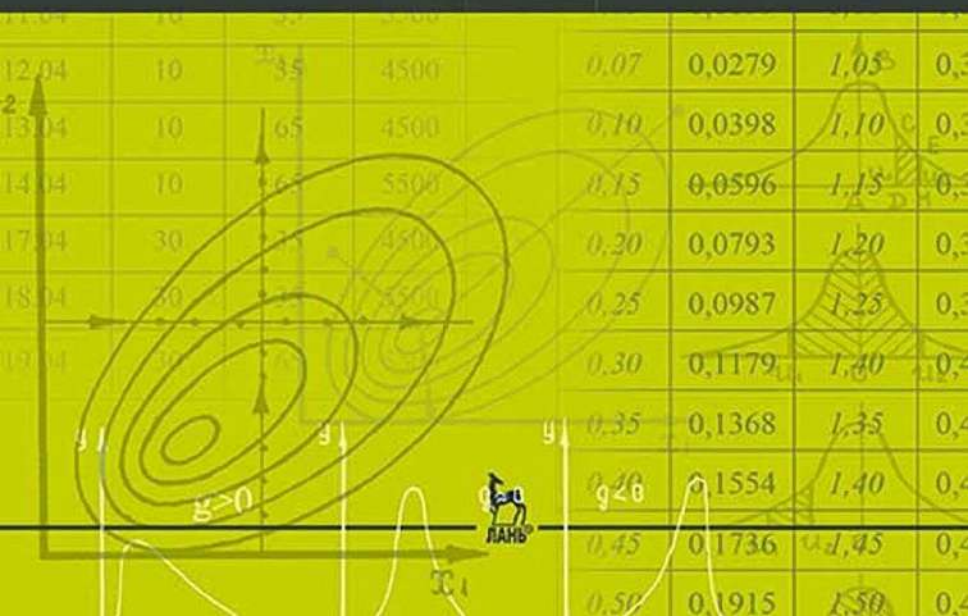


$$\Phi(U) = \frac{1}{\sqrt{2\pi}} \int_0^U e^{-\frac{t^2}{2}} dt$$

$\Phi(U)$	U	$\Phi(U)$	U	$\Phi(U)$	U	$\Phi(U)$	U
0,0040	0,90	0,3159	1,90	0,4778	1,90	0,4778	1,90
0,0120	0,95	0,3289	1,95	0,4744	1,90	0,4744	1,90
0,0199	1,00	0,3413	2,00	0,3159	1,90	0,4778	1,90
0,0279	1,05	0,3531	2,10	0,3289	1,95	0,4744	1,90
0,0398	1,10	0,3643	2,20	0,3413	2,00	0,4821	2,00
	1,15		2,30	0,3531	2,10	0,4861	2,10
				0,3643	2,20	0,4895	2,20
					2,30	0,4927	2,30

В. И. Вершинин
Н. В. Перцев

ПЛАНИРОВАНИЕ И МАТЕМАТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ХИМИЧЕСКОГО ЭКСПЕРИМЕНТА





• САНКТ-ПЕТЕРБУРГ •
• МОСКВА •
• КРАСНОДАР •
2019

В. И. ВЕРШИНИН, Н. В. ПЕРЦЕВ

ПЛАНИРОВАНИЕ И МАТЕМАТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ХИМИЧЕСКОГО ЭКСПЕРИМЕНТА

*Издание четвертое,
стереотипное*

ДОПУЩЕНО

*УМО по классическому университетскому образованию
в качестве учебного пособия для студентов,
обучающихся по специальности «Химия»
и по направлению «Химия»*



• САНКТ-ПЕТЕРБУРГ •
• МОСКВА • КРАСНОДАР •
2019

УДК 542
ББК 24я73
В 37

Вершинин В. И., Перцев Н. В.

В 37 Планирование и математическая обработка результатов химического эксперимента: Учебное пособие. — 4-е изд., стер. — СПб.: Издательство «Лань», 2019. — 236 с.: ил. — (Учебники для вузов. Специальная литература).

ISBN 978-5-8114-4120-4

Кратко изложены теоретические основы и практические рекомендации по планированию и проведению многофакторных экспериментов в химии и химической технологии. Рассмотрены методологические и метрологические аспекты научных исследований, способы обработки результатов эксперимента и решения оптимизационных задач, методы проверки статистических гипотез, построение и интерпретация математических моделей. Приведенные примеры относятся к аналитической, технической, физической и органической химии. Включены справочные материалы (статистические таблицы), рекомендации по оформлению результатов исследования и примеры контрольных заданий.

Для студентов (магистрантов) высших учебных заведений, обучающихся по направлениям «Химия» и «Химическая технология», аспирантов и преподавателей. Пособие может быть успешно использовано и в рамках бакалавриата, особенно при выполнении студентами естественнонаучных направлений выпускных квалификационных работ.

УДК 542
ББК 24я73

Рецензенты:

А. Н. СМАГУНОВА — доктор технических наук, профессор кафедры физических методов анализа Иркутского государственного университета, заслуженный деятель науки РФ;

Т. Н. ШЕХОВЦОВА — доктор химических наук, профессор, зам. зав. кафедрой аналитической химии Московского государственного университета им. М. В. Ломоносова;

А. В. ГАРМАШ — кандидат химических наук, доцент кафедры аналитической химии Московского государственного университета им. М. В. Ломоносова.

Обложка
Е. А. ВЛАСОВА

© Издательство «Лань», 2019

© В. И. Вершинин,
Н. В. Перцев, 2019

© Издательство «Лань»,
художественное оформление, 2019

ПРЕДИСЛОВИЕ

Развитие науки привело к систематическому использованию химиками математических (преимущественно статистических) методов для планирования исследований и для интерпретации количественных данных, получаемых в результате соответствующих экспериментов. Все химики понимают необходимость статистических методов, но степень их применения в разных областях химической науки неодинакова; наиболее широко такие методы используют аналитики, физико-химики и химики-технологи.

Основы математической статистики и теории вероятностей кратко излагаются студентам-химикам в курсе высшей математики (в рамках бакалавриата). При этом не рассматриваются методологические и метрологические аспекты проведения научных исследований, способы обработки экспериментальных данных и другие, важные для химика практические приложения математической статистики. Преподавателям-математикам затруднительно давать развернутые примеры и рекомендации ввиду слабого знания ими специфики химических исследований, а также потому, что курс высшей математики предлагается студентам 1–2 курсов до изучения основных химических дисциплин. Гораздо успешнее и глубже могут усвоить этот материал магистранты и аспиранты, выполняющие исследования на выпускающих кафедрах и непосредственно заинтересованные в теоретически обоснованном планировании экспериментов, грамотной обработке и правильном объяснении полученных данных.

Для студентов-химиков в некоторых университетах читаются узкоспециализированные курсы типа «Математическая обработка результатов анализа», «Методы оптимизации технологических процессов», «Факторное планирование» и т. п. При изучении таких дисциплин, как правило, недостаточно освещается общая методология эксперимента, которую должен понять и усвоить начинающий исследователь. К тому же рекомендации лекторов-химиков часто носят слишком прикладной, рецептурный характер. В стороне остаются не только теоретические обоснования, но и такие важные аспекты, как выбор факторов и параметров, выбор средств измерений, границы применимости отдельных алгоритмов, способы проверки гипотез, создания и использования математических моделей, оформление результатов эксперимента и др.

Знакомство с работой многих вузов и собственный опыт привели авторов этой книги к убеждению, что нельзя отделить учебный материал, связанный с планированием эксперимента, от статистической обработки результатов; практические же рекомендации следует давать в тесной связи с соответствующими теоретическими обоснованиями, без разрыва во времени. Естественно, при таком методическом подходе приходится частично жертвовать и строгостью математических выкладок, и рецептурными частностями в примерах. Тридцатилетний опыт преподавания курса «Планирование и обработка результатов химического эксперимента» в Омском госуниверситете показал целесообразность такого компромисса. Общая трудоемкость этой учебной дисциплины в ОмГУ составляет три зачетные единицы, включая лекции (24 ч), занятия в компьютерном классе, а также консультации по выполнению индивидуальных заданий. Содержание курса не зависит от профиля обучения студента в бакалавриате или магистратуре, профиль и уровень обучения учитываются лишь при отборе задач, входящих в индивидуальные расчетные задания.

Предыдущее (второе) издание настоящего пособия вышло в 2005 г. в Омске под редакцией одного из ведущих российских специалистов в области теории вероятностей д. ф.-м. н. В. А. Топчия. Были учтены замечания докторов наук Ю. М. Дедкова, Н. М. Островского, А. Н. Смагуновой,

Т. Н. Шеховцовой и других специалистов. В качестве приложения были представлены выдержки из статьи академика Ю. А. Золотова с рекомендациями по оформлению результатов химического эксперимента. Хотя эта статья появилась довольно давно, но рекомендации автора, уникальные по своей практической ценности и ясности изложения, полностью сохранили свою актуальность. Они полезны как студентам-химикам, пишущим свои выпускные квалификационные работы, так и опытным исследователям, готовящим очередные публикации.

При подготовке нового издания содержание книги дополнено и несколько изменено. Учтены требования образовательных стандартов по формированию компетенций. Магистрант-химик, осваивающий дисциплину «Планирование и обработка результатов химического эксперимента», должен формировать следующие профессиональные компетенции: способность вести исследования по сформулированной тематике, самостоятельно составлять план исследования и получать новые научные и прикладные результаты (ПК-1); владение теорией и навыками практической работы в избранной области химии (ПК-2); способность участвовать в научных дискуссиях и представлять полученные результаты в виде отчетов и научных публикаций (ПК-4). Учтены и другие компетенции, указанные в стандартах аспирантуры и бакалавриата по направлениям «Химия» и «Химическая технология».

В новое издание пособия включена глава «Подготовка к проведению эксперимента», где особое внимание уделено метрологическим аспектам активного эксперимента. Это потребовало заново отредактировать текст первых глав. Несколько сокращена глава 10. В «Заключении» даны краткие сведения о хемометрических алгоритмах обработки больших массивов многомерных данных. Расширен и обновлен список литературных источников. Добавлены примеры контрольных заданий (приложение 1). Приведенные в пособии контрольные задания могут быть выполнены с использованием целого ряда доступных компьютерных программ, в частности пакетов Microsoft Excel и Statistica. К сожалению, многочисленность и разнообразие подобных программ, частое обновление и существенные различия их версий, а также высокая стоимость

многих программ не позволяют дать однозначные рекомендации, а тем более детально сопоставить возможности разных программ. Выбор программ зависит и от возможностей того или иного вуза, и от предпочтений преподавателя или самого пользователя. Поэтому при переиздании пособия мы исключили характеристики компьютерных программ, предназначенных для обработки экспериментальных данных, и сознательно не указали программы, рекомендуемые для выполнения контрольных заданий.

Рассматривая алгоритмы проверки статистических гипотез и обработки многомерных данных, мы использовали результаты собственных исследований, в частности, полученные при выполнении научного проекта 16-43-550479, поддержанного грантом РФФИ. Как и в некоторых других отечественных и зарубежных учебниках, появившихся в последние годы и хорошо принятых студентами, материал излагается довольно подробно, в стиле, близком к разговорному. После знакомства с нашим пособием начинающему исследователю полезно повторно проработать материал как по учебникам более высокого теоретического уровня (например, по книге Н. И. Сидняева «Теория планирования эксперимента и анализ статистических данных»), так и по руководствам, конкретизированным для разных областей химических исследований. Так, будущим химикам-аналитикам мы рекомендуем классические руководства В. В. Налимова и К. Дерффеля, а также книги А. Н. Смагуновой, В. И. Дворкина, А. Л. Померанцева, А. В. Гармаша и других российских авторов.

Мы надеемся, что новое издание нашего учебного пособия будет полезно не только студентам и аспирантам, но и другим исследователям-химикам, независимо от их опыта и научной специальности. Замечания и предложения можно прислать авторам по электронной почте (vy-vershinin@yandex.ru).

*Заведующий кафедрой аналитической химии
Омского государственного университета им. Ф. М. Достоевского,
заслуженный работник высшей школы РФ, д. х. н. В. И. Вершинин*

*Главный научный сотрудник Омского филиала
Института математики им. С. Л. Соболева СО РАН,
д. ф.-м. н. Н. В. Перцев
Июнь 2016 г.*

Глава 1

МЕТОДОЛОГИЯ ЭКСПЕРИМЕНТА

1.1.

ЦЕЛИ НАУЧНЫХ ИССЛЕДОВАНИЙ

Прежде чем начать какое-либо исследование, химику стоит задуматься о том, какова его цель. Выделяют два вида научных исследований: фундаментальные и прикладные. Цель *фундаментальных* естественнонаучных исследований — познание природы, выявление ранее не известных, но объективно существующих закономерностей. Человечество всегда стремилось к осознанию окружающего мира, это стремление было и остается самоцелью, независимой от сиюминутных материальных потребностей. Непосредственной пользы обществу или самому ученому, фундаментальные исследования могут и не дать, однако результаты фундаментальных исследований важны не только для удовлетворения здорового любопытства ученого. Это его вклад в постепенно накапливающиеся знания человечества об окружающем мире; те обобщенные знания, на которых основывается любая деятельность.

Конечно, фундаментальные исследования могут попутно привести к практически полезным результатам, но это не обязательно. Исследования электричества привели к изобретению электромотора и электрической лампочки, а исследования гравитации ни к чему «полезному» пока

что не привели. Нельзя заранее предвидеть, будет ли вообще польза для человечества (государства, института, фирмы и т. п.) от данного фундаментального исследования, когда и как эта польза выявится. Исследования свойств урана в 1920–1930-е гг. имели чисто теоретический характер. Практического применения этот металл не имел, и исследователей, занимавшихся в советских научных институтах изучением свойств урана, упрекали в бесполезном расходовании народных средств. Тогда никто не мог предположить, какую практическую важность приобретет «урановый проект» в 1940–1950-е гг., насколько важными для СССР окажутся ранее полученные знания о химических и физических свойствах урана...

Фундаментальными исследованиями во всем мире занята лишь небольшая часть ученых. Большинство ведет *прикладные исследования*, направленные на достижение немедленной пользы (экономического эффекта). В ходе таких исследований не столько изучают природу объекта, сколько ведут *оптимизацию*, т. е. отыскивают условия, в которых объект в максимальной степени проявляет желаемые свойства, важные в практическом отношении.

Независимо от цели своих исследований, химики создают и используют модели, т. е. формализованные описания *объекта* (химической реакции, технологического процесса, способа анализа или синтеза какого-либо вещества). Весьма важно, хотя и не всегда удается довести точность описания объекта до *математической модели* — набора уравнений, по которому можно прогнозировать свойства объекта, его поведение в тех или иных условиях.

Способы проведения исследований существенно зависят от того, какое исследование проводится: фундаментальное или прикладное. В фундаментальных исследованиях математическую модель обычно создают на основе *научной гипотезы*, например, гипотезы о механизме изучаемого процесса. Построение научных гипотез будет рассмотрено в разделе 1.2. От гипотезы к модели переходят, применяя общие законы соответствующей науки и общепринятые расчетные алгоритмы [1]. Так, кинетические модели химических процессов основываются на предположениях исследователя о том, какие частицы участвуют в лимитирующей стадии реакции и как именно они взаи-

модействуют. При построении таких моделей используют фундаментальные законы химической кинетики, в частности уравнение Аррениуса. Другим примером может быть моделирование равновесного состава реакционных смесей. Его ведут на основе гипотезы о процессах, которые происходят в данной смеси, а для расчета используют условие материального баланса и закон действующих масс. Таким образом, математическая модель создается до проведения эксперимента, на основании априорных знаний и предположений исследователя. Эти модели называют *содержательными* или *физико-химическими* [2]. Главная цель последующих экспериментов — подтвердить или опровергнуть правильность модели, а тем самым и правильность исходной гипотезы.

В прикладных исследованиях возможен и даже целесообразен иной подход к моделированию, чем в фундаментальных. Модель получают без привлечения теоретических представлений о строении вещества и механизме реакции, нередко вообще без выдвижения гипотез. Модель выводят только из совокупности экспериментальных данных, используя статистические методы. Не следует переоценивать научную значимость таких моделей (их называют *эмпирическими*, *феноменологическими*, *имитационными*, а чаще всего *статистическими*), но они, безусловно, весьма полезны. С их помощью можно прогнозировать свойства объекта в новых, еще не изученных условиях. Статистическая модель позволяет также выйти на оптимальные условия проведения процесса [3], [4]. Например, с помощью такой модели можно будет предвидеть, в каких условиях будет достигнут 100% -ный выход целевого продукта. Оговоримся, что статистическое моделирование — не единственный способ оптимизации, хотя этот способ используется чаще всего. Такие модели полезны и при проведении фундаментальных исследований как лаконичный способ записи экспериментальных данных, как исходный материал для последующих теоретически обоснованных обобщений и рекомендаций.

Четко отделить фундаментальные исследования от прикладных, а проверку научных гипотез от решения прикладных (в том числе оптимизационных) задач удается далеко не всегда. Многие исследования направлены од-

новременно на достижение обеих целей^{*}. Однако большинство научных публикаций связано либо с построением теоретически обоснованных моделей, либо с оптимизацией каких-либо процессов на основе статистических моделей. Примеры приведены в таблице 1. Оба подхода к созданию моделей — *детерминированный* и *статистический* — дополняют друг друга. Исследователь выбирает один из них в зависимости от изученности объекта и от цели своей деятельности.

Т а б л и ц а 1

Примеры исследований в разных областях химии и их цели

Область науки	Фундаментальные исследования, содержательные модели	Прикладные исследования, статистические модели
Органическая химия, физическая химия	Изучение кинетики и равновесия некоторой реакции, выяснение механизма этой реакции	Подбор температуры и растворителя для достижения максимального выхода целевого продукта
Аналитическая химия	Прогнозирование характеристик хроматографического удерживания веществ с учетом состава фаз	Нахождение условий, обеспечивающих максимальное разрешение пиков на хроматограмме
Электрохимия	Выявление факторов, влияющих на скорость коррозии, изучение механизма коррозии	Поиск наиболее эффективных ингибиторов коррозии для некоторого типа сталей
Химическая технология	Изучение процесса каталитического риформинга	Подбор оптимального режима работы некоторой установки риформинга

1.2. НАУЧНЫЕ ГИПОТЕЗЫ

Одна из важнейших целей науки — установление механизмов изучаемых явлений, выявление соответствующих причинно-следственных связей. Нелегко найти ответ на вопрос «Как именно это происходит?», но еще труднее

^{*} Возможен и случай, когда единственная цель исследователя — получение диплома о высшем образовании или ученой степени. С настоящей наукой эта цель никак не соотносится, и мы ее рассматривать не будем.

объяснить, почему это происходит именно так, а не иначе. Ответ не должен напоминать известную шутку Мольера, пародирующую пустословие ученых трактатов: «Почему опиум вызывает сон? Потому что он обладает снотворным действием!» Чтобы получить ответы на подобные вопросы по существу, исследователь вначале выдвигает научную *гипотезу* — более или менее логичное предположение, объясняющее взаимосвязь наблюдаемых явлений, а затем проверяет ее. Гипотеза всегда строится на основании априорной (заранее известной исследователю) информации, т. е. на основании знаний, полученных путем анализа научной литературы, в ходе собственных наблюдений или в результате проведения предварительных опытов. При построении гипотез часто используют аналогии с другими, лучше изученными объектами, переносятся какие-то представления из смежных областей науки. Важны и логичность мышления, и интуиция.

Выдвигая научные гипотезы, молодые исследователи должны учитывать некоторые общие рекомендации.

1. Опирайтесь только на достоверные и хорошо воспроизводимые факты. Сомнительные данные, неизвестно кем, как и когда полученные, рассматривать не надо, так же как и *мнения*, не подтвержденные фактами.

2. Выдвигаемая гипотеза должна согласовываться со всеми известными свойствами исследуемой системы и не должна противоречить ни одному из наблюдаемых фактов. Даже один достоверный факт, противоречащий гипотезе, опровергает ее.

3. К результатам исследований других авторов, к их предположениям и высказываниям следует относиться критически, независимо от научных заслуг и степени известности этих авторов. Не случайно история современной химии началась с книги Р. Бойля, поучительно названной «Химик-скептик»!

4. Критическое отношение к литературным данным вовсе не означает их ненужности. Наоборот, свою гипотезу можно выдвигать только после детального ознакомления с работами всех предшественников. С их взглядами можно не соглашаться, но знать ранее полученные факты и выдвинутые на их основе гипотезы объективный исследователь обязан.

5. Не следует спешно выдвигать научные гипотезы для объяснения единичных и недостаточно изученных явлений — возможность ошибок в этом случае очень велика. Следует исключить *артефакты* — так называют явления, которые только кажутся относящимися к изучаемой системе, но обусловлены посторонними явлениями — неисправностью приборов, примесями в реактивах, ошибками в подсчетах и т. п. В истории науки было немало логичных и остроумных гипотез, придуманных для объяснения артефактов. Астроном мог бы, наверное, придумать не противоречащую известным теориям, логичную и остроумную гипотезу для объяснения природы темных продолговатых пятен, быстро перемещающихся по видимой поверхности Солнца при наблюдении его в телескоп. Но если это явление наблюдалось впервые и лишь однажды, стоит задуматься — не является ли оно артефактом? А вдруг никогда и никем не виданные движущиеся пятна были обычными мухами, ползавшими по объективу телескопа?

Многие исследователи сталкивались не только с артефактами, но и с фальсификацией экспериментальных данных. Иногда крупные ученые становились жертвами жульничества своих сотрудников, иногда сознательно или бессознательно выдавали желаемое за действительное. В такие ловушки часто попадали опытные и субъективно честные ученые, уходя из области своей бесспорной компетенции в другие области науки или в такие «опасные» и по своей сути вненаучные области, как парапсихология, астрология, глобальный пересмотр хронологии исторических событий и т. п.

6. Исследователь должен удержаться от соблазна немедленно объяснить любое непонятное ему явление действием неизвестных сил, проявлением новых законов природы или следствием существования неизвестных элементарных частиц, а также деятельностью космических пришельцев или нечистой силы. Не торопитесь объявлять, что в ходе данного исследования вами опровергнут такой-то физический закон или «устаревшая» химическая теория. Вначале следует сделать все возможное, чтобы объяснить наблюдаемые необычные явления «обычными» причинами, сформулировать свою гипотезу в рам-

ках общепринятых представлений о мире. «Сумасшедшие» гипотезы, противоречащие всем известным ранее теориям, заманчивы, но подтверждаются они очень редко. Очевидно, такие гипотезы имеют право на существование *лишь в том случае, если никакими известными причинами объяснить факты невозможно*. Данное правило в истории науки известно под названием «бритвы Оккама» — по фамилии средневекового ученого, провозгласившего принцип «Не умножайте сущностей без необходимости», отрезая тем самым фантастические объяснения явлений, легко объяснимых известными причинами.

7. Внешне логичная и не противоречащая известным законам природы гипотеза вовсе не обязательно верна. На основании одних и тех же данных можно выдвинуть множество правдоподобных гипотез, в том числе противоречащих друг другу. Так, в истории медицины выдвигалось множество гипотез о причинах рака или о механизме кариеса зубов. В химии долгое время соперничали сторонники «физической» и «химической» точек зрения на растворы, и для объяснения каждого явления, связанного с растворами, ими обязательно выдвигались противоположные гипотезы. Чтобы победить в соревновании с конкурирующими гипотезами, стать научной теорией, любая гипотеза нуждается в длительной и многократной проверке. В ходе проверки желательно работать с другими модельными объектами, применять другую технику эксперимента, другие способы расчетов — не те, что первоначально использовались. Голословное одобрение или неодобрение вашей гипотезы кем бы то ни было не имеет существенного значения, но очень важно подтверждение ваших данных в независимых экспериментах (т. е. проведенных не вами и не вашими друзьями или помощниками).

Следует учесть, что хорошая гипотеза не только объясняет уже известные факты, но и позволяет прогнозировать неизвестные явления. Подтверждение прогнозов — очень веский аргумент в пользу научной гипотезы. Так, гипотеза Д. И. Менделеева о периодической зависимости свойств элементов от их атомной массы стала *Периодическим законом* после открытия Винклером, Нильсоном и де Буабодраном новых элементов, предсказанных Менделеевым. Окончательно признана гипотеза Менделеева

была лишь через 50 лет после ее появления, в результате опытов Г. Мозли и теоретических выкладок Н. Бора.

Положительные результаты проверок подтверждают гипотезу, делают ее более вероятной, превращают в общепринятую теорию, но никакая проверка не может окончательно доказать правильность гипотезы. Ведь в будущем всегда могут обнаружиться новые факты, не согласующиеся с ранее выдвинутой гипотезой, и тогда придется строить новую гипотезу, объясняющую и эти факты; новую, более общую и точную теорию. При проверке научных гипотез применяют различные экспериментальные методы, тщательно обрабатывая полученные данные статистическими методами. Как писал один из крупнейших химиков-аналитиков XX в. И. Кольтгоф: «Теория указывает путь решения задачи, а решает эксперимент» [5]. Разумеется, проверка гипотез не сводится к статистической обработке экспериментальных данных, но без такой обработки при проверке научных гипотез химику не обойтись (см. раздел 7.1).

1.3. ПРЕДПОСЫЛКИ И СПОСОБЫ ПРОВЕДЕНИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Любое научное исследование основано на некоторых общих положениях мировоззренческого характера [2]. Начиная свое исследование, химик редко задумывается над философскими вопросами, но он не сомневается, что объекты его будущего исследования материальны, познаваемы и не зависят от каких-либо субъективных представлений о них. Любой исследователь уверен также, что объективным критерием истинности этих представлений является практика, в частности, результаты научного эксперимента.

На основе этих предпосылок любой ученый как-то организует свои действия, реализует определенный *метод исследований*. Термин «метод» происходит от греческого слова, некогда обозначавшего охоту за зверем или преследование преступника. Начиная с Платона и других древнегреческих философов, слово «метод» стало использо-

ваться для обозначения способа постижения истины. Научным методом стали называть совокупность приемов, благодаря которым ученый получает новые, ранее никому не известные знания. Основными приемами научного исследования для античных ученых стали наблюдения, логические рассуждения, а также ссылки на авторитет знаменитых предшественников, в частности, Аристотеля. Этот, довольно неэффективный метод в эпоху Возрождения сменился другим («галилеевским»), основанным на активной деятельности исследователя, который воспроизводит и моделирует изучаемые явления. Конечно, наблюдения и рассуждения сохранили свою роль и в рамках нового метода, но они уже не имеют такого исключительного значения, как в древности. В химию этот метод пришел чуть позднее, чем в физику (благодаря работам А. Л. Лавуазье и Р. Бойля), обеспечивая с тех пор плодотворное развитие нашей науки.

Начинающий исследователь-химик редко задумывается над выбором метода, он использует в своей деятельности те приемы, которые общеприняты в соответствующей области науки. Такие приемы кажутся новичку единственно возможными, поскольку так работали его учителя и так сегодня работают его многочисленные коллеги. Однако крупные ученые (Оствальд, Менделеев, Эйнштейн, Пуанкаре, Вернадский, Полинг и др.), независимо от области их научных интересов, глубоко задумывались над проблемами методологии, предлагали совершенно новые приемы и способы научной работы. Это помогало добиться успеха и в прикладных, и в фундаментальных исследованиях.

Методологию (дословно — учение о методе) в настоящее время определяют как *науку о путях приращения и упорядочивания знаний* [2]. Конкретизируя и сужая проблему, можно сказать, что методология — это *наука о способах и средствах рационализации научной деятельности*. Общая методология науки является по своей сути частью философии; она рассматривает проблему объективной научной истины, ее границы и критерии, индуктивный (от частного к общему) и дедуктивный (от общего к частному) методы познания, взаимосвязь анализа и синтеза, принципы систематизации научных знаний и т. п.

Кроме того, для каждой области научных знаний выработана своя собственная методология. В частности, выделяют методологию химии, являющуюся частью химической науки. Методология химии рассматривает общехимические понятия (например, понятие химического элемента), а также общехимические методы исследований — сравнительный, термодинамический, кинетический и квантово-химический. Можно выделить еще более узкие области — частные методологии аналитической химии [5] или органической химии [6]. В каждом случае речь идет, прежде всего, о целях, направлениях и способах организации исследований в соответствующей области науки.

Понятно, что научными исследованиями можно заниматься по-разному. В гуманитарных науках важными приемами являются: наблюдение, выборочное анкетирование, логическое сопоставление высказываний, поиск материалов в архивах и библиотеках, сравнительный (в том числе математический) анализ текстов и документов, археологические раскопки и т. п. Естествознание также не отказывается от наблюдений или изысканий в архивах, но основным способом исследований для естественных наук в целом и для химии в частности является *эксперимент*. Так называют *целенаправленное изучение некоторого процесса (явления) в точно известных и определенным образом меняющихся условиях*. Есть и другое определение: «эксперимент — это система операций, воздействий и/или наблюдений, направленных на получение информации об объекте при исследовательских испытаниях» [7]. Экспериментатор регистрирует характеристики и результаты процесса — проводит измерения, отбирает пробы, фотографирует, ведет записи в журнале, обрабатывает полученные данные, делает выводы и затем публикует результаты своих экспериментов.

Выделяют два вида эксперимента: активный и пассивный. В *пассивном* эксперименте исследователь лишь наблюдает за ходом процесса, регистрирует результаты и условия его протекания, но не инициирует этот процесс, не задает его условий и не вмешивается в ход процесса. Можно, например, в течение длительного времени измерять и записывать показатели качества бензина, выпускаемого некоторой заводской установкой. Одновременно

определяют состав сырья и/или измеряют температуру в реакционной зоне. В разные дни эти показатели несколько различаются. Затем можно обработать полученные данные и найти математическую зависимость показателей качества бензина от показателей состава сырья или от температуры.

Основную роль в химических исследованиях играет *активный* эксперимент, в этом случае исследователь целенаправленно воспроизводит исследуемый процесс, задает и поддерживает его условия. В активном эксперименте исследователь не просто наблюдает за ходом процесса и измеряет какие-то характеристики, но и многократно повторяет этот процесс в заранее заданных и тщательно поддерживаемых условиях [8]. Более того, экспериментатор меняет эти условия по заранее продуманному плану, проводит новые серии опытов, обрабатывает и анализирует полученные данные.

Для активных химических экспериментов характерно следующее:

а) исследователь-химик в состоянии *управлять* тем процессом, который он изучает, и может *воспроизводить* его в лабораторных условиях;

б) в ходе любого химического эксперимента измеряются некоторые *физические величины*, характеризующие условия и результаты отдельных опытов;

в) эксперимент в целом или отдельные опыты можно повторять произвольное число раз, причем результаты повторных опытов воспроизводятся (с учетом точности измерений). Измерение любой величины (температуры, концентрации, скорости и др.) производится с погрешностью, не превышающей некоторую заранее заданную (предельную) величину, которая в ходе эксперимента не меняется (см. раздел 3.2). Предполагается, что проведение измерений не влияет на изучаемый процесс; это допущение очень важно, но не всегда верно.

Нередко активный эксперимент нельзя провести в реальных условиях. Так, чтобы установить зависимость качества бензина от состава сырья, никто не станет целенаправленно подавать на многотоннажную заводскую установку сырье специально подобранного состава (сегодня одно сырье, завтра другое, и т. д.). Подобные эксперименты

были бы слишком дороги и технически неосуществимы. Надо либо ограничиться пассивным экспериментом, либо проводить активный эксперимент на небольшой лабораторной или полупромышленной установке. Первый путь проще и дешевле, второй путь является более экспрессным и дает больше полезной информации.

Обычно экспериментатор идет по второму пути, реализуя изучаемый процесс в *лабораторных условиях*. При этом он учитывает не все, а лишь некоторые факторы из числа влияющих на процесс в реальных условиях. Воспроизводятся лишь наиболее важные (с точки зрения экспериментатора) характеристики процесса. Отличия лабораторного эксперимента от соответствующих природных или технических процессов весьма существенны. А именно:

- лабораторный эксперимент обычно воспроизводит изучаемый процесс *в другом масштабе*. Например, химик проводит реакцию в пробирке или в небольшой колбе, а в реальных условиях та же реакция идет (или должна будет идти) в реакторе, вмещающем сотни тонн реагентов;
- лабораторный эксперимент воспроизводит изучаемый процесс заведомо *иным способом, на сходных, но не на тех же самых объектах*. Так, методику анализа сточных вод данного типа вначале проверяют на более простом объекте — приготовленном в лаборатории многокомпонентном растворе, содержащем те же токсиканты, что и сточная вода (такие растворы называют имитатами). А опасные заболевания людей и способы их лечения вначале моделируют на морских свинках, собаках и других лабораторных животных;
- лабораторный эксперимент иногда по техническим или экономическим соображениям проводят *в других (легче реализуемых) температурных или концентрационных условиях, пренебрегают влиянием некоторых факторов, ограничивают продолжительность эксперимента*. Чем больше отличий допускает экспериментатор, тем проще ему будет провести работу, но практическая ценность получаемых результатов снизится. Поэтому условия проведения

лабораторного эксперимента постепенно приближают к реальным.

Далее мы будем рассматривать проведение активных экспериментов только в лабораторных условиях, что соответствует практике химических исследований. В этом отношении химики находятся в более выгодном отношении, чем астрономы, вулканологи или политологи. Активные лабораторные эксперименты в этих областях науки пока что невозможны.

1.4. ОБЪЕКТ ИССЛЕДОВАНИЯ КАК «ЧЕРНЫЙ ЯЩИК». ФАКТОРЫ И ПАРАМЕТРЫ

Для изучения и моделирования любого объекта удобно воспользоваться кибернетическим представлением его в виде «*черного ящика*». В этом случае исследователь не рассматривает процессы, протекающие внутри объекта; он отмечает лишь исходные условия и результаты опытов. Этот подход и соответствующие ему способы планирования активного эксперимента особенно важны при недостаточно изученном механизме исследуемых процессов. Образно говоря, при неизвестном содержании «черного ящика».

На рисунке 1 буквами X_1, X_2, \dots, X_k обозначены те величины, которые мы в ходе эксперимента задаем и контролируем. Независимые величины, влияющие на поведение исследуемого объекта, называются *факторами*. Примерами факторов применительно к химическим исследованиям могут быть начальные концентрации реагентов, продолжительность термообработки, температура и давление реакционной смеси, природа катализатора и растворителя, интенсивность облучения и т. п. В каждом

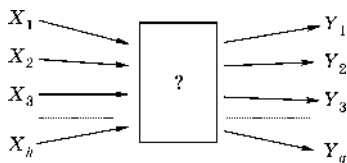


Рис. 1
Объект исследования как «черный ящик»

исследовании число таких факторов ограничено, обозначим это число символом k .

Зависящие от факторов свойства объекта, которые измеряют и оптимизируют в ходе эксперимента, называются *параметрами* (или *параметрами оптимизации*). На рисунке 1 параметры обозначены как Y_1, Y_2, \dots, Y_q . Параметры характеризуют результаты опытов и непосредственно связаны с целью исследования. В каждом исследовании число параметров ограничено, обозначим его через q . Термин *параметр* в науке используют часто, и это слово имеет много разных смысловых значений. Так, математики говорят о выборочных параметрах или о параметрах распределения, вовсе не имея в виду непосредственно измеряемые в ходе эксперимента физические величины. С другой стороны, технологи-практики иногда называют параметрами любые количественные характеристики изучаемого процесса и результаты любых измерений, т. е. и *иксы*, и *игреки*. Однако в теории планирования эксперимента разделение измеряемых характеристик объекта на факторы и параметры проводится весьма четко, путать эти термины недопустимо.

Резюмируем: объектом исследования мы называем систему, поведение которой определяется неизвестными нам заранее, но вполне определенными внутренними закономерностями. Выявить их, изучить «содержимое ящика» можно в активном эксперименте, воздействуя на объект разными способами и наблюдая, как меняются значения параметров в зависимости от значений факторов. Полученные данные позволяют рассчитать статистическую модель.

Например, при исследовании коррозии сталей в водных растворах серной кислоты мы можем менять от опыта к опыту тип стали или состав раствора, способ предварительной обработки поверхности образца, температуру и скорость перемешивания раствора, природу и концентрацию ингибитора, и т. п. Все это *факторы*. В каждом опыте будем измерять в качестве параметра скорость коррозии. Эта характеристика зависит от всех перечисленных факторов и непосредственно связана с целью эксперимента. Примером статистической модели может быть эмпирическая зависимость скорости коррозии некоторой стали от температуры, концентрации и скорости перемешивания

раствора кислоты. Эта зависимость укажет исследователю лимитирующую стадию процесса, что важно для выяснения механизма коррозии. Если же наша цель — замедление коррозии, то та же статистическая модель позволит найти условия, при которых скорость коррозии станет равной нулю, или хотя бы снизится до некоторого допустимого значения.

Будем считать математической моделью уравнение, связывающее параметр и факторы [4]. В общем виде оно записывается следующим образом: $Y = f(X_1, \dots, X_k)$. Значения параметра Y далее выделены курсивом (Y), аналогичным образом выделены значения факторов. Функцию $Y = f(X_1, \dots, X_k)$ часто называют *функцией отклика*. В случае многопараметрической задачи получают q таких уравнений, по одному на каждый параметр.

Математическую модель можно выразить не только алгебраически, но и геометрически. Так, при изучении зависимости параметра Y от фактора X ($k = 1$) каждый опыт дает пару значений (X, Y) , т. е. точку на плоскости. Зависимость Y от X изображается на плоскости в виде прямой или кривой. Зависимость Y от двух факторов ($k = 2$) можно передать плоскостью или криволинейной поверхностью в трехмерном пространстве. А для трех факторов наглядное изображение модели невозможно — пришлось бы изобразить некую поверхность в четырехмерном пространстве. Модель, учитывающая k факторов, описывается поверхностью в $(k + 1)$ -мерном пространстве. Математики без особого труда рассчитывают координаты точек, положение экстремумов и даже маршруты движения по такой *поверхности отклика*. Геометрические представления понадобятся, чтобы понять, как с помощью математической модели оптимизируют изучаемые процессы (глава 10).

Сложность исследования определяется количеством одновременно контролируемых факторов и параметров. Например, относительно простой считается работа, в которой автор изучал зависимость скорости коррозии от температуры раствора при прочих постоянных условиях, это однофакторный и однопараметрический эксперимент. Чаще изучают влияние нескольких факторов, по очереди или одновременно. Нередко для исследователя важен не

единственный параметр, а сразу несколько. Так, органик-синтетик меняет состав исходных реагентов и температуру реакции, при этом его интересует и выход целевого продукта, и его чистота, и себестоимость синтеза, и время, необходимое для проведения синтеза. Соответствующий эксперимент будет весьма сложным, причем измерение всех параметров потребует слишком больших затрат времени и средств. Сопоставление числа публикаций, посвященных моделированию и оптимизации процессов разной сложности (так называемый *наукометрический анализ*) показывает, что обычно химики изучают влияние всего 2–3 факторов, измеряя при этом единственный параметр.

1.5. ОСНОВНЫЕ ЭТАПЫ АКТИВНОГО ЭКСПЕРИМЕНТА

Любой химический эксперимент требует последовательного проведения ряда отдельных опытов. Словом «опыт» мы будем называть получение характеристик исследуемого объекта (параметров оптимизации) в каких-то определенных условиях. Опыты могут быть независимыми, в этом случае каждый опыт отличается от любого другого по условиям его проведения, т. е. в новом опыте хотя бы один фактор должен отличаться от своего значения в предыдущем опыте. Как говорят, этот фактор должен быть взят на новом уровне. Например, измерение скорости коррозии в опыте № 1 проводят при 25°C в 0,10 М растворе HCl и в присутствии бензтриазола, в опыте № 2 — в тех же условиях, но при 40°C, в опыте № 3 — при 25°C в 0,10 М растворе HCl, но в отсутствие бензтриазола. Чем сложнее объект, тем больше факторов влияет на исследуемые параметры, тем больше должно быть число независимых опытов.

Эксперимент ведут по предварительно разработанному плану, переходя от опыта к опыту в заранее определенном порядке. Каждый опыт обязательно повторяют, тщательно контролируя его условия. Если условия опытов по всем факторам остаются неизменными, то опыты называют повторными или дублирующими. Повторять опыты надо, даже если делать этого не хочется (долго, дорого, скучно

и т. п.), но иначе выводы исследователя будут ненадежными. Как правило, экспериментатор повторяет каждый независимый опыт m раз, величину m называют *кратностью*, или *повторностью*.

Любой фактор может принимать одно из своих возможных значений. Каждое из них называется *уровнем*. Число возможных уровней для большинства факторов теоретически бесконечно велико, на практике же оно ограничено: а) областью допустимых значений фактора; б) возможностями применяемой аппаратуры; в) естественными колебаниями измеряемой величины в ходе единичного опыта; г) точностью измерений. Предположим, что исследователь-химик собирается изучить влияние температуры на некоторый процесс, идущий в водном растворе при атмосферном давлении. Очевидно, такие растворы могут иметь температуру от 0 до 100°C. Предположим далее, что температуру раствора измеряют с погрешностью в один градус (такова цена деления данного термометра по всей температурной шкале). Следовательно, возможны 100 разных уровней температуры. Если температуру раствора собираются поддерживать с помощью термостата, который не может охлаждать раствор, область допустимых значений сузится, а число возможных уровней температуры уменьшится до 76°C (от 25 до 100°C). Кроме того, надо учесть, что термостат поддерживает температуру менее точно, чем ее измеряет термометр. Допустим, колебания температуры в ходе каждого опыта достигают 3°C. В таком случае число возможных уровней температуры снизится до 25°C, таким же будет и максимальное число независимых опытов. Конечно, реальное число опытов всегда меньше максимально возможного.

Для фактора «концентрация» максимальное число уровней также ограничено. Во-первых, любое вещество имеет некоторую максимально возможную растворимость S , и мы можем менять концентрацию только в пределах от 0 до S . Во-вторых, мы создаем и измеряем концентрацию с некоторой погрешностью ΔC . Максимальное число возможных уровней концентрации (p_{\max}) будет равно $S/\Delta C$. Чем точнее измерять концентрации, тем больше уровней можно использовать в ходе эксперимента, но

число концентрационных уровней в любом случае не будет бесконечно большим.

Таким образом, общий объем эксперимента (общее число опытов) определяется числом исследуемых факторов, числом уровней, отобранных для изучения влияния каждого фактора, и кратностью опытов [9]. Задав набор уровней по всем факторам, мы не только определяем условия данного опыта, но и описываем одно из возможных состояний нашего «черного ящика» — исследуемого объекта. Если каждый фактор может быть задан на p уровнях, а всего исследуется k таких факторов, то число возможных независимых опытов равно p^k . Так, трехфакторный эксперимент, в котором каждый фактор может иметь одно из четырех возможных значений, потребует $4^3 = 64$ независимых опытов. С учетом повторных измерений их число будет еще больше:

$$N = mp^k. \quad (1.1)$$

Формула (1.1) показывает, что с увеличением числа факторов и уровней объем эксперимента быстро растет. Так, для четырех факторов на 10 уровнях возможно $10^4 = 10\,000$ сочетаний и, следовательно, 10 000 независимых опытов. Для 10 факторов на четырех уровнях число возможных независимых опытов равно $4^{10} = 1\,048\,576$. Пришлось бы провести больше миллиона опытов, не считая повторных. Такой объем эксперимента просто нереален.

С учетом всего вышеизложенного можно выделить следующие стадии (этапы) активного химического эксперимента.

1. Формулировка цели и задач исследования, выбор параметров и факторов, уточнение допустимых значений факторов. На этой стадии исследователь обычно руководствуется ранее накопленной информацией об изучаемых объектах, проводит анализ соответствующей научной литературы.

2. Планирование эксперимента. В частности, на этой стадии намечают схему эксперимента и определяют его объем, учитывая число независимых опытов, их кратность и порядок их проведения.

3. Подготовка к проведению эксперимента. В частности, исследователь выбирает объекты (вещества, материалы, установки), подбирает аппаратуру, осваивает методики, проверяет воспроизводимость результатов при повторении опытов в неизменных условиях.

4. Непосредственное проведение опытов.

5. Обработка результатов эксперимента. В частности, расчет и проверка некоторой математической модели.

6. Интерпретация результатов эксперимента, оптимизация исследуемого процесса, принятие решения о дальнейших действиях. Иногда на этой стадии приходится возвращаться к пунктам 1, 2 или 3.

7. Формулировка выводов, оформление и публикация результатов.

Приведенная схема отвечает содержанию большей части химических и химико-технологических исследований, которые проводят по схеме «черного ящика». Однако не следует считать, что эта схема должна быть реализована в любом активном эксперименте. Так, в некоторых важных областях химической науки статистические методы и математические модели пока что применяют довольно редко. Примером может быть создание новых способов органического синтеза биологически активных веществ. Этапы и специфические приемы таких исследований в нашей книге не рассматриваются.

Глава 2

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

2.1.

ИСТОРИЯ И ПРИНЦИПЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Химический эксперимент не может включать все возможные состояния исследуемого объекта; перебирать все возможные сочетания факторов слишком долго и дорого. Эксперимент надо планировать, т. е. заранее выбирать, сколько и каких именно опытов надо провести, чтобы получить желаемую математическую модель (или оптимизировать некоторый процесс, или проверить некоторую научную гипотезу). Естественно, исследователь планирует свою работу так, чтобы добиться цели с минимальными затратами труда и времени, уменьшить количество опытов, но не за счет надежности или объема получаемой информации. Запишем следующее определение: «Планирование эксперимента — это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения намеченной задачи с требуемой точностью».

Для рационального планирования эксперимента нужна специальная теория. Хаотические, бесплановые исследования приводят лишь к потере времени; достичь цели без четкого плана работы можно лишь случайно, такие случаи в истории науки бывали, но крайне редко. Процитируем шуточный отрывок из учебника [4]: «Мысль о том, что эксперимент можно планировать, восходит к глубокой

древности. Наш далекий предок, убедившийся, 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. Однако он не пользовался статистическими методами, поэтому остается непонятным, как он вообще выжил и обеспечил тем самым наше существование».

Античной науке идеи активного эксперимента и статистических моделей были чужды. Современное понимание научного эксперимента начало формироваться сравнительно недавно — в эпоху Возрождения. Наиболее яркий пример — работа Г. Галилея, которой измерял время падения тел с разной высоты, поднимаясь на башни и сбрасывая оттуда специально сделанные шары. В одних опытах все факторы, кроме высоты башни, Галилей сознательно держал на постоянном уровне (бросал шары одинакового размера и веса с разной высоты). Во второй серии опытов он бросал шары одинакового размера, но разного веса с одной и той же высоты (с балкона знаменитой Пизанской падающей башни). В третьей серии бросал оттуда же шары одинакового веса, но разного размера. При этом каждый опыт многократно повторялся. По результатам этих опытов ученому удалось вывести формулу, позволявшую предсказывать время падения любого тела с любой высоты, т. е. была создана математическая модель процесса. Для своего времени это был совершенно новый подход, и современники не были от него в восторге. Затея Галилея вызывала пренебрежительные реплики и недоумение. «Неужели он всерьез думает, что, бросая с башни какие-то шары, можно опровергнуть Аристотеля, правоту суждений которого подтверждают два тысячелетия?» [10]

В химии количественные методы активного эксперимента стали применяться позднее, чем в физике. По-видимому, о целенаправленном проведении активного эксперимента по предварительно выработанному плану можно говорить, начиная с Р. Бойля. В массовом порядке химики стали использовать традиционную для естествознания галилеевскую схему поочередного варьирования факторов только в конце XIX в.

Естественно, описанная схема имеет свои недостатки. Главный из них — расходование «впустую» громадного числа опытов; при более рациональном планировании эксперимента они не понадобились бы. Вообще эффективность научных исследований невелика, их КПД, по оценкам компетентных специалистов, не превышает 2% [9]. Одной из причин является хаотичность проведения опытов, отсутствие четкого и заранее продуманного оптимального плана, что характерно для многих (не только начинающих!) исследователей.

Планирование научного эксперимента с применением статистических методов — это целая наука, которая стала складываться в первой половине XX в., когда научная работа из увлечения немногочисленных любителей стала профессиональным занятием множества узких специалистов. Первые работы в этой области были выполнены английским математиком Рональдом Фишером и его учениками около 1930 г. Основная идея Фишера заключалась в том, что влияние разных факторов на исследуемый параметр должно изучаться не по очереди, а одновременно. Причем условия каждого опыта надо выбирать с учетом необходимости статистической обработки результатов эксперимента методами регрессионного анализа. Это сокращает объем эксперимента и облегчает расчет математической модели.

В начале 1950-х гг. появились важные работы К. Уилсона и Дж. Бокса, в которых были предложены и испытаны разные планы многофакторных экспериментов. Был разработан способ, с помощью которого на основании математической модели можно было рассчитать оптимальные условия исследуемого процесса, так называемый «метод движения по градиенту». Постепенно сложилась теория многофакторного активного эксперимента, появились единая терминология и общепринятые расчетные алгоритмы [4], [7], [11]. Теория планирования эксперимента отделилась от своей прародительницы — математики, доказала свою полезность, в частности, при поиске оптимальных условий для крупномасштабных технологических процессов (эволюционное планирование). В нашей стране большой вклад в теорию планирования эксперимента внесли В. В. Налимов, Ю. В. Грановский, Ю. П. Ад-

лер; в США — Дж. Кифер. В 1970-е гг. планирование многофакторных экспериментов по Боксу — Уильсону вошло в моду, его стали применять и в тех случаях, когда можно было обойтись более простыми средствами. Позднее стало ясно, что рациональное планирование эксперимента — далеко не единственное условие достижения успеха в работе. Нездоровый ажиотаж вокруг многофакторного планирования улегся (как это обычно и бывает с новыми методами), а соответствующие алгоритмы (полный факторный эксперимент, крутое восхождение и др.) стали привычными инструментами исследователей.

Трудоемкость вычислительных процедур, необходимых для планирования многофакторного эксперимента и последующей обработки экспериментальных данных, была преодолена благодаря появлению персональных компьютеров. В настоящее время имеется большое количество (по некоторым оценкам, свыше тысячи) пакетов стандартных программ, ориентированных на статистическую обработку экспериментальных данных. Примерами являются Microsoft Excel, Stadia, Statistica, Statgraphics, SPSS, R и некоторые другие англоязычные пакеты программ, обладающие универсальными возможностями. Сегодня любой исследователь должен уметь пользоваться какой-либо программой статистической обработки данных и решения оптимизационных задач [12]. Заметим, что некоторые пакеты программ дают возможность облегчить и последнюю стадию исследования, т. е. позволяют наглядно представить и оформить полученные результаты. Появились даже специализированные программы для планирования многофакторных экспериментов и движения по градиенту, например, программа PLAN (Омск, 2000).

Однако для работы с компьютерными программами недостаточно того, что пользователь знает, как пользоваться той или иной программой, какие из ее модулей следует применять для решения той или иной статистической задачи. Экспериментатор должен понимать, можно ли применять ту или иную процедуру обработки данных при выбранной им схеме эксперимента, при данной методике измерений и т. п. Разные методики приводят к разному распределению погрешностей и требуют разных алгорит-

мов обработки данных. Исследователь-химик должен принимать обоснованные решения и правильно интерпретировать получаемые в ходе эксперимента модели. При этом выводы должны формулироваться не в математических, а в химических терминах.

2.2. ОТБОР ФАКТОРОВ И ТРЕБОВАНИЯ К НИМ

Любой эксперимент начинается с отбора исследуемых факторов. Напомним, что *фактором* называют переменную величину, характеризующую один из способов воздействия на изучаемый объект и принимающую определенное значение (уровень) в единичном опыте. Экспериментатор меняет значения каждого фактора по своему усмотрению, поэтому он должен иметь техническую возможность создавать любой его уровень (в пределах некоторой области допустимых значений) и поддерживать фактор на заданном уровне в ходе единичного опыта. Так, чтобы изучать влияние температуры на выход некоторого продукта реакции, надо иметь не только термометр, но и термостат.

