

**STATISTIK
IN DER ANALYTISCHEN
CHEMIE**

Von Dr. rer. nat. habil. K. BOERFFEL

Dozent am Institut für analytische Chemie
der TH für Chemie «Carl Schorlemmer», Leuna-Merseburg

LEIPZIG 1966

543
460

К. Доерфель

**СТАТИСТИКА
В АНАЛИТИЧЕСКОЙ
ХИМИИ**

6283

*Перевод с немецкого
И. С. ШАПЛЫГИНА
Под редакцией
В. В. НАЛИМОВА*

ИЗДАТЕЛЬСТВО «МИР» 1969

Книга служит введением в новую для химиков-аналитиков область знания. Она посвящена применению простейших методов математической статистики к задачам аналитической химии. Дается понятие о функциях распределения, сообщаются элементарные сведения о дисперсионном, регрессионном и корреляционном анализе.

Эта работа написана не статистиком, а химиком в форме, легкодоступной для химика-аналитика с элементарной математической подготовкой.

Книга будет служить настольным руководством для широкого круга химиков-аналитиков — работников научно-исследовательских институтов, заводских лабораторий, учебных заведений.

Редакция литературы по химии

ПРЕДИСЛОВИЕ

Предлагаемая вниманию читателей книга Доерфеля заинтересовала специалистов — химиков-аналитиков задолго до того, как был подготовлен ее русский перевод. Она написана значительно более популярно, чем вышедшая ранее на ту же тему работа *, которая, как показал опыт, довольно трудно воспринимается химиками-аналитиками. Можно надеяться, что книга Доерфеля будет доступна значительно более широкому кругу читателей. Большим достоинством ее являются многочисленные, хорошо подобранные примеры, иллюстрирующие применение статистических методов в аналитической химии. Стремясь к краткости и популярности, Доерфель, естественно, мог изложить только простейшие методы математической статистики, ставшие терять классическими. Однако он не рассмотрел некоторые современные, сложные и в то же время очень эффективные приемы, которые уже успешно используются в передовых лабораториях, занимающихся анализом вещества. Нам кажется, что в настоящее время не следует стремиться к изданию всеобъемлющих учебников или монографий. Скорее нужны серии руководств, освещающих различные вопросы на разном уровне. Можно надеяться, что книга Доерфеля будет самой популярной из книг такой серии, предназначенных специально для тех, кто занимается анализом вещества.

Нам представляется, что в области аналитической химии, понимаемой в широком смысле (включая и физические методы анализа), в настоящее время центральной задачей является широкое внедрение методов математической статистики и электронной вычислительной техники.

* В. В. Налимов, Применение математической статистики для анализа вещества, Физматгиз, 1960.

Рациональная организация контроля качества в сложных современных технологических процессах должна базироваться на хорошо поставленной аналитической службе. Здесь нужно учитывать и пробоотбор, и технически, а может быть и экономически обоснованные требования к точности и чувствительности, разумный выбор числа параллельных определений и способы контроля за самопроизвольным смещением результатов анализа во времени. Не слишком ли много сейчас проводят ненужных определений из-за перестраховки, из-за того, что вся организация аналитической службы базируется на каких-то устаревших, очень давно выработанных правилах, не приведенных в соответствие с современными статистическими представлениями? При разработке новых методов анализа уже давно следовало бы применять современные методы планирования эксперимента с представлением результатов поверхностями отклика. Сейчас имеется уже достаточно примеров, свидетельствующих о высокой эффективности этих методов в аналитической химии. Во всех методах анализа, заканчивающихся выдачей регистрограмм, нужно использовать специализированные вычислительные устройства, которые не только выполняют статистический анализ, но также производят все предварительные вспомогательные операции — регистрацию данных (включая сканирование спектров), перевод данных в форму, удобную для ввода в вычислительные устройства, выдачу данных (после математической обработки) в форме, удобной для экспериментатора, корректировку данных в процессе их считывания и т. д. Нужно стремиться к тому, чтобы наши аналитики в ближайшее время получили широкий набор специализированных вычислительных устройств. Вся система организации работ в аналитической химии должна быть перестроена под влиянием идей математической статистики и тех новых возможностей, которые открываются при применении электронной вычислительной техники.

В заключение считаю своим долгом выразить благодарность В. С. Дуженко за помощь при редактировании.

В. Налимов

Моему высокочтимому учителю и руководителю, ныне покойному доктору Г. Грюссу, профессору математики и технической механики Горной академии во Фрейберге, с благодарностью

ПРЕДИСЛОВИЕ АВТОРА

Математика во все возрастающей степени проникает во все сферы науки. Аналитическая химия в этом смысле не является исключением. Это относится как к теоретическим исследованиям вещества, так и к оценке полученных результатов. При этом всегда оказывается, что даже многолетний аналитический опыт нельзя переносить из одной области в другую, даже совсем близкую. Поэтому обращаются к методам математической статистики, поскольку они в состоянии дать желательное объективное суждение о надежности метода анализа или о полноте результатов анализа. При этом статистика полностью предохраняет от недооценки или переоценки результатов и тем самым помогает осмысленно интерпретировать данные анализа.

Естественно, химика-аналитика в первую очередь интересует, как он может с пользой применить методы математической статистики для своих целей. Поэтому в данной книге сделана попытка представить этот предмет с точки зрения аналитика, которого следует ознакомить с математической статистикой и дать ему инициативу к дальнейшему ее приложению. При этом старались поставить аналитические проблемы и интерпретацию математических результатов на первый план и ограничить теоретические положения до минимума. Особенно большое значение было обращено на возможно более точно вычисленные примеры применения. Они взяты из разных областей аналитической химии; перенос на аналогичные проблемы в смежных областях знаний может быть совершен без затруднений. Использованный цифровой материал большей частью основан на собственных измерениях. В некоторых случаях мне оказали помощь коллеги, работающие в промышленности. Всем им я приношу благодарность.

При обсуждении рукописи мне была оказана разносторонняя помощь. Особенно я благодарю доктора Рихтера (Институт математической статистики Технического университета в Дрездене) за критический просмотр отдельных частей рукописи и Рейнхольда — сотрудника Института математики Высшей технической школы химии «Карл Шорлеммер» в Лейна-Мерзебурге за ценные справки и советы. Критические указания были получены от профессора Аккермана из Фрейберга и профессора Гейера из Мерзебурга. Проверку примеров взяла на себя инженер-химик Штевенц. В заключение я благодарю всех коллег, которые помогали мне в дискуссионном порядке. Наконец, не могу не поблагодарить издательство за его полный понимания подход и за многие высказанные мне пожелания.

Я с благодарностью принимаю советы по улучшению книги и надеюсь, что они будут приняты во внимание в более позднем издании.

