

Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА ПРИ ПОИСКЕ ОПТИМАЛЬНЫХ УСЛОВИЙ

Программированное
введение
в планирование
эксперимента

1 9 7 1

Издательство · Наука ·
М о с к в а

- 3 -

УДК 519.24.001.5:62-50

Планирование эксперимента
при поиске оптимальных ус-
ловий. А д л е р Ю. П. и др.
Изд-во «Наука», 1971.

Планирование эксперимен-
та — новая научная дисцип-
лина. Она применяется для
решения широкого круга за-
дач: построения интерполя-
ционных моделей, изучения
кинетики и механизма явле-
ний, оптимизации процессов
и др.

Наибольшее значение для
практики имеет оптимизация
процессов (планирование экст-
ремальных экспериментов). Это-
му направлению и посвящена
монография.

Она представляет собой ввод-
ный курс по планированию
эксперимента при поиске опти-
мальных условий, написан-
ный по методу программиро-
ванного обучения.

Рассчитана на научных и ин-
женерно-технических работни-
ков. Таблиц 47. Иллюстраций
33.

О т в е т с т в е н н ы е
р е д а к т о р ы

канд. пед. наук

А. Н. ЗАХАРОВ

канд. техн. наук В. Г. ГОРСКИЙ

1-8-3

БЗ №46—1970—№16

- 4 -

Предисловие

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 и закрепить 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 вре-

мя 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: какая 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, чьи замечания были весьма полезны.

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 неизвестными, 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.

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

- применить методы планирования эксперимента при оптимизации многофакторных процессов;
- дать определение параметру оптимизации;
- » » факторам, определяющим процесс;
- » » поверхности отклика;
- » » матрице планирования;
- » » условиям ортогональности и ротатбельности;
- » » полному факторному эксперименту 2^k и дробным репликам от него;

- дать определение уравнению регрессии;
- » » основным эффектам и эффектам взаимодействия;
- » » кратчайшему пути к оптимуму;
- выбрать параметр оптимизации;
- » » 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.

Таким образом читатель приобретет навыки, необходимые для практического использования планирования эксперимента в простейших случаях.

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 статические математические модели процессов, используя факторное планирование, регрессионный анализ и движение по градиенту.

Кроме того, мы будем предполагать, что:

- задача допускает выбор одного параметра оптимизации,
- 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.

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

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*. Понадобилось еще несколько десятилетий, чтобы в начале пятидесятых 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**. 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 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 эффективный 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.

* Очерк истории развития планирования эксперимента вы сможете найти, например, в работе: Ю. П. Адлер, Е. В. Маркова. Планирование эксперимента в историческом аспекте. Информационные материалы научного Совета по кибернетике. Изд-во «Наука», М., 1970.

* R. A. Fisher. The Design of Experiments. 6-th ed, London, Oliver and Boyd, 1951.

** G. E. P. Box, K. B. Wilson. On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society, Series B, 1951, 13, № 1.

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. 9). 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., изд-во «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. Адлер «Введение в планирование эксперимента», М., изд-во «Металлургия», 1969). Для более подготовленного читателя можно рекомендовать монографию В. В. Налимова и Н. А. Чернової «Статистические методы планирования экстремальных экспериментов». М., изд-во «Наука», 1965.

При ссылках на эту книгу далее будем указывать только фамилии авторов. Такой принцип цитирования распространим и на другие источники, встречающиеся более одного раза.

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.

Характер и форма изложения материала в книге могут оказаться непривычными для читателя. Это связано с тем, что метод программированного обучения, который тоже является молодым, не нашел еще достаточно широкого применения. Мы увидели в этом методе путь к пре-

одолению упомянутых выше трудностей. Нас вдохновляли книги: «Системы сетевого планирования и управления. Программированное введение в ПЕРТ». М., изд-во «Мир», 1965 и «Основы программирования» Т. Скотта. М., изд-во «Советское радио», 1965.

Главная особенность такого способа изложения материала — это «принуждение» читателя к активному усвоению. Вы просто не сможете продвинуться вперед, пока не овладеете предыдущим материалом. Мы попытались добиться того, чтобы вы постоянно ощущали себя активным участником событий.

Следствием такой формы изложения материала должны являться четкость структуры, лаконичность и однозначность формулировок.

Вы понимаете, как трудно удовлетворить всем этим требованиям. Мы отдаем себе отчет в том, что это, вероятно, не всегда получалось у нас достаточно хорошо. Мы с благодарностью примем любые замечания, предложения и пожелания, направленные на улучшение книги.

И все же мы будем считать 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, но...

Сначала надо условиться о «правилах игры».

Памятка читателю

1. Весь текст разбит на главы, следующие в некоторой логической последовательности. В этой последовательности они и должны читаться. Текст каждой главы, в свою очередь, разбит на порции информации, пронумерованные римскими цифрами.

Порция информации, если она не последняя в главе, заканчивается вопросом (или утверждением). Вопрос (утверждение) может быть связан с данной порцией информации и со всем предыдущим текстом. За вопросом следуют возможные ответы, пронумерованные арабскими цифрами. Нумерация римских и арабских цифр в каждой главе своя.

Следуйте указаниям в конце каждой порции информации!

2. Прочтя текст порции информации и вопрос, вы должны произвести одну из следующих операций:

а) выбрать то утверждение, которое представляется вам правильным;

б) выбрать правильный ответ на поставленный вопрос;

в) заполнить пропуск.

Каждый произведенный выбор укажет вам номер ответа, к которому следует перейти.

Если, с точки зрения авторов, вы произвели правильный выбор, то у вас появится возможность продвинуться дальше; при ошибке — придется вернуться назад.

3. Не смущайтесь, если случайно ошибетесь в выборе правильного ответа. Вам обязательно объяснят вашу ошибку и дадут возможность ее исправить.

Теперь смело переходите к основному тексту и следуйте дальнейшим указаниям.

Глава первая

Основные определения

I. 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.

Мы не будем давать определение термина «эксперимент»: формулировка столь общих понятий слишком сложна, да и вряд ли входит в нашу задачу. Мы понимаем под экспериментом совокупность действий, к которым приходится обращаться, чтобы задавать природе интересующие нас вопросы. Эта совокупность может быть весьма сложной, но ее всегда можно разложить на отдельные элементы, каждый из которых мы называем опытом. Такое определение эксперимента, не являясь исчерпывающим и однозначным, будет для нас достаточным.

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

Да. Обратитесь к пункту 1.

Нет. Обратитесь к пункту 2.

Не знаю. Обратитесь к пункту 3.

1. Правильно. Если модель достаточно точно описывает объект, то эксперимент на объекте может быть заменен экспериментом на модели.

II. 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 условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью. При этом существенно следующее:

1) стремление к минимизации общего числа опытов;
2) одновременное варьирование всеми переменными, определяющими процесс, по специальным правилам — алгоритмам;

3) использование математического аппарата, формализующего многие действия экспериментатора;

4) 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 об его проведении — планирование эксперимента.

Задачи поиска оптимальных условий являются одними 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 найдутся 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, задача минимизации себестоимости продукта или задача 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 решением задачи оптимизации и поиском экстремума некоторой функции.

Давайте рассмотрим следующие две задачи.

а) Прочность бетона в значительной степени определя-

ется маркой цемента, количеством заполнителя и количеством воды. Требуется установить связь между прочностью бетона и названными факторами.

б) 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?

Экстремальной является задача а). Обратитесь к пункту 4.

Экстремальной является задача б) Обратитесь к пункту 5.

Не знаю. Обратитесь к пункту 6.

2. Вы считаете, что поставить эксперимент на математической модели невозможно. Но почему же? Эксперимент можно проводить не только непосредственно на реальном объекте, но и на его модели, которая в этом случае сама становится объектом исследования.

Вернитесь к пункту 1 и выберите другой ответ

3. Вы не знаете, можно ли планировать эксперимент на математической 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.

Вернитесь к пункту 1 и выберите правильный ответ

4. Вы считаете, что первая задача относится к задачам оптимизации. Так ли это? В данном случае важно уметь предсказывать значения прочности в зависимости от значений названных факторов, так как физическое измерение прочности может быть реализовано только через 28 суток.

Чтобы определить, является ли данная задача экстремальной, нужно сформулировать признак экстремальных задач и указать способ его проверки.

Задача является экстремальной, если ее цель — поиск экстремума некоторой функции. В данной задаче этот признак отсутствует, так как не ставилась задача полу-

чения максимальной прочности. Требовалось лишь, оперируя моделью, определить различные значения прочности при разных комбинациях факторов.

Вернитесь к пункту 11 и выберите другой ответ.

5. Вы ответили правильно.

Действительно, характерным признаком, по которому мы отличаем экстремальные задачи от всех прочих, является поиск экстремума некоторой функции. В данном случае требуется найти максимальную надежность прибора, варьируя некоторыми технологическими факторами. Следовательно, задача б) экстремальная.

Задача а) относится к иному типу задач — к задачам интерполяции.

III. Мы надеемся, что вы получили представление о характере задач, которые предстоит рассматривать на протяжении всей книги. Чтобы продвинуться дальше, нам придется определить еще ряд важных понятий, первое из которых — «объект исследования».

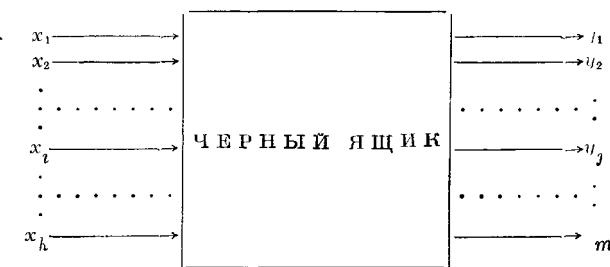


Рис. 1. Схема «черного ящика»

Для описания объекта исследования удобно пользоваться представлением о кибернетической системе, которая схематически изображена на рис. 1.

Иногда такую кибернетическую систему называют «черным ящиком»*. Стрелки справа изображают численные

* Н. Винер. Кибернетика. М., изд-во «Советское радио», 1968
У. Р. Эшби. Введение в кибернетику. М., ИЛ, 1959.

характеристики целей исследования. Мы обозначаем их буквой y 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 x 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:

$$y = \Phi(x_1, x_2, \dots, x_n),$$

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 и 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.

Естественно условиться считать, что число различных состояний определяет сложность данной системы. Зная сложность, мы сразу скажем, сколько различных опытов можно включить в эксперимент в нашей задаче. Поэтому надо научиться оценивать сложность «ящика», если известно число факторов и число уровней каждого фактора.

Мы надеемся, что вы справитесь с этой задачей и правильно ответите на следующий вопрос: какой из двух приведенных ниже «черных ящиков» сложнее?

Первый: шесть факторов на двух уровнях каждый.

Второй: три фактора на четырех уровнях каждый.

Сложнее первый. Обратитесь к пункту 7.

Сложнее второй. Обратитесь к пункту 8.

Они одинаковы по сложности. Обратитесь к пункту 9.

6. Вы не знаете, 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 а) требуется установить связь между прочностью бетона и тремя факторами. Здесь не определено, какая прочность является оптимальной, и не требуется ее оптимизировать. В задаче б) 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 а) мы будем называть интерполяционными, а типа б) — экстремальными.

При определении типа задачи нужно проявить осторожность, так как словесная формулировка может затемнять суть дела.

Теперь вы можете вернуться к пункту II и выбрать правильный ответ.

7. Ваш ответ не верен.

Шесть факторов на двух уровнях дают столько же различных состояний, сколько и три фактора на четырех уровнях, а именно 64.

Вернитесь к пункту III и выберите правильный ответ.

8. Ваш ответ не верен.

Три фактора на четырех уровнях дают столько же различных состояний, сколько и шесть факторов на двух уровнях, а именно 64.

Вернитесь к пункту III и выберите правильный ответ.

9. Вы ответили, что «черные ящики» одинаковы по сложности. Правильно. В обоих случаях число различных состояний равно 64.

IV. Вы, вероятно, потратили много времени, перебирая все возможные состояния и подсчитывая их количество. Эту задачу можно облегчить. 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 времени 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. 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*.

Возможно, плохая воспроизводимость объясняется действием фактора, систематически изменяющегося (дрейфующего) во времени. Тогда нужно обращаться к специальным методам планирования**. 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***. Эти случаи мы не будем рассматривать. Наша цель — изложение методов планирования экстремального эксперимента для воспроизводимых управляемых статических объектов.

Планирование экстремального эксперимента — это ме-

* Ю. П. Адлер, А. И. Ратнер, Г. Ф. Лецинская. Об активно-пассивном эксперименте. — Научные труды Гиредмета, 1969, т. 27. М., изд-во «Металлургия», стр. 16.

** А. Н. Лисенков. Планирование эксперимента в условиях временного дрейфа. — Труды МЭИ, 1967, том 68.

*** В. В. Налимов, Н. А. Чернова. Ук. соч., гл. I.

тод 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 он используется в химии. Детерминированный подход предполагает построение физической модели процесса на основании тщательного изучения механизма явлений (например, кинетики, гидродинамики). Такой подход 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.

Резюме

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.

Если вы чувствуете, что не запомнили основных терминов, то вернитесь к пункту I и просмотрите снова эту главу.

Если все ясно, то переходите к следующей главе.

Желаем вам успеха!

Параметр оптимизации

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, не математическом, 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 не одинаково. 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

I. 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

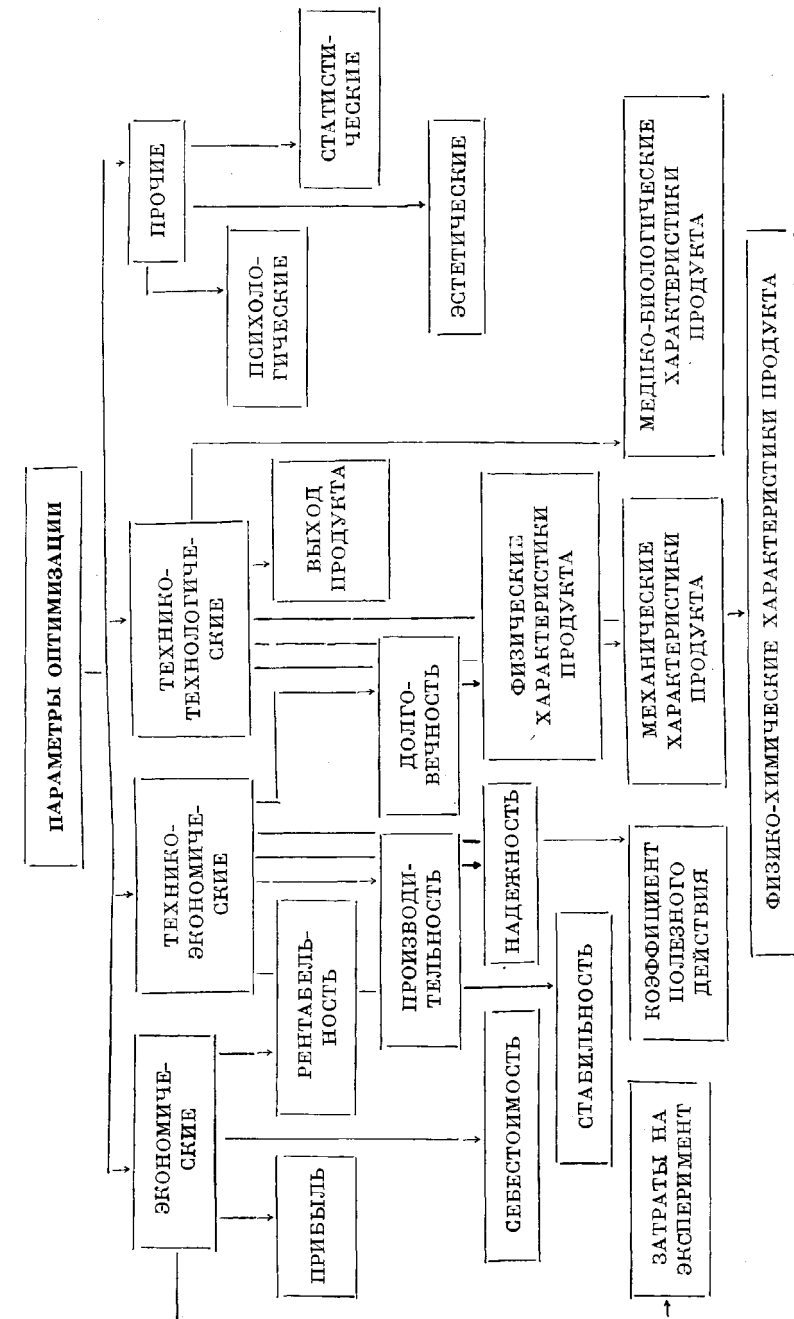


Рис. 2. Классификация параметров оптимизации

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 оптимуму 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 *.

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, про-

цент 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 (проволока, 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 задач 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 *.

Пример 1

Во время 2-й 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?

* Г. Х. Гуд, Р. Э. Макол. Ук. соч.

* Х. Гуд, Р. Э. Макол. Системотехника. Введение в проектирование больших систем. М., изд-во «Советское радио», 1962; J. von Neumann, O. Morgenstern. Theory of games and economic behaviour. Princeton. N. Y., Princeton University Press, 1947; В. М. Добкин. Выбор экономических критериев оптимизации режимных и конструктивных параметров реакторов. — Химическая промышленность, 1968, № 3.

Число сбитых самолетов. Обратитесь к пункту 1.

Потери в судах, оснащенных орудиями, по сравнению с судами без орудий. Обратитесь к пункту 2.

Если вы затрудняетесь в выборе одного из вариантов, то обратитесь к пункту 3.

1. Вы считаете, что эффективность установления орудий на торговые суда можно оценить числом сбитых самолетов. 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.

Вернитесь к пункту I и выберите другой ответ.

2. Вы полагаете, что эффективность установки орудий на торговые суда можно оценить сопоставлением потерь в судах, оснащенных орудиями, с потерями в судах без орудий. 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 окупились очень быстро.

* * *

Вы познакомились с наиболее распространенными видами параметров оптимизации. Теперь давайте подойдем к рассмотрению требований с которыми связан их выбор.

Переходите к пункту II.

3. Вы затрудняетесь в выборе параметра оптимизации для оценки эффективности установки орудий на торговых судах. Вероятно, вам не очень нравится военно-мор-

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 *). 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?

Да. Обратитесь к пункту 4.

Нет. Обратитесь к пункту 5.

Не знаю. Обратитесь к пункту 6.

4. Вы считаете, что проставление отметок по вкусовым свойствам дает возможность перейти к оптимизации. Ну что же, мы тоже так думаем. Здесь используется **ранжирование**. Если нет способа количественного измерения результата, то ранговый подход необходим.

Обратитесь к пункту III.

5. Вы считаете, что проставление отметок по вкусовым свойствам не дает возможности перейти к оптимизации. Так ли это? Нам необходимо как-то количественно представить результат приготовления пирога. Может быть,

* H. Smith, A. Rose. Subjective responses in process investigation.— Industrial and Engineering Chemistry, 1963, 55, № 7.

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 *). 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?

Да. Обратитесь к пункту 4.

Нет. Обратитесь к пункту 5.

Не знаю. Обратитесь к пункту 6.

4. Вы считаете, что проставление отметок по вкусовым свойствам дает возможность перейти к оптимизации. Ну что же, мы тоже так думаем. Здесь используется **ранжирование**. Если нет способа количественного измерения результата, то ранговый подход необходим.

Обратитесь к пункту III.

5. Вы считаете, что проставление отметок по вкусовым свойствам не дает возможности перейти к оптимизации. Так ли это? Нам необходимо как-то количественно представить результат приготовления пирога. Может быть,

* H. Smith, A. Rose. Subjective responses in process investigation.— Industrial and Engineering Chemistry, 1963, 55, № 7.

вам не нравится пятибалльная шкала? Тогда вы можете оценивать по двухбалльной (1 — вкусный пирог, 0 — невкусный пирог), десятибалльной и т. д. Тут важен принцип: необходима именно **количественная оценка** качества пирога. Эта количественная оценка носит условный (субъективный) характер. Мы ставим в соответствие качественному признаку некоторое число — **ранг**.

Если пирогу будет поставлена наивысшая оценка, то свою задачу вы сможете считать выполненной. Это один из способов количественного измерения результата, называемый **ранговым**.

Теперь вам будет нетрудно в пункте II выбрать правильный ответ.

6. Вы не рискуете принять решение о возможности оптимизации с помощью проставления отметок по вкусовым свойствам. 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.

Вернитесь к пункту II и выберите правильный ответ.

III. 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 показания прибора. Например, скорость движения машины определяется числом на спидометре. Чаще 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 пользо-

ваться значениями отклонений (или квадратов отклонений) от этого числа.

Еще одно требование, связанное с количественной природой параметра оптимизации, — о д н о з н а ч н о с т ь 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» *.

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»**.

Пример 3

При флотации сульфидной руды в лабораторных условиях изучалась эффективность применения нового ре-

* И. Винер. Ук соч.

** Ст. Вир. Кибернетика и управление производством. М., изд-во «Наука», 1965.

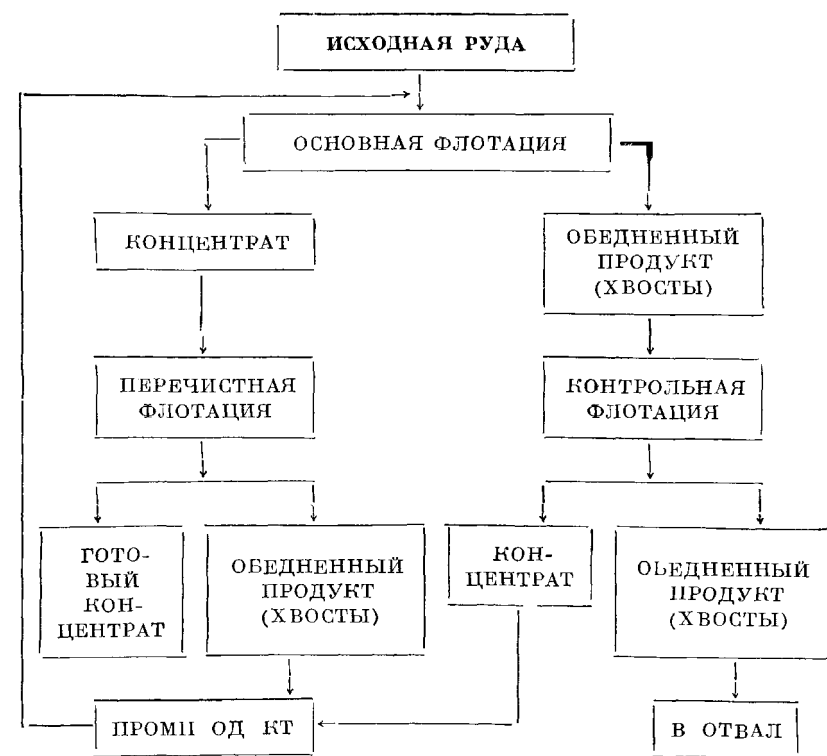


Рис. 3. Схема флотационного обогащения руды

агента-пенообразователя по схеме рис. 3*. 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?

Да. Обратитесь к пункту 7.

Нет. Обратитесь к пункту 8.

Не знаю. Обратитесь к пункту 9.

* Л. А. Барский, И. Н. Плаксин. Критерии оптимизации раздельных процессов. М., изд-во «Наука», 1967.

7. Вы ответили, что параметр оптимизации — извлечение концентрата в основной флотации — выбран обоснованно, хотя и ставилась задача оптимизации всего процесса флотационного обогащения руды. А результаты промышленных испытаний показали, что ухудшилось извлечение и качество концентрата после контрольной флотации, что повлияло на питание основной флотации и в итоге привело к снижению выхода и качества по всему циклу.

Поэтому вернитесь к пункту III и выберите другой ответ.

8. Вы ответили, что параметр оптимизации — извлечение концентрата в основной флотации — при решении задачи оптимизации всего процесса флотационного обогащения руды выбран не совсем обоснованно. Это правильный ответ, потому что существенно достижение конечной цели — получение готового концентрата (после перемешивания 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 не эффективным с точки зрения достижения конечной цели.

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

Обратитесь далее к пункту IV.

9. Вы затрудняетесь в выборе решения. Посмотрите еще раз схему процесса (рис. 3). Исследователи выбрали параметр оптимизации, который характеризует эффективность достижения промежуточной цели — повышение выхода концентрата на основной флотации. На последней

стадии, после перемешивания флотации, получается готовый концентрат. Можно ли результаты, достигнутые на промежуточной стадии, отнести ко всему процессу?

Вернитесь к пункту III и выберите одно из первых двух утверждений.

IV. Мало иметь эффективный параметр оптимизации. Надо еще, чтобы он был эффективным 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 *.

Пример выбора параметра оптимизации, обладающего полнотой, рассмотрен в работе ** для процессов зонной перекристаллизации. Обычно применяемый для этой цели

* В. М. Добкин. Ук. соч.

** В. Н. Вигдорович, Ю. П. Адлер, А. Е. Вольян. Об оценке эффективности процессов зонной перекристаллизации. — Изв. АН СССР. Metallургия и горное дело, 1964, № 2.

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. 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:

- 1) эффективным с точки зрения достижения цели;
- 2) универсальным;
- 3) количественным и выражаться одним числом;
- 4) статистически эффективным;
- 5) имеющим физический смысл, простым и легко вычисляемым;
- 6) 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.

Глава третья

Факторы

Теперь нам предстоит рассмотреть способы воздействия на оптимизируемый объект. Как вы знаете (стр. 21), 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. 7.) 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. 6.

Если число факторов больше пятнадцати, нужно обратиться к методам отсеивания несущественных факторов. Здесь можно воспользоваться формализацией априорной

информации *, методом случайного баланса **, планами Плаккетта-Бермана*** 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.

3.1. Определение фактора

I. 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.

* Ю. П. Адлер, И. Ф. Александрова, Ю. В. Грановский, В. В. Налимов. Об одном методе формализации априорной информации при планировании эксперимента. — Сб. «Планирование эксперимента». М., изд-во «Наука», 1966; Руководящие технические материалы. Экспериментально-статистические методы получения математического описания и оптимизации сложных технологических процессов (Ранговая корреляция). Вып. 3., М., ОКБА, 1966.

** В. В. Налимов, Н. А. Чернова. Ук. соч.; F. E. Satterthwaite. Random balance experimentation. — *Technometrics*, 1959, 1, № 2.

*** R. L. Plackett, I. P. Burman. The design of optimum multifactor experiments. — *Biometrika*, 1946, 33, № 4.

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 5.)

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.

В двух следующих примерах заполните пропуски.

1. Время реакции, температура, концентрация реагирующих веществ, скорость подачи вещества, величина pH — это примеры наиболее часто встречающихся... факторов.

2. Различные реагенты, абсорбенты, вулканизирующие агенты, кислоты, металлы являются примерами уровней... факторов.

Правильный ответ помещен ниже.

1. Количественных.

2. Качественных.

II. Приведем примеры, из которых будет видно, какие же факторы выбирает экспериментатор при оптимизации конкретных процессов.

Пример 1

Наш первый пример относится к исследованию процесса вулканизации бутадиен-стирольного каучука солями непредельных кислот. В планирование эксперимента были включены следующие факторы:

$\tilde{x}_1$ — температура вулканизации, °C;

$\tilde{x}_2$ — время вулканизации, мин;

$\tilde{x}_3$ — количество инициатора, вес. ч.;

$\tilde{x}_4$ — количество вулканизирующего агента, вес. ч.;

$\tilde{x}_5$ — количество окисла, вес. ч.;

$\tilde{x}_6$ — тип окисла (окись цинка или окись магния);

$\tilde{x}_7$ — тип кислотного остатка (метакрилат, малеат);

$\tilde{x}_8$ — 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. *

Пример 2

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

* А. А. Донцов, Е. В. Маркова, В. Э. Михлин, Б. А. Догадкин. Применение математико-статистического метода для оптимизации процесса вулканизации эластомеров солями ненасыщенных кислот — Каучук и резина, 1967, № 10.

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.

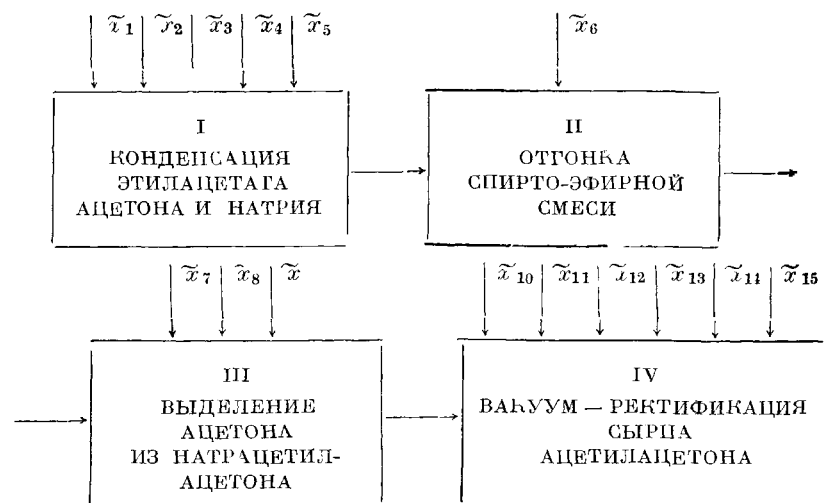


Рис. 4. Схема процесса получения ацетилацетона

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 сырца 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?

Да. Обратитесь к пункту 3.

Нет. Обратитесь к пункту 4.

Не знаю. Обратитесь к пункту 5.

3. Вы считаете, что в планирование эксперимента можно включать факторы, относящиеся к различным стадиям.

Действительно это так. Во многих случаях это является не только возможным, но и необходимым. Например, когда параметр оптимизации измеряется в конце последней стадии, как в рассматриваемом примере, а на предыдущих стадиях выходной параметр не измеряется. (Чаще всего это происходит из-за отсутствия нужных аналитических методик.)

Переходите к пункту III.

4. Ваше мнение: в планирование эксперимента нельзя включать факторы, относящиеся к различным стадиям. Почему же вы пришли к такому мнению? Каждую стадию можно представить в виде маленького «черного ящика», имеющего свои входы. Все стадии объединяются в один большой «черный ящик», на который воздействует сумма всех факторов. Конечно, не во всех случаях при оптимизации многостадийных процессов необходимо рассматривать все стадии как единое целое.

Так, например, случилось при оптимизации процесса получения никодина. Процесс состоит из двух химических стадий: получение никотинамида и оксиметилирование никотинамида *. Эти две стадии взаимосвязаны. Каж-

* Е. В. Маркова. Планирование эксперимента при оптимизации процессов тонкого органического синтеза. Диссертация на соискание ученой степени кандидата технических наук. М., 1965.

дая стадия имеет свой параметр оптимизации — выход продукта. Экспериментаторы приняли решение оптимизировать только вторую стадию, потому что выход реакции на первой стадии достигал 90—95% теоретического значения. Длительность реакции на первой стадии — десять часов. Если бы эти две стадии рассматривались как единое целое, это слишком бы усложнило экспериментирование. Однако во многих случаях представление многостадийного процесса как единого целого является совершенно необходимым. Это случается тогда, когда параметр оптимизации измеряется в конце последней стадии, как в рассматриваемом примере, или когда параметр оптимизации отдельной стадии противоречит общей цели оптимизации.

Считая, что в планирование эксперимента нельзя включать факторы, относящиеся к различным стадиям, вы неправы.

Вернитесь к пункту II и выберите другой ответ.

5. Вы не знаете, можно ли включать в планирование эксперимента факторы, относящиеся к различным стадиям.

Давайте рассуждать вместе. Многостадийный процесс можно представить как единое целое в виде большого «черного ящика».

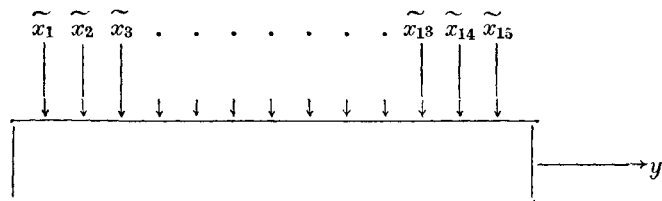


Рис. 5. Схема процесса получения ацетилацетона

В рассматриваемом примере процесс получения ацетилацетона, состоящий из четырех стадий, удобно представить как единое целое в виде одного «черного ящика» (рис. 5).

