы для изучения механизма явлений; 7) планы для построения диаграмм состав—свойство, состав—состояние. Такая классификация создается в последнее время в литературе по планированию эксперимента при изложении различных разделов этой теории. Разбиение планов на указанные группы проведено в основном по задачам исследования и методам планирования эксперимента, используемых для их решения. Естественно, что такая классификация довольно условна, но тем не менее может быть использована в качестве предварительной, чтобы помочь экспериментатору ориентироваться в различных разделах планирования эксперимента.

Заметим, что по методу анализа и виду математической модели, используемым при представлении результатов многофакторного эксперимента, все перечисленные классы планов можно объединить в три группы: 1) планы дисперсионного анализа;

2) планы регрессионного анализа; 3) планы ковариационного анализа.

Следуя Г. Шеффе [3], основные предпосылки указанных методов анализа при представлении результатов многофакторного эксперимента из N опытов можно записать в следующей форме: вектор выхода $y^{(N \times 1)}$ имеет распределение

$$\begin{aligned} y^{N \times 1} &\rightarrow N(\bar{x}^T \beta^{p \times 1}; \sigma^2 I) && \text{— в случае дисперсионного} \\ &&& \text{анализа;} \\ y^{1 \times 1} &\rightarrow N(z^T \gamma^{k \times 1}, \sigma^2 I) && \text{— в случае регрессионного} \\ &&& \text{анализа;} \\ y^{N \times 1} &\rightarrow N(\bar{x}^T \beta^{p \times 1} + z^T \gamma^{k \times 1}; \sigma^2 I) && \text{— в случае ковариационного} \\ &&& \text{анализа.} \end{aligned}$$

Здесь \bar{x}^T — транспонированная матрица независимых переменных x_{ij} , которые могут быть как количественными, так и качественными в задаче дисперсионного анализа; z^T — транспонированная матрица количественных переменных z_{ij} , пробегающих непрерывный ряд значений в задаче регрессионного анализа, а также матрица количественных и качественных неуправляемых переменных в задаче ковариационного анализа; $\beta^{(p \times 1)}$ — вектор эффектов (главные эффекты, эффекты взаимодействия, эффекты блоков и другие эффекты, например, эффекты порядка варьирования факторов, остаточные эффекты), подлежащих оценке по результатам эксперимента в задачах дисперсионного и ковариационного анализов; $\gamma^{(k \times 1)}$ — вектор коэффициентов регрессии в задачах регрессионного и ковариационного анализов; σ^2 — дисперсия ошибки эксперимента; I — единичная матрица; N — индекс нормального распределения.

В алгебраической форме уравнение модели, например для ковариационного анализа, имеет вид:

$$y = \sum_{i=1}^p b_i x_i + \sum_{j=1}^k \gamma_j z_j + \varepsilon,$$

где ε — случайная ошибка, относительно которой обычно постулируют $M(\varepsilon_i) = 0$; $M(\varepsilon_i \varepsilon_j) = 0$, т. е. ошибки некоррелированы и однородны.

Задача любого вида анализа заключается в установлении существенности эффектов исследуемых переменных на фоне этой ошибки.

Следует отметить, что хотя границы между перечисленными видами анализа не являются очень точными и общепринятыми, тем не менее для общности рассмотрение всех классов планов в рамках этих видов анализа полезно и целесообразно. Ниже дается характеристика каждого класса планов с указанием назначения планов, методов их построения, а также сведений о математической модели представления результатов эксперимента и ее анализе.

16.1. Планы дисперсионного анализа

Дисперсионный анализ — это статистический метод, с помощью которого производится разложение суммарной дисперсии на составляющие. В зависимости от числа источников дисперсии различают однофакторный и многофакторный дисперсионные анализы.

Если при постановке опытов реализуются все возможные совокупности условий, задаваемые выбранной схемой эксперимента, говорят о полных классификациях дисперсионного анализа. Если же реализуются не все возможные совокупности условий, а некоторая их часть, т. е. производится сокращение перебора вариантов, — речь идет о неполных классификациях дисперсионного анализа.

Сокращение перебора вариантов может производиться случайным образом или в соответствии с некоторыми строгими правилами. В первом случае это неполные классификации дисперсионного анализа без ограничения на рандомизацию, во втором случае — с ограничением на рандомизацию.

Полные классификации дисперсионного анализа применяются для исследования сравнительно небольшого числа факторов (обычно не более пяти), так как полный перебор вариантов требует постановки большого числа опытов, например, при варьировании пяти факторов на трех уровнях необходимо поставить 243 опыта. Число уровней может быть одинаковым для всех факторов (эксперимент типа n^m), но может быть и различным (эксперимент $n^m k^L \dots$).

Модель, с помощью которой представляются экспериментальные данные, имеет вид

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \dots + \alpha_i \beta_j + \alpha_i \gamma_k + \beta_j \gamma_k + \dots \\ \dots + \alpha_i \beta_j \gamma_k + \dots + \varepsilon_{ijk},$$

где y_{ijk} — результат эксперимента, полученный на i -м уровне первого фактора; j -м уровне второго фактора и k -м уровне третьего фактора; μ — среднее по всему множеству опытов; α_i — эффект i -го уровня первого фактора; β_j — эффект j -го уровня второго фактора; γ_k — эффект k -го уровня третьего фактора; $\alpha_i \beta_j$ — эффект взаимодействия i -го уровня первого фактора с j -м уровнем второго; $\alpha_i \gamma_k$ — эффект взаимодействия i -го уровня первого фактора с k -м уровнем третьего и т. д.; ε_{ijk} — остаточный член, с помощью которого оценивается ошибка эксперимента.

Среди неполных классификаций дисперсионного анализа с ограничением на рандомизацию наиболее популярными в планировании эксперимента являются неполноблочные планы (блок-схемы) и латинские планы. Различают полностью сбалансированные и частично сбалансированные блок-схемы.

Блок-схема, в которой пары элементов появляются определенное число раз, носит название сбалансированной неполной блок-схемы (BIB-схемы) * со следующими параметрами: v — число элементов; b — число блоков; k — число единиц в блоке; r — число блоков, которым принадлежит один и тот же элемент; λ — число повторений каждой пары элементов. Чтобы блок-схема была сбалансированной, требуется выполнение следующих условий: 1) каждый блок B_j содержит одинаковое число k элементов; 2) каждый элемент α_i принадлежит одному и тому же числу r блоков; 3) для каждой неупорядоченной пары α_i, α_j различных элементов число блоков, содержащих эту пару, равно λ .

Сбалансированное неполноблочное планирование может быть найдено для любого числа элементов v и любого размера блока k . Однако большинство BIB-схем не представляет интереса для планирования эксперимента, так как r велико. Обычно в планировании не применяются BIB-схемы с $r > 10$.

Частично сбалансированный план (частично сбалансированная блок-схема, сокращенно PBIB-схема) ** это план, в котором каждый блок содержит одно и то же число элементов и каждый элемент принадлежит одному и тому же числу блоков, но некоторые пары элементов принадлежат одному числу блоков λ_1 , в то время как другие пары — другому числу блоков λ_2 . В общем λ_1 и λ_2 могут быть любыми различными целыми числами, включая нуль. PBIB-схемы с параметрами λ_1 и λ_2 имеют два ассоциативных класса и сокращенно обозначаются PBIB(2)-схемы. В общем случае число классов равно m .