Большинство факторов в химическом эксперименте носят количественный характер, т. е. значения соответствующих величин можно измерять. Для каждого фактора экспериментатор должен иметь методику измерений, обеспечивающую достаточную точность результатов (этот вопрос будет детально рассмотрен в главе 3). Факторы могут иметь и качественный характер. Так, можно проводить один и тот же процесс, используя разные реагенты, разные растворители или разные катализаторы; можно менять от опыта к опыту некоторое оборудование и т. д. Иногда каждый реагент или катализатор можно охарактеризовать количественно, указывая какую-то его характеристику, важную для данного исследования. Например, характеризуя разные растворители, можно указать их вязкость или диэлектрическую проницаемость. Однако растворители различаются и по другим свойствам, пренебрегать этими различиями опасно. Таким образом, расчет математической модели с учетом качественных факторов серьезно осложняется. При дальнейшем изложении

учебного материала качественные факторы мы рассматривать не будем. Сведения по этому вопросу можно найти в специальной литературе [13].

Не все переменные, влияющие на объект исследования, можно использовать в качестве факторов. Из множества переменных отбирают те, которые:

1) *достоверно влияют на параметр*. Для проверки можно провести небольшой предварительный эксперимент, в котором все факторы, кроме интересующего нас фактора X_i , остаются неизменными. Проведя 5–6 опытов на разных уровнях X_i , обрабатывают их результаты методами дисперсионного анализа и узнают, существует ли статистически значимое влияние X_i на Y . Иногда такими расчетами пользуются и для ограничения числа изучаемых факторов: исследуют только те факторы, влияние которых, судя по результатам предварительных опытов, выражено в наибольшей степени;

2) *управляемы*. Это значит, что исследователь имеет возможность установить по своему усмотрению фактор на любом уровне (внутри некоторой области значений) и поддерживать его неизменным в ходе опыта. Любой фактор, которым мы не в состоянии целенаправленно управлять, должен быть зафиксирован на постоянном уровне, одном и том же для всех опытов. Увы, среди множества факторов нередко есть такие, которыми мы не только не можем управлять, но даже не можем зафиксировать (например, атмосферное давление). Случайное изменение таких факторов от опыта к опыту приведет к неточности модели. Сильное влияние неуправляемых факторов (например, скачки напряжения в сети) мешает исследованию или даже делает его невозможным;

3) *однозначны*. Фактор должен прямо указывать на способ воздействия на исследуемый объект и иметь ясный физический смысл. Неоднозначные факторы обычно являются функциями других, более простых факторов (типичные примеры — соотношение концентраций разных реагентов, суммарное содержание однотипных реагентов). Математическую модель, в которую входят неоднозначные факторы, трудно правильно интерпретировать и теоретически объяснить. Тем не менее, неоднозначные фак-

торы применяют довольно широко, особенно в химической технологии;

4) *независимы*. Изменение одного фактора не должно автоматически приводить к изменению другого. Так, бессмысленно брать в качестве одного фактора pH раствора кислоты, а в качестве второго — концентрацию этой кислоты. Из них следует выбрать тот фактор, который измеряется непосредственно. Также бессмысленно планировать эксперимент, в котором некоторая реакция проводится в газовой фазе, а в качестве факторов одновременно используются давление, температура и объем газа. Эти три величины связаны уравнением Менделеева — Клапейрона, и при известной массе газа по двум вышеуказанным величинам всегда можно вычислить третью. В тех случаях, когда имеется множество факторов, и некоторые из них взаимосвязаны, из них можно составить некоторые комбинации, которые окажутся независимыми («метод главных компонент» и другие хемометрические алгоритмы). Такой прием часто используют при оптимизации, однако при изучении механизма реакций он нежелателен, поскольку интерпретация полученной модели будет затруднена;

5) *совместимы*. Это требование означает, что все комбинации факторов внутри их областей допустимых значений осуществимы, безопасны и не приведут к изменению характера процесса. Так, в некоторых системах мы можем как угодно менять температуру при низком давлении или как угодно менять давление при низкой температуре — во всех случаях в реакторе будет образовываться один и тот же целевой продукт, хотя и с разным выходом. Но если одновременно создать высокую температуру и высокое давление, пойдет совершенно другая реакция, которая может привести к выходу реактора из строя. Несовместимость факторов ведет к сокращению их областей допустимых значений.

Выбор факторов — дело очень ответственное, требующее опыта исследовательской работы в данной отрасли науки. Не так опасно ввести «лишний» фактор, даже не отвечающий вышеперечисленным требованиям, как пропустить некоторый важный фактор, не обратить на него внимание. В этом случае эксперимент может стать невоз-

можным или, что еще хуже, привести к совершенно ошибочным выводам. Тем не менее, в 87% исследований в области органической химии исследовалось влияние только двух факторов, и лишь 3% публикаций относились к исследованиям, где было четыре и более факторов [9].

Отметим, что при моделировании химико-технологических и экологических процессов однотипные компоненты часто объединяют («агрегирование»). Так, вместо концентраций десятков разных алканов в модель риформинга включают в качестве нового фактора концентрацию псевдокомпонента «алканы», равную сумме концентраций всех или некоторых индивидуальных алканов [14].

В качестве полезного упражнения можно самостоятельно подобрать факторы для оптимизации процессов, известных любому химику. Подумайте, какие факторы могут повлиять на синтез аммиака, гидролиз сложного эфира, степень экстракционного извлечения тяжелых металлов хлороформом из водного раствора (в виде дитизонатов); результаты раздельного определения компонентов бензина методом газожидкостной хроматографии и т. п. В каждом случае важно разобраться, какие факторы можно и нужно менять в ходе активного эксперимента, а какие следует зафиксировать на постоянном уровне.

2.3.

ВЫБОР ПАРАМЕТРОВ.

ОБОБЩЕННЫЕ ПАРАМЕТРЫ ОПТИМИЗАЦИИ

В учебнике [4] раздел о параметрах оптимизации снабжен неожиданным, но очень удачным эпиграфом, найденным у автора, который вряд ли слышал о планировании эксперимента: «Приплыла к нему рыбка, спросила: “Чего тебе надобно, старче?”» (А. С. Пушкин). Действительно, выбор параметра полностью зависит от целей и желаний экспериментатора, их только надо четко осознать. В меньшей степени параметр связан со спецификой объекта исследования. Так, желая оптимизировать какой-либо химико-технологический процесс, мы обращаем основное внимание на выход целевого продукта, степень чистоты и физико-химические характеристики продукта, затраты времени. Не менее важны экономические параметры

(производительность, себестоимость, рентабельность, прибыль, энергозатраты). Возможны статистические, экологические и даже эстетические параметры. Химизм исследуемого процесса и применяемое оборудование особого значения не имеют!

Если из всех интересующих исследователя параметров можно выбрать один, наиболее существенный, то так и делают, а остальными пренебрегают или, в лучшем случае, следят, чтобы они оставались в допустимых пределах. Так, мы добиваемся максимального выхода некоторого продукта, но вводим условие, что в любых условиях примесей в этом продукте будет не больше чем 1,0%. Иногда значения второстепенных параметров учитывают в виде премий и штрафов. Например, добиваются максимальной прибыли, но прогнозируют ее уменьшение в случае выброса вредных веществ в окружающую среду, так как за это придется платить крупный штраф.

Требования к параметрам включают следующее.

1. *Количественный характер.* В отличие от факторов, параметр не может быть «качественным». Лучше всего, если параметр выражается некоторым числом и во всех опытах измеряется с помощью одного и того же прибора. Такой способ обеспечивает объективность выводов из проведенного эксперимента. Из нескольких однотипных параметров или из нескольких методик измерения одного параметра выбирают те, которые имеют лучшую сходимость.

Если подходящего измерительного прибора или методики измерений не существует, то способ объективной количественной оценки параметра должен придумать сам исследователь. Даже для таких трудно оцениваемых параметров, как эстетические или вкусовые, можно ввести некоторую эталонную шкалу, а потом выражать результат каждого опыта в баллах. Эффективность такого подхода определяется способом построения шкалы (пятибалльная годится только для успеваемости школьников, да и там она слишком грубая) и объективностью эксперта, выставяющего баллы («А судьи кто?»). Тем не менее, оценки, выраженные в баллах, широко используются в практике, например, в гидрохимическом анализе. В литературе детально описаны соответствующие шкалы, например, 0 — вода не имеет запаха, 1 — еле заметный запах, и т. д. При

оценивании параметра можно обойтись без эталонной шкалы, в этом случае каждый опыт должен оцениваться несколькими экспертами, работающими по одной методике, но независимо друг от друга. Наиболее яркие примеры — судейские оценки на соревнованиях по фигурному катанию или оценка качества пищевых продуктов несколькими дегустаторами.

Иногда вместо оценок указывают рейтинг (место в упорядоченном ряду). Например, лучшим результатом (первое место в рейтинге) эксперты признают полученный в опыте № 7, второе место занимает опыт № 4 и т. д. По этой методике идут, например, конкурсы красоты. Для химика рейтинговые оценки в качестве параметра менее удобны, чем оценки, выраженные в баллах.

2. *Эффективность и универсальность*. Из нескольких однотипных количественных оценок выбирают такую, которая наиболее эффективно соответствует конечной цели и характеризует исследуемый объект в целом, а не одну из стадий или частей его. Так, если конечная цель производителя — извлечение максимальной прибыли, то выбор себестоимости в качестве параметра неудачен. Ведь прибыль определяется не только себестоимостью, но и качеством получаемой продукции (от него зависит возможная цена и объем реализации), и способом рекламирования. Еще пример: в одном случае требуется получать продукт, имеющий как можно более высокое содержание некоторого компонента, в другом — требуется обеспечить заданное содержание компонента. В качестве параметра оптимизации можно взять найденное содержание компонента, но во втором случае в качестве параметра лучше использовать модуль или квадрат отклонения содержания компонента от заданного норматива.

3. *Однозначность параметра* означает, что любому сочетанию факторов отвечает только одно значение параметра (если не учитывать случайные погрешности измерений). Однако обратное положение было бы неверно: одно и то же значение параметра может быть получено в самых разных условиях.

4. *Простота вычислений, очевидный физический смысл*. Это требование не требует комментариев.

Если нельзя выделить один из q параметров в качестве основного, для каждого из них рассчитывают отдельную модель, а затем рассматривают их совместно. Однако при решении оптимизационных задач значительно лучше сконструировать обобщенный параметр. Это трудно, так как параметры различны по своей размерности, их нельзя складывать. Кроме того, одни свойства системы надо минимизировать (себестоимость), другие доводить до максимума (выход реакции), а третьи держать как можно ближе к некоторому заданному значению (состав продукта). Поэтому полученные в эксперименте значения частных параметров оптимизации надо выразить в безразмерном виде, а уже потом объединять их. Рассмотрим два способа вычисления обобщенных параметров.

Способ 1. Для получения безразмерных значений используем формулу

$$y_i = \left(\frac{Y_i - Y_{i0}}{Y_{i0}} \right)^2, \quad (2.1)$$

где Y_i — значение i -го частного параметра, полученное в одном из опытов; Y_{i0} — его идеальное значение; y_i — нормированный (безразмерный) i -й параметр. Обобщенный параметр Y_Σ равен сумме всех y_i :

$$Y_\Sigma = \sum \left(\frac{Y_i - Y_{i0}}{Y_{i0}} \right)^2. \quad (2.2)$$

В формуле (2.2) суммирование ведут по всем учитываемым параметрам, от $i = 1$ до $i = q$. Если в каком-либо опыте по всем частным параметрам мы достигли желаемого значения (стопроцентного выхода целевого продукта, нулевого содержания примесей, заданных физических характеристик и т. д.), то все y_i будут равны 0, нулевым будет и значение Y_Σ . Чем больше Y_Σ , тем дальше мы от идеала «в целом», хотя по каким-то частным показателям мы можем получать вполне приемлемые значения. Если отклонения от идеала по одному из параметров гораздо опаснее, чем по другому, надо вводить коэффициенты a_i

(статистические веса параметров), меняющиеся от 0 до 1. Тогда

$$Y_{\Sigma} = \sum a_i \left(\frac{Y_i - Y_{i0}}{Y_{i0}} \right)^2. \quad (2.3)$$

Понятно, что компенсировать существенный недостаток объекта по одному параметру прекрасными качествами по другому параметру нельзя. Так, при выборе квартиры избыточная площадь комнат и прекрасный вид из окон не компенсируют отсутствия водопровода и канализации. Но при использовании формул типа (2.2) или (2.3) такая «компенсация» показалась бы допустимой! Поэтому для расчета обобщенных параметров (например, при сопоставлении ценности квартир) безразмерные оценки лучше не складывать, а перемножать. Но тогда вычислять их надо не по формуле (2.1), а с помощью *функций желательности*, пользуясь изложенным ниже методом Харрингтона.

Способ 2. Для перевода частных параметров Y_i в безразмерные «желательности» d_i следует заранее составить по каждому параметру шкалу в виде графика или таблицы. Для любого параметра значения d_i должны охватывать интервал от 0 до 1, причем нулевая желательность соответствует таким значениям Y_i , которые не устраивают исследователя, а $d_i = 1$ отвечает идеальным значениям Y_i . По каждому параметру экспериментатор указывает те интервалы Y_i , которые, по его мнению, соответствуют плохим, удовлетворительным, хорошим и очень хорошим результатам. Границы интервалов исследователь выбирает на основании своего опыта и интуиции, но лучше запросить мнения компетентных экспертов и усреднить их предложения. Каждому интервалу значений Y_i ставится в соответствие определенное значение d_i , например, «очень хорошие» значения Y_i имеют желательности от 0,80 до 1,00; «хорошие» — от 0,63 до 0,80 и т. д. Любое полученное в эксперименте значение Y_i переводится в d_i с помощью заранее подготовленной номограммы (рис. 2) или каких-либо интерполяционных формул. Обобщенную функцию рассчитывают по формуле Харрингтона:

$$D = \sqrt[q]{d_1 d_2 d_3 \dots d_q}. \quad (2.4)$$

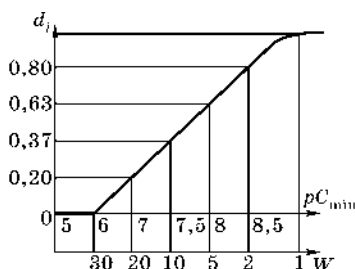


Рис. 2

Номограмма для расчета функций желательности

Пример 2-1. Методика анализа характеризуется величиной pC_{\min} — показателем предела обнаружения, а также коэффициентом вариации W . Эти частные параметры характеризуют чувствительность и воспроизводимость методики. Аналитик одновременно заинтересован в повышении pC_{\min} и в снижении W . До начала исследования некоторый элемент удавалось определять на уровне значений $pC_{\min} = 7,0$ и $W = 2$. В новых условиях получили $pC_{\min} = 8,1$ и $W = 5$. Требуется обобщенно сопоставить исходные и новые условия анализа.

Решение. По обоим частным параметрам выбираем интервалы значений и устанавливаем желательность попадания Y_i в каждый из них (табл. 2).

Таблица 2

Шкала желательностей для некоторой методики анализа

Оценка	d_i	$Y_1 = pC_{\min}$	$Y_2 = W$
Очень плохо	0	< 5	> 30
Плохо	0–0,20	5–6	30–20
Удовлетворительно	0,20–0,37	6–7	20–10
Терпимо	0,37–0,63	7–7,5	10–5
Хорошо	0,63–0,80	7,5–8	5–2
Очень хорошо	0,80–1	8–8,5	2–1
Отлично	1	$> 8,5$	< 1

По номограмме (рис. 2) методом графической интерполяции переводим величины Y_i в значения d_i . В исход-

ных условиях $d_1 = 0,37$ и $d_2 = 0,80$. Отсюда по формуле (2.4) получаем $D = \sqrt{0,37 \cdot 0,80} = 0,54$. Таким же образом находим обобщенный параметр в новых условиях анализа: $D = \sqrt{0,85 \cdot 0,63} = 0,73$. Исходные данные показывают, что переход к новым условиям анализа привел к улучшению первого и ухудшению второго параметра. Учитывая выбранные шкалы желательностей, мы должны признать, что новые условия более благоприятны для проведения анализа, чем прежние.

Если хотя бы один из частных параметров принимает недопустимое значение, обобщенная желательность примет нулевое значение. Компенсация «плохого» параметра другим («очень хорошим») окажется невозможной. Обобщенная желательность D будет равна единице только тогда, когда по всем частным параметрам в эксперименте будут получены идеальные значения Y_i .

Метод Харрингтона — удобный способ конструирования обобщенных параметров, но далеко не единственный. Официальных рекомендаций — как именно нужно конструировать обобщенные параметры — не существует.

2.4. ТРАДИЦИОННАЯ СХЕМА АКТИВНОГО ЭКСПЕРИМЕНТА И ЕЕ НЕДОСТАТКИ

Классический подход к многофакторному активному эксперименту, нацеленному на оптимизацию какого-либо процесса, иногда называют схемой Гаусса — Зейделя. Эта схема заключается в следующем.

1. На основании априорной информации и интуиции исследователь выбирает один из факторов (X_1), влияющих на поведение исследуемой системы, а также начальное значение фактора и шаг его изменения.

2. Проводят серию опытов, изменяя X_1 и сохраняя все прочие условия на постоянном уровне. Обычно меняют фактор с постоянным шагом, проводят 10–15 независимых опытов на разных уровнях. Интервал значений фактора, охваченный такой серией опытов, не должен быть узким, лучше, если величина фактора меняется не менее,

чем на порядок. В каждом опыте измеряют значение параметра (функции отклика). Данные этой серии опытов заносят в таблицу, а затем строят график $Y = f(X_1)$.

3. На основании полученных данных выбирают некоторое значение $X_1 = a$, устраивающее экспериментатора и отвечающее «наилучшему» значению Y , в дальнейшем это значение фиксируется.

4. Меняют фактор X_2 и измеряют Y при прочих постоянных условиях. В частности, при $X_1 = a$. На основании полученных данных строят зависимость $Y = f(X_2)$ и выбирают по ней «наилучшее» значение $X_2 = b$.

5. Меняют третий фактор, измеряя Y при $X_1 = a$ и $X_2 = b$ (третья серия).

Далее можно таким же образом оптимизировать четвертый и следующие факторы.

Для двухфакторного эксперимента эту схему можно показать наглядно, если значения X_1 и X_2 откладывать на осях координат, а одинаковые значения Y соединять плавными линиями, как это делают на географических картах для линий высот или глубин. Тут полная аналогия с высотами, ведь поверхность отклика в трехмерном пространстве будет выглядеть как гора с высотой Y_{\max} . Эта схема показана на рисунке 3а. Может быть, здесь изображена зависимость выхода реакции от температуры и концентрации, а может быть — зависимость скорости коррозии от pH и содержания легирующего элемента в корродирующем материале; в данном случае это не важно.

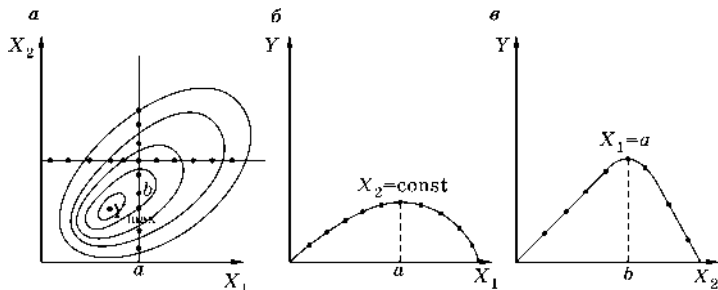


Рис. 3

*Классическая схема двухфакторного эксперимента.
Поиск экстремума поверхности отклика по Гауссу — Зейделю:*

a — поверхность отклика; б — данные первой серии опытов;

в — данные второй серии.

Проведенные серии опытов на рисунке 3а изображаются прямыми, проведенными параллельно координатным осям. Полученные зависимости Y от X_1 и X_2 показаны на рисунке 3б, в. То, что зависимость Y от X_1 , а также от X_2 описывается выпуклыми кривыми, имеющими максимум, соответствует схеме на рисунке 3а. Но, как видно из этого рисунка, проведя две серии экспериментов, мы не вышли на экстремум! Значение Y_{\max} находится вовсе не в точке с координатами (a, b) . Представьте себе туристов, которые, желая взойти на вершину горы, находящуюся на северо-востоке, пошли сначала на север, потом на восток и в результате, обогнув гору, оказались вовсе не на ее вершине. Эта аналогия очень полезна — в главе 10 мы рассмотрим этот пример снова, когда будем знакомиться с методом кругого восхождения.

Могли ли мы попасть к экстремуму, меняя факторы поочередно? Все дело в форме поверхности отклика: если бы замкнутые кривые для фиксированных значений Y имели форму окружностей, поочередное варьирование факторов действительно могло бы привести к оптимальным условиям проведения процесса. А если кривые равного отклика из-за взаимодействия факторов имеют форму эллипсов (как на рисунке 3а) или сама поверхность отклика имеет несколько локальных максимумов, то на основании проведенных серий опытов мы не только не вышли бы к «главному пику», но даже не получили бы верного представления о поверхности отклика.

То, что тысячи исследователей в течение многих десятилетий пользовались и пользуются классической схемой эксперимента, меняют факторы поодиночке, поочередно, объясняется очень просто. Во-первых, это естественное для человека стремление к упрощению, к наглядности, а что может быть нагляднее, чем варьирование одного фактора? Во-вторых, сами исследуемые объекты были относительно простыми, допускали исследование факторов порознь. Да и экспериментальная техника была довольно грубой. В результате взаимодействие факторов не обнаруживалось, и экспериментаторы достигали приемлемых для себя результатов. Но для более сложных объектов данные разных авторов начинали противоречить друг другу, приходилось проводить все новые и новые проверки.

Например, исследователь А утверждал, что скорость некоторой реакции падает при подкислении, исследователь Б — что она возрастает, а по данным исследователя С скорость вообще не зависит от рН. И правы все трое! Дело в том, что объект исследования весьма сложен, скорость реакции зависит не только от рН, но и от температуры, и от начальных концентраций реагентов, и от природы и концентрации посторонних веществ. Причем все эти факторы действуют не по отдельности, а совместно (неаддитивно), в результате влияние рН проявляется в разных условиях по-разному. А в другой, более простой реакции влияние рН не зависит от температуры и начальной концентрации; в каких бы условиях не изучать это влияние, получим одинаковую зависимость скорости от рН. Заранее понять, будут ли взаимодействовать факторы в той или иной системе, обычно не удается.

Итак, классическая схема поочередного варьирования факторов вполне применима для сравнительно простых объектов исследования, где факторы действуют независимо друг от друга и поверхность отклика не имеет каких-то изломов, локальных максимумов и т. п. В более сложных системах классическая схема либо не позволяет получить однозначных выводов, либо ее использование потребует множества опытов, вплоть до перебора всех возможных состояний исследуемой системы, что на практике невозможно.

2.5. СХЕМА И ПЛАН ПОЛНОГО ФАКТОРНОГО ЭКСПЕРИМЕНТА ВИДА 2^k

Со времен Р. Фишера разработано множество схем и планов проведения многофакторных экспериментов (полный факторный эксперимент, дробный факторный эксперимент, центральные композиционные планы, звездные планы, и т. д.), принципиально отличающихся от классической схемы Гаусса. Эти планы имеют много общего. В частности, для них характерны следующие особенности:

- а) изменение всех факторов в одной серии опытов;
- б) минимизация числа опытов за счет уменьшения числа уровней;

в) особое расположение опытов в факторном пространстве, чтобы в дальнейшем можно было легко вычислить математическую модель методами регрессионного анализа (это свойство многофакторных экспериментов называют ортогональностью, см. главу 9).

Наиболее широко применяется схема полного факторного эксперимента (ПФЭ). В предложенных в 1951 г. Боксом и Уильсоном планах ПФЭ каждый фактор задают всего лишь на двух уровнях (в классической схеме используют 10–15 уровней). Если в ПФЭ исследуется влияние температуры и концентрации, то можно провести часть опытов при $t = 30^\circ\text{C}$, а другие — при 50°C ; часть опытов — при $C = 0$, а остальные — при $C = 0,2$ моль/л. Можно провести всего четыре опыта и исчерпать все возможные сочетания уровней (табл. 3).

Т а б л и ц а 3

№ опыта	Температура, $^\circ\text{C}$	Концентрация
1	30	0
2	50	0
3	30	0,2
4	50	0,2

Т а б л и ц а 4

№ опыта	X_1	X_2	X_3
1	—	—	—
2	+	—	—
3	—	+	—
4	+	+	—
5	—	—	+
6	+	—	+
7	—	+	+
8	+	+	+

Подобные планы принято записывать в условной форме, обозначая верхний уровень каждого фактора как (+1), или просто (+), а нижний как (–1), или просто (–), т. е. кодируя натуральные значения переменных. Факторы записывают как X_1 , X_2 , и т. д. Матрица ПФЭ вида 2^3 (трехфакторный эксперимент с двумя уровнями по каждому фак-

тору) после кодирования имеет вид, представленный в таблице 4.

Видно, что восемью опытами мы исчерпали все возможные сочетания уровней. Легко показать, что при четырех факторах ПФЭ должен включать 16 опытов, при пяти — 32 и т. д. Число независимых опытов вычисляется по формуле $N = 2^k$, где k — число исследуемых факторов. Для сложных объектов с 10–15 факторами число опытов все равно будет очень большим, в этом случае вместо ПФЭ применяют дробный факторный эксперимент (так называемую *реплику*). Соответствующие планы описаны в литературе [4], [7]. Чтобы самостоятельно записать матрицу ПФЭ для любого числа факторов, полезно запомнить, что знаки (–) и (+) для первого фактора непрерывно чередуются; для второго фактора чередуются, повторяясь двойками (вначале два минуса, затем два плюса) (см. пример выше); для третьего фактора повторяются четверками (вначале четыре минуса, затем четыре плюса и т. д.), для следующего фактора — восьмерками, и т. д.

Такую матрицу записать очень легко, труднее на ее основе составить реальный план эксперимента. Для этого понадобится вначале выбрать *центр плана*, т. е. такую точку в факторном пространстве, от которой в равной степени будут удалены точки, соответствующие координатам всех опытов начальной серии. Центр плана — это та «печка», от которой начинает «танцевать» исследователь, планируя ПФЭ. Нужно выбрать координаты центра по каждому из факторов с учетом его области допустимых значений. Например, если мы имеем возможность менять температуру в пределах от 20 до 100°, то для «центрального опыта» целесообразно выбрать температуру где-нибудь в середине области допустимых значений — 50 или 60°. Так же выберем координаты «центра» по другим факторам. Реальный опыт в этой точке факторного пространства проводить не обязательно, центр плана нужен только для того, чтобы определить верхний и нижний уровень каждого из факторов.

В натуральных переменных (градусы, моль/л и т. п.) верхний уровень по X_1 соответствует значению фактора в центре плана ($X_{ю}$) плюс некоторый интервал варьирования I_1 ; нижний — значению фактора в центре плана ми-

нус тот же интервал I_1 . Так же находят верхний и нижний уровни по другим факторам. Кодированные значения фактора в принципе могут иметь и другие значения, кроме (+1) и (-1), например, при использовании полученной модели для прогнозов или для оптимизации. Связь между кодированными (безразмерными) значениями фактора (их иногда выделяют жирным курсивом) и значениями его в натуральных переменных показывает соотношение:

$$X_i = (X_i - X_{i0})/I_i. \quad (2.5)$$

Следует только помнить, что по каждому фактору мы берем свой интервал варьирования, и при его выборе учитываем ∂X_i — погрешность данного фактора. Учитывать надо и величину $(X_{i\max} - X_{i\min})$, т. е. ширину области допустимых значений. Должно выполняться условие: $\partial X_i \ll I_i < (X_{i\max} - X_{i\min})$.

Пример 2-2. Исследуется радиационная полимеризация некоторого мономера в растворе. В качестве параметра оптимизации взят выход полимера (Y , %), в качестве факторов — исходная концентрация мономера (C , г/л), температура раствора (t , °C), мощность дозы (W , усл. ед.). Прочие условия опыта фиксируются на постоянных уровнях. Судя по предварительным опытам, процесс полимеризации идет без осложнений, если C варьирует в интервале от 0 до 50, t — от 20 до 80, W — от 200 до 10 000. Факторы можно задавать с погрешностью $\partial C = 1$, $\partial t = 2$, $\partial W = 40$. Составить план ПФЭ вида 2^3 .

Решение. Выберем в качестве центра плана точку, лежащую примерно посередине области допустимых значений по каждому фактору $C_0 = 20$, $t_0 = 50$, $W_0 = 5000$. В качестве интервалов варьирования берем $\Delta C = 10$, $\Delta t = 15$, $\Delta W = 500$. Интервалы варьирования значительно превышают погрешности соответствующих факторов, но существенно меньше ширины области допустимых значений. Рассчитаем координаты нижнего и верхнего уровней: $C_- = 20 - 10 = 10$; $C_+ = 20 + 10 = 30$. Аналогично для температуры: 35 и 65, для мощности дозы: 4500 и 5500. Теперь подставим эти значения в таблицу, которую ранее записывали для ПФЭ 2^3 в общем виде. Добавим один столбец слева (NN — это номер опыта после рандомизации) и

один справа — для записи результата. Полученный план представлен в таблице 5.

Если выполнять опыты в том порядке, в каком мы их записали в плане, т. е. сначала все опыты с меньшей дозой радиации, потом — с большей дозой, то за время выполнения первых опытов могут измениться какие-то другие, неконтролируемые нами условия эксперимента. Влияние дрейфа условий мы не сможем отличить от влияния увеличенной дозы радиации. Чтобы этого не случилось, надо проводить все запланированные опыты (в том числе и повторные) в случайном порядке, как показано в таблице 6. Такой прием называют рандомизацией.

Т а б л и ц а 5

План ПФЭ вида 2^3 до рандомизации (пример 2-2)

№ опыта	NN	C	t	W	Y
1	3	10	35	4500	
2	7	30	35	4500	
3	4	10	65	4500	
4	1	30	65	4500	
5	2	10	35	5500	
6	6	30	35	4500	
7	5	10	65	5500	
8	8	30	65	5500	

Т а б л и ц а 6

Календарный план ПФЭ вида 2^3 с учетом рандомизации (пример 2-2)

Дата	C	t	W	Y
10.04	30	65	4500	
11.04	10	35	5500	
12.04	10	35	4500	
13.04	10	65	4500	
14.04	10	65	5500	
17.04	30	35	4500	
18.04	30	35	5500	
19.04	30	65	5500	

На этом планирование эксперимента можно считать завершенным. После тщательной подготовки к проведению эксперимента (см. главу 3) исследователь должен провести все намеченные опыты в соответствии с календарным планом, т. е. измерять значения Y в разных условиях и записывать их в последний столбец полученной таблицы. Разумеется, измерения проводят многократно, причем записывают в таблицу и значения Y для повторных опытов, и усредненные значения параметра для каждого независимого опыта. А в случае многопараметрического эксперимента в таблицу пришлось бы ввести дополнительные столбцы для других частных параметров, а также для обобщенного параметра оптимизации.

Глава 3

ПОДГОТОВКА К ПРОВЕДЕНИЮ ЭКСПЕРИМЕНТА

3.1.

ЧТО ТАКОЕ «ПОДГОТОВКА К ЭКСПЕРИМЕНТУ»?

Начинать проведение опытов сразу же после составления плана эксперимента не следует. Намеченный эксперимент надо тщательно подготовить. Речь идет не об изучении научной литературы, хотя это весьма важно, а о продумывании всех деталей будущего эксперимента, подготовке проб и проведении предварительных опытов. Непродуманную и плохо подготовленную работу редко удастся довести до конца, а ее результаты обычно оказываются бесполезными. Так, Робинзон Крузо потратил массу времени и сил на изготовление большой лодки, но она оказалась настолько тяжелой, что дотащить ее до берега и спустить на воду ему не удалось. Пришлось делать другую лодку, полегче. Но у Робинзона было сколько угодно свободного времени, а у современного исследователя его нет, и «метод проб и ошибок» ему не подходит!

Конечно, подготовка к эксперименту в каждой области химической науки имеет свою специфику, которая не рассматривается в этой книге. Однако можно выделить некоторые *общие элементы*, характерные для подготовки любого химического или химико-технологического эксперимента.

1. Отбор и предварительная подготовка объектов будущего исследования (синтез исходных химических со-

единений, приготовление модельных смесей, отбор природных веществ, подготовка образцов технических материалов). Допустим, мы собираемся изучать влияние переходных элементов на активность и селективность платиновых катализаторов. В этом случае мы должны заранее выбрать наиболее перспективные элементы-модификаторы; решить, в форме каких соединений и в какой концентрации мы будем вводить эти элементы в исследуемые образцы; проверить наличие соответствующих соединений в нашей лаборатории, и т. п. Придется готовить множество образцов каждого типа (модифицированных разными элементами и немодифицированных). Нередко на подготовительную работу тратится больше времени, чем на сам эксперимент!

2. Выбор способа измерения каждого фактора и параметра. Любую физическую величину можно измерять разными способами. Экспериментатор должен заранее выбрать наиболее подходящий способ измерения каждого фактора и параметра. Для этого придется сравнить точность, чувствительность, экспрессность, трудоемкость и другие характеристики известных методик. Следует также учесть доступность оборудования. Сделанный выбор должен быть окончательным. В ходе эксперимента недопустимо «улучшать» какие-либо методики, поскольку результаты разных опытов окажутся несопоставимыми. Например, нельзя в первых опытах измерять pH раствора, используя универсальную индикаторную бумагу, а в последующих применять pH-метр со стеклянным электродом. Все опыты должны быть проведены с использованием одного и того же оборудования, по одним и тем же методикам.

3. Подбор оборудования, посуды и реактивов. Нередко молодой исследователь сначала выполняет некоторый эксперимент, а уже после окончания всей работы вспоминает, каким оборудованием и реактивами он пользовался. Следует поступать наоборот! Сразу после составления плана будущего эксперимента должен быть составлен список необходимого оборудования, посуды, реактивов и всего остального, что потребуется для выполнения намеченной работы. Особенно важно выбрать подходящие измерительные приборы, провести их поверку, освоить

соответствующие методики измерений и оценить погрешности этих измерений (см. раздел 3.3).

4. Предварительные опыты. Нередко молодые исследователи слишком спешат и не проводят предварительных опытов (не входящих в заранее составленный план многофакторного эксперимента). Однако роль предварительных опытов весьма важна. Эти опыты должны подтвердить саму возможность будущего эксперимента. Так, прежде чем изучать влияние каких-либо факторов на выход некоторого целевого продукта, надо убедиться, что данный продукт действительно образуется в данных условиях! Надо также убедиться, что при повторении любого опыта выход продукта будет примерно таким же, как прежде, в тех же условиях. Известно, что исследования плохо воспроизводимых процессов приводят к ненадежным или даже ошибочным выводам. Поэтому в тех случаях, когда результаты предварительных опытов не воспроизводятся, химики отказываются от намеченного эксперимента или переходят к другим методикам [8]. Иногда меняют план эксперимента, например, отказываются от варьирования каких-то факторов. Предварительные опыты позволяют убедиться, что при изменении условий проведения опытов значения параметра достоверно меняются. Если это не так, придется заново спланировать эксперимент, увеличив интервалы варьирования факторов. В ходе предварительных опытов экспериментатор устанавливает, сколько времени потребуется ему для проведения единичного опыта и эксперимента в целом. Учтите, что продолжительность реального эксперимента всегда окажется большей, чем изначально планирует исследователь. Вспомним: «Гладко было на бумаге, а потом пошли овраги, а по ним ходить...»

3.2. СРЕДСТВА ИЗМЕРЕНИЙ

Подготовка к проведению химического эксперимента и правильная интерпретация полученных данных требуют от начинающего исследователя определенных знаний в области метрологии. Так называют науку об измерениях, методах и средствах обеспечения их единства и способах

достижения требуемой точности измерений [15]. К сожалению, метрологические аспекты научных исследований недостаточно разработаны (в отличие от измерений в промышленности). Кажется, нет даже специальных монографий на эту тему, если не считать классического, но давно устаревшего руководства [16]. К сожалению, курс метрологии не входит в традиционную программу подготовки химиков в российских университетах, за исключением подготовки химиков-аналитиков. Поэтому прежде чем изложить метрологические рекомендации, относящиеся к подготовке химического эксперимента, мы должны дать начинающему исследователю необходимый ему минимум знаний по теоретической и прикладной метрологии. Более полная информация может быть получена при обращении к учебным пособиям по метрологии (например [17]), а также к специальным руководствам по метрологии для химиков-аналитиков [11], [15], [18].

В соответствии с общепринятым определением, *измерение* — это нахождение значения физической величины опытным путем с помощью специальных технических средств. Измерения проводят, сопоставляя физическую величину с заранее принятой *шкалой*, т. е. набором возможных значений этой величины, выраженных в установленных единицах. Обеспечение единства и правильности измерений требует использования соответствующих эталонов, в том числе стандартных образцов состава и свойств.

Измерения можно классифицировать по разным признакам. В зависимости от числа повторных измерений данной величины выделяют *однократные* и *многократные* измерения. Повторение измерений проводят с целью уменьшения влияния случайных погрешностей. Не следует путать *повторные опыты* и многократные измерения. Так, можно три раза провести в одних и тех же условиях синтез некоторого вещества, получить три пробы, измерить интересующее нас свойство каждой пробы, например, показатель преломления, а потом усреднить полученные результаты. А можно провести синтез того же вещества один раз, затем три раза измерить показатель преломления полученного вещества и усреднить полученные данные. Как правило, результаты повторных опытов раз-

личаются гораздо сильнее, чем результаты повторных измерений. Понятно, что проведение повторных опытов потребует гораздо больших затрат труда и времени. Тем не менее, получить надежные результаты исследования без повторных опытов нельзя, многократные измерения их не заменяют.

Многократные измерения могут быть *равноточными* или *неравноточными*. Экспериментатор заинтересован в получении равноточных данных, они надежнее, да и обрабатывать их легче. Иногда выделяют *серийные* измерения — это многократные равноточные измерения одной и той же физической величины для серии однотипных объектов (проб). Именно такие измерения проводят в ходе многофакторного химического эксперимента.

Измерения делят на *статические* и *динамические* в зависимости от того, меняется ли соответствующая физическая величина во время ее измерения. Естественно, добиться высокой точности динамических измерений труднее.

Важнейшая классификация измерений — по способу получения результата. Все измерения делят на *прямые* и *косвенные*. Прямыми называются измерения, при которых искомое значение прямо считывают со шкалы прибора и никаких расчетов не проводят. Пример — измерение времени с помощью часов. В косвенных измерениях значения интересующей нас величины находят расчетным путем на основании их известной зависимости от других величин, которые измеряют непосредственно. Такую зависимость выражают каким-либо уравнением, например линейной или степенной зависимостью. Пример — измерение плотности раствора по результатам измерения массы раствора и объема этого раствора. Результат здесь находят путем деления, по формуле $Z = X_1/X_2$. Погрешность измерения Z зависит в основном от погрешности измерения величин X_1 и X_2 , но в ряде случаев надо учитывать и дополнительные погрешности.

Измерения проводят в соответствии с определенной методикой выполнения измерений (МВИ). Понятие МВИ определяется следующим образом: это совокупность операций и правил, выполнение которых обеспечивает получение результата измерения с погрешностью, не превышающей заданного предела. В промышленности МВИ

обычно оформляют в виде некоторого документа, официально утверждают его. В научных исследованиях это не требуется. Сам исследователь может и должен найти в литературе или самостоятельно разработать методики выполнения будущих измерений, записать их в лабораторный журнал и в дальнейшем строго соблюдать эти методики. Их приводят и в отчете о проведенном исследовании.

В любой МВИ указывается, какие средства измерений должны быть использованы, в каких условиях должны проводиться измерения, в каких пределах может находиться измеряемая данным способом физическая величина. Но самая важная характеристика МВИ — это предельная погрешность соответствующих измерений. Какую бы физическую величину мы ни измеряли, какую бы методику ни использовали, но мы должны заранее знать предельную величину абсолютной (Δ) либо относительной (δ) погрешности. Если такой информации нет ни в паспорте измерительного прибора, ни в описании МВИ, предельную величину погрешности («неопределенность») должен рассчитать сам исследователь — в ходе предварительных опытов [11], [17]. Соответствующие данные понадобятся для обоснованного выбора средств измерений, для сопоставления точности измерения разных факторов, для проверки адекватности математической модели и во многих других случаях.

Средством измерения называют техническое средство (устройство, прибор, установку и т. п.), используемое для измерений и имеющее нормированные метрологические характеристики. Например, спектрометр измеряет интенсивность излучения исследуемого объекта на разных длинах волн и регистрирует спектр — зависимость интенсивности излучения (или поглощения) от длины волны. Все средства измерений (СИ) периодически поверяют с помощью эталонов и рабочих мер, воспроизводящих некоторое значение измеряемой величины. Так, аналитические весы поверяют с помощью набора образцовых гирь, потенциометры — с помощью стандартного гальванического элемента Вестона. Ответственные научные эксперименты следует проводить с применением таких СИ, которые предварительно поверены специалистами и снабжены необходимыми сертификатами. Однако многие средства измерений

не подлежат официальной поверке (например, мерная посуда). В специальной литературе описаны методики, с помощью которых сам экспериментатор может проверить вместимость данной мерной колбы или оценить предельную погрешность измерения объема титранта с помощью данной бюретки.

Готовясь к эксперименту, следует не только оценить погрешность измерений (см. раздел 3.3), но и установить другие метрологические характеристики измерительных приборов: диапазон измерений, цену деления, чувствительность прибора и др. Для некоторых приборов важной метрологической характеристикой является *инерционная постоянная*, определяющая возможность применения этого прибора для измерения быстро меняющихся величин.

3.3. ПОГРЕШНОСТИ ИЗМЕРЕНИЙ. СПОСОБЫ ИХ СНИЖЕНИЯ

Самыми важными метрологическими характеристиками любого измерительного прибора и любой методики измерений являются предельные значения абсолютной и относительной погрешности.

Абсолютную погрешность прямого измерения находят по очевидной формуле $\Delta = X - Q$, где X — результат измерения, а Q — действительное значение измеряемой величины. Для того чтобы рассчитать Δ по вышеприведенной формуле, надо знать Q — действительное значение измеряемой величины, а оно, как правило, неизвестно. Для оценки Q можно применить другой прибор или другую методику измерений, для которых погрешность измерений пренебрежимо мала. Если эталонных способов измерения в нашем распоряжении нет, величину Δ можно считать равной цене деления. Однако оценка Δ по цене деления обычно дает лишь нижний предел абсолютной погрешности, она не учитывает сходимости данных и наличия систематической погрешности. Верхний предел величины Δ всегда больше цены деления. Можно было бы в качестве оценки этого предела использовать данные по разбросу результатов повторных измерений, но и тогда мы не учли бы систематическую погрешность, которая может

быть намного больше случайной. Наиболее точный способ — провести с помощью данного прибора и методики измерение ряда эталонов с точно известными значениями Q , найти предельную величину Δ , а потом считать, что при других измерениях величина Δ останется на том же уровне.

Абсолютную погрешность измерений учитывают при записи результатов этих измерений. Полученные результаты должны быть записаны в лабораторный журнал без всякой статистической обработки, без каких-либо поправок и без какой-либо отбраковки «ошибочных» результатов. Обработку результатов по соответствующим алгоритмам проводят позднее. Записывать результат прямого измерения принято так, чтобы предпоследняя цифра не вызывала сомнений, а последняя соответствовала бы абсолютной погрешности измерения. Так, если объем титранта записан в виде «24 мл», это означает, что измерение проведено с погрешностью порядка 1 мл. Если же пользовались более точными приборами, и погрешность измерений не превышала 0,1 мл, запись должна быть сделана с точностью до десятых долей миллилитра. Например, «24,0 мл».

Важной характеристикой методики измерений является *предел допустимых расхождений* между результатами двух повторных измерений. Выше уже указывалось, что во многих научных исследованиях по экономическим соображениям не проводят большого числа повторных опытов. Для экономии времени не проводят и большого числа повторных измерений в единичном опыте. Как правило, каждый независимый опыт повторяют (в одних и тех же условиях) только один раз, а при проведении каждого опыта ограничиваются только двумя измерениями параметра. Результаты повторных измерений (а тем более результаты повторных опытов) могут различаться довольно сильно. Не исключено, что одно из двух полученных значений является грубым промахом, но при столь малом числе измерений (опытов) этого доказать нельзя, статистические критерии отбраковки промахов требуют гораздо большего объема данных. Но можно легко определить, являются ли результаты, полученные в параллельных измерениях (или в параллельных опытах), *совместимыми*. Так называют результаты двух повторных измерений

(или повторных опытов), когда расхождения между ними не превосходят некоторый заранее установленный предел Δ_{\max} . Появление «несовместимых» результатов не дает оснований отбраковать какой-то из них, объявив его грубым промахом, но указывает экспериментатору на необходимость дополнительных измерений (или повторения опыта). Тогда, получив в тех же условиях еще несколько значений измеряемой величины, можно будет определить, какое из первоначальных измерений являлось грубым промахом, отбросить соответствующий результат, а остальные усреднить.

Величина Δ_{\max} должна указываться в МВИ или в паспорте измерительного прибора, но зачастую его приходится устанавливать самому исследователю при подготовке к выполнению эксперимента. Величина «предела допустимых расхождений» тем больше, чем больше предельная погрешность измерений по данной методике.

Способы расчета Δ_{\max} описаны во всех руководствах по метрологии. Простейший из них, применимый только при нормальном распределении результатов измерений, основан на предварительной оценке величины σ (стандартного отклонения), характеризующей воспроизводимость результатов измерений по данной методике. При нормальном распределении вероятность случайного расхождения результатов двух измерений одной и той же величины больше, чем на 4σ очень мала ($< 0,05$). Поэтому иногда считают, что $\Delta_{\max} = 4\sigma$. Иногда критерий 4σ заменяют более строгим (6σ) или менее строгим (3σ). В описании же методики исследований этот критерий выражают прямо в единицах измеряемой величины или в процентах. Например, пишут: «Если результаты анализа в повторных опытах различаются менее, чем на 2% (абс.), то их среднее арифметическое следует считать равным действительному выходу продукта реакции в данных условиях. Если же различие больше 2%, то следует провести еще два-три опыта в тех же условиях». Иногда в МВИ указывается целая таблица значений Δ_{\max} для разных диапазонов измеряемой величины, а подходящее значение этого критерия выбирают в зависимости от того, в какой диапазон попадают результаты измерений.

Погрешность косвенных измерений можно прогнозировать. Напомним, что результат косвенного измерения получают путем расчета (в частности, именно так находят результат количественного химического анализа). Понятно, что результат расчета не может быть точнее, чем исходные данные — результаты прямых измерений. Погрешность расчета включает погрешности исходных данных (*элементарные погрешности*), их суммируют по особым правилам.

Допустим, интересующую исследователя величину Z рассчитывают на основе результатов прямых измерений разнородных величин по формуле $Z = \varphi(X_1, X_2, \dots, X_n)$. Как именно это надо делать, зависит от того, какие погрешности складываются (*случайные* или *систематические*), и от того, какие действия производят с исходными данными в ходе расчета. Обозначим максимально возможные значения элементарных погрешностей каждого X символами $\Delta X_1, \Delta X_2, \Delta X_3$ и т. д., в общем случае ΔX_i . Знаки каждой погрешности не учитывают, используют их модули. Пусть основной вклад в общую погрешность всех измерений вносят систематические погрешности. Тогда результирующую погрешность (ΔZ) рассчитывают по одной из двух формул:

- при $Z = X_1 X_2$ или при $Z = X_1 / X_2$

$$\Delta Z / Z = \Delta X_1 / X_1 + \Delta X_2 / X_2; \quad (3.1)$$

- при $Z = X_1 + X_2$ или при $Z = X_1 - X_2$

$$\Delta Z = \Delta X_1 + \Delta X_2. \quad (3.2)$$

Если же погрешности измерения разных X в основном носят *случайный* характер, то при $Z = X_1 X_2$ или $Z = X_1 / X_2$ следует заранее оценить стандартные отклонения аргументов, а затем применить формулу:

$$(\sigma_z / Z)^2 = (\sigma_{X_1} / X_1)^2 + (\sigma_{X_2} / X_2)^2. \quad (3.3)$$

Можно использовать и приближенную формулу:

$$\Delta Z / Z \approx \sqrt{\left(\frac{\Delta X_1}{X_1}\right)^2 + \left(\frac{\Delta X_2}{X_2}\right)^2}. \quad (3.4)$$

Если результат косвенного измерения находят путем сложения или вычитания, т. е. $Z = X_1 + X_2$ или $Z = X_1 - X_2$, применяют другую приближенную формулу:

$$\Delta Z \approx \sqrt{(\Delta X_1)^2 + (\Delta X_2)^2}. \quad (3.5)$$

Таким образом, независимо от того, какие погрешности суммируются — случайные или систематические, при оценке точности сумм и разностей суммируют *абсолютные погрешности слагаемых*. При оценке точности произведений или частных суммируют *относительные погрешности сомножителей*.

Пример 3-1. В мерной колбе готовят по точной навеске раствор некоторого реактива. Значение концентрации (C) рассчитывают по результатам прямых измерений массы реактива (m) и объема колбы (V). Навеска реактива массой 5,000 г взята на весах, дающих систематическую погрешность порядка 0,001 г. Объем мерной колбы на 50 мл известен с точностью до 0,1 мл (предельная величина систематической погрешности измерения объема). Указать, какова точность измерения концентрации этого раствора.

Решение. Концентрацию раствора рассчитываем по формуле $C = m/V$, которая ведет к значению $C = 0,100$ г/мл. Предполагаем, что погрешности исходных измерений имеют случайный характер. Так как в ходе расчета проводили деление результатов прямых измерений, суммирование элементарных систематических погрешностей требует применения формулы (3.4). В данном случае: $(\Delta_c/C)^2 = (\Delta_m/m)^2 + (\Delta V/V)^2 = (0,001/5)^2 + (0,1/50)^2 \approx 4 \cdot 10^{-6}$, откуда $\Delta_c/C \approx 0,2\%$. Основную роль явно играет неточность объема мерной колбы, а влиянием погрешности взвешивания можно пренебречь. Так поступают, когда одно слагаемое на порядок (в 10 и более раз) меньше других.

В случаях, не требующих высокой точности расчетов, можно считать, что погрешность конечного результата косвенного измерения определяется лишь одним — наименее точным — слагаемым или сомножителем. А именно, при сложении и вычитании наименее точным слагаемым считают то, у которого была самая большая абсолютная по-

грешность, тогда $\Delta Z \approx \Delta X$. При умножении или делении наименее точным сомножителем считают тот, у которого самая большая относительная погрешность. В этом случае $\Delta Z/Z \approx \Delta X/X$. Отметим, что в тех случаях, когда результат косвенного измерения получают, умножая результат прямого измерения X на постоянный, точно известный коэффициент K , не являющийся результатом измерения, то относительная погрешность величины Z равна относительной погрешности X и не зависит от K .

Предельную величину абсолютной погрешности иногда называют *точностью*. Например, говорят, что с помощью данного термометра можно измерить температуру с точностью до 1 градуса. Однако термин «точность» является неоднозначным, его используют для обозначения самых различных величин. Лучше использовать термин «предельная величина погрешности».

Для исследователя очень важно знать и уметь применять разные *способы снижения погрешности измерений*. Прежде всего надо понять, какая из составляющих общей погрешности — случайная или систематическая — в данном случае является основной. Именно ее и следует снижать в первую очередь, а затем перейти к устранению другой составляющей, меньшей по величине. Способы устранения случайных и систематических погрешностей различны. Разумеется, существуют и общие рекомендации (тщательное выполнение всех операций, описанных в МВИ; использование приборов высокого качества, поверенных с помощью надежных эталонов; оптимизация условий проведения всех измерений и т. п.), но этого недостаточно, нужны специализированные приемы.

Снижение влияния случайных погрешностей достигается тремя основными способами: а) стабилизацией условий проведения измерений; б) увеличением числа повторных измерений; в) повышением числа и размера проб исследуемых материалов, а также числа повторных опытов. *Стабилизация условий* требует, чтобы все операции, выполняемые исследователем, проводились на одних и тех же приборах, с применением одной и той же мерной посуды, одних и тех же растворов и реагентов, одним и тем же человеком. Даже порядок смешивания реактивов должен быть

строго одним и тем же. Влияние колебаний рН раствора при переходе от пробы к пробе можно устранить с помощью буферных растворов; влияние колебаний температуры — путем термостатирования исследуемых проб. Колебания напряжения в сети не скажутся на результатах, если включать измерительный прибор через дополнительный стабилизатор. Список подобных рекомендаций можно продолжить, они специфичны для каждого метода измерений. Стабилизация условий достигается и путем автоматизации измерений — компьютерная программа, в отличие от человека-оператора, обеспечит для всех проб строго одинаковый способ измерений.