На этот «черный ящик» воздействуют 15 факторов. Почему возникла необходимость рассматривать все ста-

дии как единое целое? 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 отдельные 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.

Вернитесь к пункту II и выберите подходящий ответ.

3.2. Требования, предъявляемые к факторам при планировании эксперимента

III. 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?

Да. Обратитесь к пункту 6.

Нет. Обратитесь к пункту 7.

Не знаю. Обратитесь к пункту 8.

6. Вы хотите температуру воздуха включить как фактор в планирование эксперимента. Мы не можем согласиться с вами. 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.

Возвратитесь к пункту III и выберите другой ответ.

7. Вы ответили — нет. Совершенно правильно. Температурой окружающей среды в данном случае управлять невозможно. Ее можно только контролировать.

IV. 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 конфлюэнтному анализу*.

Факторы должны быть непосредственными воздействиями на объект. Факторы должны быть *однозначными*. 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

* Н. П. Клепиков, С. Н. Соколов. Анализ и планирование экспериментов методом максимума правдоподобия. М., изд-во «Наука», 1964.

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:

- $\bar{x}_1$ — количество едкого натра, г/мол;
- $\bar{x}_2$ — способ поддержания рН;
- $\bar{x}_3$ — время прилива вещества a , час;
- $\bar{x}_4$ — время выдержки реакционной массы, час;
- $\bar{x}_5$ — температура реакционной среды, °C;
- $\bar{x}_6$ — весовое соотношение вещества b и метанола, г/г;
- $\bar{x}_7$ — молярное соотношение вещества a и вещества b ,
 $\frac{\text{г/мол}}{\text{г/мол}}$.

(Матрица планирования для этих факторов приведена в гл. 6).

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?

Да. Обратитесь к пункту 9.

Нет. Обратитесь к пункту 10.

Не знаю. Обратитесь к пункту 11.

8. Вы ответили, что не знаете.

Ну что ж, давайте проанализируем наше определение фактора. Мы сказали, что фактор — это измеряемая переменная величина, принимающая в некоторый момент времени определенное значение. Температура воздуха является переменной величиной, принимающей определенное значение, которое можно контролировать. Если брать во внимание только эту часть определения, то можно температуру воздуха считать фактором. Действительно, для «пассивного» эксперимента, когда необходимо только контролировать переменные, температура будет факто-

ром. Но мы проводим «активный» эксперимент, вмешиваясь в процесс и поддерживая все факторы на определенных уровнях.

Факторы при планировании эксперимента — это способы воздействия на процесс. Факторами нужно управлять. Температура воздуха не отвечает этим требованиям и поэтому не может быть фактором.

Вернитесь к пункту III и выберите другой ответ.

9. Давайте проанализируем, верно ли, что скорость диффузии может быть фактором. Скорость диффузии зависит от концентрации, от 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.

В числе требований, предъявляемых факторам, есть **однозначность**. Фактор может быть непосредственным воздействием на процесс. Фактором нужно управлять. Скорость диффузии не отвечает этим требованиям. 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.

Вернитесь к пункту IV и выберите другой ответ.

10. Верно. Скорость диффузии не отвечает всем требованиям, которые предъявляются к факторам при планировании эксперимента. Скорость диффузии является функцией многих других факторов, следовательно, не выполняется требование **однозначности**. Скоростью диффузии весьма трудно управлять. Она не является непосредственным воздействием на объект исследования.

Переходите к пункту V.

11. Вы не знаете. Остановимся на этом примере более детально.

Мы рассматриваем процесс растворения твердого тела в жидкости и хотим определить, отвечает ли скорость диффузии требованиям, предъявляемым к факторам.

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. Скорость диффузии может быть выражена уравнением

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

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.

Вернитесь к пункту IV и выберите правильный ответ.

3.3. Требования к совокупности факторов

V. 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 осмолению 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 к 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: $\bar{x}_1$ — давление, атм, $\bar{x}_2$ — объем, л, $\bar{x}_3$ — температуру, °К?

Да. Обратитесь к пункту 12.

Нет. Обратитесь к пункту 13.

Не знаю. Обратитесь к пункту 14.

12. Вы ответили «да». Это, по-видимому, произошло потому, что вы были невнимательны. Требование к сово-

купности факторов — отсутствие между ними корреляции. Если это условие не выполняется, то невозможно планировать эксперимент. В данном случае один из трех факторов зависит от того, как выбраны два других. Поэтому все три фактора включить в планирование невозможно.

Вернитесь к пункту V и выберите другой ответ.

13. Вы ответили правильно. Нет никакой необходимости включать в планирование эксперимента зависимые факторы. В данном случае один фактор из трех не несет никакой дополнительной информации.

Переходите к пункту VI.

14. Вы не знаете. Рассмотрим этот вопрос более подробно.

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 .

Вернитесь к пункту V и выберите правильный ответ.

3.4. Примеры факторов

VI. 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, медицина, обогащение полезных ископаемых, пищевая и текстильная промышленность, сельское хозяйство, военное дело и др. *

Применяется планирование эксперимента и в нескольких неожиданных областях исследования, в таких, как ге-

ронтология (наука о долголетию), 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, базирясь 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* *.

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, *°C*;
- $\tilde{x}_3$ — время выдержки под давлением, *мин*;
- $\tilde{x}_4$ — температура в муфеле при прессовании, *°C*;
- $\tilde{x}_5$ — время температурной выдержки, *мин*;
- $\tilde{x}_6$ — дисперсность наполнителя, *мк*;
- $\tilde{x}_7$ — соотношение флюса и наполнителя, *г/г*;

* Э. Б. Чамлик, В. М. Никитин, Э. М. Менчер. Выделение существенных факторов при электролитическом получении алюминия. — Сб. «Проблемы планирования эксперимента». М., изд-во «Наука», 1969.

* Ю. П. Адлер, Ю. В. Грановский. Обзор прикладных работ по планированию эксперимента. Препринт № 1. МГУ, 1967.

$\tilde{x}_8$ — давление при шмотировании, кг/см^2 ;
 $\tilde{x}_9$ — дисперсность сажи, мк ;
 $\tilde{x}_{10}$ — время выдержки при шмотировании, мин ;
 $\tilde{x}_{11}$ — качество керамики оснований;
 $\tilde{x}_{12}$ — дисперсность флюсов, мк^* .

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 Na_2O), г/л ;
 $\tilde{x}_2$ — сульфитность раствора, %;
 $\tilde{x}_3$ — конечная температура варки, $^{\circ}\text{C}$;
 $\tilde{x}_4$ — продолжительность подъема температуры до конечной, мин ;
 $\tilde{x}_5$ — продолжительность варки при конечной температуре, мин^{**} .

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^{***} .

* Л. Г. Власов, В. Б. Лукьянов, Б. Г. Красильников, А. Ю. Мольтова. Применение методов планирования экстремального эксперимента в производстве резисторов. Доклад, представленный на 2-ю Всесоюзную конференцию по планированию эксперимента. М., 1968.

** Э. М. Менчер, Р. З. Пен, М. Г. Малина, Н. М. Стоит. Опыт изучения варки сульфатной целлюлозы с применением статистических методов планирования эксперимента. Там же.

*** Л. А. Барский, Ю. Б. Рубинштейн. Особенности планирования экстремальных экспериментов при исследовании разделительных процессов.— Сб. «Проблемы планирования эксперимента». М., изд-во «Наука», 1969.

Пример 8

При оптимизации процесса экстракции циркония и гафния из солянокислотных растворов в качестве независимых переменных приняты:

$\tilde{x}_1$ — концентрация металла, %;
 $\tilde{x}_2$ — концентрация кислоты, %;
 $\tilde{x}_3$ — концентрация спирта, %;
 $\tilde{x}_4$ — соотношение объемов фаз, мл/мл^* .

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 , $\tilde{x}_6$ — CaCl_2 ,
 $\tilde{x}_2$ — ZnCl_2 , $\tilde{x}_7$ — MnSO_4 ,
 $\tilde{x}_3$ — FeCl_3 , $\tilde{x}_8$ — ZnSO_4 ,
 $\tilde{x}_4$ — CuSO_4 , $\tilde{x}_9$ — $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}4\text{H}_2\text{O}$,
 $\tilde{x}_5$ — H_3BO_3 , $\tilde{x}_{10}$ — CoCl_2^{**}

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,

* Н. С. Смирнова, Ю. В. Грановский, А. Л. Каплан, И. А. Чернова, Л. Н. Комиссарова. Планирование эксперимента при изучении экстракции циркония и гафния солями.— Сб. «Проблемы планирования эксперимента». М., изд-во «Наука», 1969.

** И. М. Чирков, Л. Е. Гурина, С. С. Рылкин. Применение методов математического планирования экспериментов в микробиологических исследованиях. Там же.

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:

$\tilde{x}_{14}$ — биомицин.

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

3.5. 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 экспериментатора.

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 высокая. 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.

Глава четвертая

Выбор модели

I. 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 располагать 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.

Вспомнив то, что вы прочитали в прошлой главе о

факторах, скажите, правильно ли обратное утверждение: каждая точка на плоскости соответствует одному из возможных состояний «черного ящика»?

Да. Обратитесь к пункту 1.

Нет. Обратитесь к пункту 2.

Не знаю. Обратитесь к пункту 3.

1. Если вы считаете, что утверждение справедливо, то вам следует вернуться к началу третьей главы и еще раз рассмотреть вопрос о факторах, их видах и уровнях.

Только после этого вернитесь к пункту I и выберите другой ответ.

2. Вы правы. Обратное утверждение справедливо только в том случае, если все факторы могут принимать любые значения от $-\infty$ до $+\infty$.

II. 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

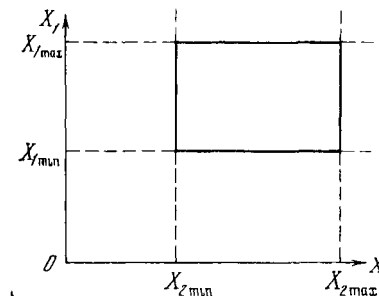


Рис. 6. Область определения факторов

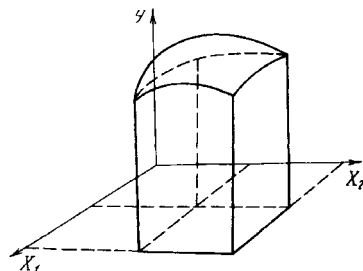
прямоугольник, внутри которого лежат точки, соответствующие состояниям «черного ящика» (рис. 6). Пунктирными линиями на рисунке обозначены границы областей определения каждого из факторов, а сплошными — границы их совместной области определения.

Чтобы указать значение параметра оптимизации, требуется еще одна ось координат. Если ее построить, то поверхность отклика будет выглядеть, как на рис. 7. 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.

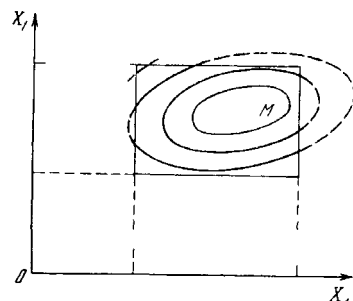
Но для двух факторов можно даже не переходить к трехмерному пространству, а ограничиться плоскостью.

Рис. 7. Поверхность отклика



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. 8). Точка M

Рис. 8. Проекция сечений поверхности отклика на плоскость



на рисунке — это и есть та оптимальная точка, которую мы ищем. Каждая линия соответствует постоянному значению параметра оптимизации. Такая линия называется линией равного отклика.

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

Как вы думаете, справедливо ли утверждение, что каждому состоянию «ящика» (т. е. каждому возможному набору значений x_1 и x_2) соответствует одно значение параметра оптимизации?

Да. Обратитесь к пункту 4.

Нет. Обратитесь к пункту 5.

Не знаю. Обратитесь к пункту 6.

3. Вы не знаете, какой ответ выбрать.

Давайте посмотрим, какие рассуждения могут лежать в основе выбора. С одной стороны, ясно, что некоторые точки плоскости соответствуют возможным состояниям «ящика». С другой стороны, должны существовать некоторые условия, при которых справедливо обратное утверждение. Очевидно, если каждый фактор может принимать любое значение от $-\infty$ до $+\infty$, то эти условия выполняются.

Теперь, если вы забыли предыдущую главу, то вернитесь к ней, а если помните, то вернитесь сразу к пункту I и выберите подходящий ответ.

4. Вы правы. Действительно 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 «ящика».

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.

III. 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 *.

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.

Переходите к параграфу 4. 1.

5. Ваш ответ неверен.

Посмотрите на рис. 8. Если найти точку, соответствующую каким-либо возможным значениям факторов, то через нее будет проходить только одна линия равных откликов. (Может оказаться, что она не показана на чертеже, но ее всегда можно провести). Нет такой точки, где бы пересекались хотя бы две линии. Значит, однозначность существует.

Поэтому вернитесь к пункту II и выберите правильный ответ.

6. Вы не знаете, что ответить. В чем причина ваших затруднений?

Возможно, вы забыли (или не усвоили), как строится «черный ящик». Тогда целесообразно вернуться к гл. 1 и повторить этот материал.

* С ней можно познакомиться, например, по книге Л. А. Растргина: «Случайный поиск в задачах оптимизации многопараметрических систем». Рига, изд-во «Зинатне», 1965. Термин «многопараметрические» здесь употреблен вместо нашего «многофакторные».

Может быть, вас смущает то, что значения параметра оптимизации определяются с ошибками и непонятно, о каком соответствии может идти речь. Если это так, то мы должны оговориться, что пока рассматриваем ситуацию без учета ошибок (мы вернемся к ним позже).

Фактически, вопрос сводится к тому, есть ли на рис. 8 точки, в которых пересекаются кривые равных откликов. Если такие точки есть, надо выбрать ответ «нет», а если точек нет — ответ «да».

Итак, вернитесь к пункту II и выберите подходящий ответ.

4.1. Шаговый принцип

IV. 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 гладкость 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 не равноценны...

На рис. 9 приведены две картинки, изображающие функции отклика для одного фактора.

Как вы думаете, в каком случае нарушены наши предпосылки?

В случае *а*. Обратитесь к пункту 7.

В случае *б*. Обратитесь к пункту 8.

Не знаю. Обратитесь к пункту 9.

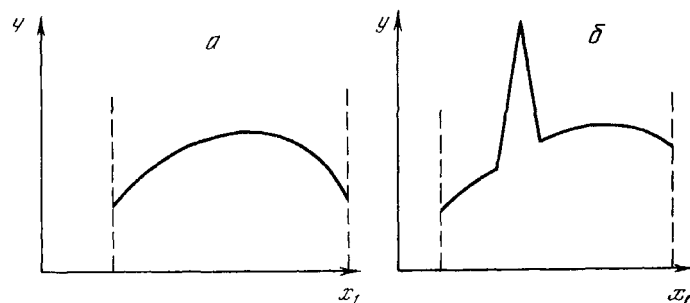


Рис. 9. Примеры функций отклика для одного фактора

7. Вы усмотрели нарушение предпосылок в случае *а*. В чем же состоят эти нарушения? На этом рисунке изображена гладкая непрерывная функция с одним оптимумом. Это как раз то, о чем мы мечтаем. Нам непонятен ваш выбор.

Вернитесь к пункту IV и прочтите снова о предпосылках.

После этого, мы надеемся, вы сделаете правильный выбор.

8. Вы правы.

На рис. 9б 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.

V. 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 условий 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

то. Такая проверка называется интерпретацией модели. Она, конечно, имеет большое значение, и мы вернемся к ней позже (в гл. 9).

На рис. 10 изображены два варианта поиска оптимума для одной и той же поверхности. Крестиками на рисунке обозначены условия опытов. В случае *a* использован подход,

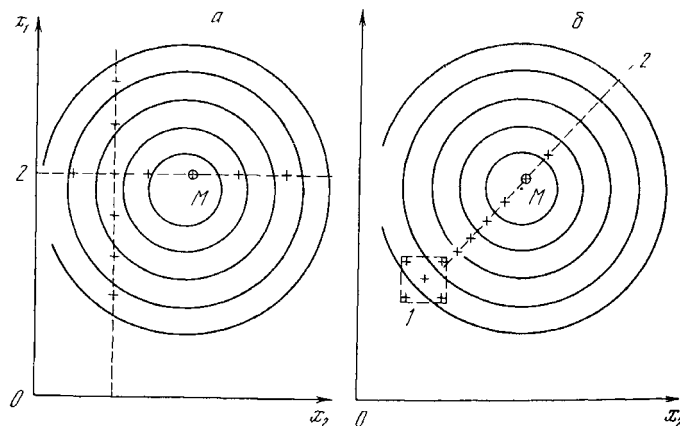


Рис. 10. Два способа поиска оптимума (пример)

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 *b* 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. 10), что в обоих случаях достигнут одинаковый результат при одинаковом суммарном количестве опытов.

Как вы думаете, всегда ли эти две процедуры эквивалентны?

Да. Обратитесь к пункту 10.

Нет. Обратитесь к пункту 11.

Не знаю. Обратитесь к пункту 12.

9. Вы не можете ответить.

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).

Если вы будете сравнивать картинки, руководствуясь этими соображениями, то, несомненно, выберите в пункте IV правильный ответ.

10. Вы утверждаете, что процедуры всегда будут эквивалентны. Это неверно.

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

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.

Вернитесь к пункту V и выберите правильный ответ.

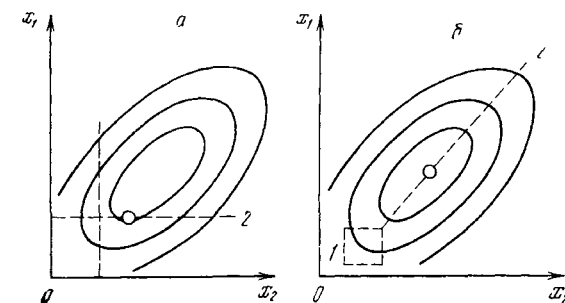


Рис. 11. Два способа поиска оптимума (продолжение)

11. Конечно.

Вот иллюстрация, которая сразу показывает правильность вашего ответа (рис. 11). Это, разумеется, только иллюстрация. В жизни не всегда удастся за один цикл достиг-

нуть оптимума. 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.

Переходите к пункту VI.

12. Вы не знаете. Не огорчайтесь, это не простой вопрос. Давайте думать вместе.

Нам важно установить, от чего зависит успех при применении каждой из процедур.

Во-первых, ясно, что успех зависит от вида поверхности отклика. Мы привели случай симметричной окружности. Здесь многие методы окажутся эквивалентными. А если, например, эллипс?

Во-вторых, влияют свойства самих процедур. Важно, какие значения факторов зафиксированы на первом этапе в процедуре *а*. Случайно может оказаться, что, например, x_2 сразу установлен на оптимальный уровень. Важно, в какой последовательности перебираются факторы: что зафиксировано раньше — x_1 или x_2 , в процедуре *б* существенно, где выбрана исходная подобласть, и т. д.

Мы надеемся, что эти рассуждения помогут вам сделать в пункте V правильный выбор.

4.2. Как выбрать модель?

VI. 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

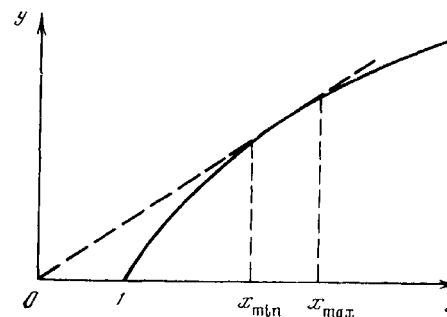


Рис. 12. График логарифмической функции

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. 8.

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

На рис. 12 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?

Проще уравнение (1). Обратитесь к пункту 13.

Проще уравнение (2). Обратитесь к пункту 14.

Не знаю. Обратитесь к пункту 15.

13. Почему вы предпочли уравнение (1)?

Возможно, вы считали, что это уравнение проще потому, что есть таблица логарифмов, по которой легко найти любое нужное значение.

Ну а если бы не было такой таблицы?

Возможно, что ваш выбор верен, но мы все же просим вас вернуться к пункту VI и рассмотреть другие варианты.

14. Почему вы предпочли уравнение (2)?

Возможно, вы считали, что это уравнение проще потому, что оно линейное алгебраическое, а первое — трансцендентное.

Ну а если точное решение не требуется и для любых практически интересных случаев есть готовые таблицы?

Возможно, что ваш выбор верен, но мы все же просим вас вернуться к пункту VI и рассмотреть другие варианты.

15. Этот ответ верен. 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».

VII. 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 сравнивать между собой классы моделей, если свойства объекта заранее не известны? 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: $y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$.

Попробуйте написать полином третьей степени.

Сравните с ответом ниже.

16. Вот правильный ответ.

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

Добейтесь того, чтобы ваш ответ получился правильным.

Поупражняйтесь в записи полиномов для 3, 4 и 5 переменных. Убедитесь в том, что вы освоили это дело.

4.3. Полиномиальные модели

VIII. 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

* В. В. Налимов, Н. А. Чернова Указ. соч.

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

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

Да. Обратитесь к пункту 17.

Нет. Обратитесь к пункту 18.

Не знаю. Обратитесь к пункту 19.

17. Вы ответили, что можно всегда использовать полином нулевой степени. К сожалению, это не так.

Вы, вероятно, не заметили, что значения этого полинома не зависят от уровней факторов. Какое бы состояние «ящика» вы ни взяли, ответ всегда один. Может случиться, что результат действительно не будет зависеть от состояния объекта, но трудно предположить, что так будет всегда. Если же модель станет предсказывать для всех условий один и тот же результат, а действительные результаты будут существенно различны, то трудно рассчитывать на адекватность.

Вот почему вы должны вернуться к пункту VIII и выбрать правильный ответ.

18. Правильно. Трудно ожидать, что результаты опытов будут всегда одинаковы независимо от уровней факторов. А если результаты различны, то такая модель не будет адекватной, т. е. не будут выполняться наши требования.

IX. 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?

Нет. Обратитесь к пункту 20.

Да. Обратитесь к пункту 21.

Не знаю. Обратитесь к пункту 22.

19. Вы затрудняетесь ответить.

Давайте разберемся. Что существенно? Чтобы модель предсказывала направление улучшения параметра оптимизации.

Если поведение объекта не зависит от его состояния, то любое состояние является оптимальным. Этому случаю и соответствует полином нулевой степени. В него не входят значения факторов.

Значит, все дело в том, чтобы оценить, является ли такая ситуация типичной. Вероятно, нет. Обычно объект реагирует на изменение состояния, и модель должна это учитывать.

Поэтому вернитесь к пункту VIII и выберите правильный ответ.

20. Если вы учли, что предполагается выполнение наших постулатов, то ваш ответ неверен.

Полином первой степени (или линейная модель) содержит информацию об изменении параметра оптимизации при изменении уровней факторов. Кроме того, он включает наименьшее возможное число коэффициентов. Вы можете возразить, что такая модель совсем не обязана быть адекватной. Это правильно. Но в наших силах добиться ее адекватности. Как?

Вернитесь к пункту IX, выберите правильный ответ и вы узнаете как.

21. Ваш ответ верен.

Полином первой степени — линейная модель — это то, что нам нужно.

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.

X. 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. 9 и 11.

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.

Переходите теперь к резюме, где мы подведем итог этой главе.

22. Ваше затруднение можно понять. Неясно, почему возник вопрос о следующем полиноме, если полином нулевого порядка отвергнут. Не лучше ли решать эти вопросы в каждом конкретном случае и не заниматься прожектерскими попытками найти универсальную рекомен-

дацию. Что толку, например, в полиноме первой степени, если объект адекватно представляется, скажем, полиномом третьей степени.

И все-таки общая рекомендация возможна. Она возможна благодаря тому, что мы хотим знать только направление градиента, для чего полином первой степени достаточен, если выполняются предпосылки и некоторые дополнительные условия.

Их мы рассмотрим после того, как вы вернетесь к пункту IX и выберете правильный ответ.

4.4. 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 параметра 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, кото-

рую 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 линейен 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.

Как выбрать область, мы рассмотрим ниже. 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, 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.

Глава пятая

Полный факторный эксперимент

«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. Весь процесс исследования можно считать состоящим из последовательности этапов, часть из которых полностью формализованы, а часть требуют «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.

5.1. Принятие решений перед планированием эксперимента

I. При выборе области эксперимента должны учитываться следующие соображения.

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*.

* Н. М. Пруткова. Автореферат кандидатской диссертации. МГУ, 1967.

Пример 1

Изучалось ионообменное разделение смесей группы редкоземельных элементов растворами иминодиуксусной кислоты. Параметр оптимизации — содержание неодима в выходном растворе (элюанте) в процентах. Рассматривалось всего два фактора: концентрация элюанта (входного раствора), % вес (x_1) и рН элюанта (x_2). Как построить область определения факторов? Начнем с x_1 . Известно, что при $x_1 > 3$ работать нельзя, так как это предел растворимости данного вещества при нормальной температуре. Значит, верхний предел $x_1 = 3$. С нижним пределом дело обстоит сложнее. Здесь нельзя указать четкую границу. Известно только, что чем ниже концентрация, тем дольше идет процесс. При $x_1 = 0,5$ 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 x_2 использовали 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 рН < 3 кислота находится в недиссоциированном состоянии, а при рН > 8 оба соединения разрушаются. Следовательно, 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. 3).

Вы видите, как не просто 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 экспе-

риментальных 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 основной 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. 13 изображена область определения для двух факторов. Кружком отмечены наилучшие условия, известные из априорной информации. Известно также, что имеется возможность дальнейшего улучшения параметра оптимизации, а данное значение нас не удовлетворяет. Как вы считаете, можно ли использовать эту точку в качестве основного уровня?

Да. Обратитесь к пункту 1.

Нет. Обратитесь к пункту 2.

Не знаю. Обратитесь к пункту 3.

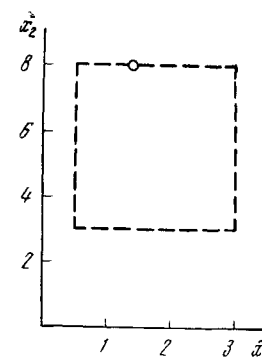


Рис. 13. Область определения двух факторов

1. Вы считаете, что эту точку можно рассматривать в качестве основного уровня. Это не так. 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 нельзя.

Вернитесь к пункту I и выберите другой ответ

2. Вы правы. Точка, лежащая на границе области определений, не может рассматриваться в качестве центра эксперимента.

II. Резюмируем наши рассуждения о принятии решений при выборе основного уровня в виде блок-схемы (см стр. 85).

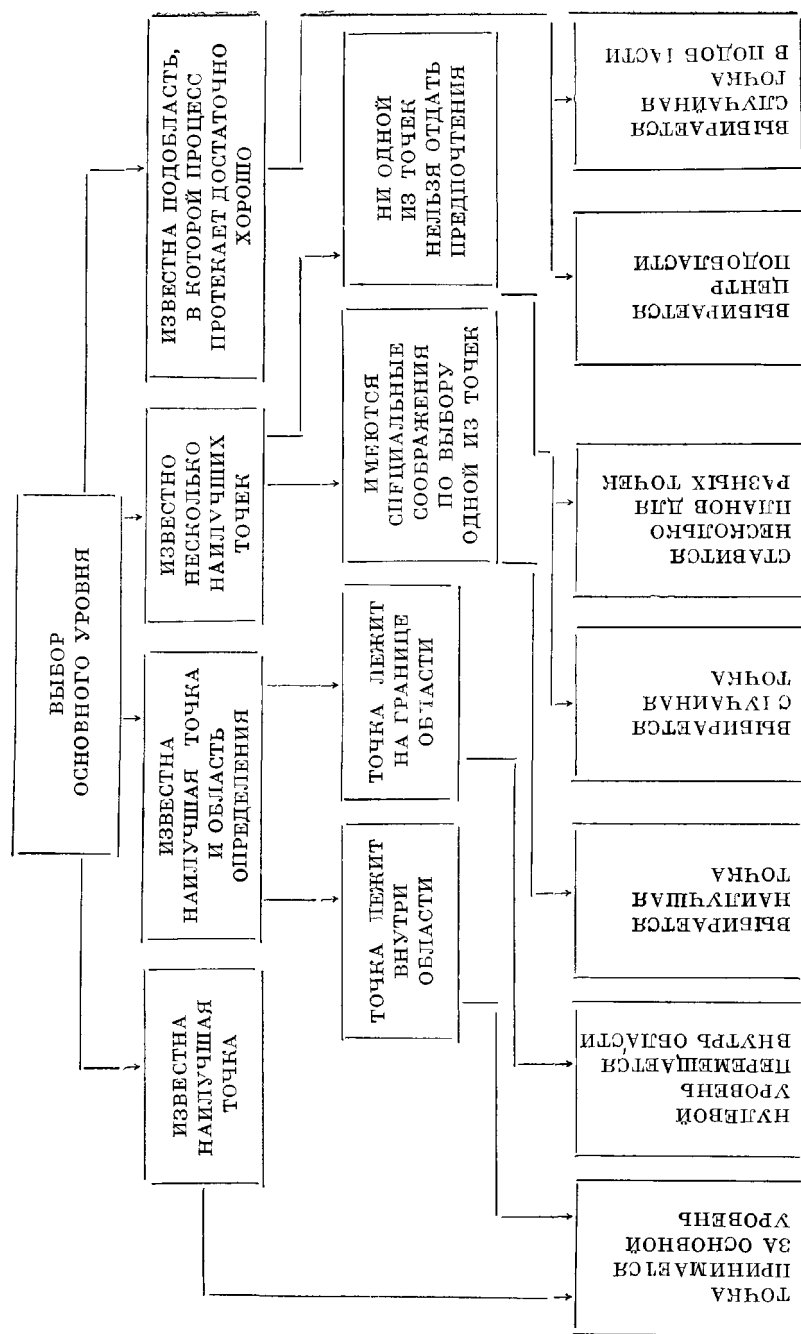
Переходите к пункту III.

3. Вы затрудняетесь в выборе ответа. Обратите внимание на то, что к нулевому уровню предъявляются следующие требования.

1. В этой точке значения параметра оптимизации должны быть наилучшими из всех известных нам значений.

2. Координаты нулевого уровня должны лежать внутри области определения, на некотором расстоянии от границы.

Если эти требования не противоречивы, то выбор ясен. Когда же они вступают в противоречие, мы вынуж-



Блок-схема принятия решений при выборе основного уровня

дены пренебрегать первым из них, так как построить план эксперимента за пределами области определения мы не можем.

Мы надеемся, что теперь, вернувшись к пункту I, вы выберете правильный ответ.

III. 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

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

где

x_j — кодированное значение фактора;

$\tilde{x}_j$ — натуральное значение фактора;

$\tilde{x}_{j0}$ — натуральное значение основного уровня;

J_j — интервал варьирования;

j — номер фактора.

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

Мы предлагаем вам поупражняться с этой формулой.

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

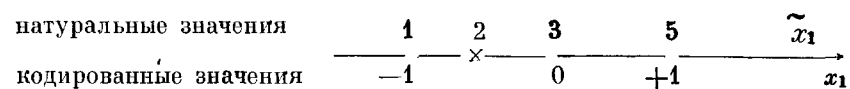
Ниже в натуральном и кодированном масштабах приводятся условия одного из опытов: $\tilde{x}_1 = 2,0$, $\tilde{x}_2 = 20$, $\tilde{x}_3 = 1,25$, $\tilde{x}_4 = 15$; $x_1 = -0,3$, $x_2 = -1$, $x_3 = 0,25$, $x_4 = 0$. Правильно ли найдены кодированные значения факторов?

Не знаю. Обратитесь к пункту 4.

Правильно. Обратитесь к пункту 5.

Неправильно. Обратитесь к пункту 6.

4. Вы не можете ответить, правильно ли проведено кодирование. Давайте рассуждать вместе. Остановимся на первом факторе. Отметим на координатной оси три уровня: нижний, основной и верхний.