Мерзебург, 1966 г.

К. Доерфель

ИСПОЛЬЗОВАННЫЕ ОБОЗНАЧЕНИЯ

При выборе обозначений приняли во внимание рекомендации номенклатурной комиссии Международного союза теоретической и прикладной химии (ЮПАК). Поэтому появились отдельные отклонения от более ранних публикаций*.

- a, b — константы прямых регрессий;
- k — число классов при эмпирических распределениях частот;
- m — число проб различного содержания;
- μ — среднее значение генеральной совокупности;
- n — число измерений;
- n_j — число параллельных определений;
- P — вероятность;
- r — коэффициент корреляции;
- R — размах варьирования;
- s — оценка средней квадратичной ошибки;
- s^2 — дисперсия;
- σ — средняя квадратичная ошибка в генеральной совокупности;
- x — измерение;
- \bar{x} — среднее значение;
- \tilde{x} — медиана;
- $\Delta\bar{x}$ — доверительный интервал.

Остальные обозначения объяснены в соответствующих местах.

* Doerffel K., Z. anal. Chem., 185, 1 (1962).

1. Виды ошибок в аналитической химии

При анализе пробы аналитик обычно проводит несколько параллельных определений. При этом отдельные результаты этих определений должны лежать как можно ближе друг к другу и соответствовать истинному содержанию пробы. Следовательно, существуют два фактора, по которым аналитик судит о своих результатах:

- 1) воспроизводимость полученных результатов;
- 2) соответствие их содержанию пробы.

Воспроизводимость зависит от *случайной ошибки метода* анализа. Чем больше случайная ошибка, тем больше разброс значений при повторении анализа, тем меньше *точность метода* анализа.

Отклонения от истинного содержания пробы вызываются *систематической ошибкой*. Метод анализа лишь тогда может дать правильные значения, когда он свободен от систематических ошибок. Случайные ошибки делают неточным результат анализа, систематические ошибки делают неверным сам анализ. Итак, *воспроизводимость* или *точность* и *правильность* метода анализа следует рассматривать порознь.

[1.1] В практикуме объемного анализа два студента титровали одну и ту же пробу. Были получены следующие значения (*мг Са*):

Первый студент			Второй студент		
121,5	122,0	121,0	125,0	126,0	126,2

У обоих студентов одинаковый случайный разброс. По сравнению с заданным содержанием пробы 125,0 *мг* результаты первого студента указывают на систематический сдвиг в сторону уменьшения и поэтому они не могут считаться «правильными».

Нередко в литературе можно встретить понятие «точность» иногда в смысле точности способа, иногда в смысле его правильности. Вследствие такой двойственности это понятие в дальнейшем не будет использовано. По этим

же причинам его не следовало бы использовать и в литературе.

Появляющиеся в аналитической химии случайные и систематические ошибки можно объяснить разными причинами. При этом обычно принимают во внимание следующие соображения:

1) большинство исследуемых веществ надо рассматривать как неоднородные. Поэтому отдельные небольшие части пробы имеют неодинаковый состав. По этой же причине результаты анализа также будут иметь случайные колебания. Вследствие неправильного одностороннего отбора проб может быть отдано предпочтение отдельной составной части, что повлечет за собой систематическое изменение состава пробы;

2) все необходимые для анализа величины, такие, например, как масса осадка или экстинкция окрашенного раствора, можно установить только с ограниченной точностью. Это определяется использованным методом измерений, величиной, подлежащей измерению, а часто и субъективными влияниями. Если исключить влияние неправильно отрегулированных измерительных приборов, то ошибки измерений появляются в основном в форме случайных отклонений. Должным подбором условий измерений эти ошибки следует сводить к минимуму;

3) особенно при классических анализах исследуемые пробы подвергают химической обработке и продукты реакции характеризуют по виду и массе. Все эти реакции следует считать равновесными процессами, при которых равновесие стремятся сдвинуть как можно дальше в направлении получения продуктов реакции. Несмотря на это, в ходе реакций появляются как случайные ошибки (например, изменение растворимости вследствие различных концентраций средней соли), так и систематические ошибки (например, явление сосаждения). Задачей аналитика является выбор реакций, наиболее пригодных для соответствующей цели.

Если отбросить ошибку, вводимую при отборе пробы, как не относящуюся непосредственно к методу анализа, то общая ошибка состоит из ошибок измерения и ошибок, вызываемых химической реакцией. В общем ошибки измерения должны быть меньше, чем ошибки метода. В то

время как ошибки измерения можно оценивать обсуждением ошибки, постоянно совершаемой при физических исследованиях, для ошибки метода этого сделать нельзя или можно сделать только в виде исключения. Описание этой ошибки и общей ошибки можно дать только с привлечением методов математической статистики.

Эти методы исходят из идеализированного представления о бесконечно большом числе измерений. Это множество результатов измерений называют *генеральной совокупностью*. Из нее выводят закономерность для процессов, кажущихся наблюдателю чисто случайными. Однако при практических измерениях всегда имеется очень ограниченное число значений. Они образуют выбор конечного объема из генеральной совокупности, называемый *выборкой*. Выборку следует брать так, чтобы возможно лучше охарактеризовать (представить) генеральную совокупность. Эта цель достигается тем лучше, чем больше объем выборки и чем удачнее сделан случайный отбор отдельных измерений.

[1.2]. Условия отбора выборки особенно удобно сравнить с принципами отбора проб в аналитике. Проба только тогда дает хороший анализ, когда выбор сделан случайно (т. е. без предпочтения определенным партиям) и когда эта проба достаточно велика.

Также принципиально важно, чтобы правильно взятая выборка имела лишь случайные отличия от генеральной совокупности. Эти случайности и вероятность их появления можно описать при помощи математической статистики. По измерениям, сделанным на выборках, можно дать сведения о поведении генеральной совокупности. Поэтому из конечного числа значений можно сделать вывод о случайной ошибке примененного метода и предсказать подобные случаи в будущем.

Если внутри серии анализов существенна только случайная ошибка, то результаты беспорядочно рассеиваются внутри небольшой области значений, несмотря на совершенно постоянные условия опыта. Наиболее правильное содержание пробы лежит внутри этой области колебаний. Случайная ошибка может иметь размерность измеряемых величин (например, мг, мг/л или может быть выражена в процентах). В этом случае говорят об *абсолютной ошибке* измерения. Но случайная ошибка может быть также от-

песена к среднему значению измеряемой величины. В этом случае говорят об *относительной ошибке* измерения. При представлении результатов измерения следует иметь в виду, что речь может идти о двух таких возможностях.