Заманчивым приемом является *увеличение числа повторных измерений*. Как будет далее показано в следующих главах, увеличение в n раз числа повторных измерений снижает случайную погрешность среднего результата в \sqrt{n} раз. Однако этот прием неудобен — химические эксперименты и без того слишком длительны и трудоемки. Как правило, экспериментатор проводит не более трех повторных измерений и не более двух повторных опытов. Кроме того, если основной вклад в общую погрешность вносит стадия пробоотбора (а так бывает нередко), то даже многократное измерение сигнала данной пробы не поможет добиться высокой точности результата. Необходимо случайным образом отбирать множество проб, измельчать их, тщательно перемешивать весь материал, уменьшать массу пробы, затем проводить многократные измерения и, наконец, усреднять полученные результаты.

Для профилактики систематических погрешностей особенно важны: выбор оптимальной методики измерений, отделение и маскирование мешающих компонентов, правильная настройка измерительных приборов, применение чистых химических реактивов, выбор наиболее подходящих эталонов, а также использование оптимального способа расчетов. Полезными приемами являются *рандомизация* и *релятивизация*. Первый термин означает параллельное использование нескольких однотипных МВИ или измерительных приборов, отобранных случайным образом. Каждый прибор дает свою систематическую погрешность измерений, но для совокупности таких при-

боров их систематические погрешности переходят в разряд случайных, а влияние последних уменьшится после усреднения полученных результатов. Второй термин означает замену абсолютных измерений относительными.

3.4. МЕТРОЛОГИЧЕСКИЕ АСПЕКТЫ МНОГОФАКТОРНОГО ЭКСПЕРИМЕНТА

Обозначим относительную погрешность i -го фактора как ∂X_i , в нее входит и погрешность измерения этого фактора, и случайные колебания (флуктуации) X_i во время опыта. Величину ∂X_i учитывают при построении плана эксперимента. А именно, соседние уровни должны отличаться не на величину ∂X_i , а значительно сильнее, иначе мы не обнаружим достоверного влияния данного фактора. Различие значений фактора при переходе от одного опыта к другому называют *шагом*. При традиционной схеме проведения эксперимента обычно величина шага постоянна. Для любого фактора шаг должен не менее чем на порядок превышать погрешность этого фактора. Так, имея термометр с ценой деления в 1° , мы не должны проводить опыты при 20° , 21° , 22° и т. д., но при достаточно точном термостатировании можем провести их через каждые 10° , например, измерять скорость реакции в растворе при 20° , 30° , 40° , 50° . Шаг эксперимента зависит от того, какую область допустимых значений имеет данный фактор, т. е. каков диапазон возможных значений X_i . Шаг должен быть намного меньше, чем ширина этого диапазона. Так, в вышеприведенном примере нецелесообразно было бы проводить опыты через каждые 50° — слишком мало опытов мы сумели бы провести. Кроме того, при переходе от опыта к опыту будет слишком сильно меняться параметр Y (скорость реакции).

Погрешности факторов (значения ∂X_i) зависят от того, с какой точностью мы измеряем соответствующие физические величины и насколько точно поддерживаем выбранное значение каждого фактора в ходе повторных опытов. Заранее отметим, что наиболее простой способ обработки результатов многофакторного анализа (регрессионный анализ) требует, чтобы значения факторов были определены

с гораздо меньшей относительной погрешностью, чем значения параметров. А измерять параметры надо как можно точнее, иначе не добиться требуемой точности модели. О возможных погрешностях измерений следует подумать до начала эксперимента, потом будет поздно. Немало важных и ответственных исследований было испорчено из-за несвоевременного стремления исполнителей улучшить методику измерений, сменить ее уже после начала эксперимента. Дело в том, что в одном и том же исследовании нельзя одновременно совместно использовать данные, полученные до и после «улучшения» методики. Нельзя же сопоставлять достижения спортсменов, если высоту планки, преодоленной на данном чемпионате одними спортсменами, измерять плохой рулеткой, а другими — хорошей!

До сих пор разные факторы и параметры, как и погрешности измерения соответствующих физических величин, рассматривались нами по отдельности. Однако при планировании будущего эксперимента погрешности всех факторов должны рассматриваться совместно. Только тогда экспериментатор сможет обоснованно и правильно выбирать методики измерений, подбирать необходимую измерительную аппаратуру, да и сами объекты исследований. Нельзя сопоставлять абсолютные погрешности разных факторов по их величине. Ведь сами факторы, как и их абсолютные погрешности, имеют разную размерность. Зато относительные погрешности всех факторов — безразмерные величины, их можно и нужно сопоставлять, учитывая следующие положения.

1. При правильном выборе аппаратуры, методики и объектов исследования *относительные погрешности разных факторов должны быть близки между собой*. Так, при проведении некоторого синтеза бессмысленно измерять температуру порядка 100°C с помощью сверхточного электронного термометра с абсолютной погрешностью $0,01^{\circ}\text{C}$ (это даст $\delta_t \approx 0,01\%$), если время реакции порядка 50 мин измеряется с точностью до 1 мин ($\delta_{\tau} \approx 2\%$), а концентрация реагента порядка 0,5 моль/л — с точностью до 0,1 моль/л ($\delta_c \approx 20\%$).

2. Если параметр Y функционально связан с несколькими факторами (например, $Y = kX_1X_2...X_k$), то относи-

тельная погрешность результата (величина δY) в основном будет определяться *погрешностью того фактора, который дает наибольшее значение относительной погрешности* ($\delta Y \approx \delta X_{\max}$). Относительная погрешность определения параметра может быть значительно больше относительной погрешности «худшего» фактора, но она не может быть меньше! Так, в приведенном примере выход реакции не удастся оценить с погрешностью, меньшей 20% отн. При повторении синтеза в тех же условиях результаты повторных экспериментов будут варьировать именно в таких пределах. Если же потребуются лучшая точность эксперимента (например, надо будет оценить выход продукта с погрешностью, не большей 2% отн.), следует изменить методику измерения концентрации, снизить величину δ_c до 1–2%.

3. Если для исследуемого объекта ожидается более сложная связь между параметром и факторами, не сводящаяся к формулам типа $Y = kX_1X_2X_3$, то для оценки точности измерения параметра Y следует учитывать не только относительную погрешность каждого фактора, но и степень его влияния на параметр. Малозначимые факторы можно измерять и задавать с меньшей точностью, чем другие, и это не вызовет существенного ухудшения точности оценки Y .

4. Бессмысленно использовать точные методики измерений и прецизионную измерительную аппаратуру, если образец, взятый для измерения свойств данного вещества, является «нетипичным» или «плохо определенным». Пропитируем классическое руководство [16]: «Было бы бесполезной тратой времени производить точные исследования какого-либо свойства воды, например ее электропроводности, на каком-нибудь случайном образце неизвестного происхождения, но с чрезвычайной тщательностью и применяя ряд экспериментальных приспособлений». Так же бесполезно было исследовать свойства «хорошо определенных» и «типичных» образцов, используя грубые методы, ведущие к получению неточных результатов. Готовясь к проведению эксперимента, надо позаботиться и о правильном отборе проб, и о высокой точности всех измерений, и о подходящих способах обработки полученных данных.

Глава 4

РЕЗУЛЬТАТ ИЗМЕРЕНИЯ КАК СЛУЧАЙНАЯ ВЕЛИЧИНА

4.1. СЛУЧАЙНЫЙ ХАРАКТЕР РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

В ходе математической обработки экспериментальных данных результаты измерений (например, значения скорости коррозии, результаты химического анализа, выходы реакций органического синтеза) считают *случайными величинами*. Это связано с двумя основными причинами:

1) при повторном измерении той же величины прежний результат, как правило, воспроизводится неточно, так как на саму эту величину и на процесс ее измерения влияет множество неучитываемых и несогласованно меняющихся факторов, что приводит к случайным погрешностям;

2) сам исследуемый объект часто имеет индивидуальные особенности (внутреннюю неоднородность или нестабильность свойств во времени), из-за чего каждая его часть, которую мы используем для измерений, случайным образом отличается по своим свойствам от других частей и от усредненных свойств объекта в целом [11], [19]. Так, найденное по некоторой методике среднее содержание некоторой примеси в пяти пробах, отобранных из партии каучука, не обязательно соответствует составу всей партии.

Средний результат анализа (\bar{x}) будет отличаться от истинного содержания этой примеси (Q) как из-за погрешностей анализа, так и вследствие случайного характера пробоотбора.

Не следует думать, что измерение с одинаковой вероятностью может привести к любому значению, от $-\infty$ до $+\infty$. Случайный характер получаемых при измерении данных вовсе не означает их хаотичности. Конечно, нельзя заранее указать, какой результат появится именно в данном измерении, но при рассмотрении серии повторных измерений одной и той же величины можно заметить четкие закономерности. Как и другие случайные величины, результат измерения имеет определенную область возможных значений (иногда говорят «область изменения» или «область варьирования»). Так, при определении SiO_2 в горной породе результат в любом случае попадает в интервал от 0 до 100%.

Полная характеристика случайной величины (в том числе результата измерения) требует, чтобы были указаны все ее возможные значения и чтобы было определено, как часто она принимает каждое из них. Для исследователя важно знать, какова вероятность получить при измерении тот или иной результат. Для разных результатов внутри области варьирования эти вероятности, как правило, неодинаковы (*неравномерное распределение*). Вероятность получения тех или иных результатов может быть задана в табличной форме или с помощью графика, или с помощью некоторых математических формул — *функций распределения* (в интегральной или дифференциальной форме). В этих формулах аргументом является значение случайной величины (или набор значений), а функцией — вероятность получить эти значения в ходе измерений. В дальнейшем мы рассмотрим, какие именно функции распределения описывают результаты реальных измерений, проведенных в ходе химического эксперимента.

Поскольку результат измерения любого параметра оптимизации является случайной величиной, постольку для математической обработки результатов измерений в химических исследованиях могут и должны использоваться методы математической статистики и теории вероятностей.

4.2. ВЫБОРКИ. РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРОК

Измерения обычно проводят так, чтобы выделить совокупности однотипных объектов. Объекты совокупности могут отличаться друг от друга по одному или нескольким признакам (соответственно *одномерные* и *многомерные совокупности*), но они должны быть сходны между собой в других отношениях, качественно однородны. Так, студенты одной и той же группы отличаются друг от друга по массе тела и цвету глаз, но сходны между собой по возрасту и получаемой в вузе специальности. Различаются между собой и повторные результаты анализа, но они однородны, если относятся к одному и тому же исследуемому материалу; характеризуют содержание одного и того же компонента; выполнены в одной и той же лаборатории, одним и тем же лаборантом, по одной и той же методике; выражены в одних и тех же единицах и близки по своей числовой величине. В приведенных примерах мы видим количественно выраженный *признак* (масса тела, содержание примеси), который меняется (варьирует) при переходе от одного объекта к другому, от одного измерения к следующему. Числовое значение признака (в частности, результат измерения) называют *вариантой*, далее будем обозначать варианты как $x_1, x_2, \dots, x_i, \dots, x_n$. Если в совокупность входит n однотипных объектов, то соответственно будет получено n вариант. Здесь n — объем совокупности.

Признак может быть выражен качественно, примером может быть видимая окраска раствора. Но всегда следует стремиться к количественному выражению признака. Например, исследуя серию растворов, желательно регистрировать не только видимую окраску каждого раствора (его цвет), но и спектры этих растворов, а затем указывать для каждого раствора положение спектрального максимума, выражая его в нанометрах или в других единицах.

Если признак может выражаться только определенным набором чисел (например, целыми числами или числами, кратными какой-то постоянной величине), совокупность называют *дискретной*. Такие совокупности часто встре-

чаются в биологии (например, число зубов у отдельных людей варьирует, но оно обязательно целочисленное!), демографии, социологии. В атомной физике и спектроскопии распространены квантованные величины; все они также образуют дискретные совокупности. Измеряемые же химиками величины обычно не квантованы, мы в основном имеем *непрерывные совокупности*, где варианты могут принимать любое значение внутри некоторого интервала. Так, при измерении pH серии водных растворов все результаты будут принадлежать к интервалу (0, 14), но каждый из них в зависимости от состава раствора может быть каким угодно, а не только целочисленным или кратным некоторой другой величине. Такие совокупности мы и будем в дальнейшем рассматривать. Для них функции распределения непрерывны и, как правило, дифференцируемы.

Следует учесть, что результаты измерения непрерывной случайной величины записываются экспериментатором с учетом погрешности измерений (цены деления и т. п.), а потому число возможных результатов ограничено. Если измерять pH с точностью до 0,1 единицы, то в области (0, 14) окажется всего 140 возможных уровней pH; если перейти к прибору, позволяющему вести измерения с точностью до 0,01 единицы, получим 1400 уровней, и т. д. Число же вариантов в совокупности результатов измерений может быть бесконечно большим, ведь варианты могут повторяться на одном и том же уровне.

Полную характеристику исследуемого материала (процесса, явления и т. д.) дает только *генеральная совокупность вариантов*, в которую входят все объекты данного вида. В частности, генеральную совокупность образуют все возможные результаты измерения некоторой величины с учетом вероятности получения каждого результата. Объем генеральной совокупности теоретически не ограничен, возможно, $n = \infty$. С другой стороны, любой исследователь по экономическим и другим соображениям проводит лишь небольшое число измерений интересующей его величины и имеет в своем распоряжении лишь ограниченное число вариантов. Подобные совокупности называют *выборками*; имеют в виду, что входящие в них результаты измерений как бы «случайно выбраны» из генеральной

совокупности всех возможных измерений той же величины. Выборочный метод — один из основных принципов научного исследования не только в химии, но и в науке в целом.

Одной и той же генеральной совокупности может соответствовать множество выборок. Варианты любой из них сходны друг с другом по своей числовой величине, но по некоторым признакам могут отличаться от других вариантов из той же генеральной совокупности. Например, по методике или по времени проведения измерений. Если в генеральной совокупности и в выборке из нее наблюдается один и тот же характер распределения вариантов, выборку называют *репрезентативной (представительной)*. Поскольку о свойствах исследуемого объекта судят не по генеральной совокупности измерений, а по некоторой выборке из нее, то эта выборка обязана быть репрезентативной. Так, желая определить соотношение между частотой рождения мальчиков и девочек, никто не подсчитывает, сколько мужчин и женщин было рождено на Земле «от сотворения мира» до начала подсчета и сколько еще их должно родиться. Исследователь-демограф считает, сколько мальчиков и девочек родилось, скажем, за месяц в некотором городе (по данным роддома или ЗАГС), т. е. делает небольшую выборку и определяет по ней интересующее его соотношение. Для проверки то же соотношение он определяет и по другим выборкам — за другие месяцы или по другим местностям. Для большей достоверности выводов рекомендуется делать больше независимых выборок и увеличивать объем каждой из них. Если результаты обработки разных выборок с достаточной точностью совпадут, выборки считают репрезентативными. Только в этом случае полученные данные можно использовать для характеристики генеральной совокупности и делать по ним выводы о свойствах объекта в целом.

Репрезентативность выборки определяется методикой измерений, в частности методикой пробоотбора. Из-за непродуманного способа получения исходных данных многие выборки оказываются нерепрезентативными, неправильно передают характер генеральной совокупности. Так, если исследователь-демограф, изучающий соотношение

численности мужчин и женщин, будет брать исходные списки не из данных переписи всего населения, а в роддомах, в штабах воинских частей или в деканате химического факультета, то как бы тщательно он потом ни проводил подсчеты, его выводы будут заведомо неверными. Об этом будут, в частности, свидетельствовать резкие отличия в исходных выборках, «штабная» и «деканатская» выборки будут совершенно непохожи друг на друга по своим характеристикам!

Для получения репрезентативных выборок каждая наука имеет свои правила. Так, нельзя получить верных выводов о составе почвы на некотором поле, анализируя единичный образец или отбирая несколько образцов только в одном месте, в агрохимии существуют определенные правила пробоотбора.

Отметим, что часто экспериментаторы регистрируют не все результаты проведенных ими измерений, а только те, которые они субъективно считают «правильными», т. е. отвечающими заранее выдвинутой гипотезе. Это грубое нарушение репрезентативности, которое приведет к совершенно неверным выводам о свойствах изучаемого объекта.

Чем больше объем репрезентативной выборки, тем ближе ее параметры к параметрам генеральной совокупности. Это положение известно под названием «закона больших чисел» (есть и другие формулировки того же закона). Данный закон долго считали препятствием для применения количественных методов в естественных науках (слишком малые выборки обычно имеют в своем распоряжении исследователи!). Иногда из-за этого обстоятельства делали даже философский вывод о принципиальной непознаваемости реальных явлений: нужны-де для этого генеральные совокупности, а у вас всегда только выборки! Но развитие микростатистики в работах В. Госсета^{*} (1908) и других авторов позволило с заданной вероятностью предсказывать свойства генеральной совокупности на основе небольших репрезентативных выборок.

^{*} Он писал под псевдонимом Стьюдент.

4.3. ВЫБОРОЧНЫЕ ПАРАМЕТРЫ

Для выборок малого объема можно вычислить некоторые параметры. Их рассчитывают без всяких априорных допущений о характере распределения вариантов; это эмпирические характеристики. В процессе вычисления выборочных параметров используют суммы (Σ) различных величин, включающих варианты $x_1, x_2, \dots, x_i, \dots, x_n$. Для упрощения записи формул принимаем, что в операции суммирования индекс i пробегает значения от 1 до n . В отдельных случаях границы индекса суммирования приводятся непосредственно в формулах или оговариваются в тексте.

Выборочные параметры разделяют на две большие группы.

1. *Характеристики центра рассеяния* (среднее арифметическое, среднее геометрическое, мода, медиана и др.). *Среднее арифметическое* $\bar{x} = 1/n \Sigma x_i$ — наиболее важная из характеристик этой группы. Если число вариантов велико и результаты были классифицированы, среднее арифметическое рассчитывают с учетом заселенности каждого класса по формуле $\bar{x} = \Sigma f_k x_{dk}$, где индекс суммирования k означает номер класса.

Доказано, что если нет систематических ошибок, то при достаточно большом ($n > 30$) числе повторных измерений $\bar{x} \approx Q$, где Q — действительное значение измеряемой величины. То есть \bar{x} приблизительно соответствует «правильному» измерению. Но при малых n не следует пренебрегать возможностью случайного отличия найденного среднего арифметического от истинного значения измеряемой величины, даже если выборка репрезентативна. Для оценки Q в этих случаях рассчитывают доверительные интервалы.

Среднее геометрическое \bar{x}_g используют в тех случаях, когда предполагают, что распределение имеет логарифмически нормальный характер, например, при определении «следов» или обработке полуколичественных результатов измерений с очень большим разбросом вариантов. Варианты

с отрицательными или нулевыми значениями при расчете \bar{x}_g отбрасывают. Применяют следующие формулы:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n} \quad \text{или} \quad \lg \bar{x}_g = 1/n \sum_{i=1}^n \lg x_i.$$

Мода (Мо) — наиболее часто встречающаяся в выборке варианта, она соответствует максимуму на гистограмме, т. е. среднему значению признака в наиболее заселенном классе. *Медиана* (Ме) — значение варианты, которое стоит точно посередине ранжированной совокупности. Так, при четном n медиана — полусумма двух вариантов, стоящих посередине ряда; при нечетном — варианта, равноотстоящая от обоих концов ряда. Так, в выборке 5, 7, 8, 10 медиана равна $(7 + 8) : 2 = 7,5$; в выборке 5, 8, 10 медиана равна 8. Моду и медиану используют для больших совокупностей, они мало чувствительны к характеру разброса вариантов относительно центра рассеяния. Приблизительное равенство моды, медианы и среднего арифметического указывает на симметричное распределение вариантов в выборке.

2. *Характеристики степени рассеяния вариантов* относительно некоторого центра рассеяния (практически всегда относительно \bar{x}). В эту группу выборочных параметров входят дисперсия, среднее отклонение, стандартное отклонение, относительное стандартное отклонение, коэффициент вариации, размах и некоторые другие характеристики.

Дисперсия обозначается как S^2 и рассчитывается по формуле

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}. \quad (4.1)$$

Величина выборочной дисперсии S^2 , полученная по результатам n повторных независимых измерений одной и той же случайной величины, при $n \rightarrow \infty$ стремится к предельному значению, называемому дисперсией генеральной совокупности и обозначаемому как σ^2 . Обычно в хи-

мических исследованиях величину σ^2 оценивают по формуле (4.2), применимой при $n > 30$:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (4.2)$$

Дисперсии используют при сравнении выборок, а также при выявлении относительного вклада разных факторов в общую погрешность измерений (дисперсионный анализ). Расчеты дисперсий по приведенным выше «классическим» формулам без компьютера или программируемого микрокалькулятора требуют слишком много времени. Для облегчения расчетов применяют ряд приемов:

1) все варианты увеличивают или уменьшают на постоянную поправку A (обычно берут целочисленную). Так, вместо того, чтобы обрабатывать выборку 101, 103, 98, 100, 99, можно воспользоваться выборкой 1, 3, -2, 0, -1. Дисперсия при таком преобразовании не изменится;

2) все варианты выборки можно заранее увеличить в 10^n раз, до целочисленных значений. Получив дисперсию по новой выборке, ее умножают на 10^{-2n} ;

3) при расчетах дисперсий можно использовать формулы, не содержащие отклонений отдельных вариантов от среднего арифметического:

$$S^2 = \frac{1}{n-1} (\sum x_i^2 - n\bar{x}^2) = \frac{1}{n-1} \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right). \quad (4.3)$$

Если случайные величины (результаты измерений) складывают или перемножают, или совершают какие-то другие действия над случайными величинами, то дисперсию результата вычисляют по особым правилам (см. [15], [17]). Иногда выборочные дисперсии обозначают иными символами (V или D).

Стандартное отклонение — основная характеристика разброса вариантов. Имеет ту же размерность, что и сами варианты. Эту величину также называют среднеквадратической ошибкой или среднеквадратичным отклонением отдельного измерения. Для характеристики небольших

выборки ($n < 30$) используют выборочное стандартное отклонение S , рассчитывая его по формуле

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}. \quad (4.4)$$

Стандартное отклонение генеральной совокупности обозначают σ и при $n > 30$ оценивают его по формуле

$$\sigma \approx \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}. \quad (4.5)$$

В обоих случаях берут только положительное значение корня.

Относительное стандартное отклонение S_r — безразмерная величина, которую находят при делении S на \bar{x} . Это очень важный показатель, характеризующий воспроизводимость (сходимость) методики измерений. *Коэффициент вариации* W (или cv) — то же, что и относительное стандартное отклонение, но выраженное в процентах. Очевидно, $W = 100S_r$.

Среднее отклонение $D = 1/n \sum |x_i - \bar{x}|$ широко использовалось в прошлом, но сейчас выходит из употребления, так как величина D не входит в формулы, описывающие теоретические распределения случайных величин. Для расчета D отклонения отдельных вариантов от среднего арифметического берут по модулю, иначе они бы при суммировании компенсировали друг друга.

Размах R равен разности $x_{\max} - x_{\min}$. Считать его легко, но по величине R нельзя судить о характере распределения вариантов в выборке. Хотя размах не является основной характеристикой выборки, в некоторых случаях можно использовать его для оценки гораздо более важного выборочного параметра — стандартного отклонения. Дин и Диксон предложили вместо трудоемкого расчета S умножать размах R на некоторый табличный коэффициент k^* , зависящий от числа вариантов в выборке. Так, по Дину и Диксону, для $n = 5$ $k^* = 0,43$. Примером может

быть выборка 14,5; 14,7; 14,3; 14,4; 14,1. В данной выборке размах равен 0,6; величина S примерно равна $0,6 \cdot 0,43 \approx 0,26$. Обычный метод расчета дает $S = 0,22$. Расхождение относительно невелико. Однако метод Дина и Диксона следует применять очень осторожно, так как правильные результаты получаются лишь при отсутствии грубых промахов (размах очень чувствителен к ним) и лишь при нормальном распределении.

Наиболее важные параметры выборки — среднее арифметическое и стандартное отклонение. Именно их используют, проверяя характер распределения.

4.4. ГИСТОГРАММЫ. ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

Выше уже указывалось, что варианты следует записывать в лабораторный журнал с учетом погрешности применяемой методики измерений (сомнительной должна быть последняя оставляемая при округлении цифра). Если варианты получены при повторном измерении свойств одного и того же объекта или при исследовании однотипных образцов, полученных из одного объекта («параллельные пробы»), то варианты записывают в порядке получения.

После записи всех вариантов некоторой совокупности исследователь обычно проверяет наличие промахов — вариант, резко отличающихся по величине от всех остальных вариантов данной совокупности, обычно из-за грубых ошибок самого исследователя, а затем исключает их. Однако с этой операцией не следует спешить. Способы отбраковки грубых промахов обычно выбирают в зависимости от того, по какому закону распределены варианты в генеральной совокупности измерений, проводимых по данной методике, а это известно далеко не всегда. Соответствующие алгоритмы будут описаны позднее.

Результаты повторных измерений в достаточно длинной их серии могут систематически сдвигаться в меньшую (или, наоборот, большую) сторону, это явление называют *дрейфом*. Дрейф вызывается не случайными погрешностями измерений, а систематически действующим фактором, например, постепенным окислением проб после их отбора.

Для выявления дрейфа используют разные статистические алгоритмы, например, разбивают исследуемую выборку на две части (в порядке получения вариантов), а затем определяют статистическую достоверность различия средних значений, полученных для этих частей (двух малых выборок). При наличии дрейфа нельзя ни усреднять варианты, ни рассчитывать другие параметры выборки.

Выборки большого объема принято ранжировать, т. е. переписывать все полученные варианты в порядке возрастания x от x_{\min} до x_{\max} , а не в порядке получения вариантов в ходе эксперимента. При этом обращают внимание на повторяемость отдельных вариантов.

Для того чтобы наглядно представить, как именно распределены варианты в выборке, выборки большого объема группируют, т. е. их варианты разносят по классам одинаковой ширины. Затем подсчитывают, сколько вариантов окажется в каждом классе. Границы классов стараются проводить по целочисленным значениям x . Число классов n_d берут от 4–5 до 10–15. Разумеется, чем больше объем выборки, тем больше может быть и n_d . Ширину класса d выбирают так, чтобы $d \cdot n_d \geq R$. Как исключение, допускается лишь увеличение ширины двух крайних классов. Чтобы не было никаких сомнений, к какому классу отнести «пограничную» варианту, верхняя граница одного класса не должна совпадать с нижней границей следующего. Эти границы должны отличаться на абсолютную погрешность измерений. После разнесения всех вариантов для каждого класса рассчитывают *заселенность* t (эту же величину называют *частотой*) и *относительную частоту* $f = t/n$. Среднее значение признака в каждом классе (x_{di}) иногда используют для расчета выборочных параметров.

После классификации получаем таблицу (*вариационный ряд*), показывающую, как часто в выборке встречаются варианты каждого класса. Полученные данные можно представить графически, в виде гистограммы. Это набор прямоугольников, высота которых равна t (или t/n), а основание — соответствующий границам класса отрезок оси абсцисс. Иногда вместо гистограмм строят *полигоны*. Так называют ломаные линии, соединяющие точки в середине каждого класса на высоте, равной t/n .

Пример 4-1. Содержание бензола в одной и той же пробе определяли 18 раз. Построить гистограмму. Найденные значения, в % :

35,0 37,5 42,3 32,3 38,4 36,6 41,7 37,5 39,1
 36,0 37,5 36,2 36,5 36,6 40,2 36,8 38,9 36,9

Р е ш е н и е. Проверка не выявила промахов и дрейфа. Ранжируем выборку:

32,3 35,0 36,0 36,2 36,5 (36,6)₂ 36,8 36,9
 (37,5)₃ 38,4 38,9 39,1 40,2 41,7 42,3

Запись (36,6)₂ означает, что этот результат в выборке повторяется два раза. Размах R равен $42,3 - 32,3 = 10$. Чтобы классифицировать выборку, введем шесть классов шириной по две единицы, начиная с 32,0. Тогда классы в сумме перекроют размах выборки. Результаты представим в виде таблицы 7, а затем построим гистограмму (рис. 4).

Т а б л и ц а 7

**Классификация вариантов в выборке результатов анализа
для примера 4-1**

Номер класса	Границы класса	Частота, m	Относительная частота, f
1	32,1–34,0	1	0,056
2	34,1–36,0	2	0,111
3	36,1–38,0	9	0,500
4	38,1–40,0	3	0,167
5	40,1–42,0	2	0,111
6	42,1–44,0	1	0,056

Изучение гистограмм позволяет выявить важные свойства совокупностей.

1. **Унимодальность.** Как правило, результаты измерений дают один максимум на гистограмме, т. е. наибольшую заселенность имеет один класс, называемый модальным. В примере 4-1 модальный класс 36–38, а мода равна 37,0. Иногда экспериментатор получает данные, при представлении которых гистограммой появляются два максимума. Однако большие по объему совокупности, имеющие

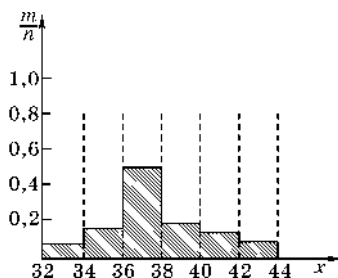


Рис. 4

*Гистограмма, характеризующая
распределение вариантов в примере 4-1*

два отчетливо выраженных максимума на гистограмме, обычно внутренне неоднородны и могут быть разбиты на две (или более) независимые однородные выборки. Так, при измерении длины ступни в большой совокупности людей (это делают для планирования выпуска обуви разных размеров) получают среднее арифметическое, равное 39,5 условных единиц. Но гистограмма, построенная по полученным данным, имеет два максимума — около 37,5 и 42,0 единиц. Причина ясна: 37,5 — мода для женской обуви*, 42,0 — для мужской. Исследователям надо было все результаты измерений разбить на две выборки, т. е. изучать распределения длины ступни у женщин и мужчин по отдельности. Заметим, что найденное по внутренне неоднородной выборке среднее значение 39,5 соответствует обуви, которая, возможно, понадобится лишь немногим мужчинам и женщинам, оно лежит в минимуме гистограммы, между горбами.

В химическом эксперименте получают внутренне неоднородные совокупности, когда часть результатовотягощена систематической ошибкой. То же может произойти, если смешивают результаты измерения одной и той же величины по двум разным методикам. Выявить такие «сюрпризы» без построения гистограмм трудно. Пример с обувью показывает, что усреднять варианты внутренне неоднородной совокупности нельзя, и вообще никакая статистическая обработка к таким совокупностям не применяется.

* Не путать с модными фасонами!

2. *Симметрия.* Как правило, реальные гистограммы симметричны. Мода расположена в середине размаха и близка к среднему арифметическому. Гистограммы могут оказаться несимметричными из-за неверно выбранной размерности признака — скажем, следовало бы брать не просто содержание какой-то микропримеси, а логарифм этого содержания. Такое преобразование первоначально измеренного признака требуется при очень сильном варьировании его, когда значения x отличаются на порядок и более. После логарифмирования гистограмма становится симметричной. Изредка асимметрия является действительным свойством выборки, не устранимым при логарифмировании (несимметричные распределения).

3. *Форма гистограмм.* Малые отклонения от центра рассеяния обычно встречаются чаще, чем большие. Крылья гистограммы снижаются в обе стороны. При многократном повторении измерений не исключены даже очень большие отклонения от центра рассеяния, они только маловероятны.

4. *Вырождение в плавную кривую.* С увеличением объема совокупности мы сможем при построении гистограммы увеличивать число классов и уменьшать ширину каждого из них. В этом случае при $n \rightarrow \infty$ гистограмма или полигон превратятся в плавную кривую (рис. 5). Эта кривая будет характеризовать уже не данную выборку, а всю генеральную совокупность, даст полную характеристику объекта по исследуемому признаку. Гистограмма по заселенностям (частотам) превратится в кривую, которую называют *дифференциальной функцией распределения* или

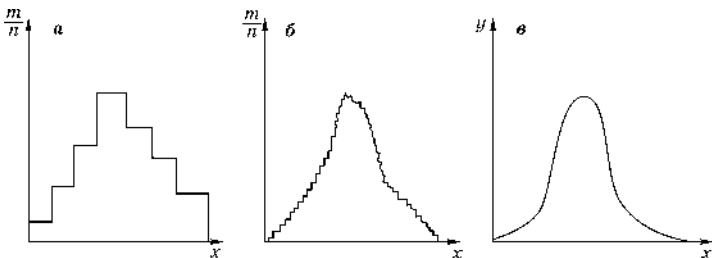


Рис. 5

Получение дифференциальной функции распределения:

$a - n = 30; n_d = 7$ классов; $b - n = 1000, n_d = 70$ классов; $c - n = \infty, n_d = \infty$.

функцией плотности вероятности. Ее описывает некоторая алгебраическая формула вида $y = f(x)$, где $y = \lim m/n$ при $n \rightarrow \infty$.

Интересно и очень важно то, что самые разные объекты исследования, самые разные методики измерений приводят всего лишь к нескольким функциям распределения. Известны нормальное распределение, логнормальное, равномерное, биномиальное, экспоненциальное; распределения Пуассона, Стьюдента, Вейбулла, Коши и некоторые другие теоретические распределения, редко описывающие результаты реальных измерений. Исследователь должен помнить, что математическую обработку экспериментальных данных проводят по-разному, в зависимости от того, какому теоретическому распределению подчиняется совокупность полученных данных. Если оно не было известно до начала статистической обработки, то придется проверить характер распределения полученных данных. Обычно для этого по выборке рассчитывают некоторые эмпирические параметры, а затем сопоставляют их с табличными характеристиками известных распределений (прежде всего с характеристиками нормального распределения). При таком сопоставлении используют некоторые численные критерии. Описание соответствующих алгоритмов приведено в следующих главах этой книги.

Глава 5

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

5.1.

ПРОВЕРКА ХАРАКТЕРА РАСПРЕДЕЛЕНИЯ

Проверять характер распределения вариант следует при разработке (освоении, аттестации) новой методики измерений. Если в описании некоторой методики указано, к какому распределению результатов она приводит, то дополнительную проверку проводить не обязательно. Для проверки потребуется сделать довольно много (обычно не менее 30) измерений и вычислить выборочные параметры S и \bar{x} . Затем проверяют, согласуются ли экспериментальные данные с предположением о нормальном (наиболее часто встречающемся на практике) характере их распределения. Для этого либо вычисляют асимметрию и эксцесс полученной выборки (упрощенный способ), либо сопоставляют по критерию Пирсона близость полученной в эксперименте гистограммы к теоретической кривой нормального распределения. Второй способ сложнее, но дает более надежные результаты (см. главу 6). Его лучше использовать при наличии компьютерной программы.

Разберем первый способ. Коэффициент асимметрии g находят по формуле

$$g = \frac{\sum (x_i - \bar{x})^3}{n \cdot S^3}. \quad (5.1)$$

Можно предварительно разнести варианты по классам, а затем рассчитывать величину g по средним значениям признака в каждом классе и относительным заселенностям классов. Этот способ используется химиками редко.

В симметричных распределениях, в частности, при нормальном распределении вариант, значение g должно быть равно 0 или незначительно от него отличаться. Так, в явно симметричной выборке 34, 32, 36 среднее значение равно 34, отклонения соответственно равны 0, -2, +2, их кубы -0, -8, +8, а сумма кубов и коэффициент асимметрии равны 0. Если $g \gg 0$ (левосторонняя асимметрия) или $g \ll 0$ (правосторонняя), то распределение несимметрично, а следовательно, не может быть нормальным (рис. 6). Асимметрия считается статистически доказанной, если значение g по модулю превосходит некоторое критическое значение $g_{\text{крит}}$, приведенное в приложении 2А. На практике распределение считают асимметричным, если коэффициент асимметрии в три и более раз превосходит по модулю критерий S_g , который вычисляют по формуле

$$S_g = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}. \quad (5.2)$$

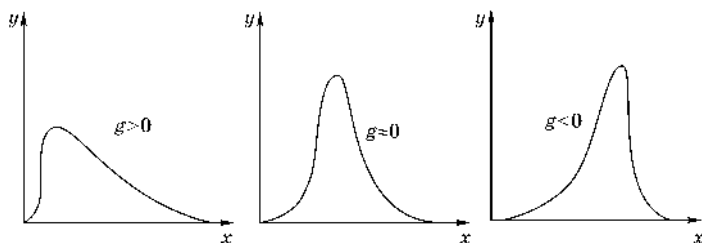


Рис. 6

Вид гистограмм при разных значениях коэффициента асимметрии

Кроме асимметрии, распределение должно быть проверено по эксцессу (рис. 7). О наличии эксцесса судят по коэффициенту E :

$$E = \frac{\sum (x_i - \bar{x})^4}{n \cdot S^4} - 3. \quad (5.3)$$

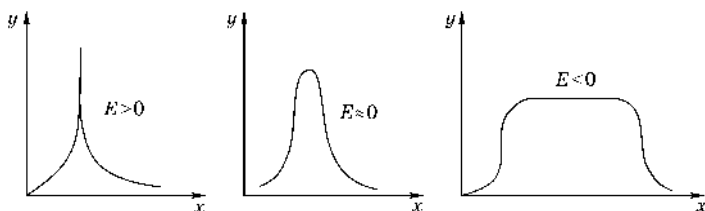


Рис. 7

Вид гистограмм при разных значениях коэффициента эксцесса

Для нормального распределения вариант величина E должна быть приблизительно равной 0. Отличие E от нуля может быть статистически незначимым, поэтому сравнивают величину E (по модулю) с его критическим значением $E_{\text{крит}}$, приведенным в таблицах (приложение 2В). Вывод о наличии эксцесса будет достоверным лишь при выполнении условия: $|E| > E_{\text{крит}}$. Другой способ — проверить выполнение условия $|E| > tS_e$, где S_e — рассчитывают по формуле (5.4), а t — коэффициент Стьюдента, зависящий от объема выборки и желаемой надежности вывода:

$$S_e = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}. \quad (5.4)$$

На практике считают достаточным, чтобы значение E превосходило (по модулю) величину S_e в три раза, это уже указывает на достоверное отличие от нормального распределения.

Если методика должна приводить к нормальному распределению, а на практике получено статистически значимое отрицательное значение E , то это говорит о внутренней неоднородности выборки, например, о наличии в части измерений систематической ошибки. Положительный эксцесс может указывать на неслучайный характер отбора вариант (например, на подгонку данных под желаемый средний результат измерения, т. е. на фальсификацию).

Если g , и E оказались незначимы, то можно считать, что методика измерений приводит к нормальному распределению. Правда, теоретически возможны и другие распределения с нулевыми значениями асимметрии и эксцесса,

поэтому нормальный характер распределения вариант нельзя считать абсолютно доказанным. Для большей гарантии можно провести проверку распределения по более строгому критерию Пирсона (см. раздел 7.4.5 и приложение 2С). Если хотя бы один из параметров (g или E) будет статистически значим, можно утверждать, что исследуемая величина распределена не по нормальному закону.

Обратите внимание, что по значениям g и E мы не можем определить, какое именно из «ненормальных» распределений имеет место в данном случае! Чтобы узнать это, придется последовательно проверять другие возможные распределения. Соответствующие алгоритмы описаны в монографиях по математической статистике [20], [21]. Для подбора подходящего распределения измеряемой величины желательно построить гистограмму и сравнить ее форму с кривыми, характерными для разных теоретических распределений.

5.2. ПРОИСХОЖДЕНИЕ И СВОЙСТВА НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

В химических исследованиях наиболее часто приходится встречаться с нормальным распределением вариант. Оно появляется в следующих случаях:

1) если отклонения x от среднего значения вызываются одновременным, несогласованным и разнонаправленным действием многих факторов примерно одинаковой значимости (следствие предельной теоремы Ляпунова);

2) если варианты представляют собой средние значения из каких-то малых выборок, относящихся к одной и той же генеральной совокупности;

3) если при повторных измерениях отклонения вариант от центра их рассеяния вызываются действием единственного фактора (меняющегося от измерения к измерению), то и в этом случае x может иметь нормальное распределение. Может, но не обязательно имеет — все дело в том, что за фактор, как именно он варьирует и как действует на процесс измерения.

В ходе химического анализа отклонения вариант обычно бывают связаны с действием множества несогласован-

ных и разнонаправленных факторов приблизительно одинаковой значимости (что соответствует случаю 1). Поэтому результаты анализов, как правило, распределены по нормальному закону. Бывают и исключения. Результаты полуколичественных анализов при очень низком содержании определяемой примеси (близком к ее пределу обнаружения) обычно подчиняются так называемому логарифмическому нормальному распределению. А если методика анализа основана на подсчете импульсов или частиц (дискретные совокупности), наблюдается распределение Пуассона. Однако чем больше число учтенных импульсов, тем ближе распределение Пуассона к нормальному, различием в характере распределений можно будет пренебречь [22].

При нормальном распределении случайной величины дифференциальная функция распределения описывается функцией Гаусса:

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (5.5)$$

где y — *плотность вероятности* — соответствует вероятности попадания единичной варианты в бесконечно малый интервал от x до $x + dx$. Вид функции распределения зависит от величин σ и μ , входящих в формулу (5.5) и называемых параметрами нормального распределения. Отметим, что каждое теоретическое распределение имеет свой набор параметров, которые не следует путать ни с выборочными параметрами, ни с параметрами оптимизации, непосредственно измеряемыми в ходе эксперимента. Число параметров произвольного теоретического распределения может быть и больше, и меньше двух. Так, распределение Пуассона характеризуется единственным параметром μ .

В формуле (5.5) σ — стандартное отклонение генеральной совокупности, характеризующее воспроизводимость данной методики измерений; μ — математическое ожидание случайной величины x , т. е. среднее арифметическое всех вариантов соответствующей генеральной совокупности. Отсутствие систематических погрешностей величина μ

соответствует действительному значению измеряемой величины, т. е. $\mu = Q$. Геометрически функция Гаусса имеет вид симметричной кривой с единственным максимумом, причем положение максимума определяется величиной μ , а высота и полуширина этой кривой (гауссианы) — величиной σ (рис. 8).

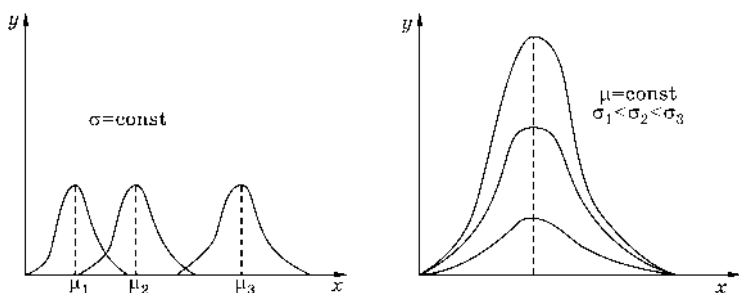


Рис. 8

Характерный вид гауссиан для разных параметров нормального распределения

Зная μ и σ для данной совокупности, можно рассчитать вероятность P попадания результата единичного измерения (величины x) в любой заданный интервал (a, b) . Для этого следует вычислить определенный интеграл

$$P(a, b) = \int_a^b y dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (5.6)$$

Рассчитывать интегралы вида (5.6) сложно, а устанавливать их табличные значения нецелесообразно, так как для каждой пары μ и σ пришлось бы составлять свою таблицу. Поэтому решая задачи, связанные с нормальным распределением, значения x выражают, пользуясь новой переменной $U = (x - \mu)/\sigma$. Вследствие такой операции на графике, показывающем распределение результатов измерений, центр распределения сдвигается в начало координат, т. е. точке с $x = \mu$ будет соответствовать $U = 0$ и максимальное значение функции y . Значению x , которое на 2σ превосходит μ , будет соответствовать $U = 2$ и т. д. Перевод x в U и обратно надо уметь делать автоматически.

После перехода к новой переменной формула (5.5) упрощается:

$$y = f(U) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{U^2}{2}}. \quad (5.7)$$

Кривые нормального распределения с разными μ и σ (например, полученные для разных методик измерения или для разных объектов) после перехода к новой переменной выражаются единственной кривой (рис. 9, справа). Площадь между этой кривой и осью абсцисс в пределах от U_1 до U_2 характеризует вероятность попадания варианты в соответствующий интервал, выраженный в единицах U . При этом вся площадь под кривой будет равна 1, ведь то, что случайная величина U примет хоть какое-нибудь значение — от $(-\infty)$ до $(+\infty)$ — абсолютно достоверное событие.

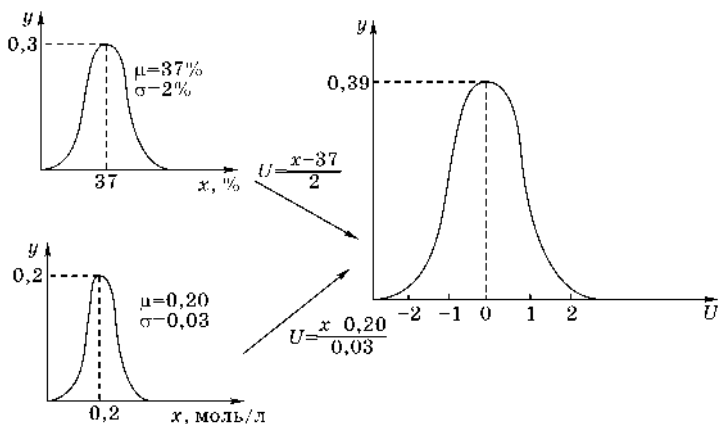


Рис. 9

Переход к безразмерным значениям U

Вероятность попадания варианты в интервал (a, b) можно рассчитать двумя способами: через гауссов интеграл ошибок либо с помощью функций Лапласа. Применим второй способ. Так как вероятность попадания варианты в интервал (a, b) после перехода к новой переменной не изменится, то при $U_2 > U_1 > 0$

$$P(a < x < b) = P(U_1 < U < U_2) = \Phi(U_2) - \Phi(U_1). \quad (5.8)$$

Значения $\Phi(U_2)$ и $\Phi(U_1)$ находят, используя таблицу значений функции Лапласа, которая задается формулой

$$\Phi(U) = \int_0^U f(u) du.$$

Значения $\Phi(U)$ для разных неотрицательных U — табличные интегралы, приведенные в статистической литературе, а в данной книге в приложении 2D. Так, $\Phi(0) = 0$; $\Phi(1) = 0,3413$; $\Phi(\infty) = 0,5$. При возрастании U величина $\Phi(U)$ монотонно и очень быстро приближается к 0,5; для значений $U > 4$ можно считать, что $\Phi(U) \approx 0,5$. Для отрицательных U функцию Лапласа находят по модулю U . Например, $\Phi(-1) = 0,3413$. Очень удобна геометрическая модель: *функциям Лапласа отвечают площади под кривой нормального распределения от оси симметрии до любого U .*

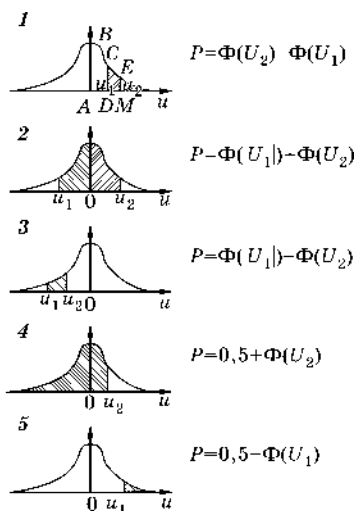


Рис. 10

Разные случаи расчета вероятностей
с применением функций Лапласа

Как видно из рисунка 10, при подсчете вероятностей значения функций Лапласа надо складывать, если U_1 и U_2 имели разные знаки, и вычитать (из большей меньшую), если знаки U_1 и U_2 были одинаковы. Общая площадь под

кривой нормированного нормального распределения с любой стороны от оси симметрии равна 0,5000. Из геометрической модели легко понять, что вероятность попадания случайной величины в симметричный интервал от $(-U)$ до $(+U)$ равна удвоенной функции Лапласа от U . Например, $P(-0,45 < U < 0,45) = 2\Phi(0,45) = 2 \cdot 0,1736 \approx 0,35$.

5.3.

РАСЧЕТЫ, ОСНОВАННЫЕ НА ПРИМЕНЕНИИ ФУНКЦИЙ ЛАПЛАСА

Если воспроизводимость измерений точно известна (величина σ дана в описании методики эксперимента или указана в паспорте прибора, или оценена экспериментатором заранее), то с помощью таблицы функций Лапласа можно решать большое число практически важных задач. Например:

1) предсказывать вероятность попадания результата эксперимента в интересующий нас интервал (a, b) ;

2) рассчитывать вероятность нахождения истинного значения измеряемой величины в любом заданном интервале значений x ;

3) рассчитывать границы доверительного интервала (a, b) , в котором с заданной вероятностью находится заранее неизвестное истинное значение измеряемой величины.

Пример 5-1. Рассчитать вероятность, с которой результат спектрального определения никеля в стали попадет в интервал 1,9–2,5%, если истинное содержание никеля составляет $Q = 2,1\%$, а методика анализа характеризуется значением $\sigma = 0,08\%$. Здесь и далее предполагается, что методика анализа не дает систематических погрешностей, т. е. $\mu = Q$, а результаты повторных анализов имеют нормальное распределение.

Решение. Задача относится к первому типу. Заданные границы интервала выражаем в безразмерных значениях U :

$$U_1 = \frac{1,9 - 2,1}{0,08} = -2,5; \quad U_2 = -\frac{2,5 - 2,1}{0,08} = 5.$$

Находим функции Лапласа:

$$\Phi(-2,5) = \Phi(2,5) = 0,4938; \quad \Phi(5) = 0,499997 \approx 0,5000.$$

Искомая вероятность равна сумме функций Лапласа, так как их аргументы имеют разные знаки. Геометрическая модель отвечает случаю 2 на рисунке 10.

$$\begin{aligned} P(1,9 < x < 2,5) &= P(-2,5 < U < 5) = \\ &= \Phi(2,5) + \Phi(5) \approx 0,4938 + 0,5000 \approx 0,9938. \end{aligned}$$

Вероятность попадания результата анализа в заданный интервал больше 99%.

Пример 5-2. Рассчитать вероятность, с которой средний результат из четырех проведенных анализов той же стали, что и в примере 5-1, даст результат, превосходящий 2%. Методика измерений та же.

Решение. Задача относится к тому же (первому) типу, что и предыдущая, но имеет два осложнения. Во-первых, в условии указана только одна граница интервала; во-вторых, речь идет о попадании в интервал не результата единичного измерения, а среднего значения. Однако эти осложнения не принципиальны. А именно, интервал значений x , указанный в задаче, на самом деле является двусторонним: от 2% до ∞ . Далее следует учесть, что \bar{x} — среднее арифметическое из n вариантов — варьирует относительно μ в меньшей степени, чем единичные варианты. Известно, что стандартное отклонение для \bar{x} равно $\sigma_{\text{ср}} = \sigma/\sqrt{n}$. В данном примере $\sigma_{\text{ср}} = 0,08/2 = 0,04$. Соответственно находим границы интервала для новых (безразмерных) координат.

$$U_1 = \frac{a - \mu}{\sigma} = \frac{2,0 - 2,1}{0,04} = -2,5; \quad U_2 = \frac{b - \mu}{\sigma} = \frac{\infty - 2,1}{0,04} = \infty.$$

Искомая вероятность равна сумме функций Лапласа от аргументов $U_1 = -2,5$ и от $U_2 = \infty$: $\Phi(|-2,5|) + \Phi(\infty) = 0,4938 + 0,5000 = 0,9938$.

Пример 5-3. Определить вероятность, с которой истинное содержание никеля в некоторой стали находится в интервале от 15 до 20%, если результат анализа оказался равным 18%, а относительное стандартное отклонение методики равно 0,10. Здесь действуют те же оговорки, что в примере 5-1.