Нужно найти кодированное значение для $\tilde{x}_1 = 2,0$. 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 $\tilde{x}_1 = 2,5$ будет $x_1 = -0,25$, для $\tilde{x}_1 = 1,5$ будет $\tilde{x}_1 = -0,75$ и т. д.)

Тот же результат можно получить и по формуле

$$x_1 = \frac{2,0 - 3,0}{2,0} = -0,5.$$

Вернитесь к пункту III, проведите вычисления и выберите правильный ответ.

5. Ваш ответ связан, по-видимому, с арифметическими ошибками.

Вернитесь к пункту III и проведите вычисления более внимательно.

6. Вы ответили правильно. В этом легко убедиться, воспользовавшись формулой перехода.

$$x_1 = \frac{2,0 - 3,0}{2,0} = -0,5; \quad x_2 = \frac{20 - 30}{10} = -1,$$

$$x_3 = \frac{1,25 - 1,5}{1} = -0,25 \quad x_4 = \frac{15 - 15}{10} = 0.$$

Запомните формулу кодирования, она вам будет нужна довольно часто.

IV. 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 подобласть, 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.

ции. 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 зависимостей, 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 не существенно отличается от разброса значений параметра оптимизации в повторных опытах. (Этот разброс, как вы узнаете в гл. 7, 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. 14, 15, 16 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. 14) 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 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?

Широкий интервал варьирования. Обратитесь к пункту 7.

Средний интервал варьирования. Обратитесь к пункту 8.

Решение неоднозначно. Обратитесь к пункту 9.

Не знаю. Обратитесь к пункту 10.

7. Вы выбрали правильный ответ. Эта ситуация обозначена на нашей схеме номером 2. Признаки ситуации определяются стрелками, направленными к данному кружочку. Стрелка, выходящая из кружочка, указывает решение.

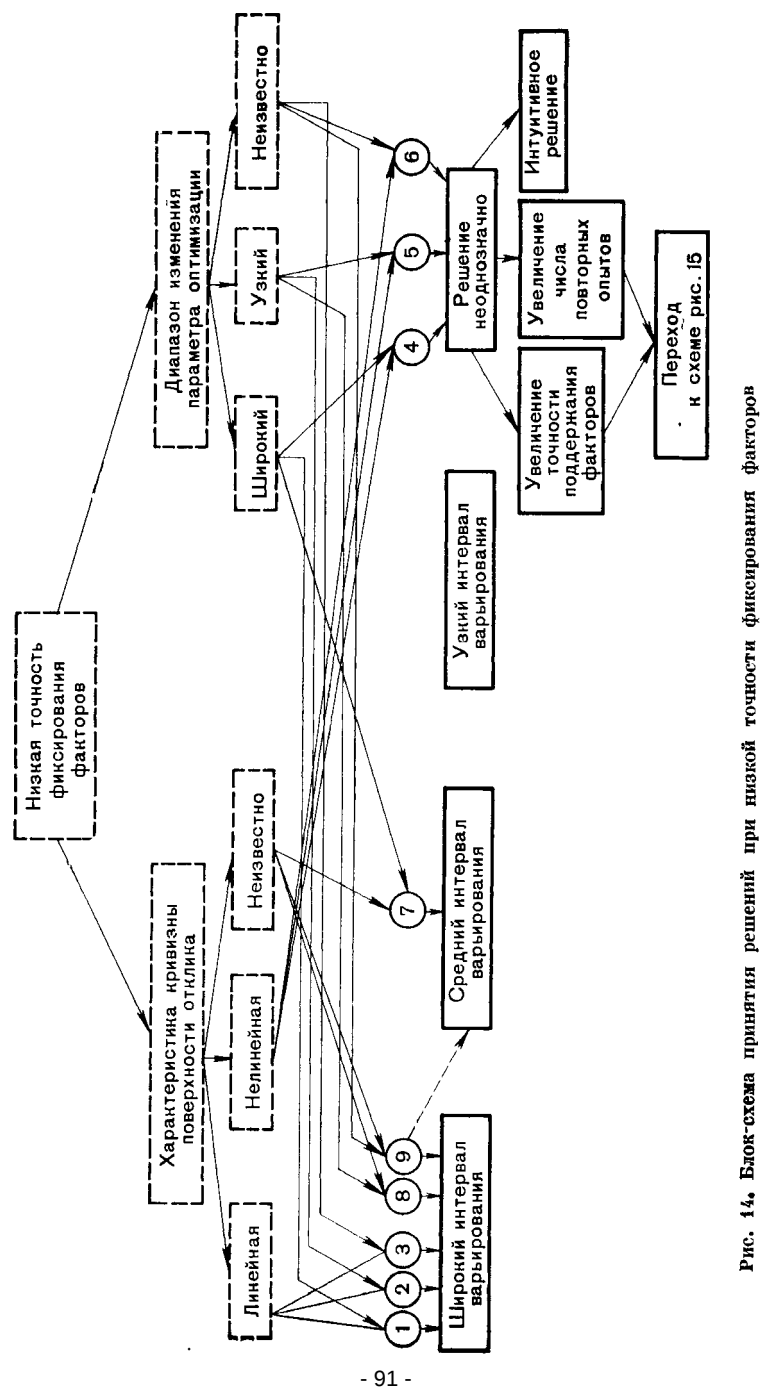


Рис. 14. Блок-схема принятия решений при низкой точности фиксирования факторов

V. 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. 15) для средней точности фиксирования факторов. Если это невозможно, то для принятия решения нет достаточных оснований и оно становится интуитивным.

Эта 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. 15 изображена блок-схема для случая средней точности фиксирования фактора. 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. 16 построена блок-схема для случая высокой точности фиксирования фактора. Сочетание высокой точности с нелинейностью поверхности всегда приводит к выбору узкого интервала. Довольно часто выбирается средний интервал и лишь в двух случаях широкий. В обеих последних блок-схемах отсутствуют неоднозначные решения.

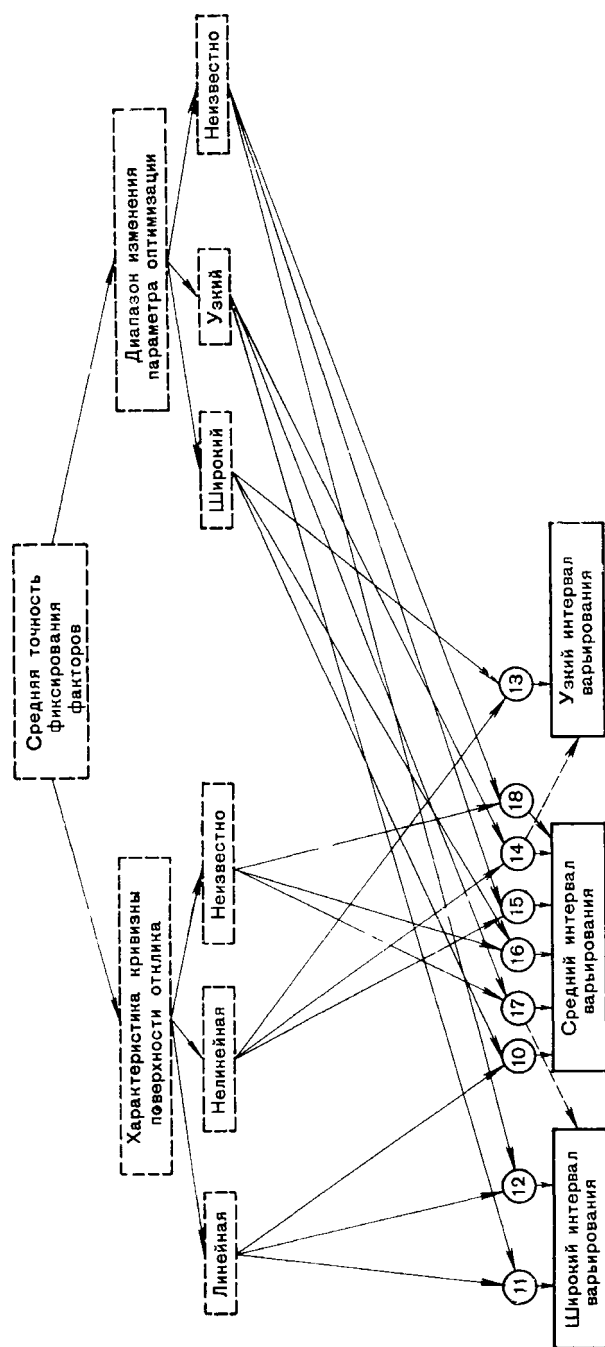


Рис. 15. Блок-схема принятия решений при средней точности фиксации факторов

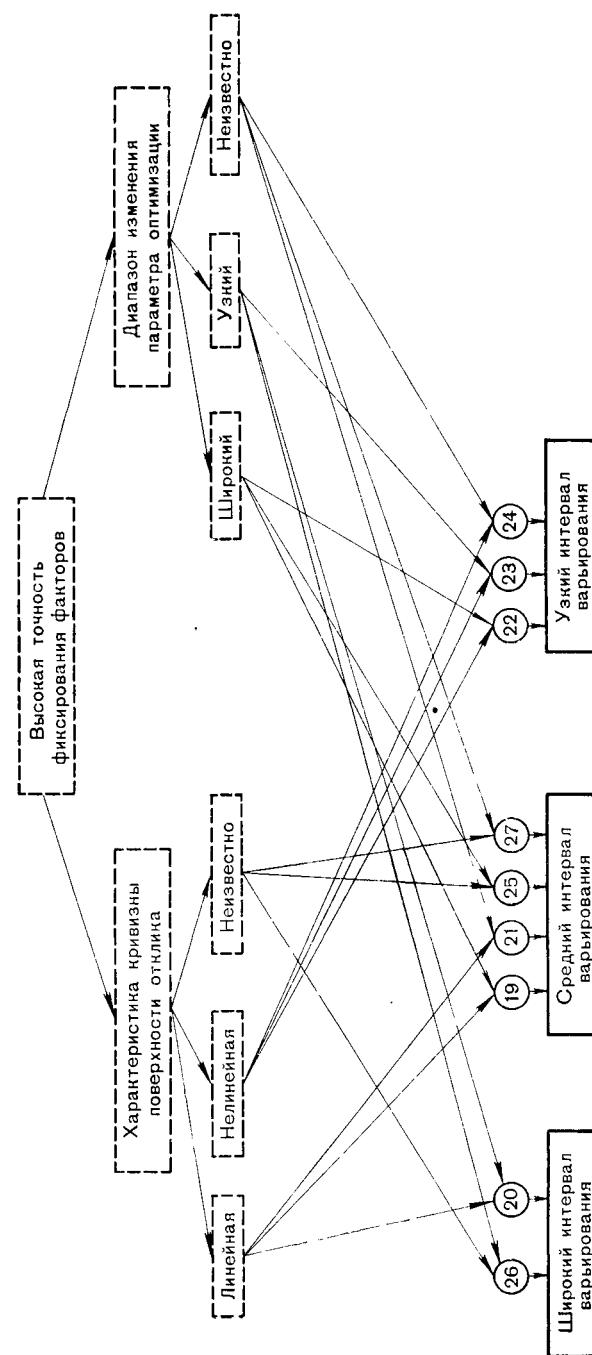


Рис. 16. Блок-схема принятия решений при высокой точности фиксации факторов

Пример 3

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

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. 15). Принимаемое решение — широкий интервал варьирования. Экспериментатор выбрал такие интервалы: $J_1 = 0,5$, $J_2 = 1,0$, что составляет 20% от области определения факторов. Это несомненно широкие интервалы варьирования. Отметим еще, что для $\bar{x}_2$ 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. 9, 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.

Переходите к пункту VI.

8. Вы предпочли средний интервал варьирования. Это неверно. Мы знаем, что поверхность линейна, поэтому расширение интервала вполне возможно. Кроме того, диапазон изменения параметра оптимизации узок, а мы стремимся получить в эксперименте различающиеся значения параметра оптимизации.

Вернитесь к пункту IV и выберите другой ответ.

9. Вы считаете, что решение неоднозначно. Это не так. Мы выбрали случай, в котором решение определяется довольно точно. Сочетание линейной поверхности с узким диапазоном изменения параметра оптимизации определяет тенденцию выбора интервала варьирования.

Вернитесь к пункту IV и подумайте о другом ответе.

10. Так как вы не знаете ответа, проведем рассуждения вместе. Вы помните, что ситуации на схеме обозначены кру-

жочками с порядковыми номерами. Они задаются двумя признаками: кривизной поверхности отклика и диапазоном изменения параметра оптимизации. Стрелки, ведущие к кружочку, показывают, какие градации признаков имеют место в данной ситуации. Какое решение принять? 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.

Мы надеемся, что теперь, вернувшись к пункту IV, вы сможете выбрать правильный ответ.

5.2. Полный факторный эксперимент

VI. 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 .

11. Пользуясь приведенной формулой, заполните следующую таблицу и сравните с ответом ниже.

k	2	3	4	5	6	7	8	9	10
N									

11. Вот правильный ответ.

k	2	3	4	5	6	7	8	9	10
N	4	8	16	32	64	128	256	512	1024

Добейтесь, чтобы таблица была заполнена правильно.

VII. 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 ниже (табл. 5.1).

Таблица 5.1

Матрица планирования эксперимента 2^2

№ опыта	x_1	x_2	y
1	-1	-1	y_1
2	$+1$	-1	y_2
3	-1	$+1$	y_3
4	$+1$	$+1$	y_4

Каждый столбец в матрице планирования называют вектор-столбцом, а каждую строку — вектор-строкой. Таким образом, в табл. 5.1 мы имеем 2 вектор-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. 17). Номера вершин квадрата соответствуют номерам опытов в матрице

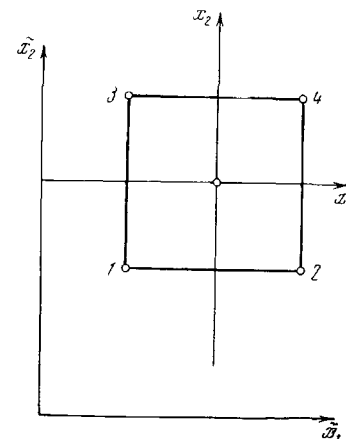


Рис. 17. Геометрическое изображение полного факторного эксперимента 2^2

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. 5.2.

Теперь вместо полной записи матрицы планирования можно пользоваться только буквенными обозначениями.

Таблица 5.2

Матрица планирования эксперимента 2³

№ опыта	x_1	x_2	Буквенные обозначения строк	y
1	-1	-1	(1)	y_1
2	+1	-1	a	y_2
3	-1	+1	b	y_3
4	+1	+1	ab	y_4

Давайте поупражняемся в этой символике. Ниже приведена матрица планирования (табл. 5.3).

Таблица 5.3

№ опыта	x_1	x_2	y
1	+1	+1	y_1
2	-1	+1	y_2
3	-1	-1	y_3
4	+1	-1	y_4

Правильно ли в буквенных обозначениях записан порядок опытов: $ab, a, (1), b$?

Да. Обратитесь к пункту 12.

Нет. Обратитесь к пункту 13.

Не знаю. Обратитесь к пункту 14.

12. Вы ответили, что порядок опытов в буквенных обозначениях записан правильно. Это не так. Запомните правило: буква записывается, если фактор находится на верхнем уровне.

Вернитесь к пункту и выберите другой ответ

13. Вы полагаете, что порядок опытов в буквенных обозначениях записан неправильно. Вы правы, так как правильная запись: $ab, b, (1), a$.

УИИ. Ниже приведена буквенная запись еще одного плана: $c, b, a, abc, (1), bc, ac, ab$.

Постройте матрицу планирования и сравните с ответом в табл. 5.4, пункт 15.

14. Вы ответили, что не знаете, правильно или нет записан порядок опытов в буквенных обозначениях. Напомним вам, что буква записывается, если фактор находится на верхнем уровне.

Посмотрите внимательно, выполняется ли это требование в записи пункта VII, и выберите правильный ответ.

15. Вот правильный ответ (табл. 5.4).

Таблица 5.4

№ 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	(1)	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 2³. Он имеет восемь опытов и включает все возможные комбинации уровней трех факторов.

IX. 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, основанные 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 :

Таблица 5.5

Построение матрицы планирования эксперимента 2^3 с помощью первого приема

№ опыта	x_1	x_2	x_3	y
1	—	—	+	y_1
2	—	+	+	y_2
3	+	—	+	y_3
4	+	+	+	y_4
5	—	—	—	y_5
6	—	+	—	y_6
7	+	—	—	y_7
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. Вы можете убедиться, что таким образом построена матрица в ответе п. 15. 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. 5.5 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 .

Как вы считаете, какое из нижеследующих утверждений верно?

а) Полный факторный эксперимент 2^3 задается координатами вершин куба.

б) Полный факторный эксперимент 2^3 задается координатами вершин тетраэдра.

Верно утверждение а). Обратитесь к пункту 16.

Верно утверждение б) Обратитесь к пункту 17.

Не знаю. Обратитесь к пункту 18.

16. Вы считаете, что геометрической интерпретацией полного факторного эксперимента 2^3 служит куб, координаты вершин которого задают условия опытов. Вы правы.

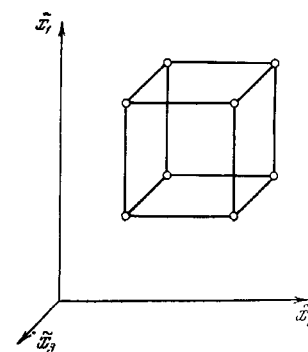


Рис. 18. Геометрическое изображение полного факторного эксперимента 2^3

Действительно, 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. 18. Куб задает область эксперимента, а центр куба является ее центром.

Х. 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.

Переходите к пункту XI.

17. Тетраэдр, которому вы отвели столь почетную роль, к сожалению, не может ее исполнить. Прежде всего потому, что у него меньше вершин, чем опытов в нашей матрице.

Вернитесь к пункту IX и выберите правильное утверждение.

18. Вы не знаете, какое утверждение выбрать.

Мы установили ранее, что для случая двух факторов строкам матрицы соответствуют точки, лежащие в вершинах квадрата. При переходе к трем факторам появляется еще одна координатная ось, на которой фиксированы два уровня нового фактора. Давайте посмотрим случай, когда уровень третьего фактора постоянен, например, -1 . Это равносильно эксперименту для двух факторов, которому, как мы знаем, соответствует квадрат. То же самое будет для другого постоянного уровня третьего фактора $(+1)$. Расстояние между плоскостями этих квадратов равно двум интервалам варьирования. Отсюда и получается куб, координаты вершин которого задают план 2^3 .

Заметим еще, что тетраэдр не подходит, так как у него меньше вершин, чем число строк в матрице. Правда, столько же вершин (8) имеет другая геометрическая фигура — октаэдр, однако его стороны не параллельны координатным плоскостям.

Вернитесь к пункту IX и выберите правильный ответ.

5.3. Свойства полного факторного эксперимента типа 2^k

XI. 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

$$\sum_{i=1}^N x_{ji}x_{ui} = 0, \quad j \neq u, \quad j, u = 0, 1, 2, \dots, k.$$

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 *.

Даны две матрицы планирования:

	x_1	x_2		x_1	x_2
	—	—		—	+
а)	+	—	б)	+	—
	—	+		—	+
	+	+		+	—

* В. В. Налимов, Н. А. Чернова. Ук. соч.

Какая из этих матриц удовлетворяет первым трем свойствам?

Не знаю. Обратитесь к пункту 19.

Матрица б). Обратитесь к пункту 20.

Матрица а). Обратитесь к пункту 21.

19. Вы не знаете, что выбрать. 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 $\sum_{i=1}^N x_{ji}x_{ui} = 0, \quad j \neq u$ выполняется, то в случае б) это не так. Действительно

$$(-1)(+1) + (+1)(-1) + (-1)(+1) + (+1)(-1) = -4 \neq 0.$$

Поэтому вернитесь к пункту XI и выберите правильный ответ.

20. Вы предпочли матрицу б). Вы ошиблись.

Вернитесь к пункту XI и проверьте еще раз, выполняется ли третье свойство.

21. Отдав предпочтение матрице а), вы сделали правильный выбор. Первые два свойства выполняются в обеих

случаях. Третье свойство для матрицы б) не выполняется так как

$$\sum_{i=1}^N x_{ji}x_{ui} = -4 \neq 0.$$

5.4. Полный факторный эксперимент и математическая модель

XII. Давайте еще раз вернемся к матрице 2^2 (табл. 5. 3). 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. 8. 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. 8. 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 значений 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. 8). А если модель нелинейна, 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. 5.6.

Таблица 5.6

№ опыта	x_0	x_1	x_2	x_1x_2	y
1	+1	+1	+1	+1	y_1
2	+1	-1	+1	-1	y_2
3	+1	-1	-1	+1	y_3
4	+1	+1	-1	-1	y_4

Очень важно, что при добавлении столбцов эффектов взаимодействий все рассмотренные свойства матриц планирования сохраняются.

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 *. Все операции по расчету приведены в табл. 5.7.

Таблица 5.7

Расчет коэффициентов регрессии по методу Йетса

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

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. 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.

Таблица 5.8

Расчетная матрица и результаты (см. пример 3)

№ опыта	x_0	x_1	x_2	x_1x_2	Буквенные обозначения	y
1	+1	-1	-1	+1	(I)	95
2	+1	+1	-1	-1	a	90
3	+1	-1	+1	-1	b	85
4	+1	+1	+1	+1	ab	82

* Ч. Хикс Основные принципы планирования эксперимента. М. изд-во «Мир», 1967.

Получено $b_0 = 88,0$, $b_1 = -2,0$, $b_2 = -4,5$, $b_{12} = -0,5$. Проверьте по методу Йетса, правильно ли рассчитаны коэффициенты, и выберите один из ответов.

Правильно. Обратитесь к пункту 22.

Неправильно. Обратитесь к пункту 23.

Не могу получить ответ. Обратитесь к пункту 24.

22. Вы ответили, что коэффициенты рассчитаны правильно. Это не так. Вот расчет

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

Коэффициент взаимодействия b_{12} получился с другим знаком.

Вернитесь к пункту XII и выберите правильный ответ.

23. Вы считаете, что коэффициенты рассчитаны неверно. Действительно, $b_{12} = 0,5$. Остальные коэффициенты рассчитаны правильно.

XIII. Обратите внимание на то, что порядок коэффициентов в последнем столбце (см. п. 22) 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 .

Попробуйте написать матрицу планирования 2^3 с учетом всех возможных взаимодействий.

Правильный ответ вы найдете в пункте 25.

24. Вам трудно получить ответ. Ну что же, давайте проделаем расчет вместе.

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

По сравнению с результатами на стр. 110 коэффициент взаимодействия b_{12} получился с другим знаком.

Вернитесь к пункту XII и выберите правильный ответ.

25. Вот правильный ответ (табл. 5.9).

Таблица 5.9

Полный факторный эксперимент 2^3

№ опыта	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	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

XIV. Вы, по-видимому, испытывали затруднения при построении столбца эффекта взаимодействия $x_1x_2x_3$. 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_1x_2 , $x_2x_3...$), тройные ($x_1x_2x_3$, $x_2x_3x_4...$) 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

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

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:

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

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.

5.5. Резюме

Первой серии 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 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 факторов 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.

Глава шестая

Дробный факторный эксперимент

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.

6.1. Минимизация числа опытов

1. Начнем с самого простого — полного факторного эксперимента 2^k . Напишем еще раз эту хорошо нам известную матрицу (табл. 6.1).

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 ин-

* 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.
O. L. Davies W. A. Hay. The construction and uses of fractional factorial designs in industrial research. Biometrics 6, 1950.

Таблица 61

Полный факторный эксперимент 2³

№ опытов	x_0	x_1	x_2	(x_3) x_1x_2	y
1	+	—	—	+	y_1
2	+	+	—	—	y_2
3	+	—	+	—	y_3
4	+	+	+	+	y_4

тервалах 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 раздельных оценок, которые мы имели в полном факторном эксперименте 2^k. Оценки смешаются следующим образом:

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

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 8 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 2³, требующего, как вы знаете, восьми опытов. Каким бы из них вы воспользовались?

Матрица № 1

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

Ответ в пункте 1

Матрица № 2

N	x_0	x_1	x_2	x_3	y
1	+	—	—	+	y_1
2	+	+	+	+	y_2
3	+	—	+	—	y_3

Матрица № 3

N	x_0	x_1	x_2	x_3	y
1	+	—	—	+	y_1
2	+	+	+	+	y_2
3	+	—	+	—	y_3
4	+	+	—	—	y_4

Ответ в пункте 2.

Ответ в пункте 3.

1. Вы избрали матрицу № 1. Давайте проверим ее свойства. Каждый вектор-столбец матрицы, кроме первого, содержит равное число +1 и —1. Очень хорошо! Это означает, что выполняется условие:

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

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

$$\sum_{i=1}^4 r_{2i} r_{3i} = -4.$$

Значит, 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 2¹. 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 раздельно оценены основные эффекты. Значит, мы потеряли информацию о двух линейных коэффициентах нашей модели. Таким планированием, как вы видите, воспользоваться невозможно.

Вернитесь к пункту 1 и выберите другой ответ.

2. Вам понравилась матрица № 2, вероятно, только потому, что она содержит всего три опыта. Если вы были бы повнимательнее, то вам бросилось бы в глаза, что 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.

Возвратитесь к пункту I и выберите другую матрицу.

3. Вы сделали правильный выбор. 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_1x_2 полного факторного эксперимента 2^3 .

II. Если мы в дополнение к столбцам матрицы № 3 вычислим еще столбцы для произведений x_1x_3 и x_2x_3 , то увидим, что элементы столбца x_1x_3 совпадут с элементами столбца x_2 , а элементы столбца x_2x_3 — с элементами столбца x_1 . Найденные нами коэффициенты будут оценками для совместных эффектов

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

Такое планирование нас вполне устраивает. Мы смешали эффекты взаимодействия с основными эффектами. (Но все основные эффекты оцениваются отдельно 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.

6.2. Дробная реплика

III. 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 $-x_1x_2$, 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 полуреplik можно получить отдельные оценки для линейных эффектов и эффектов взаимодействия, как и в полном факторном эксперименте 2^3 . Объединение этих двух полуреplik и есть полный факторный эксперимент 2^3 .

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

Т а б л и ц а 6.2

Условные обозначения дробных реplik и количество опытов

Количество факторов	Дробная реплика	Условное обозначение	Количество опытов для дробной реплики	Количество опытов 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

IV. Теперь мы предлагаем вам поупражняться и дать условные обозначения: 1/8-реплики от 2^6 , 1/16-реплики от 2^7 , 1/2-реплики от 2^5 , 1/4-реплики от 2^6 , 1/8-реплики от 2^7 , 1/16-реплики от 2^8 , а также дробных планов из 16 опытов. Подсчитайте для всех этих случаев количество опытов для дробных реплик и для полного факторного эксперимента.

Правильные ответы вы найдете в табл. 6.2.

6.3. Выбор полуреplik. Генерирующие соотношения и определяющие контрасты

V. 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. 6.3).

Таблица 6.3

Две полуреплики 2^{3-1}

I. $x_3 = x_1x_2$					II. $x_3 = -x_1x_2$				
№ опытов	x_1	x_2	x_3	$x_1x_2x_3$	№ опытов	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 о п р е д е л я ю щ и м к о н т р а с т о м. 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

$$b_1 \rightarrow \beta_1 + \beta_{23},$$

$$b_2 \rightarrow \beta_2 + \beta_{13},$$

$$b_3 \rightarrow \beta_3 + \beta_{12}.$$

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^{3-1}_{III} .

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

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

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 реп-

лику с наибольшей разрешающей способностью, так как тройные взаимодействия обычно менее важны, чем парные. Если существует информация об эффектах взаимодействий, 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^{4-1}_{IV} . 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 необходимо выбрать полуреплику 2^{4-1} . Какие полуреплики, по вашему мнению, обладают большей разрешающей способностью?

Полуреплики, заданные определяющими контрастами

$$1 = x_1x_2x_3x_4, 1 = -x_1x_2x_3x_4$$

Обратитесь к пункту 4.

Полуреплики, заданные генерирующими соотношениями

$$x_4 = x_1x_2, x_1 = -x_1x_2.$$

Обратитесь к пункту 5.

4. Вы избрали полуреплики, заданные определяющими контрастами $1 = x_1x_2x_3x_4$ и $1 = -x_1x_2x_3x_4$.

Совместные оценки здесь определяются соотношениями:

$$\begin{aligned} x_1 &= x_2x_3x_4, & x_1 &= -x_2x_3x_4, \\ x_2 &= x_1x_3x_4, & x_2 &= -x_1x_3x_4, \end{aligned}$$

$$\begin{aligned} x_3 &= x_1x_2x_4, & x_3 &= -x_1x_2x_4, \\ x_4 &= x_1x_2x_3, & x_4 &= -x_1x_2x_3, \\ x_1x_2 &= x_3x_4, & x_1x_2 &= -x_3x_4, \\ x_1x_3 &= x_2x_4, & x_1x_3 &= -x_2x_4, \\ x_1x_4 &= x_2x_3, & x_1x_4 &= -x_2x_3. \end{aligned}$$

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

V. 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^{4-1}_{IV} .

Пример 1

Этот пример относится к планированию эксперимента для отыскания оптимальных условий получения нового полимерного серу-содержащего антиоксиданта, синтезированного превращением высокомолекулярного полистирола с серой *. 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. 6. 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.

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

* Е. В. Маркова, П. Фарка, И. П. Борисова, А. А. Донцов, Б. А. Догадкин. Пластические массы, 1969, № 1,

Таблица 6.4

Уровни факторов и интервалы варьирования

Факторы	Уровни факторов			Интервал варьирования
	-1	0	+1	
$\tilde{x}_1$ — температура реакционной среды, °C	200	220	240	20
$\tilde{x}_2$ — дозировка серы в исходной смеси, вес. ч.	3	6	9	3
$\tilde{x}_3$ — время реакции, мин	40	100	160	60
$\tilde{x}_4$ — дозировка антиоксиданта в полипропилене, вес. ч. . . .	1	2	3	1

Таблица 6.5

Матрица планирования 2_{IV}^{4-1}

N	x_0	x_1	x_2	x_3	x_4	$x_1x_2 =$ $= x_3x_4$	$x_1x_3 =$ $= x_2x_4$	$x_2x_3 =$ $= x_1x_4$	y
1	+	+	+	—	—	+	—	—	10
2	+	—	—	—	—	+	+	+	9
3	+	+	—	—	+	—	—	+	15
4	+	—	+	—	+	—	+	—	25
5	+	+	+	+	+	+	+	+	26
6	+	—	—	+	+	+	—	—	14
7	+	+	—	+	—	—	+	—	5
8	+	—	+	+	—	—	—	+	20
b_i	15,00	-1,50	4,75	0,75	4,50	-0,75	0,75	2,0	

Матрица планирования, результаты эксперимента и коэффициенты регрессии показаны в табл. 6. 5.

Анализ результатов и поиск оптимальных условий приводятся в последующих главах.

Переходите к пункту VI.

5. Вы выбрали полуреплики, заданные генерирующими соотношениями $x_4 = x_1x_2$ и $x_4 = -x_1x_2$. В этом случае

определяющими контрастами 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:

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

Разрешающая способность этих полуреплик ниже, чем у планов с разрешающей способностью IV, с помощью которых линейные эффекты определяются независимо от парных взаимодействий. Избранные вами полуреплики имеют в каждой строке как четные, так и нечетные комбинации букв. Такие полуреплики не являются главными.

Вы можете быть правы при выборе такой полуреплики, если располагаете априорной информацией о большей значимости тройных взаимодействий по сравнению с парными или о незначимости трех парных взаимодействий x_2x_4 , x_1x_4 , x_1x_2 . Тогда план великолепен!

В противном случае вы ошибаетесь.

Вернитесь к пункту IV и выберите другой ответ.

VI. 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 6 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

$$1 = x_1x_2x_3x_4x_5 \text{ и } 1 = -x_1x_2x_3x_4x_5.$$

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. Каким полурепликам вы хотите отдать предпочтение?