Цепные блок-схемы являются специальным видом блок-схем. Простейшие цепные блок-схемы построены таким образом, что пара элементов в двух соседних блоках одинакова и является связующим звеном. Такое связывание блоков является основным свойством цепных блок-схем. Связывающие звенья образуются из элементов первой группы, которые повторяются в плане два или большее число раз. Элементы второй группы встречаются в плане только один раз. Цепные блок-схемы целесообразно применять в следующих ситуациях: 1) размер блока ограничен и число элементов значительно превышает этот объем; 2) сравнение элементов внутри блоков проводится с такой точностью, что достаточно одного или двух повторений; 3) все элементы можно разбить на две группы.

Элементы первой группы являются наиболее важными, и сравнение пар этих элементов желательно проводить достаточно точно. Поэтому они повторяются минимум два раза. Элементы второй группы считаются дополнительными и встречаются в плане по одному разу.

* От английского Balanced Incomplete Block.

** От английского Partially Balanced Incomplete Block.

Цепные блок-схемы используются в экспериментальных исследованиях, проводимых с высокой точностью, например в физике.

Специфическим типом неполноблочных планов являются решетчатые планы. Они могут быть полностью или частично сбалансированы и иметь форму квадрата, прямоугольника или куба. Решетчатые планы различаются также по числу величин, которые балансируются. Так, например, план в форме квадрата может быть сбалансирован по одному фактору. В этом случае он имеет одно ограничение и носит название квадратной решетки. Но план в форме квадрата может быть сбалансирован и по двум факторам. Тогда он имеет два ограничения и называется решетчатым квадратом.

В m -мерных сбалансированных решетках, имеющих квадратную форму, число элементов равно k^m , где k есть простое целое число. Параметры этих планов связаны следующими соотношениями:

$$v = k^2, \quad r = k + 1, \quad b = k(k + 1).$$

Модель, с помощью которой представляются экспериментальные данные для планов, не разбитых на реплики, имеет вид

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

а для планов, разбитых на реплики, —

$$y_{ijq} = \mu + \alpha_i + \beta_j + \gamma_q + \varepsilon_{ijq}.$$

К латинским планам мы относим латинские и гипер-греко-латинские квадраты, кубы, прямоугольники, параллелепипеды, а также сложные планы, построенные на базе латинских планов. Латинские прямоугольники, к одной из разновидностей которых относятся квадраты Юдена, имеют «двойное подчинение»: по методу построения они связаны с латинскими квадратами (их можно построить вычеркиванием определенных строк или столбцов латинских квадратов, поэтому они еще называются неполными латинскими квадратами), а по свойствам и по методам статистического анализа они близки к блок-схемам.

Латинским квадратом называется квадратная таблица из элементов (чисел или букв), такая, что каждый элемент встречается один и только один раз в каждой строке и в каждом столбце. При планировании эксперимента строчки и столбцы квадрата употребляются для обозначения уровней двух факторов, образующих факторный эксперимент типа n^2 . На него накладывается $n \times n$ латинский квадрат. Латинский квадрат является частью плана, однако в планировании эксперимента весь план принято называть латинским квадратом (рис. 39).

Результаты эксперимента представляются в виде линейной модели

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}.$$

Главными эффектами являются α_i и β_j . Они «элиминируются» группировкой элементов квадрата (γ_k).

Латинский квадрат не является обычной моделью с предположением нормальности, по которой ошибки независимы и их дисперсии равны. Статистический анализ существенно опирается на предположение аддитивности и может быть ошибочным, когда есть взаимодействия.

Два латинских квадрата называются ортогональными, если при наложении одного квадрата на другой каждая пара одинаковых элементов встречается один и только один раз. Комбинация двух ортогональных квадратов носит название латинского квадрата второго порядка. Если элементы первого квадрата обозначить латинскими буквами, а второго — греческими, то такой квадрат называется греко-латинским. Модель эксперимента имеет вид

$$y_{ijk0} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_0 + \varepsilon_{ijk0}.$$

Три ортогональных латинских квадрата образуют латинский квадрат третьего порядка или гипер-греко-латинский квадрат. Модель эксперимента

$$y_{ijk0p} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_0 + \zeta_p + \varepsilon_{ijk0p}.$$

Если имеется k ортогональных латинских квадратов, то они образуют латинский квадрат k -го порядка. Пусть дано множество S из n элементов. Латинским прямоугольником, основанном на n -множестве S , называется такая прямоугольная $r \times s$ -таблица, в которой каждая строка является s -перестановкой элементов S , а каждый столбец — r -перестановкой элементов S при $r \leq n, s \leq n$.

Квадратами Юдена называются такие латинские прямоугольники, для которых выполняются следующие соотношения: $v = b, k = r$. Модель эксперимента

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}.$$

Латинским кубом первого порядка размера n называют кубическую таблицу из n элементов, расположенных в n^3 позициях, такую, что каждый элемент входит в таблицу n^2 раз и встречается в каждой из n плоскостей, параллельных координатным плоскостям X_1OX_2 , X_1OX_3 и X_2OX_3 одинаковое для всех элементов число раз, равное n .

Если некоторый латинский куб размера n первого порядка можно наложить на другой латинский куб размера n первого по-

$\begin{matrix} \backslash \beta \\ A \end{matrix}$	β_1	β_2	β_3	β_4
α_1	ζ_2	ζ_3	ζ_4	ζ_1
α_2	ζ_1	ζ_2	ζ_3	ζ_4
α_3	ζ_3	ζ_4	ζ_1	ζ_2
α_4	ζ_4	ζ_1	ζ_2	ζ_3

Рис. 39. Латинский квадрат 4×4

рядка так, что каждый элемент одного куба встретится точно n раз с каждым элементом другого куба, то такие два куба, называют ортогональными. Два ортогональных куба, наложенных друг на друга, образуют греко-латинский куб первого порядка.

Латинский куб 2-го порядка представляет собой разноразмерный план: часть факторов имеет n уровней, а один фактор — n^2 уровней. Все линейные эффекты факторов, установленных на n уровнях, определяются с одинаковой и максимальной для данного числа N точностью, а эффект фактора на n^2 уровнях определяется с точностью в n раз меньшей.

Модель эксперимента при применении латинских кубов выглядит следующим образом:

$$y_{ijk\ldots o} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_o + \varepsilon_{ijk\ldots o},$$

где $\alpha_i, \beta_j, \gamma_k$ являются главными эффектами и «элиминируются» группировкой латинского куба (δ_o).

Разноразмерный план, у которого все факторы имеют n уровней, а один фактор — $(n - r)$ -уровней, называется латинским параллелипипедом.

Планы, построенные путем совмещения латинских квадратов или прямоугольников с факторными экспериментами типа 2^n , называются сложными совмещенными планами. При совмещении число опытов в матрице должно равняться числу клеток квадрата или прямоугольника. Для $n = 2^k$, где $k = 1, 2, \ldots$, частью сложного плана является латинский квадрат. Если же n является нечетным числом, то частью сложного плана является латинский прямоугольник $\frac{2^k}{2} \times 2^k$. При совмещении фактор, уровни которого составляют латинский план, ортогонален $2k$ факторам, задающим полный факторный эксперимент. Все 2^k уровней этого фактора встречаются с любым из двух уровней исходных $2k$ факторов одинаковое число раз.