Решение. Эта задача относится ко второму типу. Чтобы решить ее, следует прежде всего по величине σ ,

найти само стандартное отклонение σ , а затем выразить границы интервала в безразмерных переменных. Значение μ неизвестно, но очевидно, что оно варьирует относительно x в точности так же, как величина x варьирует относительно μ . Поэтому для расчета границ интервала применяем формулы, аналогичные тем, что были использованы в предыдущих примерах:

$$U_1 = \frac{a-x}{\sigma}; \quad U_2 = \frac{b-x}{\sigma};$$

$$\sigma = \sigma_r \cdot x = 0,10 \cdot 18\% = 1,8\%;$$

$$U_1 = \frac{15-18}{1,8} = -1,67; \quad U_2 = \frac{20-18}{1,8} = 1,11.$$

Находим функции Лапласа по таблицам с применением интерполяции, а затем складываем их, так как аргументы имеют разные знаки:

$$\Phi(1,67) = 0,453; \quad \Phi(1,11) = 0,368;$$

$$P(15 < \mu < 20) = 0,453 + 0,368 = 0,821.$$

Таким образом, вероятность нахождения истинного содержания никеля в указанном интервале значений составляет около 82%.

Третий тип задач — важнейший из ранее перечисленных. Это задачи, связанные с расчетом доверительного интервала. Чтобы рассчитать доверительный интервал по функциям Лапласа, имея только один результат проведенного эксперимента (величину x), надо знать характеристику воспроизводимости методики измерений (величину σ или равносильные ей σ_r или $\sigma_{\text{ср}}$). Построим симметричный интервал $x \pm \Delta x$, куда в отсутствие систематических погрешностей с вероятностью P должно попасть значение μ . Величину P называют доверительной вероятностью. Значение Δx — полуширина доверительного интервала, она равно $U\sigma$, где U — аргумент функции Лапласа, равной $P/2$. Иногда полуширину доверительного интервала называют «точностью», хотя использовать термин в данном смысле не рекомендуется. С вероятностью P можно утверждать, что

$$(x - U\sigma) < \mu < (x + U\sigma). \quad (5.9)$$

В расчетах доверительных интервалов обычно используются значения P , равные 0,90; 0,95; 0,99; 0,999. Им соответствуют значения U , равные 1,65; 1,96; 2,58; 3,29. Неизвестное μ с вероятностью 0,90 попадает в интервал от $U = -1,65$ до $U = +1,65$, чему отвечает интервал $x \pm 1,65\sigma$. Наиболее часто пользуются двухсигмовым интервалом, в него истинное значение μ попадает с вероятностью, близкой к 95%. Если доверительный интервал откладывается не относительно результата единичного эксперимента, а относительно среднего арифметического из n вариантов, полуширина доверительного интервала сокращается в \sqrt{n} раз, и с вероятностью P можно утверждать, что

$$(\bar{x} - U\sigma/\sqrt{n}) < \mu < (\bar{x} + U\sigma/\sqrt{n}). \quad (5.10)$$

В практике обработки результатов эксперимента широко известно «правило трех сигм»: отклонение истинного значения измеряемой величины от среднего арифметического из n результатов измерений не превосходит трех стандартных отклонений среднего, т. е. $3\sigma/\sqrt{n}$. Надежность этого утверждения — 0,997.

Пример 5-4. В девяти параллельных опытах некоторое соединение синтезировано со средним выходом, равным 40%. Известно, что методика синтеза характеризуется относительным стандартным отклонением 0,20. Определить выход соединения в виде доверительного интервала ($P = 0,99$).

Решение. Для $P = 0,99$ $U = 2,58$. Величину σ находим, как в предыдущем примере, она равна $0,20 \cdot 40\% = 8\%$. Полуширина доверительного интервала с учетом числа параллельных опытов равна $\Delta x = U\sigma/\sqrt{n} = 2,58 \times 8\%/3 = 7\%$. С вероятностью 0,99 можно считать, что выход соединения — от 33 до 47%.

Часто требуется решить обратную задачу — найти n — число измерений, которые надо провести по методике с известным значением σ , чтобы средний результат с вероятностью P и с точностью до Δx совпадал бы с истинным значением измеряемой величины. Можно также рассчитывать, подходит ли вообще данная методика, если число n задано, или следует перейти к более воспроизводимой,

т. е. такой, у которой значение σ меньше. Из (5.10) следует, что

$$n = U^2 \sigma^2 / (\Delta x)^2. \quad (5.11)$$

Из (5.11) можно найти одну из четырех величин (Δx , U , σ , n) по трем остальным.

Пример 5-5. Необходимо установить прочность резины на разрыв, учитывая результаты параллельных опытов. Сколько их нужно провести, если прочность резины хотят найти с точностью до пяти единиц, а методика единичного измерения характеризуется значением $\sigma = 10$ единиц. Сделать расчет для $P = 0,95$.

Решение. В данном случае $U \approx 2$. Тогда подстановка в (5.11) дает $n = 2^2 \cdot 10^2 / 5^2 = 16$. Значит, нужно провести не менее 16 повторных измерений.

Если воспроизводимость метода известна, а распределение вариант заведомо нормальное, то по экспериментальным данным легко выявить грубые промахи и систематические ошибки. Так, в серии вариант отклонение одной из них от среднего арифметического, превышающее 3σ , нельзя объяснять случайным разбросом вариант, так как вероятность такого большого случайного отклонения при нормальном распределении составляет всего 0,003. Значит, соответствующий результат следует считать грубым промахом и отбраковать. Если промахов нет, а тем не менее средний результат из n проведенных повторных измерений сильно отличается от истинного значения Q , можно предположить наличие систематической погрешности. Такой вывод делают, если средний результат проведенных измерений отличается от эталонного значения Q более, чем на $U\sigma/\sqrt{n}$, т. е. Q не входит в доверительный интервал.

Пример 5-6. Эталонный образец резины с прочностью на разрыв, равной 200 единиц, исследовали, пользуясь методикой, для которой $\sigma = 10$ единиц. Провели четыре испытания, причем были получены значения прочности, равные 205, 207, 208 и 204 единиц. Имеет ли методика систематическую ошибку?

Решение. Используем доверительную вероятность $P = 0,95$. Тогда $U \approx 2$. Рассчитаем полуширину довери-

тельного интервала $\Delta x = 2 \cdot 10 / \sqrt{4} = 10$. Среднее арифметическое из четырех проведенных измерений равно 206 единиц, оно отличается от Q только на шесть единиц, т. е. Q находится внутри доверительного интервала. Значит, систематическая погрешность измерений не выявлена, во всяком случае, она не может считаться доказанной с 95%-ной надежностью.

Глава 6

МАТЕМАТИЧЕСКАЯ ОБРАБОТКА МАЛЫХ ВЫБОРОК

6.1.

ПРЕДМЕТ МИКРОСТАТИСТИКИ

Зачастую химик, измеряющий некоторую величину (содержание компонента в пробе, выход продукта, скорость реакции, степень извлечения и т. д.), не знает ни истинного значения измеряемой величины, ни метрологических характеристик методики измерений, ни характера распределения результатов. Единственное, чем он располагает, это небольшая выборка, содержащая результаты параллельных измерений. Говоря языком статистики, задача исследователя — установить параметры генеральной совокупности (в частности, μ и σ) по параметрам небольшой выборки, т. е. по значениям \bar{x} и S . Разумеется, при правильном выборе методики измерений, т. е. в отсутствие грубых промахов и систематических ошибок, мы вправе считать, что $\bar{x} \rightarrow \mu$, $S \rightarrow \sigma$. Однако в выборку случайно могут попасть такие варианты, что вычисленное по ним среднее арифметическое будет довольно сильно отличаться от μ , а S от σ . Дополнительная погрешность, вносимая экспериментатором при оценке характеристик генеральной совокупности с помощью выборочных параметров, тем больше, чем меньше объем выборки.

Раздел математической статистики, посвященный обработке и сопоставлению малых выборок ($2 \leq n \leq 20$),

обычно называют микростатистикой. Пользуясь методами микростатистики, исследователь с заданной надежностью оценивает действительное значение измеряемой величины, представляя его в виде доверительного интервала. Сделать это можно разными способами. Так, если распределение x в генеральной совокупности предполагается нормальным, и неизвестна только величина σ , доверительные интервалы рассчитывают по формулам Стьюдента (см. раздел 6.2). Если характер распределения x в точности не известен, но известно, что это распределение симметрично, а систематические погрешности отсутствуют, для расчета доверительных интервалов применяют неравенство Чебышева в его «усиленной» форме. Если же о характере распределения вообще ничего не известно, пользуются неравенством Чебышева в канонической форме. Чем меньше при прочих равных условиях априорной информации мы имеем об измеряемой величине, тем более широкими должны быть доверительные интервалы. В частности, расчет, основанный на неравенстве Чебышева, дает значительно более широкий доверительный интервал, чем расчет по формулам Стьюдента. Последний интервал, в свою очередь, шире, чем интервал, рассчитанный по формулам нормального распределения. Химики обычно ведут расчет доверительных интервалов по методу Стьюдента, принимая (с проверкой или без нее) постулат о нормальном распределении измеряемой величины и равноценности (равноточности) всех измерений. Неравенство Чебышева и его использование для расчета доверительных интервалов описаны в специальной литературе. С ним можно познакомиться самостоятельно, например, по книгам А. К. Чарыкова [22].

Вторая задача микростатистики — расчет минимального объема выборки, необходимой для оценки μ с заданной точностью либо для надежного подтверждения (или опровержения) выдвинутой гипотезы. Вопрос о количестве параллельных опытов для экспериментатора имеет первостепенную важность.

Третья задача — проверка предположений о свойствах генеральных совокупностей, сделанных на основе ограниченных экспериментальных данных (проверка статистических гипотез). Значение проверки тем более важно, чем

меньше объем данных, на основании которых делались выводы. Для проверки гипотез созданы многочисленные алгоритмы и статистические критерии, большинство их базируется на постулате о нормальном распределении x . Алгоритмы и критерии проверки гипотез будут рассмотрены в следующей главе.

6.2. РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Если в генеральной совокупности случайная величина распределена по нормальному закону с неизвестными параметрами μ и σ , то их оценивают по малой выборке, используя распределение Стьюдента или, как его еще называют, t -распределение. Соответствующие математические формулы гораздо сложнее формул нормального распределения и ввиду своей громоздкости здесь не приводятся. Они связывают выборочное стандартное отклонение S , объем выборки n , полуширину доверительного интервала Δx и (в неявном виде) доверительную вероятность P . Формулы несколько упрощаются при переходе к безразмерной переменной t , аналогичной переменной U в нормальном распределении: $t = (\bar{x} - \mu)/S$. Случайная величина t — это мера отклонения среднего результата измеряемой величины от ее математического ожидания, оцененная в единицах выборочного стандартного отклонения.

При любом n кривая $y = \varphi(t)$ имеет вид плавной и симметричной кривой с максимумом при $t = 0$. В целом такая кривая, называемая кривой Стьюдента, похожа на кривую Гаусса, только идет более полого. Чем меньше n , тем ниже и «шире» фигура под кривой Стьюдента (рис. 11). Вероятность попадания x в интервал (a, b) соответствует площади под кривой Стьюдента, ограниченной абсциссами a и b , а после перехода к переменной t — абсциссами t_1 и t_2 .

$$P(a < x < b) = P(t_1 < t < t_2) = \int_a^b \phi(x) dx = \int_{t_1}^{t_2} \varphi(t) dt. \quad (6.1)$$

Здесь наблюдается полная аналогия с нормальным распределением. Но величина P зависит и от объема выборки.

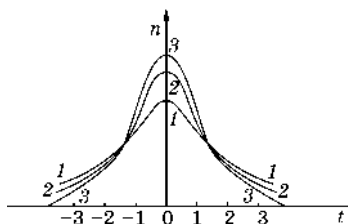


Рис. 11

Графическое представление распределения Стьюдента:

1 — при $n = 5$; 2 — при $n = 10$; 3 — при $n = \infty$ (гауссиана).

Так, при $n = 3$ в интервал $\mu \pm 2S$ (от $t = -2$ до $t = 2$) результат измерения попадет с вероятностью $P = 82\%$, при $n = 10$ эта вероятность равна $0,93$, и лишь при $n \rightarrow \infty$ $P \rightarrow 95\%$.

Отсюда очевидно, что при уменьшении n для сохранения той же вероятности (например, $P = 0,95$) границы симметричного доверительного интервала $(-t, +t)$ должны раздвигаться. При $n = 10$ $t = 2,26$; при $n = 5$ $t = 4,30$ и т. д. (рис. 12). Указанные выше значения t для заданной величины n и заданной доверительной вероятности P называют коэффициентами (или критериями) Стьюдента. Их можно вычислить, но ввиду сложности соответствующих формул обычно пользуются таблицами (см. приложение 2Е). Числовые значения коэффициента Стьюдента в справочниках приводятся как функции от числа степеней свободы (обозначаются df) и от уровня значимости (α). Число степеней свободы при расчете доверительных интервалов на единицу меньше объема выборки: $df = n - 1$. В других случаях связь между df и n иная. Например, при сопоставлении средних значений из двух выборок $df = n - 2$. Уровень

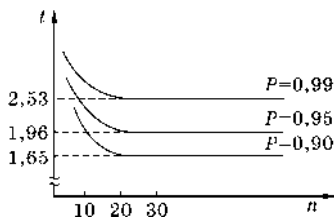


Рис. 12

Значения коэффициента Стьюдента при разных P и n

значимости дополняет доверительную вероятность P до единицы. Так, вероятности 0,95 соответствует $\alpha = 0,05$ (иногда говорят о «пятипроцентном уровне значимости»).

Если требуется более высокая доверительная вероятность (0,99 или 0,999), то при тех же n ширина доверительного интервала должна вырасти, увеличатся и значения t . Если $P = 0,99$, кривая $t = t(P, n)$ пройдет выше, чем в случае, когда $P = 0,95$. Сопоставление таблиц Стьюдента и таблиц функций Лапласа показывает, что для одинаковых значений P величина t больше, чем параметр U . Однако при $n \rightarrow \infty$ $t \rightarrow U$. Нормальное распределение — это тот предел, к которому стремится распределение Стьюдента по мере увеличения n .

6.3. ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ t-РАСПРЕДЕЛЕНИЯ

1. Допустим, по некоторой методике измеряется величина с известным значением Q (эталон). Систематические погрешности отсутствуют ($\mu = Q$), значение σ для данной методики измерений неизвестно, величина S оценена из предыдущих измерений. Тогда можно рассчитать границы интервала $\mu \pm tS$ для некоторой доверительной вероятности P , например, 0,95 или 0,99. На основании свойств t -распределения можно считать, что почти все результаты будущих измерений должны входить в этот интервал. Случайный выход результата измерения за эти границы хотя и возможен, но крайне маловероятен (вероятность такого события равна $1 - P$). На практике результат измерения, выходящий за границы указанного интервала, расценивают как грубый промах.

2. Среднее арифметическое выборки варьирует относительно μ в \sqrt{n} раз меньше, чем отдельные варианты (при нормальном распределении вариантов). Поэтому с надежностью P можно считать, что в отсутствие систематических ошибок среднее арифметическое такой выборки должно принадлежать интервалу $\mu \pm ts/\sqrt{n}$, где t — значение коэффициента Стьюдента для данных P и n .

3. Важнейшим применением t -распределения является нахождение границ доверительных интервалов. Так как варьирование неизвестного μ относительно x — результата единичного измерения — характеризуется теми же параметрами, что и варьирование x относительно μ , то доверительные интервалы можно строить вокруг единичного результата. С надежностью P можно считать, что неизвестное значение μ принадлежит интервалу $(x \pm tS)$, где t зависит от P и n .

4. Гораздо чаще доверительные интервалы строят вокруг среднего арифметического. В этом случае с надежностью P можно считать, что

$$(\bar{x} - tS/\sqrt{n}) < \mu < (\bar{x} + tS/\sqrt{n}). \quad (6.2)$$

Примером может быть выборка из пяти вариант 15, 21, 18, 20, 16. В этом случае $\bar{x} = 18$, $S = 2,45$. По таблицам Стьюдента находим: $t = 2,78$ (для $P = 0,95$ и $n = 5$). Отсюда $\Delta x = 3,1$, а округляя до целочисленных значений в соответствии с представлением исходных данных, можно считать, что $\Delta x \approx 3$. Истинное значение измеряемой величины с 95%-ной надежностью попадает в доверительный интервал от 15 до 21.

Отметим, что в данном примере, как и во многих других случаях, стандартное отклонение результата измерения — величина того же порядка, что и ошибка округления. Это существенно снижает надежность расчетов по алгоритмам микростатистики. Однако альтернативных (причем простых) алгоритмов обработки подобных выборок, к сожалению, не существует.

Стоит еще раз напомнить, что общеизвестная формула (6.2) справедлива только при выполнении трех условий: а) методика измерений не дает систематических погрешностей; б) результаты измерений в генеральной совокупности подчиняются нормальному распределению, независимы друг от друга и равноточны; в) отсутствуют дрейф и грубые промахи.

Определив границы доверительного интервала по небольшой выборке, мы можем предвидеть и степень рассеяния вариант в будущих сериях измерений той же величины. Так, при повторении опытов, описанных в разоб-

ранном выше примере, мы можем гарантировать, что 95% будущих результатов измерения должны попасть в интервал $\bar{x} \pm 2,78S$, т. е. $(18 \pm 6,8)$. Тогда варианты, меньшие 11 или большие 25, можно будет считать грубыми промахами. А любая серия из пяти будущих измерений должна дать среднее арифметическое, которое с 95%-ной вероятностью окажется внутри ранее вычисленного доверительного интервала, т. е. между 15 и 21.

При разработке новой методики анализа, изготовлении нового прибора и в других случаях необходимо заранее оценить величину стандартного отклонения σ , для чего надо провести достаточно большое число повторных измерений, что не всегда возможно. Связь между полученным в небольшой выборке значением S и стандартным отклонением генеральной совокупности, т. е. величиной σ можно представить в виде доверительного интервала:

$$\sigma = S(1 \pm q)\sqrt{\frac{n-1}{n}}, \quad (6.3)$$

где коэффициент q определяется требуемым уровнем надежности и объемом выборки. Табличные значения q имеются в литературе (например, в руководстве Румшинского [23, с. 175]). При малых n доверительный интервал для σ является асимметричным, вычислять обе его границы надо по отдельности, пользуясь описанными в литературе алгоритмами. Рассмотрим некоторые примеры.

Пример 6-1. Активность некоторого катализатора (в условных единицах) определяли в шести параллельных опытах по одной и той же методике. Получены значения 289; 284; 286; 289; 287; 287. Определить вероятность отклонения единичного результата от математического ожидания более чем на пять единиц.

Решение. Предполагаем, что результаты распределены по нормальному закону. Обозначим активность катализатора как x и вычислим выборочные параметры. $\bar{x} = 287$, $S = 1,9$. Выражаем отклонение в безразмерных единицах: $t = 5/1,9 = 2,63$. При пяти степенях свободы такому значению коэффициента Стьюдента приблизительно соответствует $P = 0,95$. Следовательно, без учета влияния ошибки округления вероятность случайного выхода еди-

ничного значения x за границы интервала (287 ± 5) равна $1 - 0,95 = 0,05$.

Пример 6-2. Сколько измерений активности катализатора (в условных единицах) следует проводить, чтобы обеспечить с надежностью 0,95 точность среднего результата в одну единицу, если величина $S = 2$?

Решение. Точно определить значение n (необходимое число опытов) можно только при известной величине генерального стандартного отклонения. В этом случае применяется формула (5.11). В данном же случае значение σ неизвестно. Предполагаем, что результаты измерений в генеральной совокупности распределены по нормальному закону, независимы и равноточны. Приближенную оценку n можно получить по таблицам [23]. В данном случае $q = \Delta x / S = 0,5$; $P = 0,95$, откуда $n = 18$.

Если бы в условии задачи не была указана величина S , то единственным способом решения было бы провести несколько (n_1) экспериментов, вычислить эмпирическую величину S , затем рассчитать полуширину доверительного интервала для заданного значения P по формуле $\Delta x_1 = tS / \sqrt{n}$ и сравнить ее с требуемой точностью измерений Δx_2 . При больших n значения t и S слабо зависят от n , и можно считать, что

$$\Delta x_1 / \Delta x_2 \approx \sqrt{n_2} / \sqrt{n_1}. \quad (6.4)$$

Если, например, величина Δx_1 окажется втрое большей, чем требуется по условию, то для соответствующего ее уменьшения надо увеличить число опытов в 9 раз по сравнению с n_1 . Чем меньше n_1 и n_2 , тем большие ошибки дает этот способ расчета, в случае 3–5 измерений применять его вообще не следует.

Рассматривая способы математической обработки результатов измерений, следует остановиться еще на одной непростой проблеме. Оформляя отчеты, статьи или диссертации, описывая методики измерений и расчетов, исследователь не имеет права использовать любые термины, руководствуясь лишь личными вкусами и ранее сформированными привычками. Так, одни предпочитают термин «стандартное отклонение», другие — «среднеквадратиче-

ское отклонение» (СКО), третьи — «среднеквадратичная погрешность», хотя это синонимы. Полуширину доверительного интервала $\Delta x = tS/\sqrt{n}$ нередко называют «точностью измерений» или «вероятной погрешностью». В настоящее время использование термина «вероятная погрешность» не рекомендуется, термин «полуширина доверительного интервала» гораздо более понятен и четко определен.

Далеко не все химики знают, что использование тех или иных терминов, связанных с измерением физических величин и статистической обработкой экспериментальных данных, регламентируется официальными документами. Основным руководством в этой области для российских специалистов является ГОСТ 5725 [24], принятый в 2002 г. В этом документе введены некоторые новые термины (например, «прецизионность измерений»), а многие другие получили новые дефиниции, соответствующие рекомендациям международных организаций (эти вопросы детально обсуждаются в учебном пособии [11]). Отметим, кстати, что сами алгоритмы статистической обработки результатов измерений также регламентируются официальными документами (ГОСТ 8.736-2011 [25]). Однако этот ГОСТ обязателен для аккредитованных испытательных лабораторий (в частности, работающих в сферах материального производства и безопасности жизнедеятельности), а применительно к научным исследованиям требования ГОСТ имеют лишь рекомендательный характер.

Глава 7

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ. ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

7.1. СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

Термин «гипотеза» используется не только для обозначения предположений о составе, строении и механизме функционирования исследуемого объекта (научные гипотезы)^{*}, но и в другом смысле — *как нуждающееся в проверке математическое предположение о свойствах генеральной совокупности данных, характеризующих этот объект*. Назовем такие гипотезы статистическими. Так, на основании измерений скорости реакции при разных температурах можно сформулировать гипотезу об экспоненциальном характере этой зависимости. Такая гипотеза будет относиться к генеральной совокупности данных о кинетике данной реакции — как уже полученных данных, так и тех, что можно будет получить в будущем.

Наиболее часто выдвигают статистические гипотезы о функции распределения некоторой генеральной совокупности данных (результатов измерений). В этих гипотезах речь идет о характере распределения вариантов, о математическом ожидании соответствующей случайной величины, о дисперсии генеральной совокупности. Гипотезы выдвигают

^{*} О научных гипотезах см. раздел 1.2.

гают на основании немногих экспериментальных данных (одной или нескольких малых выборок), а потому они нуждаются в проверке методами математической статистики.

Статистические гипотезы часто формулируют и проверяют для подтверждения или опровержения научных гипотез. Например, исследователь предполагает, что реакция пероксидного окисления некоторого органического вещества в растворе протекает по свободнорадикальному механизму. Если эта научная гипотеза верна, то скорость данной реакции должна снижаться в присутствии акцепторов свободных радикалов; хорошо известно, что они всегда замедляют реакции такого типа. Для проверки исследователь проводит «решающий эксперимент»: он тщательно измеряет скорость реакции в присутствии акцептора и без него. При повторении опытов в одинаковых условиях получаются две выборки результатов измерений. Допустим, средние значения скорости реакции в этих выборках оказались различны, в присутствии акцептора средняя скорость была несколько ниже. Но вывод из проведенного эксперимента недостоверен ввиду небольшого объема сопоставляемых выборок. Не исключено, что различие средних значений объясняется случайными погрешностями при измерении скорости реакции, а проверяемый фактор (акцептор радикалов) вовсе не влияет.

Для проверки исследователь начинает рассматривать нуль-гипотезу, обратную его гипотезе о достоверном влиянии фактора (акцептора радикалов). А именно, он проверяет значимость предположения о том, что при достаточно большом объеме выборок средняя скорость реакции в присутствии акцептора и без него будет одинаковой. Эту же статистическую гипотезу можно сформулировать более строго: «математические ожидания μ_1 и μ_2 , получаемые из сопоставляемых выборок, равны между собой, они характеризуют одну и ту же генеральную совокупность». Такую статистическую гипотезу можно проверить расчетом, используя один из известных алгоритмов, например, алгоритм сравнения средних по Стьюденту. Если в результате расчета окажется, что значимость нуль-гипотезы (α) меньше некоторого заранее выбранного критического уровня, например, менее 0,01, ее отвергают и принимают противоположное предположение (альтернативу) с заранее ого-

воренной формулировкой, например «математические ожидания μ_1 и μ_2 не равны». Последнее утверждение означает, что исследуемый фактор достоверно влияет на результаты измерений.

Результаты проверки статистических гипотез всегда имеют вероятностный характер, и выводы из проведенного исследования никогда не должны звучать безапелляционно, категорически. Например, правильным будет такой вывод: «На основании проведенного эксперимента нами установлено, что скорость реакции снижается в присутствии акцептора свободных радикалов (значимость нуль-гипотезы меньше 0,01). Этот факт согласуется с ранее выдвинутой гипотезой о радикальном механизме данной реакции». И совершенно некорректным по сути и неприемлемым по тону является другое утверждение: «Изучив влияние акцепторов свободных радикалов на скорость реакции, мы окончательно доказали свободнорадикальный механизм этой реакции».

7.2.

ОБЩИЙ АЛГОРИТМ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Как правило, при проверке гипотез сопоставляют две (или более) однотипных совокупности числовых данных. Это могут быть, например: а) результаты измерений, выполненных независимо друг от друга на двух разных приборах или по двум разным методикам, или в двух разных лабораториях; б) экспериментальные и табличные данные (или результаты некоторого расчета); в) результаты измерения свойств исследуемого объекта до и после воздействия на него. В любом случае исследователь исходит из результатов многократного измерения некоторой физической величины. Важно, чтобы варианты обеих выборок имели одинаковую размерность и одинаковый физический смысл. Объемы сопоставляемых выборок могут быть и различными. Поскольку число вариантов в сопоставляемых выборках обычно невелико, то для проверки применяют методы микростатистики. Априорная информация о характере распределения вариантов (нормальное распределение, пуассоновское распределение и т. п.) значительно облегчает проверку.

Рассмотрим примеры задач, приводящих к необходимости проверки статистических гипотез. Подобные задачи можно найти в любой области химии.

1. Скорость коррозии одного и того же образца определяли гравиметрическим методом и методом поляризационного сопротивления. Получили необходимые числовые данные (две выборки результатов). Требуется выбрать метод, дающий более воспроизводимые результаты.

2. Лекарственный препарат должен иметь определенную брутто-формулу, из которой можно рассчитать «теоретическое» содержание каждого элемента. Однако реальный препарат оказался имеющим — по данным многократно проведенного элементного анализа — несколько иное содержание тех же элементов. Считать ли несовпадение статистически незначимым или считать препарат, имеющим достоверно другой состав, т. е. фальсифицированным?

3. Определяли температуру застывания смазочного масла с добавкой (присадкой) и без нее. Средний результат с присадкой оказался чуть ниже. Считать ли присадку достоверно влияющей на свойства масла или отнести изменение средней температуры застывания на счет случайных погрешностей измерения?

Наиболее важны задачи третьего типа (проверка значимости влияния фактора на параметр оптимизации). Изменение свойств системы после воздействия на нее мы обычно принимаем за результат этого воздействия*. Но такое изменение может быть связано с одновременным влиянием других (неучтенных исследователем) факторов, либо со случайными погрешностями измерения и пробоотбора. Такую возможность обязательно следует проверить и исключить, прежде чем детально исследовать влияние фактора.

Статистические гипотезы можно разделить на *исходную* и *альтернативные* ей. Любая проверка начинается с выбора исходной гипотезы. Это может быть утверждение о

* Это не всегда так, еще древние римляне учили не путать *post hoc* (после того, как...) и *propter hoc* (вследствие того, что...). Пример: лето наступило после окончания учебного года, но не вследствие этого события!

достоверном различии сопоставляемых выборок, а может быть, наоборот, утверждение о том, что они являются выборками из одной и той же генеральной совокупности. Рекомендуется принимать в качестве исходной (или, как ее еще называют, нуль-гипотезы) ту, ошибочное принятие которой приведет к более опасным последствиям (например, большим убыткам). В химии в качестве нуль-гипотезы обычно принимают предположение об отсутствии различий. То есть считают сопоставляемые выборки, принадлежащими к одной генеральной совокупности, а наблюдаемые расхождения между ними — чистой случайностью, которая не будет повторяться в последующих экспериментах. При изучении влияния фактора нуль-гипотезе соответствует утверждение: «фактор не влияет». Аналогично в других случаях: «линейной корреляции на самом деле нет», «дрейф отсутствует», «воспроизводимость всех опытов одинакова», «математические ожидания различаются несущественно» и т. д. Именно такие нуль-гипотезы, заставляющие вспомнить название знаменитой книги Р. Бойля «Химик-скептик», мы и будем в дальнейшем проверять.

Противоположное нуль-гипотезе утверждение называют альтернативой. Примеры альтернатив: «математические ожидания достоверно отличаются», «здесь есть грубый промах», «имеется статистически значимая корреляция», «фактор достоверно влияет». Иногда одной нуль-гипотезе могут соответствовать несколько разных альтернатив. Поэтому при проверке нуль-гипотезы следует заранее точно определить, какой будет ее альтернатива.

Алгоритм проверки статистических гипотез включает следующие стадии.

1. По экспериментальным данным рассчитывают некоторый показатель $\lambda_{\text{эксп}}$. При этом используют всю совокупность данных, представленных в сопоставляемых выборках. Способ расчета каждого показателя специфичен. Расчетные формулы приводятся в руководствах по математической статистике. Для подавляющего большинства показателей — и параметрических, и непараметрических — действует правило: *чем выше λ , тем достовернее различаются сопоставляемые выборки и, соответственно, тем меньше вероятность гипотезы о случайном раз-*

личии выборок, об их принадлежности к одной и той же генеральной совокупности.

2. По соответствующим статистическим таблицам находят $\lambda_{\text{крит}}$ — максимальное значение того же показателя, которое при том же числе степеней свободы могло бы наблюдаться между выборками из одной генеральной совокупности (однородной). В таких таблицах критические значения показателя указаны для разных уровней значимости нуль-гипотезы (α): 0,10; 0,05; 0,01; ... Как правило, чем меньше α , тем $\lambda_{\text{крит}}$ больше. Выписывают $\lambda_{\text{крит}}$ для выбранного уровня значимости α или, лучше, для двух уровней α_1 и α_2 , где $\alpha_1 > \alpha_2$, а следовательно, $\lambda_1 < \lambda_2$. Обычно берут $\alpha_1 = 0,05$ и $\alpha_2 = 0,01$.

3. Сопоставляют $\lambda_{\text{эксп}}$ с критическими значениями показателя при α_1 и α_2 . Обозначим их как λ_1 и λ_2 . Могут наблюдаться три случая:

1) $\lambda_{\text{эксп}} > \lambda_2$. Высокое значение показателя свидетельствует о попадании в область, где значимость нуль-гипотезы меньше α_2 (например, $< 0,01$). Значит, ее можно отбросить и принять альтернативу. Различие между выборками считается доказанным;

б) $\lambda_{\text{эксп}}$ лежит между λ_1 и λ_2 . Данных для надежного вывода не хватает. Надо продолжить опыты, увеличить объем выборок, а затем повторить расчет;

в) $\lambda_{\text{эксп}} < \lambda_1$. Низкое значение показателя говорит о том, что нуль-гипотеза имеет значимость, большую, чем α_1 (например, $\alpha > 0,05$). Значит, мы не имеем права отбрасывать нуль-гипотезу. Хотя ее значимость может быть невысокой (0,15 или 0,20), нуль-гипотезу следует принять. Обратите внимание, в подобных случаях экспериментатор не должен утверждать, что между выборками не существует различий, что они тождественны (для доказательства тождественности с самого начала нужно было исходить из совершенно другой нуль-гипотезы). Можно лишь утверждать, что проведенный эксперимент не выявил статистически достоверных различий между выборками. Возможно, в дальнейших исследованиях такие различия обнаружатся, но пока что нуль-гипотеза сохраняется.

Многие химики и биологи считают, что при $\alpha < \alpha_2$ альтернативу следует считать доказанной с надежностью $P \geq 1 - \alpha_2$ (например [26, с. 106]). Это утверждение логично,

но со строго математической точки зрения оно некорректно. Статистические выводы, основанные на выражении $P + \alpha = 1$, корректны лишь при использовании интервальных оценок (например, при определении вероятности попадания некоторой случайной величины в границы заданного интервала, см. раздел 5.3). С другой стороны, оценка значимости альтернативы при проверке статистических гипотез вообще не является необходимой, вполне достаточно указать уровень значимости нуль-гипотезы (например, записать: $\alpha < 0,01$ или $\alpha > 0,05$). Отметим, что многие современные компьютерные программы обработки данных не сообщают пользователю числового значения показателя $\lambda_{\text{эсп}}$, а просто рассчитывают числовую величину α .

Изложенный выше общий алгоритм проверки статистических гипотез реализуется при использовании самых разных показателей. Некоторые показатели применимы только в тех случаях, когда характер распределения случайной величины в генеральной совокупности соответствующих данных заранее известен, совпадает с каким-либо теоретическим распределением (см. раздел 5.3). Соответствующие показатели называют *параметрическими критериями*. При расчете критических значений параметрических показателей учитывается характер теоретического распределения случайной величины, в частности — свойства нормального распределения. Так, если распределение экспериментальных данных подчиняется закону Гаусса, для проверки статистических гипотез используют такие параметрические критерии, как критерий Стьюдента (t), критерий Фишера (F), Q -тест Диксона, коэффициенты асимметрии, эксцесса, линейной корреляции и др. Эти параметрические критерии, хорошо известные химикам, рассматриваются в разделе 7.4.

Другие показатели (*непараметрические критерии*) не связаны с допущением о характере распределения генеральной совокупности экспериментальных данных. Параметры теоретического распределения при вычислении соответствующих показателей не применяются. Непараметрические критерии применимы для проверки статистических гипотез даже в тех случаях, когда исследователь не знает, по какому закону распределены результаты

проведенных им измерений. Это, в частности, критерии Граббса, Вилкоксона, Фридмана, критерий знаков и некоторые другие. Непараметрические критерии более сложны в математическом отношении, поэтому химики используют их довольно редко.

К сожалению, иногда начинающие исследователи не знают о существовании таких критериев, поскольку в некоторых руководствах по обработке экспериментальных данных они даже не упоминаются. Поэтому весьма часто исследователь-химик проводит обработку полученных данных с применением традиционных параметрических критериев, хотя не знает характера распределения измеряемых им величин и не имеет оснований считать его нормальным. Например, сравнивает средние значения выборок по Стьюденту или отбраковывает промахи по Q -тесту. А ведь при обработке выборок малого объема из «ненормально распределенных» генеральных совокупностей такой подход зачастую приводит к ошибочным выводам из проведенного исследования. Громадный труд экспериментатора пропадает впустую или, что еще хуже, приводит к неправильным технологическим рекомендациям и проектным решениям. Очевидно, начинающий исследователь, независимо от его специализации, должен освоить не только параметрические, но и непараметрические методы обработки данных. В настоящем пособии непараметрическим критериям посвящена глава 11.

7.3. ВОЗМОЖНЫЕ ОШИБКИ ПРИ ПРОВЕРКЕ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Решение, которое мы принимаем по описанному выше алгоритму, может быть ошибочным в двух случаях.

1. Если отбросим верную нуль-гипотезу и примем альтернативу. Это «ошибка первого рода». Радиотехники ее называют «ложным сигналом», а технологи — «риском производителя». Имеется в виду, что при такой ошибке технического контроля годная продукция отправится в брак, производитель понесет убыток, а потребитель не страдает. В химических исследованиях ошибка первого

рода приведет к категоричным (хотя и ошибочным) выводам: «обнаружено, что присадка достоверно влияет на свойства смазочного масла», «проверенное лекарство фальсифицировано», и т. п. Дальнейшей статистической проверкой подобных выводов уже не занимаются. Это делает ошибку первого рода особенно опасной. Напомним, что вывод о достоверном различии выборок делают при $\lambda_{\text{экс}} > \lambda_2$. Чтобы снизить вероятность появления таких ошибок, увеличивают критическое значение λ_2 , а для этого критический уровень значимости α_2 берут равным 0,01 или даже 0,001, в зависимости от тяжести возможных последствий ошибки первого рода.

2. Мы также совершим ошибку, если примем неверную нуль-гипотезу. Это ошибка второго рода, «пропуск сигнала» или «риск потребителя» (при соответствующей ошибке технического контроля брак выявлен не будет и потребитель получит негодную продукцию). Но надо учесть, что в корректно проведенных химических исследованиях при $\lambda_{\text{экс}} > \lambda_1$ выводы не имеют категорического характера («различие выборок не выявлено», «влияние некоторого фактора не доказано» и т. п.). Проверку обычно продолжают, используя другие способы и критерии. Поэтому ошибки второго рода в химических исследованиях обычно не столь опасны, как ошибки первого рода.

Во многих случаях вероятность обеих ошибок при выбранных значениях λ и $\lambda_{\text{крит}}$ можно оценить заранее. Однако принципы и способы таких расчетов специфичны для каждого исследования, а поэтому рассматриваться в настоящем пособии они не будут. Отметим только, что в любом случае вероятности обеих ошибок взаимосвязаны. А именно, чем меньше вероятность ошибок первого рода (α), тем больше вероятность ошибок второго рода (β) и наоборот, а соотношение между ними определяется выбором величины $\lambda_{\text{крит}}$.

В качестве иллюстрации рассмотрим идентификацию вещества в пробе сложного состава по многочисленным неспецифическим и равнозначным признакам [27]. Вещество признается присутствующим в пробе (достоверно идентифицированным), если N — число наблюдаемых совпадений в свойствах пробы и эталона (чистого вещества известной структуры) — превысит установленный кри-

терий $N_{кр}$, т. е. при $N > N_{кр}$. Так, некоторый элемент считают опознанным, если несколько линий в спектре пробы совпадут по длине волны (с точностью до погрешности измерений) с линиями эталонного спектра данного элемента. Часто считают, что достаточно даже трех совпадений; тогда условие идентификации $N > N_{кр}$ конкретизируется как $N > 2$. Однако всегда возможно случайное совпадение каких-то линий в спектре пробы с линиями элементов, отсутствующих в пробе; следовательно, возможны ошибки первого рода (ложные идентификации). Одновременное случайное совпадение 5, а тем более 10 или 20 линий в спектре пробы с линиями эталонного спектра элемента маловероятно. Вероятность ошибки первого рода будет тем меньше, чем больше линий использовано для опознания и чем более жесткий критерий $N_{кр}$ используется в методике (рис. 13, кривая α). Таким образом, для предотвращения случайных ложных идентификаций нужно увеличивать $N_{кр}$.

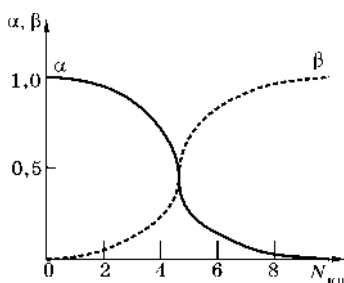


Рис. 13

Взаимосвязь вероятностей ошибок первого и второго рода при идентификации веществ по числу совпадающих спектральных линий [27]

С другой стороны, у элементов, присутствующих в пробе, некоторые линии могут не проявиться в спектре пробы. Число оставшихся линий может оказаться меньше $N_{кр}$, и присутствующий элемент не будет опознан. По мере ужесточения критерия $N_{кр}$ вероятность такого события (ошибки второго рода) растет, что показывает кривая β на рисунке 13. Следовательно, из-за возможности ошибок первого рода недопустимы слишком низкие зна-

чения $N_{кр}$, а из-за ошибок второго рода неприемлемы слишком высокие значения $N_{кр}$. Надо подобрать оптимальное значение $N_{кр}$, при котором вероятности обеих ошибок одновременно окажутся достаточно малыми, и ими можно будет пренебречь. В каждом случае оптимальное значение $N_{кр}$ может быть вычислено методами теории вероятностей. Оптимальное значение критического числа совпадений зависит от условий регистрации спектра и неодинаково для разных элементов.

Этот пример показывает, что рассчитать оптимальные значения статистических критериев для проверки гипотез и составить таблицы критических значений показателя довольно трудно. Применение же таблиц, содержащих готовые значения критериев, не вызывает затруднений, их с успехом применяет даже те исследователи, которые имеют минимальную математическую подготовку.

7.4. ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ И ИХ ПРИМЕНЕНИЕ

Рассмотрим важнейшие и широко используемые в химических исследованиях параметрические критерии, предполагающие *нормальное распределение* вариант в соответствующих генеральных совокупностях.

7.4.1. Отбраковка грубых промахов с помощью Q-теста

Любая обработка экспериментальных данных начинается с отбраковки грубых промахов. Самый простой и наименее строгий способ отбраковки — применение Q-теста, предложенного Диксоном. В данном случае нуль-гипотеза — предположение о том, что все варианты выборки относятся к одной и той же генеральной совокупности, распределенной по нормальному закону, а грубых промахов среди полученных вариант нет.

Для проверки нуль-гипотезы варианты выборки располагают в ряд по их величине — от наименьшего значе-

ния x_1 до наибольшего x_n . По экспериментальным данным рассчитывают показатель $Q_{\text{эксп}}$ (далее Q):

$$Q = \left| \frac{x_n - x_{n-1}}{x_n - x_1} \right| \quad \text{или} \quad Q = \left| \frac{x_2 - x_1}{x_n - x_1} \right|. \quad (7.1)$$

В (7.1) первая формула применяется, если сомнения вызывает наибольшая варианта, вторая — если наименьшая. В обоих случаях Q — дробь, где в числителе разность между сомнительным результатом и ближайшим к нему в ранжированном ряду вариант; в знаменателе — размах выборки. Все разности в этих формулах берутся по модулю, поэтому значения Q всегда положительны и не превосходят единицы. Табличные значения $Q_{0,05}$, $Q_{0,01}$ и т. д. есть в справочной литературе. В настоящем пособии они приведены в приложении 2F. Обычно Q -тест применяют при наличии в выборке 3–7 вариант. При большем объеме выборки для расчета Q применяет несколько более сложные формулы [15].

Пример 7-1. По результатам анализа влажность сырья в параллельных пробах оказалась (в %): 22; 25; 28; 21; 24. Сомнение вызывает наибольшая варианта 28%. $Q = (28 - 25)/(28 - 21) = 3/7 = 0,43$. Для $n = 5$ $Q_{0,05} = 0,64$; $Q_{0,01} = 0,82$. Так как $Q < Q_{0,05}$, вероятность нуль-гипотезы больше 0,05, отбросить ее и считать результат 28% грубым промахом нельзя. Вывод: «промахи не выявлены» (на уровне значимости 0,05). Если в той же выборке сомнительной вариантой было бы (вместо 28) число 48, тогда Q оказалось бы равным 0,85 и превысило критическое значение $Q_{0,01}$. Тогда мы пришли бы к выводу о том, что сомнительный результат 48% в данной выборке является грубым промахом. Уровень значимости нуль-гипотезы в этом случае был бы менее 0,01.

Отметим, что по Q -тесту в принципе нельзя доказать отсутствия промахов, можно лишь выявить (или не выявить!) имеющиеся. Если в выборке одновременно присутствует несколько промахов, обнаружить их по Q -тесту часто не удастся. В литературе описаны альтернативные Q -тесту способы отбраковки грубых промахов, в которых

используются все варианты выборки, а не три из них, как в рассмотренном методе. Эти способы несколько сложнее, но более надежны. Пример — параметрический критерий максимального относительного отклонения (см. приложение 2Г). Для решения той же задачи можно использовать и непараметрические критерии.

7.4.2.

Сравнение двух дисперсий по критерию Фишера

Этот параметрический критерий сравнения выборок по степени разброса вариант предложен Фишером и Снедекором. Он применим, если в обеих выборках отсутствуют грубые промахи и дрейф. Для расчета F находят дисперсии выборок и делят большую на меньшую, независимо от их объема. Значение $F = S_1^2 / S_2^2$ сопоставляют с табличными значениями $F_{0,05}$ и $F_{0,01}$.

Пример 7-2. Анализ бронзы в лаборатории А дал содержание Si: 0,54%, 0,57, 0,56, 0,59%. В лаборатории Б получили соответственно 0,76%, 0,70 и 0,82%. Где воспроизводимость анализа выше?

Решение. Прежде всего убеждаются в отсутствии промахов и дрейфа, затем считают дисперсии: $S^2(A) = 4,33 \cdot 10^{-4}$; $S^2(B) = 36 \cdot 10^{-4}$. Дисперсия данных по лаборатории Б в 8,3 раз выше, ее считаем «первой», получаем $F = 8,3$. Отыскивая критические значения критерия, надо учесть не только желаемый уровень значимости нуль-гипотезы, но и объем каждой выборки. В справочниках обычно вместо объема выборки n указывается число степеней свободы df , которое в данном случае равно $(n - 1)$. При двух степенях свободы (3 опыта) для первой и трех степенях свободы для второй дисперсии $F_{0,05} = 9,5$; $F_{0,01} = 31$ (см. приложение 2Н). Так как $F < F_{0,05}$, пренебречь нуль-гипотезой нельзя. Следовательно, вывод о лучшей воспроизводимости данных лаборатории А, который мы сделали бы «на глаз», статистически недостоверен. Межлабораторное расхождение по воспроизводимости результатов анализа в данных опытах не выявлено ($\alpha > 0,05$).

7.4.3.

Сравнение нескольких дисперсий.
Критерии Бартлета и Кохрена

Часто надо сравнивать не две, а несколько выборок, содержащих однотипные варианты. Исследователь хочет знать, можно ли считать соответствующие дисперсии однородными. Задачу решают с помощью параметрических критериев Бартлета или Кохрена. Критерий Кохрена G применяют, если объемы выборок одинаковы. Он представляет собой отношение наибольшей дисперсии к сумме дисперсий для всех выборок

$$G = S_{\max}^2 / \sum S_i^2. \quad (7.2)$$

Если $G < G_{0,05}$, все выборочные дисперсии S_i^2 можно считать оценками одной и той же генеральной дисперсии σ^2 , причем $\sigma^2 \approx 1/n \sum S_i^2$. Табличные значения критерия Кохрена имеются в справочной литературе и в приложении 2Ж.

Если объемы сопоставляемых выборок неодинаковы, то применяют критерий Бартлета. Формулы, по которым он рассчитывается, довольно сложны (см. [22, с. 106]). Найденные значения сопоставляют по общему алгоритму с табличными значениями. Если экспериментальное значение критерия Бартлета не превосходит критерий, взятый для $\alpha = 0,01$, следует считать все измерения равноточными, а дисперсии — однородными.

7.4.4.

Сравнение средних значений
с помощью критерия Стьюдента

Сопоставление двух выборок по среднему значению возможно, если: а) в обеих выборках отсутствуют грубые промахи и дрейф; б) измеряемая величина имеет нормальное распределение; в) дисперсии обеих выборок однородны. Поэтому до сравнения средних значений следует проверить однородность дисперсий по критерию Фишера. Применение критерия Стьюдента для сравнения средних

значений основано на предположении о равенстве математических ожиданий, т. е. о принадлежности обеих выборок к одной и той же генеральной совокупности (см. раздел 7.1). Это предположение является нуль-гипотезой. Экспериментальное значение t -критерия находят по формуле

$$t_{\text{эксп}} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_d}, \quad (7.3)$$

где в числителе стоят средние арифметические первой и второй выборок, а в знаменателе — обобщенное стандартное отклонение

$$S_d = \sqrt{\frac{S_1^2 + S_2^2}{n}} \quad (7.4)$$

при $n_1 = n_2 = n$. При различном объеме сопоставляемых выборок ($n_1 \neq n_2$) для вычисления S_d следует применять более сложную формулу, а именно:

$$S_d = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}. \quad (7.5)$$

Величину $t_{\text{эксп}}$ сравнивают с табличными значениями t -критерия для разных уровней значимости. При отборе табличных значений учитывают объемы обеих выборок. В таблицах значения t систематизированы по числу степеней свободы. При сравнении средних число степеней свободы равно

$$df = n_1 + n_2 - 2. \quad (7.6)$$

Данный способ сводится к сравнению математических ожиданий случайной величины; в частности, до и после воздействия некоторого фактора. Различие этих величин будет достоверно ($\alpha < 0,01$), если найденное значение $t_{\text{эксп}}$ окажется большим, чем $t_{0,01}$. При $t_{\text{эксп}} \leq t_{0,05}$ различие средних считается статистически недоказанным, выборки

принято считать относящимися к одной и той же генеральной совокупности. Эти выборки можно объединять и затем найти среднее значение для объединенной выборки. Однако и в этом случае не следует утверждать, что средние значения сопоставляемых выборок были одинаковы. Если же исходные выборки были получены при разных значениях некоторого фактора, и $t_{\text{экс}} \leq t_{0,05}$, не следует утверждать, что исследуемый фактор не влияет на соответствующий параметр.

Пример 7-3. Содержание вредного вещества в сточной воде до и после очистки определяли в трех параллельных пробах. Используемая при этом методика приводит к нормальному распределению результатов анализа. Полученные результаты: в пробах неочищенной воды: 24, 17 и 19 мг/л; в пробах очищенной воды: 11, 14 и 18 мг/л. Оценить эффективность очистки.

Решение. Очевидно, мы должны проверить, снижается ли в результате очистки содержание вредного вещества. Средние значения выборок не равны: $\bar{x}_1 = 20$ мг/л, $\bar{x}_2 = 14$ мг/л. На первый взгляд, налицо заметное снижение среднего содержания — на 30%! Однакостораживает плохая воспроизводимость результатов анализа — как до, так и после очистки воды воспроизводимость характеризуется значениями S_r порядка 0,2. В таком случае различие средних значений может быть следствием случайных погрешностей анализа (в частности, на стадии пробоотбора). Находим числовые значения выборочных дисперсий $S_1^2 = 13$ и $S_2^2 = 12$. Дисперсии однородны, так как $S_1^2/S_2^2 < F_{0,05}$. Для $\alpha = 0,05$ и четырех степеней свободы $t_{0,05} = 2,78$. По приближенной формуле (7.4) находим: $S_d = 2,9$ мг/л. Отсюда $t_{\text{экс}} = (20 - 14)/2,9 \approx 2,0$, что меньше критического значения. Очевидно, эффект очистки статистически недостоверен ($\alpha > 0,05$).

Значимость нуль-гипотезы для найденного значения $t_{\text{экс}}$ можно найти, пользуясь специальными таблицами, похожими на таблицы функций Лапласа. В этих таблицах приведены значения α как функция t и n [28, с. 34–35]. Так, для вышеприведенного примера значимость нуль-гипотезы для $t_{\text{экс}} = 2,0$ и $df = 4$ составляет $\alpha \approx 0,12$. Высокая значимость нуль-гипотезы не позволяет считать влияние изучаемого фактора статистически достоверным.

7.4.5.

Проверка согласия теории и эксперимента.
Критерий Пирсона

Статистический критерий χ^2 («хи-квадрат») был предложен К. Пирсоном в 1900 г. Он принципиально отличается от рассмотренных выше критериев Фишера, Стьюдента и Диксона. Хотя критерий Пирсона является параметрическим, его применяют не только в тех случаях, когда экспериментальные данные распределены по нормальному закону. Исследователь должен лишь заранее знать (или постулировать) характер распределения случайной величины в генеральной совокупности.