Полуреплика, заданная генерирующим соотношением

$$x_5 = x_1 x_3 x_4.$$

Обратитесь к пункту 6.

Полуреплика, заданная генерирующим соотношением

$$x_5 = x_1 x_2 x_3 x_4.$$

Обратитесь к пункту 7.

6. Вы предпочли полуреплику, заданную генерирующим соотношением $x_5 = x_1 x_3 x_4$. В этом случае $1 = x_1 x_3 x_4 x_5$, следовательно

$$\begin{aligned} b_1 &\rightarrow \beta_1 + \beta_{345}, & b_{12} &\rightarrow \beta_{12} + \beta_{2345}, \\ b_2 &\rightarrow \beta_2 + \beta_{12345}, & b_{13} &\rightarrow \beta_{13} + \beta_{45}, \\ b_3 &\rightarrow \beta_3 + \beta_{145}, & b_{14} &\rightarrow \beta_{14} + \beta_{35}, \\ b_4 &\rightarrow \beta_4 + \beta_{135}, & b_{15} &\rightarrow \beta_{15} + \beta_{34}, \\ b_5 &\rightarrow \beta_5 + \beta_{134}, & b_{23} &\rightarrow \beta_{23} + \beta_{1245}, \end{aligned}$$

$$\begin{aligned} b_{24} &\rightarrow \beta_{24} + \beta_{1235}, & b_{123} &\rightarrow \beta_{123} + \beta_{245}, \\ b_{25} &\rightarrow \beta_{25} + \beta_{1234}, & b_{124} &\rightarrow \beta_{124} + \beta_{235}, \\ b_{34} &\rightarrow \beta_{34} + \beta_{15}, & b_{125} &\rightarrow \beta_{125} + \beta_{234}, \\ b_{35} &\rightarrow \beta_{35} + \beta_{14}, & b_{134} &\rightarrow \beta_{134} + \beta_5, \\ b_{45} &\rightarrow \beta_{45} + \beta_{13}, & & \text{и т. д.} \end{aligned}$$

Вы должны согласиться с тем, что смешивание основных эффектов с тройными взаимодействиями, когда существуют эффекты взаимодействия более высокого порядка, нельзя признать наилучшим, если нет специальных соображений.

Вернитесь к пункту VI и выберите другую систему смешивания.

7. Вы считаете, что целесообразно воспользоваться полурепликой, заданной генерирующим соотношением

$x_5 = x_1 x_2 x_3 x_4$ и, следовательно, определяющим контрастом $1 = x_1 x_2 x_3 x_4 x_5$. В этом случае коэффициенты определяют такие смешанные оценки:

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

Вы 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 репликой ответа 6. Вы поступили совершенно правильно.

VII. Теперь мы предлагаем вам перечислить все возможные варианты выбора плана 2^{5-1} .

Правильный ответ вы найдете в пункте 8.

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

- | | |
|------------------------|--------------------------------|
| 1. $x_5 = x_1 x_2$, | 12. $x_5 = -x_3 x_4$, |
| 2. $x_5 = -x_1 x_2$, | 13. $x_5 = x_1 x_2 x_3$, |
| 3. $x_5 = x_1 x_3$, | 14. $x_5 = -x_1 x_2 x_3$, |
| 4. $x_5 = -x_1 x_3$, | 15. $x_5 = x_1 x_3 x_4$, |
| 5. $x_5 = x_1 x_4$, | 16. $x_5 = -x_1 x_3 x_4$, |
| 6. $x_5 = -x_1 x_4$, | 17. $x_5 = x_1 x_3 x_4$, |
| 7. $x_5 = x_2 x_3$, | 18. $x_5 = -x_1 x_2 x_4$, |
| 8. $x_5 = -x_2 x_3$, | 19. $x_5 = x_2 x_3 x_4$, |
| 9. $x_5 = x_2 x_4$, | 20. $x_5 = -x_2 x_3 x_4$, |
| 10. $x_5 = -x_2 x_4$, | 21. $x_5 = x_1 x_2 x_3 x_4$, |
| 11. $x_5 = x_3 x_4$, | 22. $x_5 = -x_1 x_2 x_3 x_4$. |

Мы не станем рассматривать выбор полуреplik 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 экспериментатора 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 полуреплик в определяющий контраст надо включать наибольшее число факторов.

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

6.4. Выбор 1/4-реплик.

Обобщающий определяющий контраст

VIII. Дорогой читатель, приступая к этому параграфу, вам придется запастись еще бóльшим 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-реплик.

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.

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

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-реплик, можно получить линейные эффекты, не смешанные с парными взаимодействиями.

Постарайтесь ответить на следующий вопрос. Какую из 1/4 реплик вы хотели бы избрать для дополнения 1/4-реплики, заданной определяющими контрастами $1 = x_1x_3x_4$ и $1 = x_1x_2x_3x_5$?

Реплику, заданную генерирующими соотношениями

$$x_4 = -x_1x_2 \text{ и } x_5 = -x_1x_2x_3.$$

Обратитесь к пункту 9.

Реплику, заданную генерирующими соотношениями

$$x_4 = -x_1x_3 \text{ и } x_5 = x_1x_2x_3.$$

Обратитесь к пункту 10.

Реплику, заданную генерирующими соотношениями

$$x_4 = x_1x_3 \text{ и } x_5 = x_1x_2x_3.$$

Обратитесь к пункту 11.

Не знаю. Обратитесь к пункту 12.

9. Вы хотите воспользоваться 1/4-репликой, заданной генерирующими соотношениями

$$x_4 = -x_1x_2 \text{ и } x_5 = -x_1x_2x_3.$$

В этом случае 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.

Вернитесь к пункту VIII и выберите другую 1/4-реплику.

10. Вы выбрали 1/4-реплику, заданную генерирующими соотношениями $x_4 = -x_1x_3$ и $x_5 = x_1x_2x_3$, и поступили совершенно правильно. Обобщающим контрастом в этом случае является $1 = -x_1x_3x_4 = x_1x_2x_3x_5 = -x_2x_4x_5$. Система смешивания будет

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

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

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

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

Переходите к пункту IX.

11. Вам понравилась 1/4-реплика, заданная генерирующими соотношениями $x_4 = x_1x_3$ и $x_5 = -x_1x_2x_3$ и обобщающим определяющим контрастом

$$1 = x_1x_3x_4 = -x_1x_2x_3x_5 = -x_2x_4x_5.$$

В этом случае получается следующая система смешивания:

$$\begin{aligned} x_1 &= x_3x_4 = -x_2x_3x_5 = -x_1x_2x_4x_5, \\ x_2 &= x_1x_2x_3x_4 = -x_1x_3x_5 = -x_4x_5, \\ x_3 &= x_1x_4 = -x_1x_2x_5 = -x_2x_3x_4x_5, \\ x_4 &= x_1x_3 = -x_1x_2x_3x_4x_5 = -x_2x_5, \end{aligned}$$

$$x_5 = x_1 x_3 x_4 x_5 = -x_1 x_2 x_3 = -x_2 x_4,$$

$$x_6 = x_2 x_3 x_4 = -x_3 x_5 = -x_1 x_4 x_5,$$

$$x_7 = x_3 x_4 x_5 = -x_2 x_3 = -x_1 x_2 x_4.$$

При сложении двух 1/4-реплик получаем

$$\begin{aligned} b_1 &\rightarrow \beta_1 - \beta_{34}, & b_4 &\rightarrow \beta_4 + \beta_{13}, \\ b_2 &\rightarrow \beta_2 + \beta_{1234}, & b_5 &\rightarrow \beta_5 + \beta_{1345}, & b_{15} &\rightarrow \beta_{15} + \beta_{345} \\ b_3 &\rightarrow \beta_3 + \beta_{14}, & b_6 &\rightarrow \beta_6 + \beta_{234} \quad \text{и т. д.} \end{aligned}$$

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.

Вернитесь к пункту VIII и выберите другой ответ.

12. Вы ответили, что не знаете. Не огорчайтесь. Мы поможем вам принять нужное решение.

Итак, мы рассматриваем 1/4-реплику, заданную обобщающим определяющим контрастом

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

При этом линейные эффекты смешаны с эффектами взаимодействий первого, второго и третьего порядков. Как построить эту 1/4-реплику 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 определяют парные взаимодействия в совместных оценках для линейных эффектов. Этот случай рассмотрен в пункте 10.

Однако можно представить себе и такой случай, когда целесообразно освободить линейные эффекты от эффектов взаимодействия второго порядка и только часть из линейных эффектов (например, b_2 и b_5) от парных взаимодействий. Тогда нужно выбрать 1/4-реплику таким обра-

зом, чтобы в обобщающем определяющем контрасте произведение четырех членов имело отрицательный знак, так как это произведение определяет тройные взаимодействия в совместных оценках для линейных эффектов. Подробно это описано в пункте 11.

После того как вы познакомитесь с этими двумя вариантами переходите к пункту IX.

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

Пример 2.

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. 6.6.

Параметром оптимизации служил выход реакции в процентах.

Таблица 6.6

Уровни факторов и интервалы варьирования

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

В табл. 6.7 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

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

Таблица 6.7

Матрица планирования 2^{5-2}

№ 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_i	52,300	—1,755	5,050	0,575	—2,400	0,325	

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

$$b_1 \rightarrow \beta_1 - \beta_{25} + \beta_{234} - \beta_{1345}$$

$$b_2 \rightarrow \beta_2 - \beta_{15} + \beta_{134} - \beta_{2345} \quad b_{13} \rightarrow \beta_{13} + \beta_{24} - \beta_{235} - \beta_{145}$$

$$b_3 \rightarrow \beta_3 - \beta_{45} + \beta_{124} - \beta_{1235} \quad b_{14} \rightarrow \beta_{14} - \beta_{23} - \beta_{245} - \beta_{135}$$

$$b_4 \rightarrow \beta_4 - \beta_{35} - \beta_{123} - \beta_{1245}$$

$$b_5 \rightarrow \beta_5 - \beta_{12} - \beta_{34} - \beta_{12345}$$

Была намечена вторая серия опытов для случая, если поиск оптимальных условий окажется неэффективным. Обобщающий определяющий контраст второй 1/4-реплики имеет вид $1 = x_1x_2x_3x_4 = x_1x_2x_5 = x_3x_4x_5$. Он был выбран так, чтобы знаки тройных произведений оказались противоположными знакам тех же произведений в первой четверть-реплике.

Анализ результатов приведен в гл. 8.

6.5. Реплики большой дробности

Х. 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^{5-3} можно воспользоваться вектор-столбцами трех взаимодействий, например, так:

$$1) \quad x_4 = x_1x_2, \quad x_5 = x_1x_3, \quad x_6 = x_2x_3;$$

$$2) \quad x_4 = x_1x_3, \quad x_5 = x_2x_3, \quad x_6 = x_1x_2x_3;$$

$$3) \quad x_4 = x_1x_2, \quad x_5 = x_2x_3, \quad x_6 = x_1x_2x_3,$$

$$4) \quad x_4 = x_1x_2, \quad x_5 = x_1x_3, \quad x_6 = x_1x_2x_3.$$

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_1x_2 , то нужно выбрать второе решение, если x_1x_3 — третье, а если x_2x_3 — четвертое.

Допустим, мы избрали 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

$$x_4 = x_1x_2x_3, \quad x_5 = x_1x_2, \quad x_6 = x_1x_3.$$

XI. Определите, как смешаны линейные эффекты. Ограничьтесь парными и тройными взаимодействиями.

Правильный ответ найдете в пункте 13.

13. Для 1/8-реплики с генерирующими соотношениями

$$x_4 = x_1x_2x_3, \quad x_5 = x_1x_2 \quad \text{и} \quad x_6 = x_1x_3$$

имеем следующие определяющие контрасты:

$$1 = x_1x_2x_3x_4, \quad 1 = x_1x_2x_5 \quad \text{и} \quad 1 = x_1x_3x_6.$$

Если попарно перемножить эти определяющие контрасты, то получим

$$1 = x_3x_4x_5, \quad 1 = x_2x_4x_6 \text{ и } 1 = x_2x_3x_5x_6.$$

Произведение трех определяющих контрастов равно

$$1 = x_1x_4x_5x_6.$$

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

$$1 = x_1x_2x_3x_4 = x_1x_2x_5 = x_1x_3x_6 = x_3x_4x_5 = x_2x_4x_6 = \\ = x_2x_3x_5x_6 = x_1x_4x_5x_6.$$

Получается следующая система смешивания:

$$b_1 \rightarrow \beta_1 + \beta_{25} + \beta_{36} + \beta_{234} + \beta_{456},$$

$$b_2 \rightarrow \beta_2 + \beta_{15} + \beta_{46} + \beta_{134} + \beta_{356},$$

$$b_3 \rightarrow \beta_3 + \beta_{16} + \beta_{45} + \beta_{124} + \beta_{256},$$

$$b_4 \rightarrow \beta_4 + \beta_{35} + \beta_{26} + \beta_{123} + \beta_{156},$$

$$b_5 \rightarrow \beta_5 + \beta_{12} + \beta_{34} + \beta_{236} + \beta_{146},$$

$$b_6 \rightarrow \beta_6 + \beta_{13} + \beta_{24} + \beta_{235} + \beta_{145}.$$

XII. Рассмотрим пример 1/16-реплики от 2⁷.

Пример 3.

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. 6.8. Использовалась 1/16 часть от полного факторного эксперимента 2⁷. Это дает возможность сократить число опытов до 8 вместо 128.

В табл. 6.9 приведена матрица планирования и соответствующие коэффициенты линейного уравнения.

Реплика задана генерирующими соотношениями

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

Для них имеем следующие определяющие контрасты:

$$1 = x_1x_2x_4; \quad 1 = x_1x_3x_5; \quad 1 = x_2x_3x_6;$$

$$1 = x_1x_2x_3x_7.$$

Таблица 6.8

Уровни факторов и интервалы варьирования

Факторы	Уровни факторов			Интервал варьирования
	-1	0	+1	
$\tilde{x}_1$ — количество NaOH, вносимого в реакц. массу до прилива вещества <i>a</i> , г/мол	0,0075	0,0180	0,0285	0,0105
$\tilde{x}_2$ — способ поддержания pH (вид раствора)	18%-ный раствор NaOH в метаноле	—	40%-ный водный раствор NaOH	—
$\tilde{x}_3$ — время прилива вещества <i>a</i> и раствора NaOH, час . . .	3	4,5	6	1,5
$\tilde{x}_4$ — время выдержки, час	1	2	3	1,0
$\tilde{x}_5$ — температура, °C	20	25	30	5,0
$\tilde{x}_6$ — весовое соотношение вещества <i>a</i> и метанола, г/г . . .	1:3	1:3,5	1:4	1:0,5
$\tilde{x}_7$ — молярное соотношение вещества <i>a</i> и вещества <i>e</i> , г/м/г/м	1:1	1:1,1	1:1,2	1:0,1

Таблица 6.9

Матрица планирования 2⁷⁻⁴ и коэффициенты

№ 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_i	23,590	+1,065	-5,760	+0,191	+3,210	+0,066	-1,320	-1,780	

Обобщающий определяющий контраст

$$\begin{aligned} 1 &= x_1x_2x_4 = x_1x_3x_5 = x_2x_3x_6 = x_1x_2x_3x_7 = x_2x_3x_4x_5 = \\ &= x_1x_3x_4x_6 = x_3x_4x_7 = x_1x_2x_5x_6 = x_2x_5x_7 = x_1x_6x_7 = \\ &= x_4x_5x_6 = x_1x_4x_5x_7 = x_2x_4x_6x_7 = x_3x_5x_6x_7 = x_1x_2x_3x_4x_5x_6x_7. \end{aligned}$$

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

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

$$\begin{aligned} b_1 &\rightarrow \beta_1 + \beta_{24} + \beta_{35} + \beta_{67}, & b_5 &\rightarrow \beta_5 + \beta_{13} + \beta_{46} + \beta_{27}, \\ b_2 &\rightarrow \beta_2 + \beta_{14} + \beta_{36} + \beta_{57}, & b_6 &\rightarrow \beta_6 + \beta_{23} + \beta_{45} + \beta_{17}, \\ b_3 &\rightarrow \beta_3 + \beta_{15} + \beta_{26} + \beta_{47}, & b_7 &\rightarrow \beta_7 + \beta_{34} + \beta_{25} + \beta_{16}, \\ b_4 &\rightarrow \beta_4 + \beta_{12} + \beta_{56} + \beta_{37}, \end{aligned}$$

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 специальным 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 2^3 , 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 4 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 $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7$ и число степеней свободы равно нулю. При числе факторов от 9 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 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 *.

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 4.

Изучалось влияние следующих 15 факторов на скорость хлорирования титановых шлаков в расплаве *:

$\bar{x}_1$ — расход хлора;

* В. В. Налимов Н. А. Чернова. Ук. соч.

$\tilde{x}_2$ — температура;
 $\tilde{x}_3$ — концентрация углерода в расплаве;
 $\tilde{x}_4$ — концентрация TiO_2 в расплаве;
 $\tilde{x}_5$ — $\tilde{x}_{11}$ — компоненты, определяющие состав шлака;
 $\tilde{x}_{12}$ — $\tilde{x}_{15}$ — 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. 6.10 представлены уровни факторов и интервал варьирования, матрица планирования и результаты экспериментов. В последней строке помещены коэффициенты.

6.6. Резюме

Дробные реплики находят широкое применение при получении линейных моделей. Целесообразность их применения возрастает с ростом количества факторов. В табл. 6.10 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 взаимодействий, а также от умелой стратегии экспериментирования в случае значимости некоторых взаимодействий. Априорные сведения о взаимодействиях могут оказать большую услугу экспериментатору.

* П. П. Хомяков, Ю. П. Адлер, В. В. Налимов. Выявление факторов, влияющих на скорость хлорирования титановых шлаков в расплаве. — Заводская лаборатория, 1963, 29, № 1.

Таблица 6.10

Планирование опытов при изучении процесса хлорирования титановых шлаков

Основной уровень	16,6	750	3,5	2	5,65	1,0	8,3	
Интервал варьирования	5,2	50	1,5	1	1,35	0,5	1,7	
Верхний уровень .	21,8	800	5	3	7,0	1,5	10,0	
Нижний уровень . .	11,4	700	2	1	4,3	0,5	6,6	

Кодовые обозначения переменных	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Опыт 1	+	—	—	—	—	+	+	+
» 2	+	+	—	—	—	—	—	—
» 3	+	—	+	—	—	—	+	+
» 4	+	+	+	—	—	+	—	—
» 5	+	—	—	+	—	+	—	+
» 6	+	+	—	+	—	—	+	—
» 7	+	—	+	+	—	—	—	+
» 8	+	+	+	+	—	+	+	—
» 9	+	—	—	—	+	+	+	—
» 10	+	—	+	—	+	—	+	—
» 11	+	+	+	—	+	+	—	+
» 12	+	—	—	+	+	+	—	—
» 13	+	+	—	+	+	—	+	+
» 14	+	—	+	+	+	—	—	—
» 15	+	+	+	+	+	+	+	+
» 16	+	+	—	—	+	—	—	+

Коэффициенты	75,25	8,69	3,75	1,20	28,79	—5,09	8,38	7,40
--------------	-------	------	------	------	-------	-------	------	------

При построении дробных реплик используют следующее правило: для того чтобы сократить число опытов, вводя в планирование новый фактор, 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 пользуются большой популярностью, так как позволяют производить расчет коэффициентов уравнения так же просто, как и в случае полного факторного эксперимента.

При применении дробных реплик линейные эффекты

1	6,65	5,7	4,7	7	7	8	1:1
1	1,35	1,3	1,3	3	3	2	1:0,5
2	8,0	7,0	6,0	10	10	10	1:0,5
0	5,3	4,4	3,4	4	4	6	1:2

x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	y_1	y_2	$\bar{y}$
+	+	+	—	—	—	—	+	39,2	48,5	43,85
+	+	+	+	+	+	—	—	48,0	47,5	47,75
—	—	+	+	+	—	+	—	44,7	48,5	46,6
—	—	+	—	—	+	+	+	47,2	43,7	45,45
—	+	—	+	—	+	+	—	41,7	33,4	37,55
—	+	—	—	+	—	+	+	40,5	48,2	44,35
+	—	—	—	+	+	—	+	48,2	49,2	48,7
+	—	—	+	—	—	—	—	49,4	49,5	49,45
+	—	—	—	+	+	+	—	98,5	95,0	96,75
—	+	—	+	—	+	—	+	115,0	92,5	103,75
—	+	—	—	+	—	—	—	110,0	102,5	106,75
—	—	+	+	+	—	—	+	50,4	47,0	48,7
—	—	+	—	—	+	—	—	151,0	151,1	151,05
+	+	+	—	—	—	+	—	93,5	103,7	98,6
+	+	+	+	+	+	+	+	127,0	123,5	125,25
+	—	—	+	—	—	+	+	85,0	103,0	94,0

1,29 1,66 1,65 —5,12 —3,71 7,78 —0,68 —4,99

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: $x_1 = x_2x_3$, $x_2 = x_1x_3$ и $x_3 = x_1x_2$, то определяющий контраст будет $1 = x_1x_2x_3$.

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 получаются 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.

Глава седьмая

Проведение эксперимента

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.

7.1. 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. Затем из них выбрать «главный» параметр, по которому будет проводиться крутое восхождение, а прочие параметры считать ограничениями. 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 имелась 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 представлять 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 (в %),
 y_2 — качество сульфадимизина.

Формулировка задачи оптимизации

Достижение максимального выхода сульфадимизина при качестве, удовлетворяющем требованиям фармакопей:

y_1 — является параметром оптимизации ($y_1 \rightarrow 100\%$),

y_2 — служит ограничением.

Качество продукта определяется по процентному содержанию сульфадимизина в получаемом продукте и по его температуре плавления. Согласно требованиям фармакопеи содержание основного вещества в получаемом продукте должно быть не менее 99%, а температура плавления должна находиться в пределах 196—200° С

$y'_2 \geq 99\%$, $196^\circ \text{C} < y''_2 < 200^\circ \text{C}$.

Страница 2

Факторы, определяющие процесс

$\tilde{x}_1$ — время реакции (час),

$\tilde{x}_2$ — содержание ацетилацетона в реакционной массе (%),

$\tilde{x}_3$ — содержание уксусной кислоты в реакционной массе (%),

$\tilde{x}_4$ — температура реакционной массы (°C),

$\tilde{x}_5$ — качество ацетилацетона (%),

$\tilde{x}_6$ — качество сульгина (%).

Выбор варьируемых факторов

Принято решение изменять в опытах первые три фактора. Качество ацетилацетона и сульгина решено поддерживать постоянным (таким, как на действующем производстве)

$\tilde{x}_5 = 90\%$; $\tilde{x}_6 = 98\%$.

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

Страница 3

Выбор технологии

Сульгин загружается одновременно с ацетилацетоном и уксусной кислотой. Реакция проводится при перемешивании реакционной смеси. Производится непрерывный отгон воды.

Необходимые анализы

Анализ исходного сырья: 1) ацетилацетона, 2) сульгина, 3) уксусной кислоты. (Следует описание методик анализа.)

Анализы получаемых продуктов: 1) сульфадимизина в осадке, 2) сульфадимизина в фильтрате, 3) сульгина в фильтрате. (Следует описание методик анализа.)

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). По поводу оптимального времени реакции в лабораторных условиях имеются противоречивые данные. Так, в отчете № 1 указано, что опыты проводились при $\tilde{x}_1 = 21$ час, затем время уменьшили до 12 час.

Уменьшение времени не снизило существенно выход реакции. В отчете № 2 описываются опыты с различным временем: 18, 24 и 30 час. Наилучший выход получен при $\tilde{x}_1 = 24$ часа

Влияние избытка ацетилацетона (x_2). По поводу влияния избытка ацетилацетона от стехиометрического соотношения также имеются противоречивые данные. В одном отчете указано, что содержание ацетилацетона сверх 10% является нецелесообразным, в другом — оптимальным считается 40% избытка ацетилацетона.

Таблица 7.1

Уровни факторов и интервалы варьирования

Факторы	-1	0	+1	I	Ра змерность
$\tilde{x}_1$	16	18	20	2	час
$\tilde{x}_2$	20	24	28	4	%
$\tilde{x}_3$	12	15	18	3	%

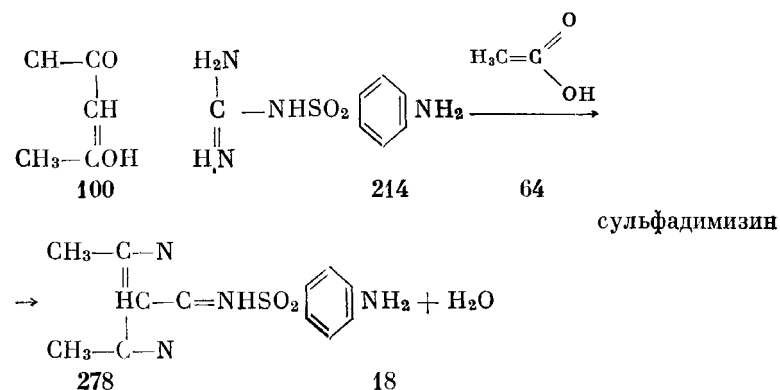
Влияние уксусной кислоты (x_3). Вопрос о влиянии процентного содержания уксусной кислоты специально не исследован. Считается, что x_3 целесообразно поддерживать около 16—17%. Предполагается, что это растворитель и его концентрация может изменяться в широких пределах.

На основании анализа имеющихся сведений решено выбрать следующие уровни и интервалы варьирования факторов (табл. 7.1).

Страница 5

Расчет компонентов реакции

ацетилацетон сульгин уксусная кислота



Расчет по стехиометрическим соотношениям на 12,5 г ацетилацетона

$$\frac{100 - 214}{12,5 - z} = \frac{214 \cdot 12,5}{100} = 26,7 \text{ г},$$

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

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

Страница 6 приведена в виде табл. 7. 2.

Имея такую таблицу, можно приступить к обработке экспериментальных данных.

I. 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

Таблица 7.2

Страница 6 Матрица планирования и результаты эксперимента

№ опытов в матрице	Случайный порядок реализации опытов	Дата, время проведения опытов	x_1		x_2		x_3		y_1 — выход реакции, %		y_2 — содержание, %	
			код	час	код	час	код	час	повтор- ные опыты	средний результат	повтор- ные опыты	средний результат
1	2	10/III, 8 час	—	16	—	20	—	12	80,23	84,08	94,2	97,05
	9	5/IV, 8 час	—	16	—	20	—	12	81,93		99,9	
2	6	24/III, 8 час	+	20	—	20	—	12	86,50	85,65	99,6	99,80
	13	14/IV, 8 час	+	20	—	20	—	12	84,80		100,0	
3	1	8/III, 8 час	—	16	+	28	—	12	82,45	82,27	99,2	99,60
	15	24/IV, 8 час	—	16	+	28	—	12	82,10		100,0	
4	7	27/III, 8 час	+	20	+	28	—	12	89,50	90,40	99,8	99,85
	10	6/IV, 8 час	+	20	+	28	—	12	91,30		99,9	
5	3	12/III, 8 час	—	16	—	20	+	18	85,10	84,95	99,5	99,65
	16	27/IV, 8 час	—	16	—	20	+	18	84,80		99,8	
6	8	1/IV, 8 час	+	20	—	20	+	18	90,30	89,95	99,6	99,65
	14	16/IV, 8 час	+	20	—	20	+	18	—		—	
7	4	15/III, 8 час	—	16	+	28	+	18	85,60	85,25	99,5	99,55
	12	12/IV, 8 час	—	16	+	28	+	18	84,90		99,6	
8	5	20/III, 8 час	+	20	+	28	+	18	88,02	88,25	99,5	99,60
	11	9/IV, 8 час	+	20	+	28	+	18	88,48		99,7	

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

Но, может быть, вы имеете другое мнение? Может быть,

t_2 — температура плавления, °C		Примечания		
повторные опыты	средний результат	Начало кипения	Начало помутнения	Выпадение осадка
196	196,5	8 час 45 мин	20 час 20 мин	20 час 40 мин
197		8 час 38 мин	20 час 18 мин	20 час 35 мин
197	196,5	8 час 38 мин	20 час 20 мин	20 час 40 мин
196		8 час 40 мин	20 час 23 мин	20 час 43 мин
195	195,5	8 час 32 мин	20 час 40 мин	20 час 55 мин
196		8 час 30 мин	20 час 45 мин	21 час
196	196,5	8 час 20 мин	20 час	20 час 32 мин
197		8 час 25 мин	20 час 15 мин	21 час
197	197,5	8 час 25 мин	19 час	19 час 15 мин
198		8 час 25 мин	19 час 05 мин	19 час 18 мин
196	196,0	8 час 30 мин	20 час 10 мин	20 час 35 мин
—		При кипении из колбы сильно выбрасывало жидкость		
196	196,5	8 час 20 мин	20 час	20 час 20 мин
197		8 час 25 мин	18 час 15 мин	18 час 35 мин
196	197,0	8 час 28 мин	18 час 45 мин	19 час
197		8 час 35 мин	20 час 45 мин	21 час
197	197,0	8 час 32 мин	21 час	21 час 45 мин
197		8 час 32 мин	21 час	21 час 45 мин

Фамилии лаборантов по сменам
Для каждого опыта
дается карточка
температуры

Опыт повторить!

вы считаете, что однофакторный эксперимент применим независимо от сложности системы?

Да. Обратитесь к пункту 1.

Нет. Обратитесь к пункту 2.

Не знаю. Обратитесь к пункту 3.

1. Вы считаете, что однофакторным экспериментом можно пользоваться при исследовании любой системы.

Давайте посмотрим, что значит однофакторный эксперимент при исследовании semifакторной системы. Итак, экспериментатор хочет исследовать влияние семи факторов на некоторый параметр оптимизации и решил прово-

дить однофакторный эксперимент. Для того чтобы построить кривую, обычно берут четыре-пять экспериментальных точек. Возьмем четыре точки. Необходимое количество опытов при реализации всевозможных комбинаций равно $N = 4^2 = 16384$. 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 застабилизировать 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 того, чтобы он мог разумно минимизировать число опытов.

Наверное вам целесообразно еще раз прочитать гл. 1, 4 и 6, а затем вернуться к пункту I и выбрать другой ответ.

2. Вы находите, что пользоваться однофакторным экспериментом при исследовании сложных систем нецелесообразно. Ну что ж, вы хорошо усвоили те основные положения, которые мы излагали в первой, четвертой и шестой главе.

Переходите к пункту II.

3. Вы ответили «не знаю». Нас это очень печалило. На протяжении всей книги, особенно в гл. 1, 4 и 6, мы пытались показать вам необходимость планирования многофакторного эксперимента. Давайте еще раз 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 детальным планированием.

Посмотрите гл. 4 и 6.

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.

Возвратитесь к пункту I и выберите правильный ответ.

II. 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. 135 (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 диэтилкарбаминоилхлоридом 1:1, $\bar{x}_2$ — соотношение между пиперазином и диэтилкарбаминоилхлоридом 1 : 1, $\bar{x}_3$ — время выдержки, 3 часа, $\bar{x}_4$ — температура, 20° C, $\bar{x}_5$ — время прилива диэтилкарбаминоилхлорида, 20 мин.

В опыте № 2 первые два фактора поддерживаются на верхних уровнях, а все остальные — на нижних. В опыте № 3 на верхних уровнях находятся x_3 и x_4 и т.д. В последнем опыте все факторы, кроме x_5 , находятся на верхних уровнях.

В результате такого многообразия условий получаются различные значения параметра оптимизации. Так, наименьшее значение 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?

Постарайтесь, пожалуйста, ответить на такой вопрос.

Находите ли вы, что условия опыта № 3 более выгодны с точки зрения величины параметра оптимизации, чем опыта № 4 ($y_3 = 48,1$, $y_4 = 46,0$)?

Да. Обратитесь к пункту 4

Нет. Обратитесь к пункту 5

Не знаю. Обратитесь к пункту 6.