Сложные планы пригодны для линейных моделей, когда $y = F(x_1, x_2, \ldots, x_n; z_1, z_2, \ldots, z_m) + \varepsilon$, где x_i — количественные факторы на двух уровнях, $i = 1, 2, \ldots, n$ и z_j — качественные факторы на числе уровней $m > 2$, $j = 1, 2, \ldots, m$.

Сложные планы позволяют: 1) варьировать количественные факторы только на двух уровнях, что является достаточным для получения линейной зависимости, когда справедлива гипотеза об отсутствии взаимодействий; 2) исключить нарушающее влияние качественных факторов при подсчете линейных эффектов количественных факторов; 3) совершить движение по градиенту для количественных факторов; 4) построить оптимальный перебор комбинаций уровней качественных факторов, если $n > 1$; 5) не превысить число опытов по сравнению с факторными планами типа 2^n .

Сравнительно редко употребляются перекрестные планы (*crossover design*), построенные на базе латинских квадратов малого размера, и переменные планы (*change-over design*), имеющие черты, связывающие их с цепными блок-схемами и с перекрестными планами [4].

16.2. Планы многофакторного анализа

Планы многофакторного анализа (ПМА) используются для оценки линейных эффектов и эффектов взаимодействий многих факторов, варьируемых на одинаковом (симметричные планы) или неодинаковом (несимметричные планы) числе уровней. Общее уравнение математической модели, представляющей результаты эксперимента по таким планам, имеет вид

$$y = b_0 + \sum_{i=1}^p b_i x_i + \sum_{i>j}^p \sum_{j>u}^p b_{ju} x_j x_u + \ldots + \sum_{j>u>\ldots>k}^p \ldots \sum_{i>j>u>\ldots>k}^p b_{ju\ldots k} x_j x_u \ldots x_k.$$

В классе ПМА различаются следующие планы.

Симметричные двухуровневые планы типа 2^k , в которых каждый фактор варьируется на двух уровнях. Эти планы наиболее освоены и распространены. Полные планы, когда реализуется весь набор из 2^k комбинаций уровней факторов, позволяют оценить все эффекты модели, включая взаимодействия. Если часть эффектов (обычно это взаимодействия высших порядков) в модели отсутствует, то используют дробные факторные планы или так называемые дробные реплики, представляющие определенную часть полного эксперимента типа 2^k . Реплика называется регулярной, если она представляет $1/2^k$ часть от 2^k эксперимента, где k — целое число. Такая реплика содержит 2^{k-p} опытов и обозначается соответствующим образом. Если реализуется часть 2^k эксперимента, не кратная степени двойки, т. е. не равная $1/2^k$, то такая реплика называется нерегулярной.

Многоуровневые симметричные планы, в которых факторы варьируются на 3, 4, \ldots , t уровнях и обозначаются соответственно как $3^k, 4^k, \ldots, t^k$ планы. Эти планы также могут быть полными и дробными.

Многоуровневые несимметричные планы, в которых факторы варьируются на различных уровнях, строятся различными способами: а) комбинированием полных и дробных планов типа $2^k, 3^k, 4^k$, т. е. планы типа $3 \times 2^k, 4 \times 2^k, 5 \times 2^k, 3 \times 2^k$ и т. д.; б) совмещением факторных планов типа 2^k с латинскими квадратами, прямоугольниками, кубами (особенностью этих планов является требование равенства числа опытов эксперимента степени двойки, т. е. $N = 2^k$); в) комбинированными методами на основе FВІВ- и VІВ-схем; г) методом теории конечных полей; д) методом

преобразования симметричных планов в соответствующие несимметричные.

Кроме перечисленных планов следует отметить планы, обеспечивающие оценку не только исследуемых факторов и их взаимодействий, но также и эффекты последовательности воздействия факторов. При использовании планов каждый объект испытывается всеми k факторами, варьируемыми на двух уровнях (1 — наличие фактора, 0 — его отсутствие) в определенной последовательности и измерение выходного показателя производится после каждого воздействия. Получаемая модель имеет вид $y_{ijk} [x_1, x_2, \dots, x_{j-1}] = m + u_{ik} + \varepsilon_{x_1 \dots x_{j-1}} + \xi_{ijk}$, где $i=1, \dots, r$ — число объектов, которые подвергались воздействию только одной определенной последовательностью $j=1, \dots, (p+1)$, j — число замеров на объекте, $k=1, \dots, p!$ — число возможных сочетаний последовательностей, m — оценка средней, u_{ik} — конечный эффект воздействия на объект факторов, приложенных в определенной k -й последовательности, ξ — эффект последовательности приложения факторов, ε — случайная ошибка измерений.

К этому же классу можно отнести также планы для оценки остаточных эффектов, перекрестные (*cross-over*) планы, планы с группировкой и планы с расщепленными деланками, которые обычно рассматриваются в группе планов дисперсионного анализа [4—7]

16.3. Планы для изучения поверхности отклика

Этот класс планов применяется для детального изучения области оптимума и участков поверхности отклика со значительной кривизной, где линейная модель становится неадекватной. Обычно для математического описания бывает достаточно полинома второго порядка, реже третьего порядка, используя планы соответственно второго и третьего порядков.

Планы 2-го порядка* позволяют получить математическое описание в виде полной квадратичной модели, содержащей кроме основных эффектов b_j все парные взаимодействия b_{ju} и квадратичные эффекты b_{jj} , т. е. всего $\frac{(k+1)(k+2)}{2}$ эффектов;

$$y = b_0 + \sum_{j=1}^k b_j x_j + \sum_{\substack{j,u \\ j \neq u}}^k b_{ju} x_j x_u + \sum_{j=1}^k b_{jj} x_j^2.$$

По методу построения планы бывают композиционные и некомпозиционные.

* Более детальная классификация планов второго порядка производится с учетом критериев оптимальности, рассмотренных нами в гл. 10 (например, A -оптимальность, D -оптимальность и т. д.).

Композиционные планы второго порядка получают путем добавления $2k$ «звездных точек» типа $(\pm \alpha, 0, \dots, 0)$, $(0, \pm \alpha, 0, \dots, 0)$, \dots , $(0, 0, \dots, \pm \alpha)$ и некоторого числа центральных точек $n_0 (0, 0, \dots, 0)$ к «ядру», образованному полным экспериментом типа 2^k или его репликой. Используют эти планы обычно на заключительном этапе исследования: при описании экспериментальной области в ситуациях, когда отсутствует априорная информация об объекте и его полиномиальную модель приходится подбирать последовательно, начиная с простейшего линейного уравнения, которое затем достраивается до полной квадратичной модели. В таких случаях применение композиционных планов оказывается наиболее выгодным по числу опытов.

Выбор величины плеча α и числа точек n_0 определяется критерием оптимальности (ортогональность, ротатабельность). При построении ортогональных и ротатабельных центральных композиционных планов (ОЦКП, РЦКП) в качестве ядра используют минимально возможные регулярные реплики 2^{k-p} , которые обеспечивают независимую раздельную оценку всех основных эффектов b_j и эффектов взаимодействия b_{ju} .

Если известно априори, что часть эффектов b_j (или b_{ju}) в модели отсутствует, то используют планы Хартли, строящиеся на основе минимальных регулярных реплик 2^{k-p} , в которых должны быть не смешаны между собой только парные эффекты b_{ju} (часть или даже все эффекты b_{ju} могут быть смешаны с эффектами b_j). Планы Хартли более экономны, чем ОЦКП и РЦКП, и рекомендуются при построении интерполяционных моделей типа [квадратичного полинома для объектов с малым уровнем шумов.