Обычно критерием Пирсона пользуются для проверки характера распределения вариантов в достаточно большой по объему выборке экспериментальных данных. Если при многократных измерениях некоторой величины частота попадания результата в i -й интервал (класс) равна Θ_i , а по предполагаемому для генеральной совокупности данных теоретическому распределению частота должна равняться T_i , показатель рассчитывают по формуле

$$\chi^2 = \sum [(\Theta_i - T_i)^2 / T_i]. \quad (7.7)$$

При соответствии теории и эксперимента величина χ^2 не должна превышать некоторого порогового значения. В качестве нуль-гипотезы принимают, что различия теоретических и экспериментальных частот статистически незначимы, носят случайный характер, т. е. эксперимент действительно приводит к предполагаемому распределению случайной величины. Будем отвергать эту гипотезу, если ее значимость окажется меньше критического уровня, например 0,01. Для этого χ^2 должно быть больше $\chi_{0,01}^2$ для данного числа степеней свободы (df). Если критерий Пирсона используется для проверки нормальности некоторого эмпирического распределения, то $df = k - 3$, где k — число классов. При проверке других теоретических распределений число степеней свободы рассчитывается по формуле, учитывающей число оцениваемых параметров каждого конкретного распределения. Табличные значения χ^2 для разных уровней значимости и разного числа степеней свободы приведены в приложении 2С.

Критерием χ^2 постоянно пользуются химики-аналитики для проверки нормальности распределения результатов анализа. Так поступают при разработке новой или при аттестации известной методики анализа. Проверка нормальности распределения по критерию χ^2 (по Пирсону) надежнее, чем проверка по асимметрии и эксцессу. При использовании критерия χ^2 требуется иметь достаточно большой объем экспериментальных данных (рекомендуется, в частности, условие $n > 50$). Данные можно группировать в 6–8 классов. Желательно, чтобы классы были одинаковой ширины, однако крайние классы могут иметь другую ширину. При выборе числа классов желательно обеспечить выполнение следующего условия: каждая теоретическая частота T_i должна быть не менее пяти. Кроме того, следует убедиться в том, что сумма всех частот Σ_i равна n , а сумма всех теоретических частот T_i равна n .

В компьютерных программах, реализующих алгоритм проверки по критерию Пирсона, теоретические частоты (заселенности) вычисляются автоматически, по формулам Гаусса или Лапласа. Многочисленные примеры проверки нормальности распределения результатов анализа по критерию Пирсона приведены в литературе по аналитической химии и метрологии. Проверка других распределений с помощью того же критерия ничем принципиально не отличается от проверки нормальности, в этих случаях используются лишь другие формулы для расчета теоретических частот.

Критерий Пирсона химики иногда используют не для проверки характера распределения, а для оценки степени расхождения экспериментальных данных с табличными (или вычисленными по некоторой теоретической модели)*. Если расхождение окажется значимым, надо дополнительно разобраться, является ли оно следствием случайных погрешностей или вызывается другими причинами, например, принципиальной непригодностью использованной модели, а значит и проверяемой научной гипотезы. Для дополнительной проверки обязательно потребуются провести повторные опыты, многократно измеряя ту

* Такой способ оценки разнородных данных не является строго корректным с математической точки зрения.

характеристику (Θ_i), которая вносила основной вклад в величину χ^2 при ее расчете по формуле (7.7). На основании результатов повторных измерений рассчитывают доверительный интервал и проверяют, попадает ли в него T_i . Если да, то расхождение Θ_i и T_i можно объяснить случайными погрешностями.

Чтобы проверить согласие теории и эксперимента, существует и другой способ, в этом случае применяют критерий Стьюдента. Допустим, при повторных измерениях получено n значений исследуемой характеристики Θ , выборка получена из генеральной совокупности нормально распределенных данных, но среднее значение $\Theta = \bar{x}$ не равно некоторой константе T — ожидаемому или «теоретическому» значению той же характеристики. Тогда показатель t рассчитывают по формуле

$$t = |\bar{x} - T| \frac{\sqrt{n}}{S}, \quad (7.8)$$

где S — выборочное стандартное отклонение. Величину t сопоставляют с табличными значениями критерия Стьюдента для выбранных уровней значимости. Считают, что при таком сравнении число степеней свободы $df = n - 1$.

Пример 7-4. Содержание азота в продукте органического синтеза по данным повторных анализов оказалось равным (%) 11,5; 11,7; 11,8; 11,3; 11,7. Судя по формуле ожидаемого продукта реакции, содержание азота в нем должно быть 11,0%. Следует ли считать расхождение статистически значимым, или его можно объяснить случайными погрешностями анализа?

Решение. Предполагаем нормальное распределение результатов анализа и применяем проверку по Стьюденту: $t = (11,6 - 11,0) \sqrt{5} / 0,2 = 6,7$; $t_{0,01} = 4,6$. Значит, расхождение среднего результата анализа и «теоретического» значения достоверно (уровень значимости нуль-гипотезы менее 0,01). Либо методика анализа дает систематическую погрешность, либо в ходе синтеза получилось не то, что ожидал исследователь.

Глава 8

ОСНОВНЫЕ ИДЕИ КОРРЕЛЯЦИОННОГО И ДИСПЕРСИОННОГО АНАЛИЗА

8.1. ПОНЯТИЕ О КОРРЕЛЯЦИИ

При проведении научных исследований нередко приходится сталкиваться с двумерными совокупностями экспериментальных данных, где объекты отличаются сразу по двум признакам — измеряемым характеристикам X и Y (табл. 8).

Т а б л и ц а 8

Примеры двумерных совокупностей

Объекты	X	Y
Дети	Масса тела	Рост
Больные гипертонией	Частота пульса	Артериальное давление
Ион-селективные электроды	Коэффициент чувствительности	Время установления равновесного потенциала
Одноатомные спирты	Молекулярная масса	Температура кипения
Природные воды	Жесткость воды	Величина pH
Легированные стали	Содержание хрома	Содержание никеля

В активном эксперименте каждый опыт также характеризуется не менее, чем двумя величинами. А именно, значениями фактора (X) и параметра (Y). Так, каждый из изучаемых Pt-содержащих катализаторов может быть охарактеризован содержанием платины (X) и активностью (Y) в некоторых заданных условиях. Исследователь должен установить, существует ли в совокупности таких катализаторов взаимосвязь величин X и Y , каков ее характер. Нередко каждый из исследованных объектов или проведенных опытов характеризуется двумя параметрами (Y_1 и Y_2), и тогда приходится выяснять, взаимосвязаны ли эти параметры или совершенно независимы друг от друга. Примером может быть та же серия однотипных катализаторов, если для каждого катализатора определены значения активности Y_1 и селективности Y_2 .

Возможно, числовые значения X и Y будут варьировать от объекта к объекту несогласованно, независимо друг от друга. Не исключен и противоположный случай, когда между значениями X и Y существует однозначная функциональная связь, заданная, например, уравнениями $Y = b_0 + b_1X$ или $Y = aX^c$. На практике часто встречается промежуточный случай, когда X и Y имеют так называемую корреляционную связь или, другими словами, X и Y являются закоррелированными. Это означает, что между изменениями этих величин существует взаимосвязь. Например, в исследуемой двумерной совокупности данных при переходе от объекта к объекту повышение X , как правило, сопровождается повышением Y (положительная корреляция). Возможна и отрицательная корреляция, когда повышение X гораздо чаще ведет к понижению Y , чем к повышению этой величины.

При наличии корреляции мы не можем однозначно вычислить для какого-либо объекта значение Y по известному значению X (или наоборот). Каждому значению X может соответствовать целый набор объектов с разными Y , а не строго определенное единственное значение Y , как было бы при функциональной зависимости. Известно, что у детей существует положительная корреляция между ростом и массой тела: чем больше рост ребенка, тем, как правило, больше и его масса. Но заранее вычислить эту

массу по измеренному значению роста нельзя, один и тот же рост будут иметь дети с совершенно разной массой. А дети одного роста могут иметь разную массу тела, здесь будут сказываться наследственность, режим питания и многие другие факторы.

За корреляцией величин X и Y часто угадывается строгая причинно-следственная связь явлений, искаженная влиянием других факторов и погрешностями измерения X и Y . В таких случаях X и Y являются неравнозначными; символом X принято обозначать причину (фактор), а символом Y — следствие (параметр). Так, между температурами кипения разных спиртов и молекулярной массой каждого из них существует положительная корреляционная связь, в которой отчетливо различимы причина и следствие: температура кипения (Y) повышается, потому что повышается молекулярная масса (X), но не наоборот. Однако причинно-следственная связь явлений для корреляции не обязательна; типичный пример такой корреляции — связь температур кипения и температур плавления ряда спиртов: оба свойства при переходе от одного спирта к другому обычно меняются в одну и ту же сторону. Здесь нет причины и следствия, обе величины согласованно меняются вследствие изменения третьей — молекулярной массы. Характеристики корреляции вычисляются одним и тем же способом, независимо от наличия или отсутствия причинно-следственной связи между закоррелированными величинами X и Y .

Задачей корреляционного анализа является проверка наличия корреляции между признаками X и Y , а также оценка степени закоррелированности. Корреляционный анализ осложняется ввиду ограниченного объема двумерной совокупности, т. е. из-за малого числа пар X и Y . Гипотеза о наличии корреляции проверяется так же, как и другие статистические гипотезы (см. раздел 7.2). Отметим, что аппарат корреляционного анализа приложим не только к двумерным, но и к многомерным совокупностям, у которых для каждого объекта измерено множество признаков X_1, X_2, \dots, X_k . Тогда выясняют, какие из признаков попарно закоррелированы, а какие не имеют корреляции с остальными.

8.2. ВИДЫ КОРРЕЛЯЦИИ

Если нанести на плоскость значения X и Y для объектов изучаемой совокупности, каждый из них даст определенную точку. Рассматривая совокупность точек, мы отчетливо увидим, к какой функциональной зависимости тяготеет корреляционная связь X и Y (рис. 14). Точки могут группироваться вокруг некоторой прямой или кривой, соответственно говорят о линейной или нелинейной корреляции. Линейная корреляция наиболее удобна для исследований, поэтому пока ограничимся рассмотрением этого случая.

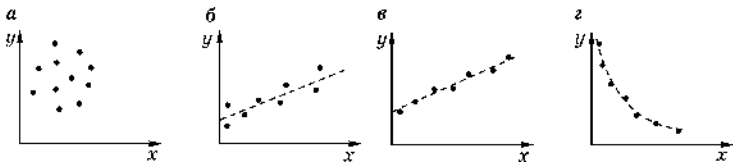


Рис. 14

Графическое истолкование корреляций:

- a* — отсутствие корреляции; *б* — слабая положительная линейная корреляция;
в — сильная положительная линейная корреляция;
г — сильная отрицательная нелинейная корреляция.

Любую линейную корреляцию можно представить себе как результат одновременного проявления линейной зависимости $Y = b_0 + b_1X$ и хаотического разброса результатов. Очевидно, доли этих факторов в разных корреляциях различны. Степень линейной корреляции (ее «силу») оценивают коэффициентом r , который меняется от -1 до $+1$. Чем больше r по модулю, тем ближе лежат точки к некоторой прямой, а корреляция ближе к функциональной зависимости. Считают, что в некоторой корреляции доля функциональной связи равна r^2 , а доля хаоса равна $1 - r^2$. Так, корреляцию с $r = 0,9$ можно представить результатом одновременного проявления линейной зависимости (81%) и хаотического разброса данных (19%). Однако такое истолкование коэффициента корреляции нельзя считать математически строгим и общепринятым.

Знак коэффициента корреляции свидетельствует о направлении изменений X и Y . При $r > 0$ увеличение X при

переходе от объекта к объекту чаще связано с одновременным увеличением Y , чем с его уменьшением (положительная корреляция). Отрицательное значение r указывает на то, что рост X в данной совокупности объектов чаще связан с уменьшением Y , чем с увеличением (отрицательная корреляция). Знак корреляции относится ко всей совокупности исследуемых объектов, но не к любой паре объектов из этой совокупности: даже при $r \approx +1$ можно найти такую пару объектов, где большему значению X соответствует меньшее значение Y .

Для выявления знака и степени корреляции должны быть использованы все имеющиеся в распоряжении исследователя данные. Или объекты («точки») должны быть выбраны из соответствующей генеральной совокупности совершенно случайным образом, без их специального отбора с целью подгонки выводов под желаемый результат. Начинаящим исследователям особенно важно усвоить недопустимость «целенаправленного отбора», который ничем не лучше прямой фальсификации экспериментальных данных.

8.3. ПРОВЕРКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ КОРРЕЛЯЦИИ

Предположим, что мы имеем n объектов двумерной нормально распределенной совокупности и получили таблицу с n парами значений X и Y , т. е. с результатами измерения двух признаков каждого объекта. Требуется выяснить, имеется ли достоверная линейная корреляция между X и Y . Коэффициент линейной корреляции (r) рассчитываем по одной из следующих формул:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \cdot \sqrt{\sum (Y_i - \bar{Y})^2}}; \quad (8.1)$$

$$r = \frac{n \sum (X_i Y_i) - \sum X_i \sum Y_i}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2] \cdot [n \sum Y_i^2 - (\sum Y_i)^2]}}. \quad (8.2)$$

Отметим, что формула (8.2) удобнее для вычислений, чем формула (8.1). Кроме того, расчет по формулам (8.1) или (8.2) можно значительно облегчить и ускорить, если кодировать значения X_i и Y_i . Могут быть использованы разные способы кодировки, в частности, величина r не изменится, если все X_i умножить или разделить на постоянное положительное число, отнять его или прибавить. То же относится к значениям Y_i .

За счет случайного характера выборки коэффициент корреляции обычно несколько отличается от 0 и в том случае, когда никакой взаимосвязи величин X и Y нет. Предположение об отсутствии линейной корреляции является нуль-гипотезой. Чтобы сделать достоверный вывод о наличии корреляции, надо, чтобы значение r по абсолютной величине превосходило бы критическое значение $r_{\text{табл}}$ зависящее от α — уровня значимости нуль-гипотезы и df — числа степеней свободы. Так как мы при вычислении r по формуле (8.1) дважды усредняли варианты (и по X , и по Y), то в данном случае $df = n - 2$. Значения $r_{\text{табл}}$ для $\alpha = 0,05$ и $\alpha = 0,01$ приведены в приложении 2L, с учетом числа степеней свободы.

При $|r| > r_{\text{табл}}$ делаем вывод о наличии линейной корреляции, вероятность ошибки первого рода при этом меньше $1 - \alpha$. При $|r| < r_{\text{табл}}$ значимость нуль-гипотезы выше критического уровня α , и отказываться от нее нельзя. Вывод должен быть следующим: «Наличие линейной корреляции не установлено ($\alpha > \alpha_{0,05}$)». Однако при $|r| < r_{\text{табл}}$ не следует считать, что величины X и Y совсем не связаны друг с другом. Возможно, корреляция между ними есть, но не линейная! Или, может быть, мы взяли для проверки слишком малое число объектов, а расширение объема обрабатываемой совокупности данных (увеличение числа точек) приведет к установлению наличия линейной корреляции.

При малом числе объектов (пар X_i и Y_i) выводы о корреляциях вообще недостаточно надежны, рекомендуется брать не менее 10 таких пар. В статистике известны способы расчета того значения n , которое необходимо для заданной надежности вывода о корреляции, но эти расчеты довольно сложны и в настоящей книге не рассматриваются.

Пример 8-1. Измерены выход продукта (W , % по массе) и давление в реакторе (P , атм) в ходе некоторого промышленного процесса. Они записаны в таблице 9. Существует ли между этими величинами линейная корреляция?

Т а б л и ц а 9

Расчет коэффициента корреляции для примера 8-1

P	W	X	Y	X^2	XY	Y^2
1,0	64	-3	-6	9	18	36
1,5	67	-2	-3	4	6	9
2,0	74	-1	4	1	-4	16
2,5	66	0	-4	0	0	16
3,0	73	1	3	1	3	9
3,5	79	2	9	4	18	81
4,0	82	3	12	9	36	144
Суммы		0	15	28	77	311

Решение. Для упрощения расчетов применим следующие кодировки: $X = 2(P - P_{\text{ср}}) = 2(P - 2,5)$; $Y = W - 70$. Рассчитываем необходимые суммы (см. табл. 9), подставляем их в формулу (8.2) и получаем $r = 0,87$. Находим в таблицах критическое значение коэффициента: для $\alpha = 0,05$ и $df = 5$ $r_{\text{табл}} = 0,7545$. Получаем, что $|r| > r_{\text{табл}}$. Следовательно, линейная корреляция достоверна на уровне значимости $\alpha < 0,05$.

Более сложен случай, когда в координатах (X , Y) точки хорошо ложатся не на прямую, а на некоторую кривую. В этих случаях можно рассчитывать по особым формулам коэффициент нелинейной корреляции, надо только заранее знать вид аппроксимирующей кривой (парабола, гипербола и т. д.). Но чаще пользуются более простым способом: вначале проводят линеаризацию данных, т. е. кодировку всех значений X и/или Y . Затем рассчитывают коэффициент линейной корреляции для новых (кодированных) переменных. Способ кодировки зависит от того, какой математической формуле соответствует кривая, к которой тяготеют точки в координатах (X , Y). Так, ожидая, что между X и Y существует связь вида $Y = aX^c$, искаженная разбросом экспериментальных данных, проводят логарифмическую линеаризацию $X = \lg X$ и $Y = \lg Y$.

(Здесь кодированные значения вариант выделены жирным наклонным шрифтом). Если наши предположения по поводу аппроксимирующей функции верны, между X и Y должна существовать линейная корреляция. Если это предположение подтверждается расчетом r , тогда можно утверждать, что между величинами X и Y также имеется корреляция, но нелинейная.

Часто на практике надо выбирать лучшую аппроксимирующую функцию. В этом случае последовательно используют несколько способов линеаризации, т. е. различные способы кодировки. Для каждого способа рассчитывают коэффициент линейной корреляции. Чем выше окажется $|r|$, тем лучше выбранная функция подходит для описания экспериментальных данных.

Так, связь между температурой (T) и скоростью (v) реакции без линеаризации дает довольно низкое значение $|r|$, корреляция оказывается недостоверной. Несколько выше будет $|r|$ при переходе к переменным $X = \lg T$ и $Y = \lg v$. Если же взять в качестве переменных $X = 1/T$ и $Y = \lg v$ (так называемые аррениусовские координаты), коэффициент r будет по абсолютной величине близок к единице. Если последний способ линеаризации оказывается лучшим, то можно сделать вывод, что экспериментальные данные приблизительно соответствуют уравнению Аррениуса: $v = ke^{-E/RT}$.

8.4. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Алгоритм сопоставления средних по критерию Стьюдента позволяет сравнивать математические ожидания случайной величины до и после воздействия фактора. Но этого недостаточно, чтобы решить важнейшую практическую задачу: влияет ли исследуемый фактор на интересующую нас величину или нет? Для ответа на такой вопрос эксперимент следует проводить, измеряя параметр при нескольких уровнях исследуемого фактора. Полученные в ходе эксперимента данные можно обрабатывать разными способами. В частности, можно рассчитать

коэффициент корреляции. Однако при таком подходе мы не учитываем воспроизводимость измерений. Более надежные результаты можно получить, многократно измеряя параметр при одном и том же уровне исследуемого фактора. Выполнив несколько серий таких измерений для разных уровней фактора, мы сможем сопоставить разброс значений параметра в повторных опытах на одном уровне с различием его средних значений на разных уровнях. Таким образом, влияние фактора будет исследовано с учетом воспроизводимости измерений параметра.

Эту идею выдвинул в 1925 г. Р. Фишер, и разработанный им способ обработки данных известен под названием *простого* или *однофакторного дисперсионного анализа*. В целом же дисперсионный анализ — это исследование неоднородного числового материала путем разложения общей дисперсии на отдельные составляющие. В основе дисперсионного анализа, как и многих других статистических методов, лежит предположение о том, что измеряемые величины имеют нормальное распределение. То есть однофакторный дисперсионный анализ — это параметрический метод.

Допустим, проводят эксперимент, задавая фактор на k уровнях, причем на фиксированном j -м уровне параметр измеряют n_j раз. Результаты измерений записывают в таблицу 10. В каждом столбце записаны результаты повторных измерений случайной величины в некоторых заданных условиях, т. е. на некотором уровне исследуемого фактора.

Т а б л и ц а 10

**Представление исходных данных
для однофакторного дисперсионного анализа**

Уровни фактора	1	2	...	k
Результаты измерений параметра	$y_{1,1}$	$y_{1,2}$...	$y_{1,k}$
	$y_{2,1}$	$y_{2,2}$...	$y_{2,k}$
	$y_{3,1}$	$y_{3,2}$...	$y_{3,k}$

	$y_{n1,1}$	$y_{n2,2}$...	$y_{nh,k}$

Число повторных измерений в каждом столбце в общем случае может быть различным, но обычно значения n_1, n_2, \dots, n_k одинаковы и равны небольшому целому числу n .

Считаем, что столбцы таблицы 10 составляют k независимых выборок, каждая из которых получена из своей генеральной совокупности с нормальным распределением. Уровни фактора соответствуют номеру выборки, записанной в данном столбце таблицы 10.

Возможно, генеральные совокупности данных, получаемых на всех исследуемых уровнях, имеют равные дисперсии σ^2 , но различные математические ожидания m_1, m_2, \dots, m_k . Различие в математических ожиданиях может быть вызвано влиянием изучаемого фактора. Нуль-гипотеза состоит в том, что $m_1 = m_2 = \dots = m_k = m$. Иначе говоря, вначале мы предполагаем, что влияние фактора на математические ожидания параметра отсутствует. Если нуль-гипотеза будет опровергнута, это будет означать, что некоторые (или все) математические ожидания отличаются друг от друга, что может быть вызвано влиянием изучаемого фактора.

Алгоритм проверки нуль-гипотезы включает несколько шагов (стадий).

На нулевом (начальном) шаге каждую выборку (столбцы из таблицы 10) проверяют на наличие грубых промахов, рассчитывают ее выборочное среднее и выборочную дисперсию, используя стандартные формулы. Выборочные дисперсии проверяют по критериям Кохрена или Бартлетта. Если все дисперсии окажутся однородными, переходят к следующим шагам.

На первом шаге проводят вспомогательные вычисления, связанные как с отдельными столбцами, так и со всей совокупностью данных, представленных в таблице 10. Обозначим через $n = n_1 + n_2 + \dots + n_k$ общее число измерений. Далее вычислим:

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{i,j}; \quad \bar{y} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} y_{i,j}, \quad (8.3)$$

где \bar{y}_j — выборочное среднее j -й выборки ($j = 1, 2, \dots, k$); \bar{y} — общее выборочное среднее по всей совокупности данных. Сумма квадратов отклонений измерений от общего среднего может быть представлена в виде

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{i,j} - \bar{y})^2 = Q_1 + Q_2,$$

где Q_1 — сумма квадратов отклонений выборочных средних \bar{y}_j от их общего среднего \bar{y} (между группами); Q_2 — сумма квадратов отклонений измерений от выборочных средних групп (внутри групп):

$$Q_1 = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2; \quad (8.4)$$

$$Q_2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{i,j} - \bar{y}_j)^2. \quad (8.5)$$

Величина $S_1^2 = \frac{Q_1}{k-1}$ характеризует рассеяние групповых средних и называется факторной (межгрупповой) дисперсией. Величина $S_2^2 = \frac{Q_2}{n-k}$ отражает рассеяние внутри групп, которое обусловлено случайными вариациями результатов измерений, и называется остаточной (внутригрупповой) дисперсией. Заметим здесь, что значения величины S_1^2 , существенно превышающие значения S_2^2 , могут быть объяснены различием средних в группах.

На втором шаге находим величину $\lambda_{\text{эксп}}$ по формуле

$$\lambda_{\text{эксп}} = \frac{S_1^2}{S_2^2}. \quad (8.6)$$

Фиксируем уровень значимости нуль-гипотезы (α) и сравниваем $\lambda_{\text{эксп}}$ с $\lambda_{\text{крит}}$, значение которого выбирается из таблиц критерия Фишера для заданного значения α и числа степеней свободы $df_1 = k - 1$, $df_2 = n - k$. Если окажется, что верно неравенство $\lambda_{\text{эксп}} \leq \lambda_{\text{крит}}$, то нуль-гипотезу принимаем. В этом случае можно считать, что все математические ожидания равны между собой, их общее значение $m_1 = m_2 = \dots = m_k = m$ оценивается величиной \bar{y} , а дисперсия σ^2 оценивается величиной S_2^2 . При выполнении

неравенства $\lambda_{\text{экс}} > \lambda_{\text{крит}}$ нуль-гипотеза отклоняется и следует считать, что среди математических ожиданий m_1, m_2, \dots, m_k есть хотя бы два, не равных друг другу.

Отклонение нуль-гипотезы может повлечь за собой необходимость применения дополнительных методов, связанных с оценкой влияния фактора на измеряемый параметр. Так, например, для поиска групп, имеющих значимое различие средних, естественно использовать попарное сравнение групп по критерию Стьюдента. Для оценки силы влияния фактора можно использовать метод Снедекора [26, с. 230].

Пример 8-2. Изучается влияние ионной силы на величину оптической плотности раствора дихромата калия при прочих постоянных условиях. В эксперименте получены следующие значения оптической плотности (табл. 11).

Т а б л и ц а 11

Представление данных для однофакторного дисперсионного анализа

Ионная сила	0,01	0,05	0,10	0,25
Результаты измерений оптической плотности	0,30	0,31	0,34	0,33
	0,28	0,30	0,32	0,34
	0,29	0,32	0,30	0,32
Выборочные средние	0,29	0,31	0,32	0,33

Сопоставление средних, полученных для разных столбцов этой таблицы, показывает, что выборочные средние возрастают по мере увеличения ионной силы: $\bar{y}_1 < \bar{y}_2 < \bar{y}_3 < \bar{y}_4$. На основании этого можно предположить, что ионная сила оказывает положительное влияние на величину оптической плотности раствора. Однако такое предположение следует рассматривать только как предварительное. Оно требует обоснования с применением однофакторного дисперсионного анализа. Для этого проверим нуль-гипотезу, состоящую в том, что влияние ионной силы на величину оптической плотности раствора отсутствует.

Р е ш е н и е. В представленной таблице имеется $k = 4$ уровня фактора (величина ионной силы), объемы выборочных данных $n_1 = n_2 = n_3 = n_4 = 3$, общее число измерений $n = 12$. Будем считать, что данные не содержат грубых

промахов, а дисперсии однородны (проверить это можно с применением простейших способов, описанных в разделе 7.4). Применяя формулы (8.3)–(8.6), получаем, что общее выборочное среднее $\bar{y} = 0,3125$, факторная дисперсия $S_1^2 = 0,000875$, остаточная дисперсия $S_2^2 = 0,000175$, $\lambda_{\text{экс}} = 5,0$. Число степеней свободы $df_1 = k - 1 = 3$, $df_2 = n - k = 8$. Критические значения таковы: $\lambda_{\text{крит}} = \lambda_1 = 4,07$ для $\alpha = 0,05$ и $\lambda_{\text{крит}} = \lambda_2 = 7,59$ для $\alpha = 0,01$. Здесь мы имеем случай, когда $\lambda_1 < \lambda_{\text{экс}} < \lambda_2$. На уровне значимости $\alpha = 0,01$ нуль-гипотеза принимается. Поэтому предварительное предположение о влиянии ионной силы на оптическую плотность раствора дихромата калия в данном эксперименте не подтверждается с необходимой надежностью. По-видимому, исследователю следует продолжить эксперимент, увеличить объем исходных данных и все расчеты повторить заново.

8.5. ДВУХФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Во многих химических исследованиях возникает ситуация, когда параметр может варьировать под действием нескольких факторов одновременно, и требуется выяснить значимость каждого из них. Важно, что факторы могут иметь как *качественный*, так и *количественный* характер. Примером может быть определение содержания кадмия в почве, причем анализируют множество проб почвы, взятых на разной глубине (количественный фактор) и в разных местах (качественный фактор). Другим примером может быть исследование активности серии однотипных катализаторов, отличающихся по дисперсности, по продолжительности предварительного прокаливания и по температуре прокаливания. Для решения этих и множества других задач используют многофакторный дисперсионный анализ. Отличие этого метода от однофакторного дисперсионным анализом состоит в том, что, как правило, нет необходимости проводить параллельные измерения. Интересующий нас параметр достаточно измерять один раз для каждого из допустимых сочетаний уровней факторов. Получаемые при этом данные образуют

многомерную таблицу, которая обрабатывается с помощью специальных формул. Рассмотрим относительно несложный случай, когда проверяется влияние только двух факторов A и B . Предполагается, что фактор A имеет k уровней, фактор B — n уровней, факторы не связаны между собой и не взаимодействуют друг с другом. Результаты измерений параметра обозначим как $y_{i,j}$. Данные представляют, как показано в таблице 12. Числа $y_{i,j}$ из таблицы 12 означают результаты однократных измерений при фиксированных сочетаниях уровней факторов B_i и A_j ($i = 1, \dots, n, j = 1, \dots, k$).

Т а б л и ц а 12

Представление данных для двухфакторного дисперсионного анализа

Уровни фактора B	Уровни фактора A			
	A_1	A_2	...	A_k
B_1	$Y_{1,1}$	$y_{1,2}$...	$y_{1,k}$
B_2	$Y_{2,1}$	$y_{2,2}$...	$y_{2,k}$
B_3	$Y_{3,1}$	$y_{3,2}$...	$y_{3,k}$
...
B_n	$Y_{n,1}$	$y_{n,2}$...	$y_{n,k}$

В ходе вычислений проверяются две основные нуль-гипотезы: 1) влияние фактора A отсутствует; 2) влияние фактора B отсутствует. Алгоритм проверки гипотез состоит в следующем.

На нулевом шаге проводят вспомогательные вычисления, связанные с отдельными столбцами и строками таблицы 12. Однако ни о какой отбраковке грубых промахов не может быть и речи — в многофакторном дисперсионном анализе мы, как правило, не имеем результатов параллельных измерений! Надо рассчитать сумму вариантов, записанных в j -м столбце (обозначим ее через Y_j^A), и сумму вариантов, стоящих в i -й строчке (Y_i^B), т. е. суммы следующего вида:

$$\begin{aligned} Y_j^A &= y_{1,j} + y_{2,j} + \dots + y_{n,j}, & j &= 1, \dots, k; \\ Y_i^B &= y_{i,1} + y_{i,2} + \dots + y_{i,k}, & i &= 1, \dots, n. \end{aligned} \quad (8.7)$$

На первом шаге вычисляем суммы квадратов величин Y_j^A и Y_i^B , деленные на число вариант в столбцах и строках соответственно:

$$Q_A = \frac{1}{n} \sum_{j=1}^k (Y_j^A)^2; \quad (8.8)$$

$$Q_B = \frac{1}{k} \sum_{i=1}^n (Y_i^B)^2. \quad (8.9)$$

На втором шаге вычисляем сумму квадратов всех вариантов, представленных в таблице 12, а именно:

$$Q_{AB} = \sum_{i=1}^n \sum_{j=1}^k Y_{i,j}^2. \quad (8.10)$$

Затем находим квадрат суммы всех вариантов, деленный на общее число вариант:

$$Q_S = \frac{1}{nk} \left(\sum_{j=1}^k Y_j^A \right)^2 = \frac{1}{nk} \left(\sum_{i=1}^n Y_i^B \right)^2. \quad (8.11)$$

На третьем шаге вычисляем выборочные дисперсии, относящиеся к измеряемой величине y , с учетом возможного влияния факторов A и B :

$$S_0^2 = \frac{Q_{AB} + Q_S - Q_A - Q_B}{(k-1)(n-1)}; \quad (8.12)$$

$$S_A^2 = \frac{Q_A - Q_S}{k-1}; \quad (8.13)$$

$$S_B^2 = \frac{Q_B - Q_S}{n-1}. \quad (8.14)$$

На четвертом шаге находим величины $\lambda_{\text{эксп}}^A, \lambda_{\text{эксп}}^B$ по формулам:

$$\lambda_{\text{эксп}}^A = \frac{S_A^2}{S_0^2}; \quad \lambda_{\text{эксп}}^B = \frac{S_B^2}{S_0^2}. \quad (8.15)$$

На последнем, пятом шаге фиксируем уровень значимости нуль-гипотезы (α) и сравниваем $\lambda_{\text{эксп}}^A$ и $\lambda_{\text{эксп}}^B$ с величиной $\lambda_{\text{крит}}$, которая выбирается из таблиц критерия Фишера при заданном α , $df_1 = k - 1$, $df_2 = (k - 1)(n - 1)$. Влияние фактора A признается значимым, т. е. нуль-гипотеза отклоняется, если выполняется неравенство $\lambda_{\text{эксп}}^A > \lambda_{\text{крит}}$. Соответственно влияние фактора B считается значимым, если $\lambda_{\text{эксп}}^B > \lambda_{\text{крит}}$. Нуль-гипотезы относительно обоих факторов принимаются (т. е. считают, что влияние факторов A и B на изучаемый параметр не обнаружено), если одновременно выполняются неравенства $\lambda_{\text{эксп}}^A \leq \lambda_{\text{крит}}$ и $\lambda_{\text{эксп}}^B \leq \lambda_{\text{крит}}$. В последнем случае неизвестное значение дисперсии σ^2 оценивается величиной $(Q_{AB} - Q_S)/(nk - 1)$.

Пример 8-3. Проводились межлабораторные испытания шести разных ингибиторов коррозии (фактор A). Эти испытания проводились по единой методике в трех различных лабораториях (фактор B). Результаты испытаний во всех случаях выразили величиной скорости коррозии, в процентах от скорости коррозии в отсутствие ингибитора. Полученные данные представлены в таблице 13.

Т а б л и ц а 13

Пример представления исходных данных
для двухфакторного дисперсионного анализа

Уровни фактора B	Уровни фактора A					
	A_1	A_2	A_3	A_4	A_5	A_6
B_1	1,90	2,30	2,60	1,80	2,00	2,05
B_2	2,02	2,10	3,20	2,70	4,00	2,40
B_3	1,60	1,50	1,80	2,60	1,90	1,70

Используя представленные данные, требуется выбрать лучший ингибитор или статистически подтвердить заметную на первый взгляд одинаковую эффективность всех исследованных ингибиторов. Попутно сопоставляют работы разных лабораторий. Естественно, результаты проверки будут более доказательными, если окажется, что выводы трех независимых лабораторий статистически совпадают, т. е. фактор исполнителя не значим.

Р е ш е н и е. Для того чтобы сопоставить между собой эффективность ингибиторов, необходимо оценить влияние

фактора A . При интерпретации получающихся результатов следует также учесть и фактор B , так как различие данных по лабораториям может быть обусловлено систематическими погрешностями, допускаемыми какой-либо лабораторией, или местными условиями. Предполагаем, что результаты измерений во всех лабораториях имеют нормальное распределение. Поэтому здесь можно и нужно использовать двухфакторный дисперсионный анализ. Рассмотрим нуль-гипотезы, состоящие в том, что остаточная скорость коррозии (в %) не зависит ни от типа ингибитора, ни от того, в какой именно лаборатории проводилось это исследование. Количество уровней факторов в этой задаче таково: $k = 6$, $n = 3$. Применяя формулы (8.7)–(8.14), получаем, что $S_0^2 = 0,247$, $S_A^2 = 0,316$, $S_B^2 = 1,245$. Используя далее формулы (8.15), находим, что $\lambda_{\text{эксн}}^A = 1,279$, $\lambda_{\text{эксн}}^B = 5,04$. Число степеней свободы равно $df_1 = 5$, $df_2 = 10$. Критические значения таковы: $\lambda_1 = 3,33$ для $\alpha = 0,05$ и $\lambda_2 = 5,64$ для $\alpha = 0,01$. На уровне значимости $\alpha = 0,05$ нуль-гипотезу относительно фактора A принимаем. Полученный результат говорит о том, что достоверных преимуществ по эффективности ни один из ингибиторов не имеет (по сравнению с другими в данной серии). Повидимому, для выбора ингибитора следует привлекать не только эффективность, но и другие характеристики (стоимость, токсичность и т. д.). Для фактора B на уровне значимости $\alpha = 0,05$ нуль-гипотеза отклоняется. Последнее следует интерпретировать как зависимость результатов испытаний от того, в какой лаборатории они проводились. Однако на уровне значимости $\alpha = 0,01$ нуль-гипотеза сохраняется, т. е. нельзя утверждать, что в разных лабораториях результаты испытаний достоверно различаются.

Таким образом, на вопрос — одинаковы ли результаты испытаний в разных лабораториях? — достаточно четкий ответ мы не получили. Однако на основной вопрос мы получили вполне определенный ответ — влияние фактора A не выявлено, можно считать, что все исследованные ингибиторы коррозии одинаково эффективны.

Глава 9

ОСНОВНЫЕ ИДЕИ РЕГРЕССИОННОГО АНАЛИЗА

9.1. ЦЕЛИ И ОГРАНИЧЕНИЯ

Если в ходе эксперимента оказывается, что некоторое свойство исследуемого объекта (например, параметр Y) связано корреляционной зависимостью с фактором X или даже с целым рядом факторов (X_1, X_2, X_3 и т. д.), то исследователь всегда пытается установить, какая же функциональная связь переменных прячется за хаосом случайных погрешностей и влиянием неконтролируемых факторов. Такую связь желательно выразить одним, пусть даже очень сложным уравнением, одновременно учитывающим взаимосвязь всех закоррелированных величин. Принято выражать искомое уравнение как зависимость одного параметра от нескольких факторов, т. е. в виде $Y = f(X_1, X_2, X_3, \dots)$. Это алгебраическое уравнение называется *уравнением регрессии* или просто регрессией. Нахождение уравнения регрессии — основная цель регрессионного анализа. Обычно этот метод применяют к тем многомерным совокупностям данных, где предполагаются причинно-следственные связи величин и где можно точно указать аргументы (факторы) и функции (параметры).

Найденное уравнение регрессии окажется статистической моделью исследуемой системы. В частности, по нему можно будет без дополнительных экспериментов прогнозировать величину Y для любой комбинации факторов

(«иксов»). Можно решать и обратные задачи, т. е. находить такую комбинацию «иксов», которая обеспечит заданное значение Y .

Регрессионный анализ — логическое продолжение корреляционного и дисперсионного. Так, проводя полное исследование некоторого объекта, вначале по данным предварительных опытов рассчитывают коэффициенты корреляции для разных пар признаков и выделяют взаимосвязанные величины. Решают, какие из них являются факторами (аргументами), а какие — параметрами (функциями). Проверяют методами дисперсионного анализа, является ли влияние выбранных факторов статистически достоверным на фоне погрешностей измерения. Только затем проводят основной эксперимент, задавая по определенному плану (см. главу 2) значения факторов и измеряя параметры. Далее начинается регрессионный анализ полученных данных. Он включает следующие операции.

I. Выбирают вид уравнения регрессии и записывают его в виде многочлена (полинома) с неизвестными коэффициентами, например в следующем виде :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2.$$

В идеальном случае коэффициенты β_i должны быть такими, чтобы при подстановке значений «иксов» в уравнение регрессии получались расчетные значения Y , совпадающие с измеренными значениями $Y_{\text{экс}}$. Для нахождения таких значений β_i нужен бесконечно большой объем экспериментальных данных. Поскольку в нашем распоряжении имеется ограниченный объем данных, рассчитывают приближенные значения (*оценки*) коэффициентов $b_i \approx \beta_i$.

II. Проверяют, все ли найденные оценки коэффициентов (значения b_i) значимы, нет ли среди них таких, которые не вносят достоверного вклада в результат расчета. Отбрасывают незначимые коэффициенты, так как для них $\beta_i = 0$.

* Коэффициенты обозначаются греческими буквами, если относятся к генеральной совокупности данных, а латинскими — их оценки, рассчитанные по малой выборке.

III. Подставляют некоторые значения X_i в полученное уравнение регрессии и проверяют, насколько совпадает вычисленное значение Y с экспериментом (со значениями $Y_{\text{эксп}}$).

Каждую из этих стадий мы в дальнейшем рассмотрим более детально.

Регрессионный анализ осложняется тем, что все измерения (и X , и Y) отягощены случайными погрешностями. Алгоритмы регрессионного анализа применимы только в тех случаях, когда исследуемая система отвечает четырем требованиям.

1. Погрешности (ошибки) происходят в основном при измерении параметров (функций), а не при задании факторов (аргументов). Во всяком случае, при повторении опытов воспроизводимость параметров не может быть лучше, чем воспроизводимость факторов, о чем можно судить по величинам коэффициентов вариации.

2. Все опыты равноточны и дают приблизительно одинаковую воспроизводимость измерения Y . Повторение опытов в любых условиях дает однородные дисперсии, т. е. $S^2(Y) \approx \text{const}$. Дисперсия Y не должна зависеть ни от значений X , ни от значений Y . Это условие не может выполняться, если величина Y меняется от опыта к опыту на несколько порядков. Классические алгоритмы регрессионного анализа применимы для обработки результатов эксперимента, если значения Y в разных условиях являются величинами одного порядка.

3. Значения Y измеряются без систематических ошибок, и при повторении опытов среднее значение величины Y стремится к ее истинному значению.

4. Полученные в повторных экспериментах при заданных и неизменных значениях всех «иксов» величины Y представляют собой случайные и независимые друг от друга величины, распределенные по нормальному закону. Таким образом, классический регрессионный анализ — параметрический метод.

Методика проведения эксперимента должна обеспечивать выполнение всех четырех перечисленных требований. Желательно перед началом опытов проверить это, в частности, убедиться в нормальном распределении результатов измерений. Кроме того, в ходе самого экспери-

мента следует время от времени убеждаться в отсутствии систематических погрешностей. Если хотя бы одно из четырех перечисленных выше требований не выполняется, связь между Y и X следует находить иными способами, более сложными, чем те, которые описаны ниже. Например, при невыполнении первого требования используют алгоритмы конфлюентного анализа, а при невыполнении последнего — так называемые *робастные (непараметрические)* методы обработки данных.

9.2. ВИДЫ РЕГРЕССИЙ. ВЫБОР УРАВНЕНИЯ РЕГРЕССИИ

Регрессию вначале задают в общем виде, а именно, в виде полинома с неизвестными коэффициентами, например:

$$Y = \beta_0 + \sum \beta_i X_i$$

(линейная регрессия);

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ij} X_i X_j, \quad i \neq j$$

(неполная регрессия второго порядка)*.

Поверхностями отклика для линейных регрессий служат плоскости в $(k + 1)$ -мерном пространстве, где k — количество факторов. Регрессиям второго порядка в том же пространстве соответствуют криволинейные поверхности. Такие регрессии используют, если есть основания предполагать неаддитивное взаимодействие факторов. Например, если Y — масса продукта синтеза, X_1 — концентрация исходного вещества, а X_2 — время синтеза, то экспериментатор вправе предположить, что масса продукта при низкой концентрации реагента накапливается в ходе опыта (зависит от времени) не так, как при высокой, т. е. действие этих факторов взаимосвязано. Это обстоятельство должно быть учтено в математической модели процесса отдельным слагаемым, включающим произведение X_1 и X_2 ,

* В литературе по планированию эксперимента часто используют и другую терминологию: регрессию с произведениями факторов называют нелинейной, а регрессией второго порядка называют только такую, где есть квадраты факторов.

а неполная линейная регрессия здесь не подойдет. Очевидно, для данного примера следует записать регрессию в виде:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2.$$

Регрессия может включать в качестве слагаемых не только парные, но и тройные и еще более сложные произведения переменных (с соответствующими коэффициентами). Если исследуемый процесс очень сложен, иногда не удастся подобрать простую модель, которая бы удовлетворительно отвечала имеющимся экспериментальным данным, не подходит даже регрессия, включающая произведение всех факторов. В этих случаях используют *нелинейные* (относительно факторов) регрессии — полиномы, содержащие слагаемые, в которые значение X входит во второй или даже в третьей степени. Например, для однофакторного эксперимента, в котором изучается зависимость $Y = f(X)$, в качестве нелинейной регрессии часто берут уравнение параболы:

$$Y = \beta_0 + \beta_1 X + \beta_{11} X^2.$$

Возможны и другие регрессии второго порядка, например:

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ij} X_i X_j + \sum \beta_{ij} X_i^2$$

(полная регрессия второго порядка).

Выбор уравнения регрессии на первый взгляд кажется произвольным. Действительно, иногда начинают с самой простой модели — линейной регрессии; убедившись в ее непригодности, проверяют полную регрессию с произведениями факторов. Если потребуется, переходят к уравнениям второго и более высоких порядков, и так далее, до тех пор, пока не отыщут подходящую (адекватную) регрессию. Такой перебор вариантов легко реализуется при использовании компьютера со специфическим программным обеспечением и не занимает много времени. Еще быстрее можно подобрать адекватную математическую модель, если с самого начала взять достаточно сложную регрессию и вычислить оценки ее коэффициентов. Если исследуемая система окажется проще, чем первоначально выбранная сложная модель, коэффициенты при «лишних» слагаемых

будут незначимыми, практически нулевыми, и модель сама упростится до необходимой степени. Однако наиболее верный — выбор уравнения регрессии на основании априорной информации об исследуемой системе, в том числе наших теоретических знаний. Так, в вышепересмотренном примере, где мы должны были найти связь массы продукта синтеза с концентрацией исходного реагента и временем синтеза, следовало бы вспомнить, какой вид имеет кинетическое уравнение любой реакции первого порядка (в интегральной форме). По этому уравнению масса продукта пропорциональна концентрации исходного реагента и времени синтеза, по крайней мере при небольшой степени протекания реакции. Поэтому мы должны были выбрать уравнение регрессии, включающее произведение этих факторов.

Для линейной регрессии гораздо проще рассчитать коэффициенты, чем для любой другой, поэтому перед проведением регрессионного анализа часто проводят *линеаризацию переменных*. Это кодировка значений X и Y , выполняемая так, чтобы прийти к линейной регрессии. Линеаризацию выполняют на основании априорной или теоретической информации об исследуемой системе. Так, имея ряд значений концентрации (C) вещества в эталонных растворах, измерив значения светопропусканий (T) этих растворов на некоторой длине волны и желая построить градуировочный график, мы не будем искать зависимость вида $T = f(C)$, зная, что эти зависимости всегда нелинейны. Значения T следует перевести в значения оптической плотности ($D = 2 - \lg T$), а уже потом находить коэффициенты линейной регрессии: $D = \beta_0 + \beta_1 C$.

Еще пример. Допустим, проведен ряд опытов, в которых одну и ту же реакцию вели при разных концентрациях реагентов (C_1 и C_2) и разных температурах (T). При этом измеряли значения скорости реакции (V). Для того чтобы отыскать методами регрессионного анализа подходящее кинетическое уравнение для данной реакции, проводим предварительную линеаризацию. Используем для этого общее уравнение химической кинетики (уравнение Аррениуса):

$$V = K_0 e^{-E/RT} C_1^{n_1} C_2^{n_2}.$$

Преобразуя его, получаем: $\ln V = \ln K_0 - E/RT + n_1 \ln C_1 + n_2 \ln C_2$. Кодлируем переменные: $Y = \ln V$; $X_1 = 1/T$; $X_2 = \ln C_1$; $X_3 = \ln C_2$. Записываем предполагаемое уравнение регрессии в виде

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

где $\beta_0 = \ln K_0$; $\beta_1 = -E/R$; $\beta_2 = n_1$; $\beta_3 = n_2$.

Следует помнить, что теоретически можно записать исходное уравнение регрессии в виде полинома любой сложности, с любым количеством коэффициентов β_i , но на самом деле мы сможем рассчитать значения лишь некоторых из них, в соответствии со следующим правилом: число рассчитываемых коэффициентов не может быть больше числа проведенных независимых опытов! Повторно проведенные в тех же самых условиях опыты не учитываются. Например, двух пар значений Y и X достаточно, чтобы найти уравнение линейной регрессии в виде $Y = \beta_0 + \beta_1 X$, где надо будет рассчитывать два коэффициента. Однако рассчитать уравнение $Y = \beta_0 + \beta_1 X + \beta_{11} X^2$ по этим данным мы не сможем, понадобится хотя бы еще один опыт, еще пара значений X и Y . Избыток же экспериментальных данных (четвертая или пятая пара значений) не мешает, напротив, в этом случае точность оценки коэффициентов увеличится.

9.3.

ОЦЕНКА КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Точные значения коэффициентов $\beta_0, \beta_1, \beta_2, \dots$ в уравнении регрессии мы сможем рассчитать только в том случае, если будем иметь бесконечно большой объем экспериментальных данных и будем пользоваться достаточно сложной моделью. Так как число опытов на самом деле невелико, то результаты расчета (числа b_0, b_1, b_2, \dots) будут близки, но не точно равны искомым коэффициентам. Значения b_i называют оценками соответствующих (неизвестных для исследователя) коэффициентов регрессии β_i . Оценки тем более точны, чем больше проведено опытов. Если регрессию рассчитывают по результатам n опытов, то считают, что $\beta_i = \lim b_i$ при $n \rightarrow \infty$. Но последнее утверждение может быть и неверно, если неверно выбран вид исходной регрес-

сии или не выполняются вышеперечисленные условия применимости регрессионного анализа.

Чтобы проверить точность оценки коэффициентов, записывают регрессию в виде $Y = b_0 + b_1X_1 + b_2X_2 + \dots$; подставляют в нее найденные оценки коэффициентов и известные значения аргументов, а затем вычисляют значения Y . Рассчитанные таким образом значения Y иногда обозначают символом \hat{Y} . Найденную величину \hat{Y} сопоставляют с $Y_{\text{эксп}}$, полученным опытным путем при тех же значениях аргументов. Чем меньше по модулю окажутся «остатки» или «невязки», т. е. разности вида $(\hat{Y} - Y_{\text{эксп}})$, тем, следовательно, лучше подобраны оценки b_i .

Существует несколько способов расчета коэффициентов b_i , выбор зависит от того, как именно оценивать близость значений \hat{Y} и $Y_{\text{эксп}}$. Основным является метод наименьших квадратов (МНК), предложенный в 1806 г. Гауссом и Лежандром. В этом случае коэффициенты выбирают так, чтобы *сумма квадратов невязок по всем значениям Y была бы минимальной*. Так, если проведено n независимых опытов, каждый из которых дает определенную величину $Y_{\text{эксп}}$, то неизвестные для нас коэффициенты b_i рассчитывают так, чтобы зависящая от них функция $F = \sum (\hat{Y} - Y_{\text{эксп}})^2$ имела бы минимальное значение. В этом случае ее частные производные по всем коэффициентам должны равняться 0:

$$\partial F / \partial b_0 = 0; \quad \partial F / \partial b_1 = 0; \quad \partial F / \partial b_2 = 0 \quad \text{и т. д.} \quad (9.1)$$

Из условия (9.1) выводятся конкретные формулы для расчета b_0 , b_1 , b_2 и других оценок коэффициентов. Формулы зависят от вида уравнения регрессии.

9.4. МНК ДЛЯ ЛИНЕЙНОЙ РЕГРЕССИИ В ОДНОФАКТОРНОМ ЭКСПЕРИМЕНТЕ

Пусть величина Y зависит от одного только аргумента X . Предполагаем линейную зависимость и записываем исходную регрессию в виде $Y = \beta_0 + \beta_1X$. Лучшими оценками коэффициентов будут такие b_0 и b_1 , при которых функция $F = \sum (b_0 + b_1X - Y_{\text{эксп}})^2$ имеет минимум, т. е. $\partial F / \partial b_0$

и $\partial F/\partial b_1$ равны 0. Суммирование ведется по всем опытам. После дифференцирования получаем:

$$\begin{aligned}\Sigma Y - nb_0 - b_1 \Sigma X &= 0; \\ \Sigma XY - b_0 \Sigma X - b_1 \Sigma X^2 &= 0.\end{aligned}$$

Эта система из двух уравнений первой степени с двумя неизвестными b_0 и b_1 легко решается подстановкой. Получаем:

$$b_0 = \frac{\Sigma X^2 \cdot \Sigma Y - \Sigma X \cdot \Sigma XY}{n \cdot \Sigma X^2 - (\Sigma X)^2}; \quad (9.2)$$

$$b_1 = \frac{n \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{n \cdot \Sigma X^2 - (\Sigma X)^2}. \quad (9.3)$$

Эти формулы химиками используются весьма часто, например в инструментальных методах анализа для расчета градуировочных зависимостей.