4 и 5. Отвечая «да» и отвечая «нет», вы дадите неправильный ответ. Когда вы ответили «да», вы руководствовались тем, что выход реакции в опыте № 3 на 2,1% выше, чем в опыте № 4. Суждения подобного рода очень часто исходят от экспериментаторов, которые не пользуются математико-статистическими методами. Экспериментаторы, вооруженные статистическими методами, поступают более 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. Тогда можно сказать «нет». И, напротив, ошибка может иметь столь малую величину, что разница в 2,1% окажется значимой. Тогда можно ответить «да». Пока мы не знаем ошибку опыта, нам ничего не остается делать, как ответить «не знаю».

Как можно вычислить ошибку опыта, вы узнаете на стр. 155. Вернитесь к пункту II и выберите правильный ответ.

6. Вы ответили «не знаю» и поступили совершенно правильно. Нам неизвестна ошибка опыта, поэтому мы не можем судить о том, значимо ли различаются результаты этих двух опытов.

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 разницы между результатами двух опытов нет.

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

7.2. Ошибки параллельных опытов

Каждый эксперимент содержит элемент неопределенности вследствие ограниченности экспериментального материала. Постановка повторных (или параллельных) опы-

тов 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

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

Отклонение результата любого опыта от среднего арифметического можно представить как разность $y_q - \bar{y}$, где y_q — 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

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

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

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

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.

* Н. Бейли. Статистические методы в биологии. ИЛ., 1962.

III. Рассказывая о расчете среднего значения, мы употребили такую формулировку: среднее арифметическое равно сумме всех n отдельных результатов, деленной на количество параллельных опытов n . Не возникло ли у вас желание наложить некоторое ограничение на такую формулировку?

Да. Обратитесь к пункту 7.

Нет. Обратитесь к пункту 8.

(Здесь, дорогой читатель, мы хотим сказать вам вслед за Пифагором: «Как ни коротки слова «да» и «нет», но все же они требуют самого серьезного размышления»).

7. Такая формулировка вызвала 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.

Переходите к пункту IV.

8. Вы не чувствуете необходимости наложить некоторое ограничение на формулировку среднего арифметического: среднее арифметическое равно сумме всех n отдельных результатов, деленной на количество параллельных опытов n . Тем не менее ограничение имеется. Приведенные формулы среднего арифметического и дисперсии справедливы, если имеет место закон нормального распределения.

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.

Отметим, однако, что наличие среди повторных опытов резко отклоняющихся результатов (так называемых грубых наблюдений) может вызвать нарушение закона нормального распределения. Поэтому нужно исключить грубые наблюдения, а затем рассчитывать среднюю арифметическую и дисперсию. Помните: даже такая простая операция, как вычисление среднего, требует определенных условий, в данном случае — нормального распределения. Надо всегда следить, чтобы не нарушались необходимые условия выполнения той или иной операции. Иначе вы рискуете принять абсурд за истину.

Вернитесь к пункту III и выберите правильный ответ.

IV. 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. Кельница «Теория ошибок измерений» (М., изд-во «Недра», 1967) (рис. 19).

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 термометры 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

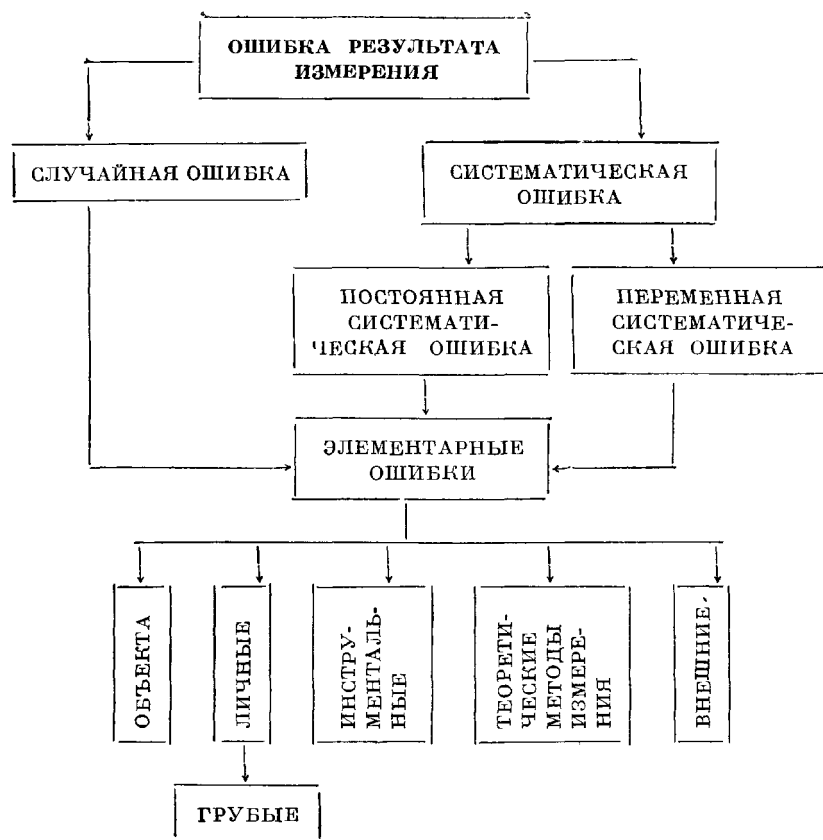


Рис. 19. Схематическое изображение компонент ошибки измерений

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. 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.

V. Для определения брака используют, например, критерий Стьюдента

$$\frac{v - \bar{y}}{s} \geq t.$$

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 мг/см² · час.

Результат первого опыта поставлен под сомнение, так как он выделяется на фоне остальных трех опытов.

Является ли первый опыт браком?

Нет. Обратитесь к пункту 9.

Не знаю. Обратитесь к пункту 10.

Да. Обратитесь к пункту 11.

9. Вы ответили «нет». По-видимому, вы не произвели необходимые расчеты и не использовали критерий Стьюдента. Если бы вы сделали все необходимое, то вам пришлось бы ответить иначе.

Вернитесь к пункту V и выберите другой ответ.

10. Вы ответили «не знаю». Давайте вместе с вами сделаем все необходимые расчеты. Итак, результат первого опыта поставлен под сомнение, потому что он выделяется на фоне остальных трех опытов. Исключим первый опыт из расчета и по остальным произведем вычисление среднего арифметического и стандарта

$$\bar{y} = \frac{2,370 + 2,710 + 2,761}{3} = 2,613,$$

$$s = \sqrt{\frac{(-0,243)^2 + 0,097^2 + 0,148^2}{3-1}} = \sqrt{\frac{0,059}{2}} \cong 0,17.$$

Если произведем проверку по критерию Стьюдента, то, получим

$$\frac{3,580 - 2,613}{0,17} = \frac{0,967}{0,17} = 5,69.$$

При числе степеней свободы $f = 2$ и уровне значимости $0,05$ $t = 4,303$. Экспериментальное значение больше табличного, поэтому сомнительный результат можно считать браком.

Вернитесь к пункту V и выберите правильный ответ.

11. Совершенно верно. Сомнительный опыт оказался браком. Об этом свидетельствует проверка по критерию Стьюдента.

Здесь показан самый простой прием, которым можно пользоваться при исключении ошибочных результатов. Рекомендуем вам познакомиться с другими примерами в статье Н. Г. М и к е ш и н о й, опубликованной в журнале «Заводская лаборатория», 1966, № 1.

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.

7.3. 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

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{q=1}^n (y_{iq} - \bar{y}_i)^2}{N(n - 1)},$$

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

$$s_{\{y\}}^2 = \frac{2 \sum_{i=1}^N (y_{i1} - \bar{y}_i)^2}{N}.$$

Дисперсию воспроизводимости проще всего рассчитывать, когда соблюдается равенство числа повторных опытов во всех экспериментальных точках. На практике весьма 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 степеней свободы

$$s_{\{y\}}^2 = \frac{s_1^2 f_1 + s_2^2 f_2 + \dots + s_k^2 f_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum_{i=1}^N f_i s_i^2}{\sum_{i=1}^N f_i},$$

где s_1^2 — дисперсия первого опыта;

s_2^2 — дисперсия второго опыта и т. д.;

f_1 — число степеней свободы в первом опыте; равному 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. 5 и 6.

Этот вопрос мы рассмотрим в гл. 8, когда будем рассказывать о расчете дисперсии адекватности.

VI. Итак, вы имеете формулы для расчета дисперсии воспроизводимости эксперимента. Казалось бы, все обстоит хорошо. И все же... Вы уже, наверное, чувствуете, что речь пойдет о «некоторых» ограничениях. Действительно, это так.

Но прежде чем рассказать о них, мы хотим обратиться к вашей эрудиции,* дорогой читатель. Как вы думаете, можно ли усреднять дисперсии, если одна из них значительно превышает остальные?

Да. Обратитесь к пункту 12.

Нет. Обратитесь к пункту 13.

Не знаю. Обратитесь к пункту 14.

12. Вы считаете, что можно усреднять неоднородные дисперсии. Наши рассуждения по поводу среднего арифметического не сделали вас более осторожным. Придется предостеречь вас: формулами

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{N(n-1)} \quad \text{и} \quad s_{\{y\}}^2 = \frac{\sum_{i=1}^N f_i s_i^2}{\sum_{i=1}^N f_i}$$

можно пользоваться только в том случае, если дисперсии однородны.

Однородность дисперсий 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.

Но сначала возвратитесь к пункту VI и выберите другой ответ.

13. Ваше мнение: неоднородные дисперсии усреднять нельзя. Вы совершенно правы. (Чувствуется математико-статистическая подготовка читателя!)

Итак, прежде чем пользоваться формулами

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{N(n-1)} \quad \text{и} \quad s_{\{y\}}^2 = \frac{\sum_{i=1}^N f_i s_i^2}{\sum_{i=1}^N f_i},$$

нужно убедиться в однородности суммируемых дисперсий. Как это делается, показано в следующем параграфе.

14. Вы не знаете, можно ли суммировать дисперсии, если одна из них значительно превышает остальные.

Вспомните наши рассуждения по поводу среднего арифметического. Они остаются в силе и в этом случае.

В гл. 4 мы рассказали вам, что после обработки экспериментальные данные представляются в виде уравнения. В следующей главе вы познакомитесь с регрессионным анализом. Так вот, одним из требований регрессионного анализа является однородность дисперсий. Если требование однородности дисперсий не выполняется, вы не имеете права пользоваться обычным регрессионным анализом.

Поэтому экспериментатору не следует забывать о проверке однородности дисперсий.

Вернитесь к пункту VI и выберите правильный ответ.

7.4. 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. 204).

VII. 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 $s_1^2 = 5,14$, $n_1 = 7$ и $f_1 = 6$,

$s_2^2 = 0,324$ для $n_2 = 6$ и $f_2 = 5$.

Как вы считаете, верна ли гипотеза об однородности этих дисперсий?

Да. Обратитесь к пункту 15.

Нет. Обратитесь к пункту 16.

Не знаю. Обратитесь к пункту 17.

15. Мы проверяем гипотезу об однородности двух дисперсий. Вы утверждаете, что эта гипотеза верна. Почему же?

По экспериментальным данным мы получили следующее дисперсионное отношение:

$$F_{\text{эксп}} = \frac{5,14}{0,324} = 15,9.$$

Для числа степеней свободы $f_1 = 6$ и $f_2 = 5$ и для уровня значимости $0,05$ $F_{\text{табл}} = 4,95$. $F_{\text{эксп}} > F_{\text{табл}}$. Это говорит о том, что две дисперсии значимо отличаются друг от друга. Весьма вероятно, что гипотеза об однородности дисперсий не верна.

Вам нужно возвратиться к пункту VII и выбрать другой ответ. Если у вас возникают затруднения в обращении с таблицей F -критерия, посмотрите наши объяснения в пункте 17.

16. Мы проверяем гипотезу об однородности двух дисперсий $s_1^2 = 5,14$ и $s_2^2 = 0,324$ при $f_1 = 6$ и $f_2 = 5$. Ваше мнение — гипотеза неверна. Совершенно правильно. Ваш ответ свидетельствует о том, что вы верно сделали вычисления и умеете обращаться с таблицей F -критерия.

Переходите к пункту VIII.

17. Вы не знаете, как ответить на этот вопрос. Вероятно, это произошло потому, что у вас возникли затруднения в обращении с таблицей F -критерия. Давайте рассуждать вместе. 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 $f_2 = 5$; $F_{\text{табл}} = 4,95$. Это значит: вероятность того, что экспериментальное значение F будет больше чем $4,95$, равна $0,05$ или 5% . Наше $F_{\text{эксп}} = 15,90$. Он 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).

Теперь вернитесь к пункту VII и выберите правильный ответ

VIII. 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

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

а затем из всех дисперсий находится наибольшая $s_{\max}^2$, которая делится на сумму всех дисперсий. Критерий Кохрена — это отношение максимальной дисперсии к сумме всех дисперсий

$$G = \frac{s_{\max}^2}{\sum_{i=1}^N s_i^2}.$$

Гипотеза об однородности дисперсий подтверждается, если экспериментальное значение критерия Кохрена не превышает табличного значения. Тогда можно усреднять дисперсии и пользоваться формулой

$$s_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{q=1}^n (y_{iq} - \bar{y}_i)^2}{N(n-1)}.$$

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 приступить к обработке экспериментальных данных. Воспользуемся этой таблицей для расчета дисперсии воспроизводимости. Перепишем ее так, чтобы было удобно производить расчет (табл. 7.3).

Дисперсия в каждом опыте равна

$$s^2 = \frac{\sum_{q=1}^2 (y_q - \bar{y})^2}{2-1} = 2(\Delta y^2).$$

Максимальная дисперсия оказалась в опыте № 4.

Проверьте гипотезу однородности дисперсий по критерию Кохрена. Если гипотеза однородности подтвердится, рассчитайте дисперсию воспроизводимости.

Правильный ответ помещен в пункте 18.

Таблица 7.3

Расчет дисперсии воспроизводимости

№№ опытов	Матрица планирования	y'	y''	$\bar{y}$	Δy	$(\Delta y)^2$	s_i^2
1	(1)	80,23	81,93	81,08	-0,850	0,722	1,444
2	a	86,50	84,80	85,65	0,850	0,722	1,444
3	b	82,45	82,10	82,27	0,175	0,031	0,062
4	ab	89,50	91,30	90,40	-0,900	0,810	1,620
5	c	85,10	84,80	84,95	0,150	0,023	0,046
6	ac	90,30	89,60	89,95	0,350	0,123	0,246
7	bc	85,60	84,90	85,25	0,350	0,123	0,246
8	abc	88,02	88,48	88,25	-0,230	0,053	0,106
$\sum_{i=1}^N (\Delta y_i)^2$						2,607	

18. 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. 7.4)

Предлагаем вам посчитать дисперсии в каждом опыте и дисперсию воспроизводимости (если не возникнет предположение, что дисперсии неоднородны).

Правильный расчет посмотрите в пункте 19.

Таблица 7.4

Матрица планирования $2^3 \cdot 1$ с различным числом повторных опытов *

№№ п/п	Матрица	y^I	y^{II}	y^{III}	y^{IV}	$\bar{y}$	Δy^I
1	c	87,31	86,01	—	—	86,65	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, %.

№№ п/п	Матрица	Δy^{II}	Δy^{III}	Δy^{IV}	$(\Delta y^I)^2$	$(\Delta y^{II})^2$	$(\Delta y^{III})^2$	$(\Delta y^{IV})^2$	f
1	c	-0,65			0,422	0,422			1
2	abc	-0,25			0,062	0,062			1
3	b	0,85	-0,35	0,15	0,422	0,723	0,122	0,022	3
4	a	0,53	-0,17		0,137	0,281	0,029		2

$$19. s_1^2 = (0,422 + 0,422)/(2 - 1) = 0,844,$$

$$s_2^2 = (0,062 + 0,062)/(2 - 1) = 0,124,$$

$$s_3^2 = (0,422 + 0,723 + 0,122 + 0,022)/(4 - 1) = 0,429,$$

$$s_4^2 = (0,137 + 0,281 + 0,029)/(3 - 1) = 0,149,$$

$$s_{\{y\}}^2 = \frac{0,844 \cdot 1 + 0,124 \cdot 1 + 0,429 \cdot 3 + 0,149 \cdot 2}{1 + 1 + 3 + 2} = \frac{2,553}{6} = 0,426.$$

В данном примере у нас не возникло предположения о неоднородности дисперсий, поскольку все они имеют одинаковый порядок.

IX. Если возникает предположение о наличии неоднородности, следует попытаться его проверить. Для этой цели можно воспользоваться критерием Бартлета. По уже зна-

комой вам формуле подсчитывается дисперсия воспроиз-

$$\text{водимости } s_{\{y\}}^2 = \frac{\sum_{i=1}^N f_i s_i^2}{\sum_{i=1}^N f_i}.$$

Далее находится величина

$$\frac{1}{c} \left\{ f \lg s_{\{y\}}^2 - \sum_{i=1}^N f_i \lg s_i^2 \right\},$$

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 неравным числом повторных наблюдений получены результаты, приведенные в табл. 7.5.

Таблица 7.5

Исходные данные для расчета критерия Бартлета

№	s_i^2	f_i
1	3,50	4
2	4,22	5
3	5,88	3
4	11,36	3

Необходимо рассчитать $s_{\{y\}}^2$ и критерий Бартлета, а затем ответить на вопрос, верна ли гипотеза об однородности дисперсии.

Гипотеза верна. Обратитесь к пункту 20.

Гипотеза неверна. Обратитесь к пункту 21.

20. Вы считаете, что гипотеза об однородности дисперсий верна. Действительно,

$$s_{\{y\}}^2 = 5,79, \quad c = 0,4850, \quad \chi_{\text{ксп}}^2 = 1,300,$$

$\chi_{\text{табл}}^2 = 7,815$ для 5%-ного уровня значимости и трех степеней свободы. $\chi_{\text{ксп}}^2 < \chi_{\text{табл}}^2$. Это свидетельствует о том, что нет значимого различия между дисперсиями.

Переходите к пункту X.

21. Ваше мнение — гипотеза об однородности дисперсий — неверна. Наверное, вы ошиблись в расчете. Давайте повторим расчет вместе.

По данным табл. 7.5 мы получаем: $\sum_1^N f_i = 15$ и $s_{\{y\}}^2 = 5,79$. Находим величину c

$$c = 0,4343 \left[1 + \frac{1}{3(3-1)} \left\{ \frac{1}{4} + \frac{1}{5} + \frac{1}{3} + \frac{1}{3} - \frac{1}{15} \right\} \right] = 0,4850.$$

Теперь мы можем определить χ^2

$$\chi^2 = \frac{1}{0,4850} \{ 15 \lg 5,79 - 4 \lg 3,50 - 5 \lg 4,22 - 3 \lg 5,88 - 3 \lg 11,36 \} = 1,370.$$

Экспериментальное значение χ^2 -критерия равно 1,370. Табличное значение для трех степеней свободы и уровня значимости 0,05 равно 7,815, и мы приходим к выводу, что дисперсии однородны.

Вернитесь к пункту IX и выберите правильный ответ.

X. Приступить 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.

7.5. Рандомизация

XI. 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. 7.6 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.

Таблица 7.6

Матрица 2^3 , не рандомизированная во времени

№ опыта	x_1	x_2	x_3	y	№ 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

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?

Да. Обратитесь к пункту 22.

Не знаю. Обратитесь к пункту 23.

Нет. Обратитесь к пункту 24.

22. Допустим, вы ставите первые четыре опыта сегодня, а последние четыре опыта — завтра. Вы уверены, что эти два дня по своим внешним условиям совершенно идентичны? Конечно, нет. А раз так, то различие между ними может вызвать систематическую ошибку, которая отразится на величине b_3 . Ведь матрица построена таким образом, что в первых четырех опытах x_3 находится на верхнем уровне.

Возвратитесь к пункту XI и выберите другой ответ.

23. Вы не знаете, как поступить. Давайте рассуждать вместе. 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 эту ошибку ϵ . Тогда четыре значения параметра оптимизации сдвинуты на величину ϵ по сравнению с истинными значениями. Пусть это будут параметры, входящие в первый блок: $y_1 + \epsilon$, $y_2 + \epsilon$, $y_3 + \epsilon$ и $y_4 + \epsilon$. Однако матрица построена так, что в первом блоке значения x_3 находятся на верхнем уровне, а во втором — на нижнем уровне. Тогда при подсчете b_3 получится следующая картина

$$b_3 = \frac{1}{8} [(y_1 + \epsilon) + (y_2 + \epsilon) + (y_3 + \epsilon) + (y_4 + \epsilon) - y_5 - y_6 - y_7 - y_8] \rightarrow \beta_3 + \frac{\epsilon}{2},$$

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.

Возвратитесь к пункту XI и выберите правильный ответ.

24. Вы совершенно правы. Данная матрица построена таким образом, что верхние и нижние значения x_3 разбивают ее на два блока: в первый блок входят опыты № 1, 2, 3 и 4, во второй — № 5, 6, 7 и 8. Если в такой же последовательности разбить опыты во времени, это чревато плохими последствиями, так как систематическая ошибка ϵ отразится на подсчете b_3 . При рандомизации вероятность такой опасности значительно уменьшается.

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 при-

водится в руководствах по математической статистике *. Фрагмент таблицы помещен на стр. 176. 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.

Выбранную случайным образом последовательность опытов не рекомендуется нарушать.

7.6. Разбиение матрицы типа 2^k на блоки

Если экспериментатор располагает сведениями о предстоящих изменениях внешней среды, сырья, аппаратуры и 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. 7.7).

В этой матрице при составлении блока 1 отобраны 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 $1 = x_1x_2x_3x_4$.

Мы предлагаем вам для данной матрицы (табл. 7.7) рассчитать все коэффициенты и посмотреть, какие коэффициенты смешаны с эффектом сырья.

Правильный расчет помещен в пункте 25.

* В. В. Налимов. Применение математической статистики при анализе вещества. М., Физматгиз, 1960, стр. 386.

Фрагмент таблицы случайных чисел

56	66	25	32	38	64	70	26	27	67	77	40	04	34	63	98	99	89	31	16	12	90	50	28	96
88	40	52	02	29	82	69	34	50	21	74	00	91	27	52	98	98	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	87	31	07	33	69	45	58	48	00	83	48
94	61	41	67	79	41	61	41	15	60	11	88	83	24	82	24	37	78	61	89	42	58	88	22	16
13	24	40	09	05	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	21	31	99	99	06	96	53	99	25	13	63
42	39	02	36	30	99	46	68	45	15	19	74	15	50	17	44	80	13	86	38	40	45	82	13	44
52	43	96	38	13	83	80	72	34	20	84	56	19	49	98	59	14	85	42	99	71	16	34	33	79
02	84	52	43	36	69	32	46	46	30	84	20	68	72	98	94	62	63	59	44	00	89	06	15	87
85	77	38	88	24	55	46	48	60	06	90	08	83	83	98	40	90	88	25	26	85	74	55	80	85
38	48	84	88	24	58	04	63	21	16	23	38	25	43	32	98	94	65	35	35	16	07	41	07	43
91	19	05	68	22	31	46	17	62	63	99	71	14	12	64	51	68	50	60	78	22	69	51	98	37
54	81	87	21	91	40	46	20	36	25	33	65	94	48	75	00	06	65	25	90	16	29	34	14	43
65	43	75	12	31	20	36	32	32	92	33	65	95	48	75	27	06	11	30	68	70	97	87	21	21
49	98	71	31	80	59	57	35	43	07	85	06	40	75	27	29	17	06	11	37	54	20	97	87	21
03	98	68	89	39	71	87	32	14	99	42	10	25	37	30	08	27	75	43	93	60	69	93	50	50
56	04	21	34	92	89	81	52	15	12	84	11	12	66	87	47	21	06	86	08	35	39	52	28	09
48	09	36	95	36	20	82	53	32	89	92	68	50	88	17	37	92	02	23	43	63	24	69	80	91
23	97	10	96	57	74	07	95	26	44	93	08	43	30	41	86	45	74	33	78	84	33	38	76	73
43	97	55	45	98	35	69	45	96	80	46	26	39	96	33	60	20	73	30	79	17	19	03	47	28
05	05	08	50	79	89	58	49	86	48	27	98	99	24	08	94	19	15	84	29	82	14	35	88	03
66	97	10	69	02	25	36	43	71	76	00	67	56	12	69	07	89	55	63	31	50	72	20	33	36
15	62	38	72	92	03	76	09	30	75	77	80	04	24	54	67	60	10	79	26	21	60	03	48	14
77	81	15	14	67	55	24	22	20	55	36	93	67	69	37	72	22	43	46	32	56	15	75	25	12
18	87	05	09	96	45	14	72	41	46	12	67	46	72	02	59	06	17	49	12	73	28	23	52	48
08	58	15	03	66	43	07	04	48	71	39	07	46	96	40	20	86	79	14	81	74	14	15	23	17
46	07	79	57	61	42	19	68	15	12	60	21	59	12	07	04	99	88	22	39	77	16	69	13	84

Таблица 77

2.5. Разбиение матрицы 2^3 на два блока

№ блока	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_6
	+	—	+	+	—	—	+	—	y_7
	+	+	+	—	+	—	—	—	y_8

$$b_0 = \frac{1}{8} [(y_1 + \varepsilon) + (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) + y_5 + y_6 + y_7 + y_8]$$

$$b_0 \rightarrow \beta_0 + \frac{\varepsilon}{2}$$

$$b_1 = \frac{1}{8} [-(y_1 + \varepsilon) + (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) - y_5 + y_6 - y_7 + y_8],$$

$$b_1 \rightarrow \beta_1;$$

$$b_2 = \frac{1}{8} [-(y_1 + \varepsilon) - (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) - y_5 - y_6 + y_7 + y_8],$$

$$b_2 \rightarrow \beta_2;$$

$$b_3 = \frac{1}{8} [(y_1 + \varepsilon) - (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) - y_5 + y_6 + y_7 - y_8],$$

$$b_3 \rightarrow \beta_3;$$

$$b_{12} = \frac{1}{8} [(y_1 + \varepsilon) - (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) + y_5 - y_6 - y_7 + y_8],$$

$$b_{12} \rightarrow \beta_{12};$$

$$b_{13} = \frac{1}{8} [-(y_1 + \varepsilon) - (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) + y_5 + y_6 - y_7 - y_8],$$

$$b_{13} \rightarrow \beta_{13};$$

$$b_{23} = \frac{1}{8} [-(y_1 + \varepsilon) + (y_2 + \varepsilon) - (y_3 + \varepsilon) + (y_4 + \varepsilon) + y_5 - y_6 + y_7 - y_8],$$

$$b_{23} \rightarrow \beta_{23};$$

$$b_{123} = \frac{1}{8} [(y_1 + \varepsilon) + (y_2 + \varepsilon) + (y_3 + \varepsilon) + (y_4 + \varepsilon) - y_5 - y_6 - y_7 - y_8],$$

$$b_{123} \rightarrow \beta_{123} + \frac{\varepsilon}{2}$$

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 (AB). Двумя другими удовлетворяющими условию опытами будут acd и abd , имеющие по две буквы со всеми взаимодействиями.

В результате получается блок № 1 (табл. 7. 8). Для определения состава следующего блока выбираем какое-либо не использованное испытание, например a , и умножаем на этот символ каждый член блока 1, получаем блок 2. Аналогичная операция проводится для определения состава блоков 3 и 4. Путем выбора неиспользованного испытания b получаем блок 3, используя d — блок 4.

Таблица 7.8

Разбиение матрицы 2^4 на четыре блока

Блок 1	Блок 2	Блок 3	Блок 4
(1)	a	b	d
bc	abc	c	bcd
acd	cd	$abcd$	ac
abd	bd	ad	ab

Теперь мы предлагаем вам записать эту матрицу в кодовом обозначении $+1$ и -1 и проверить, какие взаимодействия смешаны с межблоковым эффектом.

Вашу матрицу вы можете сравнить с матрицей в пункте 26.

В матрице табл. 7.9 можно видеть, что в каждом блоке для всех эффектов, за исключением смешанных, соблюдается равенство числа $+1$ и -1 . Следовательно, межбло-

Таблица 7.9

Матрица 2^4 , разбитая на 4 блока

№ блока	Буквен- ные обо- значения	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}
1	(1) bc acd abd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	a abc cd bd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3	b c $abcd$ ad	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4	d bcd ac ab	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

ковый 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 8 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 *.

7.7. 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, Кокрена или Бартлетта. Очень важно отбросить грубые наблюдения — брак при постановке повторных опытов.

Воспроизводимость эксперимента является одним из важнейших требований планирования эксперимента.

* *Е. В. Маркова*. Латинские квадраты в планировании эксперимента. — Заводская лаборатория, 1968, № 1.

Глава восьмая

Обработка результатов эксперимента

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*.

* Для читателей, желающих расширить свои знания по статистическим методам обработки результатов, мы рекомендуем руководства: *Е. С. Вентцель*. Теория вероятностей, М., «Наука», 1969; *К. Дюерфель*. Статистика в аналитической химии. М., «Мир», 1969; *В. П. Спиридонов, А. А. Лопаткин*. Математическая обработка физико-химических данных. МГУ, 1970; *Д. Худсон*. Статистика для физиков, 2-е изд., М., «Мир», 1970; *N. Draper, H. Smith*. Applied Regression Analysis. New York, John Wiley and Sons, 1966; *А. С. Айвазян*. Статистическое исследование зависимостей. М., «Металлургия», 1966; *Н. В. Смирнов, И. В. Дунин-Барковский*. Курс теории вероятностей и математической статистики для технических приложений. М., «Наука», 1965; *Ю. Нейман*. Вводный курс теории вероятностей и математической статистики. М., «Наука», 1968; *Л. Яноши*. Теория и практика обработки результатов измерений, М., «Мир», 1968.

8.1. 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.

I. 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 эксперимент, 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.

Как вы думаете, по какой причине возникает невязка?

Из-за ошибки эксперимента (дисперсии воспроизводимости). Обратитесь к пункту 1.

Из-за того, что линейная модель не годится. Обратитесь к пункту 2.

По обеим причинам одновременно. Обратитесь к пункту 3.

1. Вы думаете, что причина возникновения невязки — дисперсия воспроизводимости.

Несомненно, она играет здесь не последнюю роль. Не последнюю, но не единственную. Если экспериментальные точки лежат, например, строго на параболе, а вы хотите провести через них прямую, то будет ли причина отклонений только в ошибке опыта?

Вернитесь к пункту I и выберите другой ответ.

2. Вы думаете, что причина возникновения невязки — непригодность линейной модели.

Это возможно. Но ведь может быть и так, что модель пригодна, а невязка существует. Не стоит же из-за отклонений от прямой, лежащих в пределах ошибки опыта, использовать другую, более сложную функцию.

Вернитесь к пункту I и выберите другой ответ.

3. Если ответ не угадан случайно, то мы рады вашим успехам.

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.

II. 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.

Скажите, пожалуйста, возможен ли метод наименьших кубов

$$\sum_{i=1}^N \xi_i^3 = \min?$$

Да. Обратитесь к пункту 4.

Нет. Обратитесь к пункту 5.

Не знаю. Обратитесь к пункту 6.

4. Вы правы. Метод наименьших кубов возможен, 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:

$$\min \max_i |\xi_i|.$$

Предложений можно сделать сколько угодно, но мы не будем более на них останавливаться и перейдем непосредственно к МНК.

III. Когда мы ставим эксперимент, то обычно стремимся провести больше (во всяком случае не меньше) опытов, чем число неизвестных коэффициентов. Поэтому система линейных уравнений

$$\xi_i = y_i - b_0 - b_1 x_{i1}$$

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.

Сколько уравнений будет в нашем случае?

Одно. Обратитесь к пункту 7.

Два. Обратитесь к пункту 8.

Три. Обратитесь к пункту 9.

5. Почему же нельзя минимизировать сумму кубов невязок?

Мы же сказали, что условие, которое мы вводим, является произвольным. Поэтому ничто не может помешать нам построить метод наименьших кубов.

Другое дело, что оценки коэффициентов при этом будут менее точными, а вычисления более сложными.

Вернитесь к пункту II и выберите правильный ответ.

6. Вы не знаете, какой ответ предпочесть.