Для этих же целей можно использовать композиционные планы Вейстлейка с еще меньшим числом точек, построенные на основе минимально возможных нерегулярных реплик. Недостаток планов Вейстлейка — коррелированность эффектов квадратичной модели, что затрудняет их расчет и интерпретацию.

Некомпозиционные планы применяются при наличии априорной информации о существенности кривизны поверхности отклика, позволяющей начинать эксперимент сразу с реализации плана второго порядка. К их числу относятся планы типа неполного факторного эксперимента 3^k , симплексно-суммируемые планы и прочие. Первые представляют собой определенные выборки строк из матрицы 3^k . В основу их построения положен принцип комбинации матриц 2^k по сбалансированной схеме неполных блоков.

Симплексно-суммируемые планы делятся на симметричные и несимметричные в зависимости от вида суммирования. Эти планы можно также рассматривать как композиционные, потому что они получаются на базе симплекс-планов первого порядка соответствующей достройкой.

Планы третьего порядка применяются сравнительно редко, лишь в случаях, когда квадратичная модель неадекватна. Обычно

они строятся последовательно на базе планов первого и второго порядка.

Существуют также отдельные планы непоследовательного типа [5].

16.4. Планы отсеивающего эксперимента

Эти планы используются на стадии предварительных исследований для выделения существенных эффектов факторов. В этом классе различают следующие планы*.

Насыщенные планы Плакетта и Бермана, представляющие собой двухуровневые планы, образованные методом циклических сдвигов. Число опытов в планах равно числу исследуемых эффектов.

Сверхнасыщенные планы случайного баланса в зависимости от числа уровней варьирования могут быть двух- и многоуровневыми. По методу построения эти планы могут быть образованы случайным образом (например, из строк факторного эксперимента типа 2^k с помощью таблиц случайных чисел). Систематически отобранные планы обеспечивают минимальные корреляции между столбцами плана (планы Бут и Кокса и др.). При случайном балансе результаты эксперимента представляются в виде модели

$$y = b_0 + \sum_{j=1}^p b_j x_j + \sum_{j=p+1}^l a_j z_j + \varepsilon,$$

где p — число значимых эффектов, $l-p$ число отсеиваемых незначимых эффектов, ε — случайная ошибка, $N < l$ (при $p < N$) — число опытов плана.

Планы последовательного отсеивания. При последовательном отсеивании, используемом в отличие от двух вышеописанных типов планов, для задач большой размерности (число факторов до 100 и выше) все факторы на основе априорной информации делятся на группы, каждая из которых рассматривается далее как отдельный комплексный фактор. Эти группы — комплексные факторы, которые содержат только незначимые переменные, исключаются из рассмотрения после первого цикла опытов (первой проверки). Оставшиеся факторы вновь делятся на группы для проверки, и цикл опытов повторяется. Такая процедура проводится до выявления всех значимых эффектов. В процессе отсеивания комбинирование и разделение переменных по группам проводится с помощью комбинаторных планов типа ВІВ-РВІВ-схем, латинских квадратов и др.

После каждого цикла опытов получается новая информация, позволяющая выбрать оптимальные планы для реализации очеред-

ного цикла. В простейшем случае при последовательном отсеивании используется линейная модель аддитивного типа

$$E(y_{ij}) = E(m + g_{ij} + \varepsilon) = E(m + v_{j-1, s_{i+1}} + v_{j-1, s_{i+2}} + \dots + v_{j-1, s_{i+k}} + \dots + v_{j, s_{i+c}}) = m + v_{j-1, s_i} + \varepsilon,$$

где y_{ij} — результат эксперимента в j -й подгруппе и i -м цикле c -циклового эксперимента, m — общее среднее, ε — ошибка эксперимента, g_{ij} — комбинированный эффект от s_i переменных, включенных в модель. Предполагается, что только k -й из этих s_i переменных в подгруппе g_{ij} значим, т. е. имеется эффект $v_{i-1, s_{i+k}}$, который больше ошибки ε или комбинации всех остальных $s_i - 1$ незначимых переменных. Если это так, то вся группа g_{ij} должна быть полностью пересмотрена в следующем $(i+1)$ -м цикле. Если $E(y_{ij}) = E(m + g_{ip} + \varepsilon)$, то подгруппа g_{ip} , содержащая незначимые переменные, исключается из рассмотрения и далее следует $(i+1)$ -й цикл.

В более сложных случаях в модель включают эффекты взаимодействия факторов [5,6].

16.5. Планы для экспериментирования в условиях дрейфа

Блочные планы, ортогональные к дискретному дрейфу, представляют собой обычные планы типа ПМА, сбалансированные так, чтобы часть столбцов плана использовалась для оценки эффектов дискретного дрейфа независимо от эффектов исследуемых факторов.

Планы, ортогональные к непрерывному дрейфу, могут быть построены на основе таблиц полиномов Чебышева. Они используются для изучения линейных эффектов управляемых количественных факторов независимо от полиномиального дрейфа любого порядка. В случае необходимости оценки также и взаимодействий управляемых факторов используют обычные планы 2^k , отбирая те столбцы планов, которые имеют минимальные корреляции с эффектами дрейфа. К этим же планам относятся планы Кокса, предназначенные для изучения одной количественной или качественной переменной, варьируемой на двух, трех, четырех уровнях в условиях дрейфа второго и третьего порядков.

Комбинированные планы для совместного изучения количественных и качественных переменных в условиях непрерывного полиномиального дрейфа получают соответствующим комбинированием планов Чебышева и планов Кокса.

Планы для экспериментирования в условиях дрейфа используются для исключения влияния неоднородностей типа дискретного и непрерывного дрейфа на исследуемые эффекты и оценки этого влияния независимо от эффектов варьируемых факторов и составляют основу группы планов ковариационного анализа [4].

* Планы отсеивающего эксперимента в задачах отыскания биологически активных препаратов здесь не рассматриваются.

16.6. Планирование эксперимента на диаграммах состав—свойство

Специфика задачи состоит в том, что естественная координатная система — барицентрическая на симплексе. Ортогональный план построить нельзя.

Первая возможность преодоления этой трудности — преобразование системы координат в декартову размерности $k-1$, где k — число компонентов системы (факторов), и использование любых известных планов для описания области или оптимизации.

Вторая возможность — планирование на симплексе. Для аппроксимации гладких поверхностей полиномом заданной степени используются симплексные решетки Шеффе. Они задаются симметричной системой точек, число которых определяется степенью полинома. Свободный член и коэффициенты при степенях переменных (выше первой) нельзя оценить такой выборкой (альтернатива — использовать однородные полиномы, не содержащие оценок взаимодействий).

Если на симплекс наложены линейные ограничения, то необходимо строить план на произвольном выпуклом многограннике. Для этого предложены планы Мак-Лина—Андерсона, которые задаются множеством точек, лежащих в вершинах, серединах ребер и гранях многогранника.

Для включения в план дополнительных факторов применяются симплексно-центроидные планы. Возможно также совмещение симплексных решеток с факторными планами, латинскими квадратами и т. п., а также с репликами от этих планов.

Когда существует опасность смещения оценок из-за выбора низкой степени полинома, используются планы Дрейпера—Лоуренса, минимизирующие такое смещение.