[1.3]. При анализах эталонных образцов в паспорте рядом с процентным содержанием отдельных компонентов отмечают также имеющуюся случайную ошибку в виде абсолютной ошибки. Для содержания хрома в феррохроме находят, например, $(63,5 \pm 0,1)\%$ Сг. Если отсюда хотят узнать относительную ошибку, то получают

$$\frac{0,1}{63,5} \cdot 100 = 0,16\% \text{ (отн.)}$$

Особенно просто можно оперировать абсолютной ошибкой. Она дает непосредственную характеристику достоверности полученного значения. Напротив, относительная ошибка часто является наглядной для характеристики метода анализа из-за ее связи с измеряемым значением.

Систематические ошибки влияют на все измерения всегда в одинаковой степени. При этом истинное значение лежит *вне* области колебаний. Если все значения сдвинуты на одинаковую величину, то говорят о *постоянной ошибке* (например, неизвестное значение «холостого» опыта). Отклонения, которые изменяются с величиной измеряемого значения, называют *изменяющейся ошибкой*. При пропорциональности между значением измерения и величиной ошибки говорят о *линейно изменяющейся ошибке* (такую ошибку создает, например, неправильный титр при титровании жидкости). Оба вида ошибок могут, естественно, появляться одновременно. Систематические отклонения в основном даются в форме абсолютной ошибки.

Вследствие случайной ошибки значения параллельных определений при анализе беспорядочно рассеиваются вокруг истинного значения пробы. Систематическая ошибка дополнительно сдвигает значения измерений в том или ином направлении. Таким образом, метод дает «неверное значение». Случайные и систематические ошибки совершенно различно отражаются на результатах анализа. Несмотря на это, между обоими видами ошибок существуют определенные зависимости.

Ряд методов анализа известен более или менее большими систематическими положительными или отрицательными ошибками. Примером этого служит гравиметрическое определение кремниевой кислоты, при котором следует постоянно считаться с заниженными значениями. Однако это заниженное значение можно обнаружить лишь тогда, когда, например, потери, появившиеся вследствие растворимости осадка больше, чем колебания из-за случайной ошибки. Общие систематические ошибки можно найти только тогда, когда сдвиг значений измерений больше, чем случайная ошибка выбранного метода измерений.

Если анализируемую пробу исследуют в различных лабораториях, то в одних лабораториях появляются систематические положительные ошибки различной величины, а в других — систематические отрицательные ошибки различной величины. Эти систематические отклонения явно больше, чем случайная ошибка метода. Так как эти отклонения имеют разную величину и разный знак, они появляются в виде беспорядочного разброса результатов или в виде увеличенной случайной ошибки. Одновременное появление систематических ошибок разной величины и знака также вызывает увеличение случайной ошибки.

Цель всех аналитических исследований — нахождение результата, наиболее близкого к истинному содержанию пробы при помощи найденных значений анализа. Чтобы достигнуть этой цели, при выборе метода анализа мирятся с небольшой систематической ошибкой, в случае если метод дает малую случайную ошибку. Несмотря на небольшой систематический сдвиг результатов, при этом можно получить значение более близкое к истинному содержанию пробы, чем применением метода, который хотя и «правильный», но дает очень большую случайную ошибку.

Все появляющиеся при анализах ошибки — как случайные, так и систематические — можно приписать использованному методу анализа. Кроме того, на них, например, влияют условия работы в лаборатории или квалификация исполнителей. Ошибки также могут быть подвержены колебаниям времени. При всех исследованиях аналитик должен стремиться получить возможно меньшую

случайную ошибку, поддерживая постоянный контроль и постоянно исключая систематические ошибки. Последнее нужно проводить устранением причин ошибок, но не дополнительной корректурой результатов. Рекомендуемыми иногда эмпирическими коррекционными факторами являются средние значения с более или менее большим интервалом колебания, поэтому они слишком мало говорят об особом единичном случае. Кроме того, случайная ошибка всегда увеличивается вследствие «поправки», потому что в этом случае комбинируют два значения с ошибками.

Если необходимо сравнить между собой результаты большего ряда измерений — иногда это обозначается как *относительное измерение*, — то достаточно знать появляющуюся случайную ошибку. Менее важно — содержит ли метод анализа систематическую ошибку. Следует исходить из того, что эта возможная систематическая ошибка не изменяется в период исследований. Поэтому при *абсолютном определении* (например, содержания продажного продукта) следует знать как случайную ошибку, так и то, что систематическая ошибка отсутствует. Правильность метода анализа доказана только тогда, когда два разных метода дают значения, между которыми нет уловимого различия.

Во многих методах анализа оперируют эмпирическими факторами. Они лишь после калибровки приводят к «правильным» значениям (например, определение марганца по Фольгарду и Вольфу). Для калибровки используют стандартную пробу с составом, подобным исследуемой пробе. Следует установить частоту калибровки, осуществляя периодический контроль. Условные методы анализа (определение золы в горячем или определение температуры воспламенения масел) следует считать широкими стандартизованными методами исследования. Они часто облегчают взаимопонимание, например, между торговыми партнерами и обычно не дают «правильных» значений анализа в требуемом здесь смысле.

Обсуждение ошибок играет решающую роль при планах и оценках химико-аналитических исследований. Поэтому аналитик нуждается в широкой информации о возможных видах ошибок в области его работы. При рассмотрении типов ошибок аналитик при помощи математико-

статистических методов получает информацию о имеющемся в его распоряжении числовом материале. Методы математической статистики позволяют использовать «профессиональный материал» для получения ответов на определенно поставленные вопросы, например может быть поставлен вопрос о сравнении средних значений, об анализе совместных опытов или об определении систематической ошибки. Задача аналитика — выбрать из различных математико-статистических методов наиболее подходящие для решения поставленной им специальной задачи.

ЛИТЕРАТУРА

1. Ebert H., Z. Instrumentenkunde, 65, 14 (1959).
2. Gysel H., Mikrochim. Acta, 3, 266 (1953).
3. Kaiser H., Specker H., Z. anal. Chem., 149, 46 (1956).
4. Moran R. F., Ind. Engng. Chem., anal. Ed., 15, 361 (1943).

2. Эмпирические распределения частот

Многие исследования начинают со сбора обширного цифрового материала. В аналитической химии такие данные получаются, например, если проводят совместные аналитические исследования в нескольких лабораториях или если собирают данные о качестве продукта за длительный период времени. Этот цифровой материал необходимо систематизировать для дальнейших исследований, и в этом случае особенно целесообразными и наглядными оказываются графические методы. В качестве первого шага стремятся представить цифровой материал при помощи наименьшего числа специальным образом выбранных показателей. Эти известные величины позволяют простыми способами сравнивать полученные данные с другими, уже имеющимися результатами измерений.