Пример 9-1. Расчет уравнения прямолинейного градуировочного графика. Проведено измерение оптической плотности серии растворов бензола в н-гексане. Известны значения концентрации бензола в каждом растворе (C , моль/л) и соответствующие им значения оптической плотности (D) (табл. 14).

Т а б л и ц а 14

Номер раствора	$C, 10^{-5}$ моль/л	D
1	2	0,12
2	4	0,20
3	6	0,28
4	8	0,40

Требуется рассчитать методом МНК уравнение градуировочного графика и найти по нему концентрацию исследуемого раствора, для которого в данных условиях $D = 0,34$. Предполагается, что результаты измерений подчиняются нормальному распределению.

Р е ш е н и е. Проводим кодировку, упрощая исходные данные. Пусть $Y = 100D$; $X = 10^5 C$. Запишем кодирован-

ные данные в таблицу 15. Последняя строка содержит необходимые для расчета значения сумм. Подставляем найденные суммы в формулы (9.2), (9.3):

$$b_0 = \frac{120 \cdot 100 - 20 \cdot 592}{4 \cdot 120 - (20)^2} = 2;$$

$$b_1 = \frac{4 \cdot 592 - 20 \cdot 100}{4 \cdot 120 - (20)^2} = 4,6.$$

Т а б л и ц а 15

Расчет уравнения градуировочного графика по методу МНК
(пример 9-1)

X	X^2	Y	XY	\hat{Y}
2	4	12	24	11,2
4	16	20	80	20,4
6	36	28	168	29,6
8	64	40	320	38,8
$\Sigma X = 20$	$\Sigma X^2 = 120$	$\Sigma Y = 100$	$\Sigma XY = 592$	

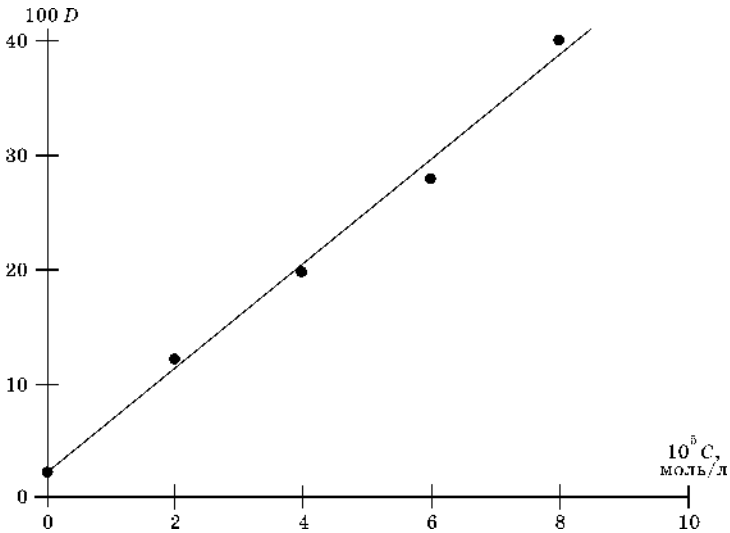


Рис. 15

Градуировочный график, соответствующий примеру 9-1

Запишем найденное уравнение регрессии в численном виде: $Y = 2 + 4,6X$. Вернемся к исходным переменным: $100D = 2 + 4,6 \cdot 10^5 C$, $D = 0,02 + 4,6 \cdot 10^3 C$. Это тоже уравнение регрессии, но уже не в кодированных, а в натуральных значениях переменных. Подстановка в это уравнение приведенных в условии задачи значений X дает набор значений \hat{Y} , приведенный в последнем столбце таблицы 15. Видно, что все значения близки к соответствующим Y (т. е. к $Y_{\text{эксп}}$). Обратите внимание, в двух случаях $Y > \hat{Y}$, а в двух других $Y < \hat{Y}$.

Используем полученное уравнение регрессии для определения концентрации бензола, приводящей в данных условиях к $D = 0,34$. Очевидно, при $D = 0,34$ $C = (0,34 - 0,02)/4,6 \cdot 10^3 = 6,957 \cdot 10^{-5} \approx 7,0 \cdot 10^{-5}$ (моль/л).

Очень удобно, когда значения X расположены с постоянным шагом, от опыта к опыту идет одинаковое приращение X . Делаем кодировку*: $X = (X - \bar{X})$.

После кодирования $\Sigma X = 0$ и тогда формулы для расчета b_0 и b_1 упрощаются:

$$b_0 = (\Sigma Y)/n; \quad (9.4)$$

$$b_1 = \Sigma XY / \Sigma X^2. \quad (9.5)$$

Если ожидается нелинейная зависимость Y от X , проводят линеаризацию путем логарифмирования, введения обратных величин и т. п. Например, для данных по потенциалу Са-селективного электрода при разных концентрациях ионов Ca^{2+} нет оснований ожидать линейной зависимости. Используя уравнение Нернста, выбираем следующий способ линеаризации: $X = \lg C$. Такой выбор позволяет записать регрессию в виде $Y = b_0 + b_1 X$, а затем использовать для расчетов простейший вариант МНК, т. е. формулы (9.2) и (9.3).

9.5.

МНК В МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТАХ

Формулы (9.2)–(9.5) пригодны только для простейших зависимостей вида $Y = \beta_0 + \beta_1 X$, т. е. для однофакторных экспериментов. Метод МНК применим и к многофакторным экспериментам, но формулы для оценки коэффици-

* В неочевидных случаях кодированные значения факторов выделены шрифтом.

ентов будут гораздо сложнее. Дело в том, что при обработке результатов многофакторных экспериментов оценки коэффициентов обычно оказываются зависимыми друг от друга. Например, для регрессии $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2$ в формулу для расчета оценки b_0 входит величина b_1 , а чтобы рассчитать последнюю, надо знать b_0 ! Однако в некоторых случаях оценки коэффициентов независимы друг от друга, сколько бы их ни было. Чтобы это стало возможным, надо специально спланировать эксперимент: менять от опыта к опыту значения всех факторов с постоянным шагом и, самое главное, закодировать факторы перед расчетом регрессии так, чтобы выполнялось условие *ортогональности*. Это значит, что просуммировав по всем опытам произведение двух любых «иксов» (например, X_1 и X_2), мы должны получить 0. Тогда применимы формулы

$$b_i = \frac{\sum X_i Y_i}{\sum X_i^2}; \quad (9.6)$$

$$b_{ij} = \frac{\sum X_{ij} Y_i}{\sum X_{ij}^2}, \quad (9.7)$$

где суммирование ведется по всем опытам. Для расчета b_0 в формулу (9.6) подставляем фиктивную переменную X_0 , всегда равную единице. Для расчета коэффициента b_{ij} , стоящего у произведения разных «иксов», в формуле (9.7) в качестве фиктивного X берут это произведение, т. е. $X_{ij} = X_i X_j$.

Пример 9-2. Скорость коррозии (V , г/л ч) измерена при трех значениях pH и двух температурах (T , °C) в шести опытах. Данные приведены в таблице 16. Требуется рассчитать по методу МНК регрессию вида $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

С учетом условия ортогональности кодируем: $X_1 = \text{pH} - 5$; $X_2 = 0,1(T - 30)$. Для упрощения записи возьмем $Y = 10^3 V$. Матрицу данных после кодировки проверяем на ортогональность. Условия $\sum X_1 X_2 = 0$; $\sum X_0 X_1 = 0$; $\sum X_0 X_2 = 0$ в данном случае выполняются. Значит, оценки коэффициентов не должны зависеть друг от друга, что позволяет применять формулы (9.6), (9.7):

$$b_0 = \Sigma(X_0 Y) / \Sigma X_0^2 = \Sigma Y / n = 60 / 6 = 10;$$

$$b_1 = \Sigma(X_1 Y) / \Sigma X_1^2 = -40 / 16 = -2,5;$$

$$b_2 = \Sigma(X_2 Y) / \Sigma X_2^2 = 15 / 6 = 2,5;$$

$$Y = 10 - 2,5X_1 + 2,5X_2.$$

Раскодируем:

$$10^3 V = 10 - 2,5(\text{pH} - 5) + 2,5 \cdot 0,1 \cdot (T - 30) = \\ = 15 - 2,5\text{pH} + 0,25T;$$

$$V = 0,015 - 0,0025\text{pH} + 0,00025T.$$

Последнее уравнение и есть искомое уравнение регрессии, выраженное в натуральных переменных. После подстановки в него указанных в условии значений всех факторов убеждаемся в довольно хорошем совпадении всех \hat{Y} с экспериментальными данными (значениями Y).

Т а б л и ц а 16

Исходные данные и их кодировка
для расчета коэффициентов регрессии

Условие задачи			После кодировки			Проверка
pH	T	V	X_1	X_2	Y	\hat{Y}
3	20	0,012	-2	-1	12	12,5
5	20	0,008	0	-1	8	7,5
7	20	0,0025	2	-1	2,5	2,5
3	40	0,018	-2	1	18	17,5
5	40	0,012	0	1	12	12,5
7	40	0,0075	2	1	7,5	7,5

Как видно из таблицы 16, «невязки» модели не превышают 0,5 единиц (порядка 5% от значений параметра). Если бы мы искали решение в виде полной линейной регрессии, т. е. старались отыскать еще и слагаемое, содержащее произведение $X_1 X_2$, то пришли бы к уравнению:

$$Y = 10 - 2,5X_1 + 2,5X_2 - 0,125X_1X_2 \\ \text{(в кодированных переменных).}$$

В этом случае расчетные значения \hat{Y} совпали бы с экспериментом еще точнее.

9.6. ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Часто оценки коэффициентов близки к 0, что указывает на слабое влияние соответствующих факторов. Для проверки значимости этих коэффициентов найденные значения b_i сопоставляют с погрешностями их определения (обозначаются как Δb_i). Погрешности рассчитывают из данных по воспроизводимости величины Y при дублировании опытов в одних и тех же условиях, при неизменности всех факторов. Пусть проведено n независимых опытов, а каждый из них дублировали m раз. Для каждого опыта находят дисперсию: $S_m^2 = \sum (Y - \hat{Y})^2 / (m - 1)$. Усреднение таких дисперсий по всем опытам (если дисперсии однородны) дает *дисперсию воспроизводимости* $S_{mn}^2 = (\sum S_m^2) / n$. Это очень важная величина, которая используется не только при проверке значимости коэффициентов регрессии, но и во многих других случаях. Число степеней свободы для этой дисперсии равно $n(m - 1)$. Если мы работаем с однофакторной регрессией вида $Y = b_0 + b_1 X$ или многофакторной линейной регрессией при ортогональном плане эксперимента, то погрешность коэффициентов легко рассчитать, исходя из S_{mn}^2 :

$$\Delta b_i = t S_{b_i}, \quad (9.8)$$

где $(S_{b_i})^2 = S_{mn}^2 / \sum X_i^2$.

Величину t в этой формуле берут с $n(m - 1)$ степенями свободы для выбранной надежности P . Иногда дублируют не все опыты, а лишь какой-то один (в центре плана). В этом случае $S_{mn}^2 \approx S_m^2$, а число степеней свободы равно $(m - 1)$.

В отдельных руководствах (например, [28]) рекомендуют проверять значимость коэффициентов регрессии по отклонению отдельных Y от рассчитанной зависимости вида $Y = f(X)$, т. е. без учета воспроизводимости результатов при дублировании измерений. Такой способ расчета Δb_i очень прост, но вызывает возражения, поэтому здесь не излагается. Лучше пользоваться описанным выше способом расчета Δb_i по S_{mn}^2 . Однако, каким бы способом ни

рассчитывалась погрешность Δb_i , для оценки значимости коэффициента проверяют: выполняется ли условие $|b_i| > \Delta b_i$. Если не выполняется, коэффициент b_i считают незначимым и соответствующее слагаемое из регрессии исключают.

9.7. ПРОВЕРКА АДЕКВАТНОСТИ РЕГРЕССИИ

Регрессия считается адекватной, если разброс между результатами параллельных опытов при неизменных значениях X будет того же порядка или больше, чем «невязки» — расхождения между предсказанными величинами \hat{Y} и соответствующими экспериментальными данными (Y). Разброс результатов параллельных опытов оценивается дисперсией воспроизводимости, невязки — дисперсией адекватности. Сопоставление двух дисперсий проводится по критерию Фишера. Алгоритм расчета дисперсии воспроизводимости описан в предыдущем разделе. Дисперсия адекватности рассчитывается, исходя из невязок.

$$S_{ad}^2 = \frac{\sum (Y - \hat{Y})^2}{n - k}. \quad (9.9)$$

Число степеней свободы для дисперсии адекватности равно числу опытов (n) минус число оцениваемых коэффициентов (k). В числителе обязательно ставится дисперсия адекватности.

$$F_{\text{экс}} = S_{ad}^2 / S_{mn}^2.$$

Если $F_{\text{экс}} > F_{\text{табл}}$, то регрессия на выбранном уровне значимости считается неадекватной. Если дисперсия неадекватна, а все расчеты верны, начинают расчет заново, но уже для более сложной регрессии, например, вместо неполной регрессии берут полную или регрессию второго порядка. Без дублирования опытов проверить адекватность регрессии невозможно. Не удастся проверить адекватность и в тех случаях, когда число опытов равно числу рассчитываемых коэффициентов ($n = k$), в этом случае формула (9.9) неприменима.

9.8. РЕГРЕССИОННЫЙ АНАЛИЗ В СЛУЧАЕ ПФЭ ВИДА 2^k

Коэффициенты регрессии особенно легко оценить по данным правильно спланированного полного факторного эксперимента вида 2^k (см. главу 2). Расчет проводится для регрессии, включающей кодированные (безразмерные) значения факторов. Схема ПФЭ ортогональна, а в этом случае все оценки вида b_i рассчитываются независимо друг от друга, какими бы ни были число коэффициентов и сложность модели. Так как в кодированных координатах факторы принимают только два значения: $(+1)$ или (-1) , то выполняется условие $\sum X_i^2 = n$. В этом случае формула (9.6) упрощается, и при расчете любого коэффициента знаменатель одинаков и равен n — числу опытов в данном эксперименте:

$$b_i = \frac{\sum X_i Y}{n}, \quad n = 2^k. \quad (9.10)$$

Как и всегда в регрессионном анализе, при расчете b_0 вводим фиктивную переменную $X_0 = 1$, а при расчете всех b_{ij} -коэффициентов перед произведениями факторов вводим фиктивные переменные $X_{ij} = X_i \cdot X_j$. На практике расчет сводится к суммированию значений Y по всем опытам, но знак перед каждым слагаемым берут из соответствующего столбца в матрице эксперимента. В качестве примера проведем оценку коэффициентов для моделирования радиационной полимеризации в трехфакторном эксперименте, план которого мы формировали в рамках примера 2-2. Результаты проведенных опытов приведены в таблице 17. Значения факторов приведены в кодированном виде.

Предполагаем, что процесс полимеризации описывается регрессией вида:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \\ + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3.$$

Результаты ПФЭ по радиационной полимеризации (пример 2-2)

№ опыта	X_0	X_1	X_2	X_3	X_{12}	X_{13}	X_{23}	X_{123}	Y	\hat{Y}
1	+	—	—	—	+	+	+	—	13	13
2	+	+	—	—	—	—	+	+	11	11
3	+	—	+	—	—	+	—	+	17	17
4	+	+	+	—	+	—	—	—	15	15
5	+	—	—	+	+	—	—	+	29	29
6	+	+	—	+	—	+	—	—	39	39
7	+	—	+	+	—	—	+	—	13	13
8	+	+	+	+	+	+	+	+	23	23

Оценки коэффициентов в соответствии с формулой (9.10) равны:

$$b_0 = \frac{\sum X_0 Y}{n} = \sum Y / 8 = (13 + 11 + 17 + 15 + 29 + 39 + 13 + 23) / 8 = 20;$$

$$b_1 = \frac{\sum X_1 Y}{n} = (-13 + 11 - 17 + 15 - 29 + 39 - 13 + 23) / 8 = 2;$$

$$b_2 = \frac{\sum X_2 Y}{n} = (-13 - 11 + 17 + 15 - 29 - 39 + 13 + 23) / 8 = -3;$$

$$b_3 = \frac{\sum X_3 Y}{n} = (-13 - 11 - 17 - 15 + 29 + 39 + 13 + 23) / 8 = 6;$$

$$b_{12} = \frac{\sum X_{12} Y}{n} = (13 - 11 - 17 + 15 + 29 - 39 - 13 + 23) / 8 = 0.$$

Аналогично получим $b_{13} = 3$; $b_{23} = -5$; $b_{123} = 0$. Обратите внимание, при суммировании знаки слагаемых чередуются так же, как знаки в соответствующем столбце исходного плана (для b_0 — только плюсы, для b_1 знаки плюс и минус чередуются, для b_2 они чередуются попарно, и т. д.). Если коэффициенты получаются не целочисленными (а так обычно и бывает), то их следует округлить с учетом точности измерения Y . В данном случае подставляем найденные оценки в исходное уравнение без округления, причем слагаемые, где коэффициент оказался равен нулю, не записываем. Получаем регрессию:

$$Y = 20 + 2X_1 - 3X_2 + 6X_3 + 3X_1X_3 - 5X_2X_3. \quad (9.11)$$

Для проверки производим подстановку и вычисляем \hat{Y} по каждому опыту. Эти значения приведены в последней колонке таблицы. Вычисленные и экспериментальные значения параметра в этом примере идеально совпали. На практике идеального совпадения обычно не бывает.

Для отбора статистически значимых коэффициентов обычно пользуются формулой (9.8), причем при обработке результатов ПФЭ она упрощается:

$$\Delta b_i = t S_{bi}; \quad (S_{bi})^2 = S_{mn}^2 / n. \quad (9.12)$$

Но в таблице 17 данные по воспроизводимости опытов отсутствовали, использовать формулу (9.8) мы не можем и все ненулевые коэффициенты в уравнении регрессии (9.11) считаем значимыми.

В некоторых случаях удобнее использовать регрессии, выраженные в натуральных значениях переменных (*интерполяционные формулы*, см. главу 10). Вспомним, что для любого фактора в кодированных переменных верхний уровень, равный единице, соответствовал центру плана (X_0) плюс интервал варьирования (ΔI). В примере 2-2 центр плана по концентрации мономера (C_0) соответствовал значению 20 г/л при интервале варьирования 10 г/л, откуда получаем $X_1 = (C - 20)/10$. Аналогично: $X_2 = (t - 50)/15$; $X_3 = (I - 5000)/500$. После подстановки этих выражений в ранее найденное уравнение (9.11) и простейших преобразований получаем регрессию, выраженную в натуральных переменных, т. е. интерполяционную формулу. Ее удобно использовать для прогнозирования результатов радиационной полимеризации в любых условиях. Адекватность исходной регрессии позволяет интерпретировать ее или применить для оптимизации исследуемого процесса.

9.9. МОДЕЛЬ НЕАДЕКВАТНА: ЧТО ДЕЛАТЬ?

Первое, что нужно делать в таких случаях — еще раз проверить ход вычислений, проводившихся для оценки коэффициентов и в ходе подстановки. Не думайте, что если расчеты проведены по стандартной программе на ЭВМ,

то ошибок не может быть. Экспериментатор мог неверно ввести данные или неверно понять машинный ответ, возможен и сбой программы. Если какие-то коэффициенты были исключены как статистически незначимые, а потом была обнаружена неадекватность, стоит проверить адекватность регрессии с неисключенными коэффициентами; ведь ошибка могла быть допущена при оценке значимости коэффициентов. Вторая возможная причина неадекватности математической модели связана с неверной оценкой воспроизводимости эксперимента. Если мы сочтем методу измерения Y более точной, чем она есть на самом деле, то дисперсия воспроизводимости будет занижена, и в сравнении с ней даже небольшие расхождения между моделью и экспериментальными данными приведут к выводу о неадекватности модели. Поэтому очень важно еще раз *проверить разброс Y в параллельных опытах*. Однако недостаточно просто повторить несколько раз подряд один и тот же опыт. Возможно, флуктуации неконтролируемых факторов (серьезнейший источник ошибок) проявляются медленно, в этом случае невоспроизводимость Y можно обнаружить только при повторении опытов с интервалом в несколько дней. Проверку следует проводить в разных условиях, так как невоспроизводимость может проявляться не во всех областях факторного пространства. Следовательно, повторная проверка воспроизводимости требует серьезного дополнительного исследования.

Еще одна возможная причина неадекватности — охват в ходе ПФЭ слишком большой области внутри факторного пространства и, соответственно, слишком большие интервалы варьирования. В этом случае следует повторить эксперимент (часть уже сделанных опытов используется, а другие проводятся заново) и рассчитать новое уравнение регрессии. Изменение интервалов варьирования часто сочетают с переносом центра плана в новую точку, ближе к оптимальному значению параметра.

Неадекватность регрессии в отсутствие арифметических ошибок часто означает, что мы «обсчитывали» данные, используя слишком простую модель. Например, в качестве модели выбрали линейную регрессию, без оценки эффектов взаимодействия, коэффициенты при произведениях факторов не вычисляли, а на самом деле факторы

действуют взаимосвязанно, что и привело к неадекватности линейной модели. В этом случае надо дополнительно оценить коэффициенты при слагаемых, описывающих парные или тройные взаимодействия, а затем ввести дополнительные слагаемые в уравнение регрессии. Если после расчета дополнительных коэффициентов нам не хватит степеней свободы для оценки дисперсии адекватности, то придется провести еще несколько независимых опытов.

Кроме того, факторы изначально могли быть неудачно выбраны. В частности, некоторые факторы могли быть связаны между собой, например, линейно зависеть друг от друга. Эта ситуация называется *мультиколлинеарностью* факторов. При неудачном выборе факторов их совместное использование не эффективно и нередко приводит к неадекватным моделям.

Для наиболее сложных объектов исследования и при достаточно большом числе проведенных опытов можно уточнить неадекватную модель, рассчитывая дополнительные слагаемые, содержащие квадраты факторов. Полученная нелинейная регрессия, скорее всего, окажется адекватной. Возможна также достройка ПФЭ до плана второго порядка и проведение дополнительных опытов, алгоритмы этих операций и последующих расчетов описаны в литературе ([29] и др.).

Глава 10

ИНТЕРПРЕТАЦИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ. ОПТИМИЗАЦИЯ ИССЛЕДУЕМОГО ПРОЦЕССА

10.1. ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

Интерпретация статистической модели фактически означает перевод полученных в ходе эксперимента данных и связывающих их уравнений с математического языка на специфический язык, привычный для химика. Выводы из проведенного эксперимента должны содержать иной набор терминов, не тот, который применялся при выводе уравнения регрессии. Речь должна идти не о дисперсиях и выборках, а о концентрациях и температурах, скоростях и выходах реакций, активности и селективности катализатора, и т. п. Интерпретировать надо математическую модель (регрессию), выраженную в кодированных переменных (в безразмерных «иксах»). После перевода регрессии в новое уравнение, связывающее между собой факторы и параметры в их натуральной форме, разобраться в значимости факторов будет гораздо труднее. Для интерпретации математической модели следует рассматривать знаки и абсолютные величины коэффициентов в уравнении регрессии. При этом стоит учесть следующие рекомендации.

1. Все факторы, которые после исключения незначимых коэффициентов вошли в адекватное уравнение регрессии, следует считать достоверно влияющими на исследуемый процесс. Например, в результате изучения влияния трех факторов — температуры, концентрации реагента и pH раствора — на выход некоторого продукта реакции получили линейную регрессию, которая после исключения незначимых коэффициентов и соответствующих им слагаемых приобрела вид:

$$Y = 1,25 + 2,32X_1 + 1,3X_3. \quad (10.1)$$

Здесь переменным X_1 и X_3 соответствуют температура и pH реакционной смеси. Оба фактора следует считать достоверно влияющими на выход продукта. Вывод из проведенного исследования можно сформулировать следующим образом: «Выход продукта растет при повышении температуры и pH раствора. Влияние концентрации реагента на выход продукта не выявлено».

2. Нередко фактор, который по интуитивному мнению исследователя или по теоретическим соображениям просто «обязан» влиять на исследуемый процесс, оказывается незначимым и не попадает в конечную форму уравнения регрессии. Примером может быть не выявленное уравнением (10.1) влияние температуры. Обычно такие итоги обработки данных говорят о неверном выборе интервала варьирования. Чтобы еще раз проверить влияние «сомнительного» фактора, следует повторить эксперимент, увеличив в несколько раз соответствующий интервал варьирования.

3. Получив в ходе полного факторного эксперимента адекватную математическую модель, не стоит торопиться и публиковать окончательные выводы. Возможно, вы еще не закончили эксперимент! Например, может оказаться, что регрессия адекватна, хотя все ее коэффициенты, кроме b_0 , незначимы. Такой случай не исключен, но маловероятен. Не следует думать, что действительно ни один из факторов не влияет на исследуемый процесс. Скорее всего, такой исход исследования связан с плохой воспроизводимостью опытов (а может быть и с выбором слишком маленьких интервалов варьирования). Эксперимент надо

продолжить, перейдя к более воспроизводимой методике измерений параметра и увеличивая интервалы варьирования факторов. Если коэффициенты снова окажутся незначимыми, это будет означать, что эксперимент изначально проводился в области факторного пространства, близкой к оптимуму, где поверхность отклика почти плоская. Найдя такую область, можно прекратить эксперимент, либо попробовать рассчитать новую модель, пользуясь планом второго порядка.

4. Исследователю всегда важно знать, имеется ли достоверное взаимодействие факторов или они действуют на параметр по отдельности, причем аддитивно. В первом случае коэффициенты при каких-то произведениях «иксов» в уравнении регрессии будут значимыми, во втором случае все такие коэффициенты будут незначимыми. Если коэффициент b_{ij} незначим, значит, действие i -го фактора не зависит от того, на каком уровне находится j -й фактор. В качестве примера возьмем уравнение регрессии, выведенное в разделе 9.8 и описывающее процесс радиационной полимеризации:

$$Y = 20 + 2X_1 - 3X_2 + 6X_3 + 3X_1X_3 - 5X_2X_3. \quad (10.2)$$

Видно, что при постоянном режиме облучения пробы (при неизменном X_3) концентрация мономера (фактор X_1) и температура раствора (фактор X_2) влияют на выход полимера (параметр Y) независимо друг от друга. Об этом говорит незначимость коэффициента b_{12} и отсутствие в (10.2) других слагаемых, содержащих произведение X_1X_2 . Переведем на химический язык: изменения температуры должны одинаково влиять на процесс полимеризации в разбавленных и концентрированных растворах мономера. Это весьма важно.

Из формулы (10.2) следует также, что влияние радиации (X_3) должно сильно зависеть от условий проведения процесса (от концентрации и температуры), так как коэффициенты b_{13} и b_{23} значимы. Значимость эффектов взаимодействий существенна как для проверки гипотез о механизме процесса, так и для оптимизации условий проведения этого процесса. Если коэффициенты вида b_{ij} незначимы, т.е. взаимодействие факторов не выявлено, *знак b_i в*

уравнении регрессии однозначно свидетельствует о характере влияния i -го фактора. Очень удобно, что знаки коэффициентов не зависят от выбора интервалов варьирования, в отличие от абсолютной величины коэффициентов. Именно знаки коэффициентов важны для выработки рекомендаций по управлению процессом.

5. Для сравнения влияния разных факторов часто сопоставляют абсолютную величину коэффициентов (значения b_i). В принципе это возможно, и такое сопоставление может дать правильные и весьма полезные для практики выводы, однако обязательно следует учесть ряд ограничений. Во-первых, *сопоставление коэффициентов нельзя проводить по регрессии, выраженной в натуральных переменных*, так как в этом случае не выполняется условие ортогональности, оценки коэффициентов не являются независимыми. Во-вторых, такой подход возможен только *при независимом действии всех факторов* (при линейности модели). Так, в вышеприведенном примере, связанном с радиационной полимеризацией, нельзя утверждать, что мощность дозы влияет на выход полимера сильнее, чем температура раствора, так как действие этих факторов взаимосвязано! В-третьих, абсолютная величина коэффициента регрессии зависит от выбранного интервала варьирования, и чем больше взять ΔI_i , тем большим окажется коэффициент b_i , независимо от действительной «силы» i -го фактора. В принципе, подобрав подходящие интервалы варьирования, всегда можно получить любой желаемый вывод! Сопоставление «силы» факторов по величине b_i правомерно, если факторы однотипны (например, в качестве факторов используются молярные концентрации разных реагентов), а интервалы варьирования этих факторов одинаковы. Но если факторы разнородны, и их числовые значения выражены в разных единицах, интервалы варьирования не могут быть одинаковыми. Поэтому правомерность сопоставления степени влияния разнородных факторов по величине b_i сомнительна даже в случае линейной модели. Сопоставляя влияние разнородных факторов по значениям регрессионным коэффициентов, исследователи не раз приходили к ошибочным выводам.

10.2. ПРИМЕНЕНИЕ АДЕКВАТНОЙ МОДЕЛИ

Можно указать три основных направления. Это: 1) оптимизация условий; 2) прогнозирование параметра для заданных условий (расчет по интерполяционной формуле); 3) выявление ранее неизвестных закономерностей, например, с целью установления механизма процесса.

Оптимизация условий проведения процесса с помощью математической модели основывается на рассмотрении знаков и абсолютной величины коэффициентов в уравнении регрессии. При этом могут быть два случая: а) исследователь стремится достичь максимального значения параметра Y ; б) исследователь стремится минимизировать этот параметр. В случае линейных регрессий, не содержащих произведений факторов, дать соответствующие рекомендации несложно. А именно, для увеличения параметра Y надо увеличивать те факторы, для которых коэффициент b_i положителен, и уменьшать те, для которых он отрицателен. Для минимизации параметра Y следует поступать наоборот.

Во многих регрессиях, содержащих произведения факторов, можно пренебречь влиянием факторов, взятых порознь. Тогда надо обратить внимание на знак эффекта взаимодействия (знак коэффициента b_{ij}). Если он положителен, для достижения максимального значения параметра следует либо одновременно повышать, либо одновременно понижать оба фактора. Более детально этот вопрос рассмотрен в разделе 10.3. Какой именно из двух этих рекомендаций воспользоваться на практике, зависит от многих обстоятельств, в частности, от технических возможностей управления исследуемым процессом.

По адекватному уравнению регрессии можно *прогнозировать параметр для любых условий проведения процесса*, для любых комбинаций факторов. Однако гораздо проще использовать для этого уравнение регрессии, включающее не кодированные (безразмерные) переменные, а натуральные значения факторов, выраженные в соответствующих единицах измерения. Такие уравнения называют *интерполяционными формулами*. Для перехода к интерполяционной формуле в первоначально получен-

ное уравнение регрессии подставляют выражения, связывающие безразмерные «иксы» с интервалами варьирования и значениями соответствующего фактора в центре плана, т. е. формулы типа (2.10).

В качестве примера рассмотрим статистическую модель экстракционного извлечения ионов меди из водного раствора в неводную фазу. Медь экстрагируется в виде некоего комплексного соединения. Допустим, двухфакторный ПФЭ привел к адекватной регрессии, в которой факторы выражены в кодированном виде:

$$Y = 97 - 1,5X_1 + 4,2X_2 + 3,7X_1X_2. \quad (10.3)$$

Здесь значения X_1 соответствуют начальной концентрации меди (мг/л) в водном растворе, но не равны им. Значения X_2 соответствуют значениям pH раствора, а Y — значениям коэффициента распределения меди (заметим, что для процесса экстракции это не единственный из возможных параметров оптимизации). Судя по уравнению (10.3), по мере увеличения pH экстрагируемость металла должна возрастать. Влияние концентрации металла неоднозначно, так как знаки коэффициентов b_1 и b_{12} различны*. Однако эти соображения следует проверять числовыми расчетами, а для этого придется «раскодировать» полученное в ПФЭ уравнение регрессии.

Допустим, центр плана ПФЭ имел координаты: $C_0 = 100$; $pH_0 = 7$; интервалы варьирования равны: $\Delta C = 20$ мг/л, $\Delta pH = 2$. Для раскодирования используем формулы, конкретизирующие общую формулу (2.10):

$$X_1 = (C - 100)/20; \quad X_2 = (pH - 7)/2.$$

Делаем подстановку:

$$D = 97 - 1,5(C - 100)/20 + 4,2(pH - 7)/2 + 3,7(C - 100)(pH - 7)/2 \cdot 20.$$

После несложных преобразований получаем:

$$D = 154,55 - 0,7225C - 7,15pH + 0,0925pH \cdot C. \quad (10.4)$$

* Теоретически коэффициент распределения не должен зависеть от концентрации экстрагируемого вещества, а на практике такое влияние нередко имеется, что связано с протеканием побочных процессов.

Тут надо обязательно провести промежуточную проверку. Подстановка в (10.4) координат центра плана приводит к $D = 97$, что соответствует величине b_0 в исходной регрессии и подтверждает правильность сделанных преобразований.

Интерполяционная формула (10.4) — удобное средство вычисления любой из трех переменных (D , C , pH) по значениям двух других. Интересно, что при pH стоит знак минус, а не плюс, как в уравнении (10.3), т. е. после раскодирования знак изменился! Это связано с нелинейностью модели.

По интерполяционной формуле можно рассчитать оптимальное значение одного из факторов (т. е. такое, при котором мы достигнем желаемой величины параметра), надо лишь задать некоторые значения других факторов, рассмотреть интересующие исследователя частные случаи. В качестве примера рассчитаем по (10.4), при каком pH следует проводить экстракцию меди, если ее концентрация будет равна 100 мг/л. После подстановки $C = 100$ получаем упрощенную формулу:

$$D = 82,3 + 2,1pH. \quad (10.5)$$

Судя по формуле (10.5), извлечение меди на указанном концентрационном уровне вполне возможно при любых реальных значениях pH (от 0 до 14), так как величина коэффициента распределения будет больше единицы. Очевидно также, что экстракционное извлечение меди будет идти тем лучше, чем больше pH . Однако при подобных расчетах следует обязательно учитывать области допустимых значений каждого фактора. Так, при высоких pH (например, при $pH = 10$) формула (10.5) явно будет неадекватна. Дело в том, что при выводе исходного уравнения (10.3) не было учтено, что в сильнощелочной среде ионы меди не экстрагируются, а связываются в гидроксид меди и переходят в осадок. Это ограничивает область допустимых значений pH .

Можно получить упрощенные интерполяционные формулы другого типа, связывающие коэффициент распределения с величиной начальной концентрации раствора при

постоянном значении рН. Так, при $\text{pH} = 5$ регрессия (10.4) преобразуется в следующее выражение:

$$D = 118,8 - 0,26C. \quad (10.6)$$

Это уравнение показывает, что экстрагируемость меди из разбавленных слабокислых растворов ($\text{pH} = 5$) слабо зависит от концентрации. Однако для концентрированных растворов (например, для $C = 1000$ мг/л) мы получили бы отрицательные значения D , что в принципе невозможно. Значит, здесь снова исследователь выходит за пределы области допустимых значений фактора, модель (10.3) явно непригодна для описания процесса экстракции меди из концентрированных растворов.

Следует учесть, что при использовании математических моделей для прогнозирования параметра для заданных условий, а также при решении обратных задач, следует отдельно рассматривать два варианта прогнозирования: интерполяцию и экстраполяцию. При интерполяции мы пользуемся моделью в той области факторного пространства, для которой эта модель была получена; при экстраполяции прогнозы относятся к совершенно другим областям факторного пространства. Так, для детей масса тела приблизительно линейно связана с возрастом. Основываясь на этой модели, можно прогнозировать массу тела. Допустим, средняя масса тела новорожденного 3 кг, двухлетнего ребенка — 13 кг, четырехлетнего — 23 кг (пример условен!). Легко построив по трем точкам линейную модель, можно прогнозировать, что средняя масса тела трехлетнего ребенка должна быть равной 18 кг. Это — типичная интерполяция, и она дает довольно точные прогнозы. А вот прогноз массы 30-летнего «ребенка» по той же формуле даст совершенно невероятное значение — 203 кг!

Следует запомнить, что экстраполяционные прогнозы возможны, но всегда сомнительны. Чем дальше в факторном пространстве отстоят условия, для которых проводится прогноз, от центра исходного плана, тем выше вероятность ошибок, тем больше величина невязок.

Хотя статистические модели создаются в основном для оптимизации и прогнозирования, эти модели и, в частности, уравнения регрессий важны и для выявления теоре-

тических закономерностей, характеризующих исследуемые объекты. В частности, уравнение (10.3) может быть полезно для выяснения механизма экстракции меди. А именно, неаддитивное влияние факторов ($b_{12} > 0$) указывает на сложный механизм процесса, не соответствующий обычным закономерностям экстракции моноядерных комплексов металлов. Такое влияние характерно для образования полиядерных комплексов. Сильное влияние pH на коэффициент распределения показывает, что важное место в механизме экстракции занимают кислотно-основные реакции. В частности, можно предположить, что при подкислении происходит разрушение экстрагируемого комплекса, что объяснило бы формулу (10.5). Таким образом, статистическая модель подталкивает исследователя к выдвижению новых научных гипотез, которые следует дополнительно проверить. Важно, чтобы теоретические модели, созданные на основании этих гипотез, согласовывались с уже полученными статистическими моделями, т. е. с результатами эксперимента.

Приведенные выше примеры не исчерпывают всех способов использования математических моделей, получаемых по результатам полного факторного эксперимента, но этих примеров достаточно, чтобы понять, как много полезной информации может «выжать» из адекватной регрессии компетентный исследователь-химик.

10.3. МЕТОД КРУТОГО ВОСХОЖДЕНИЯ КАК СПОСОБ ОПТИМИЗАЦИИ

Как уже отмечалось, целью прикладных исследований часто является оптимизация исследуемого процесса. Например, повышение выхода некоторого продукта в ходе его синтеза, повышение прочности выпускаемого материала, увеличение точности анализа, снижение скорости коррозии, снижение содержания вредных примесей и уменьшение их выброса в окружающую среду, и т. п. Требуется найти ту область факторного пространства, те условия проведения процесса, которые отвечают оптимальному значению параметра. Наиболее короткий и надежный путь к оптимуму, к искомому пику на поверхности

отклика, это движение по *градиенту функции отклика*. Для непрерывной и однозначной функции $y = f(x_1, x_2, \dots, x_k)$ градиентом называют вектор, включающий в качестве слагаемых частные производные функции по каждому из аргументов (факторов), умноженные на единичные векторы в направлении координатных осей. Оценками частных производных для функции отклика могут быть коэффициенты в уравнении регрессии. Найденное в ходе ПФЭ уравнение регрессии показывает, как именно надо одновременно менять значения всех факторов, чтобы двигаться по поверхности отклика в сторону оптимума, прямо к вершине, по самому короткому и крутому пути (рис. 16). Эта идея, предложенная Боксом и Уилсоном, легла в основу «метода крутого восхождения».

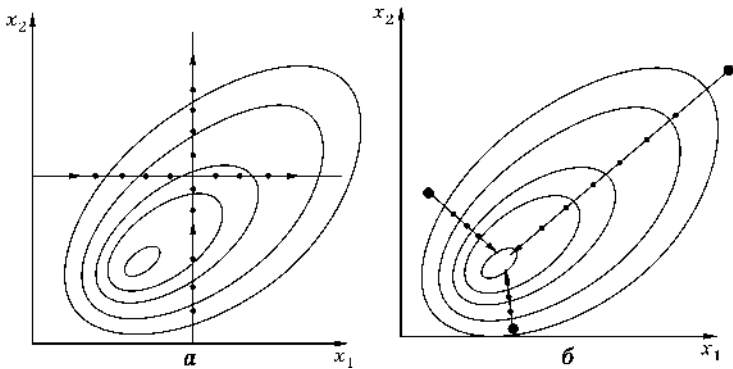


Рис. 16

Традиционная оптимизация (а) и оптимизация по методу крутого восхождения (б) для двухфакторного процесса

Если взаимодействиями факторов можно пренебречь, выйти на оптимум довольно просто. Исходное уравнение регрессии в таких случаях записывается в виде:

$$Y = b_0 + b_1X_1 + \dots + b_kX_k.$$

Градиент функции отклика:

$$\Delta Y = \frac{\partial Y}{\partial X_1}i_1 + \frac{\partial Y}{\partial X_2}i_2 + \dots + \frac{\partial Y}{\partial X_k}i_k = b_1i_1 + b_2i_2 + \dots + b_ki_k. \quad (10.7)$$

Здесь i_1 — единичный вектор по первому фактору в направлении координатной оси, аналогичные обозначения используются и по другим факторам. Из (10.7) следует, что направление градиента на плоской поверхности отклика однозначно задается значениями коэффициентов регрессии. Для исходного (кодированного) уравнения регрессии все i равны 1, формула (10.7) упрощается:

$$\Delta Y = b_1 + b_2 + \dots + b_k. \quad (10.8)$$

Изменяя значения факторов прямо пропорционально величинам найденных для них коэффициентов регрессии (с учетом их знаков), можно осуществить движение по поверхности отклика по направлению градиента, т. е. двигаться прямо к оптимуму.

Предположим, что мы проводили двухфакторный эксперимент и получили адекватную регрессию $Y = 20 + 2X_1 + 5X_2$. Также предположим, что для оптимизации процесса мы хотим добиться максимально возможного значения Y . Тогда надо одновременно увеличивать и X_1 , и X_2 , но увеличение второго фактора должно идти в большей мере! Лучше, если численные значения коэффициентов регрессии будут величинами одного порядка (симметричные оценки, симметричная функция отклика), в методе крутого восхождения это улучшает точность прогнозов.

Для определения условий новых опытов, соответствующих движению по градиенту, обычно используют уравнение регрессии, выраженное в натуральных переменных (интерполяционную формулу). В этом случае следует, в соответствии с (10.7), учитывать не только коэффициенты регрессии, но и величины интервалов варьирования по каждому фактору (выбранные при планировании эксперимента значения ΔI_i , см. главу 2). Условия каждого нового опыта должны отвечать следующему условию:

$$\begin{aligned} (X_1)_{\text{новый}} &= (X_1)_{\text{старый}} + \Delta X_1; \\ (X_2)_{\text{новый}} &= (X_2)_{\text{старый}} + \Delta X_2, \text{ и т. д.} \end{aligned}$$

На практике исследователь задает ΔX_1 — шаг по первому фактору — по своему усмотрению, а затем рассчитывает ΔX_2 по пропорции:

$$\Delta X_2 / \Delta X_1 = b_2 \Delta I_2 / b_1 \Delta I_1. \quad (10.9)$$

Так же рассчитывается величина шага по третьему и по остальным факторам:

$$\Delta X_3 / \Delta X_1 = b_3 \Delta I_3 / b_1 \Delta I_1. \quad (10.10)$$

Естественно, шаг по любому фактору должен быть гораздо больше погрешности этого фактора.

Начиная движение по градиенту, обычно исходят из того опыта в проведенном ПФЭ, в котором достигнуто наибольшее значение Y (это необязательно, можно идти и от центра плана). Например, от начальной точки на поверхности отклика сдвигаются к оптимуму, увеличив в соответствии с уравнением (10.7) величину фактора X_1 (в натуральных переменных) на два интервала варьирования по этому фактору, а величину фактора X_2 — на пять его интервалов варьирования. Далее можно сделать и второй шаг в том же направлении — к полученному после первого шага значению X_1 еще раз добавить два ΔI_1 , а к полученному значению X_2 добавить пять ΔI_2 . Можно двигаться и меньшими шагами, но степень уменьшения шага по всем факторам должна быть одинаковой.

После расчета координат каждого нового опыта совсем необязательно сразу проводить соответствующий опыт в реальности. Вначале стоит провести несколько опытов мысленно, рассчитывая по уравнению регрессии значение параметра в каждом из намеченных опытов и убеждаясь, что мы действительно получаем все лучшие и лучшие значения Y . Если Y меняется слишком медленно, увеличим шаг по всем факторам. С другой стороны, эти шаги не должны быть настолько большими, чтобы условия новых опытов по какому-либо из факторов вышли за пределы области его допустимых значений. Если это будет происходить (в мысленном эксперименте!), следует пропорционально уменьшить шаг изменения всех факторов.

Разумеется, исследователь не должен ограничиваться мысленным экспериментом. Отдельные опыты надо проверять, осуществляя процесс на практике и убеждаясь, что экспериментальные и рассчитанные значения Y примерно совпадают, т. е. модель остается адекватной. Движение по градиенту продолжают до тех пор, пока не наступает одно из трех возможных событий:

а) реализуется желаемое значение Y . В этом случае эксперимент прекращают и празднуют достижение цели;

б) расчетное значение Y перестает соответствовать измеренному в соответствующих опытах. В этом случае стоит остановиться и попробовать отыскать новое уравнение регрессии, а затем двигаться по новому градиенту;

в) по одному из факторов мы выходим за границы области допустимых значений (ОДЗ). В этом случае либо прекращают эксперимент, считая, что сделали все, что могли; либо фиксируют этот фактор на постоянном уровне (на границе ОДЗ) и проводят новый ПФЭ меньшей размерности, чтобы определить направление дальнейшего движения по поверхности отклика.

В качестве примера рассмотрим процесс каталитического отверждения мочевиноформальдегидной смолы, описанный в работе [30]. Факторы — концентрации (в %) семи компонентов катализатора, а параметр — прочность на сжатие образца полученной смолы. После проведения ПФЭ вида 2^7 была получена адекватная регрессия, не включающая эффекты взаимодействия факторов. Оценки коэффициентов оказались значимыми и симметричными. После перехода к натуральным переменным интерполяционная формула приобрела следующий вид:

$$Y = 60,75 - 2,50X_1 - 1,50X_2 + 5,75X_3 - 2,25X_4 - 2,00X_5 - 4,50X_6 + 1,75X_7.$$

В исходных опытах удавалось получить значения Y , не превышающие 74 единицы. Движение по градиенту начинали из центра исходного ПФЭ, причем шаг по X_5 по техническим причинам был выбран на уровне 0,05.

Из таблицы 18 видно, что мысленный эксперимент уже с третьего шага стал давать лучшие значения параметра, чем ранее достигнутые в ходе ПФЭ. При движении по градиенту каждый следующий шаг приводил к расчетному значению параметра, на 2,7 единиц лучше предыдущего. Экспериментальная проверка этих расчетов, разумеется, давала невязки, но тенденция подъема Y проявилась и в эксперименте, по крайней мере вплоть до шестого шага, а затем величина Y начала снижаться, что свидетельствовало о непригодности полученной модели в данной области факторного пространства. Проведенное восхождение

следует считать эффективным, так как реальную прочность смолы всего за шесть дополнительных опытов удалось поднять с 74 до 85 единиц. Приятным побочным эффектом было уменьшение в рецептуре содержания наиболее дорогих компонентов данного катализатора — компонентов X_1 и X_6 .

Т а б л и ц а 18

Расчет и результаты крутого восхождения [30]

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	$Y_{\text{расчет}}$	Y
b_i	-2,5	-1,5	5,75	-2,25	-2,00	-4,50	1,75		
ΔI_i	0,5	0,2	1,0	0,5	0,5	0,5	0,5		
ΔX_i	-0,0625	-0,015	0,287	-0,056	-0,05	-0,112	0,043		
$\Delta X_{i \text{ округл}}$	-0,06	-0,02	0,3	-0,06	-0,05	-0,12	0,04		
Центр плана	1,50	0,50	3,0	1,00	1,50	1,0	1,5	66,4	
Шаг № 1	1,44	0,48	3,3	0,94	1,45	0,88	1,54	69,1	68
Шаг № 2	1,38	0,46	3,6	0,88	1,40	0,76	1,58	71,8	
Шаг № 3	1,32	0,44	3,9	0,82	1,35	0,64	1,62	74,6	72
Шаг № 4	1,26	0,42	4,2	0,76	1,30	0,52	1,66	77,3	79
Шаг № 5	1,20	0,40	4,5	0,70	1,25	0,40	1,70	80,1	83
Шаг № 6	1,14	0,38	4,8	0,64	1,20	0,28	1,74	82,8	85
Шаг № 7	1,08	0,36	5,1	0,58	1,15	0,16	1,78	85,5	80
Шаг № 8	1,02	0,34	5,4	0,52	1,10	0,04	1,82	88,2	76

В литературе описаны и более эффективные исследования, в которых химико-технологические процессы или методики анализа оптимизировали, применяя метод крутого восхождения. В этих исследованиях величину параметра удавалось изменить очень сильно. Так, в работе

Л. Н. Комиссаровой и соавторов по экстракционному извлечению микроколичеств гафния трибутилфосфатом после трех последовательно проведенных ПФЭ ($k = 4$) и промежуточных восхождений по градиенту коэффициент распределения гафния удалось увеличить на четыре порядка [30].

Алгоритмы крутого восхождения резко осложняются, если уравнение регрессии (выраженное в безразмерных переменных) имеет значимые коэффициенты вида b_{ij} , т. е. факторы взаимодействуют между собой. Поверхность отклика не является плоской, и направление градиента от точки к точке меняется. Меняется и величина шага, поэтому координаты каждого следующего мысленного опыта следует вычислять с помощью специальной компьютерной программы.

Кроме метода крутого восхождения к оптимуму, для решения оптимизационных задач химии часто используют *симплекс-процедуру*. Алгоритмы симплексной оптимизации описаны в специальной литературе, например, в книге [31]. Достоинства этого способа подтверждены множеством публикаций в разных областях химической науки. Например, в работе [32] метод симплексной оптимизации был применен для оптимизации условий кинетического определения микрограммовых количеств молибдат-ионов. Одновременно меняли значения четырех факторов, оценивали обобщенный параметр оптимизации (разность скоростей каталитической и некаталитической реакций). Известную в аналитической химии методику удалось существенно улучшить после проведения всего 11 опытов, а именно: время анализа сократилось вдвое, а предел обнаружения снизился в три раза. При этом рассчитывать какую-либо математическую модель не потребовалось, симплексная оптимизация — чисто эмпирический метод. Отметим, что процедура симплексной оптимизации используется и во многих компьютерных программах.

Глава 11

ПРИМЕНЕНИЕ НЕПАРАМЕТРИЧЕСКИХ МЕТОДОВ И КРИТЕРИЕВ

11.1. ПОНЯТИЕ О НЕПАРАМЕТРИЧЕСКИХ МЕТОДАХ ОБРАБОТКИ ДАННЫХ

Методы обработки данных, изложенные в предыдущих главах, так или иначе опирались на свойство нормальности распределения результатов измерений. Некоторые исследователи полагают, что эти методы всегда являются «правильными» и даже единственно возможными. Однако эта точка зрения ничем не обоснована. Известна грустная шутка: «Химики убеждены, что нормальность распределения результатов теоретически доказана математиками. Математики же доказать это не могут, но полагают, что химики почему-то получают в эксперименте только нормально распределенные данные. Поэтому ни те, ни другие ничего не проверяют и всегда применяют параметрические методы».

На самом деле в химических исследованиях нередко возникают ситуации, когда распределение генеральной совокупности результатов измерений либо не известно, либо обоснованно предполагается отличным от нормального. Тогда применение описанных выше параметрических методов обработки данных (например, дисперсион-

ного и регрессионного анализа, расчета доверительных интервалов по Стьюденту и т. п.) оказывается необоснованным и может привести к ошибочным выводам. В этих случаях на помощь обычно приходят так называемые *ранговые методы*, в которых статистические критерии проверки гипотез рассчитываются на основе рангов вариант изучаемых выборок.

Поясним понятие *ранга* варианты на примере выборки, включающей n результатов измерений случайной величины X . В данном случае неважно, будет ли в химическом эксперименте величина X фактором или параметром. Обозначим выборку символом X и запишем ее в порядке получения вариант (результатов повторных измерений):

$$X: x_1, x_2, \dots, x_n. \quad (11.1)$$

Если проранжировать выборку, т. е. расставить ее варианты в порядке возрастания, то каждая варианта займет определенное место. Номер места, на котором будет стоять варианта x_i , называется ее рангом x_i^r . Если в выборке X все варианты различны по своим значениям, то ранги этих вариантов также будут различаться. Если же в выборке X имеются варианты с одинаковыми значениями, то такие варианты получают одинаковые ранги, равные среднему арифметическому мест, на которых эти варианты стоят в упорядоченной выборке. Наличие в выборке X вариант с одинаковыми значениями может быть связано с особенностями объекта исследования (дискретные совокупности), особенностями методики измерений, округлением результатов измерений и т. д. Варианты ранжированной выборки обозначим через $x_{(i)}$, $i = 1, 2, \dots, n$.