Известны модификации планов для специальных видов аппроксимирующих полиномов, например для тригонометрических полиномов.

В случае, когда поверхность отклика определяется разрывной функцией, возможна триангуляция симплекса и планирование в каждом новом многограннике [8].

16.7. Динамические задачи планирования

Можно различить два типа задач планирования: сводимые к статистическим и несводимые к статистическим (собственно динамические).

В задачах, сводимых к статистическим, выше уровень шумов, поэтому требуется большее число параллельных опытов.

Объекты исследования классифицируют по характеру дрейфа. Объекты без дрейфа — вообще не динамические. Если параметры дрейфа не известны, то эксперимент должен их оценить, одновременно с решением основной задачи.

Особенности объектов приводят к использованию преимущественно последовательных планов.

Наиболее развиты планы для промышленных экспериментов. К ним относится прежде всего эволюционное планирование Бокса (ЭВОП), которое строится на базе регулярных реплик от факторных планов и метода крутого восхождения, но с многократными повторами плана и осторожным движением в сторону градиента. Известны многочисленные модификации ЭВОП, например, вращаемое ЭВОП, позволяющее проводить локацию факторного пространства с расширяющейся сферой обследования. В планах ЭВОП используются линейные модели.

Случай непрерывного варьирования рассмотрен только для одного и двух факторов. Предложены планы Бокса—Дженкинса, основанные на модулировании входных сигналов синусоидами, ортогональными друг к другу.

В условиях автоматического управления объектом можно использовать планы адаптационной оптимизации, основанные, например, на симплекс-процедуре (отражении симплекса относительно грани, противоположной к вершине с наихудшим результатом).

Когда неуправляемых переменных несколько и их действие нельзя интерпретировать как дрейф, возникает задача активно-пассивного эксперимента (часть факторов образует план, а часть — измеряется; обработка результатов совместная).

Иногда, обычно в очень сложных ситуациях, эффективны рандомизированные последовательности опытов, т. е. планы, основанные на методах случайного поиска.

Планы для решения динамических задач — одна из слабо разработанных областей математической теории эксперимента, поэтому пока можно говорить скорее не о классификации, а о типологии этих задач и методов их решения [9], [10].

16.8. Планы для изучения механизма явлений

По уровню априорной информации различаются следующие ситуации. Модель известна, константы известны; требуется уточнение (не обязательно всех) констант. Планы такого уточняющего эксперимента можно синтезировать для различных критериев оптимальности (обычно для D -оптимальности), причем как в однократном, так и в последовательном вариантах. Так как вид модели всякий раз новый, то и план надо строить заново.

Модель известна, требуется оценка констант. Это задача интерполяции для известной функции. Выбор критерия оптимальности обеспечивает синтез плана.

Известно несколько альтернативных моделей. Требуется сравнить их, выбрать наилучшую, в некотором смысле, и оценить ее константы. Такая задача приводит к планированию дискриминирующих экспериментов. План может быть синтезирован анало-

гично предыдущим случаям. Множества факторов в моделях должны пересекаться, а лучше — совпадать.

Модель не известна. Этот случай приводит к обычной задаче аппроксимации неизвестной функции полиномом. Используется любой подходящий план. На основании результатов выдвигаются, если удастся, содержательные гипотезы о механизме. Далее все, как выше.

Для одной частной задачи — задачи химической кинетики, известен план Бокса—Хантера, который, по существу, есть реплика от плана 2^k . В этом случае отклик для условий одного опыта измеряется последовательно во времени несколько раз, т. е. отклик не число, а функция. При обработке результатов могут использоваться как полиномы, так и содержательные модели (например, модели формальной кинетики для констант скорости реакции).

Последовательные планы характерны для задач этого класса [11].

Почти все планы §§ 16.2—16.4 относятся к планам регрессионного анализа, а сложные планы с качественными и количественными факторами относятся к планам ковариационного анализа.

Литература

1. А. Н. Лисенков, Е. В. Маркова, Ю. П. Адлер. О классификации экспериментальных планов Информационные материалы Совета по кибернетике. М., 1970, № 8 (45), с. 21.
2. А. Н. Лисенков. Основные принципы и методы планирования многофакторных медико-биологических экспериментов. — В сб. «Применение математических методов в медико-биологических исследованиях». ИППЭ АМН СССР, 1972, 20, с. 10.
3. Г. Шеффе. Дисперсионный анализ. М., Физматгиз, 1963.
4. Е. В. Маркова, А. Н. Лисенков. Планирование экспериментов в условиях неоднородностей. М., «Наука», 1973.
5. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. М., «Наука», 1965.
6. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
7. А. Н. Лисенков, Г. Х. Кодкин. Несимметричные факторные планы и методы их построения. Информационные материалы Совета по кибернетике. М., 1970, № 8 (45), с. 45.
8. Применение математических методов для исследования многокомпонентных систем. Сб. под ред. И. Г. Зедгинидзе и др. М., «Металлургия», 1974.
9. В. Г. Горский, Ю. П. Адлер. Планирование промышленных экспериментов. М., «Металлургия», 1974.
10. Д. Химмельблау. Анализ процессов статистическими методами. М., «Мир», 1974.
11. В. В. Федоров. Теория оптимального эксперимента. М., «Наука», 1971.

Глава семнадцатая

ОЧЕРК ПО ИСТОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

История... Читай и плачь.

Боконом

Планирование эксперимента — продукт нашего времени, однако истоки его теряются в глубине веков. Давайте попытаемся проследить за течением этих ручейков, слившихся сегодня в реку планирования эксперимента.

Существует известный произвол в определении того, относится ли то или иное событие или явление к истокам планирования эксперимента или нет. Поэтому отбор материала в известной мере определяется вкусами и информированностью авторов.

Понимая существующие трудности, начнем все же рассмотрение истории, ибо как иначе можно узнать о будущем?

17.1. Предыстория

Истоки планирования эксперимента уходят в глубокую древность и связаны с числовой мистикой, пророчествами и суевериями. Это собственно не планирование физического эксперимента, а планирование числового эксперимента, т. е. расположение чисел так, чтобы выполнялись некоторые строгие условия, например, на равенство сумм по строкам, столбцам и диагоналям квадратной таблицы, клеточки которой заполнены числами натурального ряда. Такие условия выполняются в магических квадратах, которым, по-видимому, принадлежит первенство в планировании эксперимента.

Согласно одной легенде примерно в 2200 г. до н. э. китайский император Ю выполнял мистические вычисления с помощью магического квадрата, который был изображен на панцире божественной черепахи.

Квадрат императора Ю		
4	9	2
3	5	7
8	1	6

Клетки этого квадрата заполнены числами от 1 до 9, и суммы чисел по строкам, столбцам и главным диагоналям равны 15.

В 1514 г. немецкий художник Альбрехт Дюрер изобразил магический квадрат в правом углу своей знаменитой гравюры

аллегии «Меланхолия». Два числа в нижнем горизонтальном ряду (15 и 14) составляют год создания гравюры. В этом состояло своеобразное «приложение» магического квадрата.

Квадрат Дюрера			
16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

В течение нескольких веков построение магических квадратов занимало умы индийских, арабских, немецких, французских математиков.

В настоящее время магические квадраты используются при планировании эксперимента в условиях линейного дрейфа, при планировании экономических расчетов и составлении рационов питания, в теории кодирования и т. д.