2.1. Обсуждение эмпирических распределений частот

Понимание обширного цифрового материала облегчается систематизацией. Весьма ценным часто является представление измеренного значения в зависимости от частоты его появления при измерениях. При этом можно представлять результаты измерений в виде отдельных точек на линейно поделенной (одномерной) цифровой прямой и затем судить о плотности точек. Вследствие такой формы представления подобное распределение называют *одномерным распределением*.

Более наглядной является ступенчатая диаграмма, на которой графически представляют величину измеренного значения x , распределенную по классам в зависимости от частоты ее появления h . Наивысший столбец такой ступенчатой диаграммы показывает измеренное значение, появляющееся чаще всего, и соответствует вертикальным прямым в области с наибольшей плотностью точек в диа-

граммах первого типа. Для построения ступенчатой диаграммы группируют отдельные значения в k классов с шириной класса d . Число классов k должно быть примерно равно корню квадратному из общего числа измеренных значений, которое, однако, не должно быть меньше пяти и больше двадцати. Если выбрать слишком мало классов, то можно потерять характерную деталь распределения частоты; слишком подробным делением на классы можно затушевать общую картину небольшими случайными отклонениями. При выборе границ класса следует обращать внимание на то, чтобы верхняя граница класса была меньше, чем нижняя граница прилегающего соседнего класса (ср. пример [2.1]). Когда результаты анализов одной пробы упорядочивают описанным способом при безукоризненных условиях опыта чаще всего получают симметричное распределение с одним максимумом частоты. Асимметричные распределения частот со сдвинутым влево или вправо максимумом указывают на недостатки условий опыта или на неправильное деление оси абсцисс.

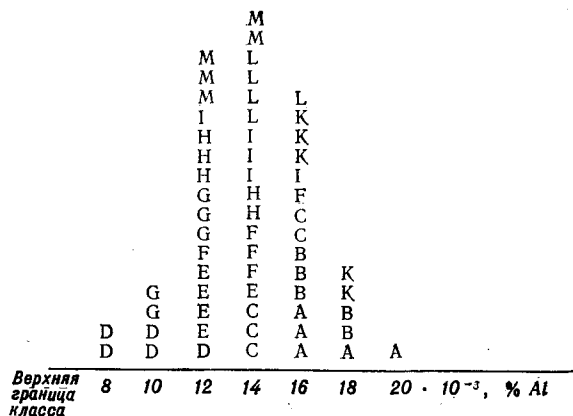
[2.1]. Совместное определение алюминия в стали проводили двенадцать лабораториями. Каждая лаборатория дала пять полученных в разные дни значений анализов, которые систематизированы следующим образом (% Al):

Лаборатория А	0,016	0,015	0,017	0,016	0,019
Лаборатория В	0,017	0,016	0,016	0,016	0,018
Лаборатория С	0,015	0,014	0,014	0,014	0,015
Лаборатория D	0,011	0,007	0,008	0,010	0,009
Лаборатория E	0,011	0,011	0,013	0,012	0,012
Лаборатория F	0,012	0,014	0,013	0,013	0,015
Лаборатория G	0,011	0,009	0,012	0,010	0,012
Лаборатория H	0,011	0,011	0,012	0,014	0,013
Лаборатория I	0,012	0,014	0,015	0,013	0,014
Лаборатория K	0,015	0,018	0,016	0,017	0,016
Лаборатория L	0,015	0,014	0,013	0,014	0,014
Лаборатория M	0,012	0,014	0,012	0,013	0,012

В общем имеется $n = 60$ значений. Низшее значение дано лабораторией D с $x_{D2} = 0,007\%$ Al, высшее лабораторией A с $x_{A5} = 0,019\%$ Al. Сгруппировав данные в $k = 7$ классам с шириной класса $d = 0,002\%$ Al, получаем $k \sim \sqrt{n}$. Первый класс включает значение 0,007 и 0,008% Al, второй — 0,009

и 0,010% Al и т. д. При этом делении на классы получают распределение частот, приведенное на рис. 2.1*.

Несмотря на различное происхождение значений, распределение частот получается с выраженным максимумом. Поразительно



Р и с. 2.1. Распределение частот при совместном определении алюминия в стали двенадцатью лабораториями.

то, что значения, повторяющиеся в отдельных лабораториях, лежат рядом, в то время как при сравнении данных отдельных лабораторий можно наблюдать довольно большие расхождения.

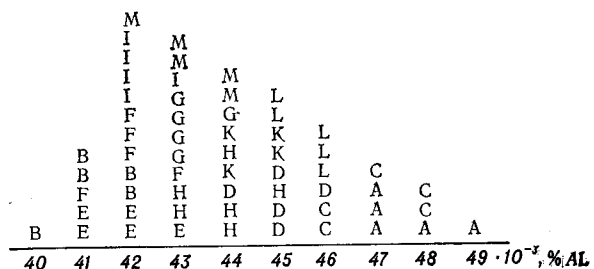
Из формы распределения частот можно получить качественную картину появляющейся случайной ошибки. При большой случайной ошибке получаются широкие распределения, при малой случайной ошибке кривая распределения становится узкой и острой (предполагаются сравнимые классы). Однако никаких сведений о возможной систематической ошибке в этом случае не получают, так как она не изменяет вид распределения.

Зато неоднородные систематические ошибки часто определяются весьма своеобразным путем. Если, например, при совместном опыте нескольких лабораторий в результатах части лабораторий появляется одинаковая

* Для практических работ часто целесообразно подсчитать эмпирические распределения частот, повернутые на 90°. Значение x тогда, следовательно, изображается сверху вниз, частот h — слева направо.

по величине и знаку систематическая ошибка, то распределения частот получаются с двумя (или несколькими) максимумами. Второй максимум может служить «плечом» главного максимума; он как бы образует ложное распределение, когда систематический сдвиг не очень велик. Распределение таких сложных распределений во многих случаях облегчается частотной сеткой*.

[2.2]. В определении алюминия в магниевом сплаве участвовало двенадцать лабораторий. При исследовании были применены два разных метода анализа, а именно осаждение алюминия в виде 8-оксихинолята (лаборатории В, F, G, I, M) и осаждение фосфата



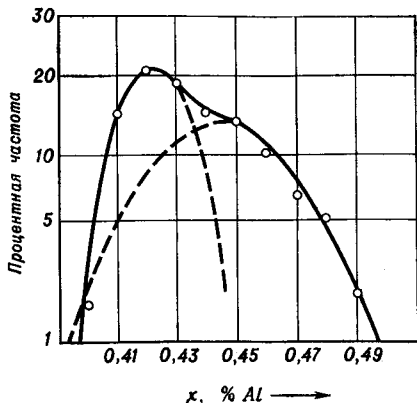
Р и с. 2.2. Асимметричное распределение частот при совместном определении алюминия двумя разными методами.