Рассмотрим пример. Выборка значений X имеет вид:

$$x_1 = 7,1, \quad x_2 = 8,2, \quad x_3 = 5,1, \quad x_4 = 3,5, \quad x_5 = 8,2.$$

После ранжирования варианты этой выборки будут располагаться в таком порядке: x_4, x_3, x_1, x_2, x_5 . Находим ранги вариант:

$$x_4^r = 1, \quad x_3^r = 2, \quad x_1^r = 3, \quad x_2^r = x_5^r = (4+5)/2 = 4,5.$$

Ранговые методы оперируют не с самой выборкой X , а с рангами ее вариантов. Ранги позволяют «забыть» о числовых значениях вариантов, но содержат в себе информацию о расположении вариантов выборки. Особенность ранговых методов состоит в том, что они используются в тех ситуациях, когда исследователь ничего не знает о параметрах измеряемой случайной величины (о ее математическом ожидании, дисперсии и др.), о законе ее распределения и т. д. Отсюда возникает более общее и более распространенное название этих методов — *непараметрические методы*. Непараметрические методы приводят к новым критериям проверки статистических гипотез, эти критерии не зависят от вида распределения измеряемой случайной величины.

Важной стороной непараметрических методов является возможность проверки нуль-гипотез по выборкам малого объема. Однако в этом случае необходимо располагать таблицами критических значений статистических показателей (критериев). Для выборок малого объема такие таблицы построены путем компьютерной обработки множества числовых примеров. Для выборок большого объема критические значения были найдены математиками с помощью различных теоретических приближений. Таблицы непараметрических критериев менее популярны, чем таблицы Стьюдента или Фишера, но все же вполне доступны, они приведены в специальных справочниках. Обычно составители таблиц вводят единственное ограничение в применимости непараметрических методов и критериев. Это ограничение состоит в том, что результаты измерений должны представлять собой значения случайных величин с непрерывными функциями распределения.

В последующих разделах представлено несколько непараметрических критериев, которые используются на практике чаще других. Более подробное описание этих и других критериев, а также соответствующие статистические таблицы можно найти в работах [33]–[36].

* Термин «параметр» в данном случае имеет совершенно другой смысл, чем при планировании эксперимента и последующей оптимизации условий какого-либо процесса. Непараметрические методы применимы для обработки любых данных как факторов, так и параметров оптимизации.

11.2. КРИТЕРИИ ОТБРАКОВКИ ПРОМАХОВ

Критерий Смирнова^{*}. Проранжируем выборку X из (11.1). Если сомнения вызывает «подозрительно большая» варианта $x_{(n)}$, последняя в ранжированной выборке, то проверяют нуль-гипотезу, состоящую в том, что $x_{(n)}$ не является грубой ошибкой (промахом). Вычисляют величину

$$\lambda_{\text{эксп}} = \frac{x_{(n)} - \bar{x}}{\sigma}, \quad (11.2)$$

где \bar{x} — среднее арифметическое выборки, включающей сомнительную варианту; σ — стандартное (среднеквадратическое) отклонение, рассчитываемое по формуле (4.2). Затем сравнивают $\lambda_{\text{эксп}}$ с величиной $\lambda_{\text{крит}}$, которая выбирается из соответствующих таблиц критических значений этого показателя, с учетом выбранного уровня значимости нуль-гипотезы (α). Значения $\lambda_{\text{крит}}$ приведены в приложении 2М. При выполнении неравенства $\lambda_{\text{эксп}} \leq \lambda_{\text{крит}}$ нуль-гипотезу принимают. В случае $\lambda_{\text{эксп}} > \lambda_{\text{крит}}$ нуль-гипотезу отклоняют.

Если в качестве «подозрительной» взять варианту $x_{(1)}$, то следует вычислить величину

$$\lambda_{\text{эксп}} = \frac{\bar{x} - x_{(1)}}{\sigma} \quad (11.3)$$

и аналогичным образом сравнить ее с величиной $\lambda_{\text{крит}}$.

Пример 11-1. Результаты анализа характеризуют влажность сырья (в %):

X : 12, 11, 13, 12, 14, 12, 18, 15, 14, 13.

«Подозрительной» является варианта $x_{(10)} = 18$. Следует ли ее сохранять? Характер распределения результатов измерения влажности неизвестен.

^{*} Этот критерий правильнее называть критерием Смирнова — Граббса, но такое название может привести к путанице, поскольку есть и другой критерий Граббса (см. ниже).

Решение. Применяем критерий Смирнова. В данном случае $\bar{x} = 13,4$, $\sigma = 1,91$. Из формулы (9.2) находим, что $\lambda_{\text{эксп}} = 2,41$. Из таблицы (приложение 2М) получаем, что $\lambda_{\text{крит}} = 2,294$ при уровне значимости $\alpha = 0,05$. Поскольку $\lambda_{\text{эксп}} > \lambda_{\text{крит}}$, нуль-гипотеза отклоняется. Принимают, что варианта $x_{(10)} = 18$ является грубой ошибкой и ее следует отбраковать (вероятность ошибки первого рода при таком решении не превосходит 0,05). Поэтому в ходе обработки результатов анализа данный результат не учитывают.

Критерий Граббса позволяет решать те же задачи и также опирается на рассмотрение ранжированной выборки. Однако величина $\lambda_{\text{эксп}}$ вычисляется в этом случае по другим формулам. Кроме того, отклоняют нуль-гипотезу, когда выполняется неравенство $\lambda_{\text{эксп}} < \lambda_{\text{крит}}$. В качестве претендентов на грубую ошибку будем рассматривать «подозрительные» варианты $x_{(n)}$ и $x_{(1)}$ — наибольшую и наименьшую в ранжированной выборке. Проверяют нуль-гипотезы, состоящие в том, что $x_{(n)}$ или $x_{(1)}$ не является грубой ошибкой. Для этого вычисляют средние арифметические исследуемой выборки обычным способом (учитывая все варианты):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{(i)}.$$

Затем повторяют расчет, не учитывая сомнительную варианту. А именно:

$$\bar{x}_{(n)} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_{(i)}; \quad (11.4)$$

$$\bar{x}_{(1)} = \frac{1}{n-1} \sum_{i=2}^n x_{(i)}. \quad (11.5)$$

Далее для проверяемой варианты находят значение показателя:

$$\lambda_{\text{эксп}}^{(n)} = \frac{\sum_{i=1}^{n-1} (x_{(i)} - \bar{x}_{(n)})^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}; \quad (11.6)$$

$$\lambda_{\text{эксп}}^{(1)} = \frac{\sum_{i=2}^n (x_{(i)} - \bar{x}_{(1)})^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}. \quad (11.7)$$

Очевидно, при наличии грубых промахов средние арифметические, рассчитанные с учетом и без учета промаха, будут существенно различны, и показатели, рассчитанные по формулам (11.6) или (11.7), станут заметно отличаться от единицы. Нуль-гипотеза относительно $x_{(n)}$ или $x_{(1)}$ отклоняется, и проверяемую варианту считают промахом, если $\lambda_{\text{эксп}}^{(n)}$ или $\lambda_{\text{эксп}}^{(1)}$ будут меньше $\lambda_{\text{крит}}$. Величину $\lambda_{\text{крит}}$ выбирают из таблиц критерия Граббса (приложение 2N). При выполнении неравенства $\lambda_{\text{эксп}} \geq \lambda_{\text{крит}}$ нуль-гипотезу принимают. Обобщением критерия Граббса является критерий Титъена — Мура, который используется при одновременном поиске нескольких грубых промахов. Этот интересный критерий описан в специальной литературе (см. [35, с. 303]).

Пример 11-2. Рассмотрим данные из примера 11-1 еще раз и проверим по критерию Граббса, не является ли варианта $x_{(10)} = 18$ грубым промахом.

Р е ш е н и е. Вычисляя $\bar{x}_{(n)}$ и $\lambda_{\text{эксп}}^{(n)}$ по формулам (9.4) и (9.6), находим $\lambda_{\text{эксп}}^{(n)} = 12,89/36,4 = 0,3541$. При объеме выборки $n = 10$ и уровне значимости $\alpha = 0,05$ величина $\lambda_{\text{крит}} = 0,4154$. Так как верно неравенство $\lambda_{\text{эксп}}^{(n)} < \lambda_{\text{крит}}$, то нуль-гипотезу отклоняем и считаем, что $x_{(10)} = 18$ является грубым промахом. Оба способа проверки (по Смирнову и по Граббсу) дали один и тот же результат.

11.3.

КРИТЕРИИ РАЗЛИЧИЯ ВЫБОРОК

Критерий Вилкоксона — Манна — Уитни является ранговым. Его применяют для обработки выборок вида:

$$X: x_1, x_2, \dots, x_n; \quad Y: y_1, y_2, \dots, y_k, \quad (11.8)$$

представляющих собой результаты измерения однотипных величин X и Y . Объемы выборок (n и k) в общем случае могут быть различными. Критерий используется для проверки нуль-гипотезы, состоящей в том, что величины X и Y имеют одинаковые математические ожидания. Отклонение нуль-гипотезы можно интерпретировать как результат вмешательства некоторых факторов, воздействующих на измеряемые величины и приводящих к достоверным различиям в их математических ожиданиях и, следовательно, в функциях распределения. Полагаем, что выборки X и Y получены независимо друг от друга. В отличие от рассматриваемого далее критерия знаков, здесь отсутствует какая-либо привязка вариант x_i и y_i друг к другу.

Суть метода — в оценке степени «перемешивания» вариант различных выборок при их объединении. «Хорошее перемешивание» объединенной выборки означает, что измеряемая величина в обеих выборках имела одно и то же математическое ожидание. Это можно интерпретировать как отсутствие влияния «факторов сдвига». Для расчета показателя поступают следующим образом. Выборки X и Y объединяют в одну выборку и располагают ($n + k$) ее вариант в порядке возрастания. Получают ранжированный ряд и находят ранг каждой варианты этого ряда. Обозначают через R_x сумму рангов вариант, относившихся ранее к выборке X , а через R_y — сумму рангов вариант, относившихся ранее к выборке Y . Используя R_x и R_y , находят вспомогательные величины:

$$\begin{aligned} W_y &= nk + \frac{k(k+1)}{2} - R_y; \\ W_x &= nk + \frac{n(n+1)}{2} - R_x, \end{aligned} \quad (11.9)$$

причем для контроля правильности вычислений следует проверить выполнение соотношения $W_x + W_y = nk$. Для дальнейших расчетов берут меньшую из величин W_x и W_y , обозначая ее как W_{xy} . Расчет показателя $\lambda_{\text{эксп}}$ облегчается при $n > 8$ и $k > 8$, т. е. при сопоставлении выборок относительно большого объема. В этом случае показатель $\lambda_{\text{эксп}}$ вычисляют по формуле

$$\lambda_{\text{эксп}} = \frac{|W_{xy} - 0,5nk|}{\sqrt{nk(n+k+1)/12}}. \quad (11.10)$$

Теперь сравнивают $\lambda_{\text{эксп}}$ с $\lambda_{\text{крит}}$, выбирая критерий из таблицы значений функции Лапласа при выбранном уровне значимости нуль-гипотезы (α). В частности, $\lambda_{\text{крит}} = \lambda_1 = 1,96$ для $\alpha = 0,05$; $\lambda_{\text{крит}} = \lambda_2 = 2,576$ для $\alpha = 0,01$. Если окажется, что $\lambda_{\text{эксп}} \leq \lambda_1$, то нуль-гипотезу принимают. Это будет означать, что рассматриваемые величины X и Y имеют одинаковые математические ожидания, а также одинаковые медианы. В противном случае, когда $\lambda_{\text{эксп}} > \lambda_2$, нуль-гипотезу отклоняют.

Если объемы выборок не удовлетворяют условию $n > 8$, $k > 8$, то для вычисления $\lambda_{\text{эксп}}$ используют более сложные формулы, а $\lambda_{\text{крит}}$ находят по другим таблицам (см., например, руководства [33], [34]).

Пример 11-3. Предложены два метода (А и Б) проведения некоторой реакции и определены значения выхода продукта (в % от теоретически возможного). На уровне значимости $\alpha = 0,05$ требуется проверить нуль-гипотезу об одинаковой эффективности обоих методов. Применение метода А привело в повторных опытах к выборке:

X: 50,0; 41,0; 48,5; 60,0; 46,0; 60,0; 51,0; 42,5;
62,5; 54,0; 42,5; 46,0.

Применение метода Б привело к следующим результатам:

Y: 38,5; 40,0; 47,5; 51,0; 63,0; 50,0; 63,0; 57,5; 59,0; 51,0.

Решение. Используем критерий Вилкоксона — Манна — Уитни. Объемы выборок ($n = 12$, $k = 10$) позволяют применить формулу (11.10). Вначале по формулам (11.9) получаем, что $W_x = 68,5$, $W_y = 51,5$. Проверка подтверждает правильность расчета: $W_x + W_y = n \cdot k = 120$. Обозначим $W_y = 51,5$ через W_{xy} . Применяя формулу (11.10) получаем $\lambda_{\text{эксп}} = 0,56$. Значение критерия для $\alpha = 0,05$ равно 1,96. Так как $\lambda_{\text{эксп}} \leq \lambda_{0,05}$, нуль-гипотеза принимается. Можно считать, что эффективность методов А и Б для данного синтеза приблизительно одинакова ($\alpha > 0,05$).

Следует обратить внимание, что ту же задачу можно было бы решить с применением более простого критерия Стьюдента (см. раздел 7.4), но тогда надо было бы вначале проверить нормальность распределения вариантов в обеих выборках (например, по эксцессу и асимметрии) и однородность обеих дисперсий (например, по Фишеру). Использование критерия Вилкоксона — Манна — Уитни этих предварительных операций не требует.

Критерий знаков. Этот непараметрический критерий применяется для обработки так называемых *парно связанных выборок*:

$$X: x_1, x_2, \dots, x_n; \quad Y: y_1, y_2, \dots, y_n. \quad (11.11)$$

Такие выборки имеют один и тот же объем, относятся к одним и тем же объектам исследования и возникают вследствие измерения однотипных величин по двум разным методикам, или в разных условиях, или в разное время (до и после некоторого воздействия на объекты). Например, для десяти разных катализаторов определяют удельную поверхность каждого из них, применяя две разных методики. Результат, полученный по первой методике, обозначим как x_i , по второй — как y_i . Будем называть такой процесс парным измерением. Полученные результаты парных измерений (значения x и y) образуют парно связанные выборки.

Варианты выборок парно сравнивают по принципу «больше — меньше», определяя знак разности ($x - y$). Сравнение вариантов производится только внутри пары (x_i, y_i), относящейся к i -му парному измерению. Комбинации пар в этом методе не рассматриваются. В частности, не рассчитываются и не сопоставляются средние значения разных x_i или разных y_i .

Различие x_i и y_i внутри пары можно интерпретировать как результат вмешательства некоторого фактора, который приводит к уменьшению (или увеличению) возможных значений величины y по отношению к величине x . Если фактор приводит к уменьшению значений y , разности ($x_i - y_i$) будут в основном положительными. Если же фактор ведет к увеличению y , разности будут в основном отрицательными. На подсчете числа разностей, имеющих

тот или иной знак, и основан критерий знаков. При этом не важно, насколько будут сдвинуты в каждой паре значения y по сравнению с x .

Проверяемая нуль-гипотеза состоит в том, что величины x и y имеют одну и ту же функцию распределения (а значит, одно и то же математическое ожидание), т. е. «фактор сдвига» не действует. Если это так, то вероятности появления положительных или отрицательных разностей вида $(x - y)$ будут одинаковы и равны 0,5. В противном случае вероятности будут различны.

Для проверки нуль-гипотезы используют следующий алгоритм. По выборкам X и Y из (11.11) составляют разности $(x_i - y_i)$, где $i = 1, \dots, n$. Все нулевые разности исключают из рассмотрения. Количество ненулевых разностей обозначают через m . Число знаков «+» в ненулевых разностях обозначают через r . Зафиксируем уровень значимости нуль-гипотезы (α). Нуль-гипотезу отклоняют, если выполняется неравенство:

$$\lambda_{\text{эксп}} = \frac{r}{m - r + 1} > \lambda_{\text{крит}}, \quad (11.12)$$

где $\lambda_{\text{крит}}$ выбирают из таблиц критерия Фишера при числе степеней свободы

$$df_1 = 2(m - r + 1); \quad df_2 = 2r$$

и заданном α . При выполнении (11.12) полагают, что величины x и y имеют различные функции распределения, т. е. «фактор сдвига» достоверно действует, уменьшая значения y по отношению к значениям x .

Нуль-гипотеза также будет отклонена, если выполняется неравенство

$$\lambda_{\text{эксп}} = \frac{m - r}{r + 1} > \lambda_{\text{крит}}, \quad (11.13)$$

где $\lambda_{\text{крит}}$ выбирается из таблиц критерия Фишера при числе степеней свободы:

$$df_1 = 2(r + 1); \quad df_2 = 2(m - r)$$

и заданном α . В этом случае также полагаем, что величины x и y имеют различные функции распределения, но

«фактор сдвига» достоверно повышает результаты измерений величины y по сравнению с x . Если же оба неравенства (11.12) и (11.13) не выполняются, то нуль-гипотеза принимается.

В конкретных задачах можно рекомендовать следующую очередность проверки неравенств (11.12), (11.13). Если в разностях $(x_i - y_i)$ встречается большинство знаков «+», то сначала применяется проверка по (11.12), если же большинство знаков «-», то вначале используется (11.13).

Пример 11-4. Некоторая присадка улучшила термостабильность семнадцати образцов масел определенной группы, ухудшила три и не повлияла на шесть образцов. Можно ли считать такую присадку эффективной для группы масел в целом?

Решение. Объем выборки $n = 17 + 3 + 6 = 26$. Количество ненулевых разностей $m = 17 + 3 = 20$. Число знаков «+» в ненулевых разностях $r = 17$. Применяя формулу (11.12), получим, что $\lambda_{\text{эксп}} = 17/4 = 4,25$. Число степеней свободы: $df_1 = 8$, $df_2 = 34$. Критические значения таковы: $\lambda_{\text{крит}} = \lambda_1 = 2,3$ для $\alpha = 0,05$ и $\lambda_{\text{крит}} = \lambda_2 = 3,3$ для $\alpha = 0,01$. Здесь мы имеем случай, когда $\lambda_{\text{эксп}} > \lambda_2$. Поэтому можно считать присадку достоверно повышающей термостабильность масел данной группы ($\alpha < 0,01$). Для сравнения можно повторить расчет при другом количестве «плюсов». Оказывается, при семи «плюсах» и трех «минусах» вывод об эффективности присадки был бы недостоверен, а при одиннадцати «плюсах» и трех «минусах» величина $\lambda_{\text{эксп}}$ попала бы между λ_1 и λ_2 . В последнем случае следовало бы продолжить эксперимент, увеличить объем выборки и все расчеты повторить заново, включая новые данные.

В статистической литературе можно найти таблицы критического числа «плюсов» для разного числа ненулевых разностей [26]. В данном пособии они приведены в качестве приложения 2К. Применение готовых таблиц облегчает использование критерия знаков, но несколько снижает надежность выводов.

Критерий равенства дисперсий по своему предназначению похож на параметрический критерий Фишера — здесь также сопоставляются две однотипные выборки X и Y по степени разброса их вариант. При этом парное связывание вариант отсутствует. Предполагают, что математи-

ческие ожидания или медианы величин x и y равны. Данное предположение может быть заранее проверено с помощью критерия Вилкоксона — Манна — Уитни.

Выбирают нуль-гипотезу, состоящую в том, что величины x и y имеют одинаковые дисперсии. Для ее проверки будем опираться на выборки (11.8), объемы которых таковы, что $n > 10$, $k > 10$, причем $n \geq k$. Последнее неравенство всегда можно обеспечить при планировании эксперимента, устанавливая, что выборка X должна содержать не меньше значений, чем выборка Y .

Для проверки нуль-гипотезы поступим следующим образом. Выборки X и Y объединим в одну выборку Z и расположим $(n + k)$ ее вариант z_i в порядке возрастания (принадлежность вариант z_i к исходным выборкам X или Y запоминается). Каждой варианте z_i присваивается номер z_i^u , задаваемый так: наименьшему значению присваивается номер 1, два наибольших значения получают номера 2 и 3, номера 4 и 5 получают следующие наименьшие значения и т. д. Иначе говоря, номера z_i^u расставляются следующим образом:

$$1, 4, 5, 8, 9, \dots, 7, 6, 3, 2. \quad (11.14)$$

Если в выборках X и Y имеются совпадающие варианты, то они получают номер, равный среднему арифметическому их номеров. Это значит, что в перечне номеров для (11.14) могут появляться нецелые значения. Обозначим через r_y сумму номеров вариант выборки Y , полученных описанным способом. Используя r_y , составим величину

$$\lambda_{\text{экс}} = \frac{0,5|2r_y - 1 - k(n + k + 1)|}{\sqrt{n(n + k + 1)/12}}. \quad (11.15)$$

Сравним $\lambda_{\text{экс}}$ с $\lambda_{\text{крит}}$, выбираемым из таблицы функции Лапласа при заданном α . Если окажется, что $\lambda_{\text{экс}} \leq \lambda_{\text{крит}}$, то нуль-гипотезу принимаем. В противном случае ($\lambda_{\text{экс}} > \lambda_{\text{крит}}$) нуль-гипотезу отклоняем. При $n \leq 10$ или $k \leq 0$ описанный критерий применять не рекомендуется.

Пример 11-5. Определение хрома в одной и той же низколегированной стали проведено по двум разным методикам. Обе методики свободны от систематических погрешностей, т. е. математическое ожидание результата анализа

в обоих случаях равно истинному содержанию хрома в стали. Получены две выборки с объемами $n = 12$ и $k = 10$:

Методика А: 1,50; 1,41; 1,48; 1,60; 1,46; 1,60;
1,51; 1,42; 1,62; 1,54; 1,42; 1,46;

Методика Б: 1,38; 1,40; 1,47; 1,51; 1,63; 1,50;
1,63; 1,57; 1,59; 1,51.

Требуется сравнить методики по воспроизводимости результатов анализа.

Р е ш е н и е. Проверяем нуль-гипотезу о равенстве дисперсий по вышеописанному критерию. Упорядочивая объединенную выборку по возрастанию вариантов, получим следующее расположение ее вариантов с учетом их прежней принадлежности к выборке X (методика А) или к выборке Y (методика Б):

$$y_1, y_2, x_2, x_8, x_{11}, x_5, x_{12}, y_3, x_3, x_1, y_6, x_7, y_4, y_{10}, x_{10}, \\ y_8, y_9, x_4, x_6, x_9, y_5, y_7.$$

Номера вариант исходных выборок, найденные способом (9.14), равны:

$$y_1^u = 1, y_2^u = 4, x_2^u = 5, x_8^u = 8,5, x_{11}^u = 8,5, x_5^u = 12,5, \\ x_{12}^u = 12,5, y_3^u = 16, x_3^u = 17, x_1^u = 20,5, y_6^u = 20,5, x_7^u = 19,7, \\ y_4^u = 19,7, y_{10}^u = 19,7, x_{10}^u = 15, y_8^u = 14, y_9^u = 11, x_4^u = 8,5, \\ x_{76}^u = 8,5, x_9^u = 6, y_5^u = 2,5, y_7^u = 2,5.$$

Сумма номеров элементов выборки Y равна $r_y = 110,9$. Используя формулу (11.15), находим, что $\lambda_{\text{эксн}} = 0,237$. Для $\alpha = 0,05$ $\lambda_1 = 1,96$. Очевидно, $\lambda_{\text{эксн}} < \lambda_1$. Следовательно, нуль-гипотеза принимается ($\alpha > 0,05$), методики А и Б можно считать примерно одинаковыми по воспроизводимости результатов.

11.4.

НЕПАРАМЕТРИЧЕСКИЕ АНАЛОГИ ДИСПЕРСИОННОГО АНАЛИЗА

Критерий Краскела — Уоллиса. При изучении различных процессов можно столкнуться с ситуацией, когда фактор А предположительно влияет на результат эксперимента, т. е. на некоторый параметр оптимизации Y .

Причем фактор принимает определенные значения (уровни), которые могут выражаться как качественно, так и в числовой форме. Результаты всех опытов отягощены погрешностями измерений, поэтому каждый независимый опыт повторяют несколько раз. Характер распределения повторных результатов для данной методики измерений неизвестен. Надо установить, действительно ли фактор A достоверно влияет на Y . Задача решается методами, близкими к однофакторному варианту дисперсионного анализа (см. раздел 8.4).

В качестве примера рассмотрим случай, когда фактор A имеет три разных уровня (ход дальнейших рассуждений не зависит от того, сколько уровней фактора используется). Допустим, мы получили результаты измерений, представленные в таблице 19. Символами x , y , z обозначили результаты измерений одного и того же параметра, полученные когда фактор A находится соответственно на первом, втором и третьем уровнях. Числа n_1 , n_2 , n_3 указывают на объемы соответствующих выборок, они могут быть одинаковыми или различными. Символом $N = n_1 + n_2 + n_3$ обозначили общее число измерений. Полагаем, что выборки, представленные в строчках таблицы 19, получены независимо друг от друга.

Т а б л и ц а 19

Представление исходных данных для однофакторного анализа

Уровни фактора A	Результаты измерений параметра				
1	x_1	x_2	x_3	...	x_{n_1}
2	y_1	y_2	y_3	...	y_{n_2}
3	z_1	z_2	z_3	...	z_{n_3}

Рассмотрим нуль-гипотезу, состоящую в том, что величины x , y , z имеют одинаковые математические ожидания. Это означает, что фактор A не оказывает достоверного влияния на величину измеряемого параметра. Будем проверять эту нуль-гипотезу, опираясь на критерий Краскела — Уоллиса.

На первом шаге объединим все варианты каждой строки в одну выборку и упорядочим ее по возрастанию вариантов. Найдем ранги вариант проранжированной выборки (табл. 20).

Ранги вариант в объединенной выборке

Уровни фактора A	Ранги результатов измерений				
1	x_1^r	x_2^r	x_3^r	...	x_{n1}^r
2	y_1^r	y_2^r	y_3^r	...	y_{n2}^r
3	z_1^r	z_2^r	z_3^r	...	z_{n3}^r

Для каждой строки таблицы вычисляем сумму рангов вариант, т. е. находим вспомогательные величины:

$$\begin{aligned} R_x &= x_1^r + x_2^r + x_3^r + \dots + x_{n1}^r; \\ R_y &= y_1^r + y_2^r + y_3^r + \dots + y_{n2}^r; \\ R_z &= z_1^r + z_2^r + z_3^r + \dots + z_{n3}^r. \end{aligned} \quad (11.16)$$

На втором шаге находим величину $\lambda_{\text{эксп}}$ по формуле

$$\lambda_{\text{эксп}} = \frac{12}{N(N+1)} \left(\frac{R_x^2}{n_1} + \frac{R_y^2}{n_2} + \frac{R_z^2}{n_3} \right) - 3(N+1). \quad (11.17)$$

На третьем сравниваем $\lambda_{\text{эксп}}$ и $\lambda_{\text{крит}}$. Величина $\lambda_{\text{крит}}$ задается следующим образом. Если объемы выборок из таблицы 19, т. е. числа n_1, n_2, n_3 велики, то $\lambda_{\text{крит}}$ выбираем из таблиц критических значений для χ^2 -критерия Пирсона при числе степеней свободы $df = 2$. В частности, $\lambda_{0,05} = 5,99$; $\lambda_{0,01} = 9,21$. Если же объемы выборок n_1, n_2, n_3 малы, то $\lambda_{\text{крит}}$ выбираем из таблиц критических значений для данного критерия (приложение 20). Отклоняем нуль-гипотезу, если $\lambda_{\text{эксп}} > \lambda_{\text{крит}}$; сохраняем, если $\lambda_{\text{эксп}} \leq \lambda_{\text{крит}}$. Отклонение нуль-гипотезы можно интерпретировать как достоверное влияние рассматриваемого фактора на математические ожидания величин x, y, z , т. е. на исследуемый параметр. Отметим, что формула (11.17) применяется в случае, когда данные таблицы 19 не содержат вариант, одинаковых по своим значениям. В противном случае величина $\lambda_{\text{эксп}}$ корректируется с учетом совпадающих вариант. Подробности описаны в статистической литературе (см. [33], [34], [36]).

Критерий Фридмана. Естественным развитием метода однофакторного анализа является двухфакторный анализ, в котором рассматривается классификация результатов

измерений некоторого параметра по возможному влиянию факторов A и B . Примем, что фактор A имеет k уровней, фактор B — n уровней, а исследователь располагает набором данных, приведенных в таблице 21.

Т а б л и ц а 21

Представление исходных данных для двухфакторного анализа

Уровни фактора B	Уровни фактора A				
	A_1	A_2	A_3	...	A_k
B_1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$...	$x_{1,k}$
B_2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$...	$x_{2,k}$
B_3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$...	$x_{3,k}$
...
B_n	$x_{n,1}$	$x_{n,2}$	$x_{n,3}$...	$x_{n,k}$

Числа $x_{i,j}$ из таблицы 21 означают результаты конкретных измерений при фиксированных уровнях факторов B и A , $i = 1, \dots, n$; $j = 1, \dots, k$. Главное отличие данных из таблицы 21 от данных таблицы 19 заключается в следующем. В задачах однофакторного анализа каждая строка таблицы 19 образует выборку из конкретной генеральной совокупности. В задачах двухфакторного анализа каждая строка таблицы 21, вообще говоря, выборку не образует. Это связано с тем, что фактор A может оказывать влияние на распределение величин $x_{i,j}$, стоящих в конкретной строке таблицы 21. Например, на математическое ожидание или на распределение генеральной совокупности.

Проверяемая нуль-гипотеза в задачах двухфакторного анализа состоит в предположении, что влияние фактора A отсутствует (при этом B рассматривается как фактор, определяющий условия проведения эксперимента). Алгоритм проверки нуль-гипотезы по критерию Фридмана включает несколько шагов.

На первом шаге проводим ранжирование данных, представленных в строках таблицы 21. Каждая строка рассматривается отдельно от остальных. Ранги элементов записываем, как показано в таблице 22.

Таблица рангов для двухфакторного анализа

Уровни фактора В	Уровни фактора А				
	A_1	A_2	A_3	...	A_k
B_1	$x_{1,1}^r$	$x_{1,2}^r$	$x_{1,3}^r$...	$x_{1,k}^r$
B_2	$x_{2,1}^r$	$x_{2,2}^r$	$x_{2,3}^r$...	$x_{2,k}^r$
B_3	$x_{3,1}^r$	$x_{3,2}^r$	$x_{3,3}^r$...	$x_{3,k}^r$
...
B_n	$x_{n,1}^r$	$x_{n,2}^r$	$x_{n,3}^r$...	$x_{n,k}^r$

На втором шаге находим сумму рангов по отдельным столбцам таблицы 22:

$$\begin{aligned}
 R_1 &= x_{1,1}^r + x_{2,1}^r + x_{3,1}^r + \dots + x_{n,1}^r; \\
 R_2 &= x_{1,2}^r + x_{2,2}^r + x_{3,2}^r + \dots + x_{n,2}^r; \\
 &\dots \\
 R_k &= x_{1,k}^r + x_{2,k}^r + x_{3,k}^r + \dots + x_{n,k}^r.
 \end{aligned} \tag{11.18}$$

Далее вычисляем величину $\lambda_{\text{экср}}$, заданную формулой

$$\lambda_{\text{экср}} = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1). \tag{11.19}$$

На третьем шаге фиксируем уровень значимости нуль-гипотезы (α) и сравниваем величину $\lambda_{\text{экср}}$ с $\lambda_{\text{крит}}$, которая выбирается из таблиц критических значений для критерия Фридмана (см. приложение 2Р). Если параметр n достаточно велик, то $\lambda_{\text{крит}}$ можно взять из таблиц критических значений для χ^2 -критерия Пирсона при числе степеней свободы $df = k - 1$. При выполнении неравенства $\lambda_{\text{экср}} \leq \lambda_{\text{крит}}$ нуль-гипотезу принимаем. Значит, в этом случае нет оснований говорить о влиянии фактора А на результаты измерений параметра при условиях, определенных уровнями фактора В. Если окажется, что $\lambda_{\text{экср}} > \lambda_{\text{крит}}$, то нуль-гипотезу отклоняем и считаем, что такое влияние имеет место. В частности, влияние фактора А проявляется в изменении математических ожиданий результатов измерений. Отметим, что формула (11.19) применяется

в случае, когда данные таблицы 21 не содержат одинаковых по значениям вариант или количество групп совпадающих вариант не значительно. В противном случае величина $\lambda_{\text{экс}}$ корректируется с учетом таких групп (подробнее см. [33]–[35]). Аналогичным образом проверяется еще одна нуль-гипотеза, а именно: что влияние фактора B отсутствует (при этом фактор A рассматривается как фактор, определяющий условия проведения эксперимента). Для проверки этой гипотезы достаточно поменять местами факторы A и B , преобразовать соответствующим образом таблицы 21, 22 и провести необходимые вычисления.

Пример 11-6. Вернемся к примеру 8-3, в котором описывались результаты исследования шести ингибиторов коррозии, проведенные в трех разных лабораториях. Для удобства изложения заново представим таблицу результатов (табл. 23).

Т а б л и ц а 23

Результаты межлабораторного эксперимента по ингибиторам

Уровни фактора B (данные разных лабораторий)	Уровни фактора A (скорость коррозии в присутствии разных ингибиторов)					
	A_1	A_2	A_3	A_4	A_5	A_6
B_1	1,90	2,30	2,60	1,80	2,00	2,05
B_2	2,02	2,10	3,20	2,70	4,00	2,40
B_3	1,60	1,50	1,80	2,60	1,90	1,70

Ранее мы предполагали, что выборочные данные получены из генеральной совокупности с нормальным распределением. Однако такое предположение невозможно проверить, опираясь только на таблицу 23. Поэтому для решения поставленной ранее задачи о выборе ингибитора следует применить один из критериев непараметрической статистики, а именно критерий Фридмана.

Р е ш е н и е. Для того чтобы сопоставить между собой эффективность ингибиторов, необходимо оценить влияние фактора A . При интерпретации получающихся результатов следует учесть фактор B , так как различие данных по лабораториям может быть связано с различием методики исследований. Рассмотрим нуль-гипотезу, состоящую в том, что математические ожидания остаточной скорости

коррозии не зависят от типа ингибитора. Принятие нуль-гипотезы будет означать отсутствие влияния интересующего нас фактора A при условии, что измерения проводились в лабораториях B_1, B_2, B_3 . Найдем ранги вариантов, стоящих в первой, второй и третьей строках приведенной таблицы. После этого получаем следующую таблицу, отражающую указанные ранги (табл. 24).

Т а б л и ц а 24

Результаты межлабораторного эксперимента по ингибиторам, представленные в ранговой форме

Уровни фактора B	Уровни фактора A					
	A_1	A_2	A_3	A_4	A_5	A_6
B_1	2	5	6	1	3	4
B_2	1	2	5	4	6	3
B_3	2	1	4	6	5	3

В данном случае $k = 6$, $n = 3$. Применяя формулу (11.18), находим, что сумма квадратов рангов по всем столбцам равна $5^2 + 8^2 + 15^2 + 11^2 + 14^2 + 10^2 = 731$. Используя формулу (11.19), получаем $\lambda_{\text{экс}} = 6,63$. Задавая $\alpha = 0,05$, из таблицы Р приложения 2 находим, что $\lambda_{\text{крит}} = 9,86$. В данном случае $\lambda_{\text{экс}} < \lambda_{\text{крит}}$. Следовательно, нуль-гипотезу принимаем. Таким образом, оба подхода (двухфакторный дисперсионный анализ и критерий Фридмана) говорят о том, что ни один из изученных ингибиторов не имеет достоверных преимуществ перед другими.

11.5.

ПРОВЕРКА ВЗАИМОСВЯЗИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЯ ДВУХ ВЕЛИЧИН

Критерий Спирмена. Одной из важнейших задач обработки данных является изучение возможной взаимосвязи между попарно связанными результатами измерений некоторых случайных величин x и y . Исходные данные могут быть записаны в виде попарно связанных выборок X и Y :

$$X: x_1, x_2, \dots, x_n; \quad Y: y_1, y_2, \dots, y_n. \quad (11.20)$$

Каждая пара вариант (x_i, y_i) характеризует некоторый объект, для которого известны (измерены) два признака.

Нередко взаимосвязь между результатами измерений величин x и y по всей совокупности исследуемых объектов нельзя выразить функциональной зависимостью, т. е. некоторым уравнением вида $y = f(x)$, но в совокупности данных легко обнаруживается монотонная зависимость. Это означает, что возрастанию одной величины, как правило, соответствует возрастание (или убывание) другой. Понятие монотонной зависимости близко понятию корреляции (см. раздел 8.1), но является более общим.

Проверка наличия монотонной зависимости требует вычисления коэффициента ранговой корреляции по Спирмену. Известно, что числовые значения выборочного коэффициента корреляции для любых попарно связанных выборок лежат в промежутке $-1 \leq R_s \leq +1$. Если между величинами x и y имеется строго монотонная возрастающая зависимость, то $R_s = +1$, если строго монотонная убывающая зависимость, то $R_s = -1$. Значения R_s , близкие к нулю, говорят об отсутствии монотонной зависимости.

Проверяемая нуль-гипотеза состоит в том, что между величинами x и y монотонная зависимость отсутствует. Алгоритм проверки следующий.

На первом шаге подсчитывают ранги вариант выборок X и Y из (11.20). Полученные результаты представляют в следующем виде:

$$X^r: x_1^r, x_2^r, \dots, x_n^r; \quad Y^r: y_1^r, y_2^r, \dots, y_n^r. \quad (11.21)$$

На втором шаге вычисляют выборочный коэффициент ранговой корреляции

$$R_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (x_i^r - y_i^r)^2. \quad (11.22)$$

На третьем шаге вычисляют величину статистического показателя:

$$\lambda_{\text{эксп}} = |R_s| \cdot \sqrt{\frac{n-2}{1-R_s^2}}. \quad (11.23)$$

Формула (11.23) предполагает, что $|R_s| \neq 1$ и объем данных достаточно велик ($n \geq 9$). В этом случае величину $\lambda_{\text{крит}}$ выбирают из таблиц Стьюдента при числе степеней свободы $df = n - 2$ и заданном α . Если окажется, что $\lambda_{\text{эксп}} > \lambda_{\text{крит}}$, то нуль-гипотеза отклоняется. Можно говорить о том, что между величинами x и y имеется ранговая корреляционная зависимость. При этом, если $R_s > 0$, то зависимость имеет характер монотонно возрастающей, а при $R_s < 0$ — монотонно убывающей. Если выполнено неравенство $\lambda_{\text{эксп}} \leq \lambda_{\text{крит}}$, то нуль-гипотеза принимается; считается, что между величинами x и y нет монотонной зависимости (ранговой корреляции). Заметим, что в этом случае зависимость между x и y все же может иметь место, но выражаться она будет в другой (не монотонной) форме. При малом объеме данных ($n < 9$) применять критерий Спирмена не рекомендуется.

Пример 11-7. Получены данные по временам удерживания 17 веществ в двух разных хроматографических колонках. Методика хроматографирования строго постоянна. Времена удерживания всех веществ выражены в условных единицах (миллиметры диаграммной ленты самописца). Значения времен удерживания веществ в первой колонке обозначены символом x , во второй — y . Проверить по критерию Спирмена наличие монотонной связи (ранговой корреляции) между значениями x и y . Для удобства анализа экспериментальные данные (x_i, y_i) упорядочены по значениям x и записаны в таблице 25.

Т а б л и ц а 25

Проверка взаимосвязи характеристик хроматографического удерживания 17 однотипных веществ на двух колонках

i	X	Y	i	X	Y	i	X	Y
1	28	5,3	7	45	39,3	13	73	94,8
2	29	9,2	8	51	52,7	14	75	101,9
3	32	15,2	9	53	55,4	15	80	89,5
4	35	20,7	10	58	64,3	16	83	114,8
5	40	21,7	11	64	76,0	17	93	137,4
6	44	36,5	12	65	79,1	18	—	—

Р е ш е н и е. Вычисляем ранги вариант x_i , y_i и коэффициент ранговой корреляции по формуле (11.22). Получаем $R_s = 0,993$. Используя (11.23), находим, что $\lambda_{\text{эксп}} = 32,5$. В таблице Стьюдента величина $\lambda_{\text{крит}} = 2,95$ при числе степеней свободы $df = 17 - 2 = 15$ и $\alpha = 0,01$. Поскольку верно неравенство $\lambda_{\text{эксп}} > \lambda_{\text{крит}}$, то нуль-гипотеза отклоняется. Можно утверждать, что существует монотонно возрастающая зависимость между временами хроматографического удерживания веществ данного типа на двух использованных колонках.

Из таблицы 25 сразу было видно, что практически для всех пар веществ росту X соответствовало возрастание Y . Однако в других случаях монотонная зависимость (ранговая корреляция) может быть не столь очевидной, и проверка по Спирмену будет иметь гораздо большую актуальность.

Отметим, что числовые данные примера 11-7 можно было бы обработать не только по критерию Спирмена, но и воспользовавшись традиционным алгоритмом корреляционного анализа. В этом случае коэффициент линейной корреляции равнялся бы 0,988, достоверно превышая критический уровень. То есть в координатах (x, y) точки, представляющие разные вещества, группируются относительно некоторой прямой. Но применять такой метод можно было бы только при условии, что результаты измерений характеристик удерживания подчиняются нормальному распределению, а в условии данной задачи об этом ничего не сказано. С другой стороны, при проверке ранговой корреляции (в отличие от проверки линейной корреляции) не требуется априорно указывать вид функциональной зависимости (линейная или какая-либо нелинейная), которая связывала бы величины x и y в отсутствие погрешностей измерений и неконтролируемого воздействия других факторов. Поэтому критерий Спирмена — гораздо более универсальный критерий, чем коэффициент линейной корреляции.

Установление наличия монотонной зависимости (ранговой корреляции) не позволяет прогнозировать свойства объектов, однотипных ранее исследованным. В частности, если в примере 11-7 взять $x_{18} = 120$, большее, чем x_{17} , можно уверенно ожидать, что ожидаемое значение y_{18} будет не

меньше, чем $y_{17} = 137,4$, но предвидеть точное значение y_{18} мы не сможем. Выполнение обратного неравенства ($y_{18} < y_{17}$) будет свидетельствовать об ошибках при измерении, о нарушении условий проведения эксперимента или, что бывает чаще всего, о неконтролируемом воздействии других факторов. Так, в примере 11-7 закономерное изменение времен удерживания, установленное для данных колонок по 17 веществам одного класса, могло бы резко нарушиться, если бы следующее, 18-е вещество относилось бы к иному классу веществ.

В заключение подчеркнем, что одни и те же задачи можно решать и параметрическими и непараметрическими методами. Так, коэффициент ранговой корреляции и критерий Спирмена используют для решения задач, которые при нормальном распределении вариант решались бы методами корреляционного анализа. Критерий равенства дисперсий аналогичен критерию Фишера, критерий Смирнова — Граббса — Q -тесту и т. д. Существуют также многочисленные и достаточно сложные непараметрические методы, которые позволяют решать те задачи, которые при нормальном распределении вариант решались бы методами регрессионного анализа. В настоящем пособии такие методы не рассматриваются, но с ними можно познакомиться, используя специальную литературу, например монографии [34], [35].

ЗАКЛЮЧЕНИЕ

Приведенный в этой книге теоретический материал преимущественно относится к статистическому моделированию сравнительно несложных химических процессов, изучаемых и/или оптимизируемых в ходе многофакторного активного эксперимента. Несомненно, это очень важная задача, с ней постоянно сталкиваются многие исследователи, независимо от того, с какими именно объектами они работают. Однако эта проблема далеко не единственная; статистические модели и, в частности, многомерные регрессии применяют не только для оптимизации каких-либо процессов, но и во многих других случаях. Можно назвать моделирование работы технологических установок, поиск спектро-структурных корреляций, прогнозирование биологической активности веществ по структуре их молекул, прогнозирование состава объектов с заданными свойствами и др. Но, пожалуй, наиболее интересной и многообещающей областью исследований на стыке химии и математики является *хеометрика*. Это научное направление является развитием классических методов статистической обработки экспериментальных данных. Поэтому в конце нашей книги мы хотели бы кратко рассказать о возможностях хеометрики и дать необходимые ссылки и рекомендации тем, кто захочет использовать хеометрические алгоритмы в своих будущих исследованиях.

Хеометрика возникла в 1970-х гг. По мнению ее основоположников (С. Волд, Б. Ковальски, Р. Бреретон), целью хеометрических исследований является извлечение важной для химика информации из больших массивов данных,

полученных в ходе эксперимента и представленных в виде двумерных или многомерных матриц [39]. Хемометрические алгоритмы стали активно использоваться специалистами на рубеже веков, когда лаборатории получили автоматизированные приборы, быстро и точно измеряющие целый ряд параметров и записывающие дискретные результаты измерений в цифровой форме (спектры, масс-спектры, хроматограммы, вольтамперные кривые и т. п.). Одновременно появились многоцелевые компьютерные программы для статистической обработки многомерных данных, примером может быть программа The Unscrambler Client 9.8. Хемометрические алгоритмы позволяют быстро выявить скрытые связи между разными переменными, формировать адекватные многомерные модели (без специального планирования эксперимента!) и успешно решать с их помощью всевозможные идентификационные и классификационные задачи.

Классификационные задачи и способы их решения студенты-химики обычно не изучают, но такие задачи широко распространены в практике (маркировочный анализ промышленной продукции, криминалистическая экспертиза, лабораторная диагностика и др.). Для решения таких задач был создан метод распознавания образов и многочисленные хемометрические алгоритмы, реализующие его. Особую популярность приобрели два алгоритма: метод главных компонент (ГК) и метод проекции на латентные структуры (ПЛС). Последний метод нечувствителен к отклонениям от аддитивности сигналов, вызванным взаимодействием компонентов. Известны и другие проекционные методы обработки многомерных числовых совокупностей, в том числе алгоритмы декомпозиции и сопоставления спектров.

Особенно широко применяют хемометрические алгоритмы химики-аналитики (см. обзор [40]) и химики-технологи, создавшие особую «хемометрическую» систему управления процессами (РАТ) [41]. Аналитики используют хемометрические алгоритмы в основном для обработки, сопоставления и интерпретации совокупностей аналитических сигналов. Цель — извлечение информации о качественном и количественном составе исследуемых систем, а также оценка достоверности этой информации.

Вероятно, крупнейшим достижением в этой области можно считать «электронный язык» — мультисенсорную систему анализа жидкостей. Она была создана в середине 1990-х гг. группой исследователей из Санкт-Петербургского университета под руководством Ю. Г. Власова [42]. Системы типа «электронный язык» теперь применяют для распознавания и классификации пищевых продуктов (соки, минеральные воды, вина, растительные масла, кофе, чай и т. п.), лекарственных препаратов, поверхностно-активных веществ и даже нефтепродуктов. Кроме того, с помощью «электронного языка» объективно оценивают вкус напитков (замена дегустаторов!), выявляют фальсификаты, контролируют процесс очистки сточной воды и решают многие другие прикладные задачи.

Хеометрические алгоритмы применяют и в количественном анализе для построения многомерных градуировок, связывающих массив аналитических сигналов (например, оптические плотности при разных длинах волн) с массивом данных по содержанию компонентов в многокомпонентных градуировочных образцах. Примером может быть используемая в нефтехимической промышленности методика экспрессного спектрометрического анализа бензинов без их разделения, по спектрам поглощения пробы в ближней ИК-области [43]). Для создания многомерной градуировки авторы использовали метод ПЛС. В дальнейшем по спектру бензина определяются содержания аренов, алканов и циклоалканов, причем погрешности анализа не превышают 8% отн. Продолжительность анализа единичной пробы — всего 5–10 мин.

ПРИЛОЖЕНИЯ

Приложение 1

ПРИМЕРЫ КОНТРОЛЬНЫХ ЗАДАНИЙ

Освоение изложенного в настоящем пособии учебного материала требует его закрепления в ходе выполнения нескольких индивидуальных контрольных заданий. Такие задания в обязательном порядке выполняют студенты (или аспиранты) ряда российских вузов, где в программу подготовки химиков введено изучение дисциплины «Планирование и математическая обработка результатов эксперимента». Каждое задание должно отражать материал одного из разделов данной дисциплины. Так, в Омском госуниверситете им. Ф. М. Достоевского все студенты-химики выполняют по три контрольных задания, а именно:

1) статистическая обработка результатов измерений (одномерные выборки);

2) корреляционный, дисперсионный и регрессионный анализ;

3) планирование многофакторного эксперимента. Построение, проверка и интерпретация статистических моделей. Метод крутого восхождения.

Каждое задание включает от 3 до 6 типовых расчетных задач. Студент выполняет свое задание самостоятельно, вне сетки расписания, дома или в компьютерном классе. На выполнение одного задания отводится примерно три недели. Решение задач требует знания теоретического материала, поиска недостающих данных в литературе, применения компьютерных расчетных программ, осмысления

полученных результатов. При выполнении задания студент может использовать любые компьютерные программы и любые учебные пособия, а также может получить консультацию преподавателя. В дальнейшем студент должен будет оформить развернутый отчет, сдать его на проверку, а затем пройти индивидуальное собеседование с преподавателем по итогам выполнения данного задания. Расчетные задания помогают организовать текущий и рубежный контроль усвоения материала и сформированности компетенций.

Желательно, чтобы химическое содержание индивидуальных заданий было бы связано с профилем подготовки данного магистранта или аспиранта. Именно такие (профилизованные) задания приведены далее в качестве примеров. С другой стороны, способы решения задач должны основываться на одних и тех же математических алгоритмах, независимо от того, идет ли в задаче речь об анализе, синтезе или другом процессе. Разумеется, расчетные задачи имеют условный характер, они не претендуют на воспроизведение условий и результатов реальных химических или технологических исследований. Приводимые далее примеры контрольных заданий апробированы в учебном процессе, с ними успешно справлялись студенты-химики разных вузов.

Расчетное задание № 1 (вариант для будущих химиков-аналитиков)

1. Для изучения загрязненности почв пестицидами планируется использовать некоторую методику хроматографического анализа. Перед началом эксперимента проведена метрологическая аттестация методики. Было отобрано 20 проб одной и той же почвы. Получены следующие результаты — содержания некоторого пестицида (мг/кг): 2,0; 2,5; 2,3; 2,4; 2,6; 2,4; 1,4; 2,6; 2,5; 2,6; 2,5; 2,3; 2,6; 2,3; 2,2; 2,3; 2,4; 2,8; 2,0; 2,5. Имеются ли в данной выборке грубые промахи? Для проверки используйте не менее двух разных критериев. При каких условиях возможно применение каждого из выбранных критериев, одинаковую ли надежность отбраковки промахов они обеспечивают? Охарактеризуйте содержание пестицида в данной почве в виде доверительного интервала.

2. При определении свинца в воде атомно-абсорбционным методом получены следующие результаты (мкг/л): 48; 65; 59; 58; 59; 29; 37; 88; 34; 48; 56; 20; 29; 57; 28; 29; 27; 36; 28; 58; 28; 56; 23; 56; 28; 29; 27; 56; 56; 18; 29; 25; 56; 36; 58; 29; 28; 65; 45; 56. Пробы отобраны в разные дни из одного источника и проанализированы по одной и той же методике. Постройте гистограмму и сделайте вывод о характере распределения экспериментальных данных. Можно ли считать, что методика приводит к нормальному распределению результатов? В ходе расчетов используйте не менее двух разных критериев.

3. В ходе проведения арбитражного анализа бензина на содержание в нем свинца (мг/л) в лаборатории А и Б отосланы пробы, взятые из одного резервуара. В ходе определений была использована спектрофотометрическая методика. Получили следующие результаты: А: 0,026; 0,027; 0,034; 0,024; 0,026; 0,028; 0,020; 0,029; 0,025; 0,026. Б: 0,029; 0,029; 0,024; 0,029; 0,033; 0,024; 0,022; 0,023; 0,024; 0,025. Сравните результаты работы двух лабораторий по средним значениям соответствующих выборок и по прецизионности. Можно ли указать, какая лаборатория получила более правильные результаты?