Построение магических квадратов является задачей комбинаторного анализа, основы которого в его современном понимании заложены Г. Лейбницем. Он не только рассмотрел и решил основные комбинаторные задачи, но и указал на большое практическое применение комбинаторного анализа: к кодированию и декодированию, к играм и статистике, к логике изобретений и логике геометрии, к военному искусству, грамматике, медицине, юриспруденции, технологии и к комбинации наблюдений. Последняя область применения наиболее близка к планированию эксперимента.

Г. Лейбниц проектировал создание «обобщенной математики», используя которую можно свести понятия к символам, символы к числам и благодаря этому подвергнуть понятие механическому вычислению. Основная роль при этом отводилась комбинаторике.

Этот проект во времена Лейбница казался абсурдным и фантастическим, но в настоящее время, благодаря развитию вычислительной техники и математической логики, часть этого плана уже реализуется. И в его реализации немалую роль сыграло планирование эксперимента.

Одной из комбинаторных задач, имеющей прямое отношение к планированию эксперимента, занимался известный петербургский математик Л. Эйлер. В 1779 г. он предложил задачу о 36 офицерах как некоторый математический курьез.

Он поставил вопрос, можно ли выбрать 36 офицеров 6 рангов из 6 полков по одному офицеру каждого ранга от каждого полка и расположить их в каре так, чтобы в каждом ряду и в каждой шеренге было бы по одному офицеру каждого ранга и по одному от каждого полка. Задача эквивалентна построению парных ортогональных 6×6 квадратов. Оказалось, что эту задачу решить невозможно. Эйлер высказал предположение, что не существует пары ортогональных квадратов порядка $n \equiv 2 \pmod{4}$.

Задачей Эйлера, в частности, и латинскими квадратами вообще занимались впоследствии многие математики, однако почти никто из них не задумывался над практическим применением латинских квадратов. Тем больший интерес могут представить работы немецкого математика Э. Шредера, который в конце XIX в. применил латинские квадраты в алгебре логики при разработке теории алгоритмов и исчислений.

Для того чтобы полностью выписать все формулы, входящие в какой-либо алгоритм, Шредер поставил задачу отыскать все возможные исчисления и получить их полный перебор. При этом ему стало ясно, что сделать полный перебор в большинстве случаев практически невозможно и нужно думать о том, как его уменьшить, как доказать, что данная система функциональных уравнений замкнута относительно логических следствий из них и т. д. Для сокращения перебора Шредер использовал латинские квадраты, которыми он занимался с большим увлечением. Эта работа использовалась Шредером для получения чисто логических результатов — для доказательства независимости закона дистрибутивности конъюнкции относительно дизъюнкции от системы более простых логических законов.

В 1900 г. П. Терри подтвердил путем систематического пересчета справедливость предположения Эйлера для $n=6$. Р. Боуз, С. Шрикханде и Е. Паркер доказали теорему о существовании пары ортогональных квадратов порядка $n \equiv 2 \pmod{4}$ для $n \neq 6$, и предположение Эйлера оказалось справедливым только для случая, когда $n=6$. Так эйлеровская задача о 36 офицерах способствовала интенсивному изучению теории латинских квадратов многими математиками в различных странах в течение последних 150 лет.

В настоящее время латинские квадраты являются одним из наиболее популярных способов ограничения на рандомизацию при наличии источников неоднородностей дискретного типа в планировании эксперимента. Группировка элементов латинского квадрата, благодаря своим свойствам (каждый элемент появляется один и только один раз в каждой строке и в каждом столбце квадрата), позволяет защитить главные эффекты от влияния источника неоднородностей. Широко используются латинские квадраты и как средство сокращения перебора в комбинаторных задачах.

Говоря об истоках планирования эксперимента с ограничением на рандомизацию, уместно вспомнить имя еще одного немецкого математика XIX в., положившего начало теории блок-схем или неполоблочных планов, как принято говорить в планировании эксперимента. Блок-схемы впервые изучались И. Штейнером в 1850 г. с точки зрения комбинаторики как тактические конфигурации. В планировании эксперимента блок-схемы стали использоваться значительно позже, о чем расскажем далее.

Мы остановились на предыстории планирования эксперимента, связанной с комбинаторным анализом. Обратимся теперь к другим областям знаний. В европейской философии основы методоло-

гии научных исследований закладывались Р. Декартом, Ф. Баконом и Дж. Миллем. Так, Милль в 1843 г., рассматривая причинные связи явлений, формулировал известные методы опытного исследования: метод сходства, метод различия, метод сопутствующих явлений и др. Эти методы представляли собой простейшие правила, которые помогали устанавливать взаимосвязи между причинами и следствиями (и наоборот). Теория Милля была основана на весьма сильных упрощающих предположениях (постулировался набор факторов, игнорировались взаимодействия между ними и случайные воздействия). При этом не использовался математический аппарат для количественных оценок.

В математике закладывались основы количественных методов исследования. Работа шла в нескольких направлениях. П. Ферма и Б. Паскаль в переписке обсуждали задачи теории вероятностей (XVII в.). Тейлор, Маклорен и Л. Эйлер (снова Эйлер!) (XVIII в.) разработали аппарат теории приближения функций степенными рядами. А. Коши (XIX в.) развил метод градиента для решения систем совместных уравнений. Ю. Кели создал матричные обозначения.

А. Лежандр и К. Гаусс создали и обосновали с вероятностных позиций метод наименьших квадратов. Позже все эти направления соединились в планировании эксперимента.

Обсуждая нашу тему, нельзя не коснуться положения дел в физике. Здесь появляются работы Дж. Максвелла — метод градиента проникает в физику. Дж. Гиббс вводит представление о поверхности отклика и вместе с Л. Больцманом утверждает, что природе свойственны скорее статистические, чем детерминированные закономерности.

Начало XX в. знаменуется созданием английской статистической школы (К. Пирсон), получением важных результатов Стьюдента (В. Госсета — английского статистика) по статистике малых выборок [1—4].

17.2. Становление

Возникновение современных статистических методов планирования эксперимента связано с именем Р. Фишера. С 1918 г. он начал свою известную серию работ на Рочемстедской агробиологической станции в Англии. В 1935 г. появилась его монография «Design of Experiments», давшая название всему направлению.

Среди методов планирования первым был дисперсионный анализ (кстати, Фишеру принадлежит и термин «дисперсия»). Фишер создал основы этого метода, описав полные классификации дисперсионного анализа (однофакторный и многофакторный эксперименты) и неполные классификации дисперсионного анализа без ограничения и с ограничением на рандомизацию. При этом он широко использовал латинские квадраты и блок-схемы. Вместе

с Ф. Йетсом он описал их статистические свойства. В 1942 г. А. Кипен рассмотрел планирование по латинским кубам, которое явилось дальнейшим развитием теории латинских квадратов. Затем Р. Фишер независимо опубликовал сведения об ортогональных гипер-греко-латинских кубах и гипер-кубах. Вскоре после этого (1946—1947 гг.) Р. Рао рассмотрел их комбинаторные свойства. Дальнейшему развитию теории латинских квадратов посвящены работы Х. Манна (1947—1950 гг.). Здесь интересно отметить, что начиная с работ Р. Муфанта (1935 г.) развивается теория квазигрупп, т. е. множеств E с бинарной операцией на E , для которой существуют обе обратные ей операции. Теория квазигрупп имеет прямую связь с теорией латинских квадратов. Так, если область определения конечна и задана списком ее элементов, то эта операция может быть эффективно определена квадратной таблицей с двумя входами, в каждом столбце и в каждой строке которой без повторений помещены все элементы из области E . Такая таблица является латинским квадратом.