алюминия (лаборатории А, С, D, E, H, K, L). Распределение частот оказалось с левосторонним максимумом (рис. 2.2). Для проверки в частотной сетке подсчитывали относительные (процентные) частоты в каждом классе. Получена следующая схема:

Класс, %	Частота		Класс, %	Частота	
	абсолютная	процентная		абсолютная	процентная
0,40	1	1,7	0,46	6	10,0
0,41	5	8,3	0,47	4	6,7
0,42	12	20,0	0,48	3	5,0
0,43	11	18,3	0,49	1	1,7
0,44	9	15,0			
0,45	8	13,3		60	100,0

* Для этого используют бумагу компании «Schäfers Feinpapier», Плауэн (Фогтланд), № 606. О конструкции частотной сетки см. стр. 48.

При внесении этих значений в частотную сетку (рис. 2.3) становится ясным, что распределение на рис. 2.2 состоит из двух частей. Они различаются своими максимумами и шириной, и без труда можно согласовать оба использованных метода. Вследствие этого



Р и с. 2.3. Разбиение распределения на две части при помощи частотной сетки.

очевидно, что оба способа не дают одинакового значения. Фосфатный метод в противоположность оксихинолиновому дает слишком высокие значения; кроме того, при фосфатном методе, возможно, наблюдаются наибольшие случайные отклонения.

Эти разработанные Давесом и Беккелем [2] методы разложения на составные части помогают практику легко обнаружить неоднородность цифрового материала в тех случаях, когда следует предполагать такую неоднородность. Однако их теоретические положения являются спорными, так как многие функции, описывающие распределения частот, можно представить как сумму частных частотных функций, не приписывая это смешанному распределению. Поэтому при применении частотной сетки постоянно следует проверять, насколько найденные результаты согласуются с экспериментами. В большинстве случаев числовые значения из многопикового распределения негодны для дальнейшего использования. Следует выяснить причины появившейся систематической ошибки и устранить их. Сообразно с этим опыт повторяют при безукоризненных условиях.

При совместных опытах нескольких лабораторий могут наблюдаться асимметричные распределения с лево- или правосторонним максимумом частот, если результаты отдельных лабораторий сопровождаются систематически ошибками одинакового знака, но различной величины.

[2.3]. Известно, что при определении кремния в стали существует опасность получения заниженного результата. На рис. 2.4 приведено два типа распределения частот для результатов совместных опытов по определению кремния. Необходимо было исследовать две пробы стали с разным содержанием кремния. Получилось два

G	H					
G	G	H			F	G
G	F	H			F	G
F	F	F			F	G
E	F	C	D		F	E
E	E	B	D	D	D	E
E	E	B	D	C	C	E
A	C	A	D	B	C	D
A	C	A	D	B	C	B
					A	A
					A	B
					A	B

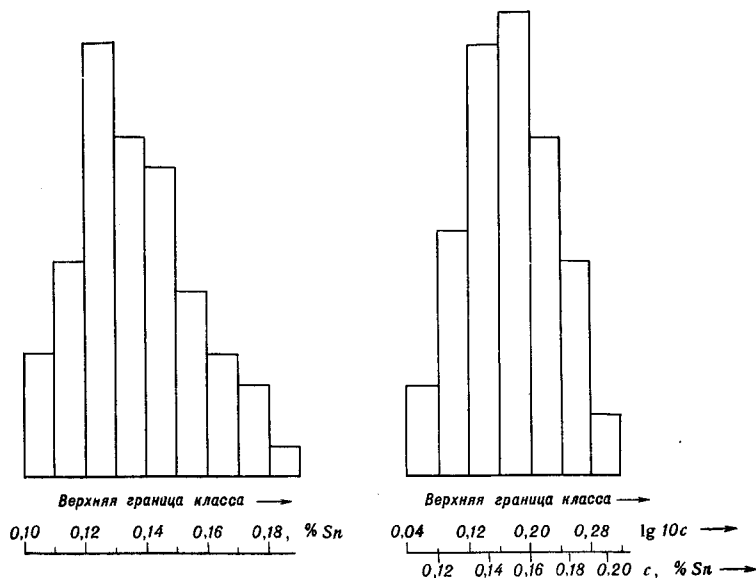
Р и с. 2.4. Асимметричное распределение частот, полученное при совместном определении кремния.

явно разных левосторонних распределения. Сдвиг распределения особенно сказывается в пробе 1 с более низким содержанием кремния (ср. пример [2.8]). Поэтому можно предполагать, что систематическое отклонение обнаруживается в форме постоянной ошибки (ср. стр. 14).

Асимметричные распределения могут также возникать, в случае если линейное деление оси свойства (абсцисса) выполнено неправильно из методологических соображений. Подобные кажущиеся асимметричными распределения могут переходить в симметричную форму, когда ось свойства делат логарифмически.

[2.4]. Для методических исследований было проведено 60 спектральных определений одинаковых проб оловянной руды. Распределение частот измеренных значений показывает левостороннюю асимметрию (рис. 2.5). Однако она исчезает, если логарифмировать найденное содержание и представить распределение частот для логарифмов содержания.

Подобные логарифмические распределения встречаются чаще, чем можно предположить. Однако часто их не распознают, так как при незначительной случайной ошиб-



Р и с. 2.5. Распределение частот в линейном и логарифмическом масштабах при спектрохимических определениях олова.

ке метода их отличие от распределения с линейной шкалой незначительно. С появлением логарифмических распределений следует считать:

- 1) при анализе с очень малым содержанием (анализ следов);
- 2) при исследованиях внутри очень широкой области концентраций (несколько десятков процентов);
- 3) при очень большой случайной ошибке (например, в полуколичественном спектральном анализе);
- 4) при измерениях времени.

При существовании логарифмических распределений следует вводить логарифмы измеренных значений.

Обсуждение эмпирических распределений частот описанным выше образом может всегда давать только первые ориентировочные указания. Даже при наличии достаточного числа измерений ($n > 40$) с достаточной надежностью можно предсказывать только явно выраженные простые явления. При анализе результатов коллективных исследований необходимо особенно много измерений (ср. пример [2.2]). Следует стремиться к тому, чтобы при совместных опытах каждая лаборатория давала примерно 30 значений. Безусловная осторожность необходима при оценке небольших эффектов, так же как и при малом числе измерений. Здесь следовало бы (чтобы не ошибиться) обратиться к описываемому в дальнейшем математическому способу проверки (ср. разд. 7.6).