Видимо, вам непривычно утверждение о том, что выбор требования произволен. Часто принято считать, что математика — наука, чуждая произвола. Но это не так.

С точки зрения модели все постулаты, которые мы используем, произвольны. Их выбор делается на основе каких-то внешних соображений. Поэтому мы можем построить не только МНК, но и метод наименьших кубов, и любые другие методы. Они не будут равноценными с точки зрения точности полученных оценок, трудности вычислений и т. п.

Вернитесь к пункту II и выберите правильный ответ.

7. Почему же одно?

Наше уравнение регрессии имеет вид

$$y = b_0 + b_1 x_1.$$

В нем два неизвестных коэффициента. Значит, применяя МНК, мы получим два уравнения.

Вернитесь к пункту III и выберите правильный ответ.

8. Правильно. Для определения двух неизвестных коэффициентов требуется 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.

$$\left. \begin{aligned} \frac{\partial U}{\partial b_0} &= 0 \\ \frac{\partial U}{\partial b_1} &= 0 \end{aligned} \right\}.$$

Вот откуда берутся наши уравнения для определения коэффициентов. Теперь, как говорится, дело техники.

Попробуйте записать эти уравнения в явном виде и сравните с ответом в пункте 10.

9. Почему же три?

Наше уравнение регрессии имеет вид: $y = b_0 + b_1 x_1$.

В нем два неизвестных коэффициента. Значит, применяя МНК, мы получим два уравнения.

Вернитесь к пункту III и выберите правильный ответ.

10. Вот наши уравнения

$$\left. \begin{aligned} -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_{1i}) &= 0; \\ -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_{1i}) x_{1i} &= 0. \end{aligned} \right\}$$

Для вычислений удобно раскрыть скобки и провести простые преобразования, которые дают

$$\left. \begin{aligned} Nb_0 + \sum_{i=1}^N x_{1i} b_1 &= \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{1i} b_0 + \sum_{i=1}^N x_{1i}^2 b_1 &= \sum_{i=1}^N y_i x_{1i}. \end{aligned} \right\}$$

Запишите теперь в общем виде формулы для b_0 и b_1 и сравните с правильными формулами в пункте 11.

11. Окончательные формулы для вычисления коэффициентов регрессии, которые удобно находить с помощью определителей, имеют вид

$$b_0 = \frac{\sum_{i=1}^N y_i \cdot \sum_{i=1}^N x_{1i}^2 - \sum_{i=1}^N y_i x_{1i} \cdot \sum_{i=1}^N x_{1i}}{N \cdot \sum_{i=1}^N x_{1i}^2 - \left(\sum_{i=1}^N x_{1i} \right)^2},$$

$$b_1 = \frac{N \cdot \sum_{i=1}^N y_i x_{1i} - \sum_{i=1}^N y_i \sum_{i=1}^N x_{1i}}{N \cdot \sum_{i=1}^N x_{1i}^2 - \left(\sum_{i=1}^N x_{1i} \right)^2}.$$

Посмотрим теперь, как вычисляются суммы, входящие в эти формулы.

Результаты эксперимента представляются следующей матрицей:

Т а б л и ц а 8.1

Опыты	x_1	y
1	x_{11}	y_1
2	x_{12}	y_2
.	.	.
.	.	.
.	.	.
i	x_{1i}	y_i
.	.	.
.	.	.
N	x_{1N}	y_N

Для выполнения вычислений ее расширяют, как представлено в табл. 8.2.

Таблица 8.2

Опыты	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	$\bar{x}_1$	$\bar{y}$					

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.

IV. 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.

Какая из проверок является на ваш взгляд наиболее полной, наиболее жесткой?

Первая. Обратитесь к пункту 12.

Вторая. Обратитесь к пункту 13.

Они равноценны. Обратитесь к пункту 14.

12. Вы ответили, что первая проверка наиболее полная. Это не верно.

Первое соотношение проверяет только правильность вычислений в таблице. Но у вас еще остается масса возможностей для ошибок при вычислении коэффициентов.

Вернитесь к пункту IV и выберите правильный ответ.

13. Вы думаете, что вторая проверка более полная. Вы правы. Она проверяет не только вычисления сумм, но и вычисления коэффициентов.

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!

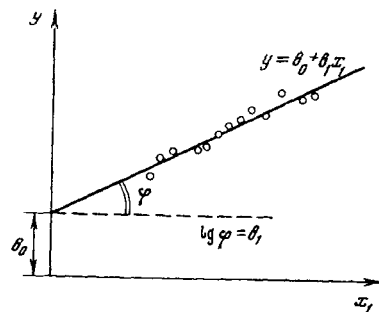


Рис. 20. Линейное уравнение регрессии

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. 20).

V. Выделим для удобства рассмотрения несколько экспериментальных точек и отрезок нашего уравнения в большем масштабе (рис. 21).

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. Как, по-вашему, мы мини-

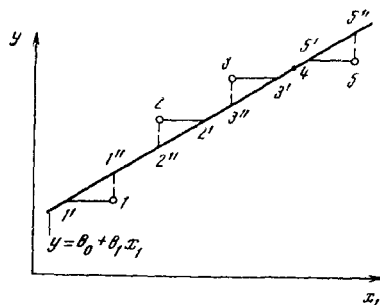


Рис. 21. Линейное уравнение регрессии (фрагмент)

мизировали сумму квадратов сплошных отрезков или пунктирных?

Сплошных. Обратитесь к пункту 15.

Пунктирных. Обратитесь к пункту 16.

14. Если вы думаете, что проверки равноценны, то вы ошибаетесь. Первая из них обеспечивает безошибочный подсчет всех сумм в таблице, но оставляет возможности для ошибок в вычислениях коэффициентов. Вторая проверка обеспечивает правильность всех вычислений.

Вернитесь к пункту IV и выберите правильный ответ.

15. Вы сказали, что мы минимизировали сумму сплошных отрезков. Вы ошиблись.

Вы были бы правы, если бы наше уравнение регрессии имело вид

$$x_1 = b_0 + b_1 y.$$

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.

Вернитесь к пункту V и выберите правильный ответ.

16. Конечно. Мы минимизируем сумму квадратов разностей $y - \hat{y}$, поэтому надо выбирать вертикальные отрезки.

Чтобы справедливым стало первое утверждение, надо записать уравнение регрессии в виде

$$x_1 = b_0 + b_1 y.$$

Тогда, поменяв местами x_1 и y во всех уравнениях, можно найти новые коэффициенты. 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.

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. 8.3).

Таблица 8.3

Опыты	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. Мы не будем этим заниматься, отсылая желающих к литературе *. Мы воспользуемся тем, что наши матрицы планирования ортогональны. Если вы забыли это понятие, то обратитесь к стр. 103 и повторите его. Далее будем рассматривать только этот случай, который позволяет резко упростить

* В. В. Налимов, Н. А. Чернова. Указ. соч.

вычисления, что составляет одно из преимуществ планирования эксперимента.

VI. Можно показать *, что для любого числа факторов коэффициенты будут вычисляться по формуле

$$b_j = \frac{\sum_{i=1}^N y_i x_{ji}}{N}.$$

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. 188. В первой формуле

$$\sum_{i=1}^N x_{1i} = 0$$

в силу симметричности плана. Поэтому после сокращения формулы приобретают вид

$$b_0 = \frac{\sum_{i=1}^N y_i}{N}, \quad b_1 = \frac{\sum_{i=1}^N y_i x_{1i}}{\sum_{i=1}^N x_{1i}^2},$$

где

$$\sum_{i=1}^N x_{1i}^2 = N,$$

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.

Вычислите, пожалуйста, коэффициенты линейного уравнения регрессии для примера 4 гл. 5 и сравните с результатами в пункте 17.

* В. В. Налимов, Н. А. Чернова. Указ. соч.

17. Вот искомое уравнение

$$y = 88,0 - 2,0x_1 - 4,5x_2.$$

Добейтесь, чтобы все результаты совпадали.

При вычислениях линейных моделей по дробным репликам никаких особенностей не появляется. Все точно так же.

Убедитесь в этом, вычислив линейные модели для всех примеров гл. 6, и сверьте с ответами в пункте 18.

Вы не должны пренебрегать этими упражнениями. Их трудоемкость не должна вас отпугивать, ибо «трудно в ученье — легко в бою».

18. Вот ответы

$$1) y = 15,00 - 1,50x_1 + 4,75x_2 + 0,75x_3 + 4,50x_4,$$

$$2) y = 52,300 - 1,755x_1 + 5,050x_2 + 0,575x_3 - 2,100x_4 + 0,325x_5,$$

$$3) y = 23,590 + 1,065x_1 - 5,700x_2 + 0,191x_3 + 3,210x_4 + 0,066x_5 - 1,320x_6 - 1,780x_7.$$

Добейтесь совпадения всех результатов и перейдите к пункту VII.

VII. Дополнительные трудности возникают, если мы хотим найти коэффициенты неполного квадратного уравнения (если нас интересуют эффекты взаимодействия). Тогда уравнение регрессии будет иметь вид

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + b_{12}x_1x_2 + b_{13}x_1x_3 + \dots + b_{k-1,k}x_{k-1}x_k.$$

Конечно, можно интересоваться не всеми эффектами взаимодействия, а только определенными. В полном факторном эксперименте можно оценить все взаимодействия. Для дробных реплик это не так.

Скажите, если вы построили полуреплику 2^{4-1} с определяющим контрастом $1 = x_1 x_2 x_3 x_4$, можно ли оценить независимо коэффициенты b_{12} и b_{34} ?

Да. Обратитесь к пункту 19.

Нет. Обратитесь к пункту 20.

Не знаю. Обратитесь к пункту 21.

19. Вы ответили, что такие оценки можно получить. Мы не удовлетворены тем, как вы проработали гл. 6. Вам придется вернуться к ней и разобраться в смешанных эффектах. В рассматриваемом случае имеет место соотношение $x_1x_2 = x_3x_4$, следовательно, отдельные оценки b_{12} и b_{34} невозможны.

После гл. 6. вернитесь к пункту VII и выберите правильный ответ.

20. Верно

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, $u \neq j$.

Вычислите, пожалуйста, эффект взаимодействия для примера из гл. 5 и сравните с ответом в пункте 22.

21. Вы не знаете. Ну что ж. Давайте совершим экскурсию в гл. 6 и вспомним смешанные эффекты.

Так как имеет место соотношение $x_1x_2 = x_3x_4$, то отдельные оценки b_{12} и b_{34} невозможны.

После гл. 6 вернитесь к пункту VII и выберите правильный ответ.

22. Вот ответ: $b_{12} = -0,5$.

Добейтесь совпадения результатов.

Обратите внимание, что в силу ортогональности эффекты взаимодействия оцениваются независимо от линейных эффектов.

8.2. 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.

VIII. 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?

Нарушения нормального распределения не приводят ни к каким последствиям. Обратитесь к пункту 23.

Нарушения нормального распределения приводят к изменению вероятностей, с которыми справедливы наши утверждения при проверке гипотез. Обратитесь к пункту 24.

23. Вы думаете, что регрессионный анализ не статистический метод? Ведь статистика существенно связана с законами распределения, и их нарушения немедленно сказыв-

ваются при проверке гипотез. Если бы это было не так, то первый постулат нам бы просто не понадобился.

Вернитесь к пункту VIII и выберите правильный ответ.

24. Вы, конечно, правы.

При нарушении нормальности мы лишаемся возможности установления вероятностей, с которыми справедливы те или иные высказывания. В этом таится большая опасность. Мы рискуем загипнотизировать себя численными оценками и вероятностями, за которыми ничего не стоит. Это даже хуже волюнтаризма. Вот почему надо очень внимательно относиться к возможным нарушениям предпосылок.

IX. Второй постулат. Дисперсия y не зависит от абсолютной величины y . С этим требованием мы уже встречались в седьмой главе. Если вы забыли, то вернитесь на стр. 165 и повторите.

Выполнимость этого постулата проверяется с помощью критериев однородности дисперсий в разных точках факторного пространства. Нарушение этого постулата недопустимо.

Что, по вашему мнению, следует предпринять, если однородность дисперсий все же отсутствует?

Отказаться от решения задачи. Обратитесь к пункту 25.

Искать преобразование y , делающее дисперсии однородными. Обратитесь к пункту 26.

25. Вы настроены слишком пессимистично. Не следует опускать руки перед такими препятствиями. Это, конечно, неприятное затруднение, но что-нибудь обязательно придумаем.

Давайте обратимся к пункту 26 и рассмотрим другое предложение.

26. Это разумное предложение. Всегда существует такое преобразование y , которое делает дисперсии однородными. Увы, его не всегда легко найти. Довольно часто помогает логарифмическое преобразование, с которого обычно начинают поиски.

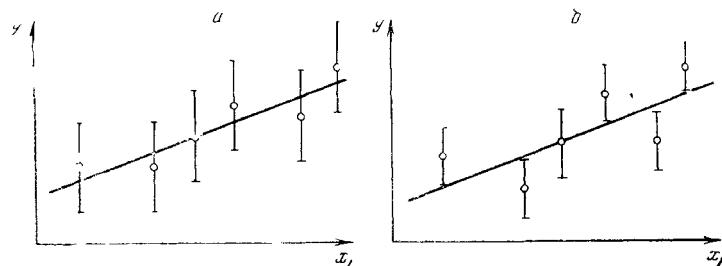


Рис. 22. Иллюстрация к проверке адекватности

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.

8.3. Проверка адекватности модели

Первый вопрос, который нас интересует после вычисления коэффициентов модели, это проверка ее пригодности. Мы будем называть такую проверку *проверкой адекватности модели*.

Х. Выше (рис. 22, а, б) приведены два рисунка с одинаковым расположением экспериментальных точек и, следовательно, одинаковым разбросом относительно линии регрессии, но с различным средним разбросом в точках (с различной дисперсией воспроизводимости).

Разброс в точках показан, как это иногда делается, отрезками прямых, составляющих доверительный интервал, равный $\pm 2s_{\{y\}}$.

Как вы думаете, в каком случае модель можно считать адекватной?

В первом случае. Обратитесь к пункту 27.

Во втором случае. Обратитесь к пункту 28.

Не знаю. Обратитесь к пункту 29.

27. Вы правы.

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.

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).

XI. 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* ($s_{ад}^2$).

$$s_{ад}^2 = \frac{\sum_{i=1}^N \Delta y_i^2}{f}.$$

Скажите, чему равно число степеней свободы для $s_{ад}^2$ в следующем случае: план 2^{4-1} и 4 параллельных опыта в нулевой точке для вычисления ошибки опыта, модель линейная?

7. Обратитесь к пункту 30.
4. Обратитесь к пункту 31.
3. Обратитесь к пункту 32.

28. Вы отдали предпочтение второму случаю. Это неверно.

Здесь мы имеем дело с точным экспериментом, поэтому отклонения от линии регрессии, которые мы наблюдаем, трудно признать случайными.

На первом рисунке эксперимент достаточно груб, чтобы можно было удовлетвориться уравнением прямой линии. Это не приводит к противоречиям.

Вернитесь к пункту X и выберите правильный ответ.

29. Вы не знаете, какой вариант предпочесть.

Давайте вернемся к смыслу проверки адекватности. Мы сказали, что хотим сравнить разброс относительно линии регрессии с разбросом в точке. Чем больше будет различие между ними, тем труднее признать отклонения случайными. Тогда придется утверждать, что выбранная модель не пригодна (не адекватна) и требуется другая, более сложная модель.

Вернитесь теперь к пункту X и выберите другой ответ.

30. У вас получилось 7 степеней свободы. Это неверно.

Вероятно, вы рассуждали так. Прделано 12 опытов: $2^{4-1} = 8$ плюс 4 нулевых. В уравнение входит 5 коэффициентов. Следовательно, $f = 12 - 5 = 7$. Вы не учли, что параллельные опыты нельзя считать самостоятель-

ными, так как они дублируют друг друга. Поэтому они все дают одну степень свободы.

Вернитесь к пункту XI и выберите другой ответ.

31. Вы насчитали 4 степени свободы. Это неверно. Такой 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 входить в число степеней свободы.

Вернитесь к пункту XI и выберите правильный ответ.

32. Вы ответили правильно.

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.

В статистике разработан критерий, который очень удобен для проверки гипотезы об адекватности модели. Он называется *F-критерием Фишера* и определяется следующей формулой:

$$F = \frac{s_{ад}^2}{s_{\{y\}}^2}.$$

Величину, стоящую в числителе этой формулы, мы только что научились считать, а знаменатель — старый знакомый (см. гл. 7) — 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 соответствующих таблиц, который может удовлетворить ваши нужды не только в упражнениях этой книги, но и в большинстве случаев практики, приведен ниже (табл. 8.4).

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.

Пользуясь таблицами и найденной ранее линейной моделью (гл. 5, пример 4, стр. 109), проверьте адекватность, если $s_{y_1}^2 = 0,625$.

Сравните полученный ответ с ответом в пункте 33.

33. Вот ответ:

$$F = 1,6, F_{\text{табл}}(1,4) = 7,7.$$

Убедитесь, что вы не ошиблись.

Способ расчета дисперсии адекватности, который мы рассмотрели выше, применим в том случае, если опыты в матрице планирования не дублируются, а информация о дисперсии воспроизводимости извлекается из параллельных опытов в нулевой точке или из предварительных экспериментов.

Для нас еще важны два случая: 1) опыты во всех точках плана дублируются одинаковое число раз (равномерное дублирование), 2) число параллельных опытов не одинаково (неравномерное дублирование).

Таблица 8.4

Значения F -критерия Фишера при 5%-ном уровне значимости

$f_1 \backslash f_2$	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	1,8
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

В первом случае дисперсию адекватности нужно умножать на n , где n — число повторных опытов

$$s_{\text{ад}}^2 = \frac{n \sum_{i=1}^N \Delta y_i^2}{f}.$$

Такое видоизменение формулы вполне естественно. Чем больше число параллельных опытов, тем с большей достоверностью оцениваются средние значения. Поэтому требования к различиям между экспериментальными и расчетными значениями становятся более жесткими, что отражается в увеличении F -критерия.

Во втором случае, когда приходится иметь дело с неравномерным дублированием, положение усложняется. Даже когда экспериментатор задумал провести равное число параллельных опытов, часто не удается по тем или иным причинам все их реализовать. Кроме того, иногда приходится отбрасывать отдельные опыты как выпадающие наблюдения.

При неравномерном дублировании нарушается ортогональность матрицы планирования и, как следствие, изменяются расчетные формулы для коэффициентов регрессии и их ошибок, а также для дисперсии адекватности.

Для дисперсии адекватности можно записать общую формулу

$$s_{ад}^2 = \frac{\sum_{i=1}^N n_i (\bar{y}_i - \hat{y}_i)^2}{f},$$

где N — число различных опытов (число строк матрицы);

n_i — число параллельных опытов в i -й строке матрицы;

$\bar{y}_i$ — среднее арифметическое из n_i параллельных опытов;

$\hat{y}_i$ — предсказанное по уравнению значение в этом опыте.

Смысл этой формулы очень прост: различию между экспериментальным и расчетным значением придается тем больший вес, чем больше число повторных опытов.

Для b -коэффициентов нельзя записать универсальную расчетную формулу. Все зависит от того, какой был план и как дублировались опыты. Всякий раз приходится делать специальные расчеты, пользуясь методом наименьших квадратов. На простом спланированном примере мы проиллюстрируем такой расчет.

Пусть ставился полный факторный эксперимент 2^3 с двумя параллельными опытами в каждой строке. Один из параллельных опытов пришлось отбросить. Матрица планирования имеет вид табл. 8.5.

Запишем систему уравнений для определения коэффициентов регрессии

$$\begin{aligned} 7b_0 - 1b_1 + 1b_2 &= 19, \\ -1b_0 + 7b_1 + 1b_2 &= -7, \\ 1b_0 + 1b_1 + 7b_2 &= -7. \end{aligned}$$

Таблица 8.5

№	x_0	x_1	x_2	y'	y''	$\bar{y}$	$\hat{y}$	$\bar{y} - \hat{y}$	$(\bar{y} - \hat{y})^2$	$n(\bar{y} - \hat{y})^2$
1	+	—	—	4,5	5,5	5,0	4,59	0,41	0,16	0,32
2	+	+	—	3,0	—	3,0	3,79	0,79	0,64	0,64
3	+	—	+	2,0	2,0	2,0	1,89	0,11	0,01	0,02
4	+	+	+	0,5	1,5	1,0	1,09	0,09	0,01	0,02

$$\sum_{i=1}^N = 1,00$$

Эту запись легко получить с помощью матричных операций *. Для этого нужно матрицу планирования умножить слева на транспонированную матрицу

$$\begin{pmatrix} + & + & + & + & + & + & + \\ - & + & - & + & - & - & + \\ - & - & + & + & - & + & + \end{pmatrix} \begin{pmatrix} + & - & - \\ + & + & - \\ + & - & + \\ + & + & + \\ + & - & - \\ + & - & + \\ + & + & + \end{pmatrix} = \begin{pmatrix} 7 & -1 & 1 \\ -1 & 7 & 1 \\ 1 & 1 & 7 \end{pmatrix}.$$

Элементы полученной матрицы представляют собой коэффициенты уравнений. А числа в правой части системы уравнений есть

$$\sum_{i=1}^7 y_i = 19; \quad \sum_{i=1}^7 y_i x_{1i} = -7; \quad \sum_{i=1}^7 y_i x_{2i} = -7.$$

Чтобы вычислить b -коэффициенты, надо получить обратную матрицу системы уравнений. В нашем случае она имеет вид

$$\frac{1}{40} \begin{pmatrix} 6 & 1 & -1 \\ 1 & 6 & -1 \\ -1 & -1 & 6 \end{pmatrix}.$$

Теперь для расчета b -коэффициентов нужно подсчитать сумму произведений соответствующих элементов столбца свободных членов

* Читатель, незнакомый с матричной алгеброй, может обратиться, например, к справочнику А. П. Мишина, И. В. Проскуряков «Вышая алгебра» М., Физматгиз, 1962.

системы уравнений на элементы строки обратной матрицы

$$b_0 = \frac{6}{40} \cdot 19 + \frac{1}{40}(-7) + \left(-\frac{1}{40}\right)(-7) = 2,84,$$

$$b_1 = \frac{1}{40} \cdot 19 + \frac{6}{40}(-7) + \left(-\frac{1}{40}\right)(-7) = -0,4,$$

$$b_2 = \left(-\frac{1}{40}\right) 19 + \left(-\frac{1}{40}\right)(-7) + \frac{6}{40}(-7) = -1,35.$$

Проверим теперь адекватность линейной модели по приведенной выше формуле. Вспомогательные расчеты представлены в последних четырех столбцах табл. 8.5. Дисперсия воспроизводимости равна: $s_{\text{воспр}}^2 = 0,33$ с тремя степенями свободы (расчетная формула дана в гл. 7), а дисперсия адекватности: $s_{\text{ад}}^2 = \frac{1,00}{4-3} = 1,00$ с одной степенью свободы; $F_{\text{эксп}} = \frac{1,00}{0,33} = 3$, $F_{\text{табл}}(1, 3; 0,05) = 10,1$. Следовательно, модель можно признать адекватной.

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.

8.4. 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 $s_{\{b_i\}}^2$. Она определяется в нашем случае по формуле *

$$s_{\{b_i\}}^2 = \frac{s_{\{y\}}^2}{N}.$$

* В. В. Налимов, Н. А. Чернова. Указ. соч.

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)

$$\Delta b_j = \pm ts \{b_j\}.$$

Здесь t — табличное значение критерия Стьюдента при числе степеней свободы, с которыми определялась $s_{\{y\}}$, и выбранном уровне значимости (обычно 0,05); $s_{\{b_j\}}$ — квадратичная ошибка коэффициента регрессии

$$s_{\{b_j\}} = \sqrt{s_{\{b_j\}}^2}.$$

II. 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.

Таблица 8 6

Значения t -критерия Стьюдента при 5%-ном уровне значимости

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

Доверительный интервал задается верхней и нижней границами $b + \Delta b_j$ и $b_j - \Delta b_j$.

Для отыскания значений t -критерия можно воспользоваться таблицами, фрагмент из которых приведен в табл. 8.6.

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.

Скажите, какой из них значим?

№	b_j	Δb_j
1	5,3	$\pm 0,5$
2	5,3	$\pm 2,6$

Значим первый. Обратитесь к пункту 34.

Значим второй. Обратитесь к пункту 35.

Не знаю. Обратитесь к пункту 36.

34. Вы не правы.

По-вашему, выходит, что чем больше доверительный интервал, тем лучше, тем более значим коэффициент. 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.

Вернитесь к пункту XII и выберите правильный ответ.

35. Верно.

Во втором случае доверительный интервал значительно уже. Для проверки значимости можно использовать рабочее правило: если абсолютная величина коэффициента больше, чем доверительный интервал, то коэффициент значим.

Вам необходимо выработать навык проверки значимости, поэтому проверьте значимость коэффициентов линейных моделей для примеров главы шестой. Для расчета вам понадобятся значения дисперсий воспроизводимости. Вот они:

1. Пример 1, стр. 124.

$$s_{\text{воспр}}^2 = 3,06.$$

2. Пример 2, стр. 134.

$$s_{\text{воспр}}^2 = 1,0.$$

3. Пример 3, стр. 137.

$$s_{\text{воспр}}^2 = 1,0.$$

4. Пример 4, стр. 141.

$$s_{\text{воспр}}^2 = 19,92.$$

Сравните свои ответы с правильными значениями в пункте 37.

36. Вы не знаете. Что же вас затруднило? Рассуждение простое: чем больше доверительный интервал, тем менее вероятно, что коэффициент значим. Значит...

Вернитесь к пункту XII и выберите правильный ответ.

37. Вот ответы:

1. $s_{b_j} = 0,62$. Незначим только b_3 .

2. $s_{b_j} = 0,35$. Незначимы b_3 и b_5 .

3. $s_{b_j} = 0,35$. Незначимы b_3 и b_5 .

4. $s_{b_j} = 1,11$. Незначимы b_3 , b_8 , b_9 , b_{10} , b_{14} .

Убедитесь, что вы не ошиблись.

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.

Убедитесь в этом на всех примерах с нечетными номерами из гл. 6. Ответ вы снова найдете в пункте 37.

Так производится проверка значимости коэффициентов.

8.5. 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.

Глава девятая

Принятие решений после построения модели

9.1. 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. Больший 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.

I. 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 не значимы, 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. Инвариантными к изменению интервалов остаются знаки линейных коэффициентов регрессии. Однако и они изменяться на обратные, если при движении по градиенту (гл. 10) 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. 86). 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. 81) $y \rightarrow \max$.

Вспомним, что $\bar{x}_1$ — концентрация промывающего раствора (элюанта), $\bar{x}_2$ — рН этого же раствора, y — процентное содержание неодима в выходящем растворе — элюате.

После обработки экспериментальных данных получено уравнение регрессии

$$y = 88,0 - 2,0x_1 - 4,5x_2, \quad s\{b\} = 0,30.$$

Какое утверждение вы считаете правильным:

К увеличению параметра оптимизации приводит увеличение значений факторов. Обратитесь к пункту 1.

К увеличению параметра оптимизации приводит уменьшение значений факторов. Обратитесь к пункту 2.

Не ясно, что приводит к увеличению параметра оптимизации. Обратитесь к пункту 3.

1. Вы считаете, что с ростом значений факторов параметр оптимизации увеличивается. Давайте сделаем вычисления. Сначала подставим в уравнение значение $x_1 = -1$, $x_2 = -1$, а затем $x_1 = +1$ и $x_2 = +1$, получим результат

$$y = 88,0 - 2,0(-1) - 4,5(-1) = 94,5,$$

$$y = 88,0 - 2,0(+1) - 4,5(+1) = 81,5.$$

Теперь вы можете вернуться к пункту 1 и выбрать правильный ответ

2. Правильно. Так как знаки коэффициентов отрицательны, то 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.

II. Теперь мы получили основу для перехода к следующему этапу. На основе априорных сведений обычно имеются 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 харак-

тер. 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 выдвиганию 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).

Как вы считаете, какие сочетания значений факторов приводят к росту параметра оптимизации, если эффект взаимодействия факторов x_1 и x_2 имеет отрицательный знак?

Это достигается при $x_1 = +1$ и $x_2 = +1$.

Обратитесь к пункту 4.

Это достигается при $x_1 = -1$ и $x_2 = -1$.

Обратитесь к пункту 5.

Это достигается при $x_1 = +1$ и $x_2 = -1$.

Обратитесь к пункту 6.

Это достигается при $x_1 = -1$ и $x_2 = +1$.

Обратитесь к пункту 7.

Не знаю. Обратитесь к пункту 8.

3. Так как вы не знаете, что ответить, давайте сделаем вычисления. 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 $x_2 = -1$

$$y = 88,0 - 2,0 (-1) - 4,5 (-1) = 94,5.$$

Вы видите, что уменьшение значений факторов действует благоприятно.

Возвратитесь к пункту I и выберите правильный ответ.

4 и 5. Утверждая, что к росту параметра оптимизации приводят сочетания $x_1 = +1$ и $x_2 = +1$ или $x_1 = -1$ и $x_2 = -1$, вы ошибаетесь в обоих случаях.

Вероятно, вы невнимательно читали стр. 215, где обсуждались эти случаи.

Вернитесь к пункту II, прочтите его внимательно и выберите другой ответ.

6 и 7. Вы правы. Любая комбинация разных знаков x_1 и x_2 приводит к росту параметра оптимизации, если эффект взаимодействия отрицателен.

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. 146). 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

$$y = 85,975 + 2,588 x_1 + 0,568 x_2 + 1,125 x_3 - 0,588 x_1 x_3 - 0,918 x_2 x_3, \\ (s \{b\} = 0,28)$$

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 $x_1 x_2 x_3$, 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 взаимо-

действия 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.

Переходите к следующему параграфу.

8. Вы не можете ответить.

Давайте посмотрим на уравнение регрессии

$$y = b_0 + b_1 x_1 + b_2 x_2 - b_{12} x_1 x_2.$$

При подсчете предсказанного значения параметра оптимизации в уравнение регрессии подставляются кодированные значения факторов. Для увеличения параметра оптимизации нужно, чтобы член $b_{12} x_1 x_2$ имел знак плюс. Но коэффициент b_{12} отрицательный. Поэтому произведение $x_1 x_2$ тоже должно быть отрицательно. Это возможно если знаки x_1 и x_2 различны.

Вернитесь к пункту II и выберите правильные ответы.

9.2. Принятие решений после построения модели процесса

Нам придется принимать решения в сложных ситуациях. Решения зависят от числа факторов, подробности 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. 5), где удалось рассмотреть все варианты. Ситуации будем различать по адек-

ватности и неадекватности модели, значимости и незначимости коэффициентов регрессии в модели, информации о положении оптимума.

III. Обсудим сначала принятие решения для адекватного линейного уравнения регрессии.

Линейная модель адекватна. Здесь возможны 3 варианта.

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 *.

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.

1. Методом «перевала» — у исходной реплики изменяют знаки на обратные. В этом случае основные эффекты оказываются не смешанными с парными эффектами взаимодействия.

2. Переходом к полному факторному эксперименту.

3. Переходом к реплике меньшей дробности.

4. 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 случай: 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

* В. В. Налимов, Н. А. Чернова. Указ. соч.

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. 23). В блок-схеме пунктирными линиями обведены ситуации, сплошными линиями — принимаемые решения.

Рассмотрим пример.

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.

Параметр оптимизации — прочность волокна. Условия, матрица планирования и результаты этих дорогостоящих и трудоемких опытов приведены в табл. 9.1*.

Здесь использована 1/8-реплика от полного факторного эксперимента 2^6 с генерирующими соотношениями $x_4 = x_1x_2x_3$, $x_5 = -x_1x_3$, $x_6 = -x_2x_3$.

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.

* Н. С. Иванов, Е. Н. Марина, Д. Ф. Фильберт, С. Я. Межилова, Ю. П. Адлер. Применение математической статистики при исследовании процесса формования и вытягивания полипропиленового волокна. — Сб. «Карбоцепные волокна» М., изд-во «Химия», 1966.