Первое глубокое математическое исследование блок-схем выполнено Р. Боузом в 1939 г. Вначале была разработана теория сбалансированных неполноблочных планов (BIB-схем). Затем Р. Боуз, К. Нер и Р. Рао обобщили эти планы и разработали теорию частично сбалансированных неполноблочных планов (PBIB-схем). С тех пор изучению блок-схем уделяется большое внимание как со стороны специалистов по планированию эксперимента (Ф. Йетс, Г. Кокс, В. Кохрен, В. Федерер, К. Гульден, О. Кемпторн и многие др.), так и со стороны специалистов по комбинаторному анализу (Боуз, Ф. Шимамото, В. Клаисворси, С. Шрикханде, А. Гоффман и др.). Их усилия были направлены на разработку способов построения блок-схем, составления каталогов планов и на решение вопроса классификации PBIB-схем по ассоциативным схемам.

Исследования Р. Фишера, проводившиеся в связи с работами по агробиологии, знаменуют начало первого этапа развития методов планирования эксперимента. Фишер разработал метод факторного планирования. Йетс предложил для этого метода простую вычислительную схему. Факторное планирование получило широкое распространение. Особенностью полного факторного эксперимента является необходимость ставить сразу большое число опытов. Это вполне пригодно в агробиологии, но связано со значительными трудностями в технических приложениях.

В 1945 г. Д. Финни ввел дробные реплики от факторного эксперимента. Это позволило резко сократить число опытов и открыло дорогу техническим приложениям планирования. Другая возможность сокращения необходимого числа опытов была показана в 1946 г. Р. Плакеттом и Д. Берманом, которые ввели насыщенные факторные планы.

Г. Хотеллинг в 1941 г. обратил внимание на то, что многие практические, особенно технические, задачи можно рассматривать как

экстремальные. Он предложил находить экстремум по экспериментальным данным с использованием степенных разложений и градиента (подробно для случая одной независимой переменной). Проблемы, поставленные второй мировой войной, привели к интенсификации исследований в этом направлении.

Следующим важным этапом было введение принципа последовательного шагового экспериментирования. Этот принцип, высказанный в 1947 г. М. Фридманом и Л. Сэвиджем *, позволил распространить на экспериментальное определение экстремума известный в математике прием — итерацию. Идейно их работа примыкает к работам А. Вальда по последовательному анализу, сформировавшемуся в 1943—1950 гг.

Чтобы построить современную теорию планирования эксперимента, не хватало одного звена — формализации объекта исследования. Это звено появилось в 1947 г. после создания Н. Винером теории кибернетики. Кибернетическое понятие «черный ящик», играет в планировании важную роль. Кроме того, кибернетика стимулировала развитие вычислительной техники, дав в руки исследователей инструмент, позволяющий быстро решать возникающие при планировании математические задачи [3, 5].

17.3. Развитие

В 1951 г. работой американских ученых Дж. Бокса и К. Уилсона начался новый этап развития планирования эксперимента. Эта работа подытожила предыдущие. В ней ясно сформулирована и доведена до практических рекомендаций идея последовательного экспериментального определения оптимальных условий проведения процессов с использованием оценки коэффициентов степенных разложений методом наименьших квадратов, движения по градиенту и отыскания интерполяционного полинома (степенного ряда) в области экстремума функции отклика («почти стационарной» области).

В 1954—1955 гг. Дж. Бокс, а затем Дж. Бокс и П. Юл показали, что планирование эксперимента можно использовать при исследовании физико-химических механизмов процессов, если априори высказаны одна или несколько возможных гипотез. Здесь планирование эксперимента пересекалось с исследованиями по химической кинетике. Интересно отметить, что кинетику можно рассматривать как метод описания процесса с помощью дифференциальных уравнений, традиции которого восходят к И. Ньютону. Описание процесса дифференциальными уравнениями, называемое детерминистическим, нередко противопоставляется статистическим моделям. Указанные работы наметили конкретный путь сближения этих подходов. Если кинетическая модель уже известна, возникает вопрос о планировании эксперимента для уточнения полученных ра-

нее констант. Это направление также получило развитие в последние годы, в частности в работах Н. П. Клепикова и С. Н. Соколова, а затем В. В. Федорова в связи с задачами ядерной физики.

Если для первого этапа развития планирования эксперимента характерны полевые и лабораторные агробиологические исследования, то второй этап — это этап лабораторных, главным образом химических, исследований. Хотя развитие этого направления еще не закончено, но уже можно говорить о третьем этапе — этапе промышленных экспериментов. Он начался в 1957 г., когда Бокс модифицировал свой метод, приспособив его для использования в промышленности. Этот метод стал называться «эволюционным планированием».

Почти одновременно с эволюционным планированием Бокс и Дж. Хантер сформулировали принцип ротатбельности для описания «почти стационарной» области, развивающейся в настоящее время в важную ветвь теории планирования эксперимента. В той же работе показана возможность планирования с разбиением на ортогональные блоки, указанная ранее независимо де Бауном. Дальнейшим развитием этой идеи было планирование, ортогональное к неконтролируемому временному дрейфу, которое следует рассматривать как важное открытие в экспериментальной технике — значительное увеличение возможностей экспериментатора.

Слабым местом было отсутствие метода оценки риска, связанного с тем, что в рассмотрение не включен какой-либо важный фактор. Для преодоления этой трудности в 1957—1959 гг. Ф. Сатерзвайтом был развит метод «случайного баланса» *. Этот метод возник не в связи с общим ходом развития статистических идей, а скорее как прием формализации психо-физиологических методов решения сложных задач человеком. Поэтому он вызвал острую дискуссию.

В 1958 г. Г. Шеффе предложил новый метод планирования эксперимента для изучения физико-химических диаграмм состав—свойство. В этом методе, названном методом «симплексной решетки», используется то обстоятельство, что состав многокомпонентной системы дается точкой в правильном симплексе. Здесь произошло пересечение с химической топологией, созданной Н. С. Курнаковым [6].

Симплексы оказались пригодными и для построения ротатбельных планов, что было показано в 1960 г. Дж. Боксом и Д. Бенкиным. С. Адельманом в 1961 г. обобщена идея Финни о дробных репликах. Им построены нерегулярные дробные реплики, которые в ряде случаев позволяют существенно сократить число опытов.

С развитием средств автоматизации и вычислительной техники возник вопрос о возможности использования какого-либо алго-

* Очень кратко эта идея обсуждена также Г. Хотеллингом.

* Первый план со случайными уровнями факторов (для 30 факторов) был, вероятно, реализован Е. Широком в 1949 г. Первый такой эксперимент Ф. Сатерзвайт провел в 1952 г. (для 14 факторов), а теоретические основы развил в 1955—1956 гг.

ритма планирования при автоматическом управлении процессом с помощью вычислительной машины. Этот вопрос был теоретически решен в 1962 г. В. Спендлеем, Дж. Хекстом и Ф. Химсворстом, Боксом и Дж. Дженкинсом, создавшими основы теории адаптационной оптимизации, в которой сочетаются идеи планирования и теория случайных функций.