2.2. Статистические характеристики

Для непосредственной дальнейшей оценки эмпирических распределений цифровой материал, полученный из опыта, следует характеризовать числовыми показателями. Для этого служат средние значения и мера разброса. Распределение частот задается только при помощи двух этих величин. Поэтому наиболее часто приводимые данные среднего значения оказываются недостаточными. Они должны постоянно дополняться данными, относящимися к разбросу.

2.2.1. СРЕДНИЕ ЗНАЧЕНИЯ

При оценке результатов анализа часто пользуются только арифметическим или геометрическим средним, или также медианой. Выбирают то или иное среднее значение, соответствующее качеству выполненных измерений и поставленной задаче.

Важно всегда пользоваться одним и тем же средним значением для результатов, которые должны сравниваться между собой.

Арифметическое и геометрическое среднее. Допустим, что имеются отдельные значения x_1, x_2, \dots, x_n из выборки объема n (с общим x_i). Тогда арифметическое среднее

значение следует вычислять по формуле

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n} \quad (2.1)$$

Поскольку выполнено достаточное число измерений, арифметическое среднее \bar{x} в большинстве случаев представляет собой хорошее приближение к среднему значению μ в генеральной совокупности.

Многие методы анализа дают результаты в виде логарифма искомого содержания (например, спектральный анализ, ср. пример [2.4]).

Пересчет на искомое содержание чаще всего производят особенно просто путем логарифмического деления концентрационной оси калибровочного графика. Для определения средних значений в подобных случаях следует пользоваться логарифмами (а не числами). При $x = \lg X$ получаем

$$\bar{x}_{\lg} = \frac{\lg X_1 + \lg X_2 + \dots + \lg X_n}{n} = \lg \bar{X} \quad (2.2)$$

$$\bar{X} = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n} \quad (2.3)$$

Также можно представлять результаты вместо арифметического геометрическим средним. Его числовое выражение лежит всегда ниже арифметического среднего. Однако для практических целей эта разница несущественна, когда случайная ошибка метода анализа достаточно мала.

[2.5]. При количественном спектрохимическом определении олова в бедных оловянных рудах в одной пробе найдено следующее содержание (Sn, %):

$$X_i = \begin{matrix} 0,19_2 \\ 0,24_3 \\ 0,15_7 \\ 0,25_5 \\ 0,31_9 \end{matrix}$$

Значения пересчитывают по $x = \lg 10X$ и получают

$$x_i = \begin{matrix} 0,283 \\ 0,386 \\ 0,196 \\ 0,407 \\ 0,504 \end{matrix}$$

Подстановка этих логарифмов в уравнение (2.2) приводит к $\bar{x}_{\lg} = 0,355 = \lg \bar{X}$. Потенцированием в качестве геометрического среднего находят $\bar{X} = 0,227\%$ Sn. Арифметическое среднее оказывается равным $0,23_3\%$ Sn.

Арифметическое (и геометрическое) среднее нельзя вычислять из распределения с несколькими максимумами. При составлении среднего можно комбинировать только значения из сравнимых измерений. В общем в основе должно лежать среднее значение по крайней мере трех отдельных измерений. При вычислении среднего ни в коем случае нельзя отбрасывать самое низкое или самое высокое выпадающее измерение, иначе среднее окажется грубой ошибкой (ср. 7.5). Это строгое правило можно несколько смягчить при анализах, проводимых в одной лаборатории, если отброшенное значение замещается по меньшей мере тремя последующими измерениями. Вычисление среднего не допустимо, если ось свойства графика распределения частот не делится линейно или если значения измерений при распределении во времени имеют возрастающую или понижающуюся тенденцию («тренд»). При более или менее сильно разбросанных значениях арифметическое среднее плохо представляет ряд измерений.

Медиана (срединное значение). Для определения медианы результаты измерений упорядочиваются по величине. Так, при выборке из n измерений получают $x_1 < x_2 < \dots < x_n$. Медиану затем определяют вычислением.

Если n нечетное число, то \tilde{x} равен срединному члену ряда. При четном числе наблюдений медиана равна среднему арифметическому обоих срединных значений упорядоченного по величинам ряда наблюдений. Например,

$$\begin{aligned} \text{для } n=3 \quad \tilde{x} &= x_2 \\ \text{для } n=4 \quad \tilde{x} &= \frac{x_2 + x_3}{2} \end{aligned} \quad (2.4)$$

Медиана в противоположность арифметическому среднему оказывается нечувствительной к крайним (резко выделяющимся) значениям измерений. Поэтому ее можно использовать для характеристики небольшой серии измерений ($n < 10$), при которых появляются такие резко выделяющиеся значения. В аналитической химии это

явление обусловлено особенностями метода. Резко выделяющиеся измерения появляются при количественном спектральном эмиссионном анализе порошков или при количественной инфракрасной спектрофотометрии, проводимой методом таблеток из КВг. Вопреки тенденционным крайним значениям ряда наблюдений медиану \tilde{x} считают в этих случаях надежной оценкой средних значений генеральной совокупности. Правда, необходимо подсчитывать медиану для всех проб, исследованных данным методом, и не проводить сопоставление среднего значения с медианой. При сериях измерений, состоящих из большого числа измерений ($n > 10$), медиана \tilde{x} дает все же только плохую оценку среднего значения, так как здесь учитывается только одно или два измерения из целой серии. При наличии логарифмического распределения медиану вычисляют для логарифмов измеренных значений.

[2.6]. При получении продуктов из полиакрилонитрила остаточное содержание нитрила определяют методом инфракрасной спектрофотометрии.

Пробы готовят прессованием с КВг; для каждой пробы выполняют четыре параллельных определения. Для одной пробы получены следующие значения экстинкции (упорядоченные по возрастанию).

$$E = \begin{matrix} 0,625 \\ 0,665 \\ 0,673 \\ 0,680 \end{matrix}$$

По уравнению (2.4) в качестве медианы получают

$$\tilde{E} = \frac{0,665 + 0,673}{2} = 0,669$$

В качестве арифметического среднего находят в этом случае неверное слишком низко лежащее значение

$$\bar{E} = \sum E_i/n = 0,661$$

При использовании медианы остается лежащее в стороне, не влияющее на результат значение $E_1 = 0,625$.

Между средним значением и медианой, полученными из большого числа определений, наблюдается обычно только небольшая разница; это следует из симметричного

распределения частот. Большая разница между $|\bar{x} - \tilde{x}|$ указывает на кажущуюся или действительную асимметрию распределения или на лежащее в стороне крайнее значение.

2.2.2. ГРАНИЦЫ РАЗБРОСА

Отдельные измеренные или наблюдаемые значения распределения частот более или менее кучно разбросаны вокруг среднего значения. Описание этого разброса является второй характеристикой цифровых данных. В качестве границ разброса в аналитической химии почти без исключения используют квадратичную ошибку, иначе называемую стандартным отклонением, и размах варьирования. Обе меры разброса следует выбирать в соответствии с преследуемой целью.