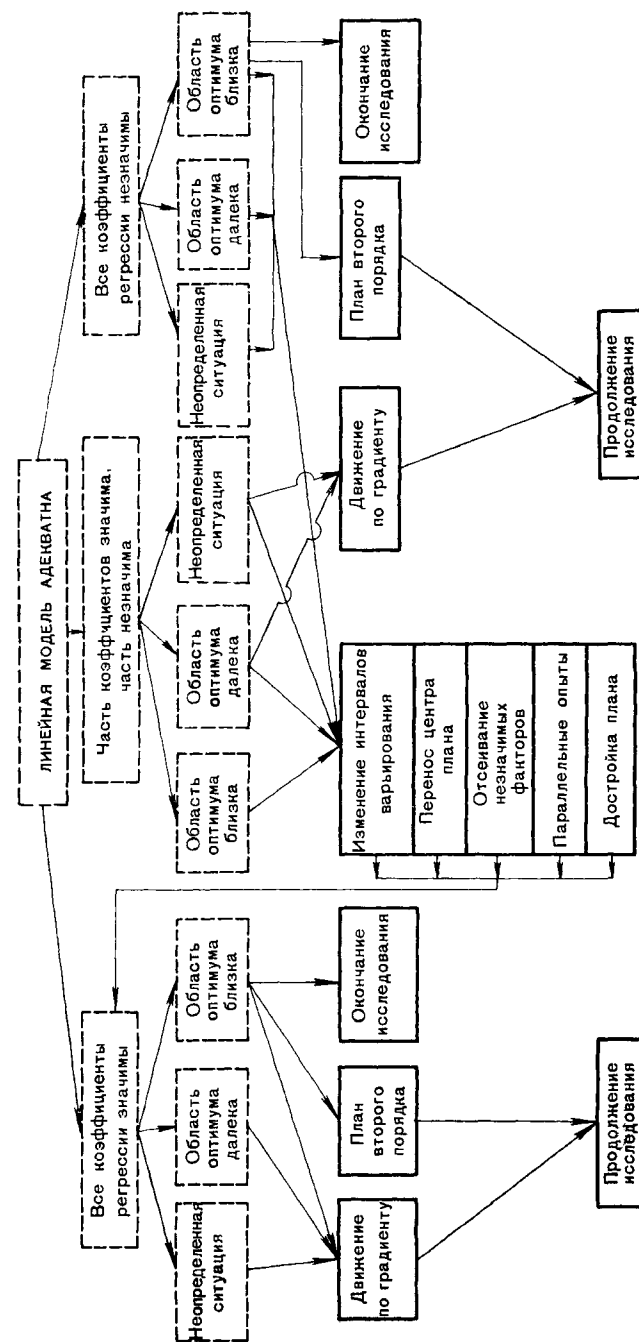


Рис. 23. Блок-схема принятия решений в задаче определения оптимальных условий, линейная модель адекватна

Таблица 9.1

Матрица планирования и результаты опытов

	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$	$\tilde{x}_6$	y
Основной уровень	300	50	2,40	150	0,35	7,2	
Интервал варьирования	10	7	0,47	5	0,12	0,3	
Верхний уровень	310	57	2,87	155	0,47	7,5	
Нижний уровень	290	43	1,93	145	0,23	6,9	
Кодированные значения факторов	x_1	x_2	x_3	x_4	x_5	x_6	
Опыты 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

Какое решение представляется вам целесообразным?

Движение по градиенту. Обратитесь к пункту 9.

Расширение интервалов варьирования. Обратитесь к пункту 10.

Затрудняюсь дать ответ. Обратитесь к пункту 11.

9. Вы считаете, что целесообразно принять решение о движении по градиенту. Оценим это предложение. Из шести коэффициентов регрессии только три оказались значимыми, так что движение может быть не эффективным. Далее, здесь 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.

Переходите к пункту IV.

10. Вы полагаете, что имеет смысл 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, лучше принять другое решение.

Вернитесь к пункту 9 и познакомьтесь с ним.

11. Вы затрудняетесь в выборе одного из двух решений: движение по градиенту или устранение незначимости эффектов. Что нужно учитывать? Из шести коэффициентов регрессии только три коэффициента оказались значимыми, эффекты сильно смешаны. Двигаться по градиенту в этих условиях едва ли целесообразно. Но устранение незначимости эффектов требует дополнительного проведения не менее 8 опытов. Опыты длительны и дороги. В крутом восхождении мы рискуем напрасно поставить 2—3 опыта.

Учитывая это, вернитесь к пункту III и выберите одно из решений.

IV. Рассмотрим пример другой ситуации.

Пример 4

В задаче ионнообменного разделения неодима и празеодима (стр. 109) получено адекватное уравнение регрессии: $y = 88,0 - 2,0 x_1 - 4,5 x_2$, $s\{b\} = 0,30$. Все коэффициенты регрессии значимы, область оптимума близка (наилучший опыт серии $y_1 = 95\%$). Цель исследования — получение выхода 99—100%, число опытов лимитировано.

Какое решение представляется вам целесообразным?

Движение по градиенту. Обратитесь к пункту 12.

Окончание исследования. Обратитесь к пункту 13.

План второго порядка. Обратитесь к пункту 14.

12. Вы полагаете, что следует переходить к движению 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.

В данной задаче было использовано движение по градиенту, расчет которого приведен в гл. 10.

Остается только упомянуть о задаче построения ин-терполяционной формулы: цель исследования достигнута, если получена адекватная модель.

V. 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. 23, возможны варианты окончания исследования и перехода к построению плана второго порядка.

На рис. 24 приведена блок-схема принятия решений в задаче оптимизации для случая, когда линейная модель неадекватна.

Рассмотрим пример.

Пример 5

Оптимизировался процесс получения фармацевтического препарата (карбометоксисульфанилгуанидина).

В качестве факторов были выбраны:

$\bar{x}_1$ — отношение растворителя к основному веществу, г/л,

$\bar{x}_2$ — температура реакционной массы, °C,

$\bar{x}_3$ — время реакции, мин.

Параметр оптимизации — выход продукта в процентах. Условия, матрица планирования и результаты опытов — в табл. 9.2.

Получены следующие результаты: $s\{b\} = 0,12$,

$b_0 = 23,28$; $b_3 = 9,36$; $b_{23} = 3,77$;

$b_1 = 1,78$; $b_{12} = 0,17$; $b_{123} = 1,00$;

$b_2 = 10,23$; $b_{13} = -0,79$; $s^2\{y\} = 0,97$.

Линейное уравнение регрессии оказалось неадекватным: $F_{\text{экср}} = 32,74$ при табличном значении 4,12. Область оптимума далека.

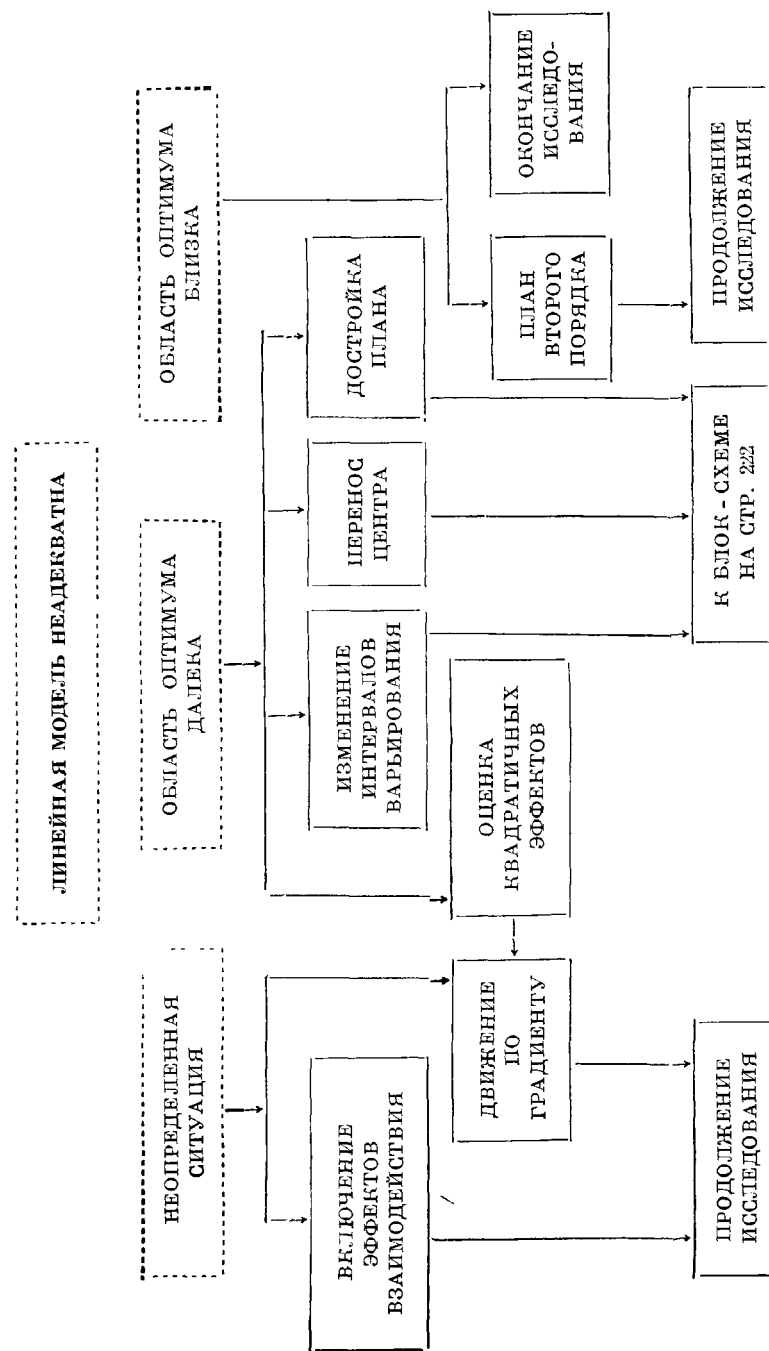


Рис. 24. Блок-схема принятия решений в задаче определения оптимальных условий, линейная модель неадекватна

Таблица 9.2

Матрица планирования и результаты опытов

	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	y
Основной уровень	0,7	135	30	
Интервал варьирования . . .	0,2	5	15	
Верхний уровень	0,9	140	45	
Нижний уровень	0,5	130	15	
Кодированные значения факторов	x_1	x_2	x_3	
Опыты				
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

Какое решение кажется вам целесообразным?

Постановка новой серии опытов, связанная с изменением интервалов варьирования и переносом центра. Обратитесь к пункту 15.

Движение по градиенту. Обратитесь к пункту 16.

Не знаю. Обратитесь к пункту 17.

13. Вы считаете, что 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 ($y_1 = 95\%$).

Вернитесь к пункту IV и выберите другой ответ.

14. Вы полагаете, что следует переходить к планированию второго порядка. По условию задачи важно было увеличить выход за счет двух-трех опытов. Этому отвечает движение по градиенту.

Вернитесь в пункте IV к одному из вариантов.

15. Вы считаете, что следует изменить интервалы варьирования факторов и получить адекватную модель. Что же, в данной ситуации это 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 8 опытов.

Опыты дороги, поэтому вам следует в пункте 16 ознакомиться и с другой процедурой.

16. Вы полагаете, что следует переходить к движению по градиенту. Проанализируем это 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.

Обратитесь к пункту VI.

17. Вы не знаете, какое выбрать решение: постановку новой серии опытов или движение по градиенту. Прежде всего оценим затраты, связанные с принятием первого или второго решения. В первом случае требуется реализовать еще 8 опытов, во втором — два-три опыта. Конечно, первое решение представляется достаточно корректным, во втором случае имеется риск не получить улучшения результатов. Но, так как опыты в данном случае длительны и дороги, попытка экономии числа опытов является разумной.

Поэтому вернитесь к пункту V и выберите один из вариантов решения.

VI. При выполнении этой работы исследователи выбрали движение по градиенту и улучшили результаты в 2 раза (см. гл. 10).

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.

9.3. 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 добавления эффектов взаимодействия может не хватить степеней свободы для проверки гипотезы адекватности и потребуется реализация еще двух-трех опытов внутри области эксперимента.

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. 220): метод «перевала», достройка до полного факторного эксперимента, до дробной реплики, для которой ранее смешанные эффекты становятся «чистыми», достройка до плана второго порядка.

Еще один, хотя и не очень распространенный, прием — образование зависимых и независимых переменных *. Этот вопрос был упомянут в гл. 2. Его подробное рассмотрение выходит за рамки нашей книги.

Наконец, если не удалось все же 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

* В. В. Налимов, Н. А. Чернова. Указ. соч.

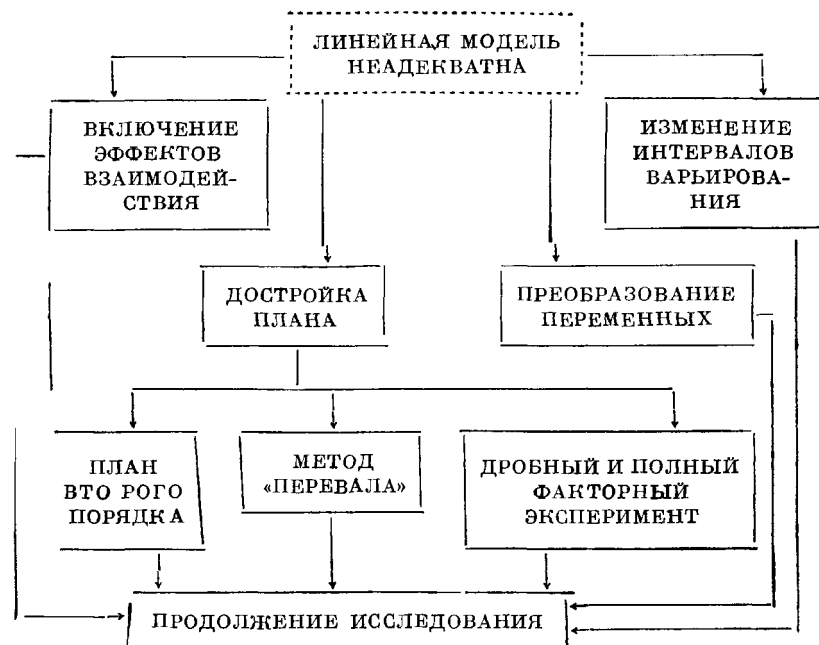


Рис. 25. Блок-схема принятия решений в задаче построения интерполяционной формулы, линейная модель неадекватна

неадекватна (рис. 25). Если линейная модель адекватна, то задача решена.

Приведем пример.

Пример 6

Проводилось исследование по получению математической модели ядичного экстрактора. В качестве факторов были выбраны:

- $\tilde{x}_1$ — диаметр турбинки, мм,
- $\tilde{x}_2$ — скорость вращения турбинки, об/мин,
- $\tilde{x}_3$ — температура, °C,
- $\tilde{x}_4$ — концентрация свободной кислоты в водном растворе, г/кг/л,
- $\tilde{x}_5$ — высота слоя жидкости в ячейке, мм,
- $\tilde{x}_6$ — соотношение фаз в эмульсии.

Параметр оптимизации — продолжительность полного рас-
слаивания, мин. Условия, матрица планирования и результаты опы-
тов приведены в табл. 9.3. Использована 1/4-реплика от полного

факторного эксперимента 2⁶. Линейное уравнение регрессии оказа-
лось неадекватным. Затем были введены 3 несмешанных между собой
эффекта взаимодействия факторов, имеющих наибольшую абсолют-
ную величину

$$y = 12,16 + 0,53 x_1 + 0,53 x_2 - 1,38 x_3 - 3,22 x_4 + 1,44 x_5 - \\ - 0,62 x_6 - 0,84 x_1 x_4 - 0,50 x_1 x_6 - 0,78 x_2 x_4.$$

Это уравнение адекватно описывает процесс, $s^2 \{y\} = 0,39$.

Таблица 9.3

Матрица планирования и результаты опытов

	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$	$\tilde{x}_6$	y
Основной уровень	90	600	26	0,40	195	0,8115	
Интервал варьирования	10	100	4	0,29	25	0,0975	
Верхний уровень	100	700	30	0,69	220	0,909	
Нижний уровень	80	500	22	0,11	170	0,714	
Кодированные значения факторов	x_1	x_2	x_3	x_4	x_5	x_6	
Опыты							
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

Рассчитанное значение $F_{\text{эксп}} = 2,4$ при табличном значении $F = 2,7$. Уравнение было использовано при проектировании про-
мышленного аппарата.

Вот один из возможных приемов построения интерполяционной
модели.

9.4. 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 коэффициентами регрессии возможны: движение по градиенту, план второго порядка, окончание исследования. Если часть коэффициентов регрессии не значима, 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, выбирают 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 варьирования, перенос центра плана, достройка плана, 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 квадра-

тичных 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.

Глава десятая

Крутое восхождение по поверхности отклика

Решения, которые обсуждались в предыдущей главе, направлены на то, чтобы обеспечить эффективное движение по градиенту. Давайте посмотрим, как на практике осуществить это движение.

10.1. Движение по градиенту

Посмотрите на рис. 26. 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 А 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. 26 это направление AB , перпендикулярное линиям уровня. Градиент непрерывной однозначной 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

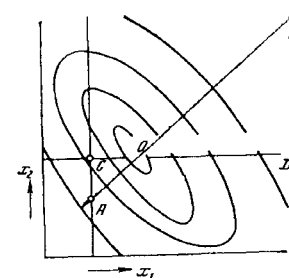


Рис. 26. Движение по поверхности отклика [методами однофакторного эксперимента и градиента]

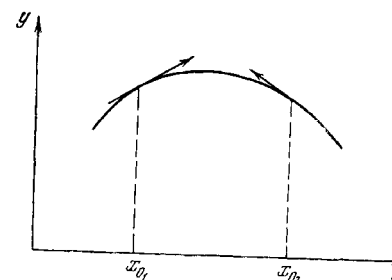


Рис. 27. Зависимость знака градиента от формы поверхности и положения нулевой точки

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**.

1. 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. 27).

Пусть интервал варьирования некоторого значимого фактора равен 10 единицам. Как вы считаете, изменится ли составляющая градиента, если в качестве единицы измерения воспользоваться вначале миллиметром, а затем дюймом?

Да. Обратитесь к пункту 1.
 Нет. Обратитесь к пункту 2.
 Не знаю. Обратитесь к пункту 3.

1. Правильно. Составляющая градиента не инвариантна к выбору размерности фактора. 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 точки 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.

Переходите к параграфу 10.2.

2. Вы полагаете, что выбор размерности не существен. Это не так. Переход от миллиметров к дюймам эквивалентен значительному увеличению интервала варьирования.

Вернитесь к пункту 1 и выберите другой ответ.

3. Так как вы не знаете, что ответить, давайте посмотрим, к чему приводит переход от миллиметров к дюймам.

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.

Вернитесь к пункту 1 и выберите правильный ответ.

10.2. Расчет крутого восхождения

Технику расчета крутого восхождения удобно рассмотреть на простейшем примере в случае одного фактора (рис. 28).

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.

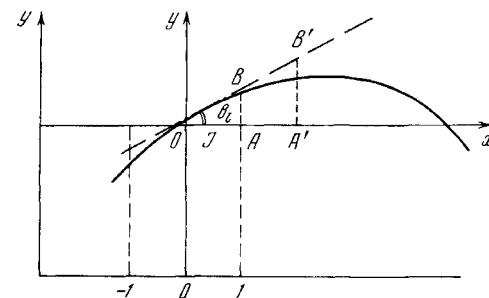


Рис. 28. Расчет координат точек в направлении градиента

Обобщение на случай k факторов делается механически, так как все эффекты независимы друг от друга. Существенно только соотношение произведений коэффициентов на соответствующие 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. Шаги получаются, если 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 требует значительного числа опытов при движении к оптимуму. Большой шаг увеличивает вероятность проскока области оптимума. Во всяком случае, аналогично выбору интервалов варьирования (гл. 5), нижняя граница задается возможностью фиксирования двух соседних опытов, а верх-

няя 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.

Рассмотрим пример расчета крутого восхождения для процесса ионнообменного разделения смеси редкоземельных элементов (см. стр. 81, 224).

Пример 1

В табл. 10.1 приведены условия, матрица планирования и результаты серии опытов, а также расчет крутого восхождения.

II. Приведем этапы расчета крутого восхождения.

1. Расчет составляющих градиента.

$$b_1 \times J_1 = -1,0, \quad b_2 \times J_2 = -4,5.$$

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. 10.1 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. Получа-

Таблица 10.1

Матрица планирования, результаты и расчет крутого восхождения

	$\bar{x}_1$	$\bar{x}_2$	y^*
Основной уровень	1,5	7,0	
Интервал варьирования	0,5	1,0	
Верхний уровень	2,0	8,0	
Нижний уровень	1,0	6,0	
Кодированные значения переменных	x_1	x_2	
Опыты			
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	
$b_j \times$ интервал варьирования	-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.

ем серию опытов крутого восхождения (в табл. 10.1 — 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

Правильно. Обратитесь к пункту 4.
Неправильно. Обратитесь к пункту 5.
Не знаю. Обратитесь к пункту 6.

4. Вы считаете, что серия мысленных опытов рассчитана правильно. Вы правы. Действительно, мы уменьшили составляющую градиента 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 на стр. 240 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. 10.1). Мы собираемся для оценки параметра оптимизации использовать уравнение регрессии: $y = 88,0 - 2,0 x_1 - 4,5 x_2$. Однако в табл. 10.1 приведены натуральные 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. 86): $x_j = (\bar{x}_j - \bar{x}_{j0})/J_j$, $x_1 = -0,6$, $x_2 = -1,5$. 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.

Рассмотрим еще один пример по расчету крутого восхождения для процесса экстракции гафния трибутилфосфатом.

Пример 2.

В табл. 10.2 так же, как и в табл. 10.1, приведена стандартная форма представления условий, матрицы планирования и результатов опытов вместе с крутым восхождением.

5. Вы полагаете, что серия мысленных опытов рассчитана неправильно. Это не так.

Еще раз прочтите на стр. 238 о пропорциональном изменении составляющих градиента и в пункте II выберите другой ответ.

6. Вы не знаете, правильно ли рассчитана серия опытов крутого восхождения. Проведем этот расчет. Составляющие градиента уменьшаются в 5 раз: $b_j \times J_j = -1,0$, $\frac{1,0}{0,2} = 5$. Следовательно, шаг по $\bar{x}_2$ будет равен $-0,9$, с округлением $-1,0$.

Теперь составьте серию опытов крутого восхождения последовательным прибавлением шагов к основному уровню факторов ($\bar{x}_1 = 1,5$, $\bar{x}_2 = 7,0$), сравните с серией опытов в пункте II и там же выберите другой ответ.

Остались не рассмотренными 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 -коэффициенты 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

Таблица 10.2

Матрица планирования, результаты опытов и крутое восхождение

	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$y \cdot 10^3$ *
Основной уровень	3,0	30	1,5	15	
Интервал варьирования	2,0	40	1,0	10	
Верхний уровень	5,0	40	2,5	25	
Нижний уровень	1,0	20	0,5	5	
Кодированные значения факторов Опыты	x_1	x_2	x_3	x_4	y
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	0,0495	0,0203	-0,0137	-0,0066	
$b_j \times I_j$	0,0390	0,203	-0,0137	-0,066	
Шаг при изменении $\tilde{x}_2$ на 5	0,9605	5	-0,3399	-1,6256	
Округление	1,0	5,0	-0,3	-2,0	
Опыты					
9	4,0	35,0	1,2	13	
10	5,0	40,0	0,9	11	
11	6,0	45,0	0,6	9	
12	7,0	50,0	0,3	7	
13	8,0	55,0	0	5	
14	9,0	60,0	0	3	

* Среднее значение и 2-х параллельных опытов, рандомизированных во времени.

выгоднее вновь поставить эксперимент, изменив интервалы варьирования, а не двигаться по градиенту. Так, в работе * после первой серии опытов получено следующее уравнение регрессии:

$$y \cdot 10^3 = 22,48 - 1,01x_1 + 19,46x_2 + 1,46x_3 - 1,05x_4.$$

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:

$$y \cdot 10^3 = 65,5 + 19,5x_1 + 20,3x_2 - 13,8x_3 - 6,6x_4.$$

На этот раз функция оказалась симметричной и было реализовано крутое восхождение, показанное на странице 243.

III. Вы видели, что при расчете градиента движение начиналось из нулевой точки. Как вы считаете, можно ли начинать движение из наилучшего опыта матрицы планирования, если использовать только линейные коэффициенты регрессии?

Можно. Обратитесь к пункту 7.

Нельзя. Обратитесь к пункту 8.

Не знаю. Обратитесь к пункту 9.

7. Вы полагаете, что можно двигаться из наилучшей точки матрицы планирования. К сожалению, это неверно. Вы, вероятно, забыли, что направление градиента единственно. Но, может быть, вы считаете, что мы ошибались, отдавая предпочтение нулевой точке?

Неизвестная функция отклика разлагалась в ряд Тейлора в окрестностях нулевой точки, к которой и относится оценка градиента.

Вернитесь к пункту III и выберите другой ответ.

* Л. Н. Комиссарова, Ю. В. Грановский, Н. М. Пруткова, Ю. П. Адлер, В. В. Налимов, В. И. Спицын. Определение оптимальных условий экстракции микроколичеств гафния трибутилфосфатом. — Заводская лаборатория, 1963, 29, № 1.

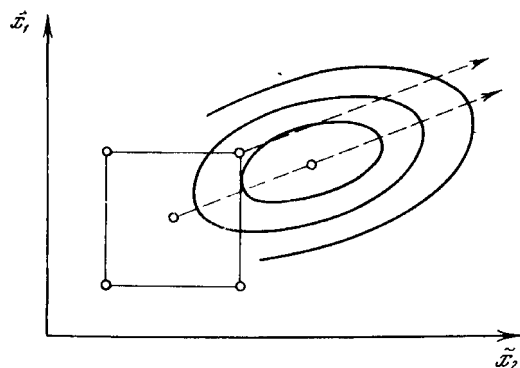


Рис. 29. Движение по градиенту из нулевой точки и из наилучшей точки плана

8. Вы правы. Направление градиента определяется единственным способом, и движение должно начинаться из нулевой точки. На рис. 29 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.

Переходите к параграфу 10.3.

9. Вы не знаете, что ответить. Чтобы помочь вам найти правильный ответ, напомним:

- а) направление градиента определяется однозначно;
- б) функция отклика разлагается в ряд Тейлора в окрестности нулевой точки, центре эксперимента. К этой точке и относятся оценки составляющих градиента.

Вернитесь к пункту III и постарайтесь выбрать правильный ответ.

10.3. 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.

IV. 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 1—2 опыта выполняют в области эксперимента.

Как вы считаете, для мысленных опытов процесса ионообменного разделения (табл. 10.1) с какого опыта имеет смысл ставить эксперимент?

10. Начиная с пятого опыта ($\tilde{x}_1 = 1,4$, $\tilde{x}_2 = 6,5$). Обратитесь к пункту 10.

11. Начиная с седьмого опыта ($\tilde{x}_1 = 1,2$, $\tilde{x}_2 = 5,5$). Обратитесь к пункту 11.

12. Не знаю. Обратитесь к пункту 12.

10. Вы полагаете, что начинать реализацию мысленных опытов следует с пятого опыта, условия которого $\tilde{x}_1 = 1,4$, $\tilde{x}_2 = 6,5$. Но ведь значения факторов для этого опыта находятся внутри области эксперимента (основной уровень $\tilde{x}_1 = 1,5$, $\tilde{x}_2 = 7,0$, $J_1 = 0,5$, $J_2 = 1,0$). Так как модель адекватна, то значение параметра оптимизации для этого опыта можно получить просто расчетом: $\hat{y}_1 = 90,65$. Поэтому ваш ответ неверен.

Вернитесь в пункте IV к одному из вариантов.

11. Вы полагаете, что начинать следует с седьмого опыта ($\tilde{x}_1 = 1,2$, $\tilde{x}_2 = 5,5$). Вы совершенно правы. Этот опыт по фактору x_2 уже выходит за пределы области эксперимента.

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 последовательной стратегией, так как минимизация числа опытов приобретает большую актуальность.

V. 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 (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 5-й опыт в серии (стр. 243) $y_5 = 15,1$, а реализованный десятый опыт $y_{10} = 30,4$, 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. 221) 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 $b_0 = 55,9$, максимальный выход — $y = 65,3$. Можно ли было при последовательном поиске уменьшить число опытов?

Можно. Обратитесь к пункту 13.

Нельзя. Обратитесь к пункту 14.

Не знаю. Обратитесь к пункту 15.

12. Вы не можете дать ответ, с какого опыта следует начинать реализацию крутого восхождения. Прежде всего надо установить, адекватна или неадекватна модель в серии опытов. Для нашего случая модель адекватна (стр. 239). Условия первого опыта хотя бы по одному из факторов должны выходить за область эксперимента. Область эксперимента задается значениями основных уровней факторов и интервалами варьирования: $\bar{x}_1 = 1,5$, $\bar{x}_2 = 7,0$, $J_1 = 0,5$, $J_2 = 1,0$. Выполняется ли это условие для первого опыта?

Вернитесь в пункте IV к одному из вариантов.

13. Вы считаете, что при последовательном поиске можно было бы сократить число опытов. Вы правы. Действительно, уже первый опыт не дал улучшения результатов. Однако возражений против проведения 2-го или даже 3-го

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 (стр. 226).

Таблица 10.3

Крутое восхождение при оптимизации процесса получения карбометоксисульфанилгуанидина

Факторы	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	y
$b_j \times J_j$	0,356	51,40	140,40	
Шаг при изменении $\tilde{x}_2$ на 5	0,0346	5	13,60	
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

Крутое восхождение оказалось весьма эффективным: средний выход в предыдущей серии опытов $b_0 = 23,28$, максимальный выход $y = 46,80$ (стр. 228).

Пример 4

Оптимизация процесса получения производного пиперазина (табл. 10.4)

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.

14. Вы считаете, что при последовательном поиске нельзя было бы уменьшить число опытов крутого восхождения.

Таблица 10.4

Крутое восхождение

Факторы	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{x}_5$	$y, \%$
Интервал варьирования $\times b_j$	$-1,775 \cdot 0,025 = -0,444$	$5,05 \cdot 0,25 = 1,26$		$-2,10 \cdot 5 = -10,5$		
Шаг при $x_5 = 40$	-0,0845	-0,24		-2		
Округление значения шага	0,08	0,25	—	-2		
Нулевой уровень	1,25	1,25	4,00	25	40	55,0
Мысленный опыт № 9	1,17	1,50	4,00	23	40	
То же № 10	1,09	1,75	4,00	21	40	
Реализуемый опыт № 11	1,01	2,00	4,00	19	40	68,8
То же № 12	0,93	2,25	4,00	17	40	74,3
То же № 13	0,85	2,50	4,00	15	40	66,8

Вы ошиблись. Если верно предположение об одноэкстремальной поверхности отклика и получены последовательно величины 60,1; 59,7; 57,8, то максимум не может быть в следующих опытах.

Вернитесь в пункте V к одному из вариантов.

15. Вы не знаете, можно ли уменьшить число опытов при последовательном поиске.

Мы просим вас в пункте V еще раз ознакомиться с различными вариантами последовательной постановки опытов в крутом восхождении и выбрать правильный вариант.

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

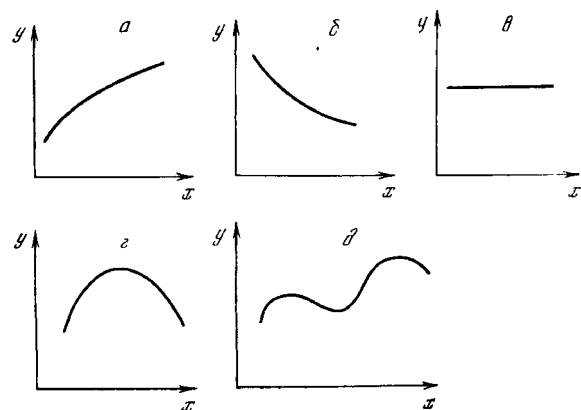


Рис. 30. Ситуации при движении по градиенту

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 градиенту 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. 30.