Следует отметить, что ряд научных концепций, развивающихся параллельно с планированием эксперимента, оказывает влияние на его развитие: к их числу, кроме уже упоминавшихся, относятся теория принятия решений, теория игр, линейное, нелинейное и динамическое программирование, правдоподобные рассуждения, введенные Д. Поля. Планирование эксперимента входит как составная часть в более общие концепции: исследование операций и системотехнику, теорию управляющих систем.

Современному состоянию планирования эксперимента и развитию этого направления в СССР посвящены две статьи в Информационных материалах научного совета по комплексной проблеме «Кибернетика» № 10 1969 г. Это — статья В. В. Налимова «Решенные и нерешенные проблемы планирования эксперимента» и статья Ю. П. Адлера, Ю. В. Грановского, Е. В. Марковой «Некоторые аспекты развития исследований по планированию эксперимента».

Отметим, однако, что наиболее характерной чертой, пожалуй, является объединение подхода Дж. Бокса с подходом американского математика Дж. Кифера, рассматривавшего формальные аспекты теории планирования. Значительные усилия для такого объединения делаются в СССР под руководством В. В. Налимова.

Растут возможности методов планирования эксперимента, расширяется сфера их приложения. Надо полагать, что развитие уже достигло такого состояния, когда можно писать с большой буквы — «Математическая теория эксперимента», когда накоплен значительный практический опыт.

Л и т е р а т у р а

1. Ю. П. Адлер, Е. В. Маркова. Планирование эксперимента в историческом аспекте. Информационные материалы Совета по «Кибернетике». М., 1970, № 8 (45), с. 13.
2. М. Холл. Блок-схемы. В сб. «Прикладная комбинаторная математика», под ред. Э. Беккенбаха. М., «Мир», 1968.
3. Е. В. Маркова. Неполноблочные планы. Препринт № 15. М., Изд-во МГУ, 1970.
4. Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. О принятии решений в неформализованных ситуациях. В сб. «Методологические проблемы кибернетики» (материалы к Всесоюзной конференции). т. 2. М., Совет по Кибернетике, 1970.
5. Ю. П. Адлер. Введение в планирование эксперимента. М., «Металлургия», 1969.
6. Применение математических методов для исследования многокомпонентных систем. Сб. под ред. И. Г. Зедгенидзе и др. М., «Металлургия», 1974.

ОГЛАВЛЕНИЕ

Предисловие ко второму изданию	5
Предисловие к первому изданию	7
Цели книги	9
Ограничения	10
Введение	11
Глава первая. Основные определения	14
Глава вторая. Параметр оптимизации	20
2.1. Виды параметров оптимизации (20). 2.2. Требования к параметру оптимизации (23). 2.3. О задачах с несколькими выходными параметрами (28). 2.4. Резюме (30)	
Глава третья. Обобщенный параметр оптимизации	32
3.1. Простейшие способы построения обобщенного отклика (32). 3.2. Шкала желательности (36). 3.3. Преобразование частных откликов в частные функции желательности (39). 3.4. Обобщенная функция желательности (41). 3.5. Резюме (45)	
Глава четвертая. Факторы	47
4.1. Определение фактора (48). 4.2. Требования, предъявляемые к факторам при планировании эксперимента (51). 4.3. Требования к совокупности факторов (54). 4.4. Примеры факторов (54). 4.5. Резюме (56)	
Глава пятая. Выбор модели	58
5.1. Шаговый принцип (60). 5.2. Как выбрать модель? (64). 5.3. Полиномиальные модели (65). 5.4. Резюме (67)	
Глава шестая. Полный факторный эксперимент	69
6.1. Принятие решения перед планированием эксперимента (69). 6.2. Полный факторный эксперимент типа 2^k (80). 6.3. Свойства полного факторного эксперимента типа 2^k (84). 6.4. Полный факторный эксперимент и математическая модель (85). 6.5. Резюме (91)	
Глава седьмая. Дробный факторный эксперимент	93
7.1. Минимизация числа опытов (93). 7.2. Дробная реплика (95). 7.3. Выбор полуреplik. Генерирующие соотношения и опреде-	

ляющие контрасты (96). 7.4. Выбор 1/4-реплик. Обобщающий определяющий контраст (102). 7.5. Реплики большой дробности (106). 7.6. Резюме (111)

Глава восьмая. Проведение эксперимента 113

8.1. Анкета для сбора априорной информации (113). 8.2. Реализация плана эксперимента (115). 8.3. Ошибки параллельных опытов (122). 8.4. Дисперсия параметра оптимизации (127). 8.5. Проверка однородности дисперсий (129). 8.6. Рандомизация (133). 8.7. Разбиение матрицы типа 2^k на блоки (136). 8.8. Резюме (140)

Глава девятая. Обработка результатов эксперимента 141

9.1. Метод наименьших квадратов (141). 9.2. Регрессионный анализ (148). 9.3. Проверка адекватности модели (149). 9.4. Проверка значимости коэффициентов (153). 9.5. Резюме (154)

Глава десятая. Матричный подход к регрессионному анализу . . . 156

10.1. Метод наименьших квадратов для одного фактора (156). 10.2. Некоторые операции над матрицами (158). 10.3. Обобщение метода наименьших квадратов на многофакторный линейный случай (164). 10.4. Статистический анализ (167). 10.5. Взвешенный метод наименьших квадратов и статистический анализ (172). 10.6. Критерии оптимальности планов (184)

Глава одиннадцатая. Принятие решений после построения модели 191

11.1. Интерпретация результатов (191). 11.2. Принятие решений после построения модели процесса (195). 11.3. Построение интерполяционной формулы. Линейная модель неадекватна (203). 11.4. Резюме (204)

Глава двенадцатая. Крутое восхождение по поверхности отклика 207

12.1. Движение по градиенту (207). 12.2. Расчет крутого восхождения (209). 12.3. Реализация мысленных опытов (214). 12.4. Резюме (218)

Глава тринадцатая. Принятие решения после крутого восхождения 219

13.1. Крутое восхождение эффективно (219). 13.2. Крутое восхождение неэффективно (223). 13.3. Резюме (229)

Глава четырнадцатая. Обсуждение результатов 231

14.1. Чем кончается эксперимент? (231). 14.2. Перспективы (235). 14.3. Резюме (235)

Глава пятнадцатая. Как решать задачу оптимизации. Конкретный пример 237

15.1. Предпланирование эксперимента (237). 15.2. Выбор условий проведения опытов (241). 15.3. Выбор и реализация плана (первая серия) (244). 15.4. Обработка результатов эксперимента (246). 15.5. Интерпретация результатов. Принятие решений после построения модели (первая серия) (249). 15.6. Реализация плана (вторая серия) (250). 15.7. Интерпретация результатов. Принятие решений после построения модели (вторая серия) (252)

Глава шестнадцатая. О классификации экспериментальных планов 254

16.1. Планы дисперсионного анализа (256). 16.2. Планы многофакторного анализа (261). 16.3. Планы для изучения поверхности отклика (262). 16.4. Планы отсеивающего эксперимента (264). 16.5. Планы для экспериментирования в условиях дрейфов (265). 16.6. Планирование эксперимента на диаграммах состав — свойство (266). 16.7. Динамические задачи планирования (266). 16.8. Планы для изучения механизма явлений (267)

Глава семнадцатая. Очерк по истории планирования эксперимента 269

17.1. Предыстория (269). 17.2. Становление (272). 17.3. Развитие (274)