Средняя квадратичная ошибка (стандартное отклонение). Среднюю квадратичную ошибку выборки определяют посредством

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (2.5)$$

где x_i — единичное значение; \bar{x} — среднее из всех x_i ; n — число всех измерений.

Она является мерой разброса, используемой в аналитической химии почти всегда и характеризующей случайную ошибку метода анализа (но не единичные значения, ср. гл. 6). Средняя квадратичная ошибка s является лучшим приближением для соответствующей величины σ в генеральной совокупности. Ее квадрат s^2 (соответственно σ^2) называют *дисперсией*.

Сумма квадратов в числителе равенства (2.5) подсчитывается чаще всего не по этой определяющей формуле. Путем преобразования* получают

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad (2.6a)$$

$$= \sum x_i^2 - n\bar{x}^2 \quad (2.6b)$$

* Указанное равенство (2.6) находят следующим путем:
 $\sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2$
 По равенству (2.1) $\bar{x} = \sum x_i/n$, отсюда подстановкой получают равенства (2.6a) и (2.6b).

Равенство (2.6а) дает только малую ошибку после округления. Поэтому оно особенно подходит для цифровых вычислений. Равенство (2.6б) используют при работе с малыми вычислительными машинами. Целесообразно всегда применять одинаковые формулы для подсчета суммы квадратов. В дальнейшем будет всегда использоваться равенство (2.6а). При вычислениях результаты измерений преобразуют так, чтобы отбросить лишние цифры и опустить запятые. Последние потом снова восстанавливаются. При вычислениях пользуются таблицей квадратов чисел (см. табл. 12.7).

Величина $n - 1$, стоящая в знаменателе равенства (2.5), называется *числом степеней свободы*. Эту величину можно интерпретировать как число контрольных измерений, которым соответствует результат, уже полученный из измерений. Число степеней свободы в дальнейшем обозначается f .

При наличии логарифмического распределения (ср. пример [2.4]) подсчитывается средняя квадратичная ошибка s_{lg} для логарифмов измеренных величин (ср. пример [5.3]).

[2.7]. Из десяти определений марганца требуется подсчитать среднюю квадратичную ошибку в одинаковых пробах. Получены следующие значения (Мп, %):

0,69	0,70	0,67	0,66	0,67
0,68	0,67	0,69	0,68	0,68

Результаты измерений преобразуют по формуле $X = 100x - 68$; вследствие этого отбрасывают запятую при подсчете. Вычитанием значения 68, примерно соответствующего арифметическому среднему, сводят результаты измерения к малым числам; кроме того, последний член в уравнении (2.6а) становится неизвестным. Получаются следующие преобразованные значения:

+1	+2	-1	-2	-1	$\Sigma X_i = -1$
0	-1	+1	0	0	$n = 10$

Из уравнения (2.6а) находят сумму квадратов

$$\Sigma (X_i - \bar{X})^2 = 1^2 + 2^2 + 1^2 + \dots - \frac{(-1)^2}{10} = 13$$

$$S = \sqrt{13/9} = 1,2$$

Затем делают обратное преобразование, причем вычитаемый член не принимают во внимание, и получают в качестве средней квадратичной ошибки

$$s = 0,01\% \text{ Мп}$$

при $f = 9$ степеней свободы

Скорость вычисления по уравнению (2.6а) можно значительно увеличить применением таблицы обратных величин (табл. 12.8). Вычисляют Σx_i , возводят в квадрат и умножают на $1/n$. Результат с отрицательным знаком задают вычислительной машине, а затем прибавляют единичные значения x_i^2 , взятые из табл. 12.7 (ср. гл. 10). **Размах варьирования.** Разница между наибольшим и наименьшим значением ряда измерений определяется как **размах варьирования R** . Следовательно,

$$R = x_{\text{макс}} - x_{\text{мин}} \quad (2.7)$$

Размах варьирования особенно подходит для характеристики рассеяния при выборке малого объема ($n < 10$). При наличии большого числа измерений ($n > 10$) размах варьирования является плохой оценкой рассеяния в генеральной совокупности, так как он (в противоположность средней квадратичной ошибке) учитывает только два значения целой серии измерений. На величину размаха варьирования влияет объем выборки; при остающейся одинаковой случайной ошибке R возрастает с увеличивающимся числом измерений. При определенных предположениях по размаху варьирования выборки можно получить представление о средней квадратичной ошибке генеральной совокупности (ср. разд. 5.1).

2.2.3. АСИММЕТРИЯ И ЭКСЦЕСС

В разд. 2.1 указаны некоторые причины, которые могут вести к якобы асимметричному распределению. Все соответствующие распределения частот можно при помощи подходящих преобразований (например, переход к логарифмическому масштабу) перевести в симметричные распределения. Поэтому речь идет не об истинной асимметрии. Истинная асимметрия имеет место, если при достаточно большом числе измерений и после ликвидации всех технических или возможных математических причин асимметрия сохраняется. Такое распределение, кроме среднего значения и границ разброса, характеризуют дополнительно асимметрией ρ . Она определяется формулой

$$\rho = \frac{\Sigma n_j (x_i - \bar{x})^3}{ns^3} \quad (2.8)$$

где n_j — число измеренных значений в j -м классе; n — число всех измеренных значений.

Асимметрия является безразмерной величиной для симметричного распределения $\rho = 0$. Левосторонняя асимметрия оказывается при $\rho > 0$, правосторонняя асимметрия при $\rho < 0$.

[2.8]. В качестве примера эмпирического распределения частот с истинной асимметрией часто приводят данные по определению кремния. Для распределения, представленного на рис. 2.4, требуется определить асимметрию. Из значений для пробы 1 (первое распределение) получается следующая схема расчета:

x_i	$X_i = 100 x_i - 50$	n_j	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$n_j (X_i - \bar{X})^3$
0,52	+2	9	-1,75	-5,3594	-48,2346
0,53	+3	11	-0,75	-0,4219	-4,6409
0,54	+4	9	+0,25	+0,1563	+1,4067
0,55	+5	5	+1,25	+1,9531	+9,7655
0,56	+6	4	+2,25	+11,3906	+45,5624
0,57	+7	2	+3,25	+34,3281	+68,6562

$$\sum X_i = 150 \quad n = 40 \quad \sum n_j (X_i - \bar{X})^2 = +72,5153$$

$$\bar{X} = 3,75$$

$$\sum X_i^2 = 646 \quad (\sum X_i)^2/n = 563 \quad S = \sqrt{\frac{646 - 563}{39}} = 1,46$$

$$nS^2 = 124,4854$$

$$\rho_1 = \frac{72,5153}{124,4854} = +0,58$$

Для второго распределения частот таким же путем получают $\rho_2 = +0,09$. Асимметрия распределения также проявляется особенно четко при малом содержании пробы 1 (ср. пример [2.3]) *