Наиболее благоприятные случаи даны на рис. 30, 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 б, где нарушена предпосылка одноэкстремальности.

Случай в иллюстрирует неэффективное крутое восхождение. Вместо ожидаемого увеличения параметра опти-

мизации 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.

10.4. Резюме

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 1—2 опыта выполняют в области эксперимента. Возможно проведение сразу всех мысленных опытов. Более экономная процедура состоит в проведении 2—3-х 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.

Глава одиннадцатая

Принятие решений после крутого восхождения

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.

11.1. 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 построить 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.

I. Область оптимума не достигнута. В этом случае ставится линейный план следующего цикла и исследование продолжается.

В предыдущей главе на стр. 249 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 %.

Какому из приведенных ниже решений вы можете отдать предпочтение?

Окончить исследование. Обратитесь к пункту 1.

Построить план второго порядка для исследования области оптимума. Обратитесь к пункту 2.

Построить линейный план следующего цикла крутого восхождения. Обратитесь к пункту 3.

1. Вы отдаете предпочтение решению окончить исследование. Мы обратили ваше внимание на предположение, что

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.

Вам нужно возвратиться к пункту I и выбрать другое решение.

2. Вы решили приступить к построению плана второго порядка для исследования области оптимума. Область оптимума достаточно далека — имеется резерв более чем 20%. Еще использованы не все попытки пробиться в область более высоких выходов.

Возвратитесь к пункту I и подумайте о выборе другого ответа.

3. Ваше 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 в данной работе.

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. 5).

Обратимся к примеру оптимизации процесса получения карбометоксисульфанилгуанидина (матрица планирования первого цикла приведена на стр. 228, крутое восхождение — на стр. 249).

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 увеличить x_1 нежелательно. В ходе работы появились новые соображения в пользу существенного уменьшения значений x_1 . 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. 11.1.

Таблица 11.1

Вторая серия опытов на стадии крутого восхождения

	x_1	x_2	x_3	
+1	0,7	166	70	
0	0,5	160	55	
—1	0,3	154	40	
Интервал варьирования	0,2	6	15	

№ опытов	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

Мы предлагаем вам подсчитать для этой матрицы коэффициенты, дисперсию адекватности линейной модели и критерий Фишера и ответить, адекватно ли линейное приближение ($s_{\{y\}}^2 = 0,97$).

Правильный ответ посмотрите в пункте 4.

4. Уравнение регрессии получается в виде полинома первой степени: $y = 75,91 - 3,524x_1 + 3,247x_2 + 2,436x_3$.

Таблица 11.2

Расчет дисперсии линейной модели

$\bar{y}$	$\hat{y}$	$\bar{y} - \hat{y}$	$(\bar{y} - \hat{y})^2$
76,00	78,06	2,06	4,24
74,05	71,56	2,49	6,25
80,90	78,61	2,29	5,29
73,00	73,74	0,74	0,55
76,81	73,19	3,62	13,10
62,65	66,63	3,98	15,84
81,40	80,24	1,16	1,34
82,40	85,10	2,70	7,29

$$s_{\text{адекв}}^2 = \frac{53,90}{5} = 13,47,$$

$$F_{\text{эксп}} = \frac{13,47}{0,97} = 13,08,$$

$$F_{\text{табл}} = 4,12.$$

Линейное приближение неадекватно.

Ниже приведены два решения. Какому из них вы отдадите предпочтение?

Движение по градиенту. Обратитесь к пункту 5.

Переход к нелинейному планированию. Обратитесь к пункту 6.

5. Вы предлагаете движение по градиенту, несмотря на то, что линейное приближение неадекватно и область оптимума близка. Нужно предостеречь вас: вероятность успеха мала. Конечно, можно рассчитать опыты и некоторые из них реализовать. Ведь 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.

Что получилось при крутом восхождении, вы можете увидеть в пункте II, табл. 11.3.

6. Формально вы правы. Линейное приближение неадекватно, и мы находимся в области довольно высоких выходов реакции. Это говорит о том, что область оптимума близка.

В таких случаях следует переходить к нелинейному планированию. Однако, если опыты 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.

В пункте II вы можете увидеть, что дала реализация опытов на стадии крутого восхождения.

II. Таблица 11.3

Крутое восхождение после второй серии опытов

	x_1	x_2	x_3	y
$b_j \times J_j$	$-3,524 \times 0,2 =$ $= -0,705$	$3,247 \times 6 =$ $= 19,436$	$2,436 \times 15 =$ $= 36,540$	
Шаг при изменении x_1 на 0,1	-0,1	2,73	5,17	
Округление шага № 9 (нулевой уровень)	-0,1	3	5	
№ 10	0,5	160	55	77,3
№ 11	0,4	163	60	82,2
№ 12	0,3	166	65	—
№ 13	0,2	169	70	83,6
№ 14	0,1	172	75	84,8
	0	175	80	36,9

Выход реакции повышен примерно до 85%. По сравнению с выходом в центре эксперимента это — увеличение на 7,6%, а по сравнению с лучшим опытом в матрице — всего лишь на 2,4% ($s_{iy}^2 = 0,97$). Тем не менее крутое восхождение оказалось весьма примечательным. В опыте № 13 максимальный выход получен при значении $x_1 = 0,1$. До планирования эксперимента считалось, что процесс получения карбометоксисульфанилгуанидина может успешно протекать при $x_1 > 0,7$, а выход реакции более 70% был неизвестен.

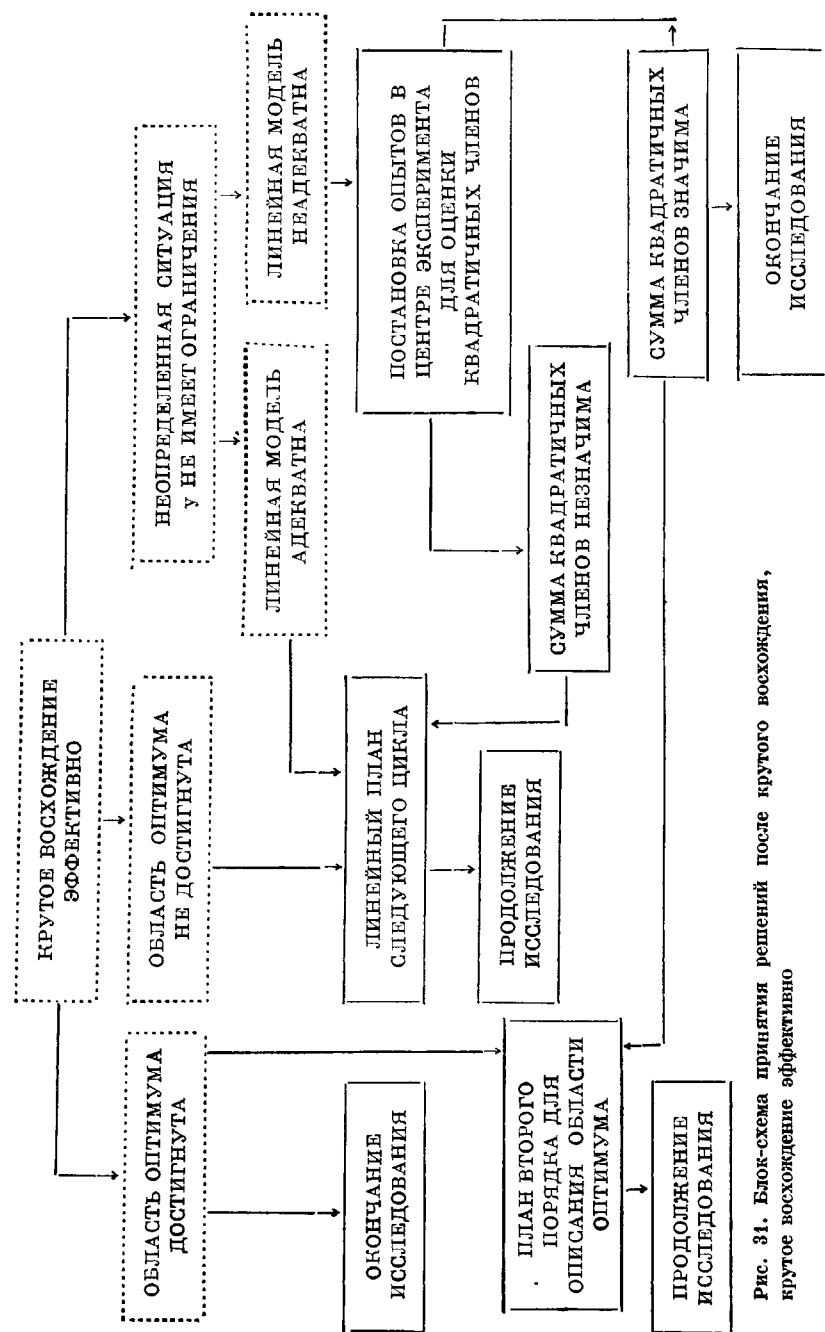


Рис. 31. Блок-схема принятия решений после крутого восхождения, крутое восхождение эффективно

В данном случае крутое восхождение в некорректном применении (в условиях нелинейности и близости к области оптимума), привело в область факторного пространства, где при $x_1 = 0,1$ 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 блок-схеме рис. 31.

11.2. Крутое восхождение неэффективно

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. 32.

Рассмотрим каждую ситуацию отдельно.

III. 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. 227, чтобы выяснить причины неадекватности линейной модели.

Пример 1.

Имеется следующая ситуация: исходной план — полуреблика, линейная модель неадекватна, крутое восхождение оказалось неэффективным, область оптимума близка. Параметром оптимизации является выход полезного продукта. Максимально возможный выход — 100%. При

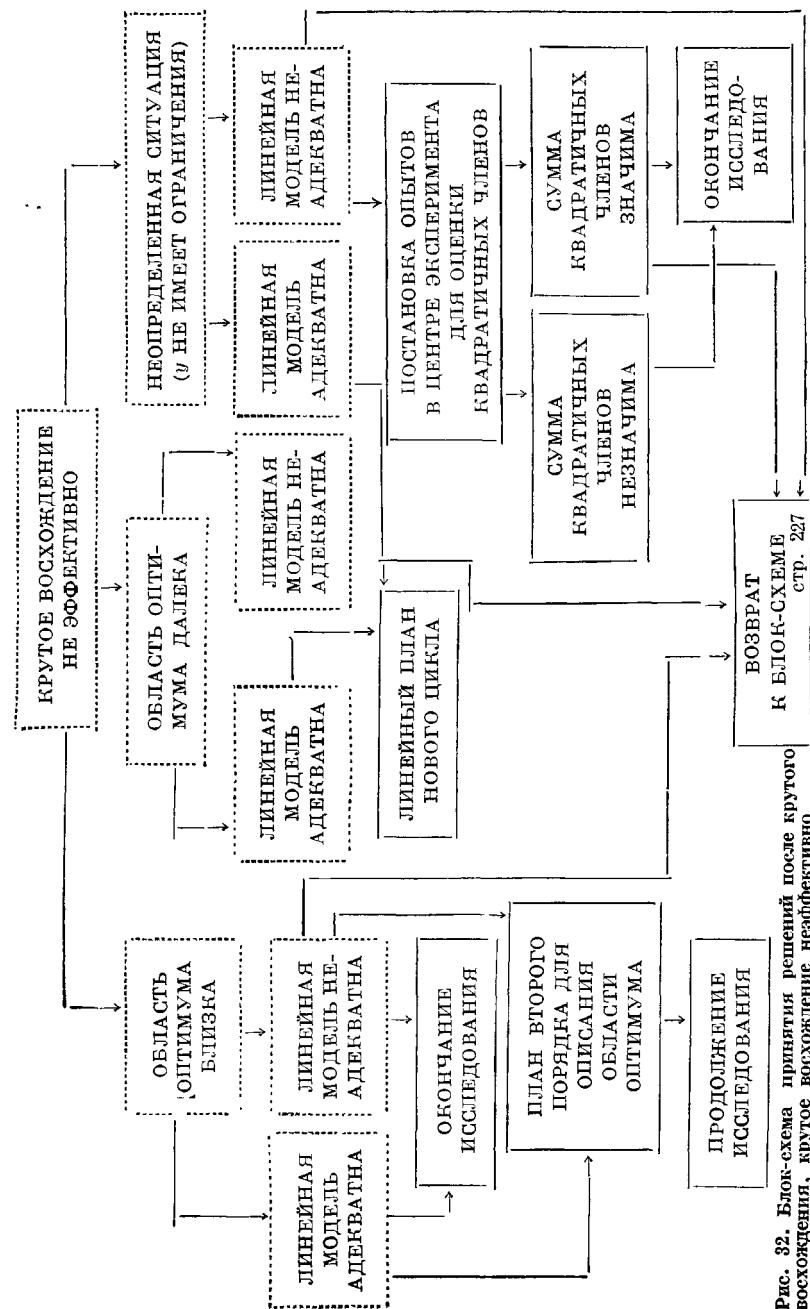


Рис. 32. Блок-схема принятия решений после крутого восхождения, крутое восхождение неэффективно

реализации полуреплики получен наибольший выход — 80 %. Ошибка опыта — 1 %.

Какому из трех решений вы отдадите предпочтение?

Окончить исследование. Обратитесь к пункту 7.

Перейти к нелинейному планированию второго порядка. Обратитесь к пункту 8.

Достроить полуреплику до полного факторного эксперимента. Обратитесь к пункту 9.

7. Вы решили 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 ставилась цель только приблизиться к области высокого выхода.

Возвратитесь к пункту III и подумайте о выборе другого решения.

8. Вы решили достроить 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 целесообразным. 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.

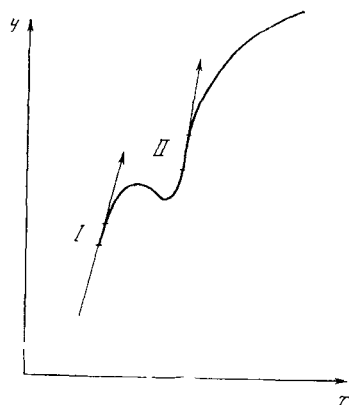
Возвратитесь к пункту III.

9. Вы решили 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 кру-

Рис. 33. Иллюстрация крутого восхождения для случая: «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



тое 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 отклика гладкая и одноэкстремальная. В действительности она может иметь, например, вид, показанный на рис. 33.

В таких случаях целесообразно передвинуться в другую область факторного пространства и построить линейный план второго цикла крутого восхождения.

IV. Область оптимума далека. Линейная модель неадекватна. Здесь возможно единственное решение — возвратиться к блок-схеме гл. 9 и выяснить причины неадекватности линейной модели. Напомним некоторые причины, вследствие которых крутое восхождение могло оказаться неэффективным.

а) Интервалы варьирования выбраны неудачно.

б) Исходная модель строилась по полуреplikе. Нужно достроить полуреplikу до полного факторного эксперимента, получить отдельные оценки для всех коэффициентов регрессии и совершить новое крутое восхождение.

в) Исходная модель строилась по дробной реплике 2^{n-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 2.

Перед вами (табл. 11.4) план 2^{4-1} с генерирующим соотношением $x_4 = x_1x_2x_3$. Этот план применялся при оптимизации процесса получения новокаина. Здесь x_1 — время реакции, мин, x_2 — температура реакционной среды, °C, x_3 — избыток натриевой соли парааминобензойной кислоты, %, x_4 — концентрация натриевой соли парааминобензойной кислоты, %. Параметром оптимизации является выход реакции, %.

В этом примере линейное приближение оказалось неадекватным и крутое восхождение неэффективным. Какое решение целесообразно принять?

Построить новый план, уменьшив интервалы варьирования. Это даст возможность избавиться от эффектов взаимодействий и, возможно, сделать линейное приближение адекватным. Обратитесь к пункту 10.

Не знаю. Обратитесь к пункту 11.

Достроить линейный план до плана второго порядка. Обратитесь к пункту 12.

Достроить полуреplikу до полного факторного эксперимента с тем, чтобы освободить линейные эффекты от смешивания с взаимодействиями второго порядка. Обратитесь к пункту 13.

10. Вы решили построить новый план, уменьшив интервалы варьирования, с надеждой получить адекватную линейную модель.

Обратите внимание на величину $ts_{\{b_j\}}$.

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, может оказаться, что ни один линейный эффект не выделится на фоне ошибок.

Вернитесь к пункту IV и выберите другой ответ.

Таблица 11.4

Матрица планирования 24^{-1} и результаты эксперимента

№ опытов	x_0	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$x_1x_2 + x_3x_4$	$x_1x_3 + x_2x_4$	$x_2x_3 + x_1x_4$	y
	-1	45	50	0	3,5				
	0	60	60	4	6,5				
	+1	75	70	8	9,5				
		15	10	4	3				
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_i	58,65	0,70	0,15	-0,75	-3,05	-0,80	0,10	0,65	

$s_{\text{воспр}}^2 = 1:$

$s_{\{b_i\}}^2 = 0,125$

$2,3 s_{\{b_i\}} = 0,81.$

11. Вы не знаете, как поступить. Давайте рассуждать вместе. Мы имеем полуреплику 2^{4-1} с генерирующим соотношением $x_4 = x_1 x_2 x_3$. При этом получаются следующие смешанные оценки:

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

Принимая такую систему смешивания, мы исходим из предположения, что эффекты взаимодействия второго порядка близки к нулю (во всяком случае они меньше, чем эффекты взаимодействия первого порядка), поэтому они мало исказят истинные значения линейных эффектов. Вследствие этого движение по градиенту не будет нарушено. Линейное приближение оказалось неадекватным и крутое восхождение — неэффективным. Как поступать в таком случае? Имеются следующие возможности. Первая — построить новый план, уменьшив интервалы варьирования факторов в надежде получить адекватную линейную модель.

Вторая — достроить полуреплику до полного факторного эксперимента с тем, чтобы получить все оценки коэффициентов раздельными.

Однако уменьшение интервалов варьирования приведет к потере значимости линейных эффектов. Даже при существующих интервалах варьирования значим только один линейный эффект.

Можно было бы дать противоположную рекомендацию — увеличить интервалы варьирования факторов.

Возвратитесь к пункту IV и выберите нужное решение.

12. Вы решили достроить линейный план до плана второго порядка. Такое решение можно было принять, если бы эксперимент проводился в области более высоких значений параметра оптимизации.

Возвратитесь к пункту IV и выберите другой ответ.

13. Вы хотите достроить полуреплику до полного факторного эксперимента и совершить новое крутое восхождение при неискаженных линейных оценках. Мы разделяем ваше мнение. Вот как выглядит достроенный план (табл. 11.5).

Таблица 11.5

Матрица планирования ξ^4

№ п/п	x_0	x_1	x_2	x_3	x_4	$y, \%$	№ п/п	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	+	—	+	+	—	61,0	15	+	—	+	+	+	66,0
8	+	+	+	+	—	71,5	16	+	+	+	+	+	55,4

Полученное уравнение регрессии имеет вид

$$y = 60,85 + 0,62x_1 + 0,72x_2 + 1,36x_3 - 2,99x_4 - \\ - 0,30x_1x_2 + 0,72x_1x_3 - 0,25x_1x_4 + 0,93x_2x_3 - \\ - 0,65x_2x_4 + 0,78x_3x_4 + 0,034x_1x_2x_3 - 2,13x_1x_2x_4 - \\ - 0,43x_1x_3x_4 - 0,12x_2x_3x_4 - 2,15x_1x_2x_3x_4,$$

$$s_{\text{воспр}}^2 = 1,67; \quad s_{\{b_j\}}^2 = 0,104; \quad s_{\{b_j\}} = 0,325, \\ 2s_{\{b_j\}} = 0,65.$$

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. 11.6).

Таблица 11.6

Крутое восхождение ($x_1 = 60$ мин)

	x_2	x_3	x_4	$y, \%$
$b_1 \times J_1$	$0,72 \cdot 10 = 7,2$	$1,36 \cdot 4 = 5,44$	$-2,99 \cdot 3 = -8,97$	
Шаг	0,802	0,606	—1	
Округленный шаг	0,8	0,6	—1	
№ 17	60	4	6,5	
№ 18	60,8	4,6	5,5	
№ 19	61,6	5,2	4,5	
№ 20	62,4	5,8	3,5	65,8
№ 21	63,2	6,4	2,5	79,3
№ 22	64,0	7,0	1,5	63,3

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: $b_{11} = 5,7$, $b_{22} = -5,5$, $b_{11} + b_{22} = +0,4$. Сумма незначима, 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.

11.3. 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. 9. 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 исследование. При незначимой сумме квадратичных членов нужное решение выбрать довольно трудно. Наиболее типичным является построение линейного плана нового цикла.

Глава двенадцатая

Обсуждение результатов

Вот мы и добрались до последней главы. 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. Надо, наконец, указать литературу.

12.1. 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 лишнее 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.

I. Давайте рассмотрим пример (см. гл. 9, пример 6). Исследовался аппарат для экстракционного разделения редкоземельных элементов. Задача ставилась так: найти

интерполяционную формулу, пригодную для расчета смесительно-отстойных экстрактов 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}):

$$y = 12,16 + 0,53x_1 + 0,53x_2 - 1,38x_3 - 3,22x_4 + 1,44x_5 - 0,62x_6 - 0,84x_1x_2 - 0,50x_1x_6 - 0,78x_2x_4.$$

Проверка полученных рекомендаций в аппарате размером 325×325 мм, что представляло практический интерес, дала отклонения от уравнения в разных точках 5 — 7%.

Как вы считаете, можно ли признать работу законченной? (Имеется в виду переход от размеров 300×200 к 325×325 мм.)

Да. Обратитесь к пункту 1.

Нет. Обратитесь к пункту 2.

Не знаю. Обратитесь к пункту 3.

1. Вы думаете, что работу можно считать оконченной. На чем основывается ваше утверждение?

Очевидно, оно не может быть верным во всех случаях. Ответ зависит от того, является ли указанная точность достаточной. Но вам это не известно.

Вернитесь к пункту 1 и выберите другой ответ.

2. Вы предлагаете продолжить исследование. Но почему? Быть может, полученная точность удовлетворяет исследователей. Тогда можно считать, что работа окончена.

Вы не знаете, какая точность достаточна, поэтому вернитесь к пункту 1 и выберите другой ответ.

3. Вы не знаете, что сказать относительно окончания данного исследования. Это вполне понятно. Чтобы ответить, надо знать, какая точность удовлетворяет исследователя.

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.

II. Теперь нам предстоит менее приятное занятие — рассмотреть случаи, когда цели не удалось достигнуть. 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 они 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 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. Какой из диагнозов вы бы предпочли?

Имеет место регулярный временной дрейф. Обратитесь к пункту 4.

Не включен в рассмотрение важный фактор, произвольно изменяющийся от серии к серии. Обратитесь к пункту 5.

Оба эти обстоятельства действуют одновременно. Обратитесь к пункту 6.

Не знаю. Обратитесь к пункту 7.

4. Вы полагаете, что имеет место временной дрейф. Это очень правдоподобное объяснение.

Действительно, 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. Во всяком случае гипотеза о дрейфе должна быть высказана

Вернитесь к пункту II и рассмотрите другие возможности.

5. Вы полагаете, что существует важный неучтенный фактор, который свободно меняется. Это вполне возможно.

Если фактор пропущен и он флуктуирует, как ему вздумается, то может возникнуть наша ситуация. К сожалению, интерпретация не является однозначной, ибо та же картина возможна и в случае временного дрейфа. Во всяком случае гипотеза о неучтенном факторе должна быть высказана.

Вернитесь к пункту II и рассмотрите другие возможности.

6. Жизнь рисуется вам в мрачных тонах. Вы проводите стратегию осторожного экспериментатора: лучше сразу предположить худшее и искать выход, чем потом разочаровываться.

В этом есть резон. Мы действительно не можем сказать, по какой причине наблюдается данная картина.

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.

Вернитесь к пункту II и рассмотрите другие возможности.

7. Вы не знаете, что выбрать. Ну что же, давайте посмотрим, откуда взялись предложенные варианты.

Ясно, что нестабильная картина связана с какими-то явлениями, не описываемыми моделью. Мы не можем предположить, что причиной всему просто дисперсия воспроизводимости, так как крутое восхождение оказывалось эффективным. Можно высказать гипотезу о дрейфе. Если, например, по мере работы объекта его качество постепенно ухудшается, то, повторив опыт в идентичных условиях через значительное время, мы получим значимо отличный результат.

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.

Итак, если вы не обращались к другим вариантам, то, прежде чем читать дальше, вернитесь к пункту II и познакомьтесь с ним. Если это уже сделано, то давайте рассмотрим еще пример.

III. Многократно реализовав шаговую процедуру, вы получили некоторый наилучший результат, который, однако, хуже требуемого. Опыты хорошо воспроизводятся. Все дальнейшие попытки улучшения ни к чему не привели.

Какое из перечисленных ниже объяснений вам больше нравится?

Достигнут предел возможностей данного объекта, предъявляемым требованиям он не может удовлетворить. Обратитесь к пункту 8.

На поверхности отклика существует несколько экстремумов и найден не наилучший. Обратитесь к пункту 9.

Не знаю. Обратитесь к пункту 10.

8. Вы полагаете, что при постановке задачи возможности объекта были переоценены и из него ничего больше извлечь нельзя.

Если ни в теории, ни у аналогичных объектов не известны требуемые результаты, то вы, возможно, правы. (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.)

Кроме того, возможно, что поверхность, отклика имеет несколько экстремумов, а мы нашли один из них, ближайший к исходной точке.

Поэтому вернитесь к пункту III и познакомьтесь с другими возможностями.

9. Вы предполагаете, что существует несколько экстремумов и мы просто нашли не самый лучший. Это вполне возможно.

Во всяком случае, 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.

Вернитесь к пункту III и познакомьтесь с другими возможностями.

10. Вы не знаете, какое объяснение выбрать.

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 по-

бочных реакций выход более 80% невозможен. Тогда приходится пересматривать представления о возможностях объекта. По условию с воспроизводимостью все в порядке. Значит, можно предполагать, что существуют другие экстремумы, т. е. нарушена предпосылка об уни-модальности.

Как уже отмечалось выше (см. гл. 4), 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. 9 и 11).

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?

12.2 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 аппа-

рат 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.

Встав на путь изучения методов 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 написали следующий, последний параграф этой главы.

12.3. Что читать?

Конкретные ссылки содержатся прямо в тексте. Здесь мы хотим указать некоторые общие работы. Этот перечень не претендует на полноту, его задача — служить ориентиром. Кроме того, ограничимся только работами, которые изданы на русском языке. Чтобы это ограничение не казалось существенным, укажем на обзор: Ю. П. Адлера, Ю. В. Грановский. «Обзор прикладных работ по планированию эксперимента». М., Изд-во МГУ, 1967. В этом обзоре содержатся ссылки на 800 работ, в которых планирование эксперимента использовалось в самых разнообразных областях исследования.

Обзор более ранних работ содержится в уже знакомой вам книге: В. В. Налимов, Н. А. Чернова. «Статистические методы планирования экстремальных экспериментов». В книге рассматривается большинство современных методов планирования. Следует, однако, отметить, что она рассчитана на читателя с повышенной математической подготовкой.

Мы упоминали и книгу Ю. П. Адлера «Введение в планирование эксперимента». М., изд-во «Металлургия», 1969, которая предназначена для первого знакомства с предметом.

Укажем еще книги: Ч. Х и К. С. «Основные принципы планирования эксперимента». Изд-во «Мир», 1967 и Д. Дж. Уайлд. «Методы поиска экстремума». М., изд-во «Наука», 1967. Совсем разные, они обе вводят читателя в мир планирования эксперимента.

Много конкретных работ содержится в сборниках материалов первого и второго всесоюзных совещаний по планированию эксперимента: «Планирование эксперимента». «Наука» 1966. «Проблемы планирования эксперимента», «Наука», 1969.

Регулярно публикует работы журнал «Заводская лаборатория» в разделе «Математические методы исследования».

Мы надеемся, что этих указаний достаточно, чтобы вы смогли двигаться дальше сами.

12.4. Последнее 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 подходящего плана — дробной реплики 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. Мы рассмотрели ряд типичных случаев.

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.

Но подождите отбрасывать книгу, ответьте, пожалуйста, на последний вопрос: как вы думаете, можно ли, прочтя эту книгу, практически планировать эксперимент?

Нет. Обратитесь к пункту 11.

Не знаю. Обратитесь к пункту 12.

Да. Обратитесь к пункту 13.

11. Очень жаль, что вы напрасно потратили время. Сдайте книгу в букинистический магазин и найдите какую-нибудь более полезную для вас.

12. Вы не знаете, можно ли использовать то, о чем вы прочли, на практике.

Мы тоже не знаем.

Давайте обратимся к практике и попробуем. Быть может, вам будет сопутствовать удача. Во всяком случае, мы желаем вам этого.

13. Мы благодарим вас за комплимент, однако не торопитесь с выводами. Обратитесь теперь к практике и, когда вам представится случай, воспользуйтесь тем, что вы узнали из этой книги.

Желаем вам успеха!

Эксперимент окончен, да здравствует эксперимент!

ОГЛАВЛЕНИЕ

Предисловие	5
Цель книги	7
Ограничения	9
Введение	11
Глава первая. Основные определения	16
Глава вторая. Параметр оптимизации	27
2.1. Виды параметров оптимизации (27). 2.2. Требования к параметру оптимизации (32). 2.3. Резюме (40)	
Глава третья. Факторы	41
3.1. Определение фактора (42). 3.2. Требования, предъявляемые к факторам при планировании эксперимента (48). 3.3. Требования к совокупности факторов (54). 3.4. Примеры факторов (55). 3.5. Резюме (59)	
Глава четвертая. Выбор модели	61
4.1. Шаговый принцип (66). 4.2. Как выбрать модель? (71). 4.3. Полиномиальные модели (74). 4.4. Резюме (78).	
Глава пятая. Полный факторный эксперимент	80
5.1. Принятие решения перед планированием эксперимента (80). 5.2. Полный факторный эксперимент (96). 5.3. Свойства полного факторного экспери-	

мента типа 2^k (103). 5.4. Полный факторный эксперимент и математическая модель (106). 5.5. Резюме (113)

Глава шестая. Дробный факторный эксперимент	116
6.1. Минимизация числа опытов (116). 6.2. Дробная реплика (119). 6.3. Выбор полуреplik. Генерирующие соотношения и определяющие контрасты (121). 6.4. Выбор $1/4$ -реплика. Обобщающий определяющий контраст (129). 6.5. Реплики большой дробности (136). 6.6. Резюме (141)	
Глава седьмая. Проведение эксперимента	145
7.1. Реализация плана эксперимента (145). 7.2. Ошибки параллельных опытов (155). 7.3. Дисперсия параметра оптимизации (161). 7.4. Проверка однородности дисперсий (165). 7.5. Рандомизация (172). 7.6. Разбиение матрицы типа 2^k на блоки (175). 7.7. Резюме (181)	
Глава восьмая. Обработка результатов эксперимента	182
8.1. Метод наименьших квадратов (183). 8.2. Регрессионный анализ (197). 8.3. Проверка адекватности модели (199). 8.4. Проверка значимости коэффициентов (207). 8.5. Резюме (211)	
Глава девятая. Принятие решений после построения модели	212
9.1. Интерпретация результатов (212). 9.2. Принятие решений после построения модели процесса (218). 9.3. Построение интерполяционной формулы, линейная модель неадекватна (230). 9.4. Резюме (233)	
Глава десятая. Крутое восхождение по поверхности отклика	235
10.1. Движение по градиенту (235). 10.2. Расчет крутого восхождения (238). 10.3. Реализация мысленных опытов (246). 10.4. Резюме (252)	

Глава одиннадцатая. Принятие решений после крутого восхождения	253
11.1. Крутое восхождение эффективно (253). 11.2. Крутое восхождение неэффективно (260). 11.3. Резюме (268)	
Глава двенадцатая. Обсуждение результатов	270
12.1. Чем кончается эксперимент? (270). 12.2. Перспективы (277). 12.3. Что читать? (278). 12.4. Последнее резюме (279)	