4. При хроматографическом определении содержания этанола в пищевом спирте среднее содержание этанола в 14 параллельных пробах составило 89%. Результаты распределены по нормальному закону. Для данной методики коэффициент вариации составляет 5%. А. Рассчитайте содержание этанола в виде доверительного интервала с надежностью 0,90. Б. Оцените вероятность того, что в единичном опыте результат анализа окажется меньше 85%, что приведет к признанию спирта не соответствующим ГОСТ.

5. Вискозиметрическое определение вязкости некоторого масла должно проводиться с точностью до четырех единиц. Величина стандартного отклонения при однократном определении составляет 10 единиц. Обеспечит ли проведение трех повторных измерений и усреднение результатов оценку вязкости с требуемой точностью, если надежность результатов должна быть равна 0,95?

6. При проведении полного факторного эксперимента параметр оптимизации определяли по методике, для ко-

торой коэффициент вариации составляет 2%. Значения параметра определяли после каждого из 16 опытов, проведенных в разных условиях. Получены следующие значения параметра (в условных единицах): 3,5; 3,8; 3,7; 3,9; 4,1; 4,4; 4,3; 4,2; 5,0; 5,2; 5,6; 5,0; 4,8; 5,1; 6,7; 4,8. Установите, есть ли достоверное изменение параметра при переходе от одних условий к другим, или все различия полученных результатов можно объяснить случайными погрешностями при его измерении. Повторите решение задачи, считая, что методика измерений параметра характеризуется коэффициентом вариации, равным 20%.

Расчетное задание № 2 (вариант для будущих физико-химиков)

1. Проверьте наличие линейной корреляции между скоростью некоторой реакции v (выражена в условных единицах) и концентрацией (C , ммоль/л) реагента:

v	6,08	5,30	7,99	7,89	7,04	4,32	5,53	5,48	5,15
C	7,73	9,10	3,83	4,87	3,99	5,94	3,65	3,68	3,86

2. Адсорбция некоторого газа на активированном угле (C , ммоль на 1 г угля) в зависимости от давления газа (ГПа) описывается следующими данными:

P	65	124	182	241	305	362	421	483	541
C	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0

Выберите теоретическое уравнение (модель), наилучшим образом описывающее полученные данные (после их линеаризации). Можно воспользоваться моделью Генри, Ленгмюра или Фрейндлиха. По выбранной модели прогнозируйте величину адсорбции при $P = 342$ ГПа.

3. При изучении влияния температуры T на константу скорости некоторой реакции (K) получены следующие данные:

T	291	309	327	345	363	381	399
K	0,02	0,12	0,58	2,46	9,06	29,46	86,18

Проверьте, существует ли достоверная линейная корреляция между: 1) величиной константы скорости и тем-

пературой; 2) величиной натурального логарифма константы и обратной температурой. Выполняется ли закон Аррениуса для данной реакции? Определите коэффициенты уравнения Аррениуса. Рассчитайте по найденной модели константу скорости при $T = 354$.

4. Предполагается, что примеси олова в поверхностном слое железо-ванадиевого катализатора могут влиять на активность катализатора в некоторой реакции. На лабораторной установке изучали кинетику этой реакции с применением четырех однотипных катализаторов с разным содержанием олова. Для каждого катализатора проводили по пять параллельных опытов. Получены следующие результаты:

$C_{\text{Sn, \%}}$	Скорость реакции, условные единицы				
0,0	38,1	33,0	36,9	34,2	34,8
0,2	41,7	43,5	38,2	35,1	37,8
0,4	39,5	41,9	38,6	46,2	42,0
0,6	42,0	44,8	44,0	48,6	43,8

Проверьте, существует ли достоверное ($P = 0,95$) влияние олова на скорость реакции. Воспользуйтесь методом однофакторного дисперсионного анализа.

Расчетное задание № 3 (вариант для будущих химиков-технологов)

1. Исследуется процесс превращения метанола в формальдегид. Необходимо получить статистическую модель, связывающую активность некоторого катализатора с его дисперсностью и пористостью, а также с температурой процесса. В распоряжении экспериментатора есть образцы катализатора с различным средним радиусом зерна R (в пределах от 0,05 до 1 см) и с разным средним радиусом пор r_i (в пределах от 50 до 500 нм). Возможно проведение реакции в интервале температур от 200 до 600°C. Самостоятельно задайте погрешности измерений дисперсности и пористости и выберите координаты центра плана и интервалы варьирования факторов. Предложите рабочий план проведения полного факторного эксперимента. Перечис-

лите факторы, которые следует держать на постоянном уровне.

2. Исследуют процесс каталитического разложения метана. Определены значения степени конверсии (R) при разных значениях M — молярного соотношения реагирующих веществ и разных значениях, T — температуры в реакторе. В ходе полного факторного эксперимента получены следующие результаты:

№	$M, \text{CH}_4 : \text{H}_2\text{O}$	$T, ^\circ\text{C}$	Степень конверсии, R
1	1,0	600	0,19
2	0,4	600	0,85
3	1,0	800	0,66; 0,71; 0,69; 0,68
4	0,4	800	0,95

Рассчитайте математическую модель процесса. Укажите, действуют ли факторы M и T независимо друг от друга. Рассчитайте степень конверсии при молярном соотношении $M = 0,5$ и температуре $T = 730^\circ\text{C}$. Взяв условия этого опыта в качестве исходной точки, проведите оптимизацию ($R \rightarrow 1$) методом крутого восхождения.

3. Процесс получения некоторого продукта адекватно описывается интерполяционной формулой:

$$M = 26,5 + 0,72T + 37,9P - 40,9C - 15CP,$$

где M — масса продукта (кг), получаемого за один час в некотором реакторе; T — температура ($^\circ\text{C}$); P — давление (атм); C — концентрация некоторой ингибирующей примеси в используемом сырье (г/л). При каких температурах будет достигнута максимально возможная производительность реактора (300 кг/ч) при атмосферном давлении, если: а) примеси в сырье не будет; б) содержание примеси будет составлять 2 г/л; в) содержание примеси будет составлять 10 г/л?

Приложение 2

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

А. Критические значения коэффициента асимметрии (по данным [26, с. 284]).

n	$\alpha = 0,05$	$\alpha = 0,01$
25	0,71	1,06
30	0,66	0,98
35	0,62	0,92
40	0,59	0,87
45	0,56	0,82
50	0,53	0,79
100	0,39	0,57

В. Критические значения коэффициента эксцесса (по данным [26, с. 285]).

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
11	0,89	0,91	0,94
16	0,87	0,89	0,91
21	0,86	0,88	0,90
31	0,85	0,86	0,88
41	0,84	0,85	0,87
51	0,84	0,85	0,86
101	0,83	0,83	0,85

Проверка нормальности распределения по критериям асимметрии и эксцесса при малых n ненадежна!

С. Критические значения χ^2 -критерия Пирсона в зависимости от уровня значимости (α) и числа степеней свободы (df) (по данным [20, с. 525–526]).

df	$\alpha = 0,99$	$\alpha = 0,95$	$\alpha = 0,90$	$\alpha = 0,50$	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
2	0,02	0,10	0,21	1,39	4,60	5,99	9,21
3	0,11	0,35	0,58	2,37	6,25	7,82	11,34
4	0,30	0,71	1,06	3,36	7,78	9,49	13,28
5	0,55	1,14	1,61	4,35	9,24	11,07	15,09
6	0,87	1,64	2,20	5,35	10,64	12,59	16,81
7	1,24	2,17	2,83	6,35	12,02	14,07	18,48
8	1,65	2,73	3,49	7,34	13,36	15,51	20,09
9	2,09	3,32	4,17	8,34	14,68	16,92	21,67
10	2,56	3,94	4,86	9,34	15,99	18,31	23,21
15	5,23	7,26	8,55	14,34	22,31	25,00	30,58
20	8,26	10,85	12,44	19,34	28,41	31,41	37,57
50	29,71	34,76	37,69	49,33	63,17	67,50	76,15

Д. Значения функции Лапласа.

Приведены значения функции $\Phi(U)$, характеризующие вероятность попадания случайной величины в интервал $(0, U)$ при ее нормальном распределении [22, с. 164]. Геометрически функция Лапласа соответствует доле площади под кривой нормированного нормального распределения, ограниченной абсциссами $(0, U)$. Величина $\Phi(U)$ вычисляется как определенный интеграл:

$$\Phi(U) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^U e^{-\frac{u^2}{2}} du.$$

U	$\Phi(U)$	U	$\Phi(U)$	U	$\Phi(U)$
0,01	0,0040	0,10	0,0398	0,30	0,1179
0,03	0,0120	0,15	0,0596	0,35	0,1368
0,05	0,0199	0,20	0,0793	0,40	0,1554
0,07	0,0279	0,25	0,0987	0,45	0,1736

Продолжение табл.

U	$\Phi(U)$	U	$\Phi(U)$	U	$\Phi(U)$
0,50	0,1915	1,40	0,4032	2,20	0,4861
0,55	0,2088	1,35	0,4115	2,30	0,4893
0,60	0,2257	1,40	0,4192	2,40	0,4918
0,65	0,2422	1,45	0,4265	2,50	0,4938
0,70	0,2580	1,50	0,4332	2,60	0,4953
0,75	0,2734	1,55	0,4394	2,70	0,4965
0,80	0,2881	1,60	0,4452	2,80	0,4974
0,85	0,3023	1,65	0,4505	2,90	0,4981
0,90	0,3159	1,70	0,4554	3,00	0,49865
0,95	0,3289	1,75	0,4599	3,10	0,49903
1,00	0,3413	1,80	0,4641	3,20	0,49931
1,05	0,3531	1,85	0,4678	3,40	0,49966
1,10	0,3643	1,90	0,4713	3,60	0,49984
1,15	0,3749	1,95	0,4744	3,80	0,499928
1,20	0,3849	2,00	0,4772	4,00	0,499968
1,25	0,3944	2,10	0,4821	5,00	0,499997

Е. Критические значения коэффициента Стьюдента (по данным [22, с. 165]).

Приведены значения аргумента функции $\Phi(t, df)$ для разного числа степеней свободы df и некоторых значений уровня значимости $\alpha = 1 - P$, где P соответствует вероятности попадания случайной величины t в интервал $(0, t)$ для распределения Стьюдента при заданном числе степеней свободы. Геометрически функция $\Phi(t, df)$ соответствует доле площади под кривой Стьюдента, ограниченной абсциссами $(0, t)$.

df	$\alpha = 0,10$ $P = 0,90$	$\alpha = 0,05$ $P = 0,95$	$\alpha = 0,01$ $P = 0,99$
1	6,31	12,7	63,6
2	2,92	4,30	9,93
3	2,35	3,18	5,84
4	2,13	2,78	4,60

Продолжение табл.

df	$\alpha = 0,10$ $P = 0,90$	$\alpha = 0,05$ $P = 0,95$	$\alpha = 0,01$ $P = 0,99$
5	2,02	2,57	4,03
6	1,94	2,45	3,71
7	1,90	2,37	3,50
8	1,86	2,31	3,36
9	1,83	2,26	3,25
10	1,81	2,23	3,17
15	1,75	2,13	2,95
20	1,73	2,09	2,85
25	1,71	2,06	2,79
30	1,70	2,04	2,75
∞	1,65	1,96	2,58

Г. Критические значения Q-теста для разных уровней значимости α (по данным [37, с. 38]).

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
3	0,94	0,98	0,99
4	0,76	0,85	0,93
5	0,64	0,73	0,82
6	0,56	0,64	0,74
7	0,51	0,59	0,68
8	0,47	0,54	0,63
9	0,44	0,51	0,60
10	0,41	0,48	0,57

Использование этого параметрического критерия описано в разделе 6.4.1. Если выборка включает от 3 до 7 вариантов, используются формулы:

$$Q = \left| \frac{X_n - X_{n-1}}{X_n - X_1} \right| \quad \text{или} \quad Q = \left| \frac{X_2 - X_1}{X_n - X_1} \right|.$$

В случае выборки, включающей восемь и более вариантов, используются несколько иные формулы, а именно:

$$Q = \left| \frac{X_n - X_{n-1}}{X_n - X_2} \right| \quad \text{или} \quad Q = \left| \frac{X_2 - X_1}{X_{n-1} - X_1} \right|.$$

В обоих случаях первая формула используется для проверки наибольшей варианты X_n , а вторая — для наименьшей варианты X_1 . При $n > 10$ лучше использовать метод максимального отклонения. Приведенные в таблице значения Q -теста предназначены для одновременной отбраковки и слишком больших, и слишком маленьких вариантов (двусторонний критерий). Эти же значения используются в компьютерных программах, предназначенных для статистической обработки данных (например, в программах серии [38]) при отбраковке грубых промахов. Однако в статистической литературе приводятся и другие табличные значения Q -теста (например, в [11], [20]), но в этом случае он используется как односторонний критерий, например, для проверки вариант, высказывающих только в сторону больших значений.

Г. Критические значения максимального относительного отклонения $t_{\text{крит}}$ (по данным [22, с. 165–166]).

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
3	1,41	1,41	1,41
4	1,65	1,69	1,72
5	1,79	1,87	1,96
6	1,89	2,00	2,13
7	1,97	2,09	2,27
8	2,04	2,17	2,37
9	2,10	2,24	2,46
10	2,15	2,29	2,54
15	2,33	2,49	2,80
20	2,45	2,62	2,96
25	2,54	2,72	3,07

Результат измерения X считают грубым промахом, если вычисленная величина

$$\tau_{\text{эксп}} = \frac{|X - \bar{X}|}{S}$$

превышает критическое значение τ для выбранного уровня значимости α с учетом объема выборки n . Величину S следует рассчитывать с учетом сомнительной варианты. Критерий применяют для обработки нормально распределенных данных. Иногда рекомендуют вычислять не обычное, а «исправленное» значение τ :

$$\tau_{\text{эксп}} = \frac{|X - \bar{X}|}{S \cdot \sqrt{\frac{n-1}{n}}}$$

и сопоставлять его с вышеприведенными критическими значениями.

Н. Критические значения критерия Фишера.

1. Для уровня значимости 0,05.

$\frac{df_2}{df_1}$	1	2	3	4	5	6	10	16	24	∞
1	161	200	216	225	230	234	242	246	249	254
2	18,5	19,0	19,2	19,3	19,3	19,3	19,4	19,4	19,4	9,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,8	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	6,0	5,8	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,6	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,1	3,9	3,8	3,7
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	3,0	2,9	2,7
14	4,6	3,7	3,3	3,1	3,0	2,9	2,6	2,4	2,3	2,1
24	4,3	3,4	3,0	2,8	2,6	2,5	2,3	2,1	2,0	1,7
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,6	1,5	1,0

2. Для уровня значимости 0,01.

$\frac{df_2}{df_1}$	1	2	3	4	5	6	10	16	24	∞
1	4052	4999	5403	5625	5764	5859	6056	6169	6234	6366
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,5	99,5

Продолжение табл.

df_2/df_1	1	2	3	4	5	6	10	16	24	∞
3	34,1	30,8	29,5	28,7	28,2	27,9	27,2	26,8	26,6	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	14,5	14,1	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,1	9,7	9,5	9,0
6	13,7	10,9	9,8	9,2	8,8	8,5	7,9	7,5	7,3	6,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,3	4,9	4,7	4,3
14	8,9	6,5	5,6	5,0	4,7	4,5	3,9	3,6	3,4	3,0
24	7,8	5,6	4,7	4,2	3,9	3,7	3,2	2,9	2,7	2,2
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,2	2,0	1,8	1,0

Приведенные критические значения соответствуют данным [19, с. 246–249], но округлены до десятых долей. Дисперсии неоднородны, если отношение большей из них (первой) к меньшей (второй) превышает критическое значение F для выбранного уровня значимости с учетом числа степеней свободы df_1 и df_2 .

Ж. Критические значения критерия Кохрена [11].

1. Уровень значимости 0,05.

N	Значения критерия для разного числа степеней свободы								
	1	2	3	4	5	6	7	8	9
2	0,998	0,975	0,940	0,906	0,877	0,853	0,833	0,816	0,801
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	0,633	0,617
4	0,906	0,768	0,684	0,629	0,589	0,560	0,536	0,517	0,502
5	0,841	0,684	0,598	0,544	0,506	0,478	0,456	0,439	0,424
6	0,781	0,616	0,532	0,480	0,445	0,418	0,398	0,382	0,368
7	0,727	0,561	0,480	0,431	0,397	0,373	0,354	0,338	0,326
10	0,602	0,445	0,373	0,331	0,303	0,282	0,267	0,254	0,244
15	0,471	0,335	0,276	0,242	0,220	0,203	0,191	0,182	0,174
20	0,389	0,271	0,221	0,192	0,174	0,160	0,150	0,142	0,136

2. Уровень значимости 0,01.

N	Значения критерия для разного числа степеней свободы								
	1	2	3	4	5	6	7	8	9
2	0,999	0,995	0,979	0,959	0,937	0,917	0,900	0,882	0,867
3	0,993	0,942	0,883	0,834	0,793	0,761	0,734	0,711	0,691
4	0,968	0,864	0,781	0,721	0,676	0,641	0,613	0,590	0,570
5	0,928	0,788	0,696	0,633	0,587	0,553	0,526	0,504	0,485
6	0,883	0,722	0,626	0,564	0,520	0,487	0,461	0,440	0,423
7	0,838	0,664	0,518	0,508	0,466	0,435	0,410	0,391	0,375
10	0,717	0,536	0,447	0,393	0,357	0,331	0,311	0,294	0,281
15	0,575	0,407	0,332	0,288	0,259	0,239	0,2231	0,210	0,200
20	0,480	0,330	0,265	0,229	0,205	0,188	0,175	0,165	0,157

При сопоставлении N дисперсий одинакового объема (по n параллельных опытов в каждой из N выборок) они считаются неоднородными при заданном уровне значимости, если $G > G_{\text{табл}}$, где G — отношение максимальной дисперсии к сумме всех дисперсий. Число степеней свободы $df = n - 1$. В противном случае значимость нуль-гипотезы об однородности дисперсий превышает заданный уровень, все дисперсии считаются однородными.

К. Критерий знаков (по данным [26, с. 283]).

Указаны критические значения количества «плюсов», для разного числа m ненулевых разностей и для разных уровней значимости при сопоставлении попарно связанных выборок.

m	$\alpha = 0,05$	$\alpha = 0,01$	m	$\alpha = 0,05$	$\alpha = 0,01$	m	$\alpha = 0,05$	$\alpha = 0,01$
7	7	—	14	12	13	25	18	20
8	8	8	15	12	13	30	21	23
9	8	9	16	13	14	40	27	29
10	9	10	17	13	15	50	33	35
11	10	11	18	14	15	60	39	41
12	10	11	19	15	16	80	50	52
13	11	12	20	15	17	100	61	64

Л. Критические значения коэффициента линейной корреляции (по данным [26, с. 285]).

n	$\alpha = 0,05$	$\alpha = 0,01$	n	$\alpha = 0,05$	$\alpha = 0,01$	n	$\alpha = 0,05$	$\alpha = 0,01$
3	0,997	—	11	0,60	0,74	19	0,46	0,58
4	0,95	0,99	12	0,58	0,71	20	0,44	0,56
5	0,88	0,96	13	0,55	0,68	21	0,43	0,55
6	0,81	0,92	14	0,53	0,66	22	0,42	0,54
7	0,75	0,87	15	0,51	0,64	23	0,41	0,53
8	0,71	0,83	16	0,50	0,62	24	0,40	0,52
9	0,67	0,80	17	0,48	0,61	25	0,40	0,51
10	0,63	0,77	18	0,47	0,59	30	0,36	0,46

При $n < 7$ корреляционный анализ ненадежен!

М. Критические значения критерия Смирнова (по данным [35, с. 302]) в зависимости от уровня значимости α и объема выборки n .

n	$\alpha = 0,05$	$\alpha = 0,025$	n	$\alpha = 0,05$	$\alpha = 0,025$
3	1,412	1,414	15	2,493	2,638
4	1,689	1,710	16	2,523	2,670
5	1,869	1,917	17	2,551	2,701
6	1,996	2,067	18	2,577	2,728
7	2,093	2,182	19	2,600	2,754
8	2,172	2,273	20	2,623	2,778
9	2,237	2,349	21	2,644	2,801
10	2,294	2,414	22	2,664	2,823
11	2,343	2,470	23	2,683	2,843
12	2,387	2,519	24	2,701	2,862
13	2,426	2,562	25	2,717	2,880
14	2,461	2,602	26	2,734	2,897

Н. Критические значения критерия Граббса (по данным [35, с. 304]) в зависимости от уровня значимости α и объема выборки n .

n	$\alpha = 0,05$	$\alpha = 0,025$	n	$\alpha = 0,05$	$\alpha = 0,025$
4	0,0494	0,0248	15	0,5559	0,5030
5	0,1270	0,0808	16	0,5755	0,5246
6	0,2032	0,1453	17	0,5933	0,5442
7	0,2696	0,2066	18	0,6095	0,5621
8	0,3261	0,2616	19	0,6243	0,5785
9	0,3742	0,3101	20	0,6379	0,5937
10	0,4154	0,3526	21	0,6504	0,6076
11	0,4511	0,3901	22	0,6621	0,6206
12	0,4822	0,4232	23	0,6728	0,6327
13	0,5097	0,4528	24	0,6829	0,6439
14	0,5340	0,4792	25	0,6923	0,6544

О. Критические значения критерия Краскела — Уоллиса (по данным [33, с. 307, 308]) в зависимости от уровня значимости α и объемов выборок n_1, n_2, n_3 .

n_1	n_2	n_3	$\alpha = 0,05$	$\alpha = 0,01$
5	4	3	5,656	7,445
5	5	3	5,705	7,578
5	5	4	5,666	7,823
5	5	5	5,780	8,000
6	3	2	5,348	6,970
6	3	3	5,615	7,410
6	4	3	5,610	7,500
6	4	4	5,681	7,795
6	5	4	5,661	7,936
6	5	5	5,729	8,028
6	6	5	5,765	8,124
6	6	6	5,801	8,222
7	7	7	5,819	8,378
8	8	8	5,805	8,465

Р. Критические значения критерия Фридмана (по данным [33, с. 309–311]) в зависимости от уровня значимости α и числа уровней для факторов k и n .

k	n	$\alpha = 0,05$	$\alpha = 0,01$
3	5	6,40	8,40
3	6	7,00	9,00
3	7	7,14	8,86
3	8	6,25	9,00
4	4	7,80	9,60
4	5	7,80	9,96
4	6	7,60	10,20
5	3	8,53	10,13
5	4	8,80	11,20
5	5	8,96	11,68
6	3	9,86	11,76

Приложение 3

ПОДГОТОВКА НАУЧНЫХ ПУБЛИКАЦИЙ

Выдержки из статьи акад. Ю. А. Золотова, опубликованной в «Журнале аналитической химии», 1977. Т. 37. № 1. С. 170–178*.

Научные статьи — основная форма представления результатов фундаментальных и часто прикладных исследований. Основная, но, конечно не единственная — есть отчеты для узкого круга читателей, доклады, монографии, диссертации, заявки на изобретения. Результатом научной работы может быть новый прибор, внедренный процесс, метод, материал. Подчас все это важнее статей по тому же вопросу, однако значение журнальных публикаций все равно очень велико.

Практически любой научный сотрудник пишет статьи, и он должен уметь это делать. Методике научной работы, в том числе подготовке публикаций, учат мало, начинающие сотрудники испытывают в этом деле естественные трудности. Есть категория научных сотрудников, вообще не любящих писать, взирающих на подготовку очередного манускрипта как на тяжелое испытание. Пишут они обычно натужно, медленно, без блеска. Распространен и другой персонаж, особенно среди тех, кто делает первые шаги в науке. Его одолевает «влечение, род недуга», он мечтает

* Воспроизведено нами с разрешения Ю. А. Золотова.

о публикации, еще толком не начав работы, а писать статью начинает после того как получил первые пять экспериментальных точек. Крайности, однако, как всегда, нежелательны.

Статьи иногда путают с отчетами. Отчет о проведенной работе включает описание того, что делалось, что получилось. Однако далеко не все, что делалось, привело к существенным научным результатам, которые важны и интересны для достаточно широкого круга специалистов. Интересное и важное для нас может быть ненужным для других. Здесь должны быть проявлены объективность, здравый смысл, эрудиция и скромность научного сотрудника. Но вот накоплен обширный и ценный материал, факты проверены, а наиболее важные проверены не раз. Цифровые данные обработаны, обобщены. Найдено разумное истолкование фактов и закономерностей. Еще раз сопоставив свои данные и оформившиеся выводы с тем, что есть в литературе, приходим к выводу, что налицо заметный шаг вперед. У нас либо совокупность свежих и любопытных фактов, либо новая закономерность, либо теория, строгое доказательство чужой гипотезы, новый методический прием, либо все это вместе. Надо писать статью, и мы приступаем к работе. Статья не самоцель; это не более чем средство информации о результатах научной работы.

Название статьи. Выбрав журнал, мы должны дать название готовящейся работе. Можно озаглавить статью и потом, когда она будет написана, но лучше это сделать вначале. Тогда и материал будет подаваться под определенным, заранее установленным ракурсом. В эпоху информационного взрыва информативная ценность названия статьи повысилась. В заглавие должны входить ключевые слова, отражающие главные для статьи понятия. Оно должно по возможности полно, точно и достаточно лаконично выразить предмет работы. Не очень хорошо называть статью, скажем, так: «Способ определения циркония». Лучше писать конкретно, но недлинно, например: «Фотометрическое определение циркония с ксиленоловым оранжевым». В последнем заголовке каждое слово несет смысловую нагрузку, а вместе они почти исчерпывающе определяют предмет публикации.

План статьи и ее основной текст. Типовой план статьи включает следующие разделы:

- 1) введение;
- 2) экспериментальная часть;
- 3) полученные результаты;
- 4) обсуждение результатов;
- 5) выводы;
- 6) литература.

Теоретические статьи и обзоры пишут, конечно, по-другому, да и вообще схема статьи зависит от ее предмета. Например, план публикации, посвященной разработке новой методики, должен включать раздел, где эта методика (пропись) будет изложена (очевидно, после «Обсуждения результатов»).

Во введении дадим краткую характеристику значения и состояния вопроса, определим цель работы. Хорошо написать введение — нелегкая задача. Информация, которую здесь нужно дать, должна быть краткой, но емкой, изложенной четко. Важно правильно процитировать наших предшественников, не стараясь принизить их роль небрежным упоминанием вскользь. Формулируя задачу статьи, точнее работы, результаты которой приводятся, не нужно, видимо, повторять того, что будет написано в выводах.

Экспериментальная часть посвящается описанию того, как мы работали: какие использовали реактивы, приборы, как очищали препараты и готовили растворы, какова была методика исследования. Следующий раздел — основной в статье; здесь мы излагаем полученные результаты. Факты, по возможности только факты. Приводятся не все полученные результаты, а только основные. Излагаются они прежде всего словами, но нам хорошо известны и другие средства — рисунки, таблицы. Нужно рационально использовать их, но главным образом в тех случаях, когда словами изложить нужные сведения невозможно или это получается слишком длинно. Один и тот же материал не принято дублировать на рисунке и в таблице. Неплохо вводить более мелкие подразделы с заголовками в подбор, т. е. в основной строке, но со шрифтовым выделением. Наличие таких разделов организует автора и облегчает чтение статьи.

Те данные, которые целесообразно подвергнуть математической обработке, должны быть обработаны, например, свернуты в формулу. Часто приводят, скажем, градиуровочные графики в виде прямой линии на рисунке; во многих случаях компактнее выглядит уравнение соответствующей прямой.

В разделе «Обсуждение результатов» мы анализируем полученные данные, объясняем их, сопоставляем с литературными сведениями и оцениваем перспективы дальнейшей работы. К сожалению, некоторые авторы недостаточно четко отделяют факты от их интерпретации, а свои фактические результаты иногда дают вперемешку с литературными, не делая между ними четкого различия. В отдельных случаях такое перемешивание почти неизбежно, но обычно его следует избегать.

На что делать упор — на факты или на идейную сторону работы, зависит от многих факторов, в том числе от чисто субъективных причин. Большую роль играет просто стиль работы данного научного сотрудника. Отношение к факту и идее у многих различное. Сопоставим, например, мнения В. Г. Белинского и В. И. Вернадского. Белинский: «В науке должно искать идеи. Нет идеи, нет и науки. Знание фактов только потому и драгоценно, что в фактах скрываются идеи: факты без идей — сор для головы и памяти». Вернадский: «Точно, научно установленный факт по существу всегда дает больше, чем основанная на нем, его объясняющая теория. Он верен и для будущей теории и в исторической смене теорий он остается неизменным».

Терминология. Язык статьи. Каждый автор должен стремиться к использованию современной научной терминологии и системы обозначений. Химическую терминологию разрабатывает Международный союз теоретической и прикладной химии, однако эти публикации не охватывают всех аспектов. Во многих случаях приходится при первом упоминании термина или обозначения указывать, что под ним подразумевается.

Установилась традиция писать научные работы языком нейтральным, лишенным эмоций и образности, с преобладанием безличных оборотов, пассивных конструкций. Вероятно, это в известной мере оправдано. Такой язык призван, очевидно, подчеркнуть объективность научного

знания, оттенить беспристрастность, осторожность и ответственность исследователя. Однако даже такой «подсушенный» язык может быть хорошим и плохим. Недостатки языка современных научных статей включают: многословие, неумеренное использование вспомогательных глаголов (проводили, осуществляли и т. п.) и соответствующих отглагольных существительных (проведение и т. д.), неумелое использование пассивных конструкций (например, пишут «раствор титруется»), неловкое построение длинных фраз, неточное акцентирование, нанизывание родительных падежей.

Можно взять кусочек текста и проанализировать его.

«Нами проведено исследование по изучению процесса осаждения сульфата свинца. Оценка возможности проведения процесса осаждения сульфата свинца показала, что этот процесс вполне возможен. Осаждение осуществляется в результате добавления раствора нитрата свинца к раствору серной кислоты. Осаждение сульфата свинца происходит быстро, причем наблюдается осветление раствора над осадком. Исходя из результатов этих проведенных опытов, нам представлялось интересным провести работу по разработке гравиметрического метода определения сульфатов.

Берут анализируемый раствор и производят его нагревание до 60°C, после чего осуществляют постепенное приращивание 0,1 М раствора нитрата свинца. После фильтрования производят прокалывание и взвешивание осадка. Метод является очень точным».

Здесь множество языковых небрежностей и ошибок. В первом абзаце «исследование по изучению» — это масло масляное, достаточно либо «исследования», либо «изучения». «Нами проведено исследование» — не ошибка, но проще сказать «мы исследовали», без лишней пассивной конструкции. Слово «процесса» не нужно, можно написать просто «осаждение». Весь первый абзац лучше выглядел так: «Мы исследовали осаждение сульфата свинца».

Во втором абзаце многословие; мы уже знаем, об осаждении какого соединения идет речь, поэтому повторять «сульфата свинца» не нужно. Опять необязательное слово «процесс». Здесь мы имеем пример нанизывания родительных падежей: «Оценка (чего?) возможности (чего?)

проведения (чего?) процесса (чего?) осаждения (чего?) сульфата (чего?) свинца». В следующем абзаце неудачные, тяжелые фразы. Целесообразнее избегать выражения типа «осаждение осуществляется», «осаждение происходит», «наблюдается осветление». В четвертом абзаце есть некорректность, связанная с деепричастным оборотом; этот оборот плохо сочетается с безличным предложением. Здесь же «работа по разработке» — все равно что «исследование по изучению». Глагол «является» часто не украшает фразы.

Теперь мы можем отредактировать наш отрывок. Вот что получается.

«Мы исследовали осаждение сульфата свинца. Предварительная проверка показала, что осаждение вполне возможно. При добавлении раствора нитрата свинца к раствору серной кислоты осадок выпадает быстро, причем раствор над осадком светлеет. Результаты этих опытов мы использовали для разработки гравиметрического метода определения сульфатов. К анализируемому раствору, нагретому до 60°C, прибавляли постепенно 0,1 М раствор нитрата свинца. Осадок фильтровали, прокаливали и взвешивали. Метод очень точен».

Текст сокращен почти вдвое, ни одной мысли не утеряно, более того, материал лучше воспринимается.

Рисунки. Один из самых удобных способов представления информации — рисунок. Однако в статье не должно быть слишком много рисунков, обычно журналы ограничивают их число.

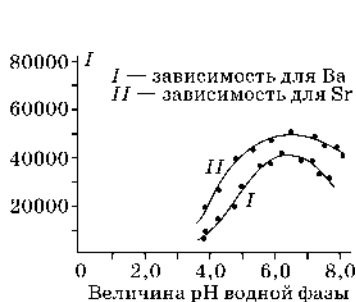


Рис. 1

Неправильно
подготовленный рисунок

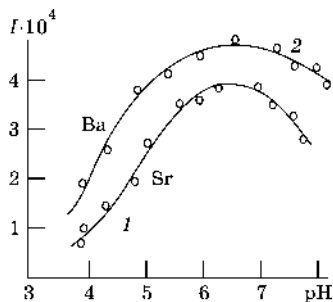


Рис. 2

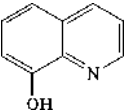
Правильно
подготовленный рисунок

Рассмотрим рисунок 1. Он сделан неправильно. Кривые занимают на нем немного места, значительная часть площади «не работает». Сами кривые принято обозначать арабскими цифрами, а не римскими, причем позиции нумеруют обычно по часовой стрелке. По оси ординат не следовало писать громоздких чисел: 80 000, 60 000 и т. д., лучше использовать более компактную форму: $8 \cdot 10^4$ и т. п. В числах по оси абсцисс лишние нули: не проще ли написать 0; 1; 2 и т. д.? Обозначение кривых дано прямо на рисунке; обычно издательства требуют переносить это в подпись. Величина I на оси ординат дана справа от оси, а нужно слева. Есть и другие неточности. Если мы учтем все сделанные замечания, получим рисунок 2, который удовлетворяет всем требованиям.

Таблицы. Редкая статья обходится без таблиц. Помещать в них следует по возможности обобщенные, итоговые данные, а не результаты первичных измерений, которым место в лабораторном журнале. Желательно приводить в таблицах только самые необходимые величины.

Головка таблицы не должна походить на небоскреб, два-три этажа — ее предел. Обычно излишней бывает графа «Номер по порядку». Если исходная величина повторяется во всех строках таблицы, соответствующую графу устраняют, а нужные сведения дают в заголовке таблицы или в тексте. Не следует приводить в таблицах структурные формулы. Таблица должна иметь заголовок, даже если о ней много говорится в тексте. Заголовку предшествует слово «Таблица» с соответствующим номером (арабские цифры); знак № не ставится. Если в статье одна таблица, слово «таблица» вообще не пишут, а в тексте его приводят полностью. Если же таблица имеет номер, в тексте пишут сокращенно: табл. 2. Желательно, чтобы таблицы были компактными, не вытянутыми по горизонтали или по вертикали. Вот пример не вполне удачно составленной таблицы.

Т а б л и ц а IV

№ п/п	Ре- агент	Формула реагента	Растворимость			
			реагента		комплекса с ионами Cu ²⁺	
			в воде	в CHCl ₃	в воде	в CHCl ₃
1	Дити- зон	$ \begin{array}{l} \text{S} = \text{C} \begin{cases} \text{NH} - \text{NH} - \text{C}_6\text{H}_5 \\ \text{N} = \text{N} - \text{C}_6\text{H}_5 \end{cases} \end{array} $	0,01 М	0,2 М	0,01 М	0,2 М
2	Окси- хино- лин		0,1 М	0,3 М	0,01 М	2 М

Видимо, не нужна первая графа — номера по порядку. Едва ли следовало вводить в таблицу структурные формулы химических соединений. Головка таблицы сложновата. В последних четырех графах после чисел везде указана молярность — ее нужно указать один раз в названии таблицы. По-разному обозначены химические соединения, в головке: вода — словом, хлороформ — формулой. Отсутствует заголовок таблицы, ее номер дан римскими цифрами. Таблицу эту мы легко можем отредактировать.

Т а б л и ц а 4

Растворимость (S , моль/л) реагентов (НА) и их комплексов с медью (CuA₂)

Реагент	$S_{\text{НА}}$		S_{CuA_2}	
	в H ₂ O	в CHCl ₃	в H ₂ O	в CHCl ₃
Дитизон	0,01	0,2	0,02	1,0
Оксихинолин	0,1	0,3	0,01	2,0

Иногда, чтобы представить 4–5 чисел, строят сложную таблицу, а между тем об этих величинах проще сказать прямо в тексте.

Выводы. В большинстве статей полагаются выводы или аннотации. Аннотацию дают перед основным текстом, вы-

воды — после него. Выводы даются с заголовком, аннотацию обычно без заголовка, ее отделяют от основного текста отбивкой, т. е. небольшим пропуском. В одних журналах приняты аннотации, в других — выводы. В выводах должны быть кратко и четко сформулированы основные результаты работы, по возможности конкретно, иногда с цифрами. Такие выводы несколько приближаются к реферату. Если мы определили константу устойчивости, не стоит писать «определена константа устойчивости», резоннее сказать «константа устойчивости равна» и дать численное значение. Не нужны слова типа «В результате проведенного исследования установлено, что...», или «Показано, что...». Сокращения названий в выводах нежелательны. В большинстве случаев нет смысла ставить номера абзацев. Более того, иногда делить выводы на абзацы вообще нет нужды. Краткость в научной работе ценится очень высоко.

Ссылки на литературу. Почти в каждой статье есть ссылки на литературу. У научных работников различных специальностей сложились свои традиции оформления библиографии. Биологи и медики часто располагают литературные источники по алфавиту с указанием полного названия цитируемой статьи, а в тексте указывают фамилию автора этой статьи и в скобках — год опубликования. У химиков более лаконичная система: ссылки даются в порядке упоминания их в тексте, названия статей не приводятся. Даются только названия книг и диссертаций. При цитировании книги в большинстве случаев следует указывать нужные страницы.

В тексте статьи ссылку на литературный источник дают в разных журналах по-разному. Часто дают номер ссылки в квадратных скобках на основной строке. Начинающие научные работники иногда поддаются искушению «показать ученость свою» и в короткой статье дают по 30–40 ссылок. Как правило, в таком массивном цитировании нет необходимости. Целесообразно указывать работы, непосредственно относящиеся к теме статьи, а также обзоры и монографии, если они имеются. Совершенно обязательны ссылки на работы, которые являются прямыми предшественниками нашего исследования.

СПИСОК ЛИТЕРАТУРЫ

1. *Сивоконь, М. Е.* Методологические проблемы естественнонаучного эксперимента. — М. : МГУ, 1968. — 370 с.
2. *Макареня, А. А.* Методология химии / А. А. Макареня, В. П. Обухов. — М. : Просвещение, 1985. — 160 с.
3. *Бэйнз, А.* Организация исследований в химии и химической промышленности. Условия, цели и стратегия / А. Бэйнз, Ф. Бредбери, С. Саклинг. — М. : Химия, 1974. — 336 с.
4. *Адлер, Ю. П.* Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. — М. : Наука, 1976. — 279 с.
5. *Золотов, Ю. А.* История и методология аналитической химии / Ю. А. Золотов, В. И. Вершинин. — М. : Академия, 2007. — 464 с.
6. *Кузнецов, В. И.* Диалектика развития химии. — М. : Наука, 1973. — 328 с.
7. *Сидняев, Н. И.* Теория планирования эксперимента и анализ статистических данных. — М. : ИД Юрайт, 2011. — 399 с.
8. *Романенко, В. Н.* Книга для начинающего исследователя химика / В. Н. Романенко, А. Г. Орлов, Г. В. Никитина. — Л. : Химия, 1987. — 280 с.
9. *Маркова, Е. В.* Математическое планирование химического эксперимента / Е. В. Маркова, А. Е. Рохваргер. — М. : Знание, 1971. — 31 с.
10. *Штекли, А.* Галилей. — М. : Молодая гвардия, 1972. — С. 19.
11. *Смагунова, А. Н.* Методы математической статистики в аналитической химии : учеб. пособие / А. Н. Смагунова, О. М. Карпукова. — Ростов-на-Дону : Феникс, 2012. — 347 с.
12. *Волкова, П. А.* Статистическая обработка данных в учебно-исследовательских работах / П. А. Волкова, А. Б. Шипунов. — М. : Экспресс, 2008. — 60 с.
13. *Хикс, Ч.* Основные принципы планирования эксперимента. — М. : Мир, 1967. — 406 с.
14. *Островский, Н. М.* Кинетика дезактивации катализаторов. Математические модели и их применение. — М. : Наука, 2001. — 334 с.
15. *Дворкин, В. И.* Метрология и обеспечение качества химического анализа. — М. : Изд-во МИТХТ, 2014. — 416 с.

16. *Оствальд, В.* Физико-химические измерения / В. Оствальд, Р. Лютер, К. Друкер; пер. с нем. — Ч. 1. — Л.: ОНТИ-Химтеорет, 1935. — 379 с.
17. *Сергеев, А. Г.* Метрология, стандартизация, сертификация / А. Г. Сергеев, М. В. Латышев, В. В. Терегеря. — М.: Логос, 2005. — 560 с.
18. *Гармаш, А. В.* Метрологические основы аналитической химии / А. В. Гармаш, Н. М. Сорокина. — М.: МГУ, 2012. — 47 с.
19. *Дерффель, К.* Статистика в аналитической химии. — М.: Мир, 1994. — 267 с.
20. *Джонсон, Н.* Статистика и планирование эксперимента в технике и науке / Н. Джонсон, Ф. Лион. — М.: Мир, 1980. — Т. 1. — Методы обработки данных. Гл. 1, 2, 5, 8.
21. *Общая теория статистики* / под ред. А. Я. Боярского. — М.: Изд-во МГУ, 1985. — 375 с.
22. *Чарыков, А. К.* Математическая обработка результатов химического анализа. — Л.: Химия, 1984. — 167 с.
23. *Румшицкий, А. З.* Математическая обработка результатов эксперимента. — М.: Наука, 1971. — 192 с.
24. ГОСТ Р ИСО 5725-2002. Точность (правильность и прецизионность) методов и результатов измерений. — М.: Госстандарт, 2009. — 32 с.
25. ГОСТ Р 8.736-2011. Измерения прямые многократные. Методы обработки результатов измерений. Основные положения. — М.: Стандартинформ, 2013. — 19 с.
26. *Лакин, Г. Ф.* Биометрия. — М.: Высш. шк., 1980. — 292 с.
27. *Вершинин, В. И.* Компьютерная идентификация органических соединений / В. И. Вершинин, Б. Г. Дерендяев, К. С. Лебедев. — М.: Академкнига, 2002. — 197 с.
28. *Физико-химические методы анализа: практ. пособие* / под ред. В. Б. Алексовского и К. Б. Яцимирского. — Л.: Химия, 1971. — С. 24–54.
29. *Рузинов, Л. П.* Планирование эксперимента в химии и химической технологии / Л. П. Рузинов, Р. И. Слободчикова. — М.: Химия, 1980. — 280 с.
30. *Грановский, Ю. В.* Основы планирования экстремального эксперимента для оптимизации многофакторных технологических процессов. — М.: МИНХ, 1971. — 71 с.
31. *Горский, В. Г.* Новые идеи в планировании эксперимента / В. Г. Горский, В. З. Бродский. — М.: Наука, 1969. — С. 54.
32. *Вершинина, О. И.* Симплексная оптимизация с обобщенным параметром при кинетическом определении молибдена / О. И. Вершинина, И. Л. Надеяева, В. И. Вершинин // Изв. ВУЗов СССР. Химия и химическая технология. — 1986. — Т. 29. — № 8. — С. 35–37.
33. *Ликеш, И.* Основные таблицы математической статистики / И. Ликеш, И. Ляга. — М.: Финансы и статистика, 1985. — 356 с.
34. *Холлендер, М.* Непараметрические методы статистики / М. Холлендер, Д. Вульф. — М.: Финансы и статистика, 1983. — 518 с.
35. *Дубров, А. М.* Многомерные статистические методы / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. — М.: Финансы и статистика, 2000. — 352 с.
36. *Тюрин, Ю. Н.* Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. — М.: Инфра-М, 2003. — 544 с.

-
37. *Фритц, Дж.* Количественный анализ / Дж. Фритц, Г. Шенк. — М. : Мир, 1978. — 556 с.
 38. *Петрук, Е. А.* Расчетные программы для ПЭВМ в вузовском курсе аналитической химии / Е. А. Петрук, В. В. Галкин, В. И. Вершинин // Журнал аналитической химии. — 1994. — Т. 49. — № 8. — С. 393–398.
 39. *Родионова, О. Е.* Хемометрика: достижения и перспективы / О. Е. Родионова, А. Л. Померанцев // Успехи химии. — 2006. — Т. 75. — № 4. — С. 302–321.
 40. *Вершинин, В. И.* Хемометрика в работах российских аналитиков // Журнал аналитической химии. — 2011. — Т. 66. — № 11. — С. 1124–1134.
 41. *Pomerantsev, A. L.* Process analytical technology: a critical view of the chemometricians / A. L. Pomerantsev, O. Ye. Rodionova // Journal of Chemometrics. — 2012. — V. 26. — № 6. — P. 299–310.
 42. *Власов, Ю. Г.* Мультисенсорные системы типа электронный язык — новые возможности создания и применения химических сенсоров / Ю. Г. Власов, А. В. Легин, А. М. Рудницкая // Успехи химии. — 2006. — Т. 75. — № 2. — С. 141–150.
 43. *Вершинин, В. И.* Определение суммарных содержаний парафинов, нафтен и аренов по светопоглощению бензина в ближней ИК-области / В. И. Вершинин, Е. В. Коптева, В. В. Троицкий // Заводск. лабор. Диагностика материалов. — 2005. — Т. 71. — № 11. — С. 10–15.

ОГЛАВЛЕНИЕ

Предисловие	5
-------------------	---

Глава 1

Методология эксперимента	9
1.1. Цели научных исследований	9
1.2. Научные гипотезы	12
1.3. Предпосылки и способы проведения научных исследований	16
1.4. Объект исследования как «черный ящик». Факторы и параметры	21
1.5. Основные этапы активного эксперимента	24

Глава 2

Планирование эксперимента	28
2.1. История и принципы планирования эксперимента	28
2.2. Отбор факторов и требования к ним	32
2.3. Выбор параметров. Обобщенные параметры оптимизации	35
2.4. Традиционная схема активного эксперимента и ее недостатки	41
2.5. Схема и план полного факторного эксперимента вида 2^k	44

Глава 3

Подготовка к проведению эксперимента	50
3.1. Что такое «подготовка к эксперименту»?	50
3.2. Средства измерений	52
3.3. Погрешности измерений. Способы их снижения	56
3.4. Метрологические аспекты многофакторного эксперимента	63

Глава 4

Результат измерения как случайная величина	66
4.1. Случайный характер результатов измерений	66
4.2. Выборки. Репрезентативность выборок	68
4.3. Выборочные параметры	72
4.4. Гистограммы. Функции распределения	76

Глава 5

Нормальное распределение результатов измерений	82
5.1. Проверка характера распределения	82
5.2. Происхождение и свойства нормального распределения	85
5.3. Расчеты, основанные на применении функций Лапласа	90

Глава 6

Математическая обработка малых выборок	96
6.1. Предмет микростатистики	96
6.2. Распределение Стьюдента	98
6.3. Примеры использования t -распределения	100

Глава 7

Проверка статистических гипотез.	
Параметрические критерии	105
7.1. Статистические гипотезы	105
7.2. Общий алгоритм проверки статистических гипотез	107
7.3. Возможные ошибки при проверке статистических гипотез	112
7.4. Параметрические критерии и их применение	115
7.4.1. Отбраковка грубых промахов с помощью Q -теста	115
7.4.2. Сравнение двух дисперсий по критерию Фишера	117
7.4.3. Сравнение нескольких дисперсий. Критерии Бартлетта и Кохрена	118
7.4.4. Сравнение средних значений с помощью критерия Стьюдента	118
7.4.5. Проверка согласия теории и эксперимента. Критерий Пирсона	121

Глава 8

Основные идеи корреляционного и дисперсионного анализа	124
8.1. Понятие о корреляции	124
8.2. Виды корреляции	127
8.3. Проверка значимости коэффициентов линейной корреляции	128
8.4. Однофакторный дисперсионный анализ	131
8.5. Двухфакторный дисперсионный анализ	136

Глава 9

Основные идеи регрессионного анализа	141
9.1. Цели и ограничения	141
9.2. Виды регрессий. Выбор уравнения регрессии	144
9.3. Оценка коэффициентов регрессии	147
9.4. МНК для линейной регрессии в однофакторном эксперименте	148

9.5. МНК в многофакторных экспериментах	151
9.6. Оценка значимости коэффициентов регрессии	154
9.7. Проверка адекватности регрессии	155
9.8. Регрессионный анализ в случае ПФЭ вида 2^k	156
9.9. Модель неадекватна: что делать?	158

Глава 10

Интерпретация математической модели.

Оптимизация исследуемого процесса	161
10.1. Интерпретация уравнения регрессии	161
10.2. Применение адекватной модели	165
10.3. Метод крутого восхождения как способ оптимизации	169

Глава 11

Применение непараметрических методов и критериев 176

11.1. Понятие о непараметрических методах обработки данных	176
11.2. Критерии отбраковки промахов	179
11.3. Критерии различия выборок	181
11.4. Непараметрические аналоги дисперсионного анализа	188
11.5. Проверка взаимосвязи результатов измерения двух величин	194

Заключение 199

Приложения 203

Список литературы 230

*Вячеслав Исаакович ВЕРШИНИН
Николай Викторович ПЕРЦЕВ*

ПЛАНИРОВАНИЕ И МАТЕМАТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ХИМИЧЕСКОГО ЭКСПЕРИМЕНТА

Учебное пособие

Издание четвертое, стереотипное

Зав. редакцией
естественнонаучной литературы *М. В. Рудкевич*

ЛР № 065466 от 21.10.97
Гигиенический сертификат 78.01.10.953.П.1028
от 14.04.2016 г., выдан ЦГСЭН в СПб

Издательство «ЛАНЬ»

lan@lanbook.ru; www.lanbook.com
196105, Санкт-Петербург, пр. Юрия Гагарина, д. 1, лит. А.
Тел./факс: (812) 336-25-09, 412-92-72.
Бесплатный звонок по России: 8-800-700-40-71

ГДЕ КУПИТЬ

ДЛЯ ОРГАНИЗАЦИЙ:

*Для того, чтобы заказать необходимые Вам книги, достаточно обратиться
в любую из торговых компаний Издательского Дома «ЛАНЬ»:*

по России и зарубежью

«ЛАНЬ-ТРЕЙД». 196105, Санкт-Петербург, пр. Ю. Гагарина, д. 1, лит. А.
тел.: (812) 412-85-78, 412-14-45, 412-85-82; тел./факс: (812) 412-54-93
e-mail: trade@lanbook.ru; ICQ: 446-869-967

www.lanbook.com

пункт меню «Где купить»
раздел «Прайс-листы, каталоги»

в Москве и в Московской области

«ЛАНЬ-ПРЕСС». 109387, Москва, ул. Летняя, д. 6
тел.: (499) 178-65-85, 722-72-30; e-mail: lanpress@lanbook.ru

в Краснодаре и в Краснодарском крае

«ЛАНЬ-ЮГ». 350901, Краснодар, ул. Жлобы, д. 1/1
тел.: (861) 274-10-35; e-mail: lankrd98@mail.ru

ДЛЯ РОЗНИЧНЫХ ПОКУПАТЕЛЕЙ:

интернет-магазин

Издательство «Лань»: <http://www.lanbook.com>

магазин электронных книг

Global F5: <http://globalf5.com/>

Подписано в печать 26.02.19.

Бумага офсетная. Гарнитура Школьная. Формат 84×108^{1/32}.
Печать офсетная. Усл. п. л. 12,39. Тираж 100 экз.

Заказ № 202-19.

Отпечатано в полном соответствии с качеством
предоставленного оригинал-макета.
в АО «Т8 Издательские Технологии».
109316, г. Москва, Волгоградский пр., д. 42, к. 5.