Недостатки в условиях проведения опыта могут привести к тому, что кривая распределения частот будет иметь слишком заостренный, или, наоборот, слишком

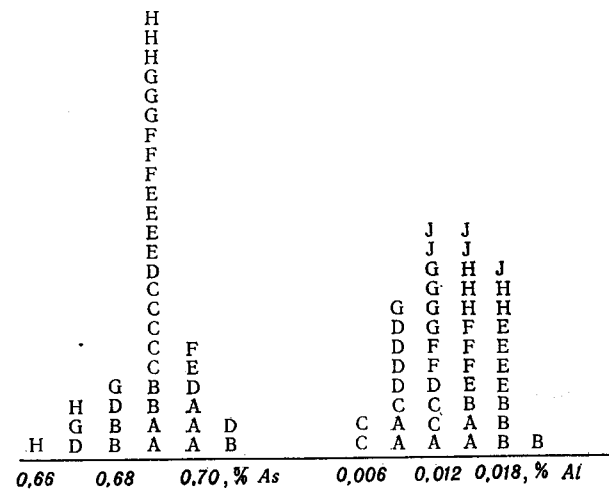
* Приведенные здесь примеры надо рассматривать лишь как некоторую иллюстрацию. Надо проявлять очень большую осторожность при определении асимметрии и эксцесса по малым выборкам. Нужно всегда оценивать степень точности полученных значений так, как это показано в книге Н. В. Смирнова и И. В. Дунина-Барковского (см. стр. 277, 1963). — Прим. ред.

пологий максимум. Заостренное распределение получают, например, если выборка была сделана неслучайно; пологое распределение возникает, если при постановке совместного опыта в отдельных лабораториях преобладали весьма различные условия работы. Подобные искаженные распределения частот характеризуют дополнительным показателем — эксцессом ϵ , который определяется формулой

$$\epsilon = \frac{\sum n_j (x_i - \bar{x})^4}{ns^4} - 3 \quad (2.9)$$

Заостренное распределение частот дает значение $\epsilon > 0$, пологое распределение ведет к $\epsilon < 0$.

[2.9]. Два примера распределений с ненулевым эксцессом приведены на рис. 2.6. Заостренное распределение результатов анализа мышьяка показывает, что здесь была нарушена случайность. Опре-



Р и с. 2.6. Распределение частот с эксцессом, отличным от нуля.

деление малых количеств алюминия в стали особенно сильно подвержено влиянию небольших систематических ошибок, связанных со спецификой работ различных лабораторий. Поэтому при совместно проводимых анализах алюминия часто получают пологие распределения.

Подсчет отклонений для первого распределения дается следующей схемой:

x_i	$X_i = 100x_i - 69$	n_j	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$n_j (X_i - \bar{X})^2$
0,66	-3	1	-3	+81	81
0,67	-2	3	-2	+16	48
0,68	-1	4	-1	+1	4
0,69	0	23	0	0	0
0,70	+1	7	+1	+1	7
0,71	+2	2	+2	+16	32

$$\sum X_i = -2 \quad n = 40 \quad \sum n_j (X_i - \bar{X})^2 = 172$$

$$\bar{X} \sim 0$$

$$\sum X_i^2 = 40 \quad (\sum X_i)^2 / n \sim 0 \quad S = \sqrt{40/39} \sim 1$$

$$nS^2 = 40$$

$$\varepsilon_1 = \frac{172}{40} - 3 = +1,30$$

Подобным же путем для второго распределения частот (А1 в стали) получается $\varepsilon_2 = -0,88$.

2.2.4. ТАБЛИЧНОЕ ВЫЧИСЛЕНИЕ АРИФМЕТИЧЕСКОГО СРЕДНЕГО И СРЕДНЕЙ КВАДРАТИЧНОЙ ОШИБКИ

Вычисление \bar{x} и s может быть особенно трудным при большом числе наблюдений. Очень простое постепенное вычисление этих величин возможно при помощи трафарета — «Таблица для статистической оценки результатов измерений»*. В соответствующие графы этого трафарета вносят каждое измеренное значение в виде штриха. Измеренные значения преобразуют таким образом, чтобы верхняя граница наиболее часто встречающегося класса обозначалась $X = 0$. Затем вычисление ведут по уравнениям (2.1), (2.5) и (2.6) описанным выше способом.

[2.10]. Расчет \bar{x} и s для содержания железа в фосфорной кислоте показан на рис. 2.7**. Путь расчета без детального обсуждения

* Поставляется фирмой «Schäfers Feinpariere», Плауэн (Фогтланд), № 630.

** Взятые за основу значения предоставлены автору Андерсом («Азотный завод Пистерлиц»).

Штриховая таблица для статистической оценки результатов измерений		№	Дата				
		замечания:					
объект: определение H_3PO_4 -Fe		условия проверки:					
верхние границы класса	абсолютная частота		номер класса m	$m \cdot h_m$	$m^2 \cdot h_m$	частота, %	Σ , %
	штриховая таблица	число h_m					
(1)	(2)	(3)	(4)	(5) = (3) × (4)	(6) = (4) × (5)	(7)	(8)
0,001		4	-2	-8	+16	3,9	3,9
0,002		33	-1	-33	+33	31,7	35,6
0,003		44	0	0	0	42,3	77,9
0,004		20	+1	+20	+20	19,2	97,1
0,005		3	+2	+6	+12	2,9	100,0
ширина класса $d=0,001$		принятое среднее значение (середины классов-класс $m=0$) $\bar{x}_a = 0,0025$	$n=104$	$\Sigma m \cdot h_m = 41$	$\Sigma h_m = 26$	$81 = Q$	
среднее значение: $\bar{x} = \bar{x}_a + \frac{d}{n} P = 0,0025 - \frac{0,001}{104} \cdot 15$							$\bar{x} = 0,0024$
дисперсия: $s^2 = \frac{d^2}{n-1} \cdot [Q - \frac{P^2}{n}] = \frac{10^{-6}}{103} [81 - \frac{225}{104}] = 10^{-8} [81 - 2,163]$							
средняя квадратичная ошибка: $s = \sqrt{s^2} = 0,00087$							
коэффициент вариации: $v = \frac{s}{\bar{x}} \cdot 100\% = 100 \cdot 0,00087 / 0,0024 = 36,25\%$							$v =$ %
							прочие значения:

Рис. 2.7. Расчет среднего значения и средней квадратичной ошибки при помощи трафарета — «Таблица для статистической оценки результатов измерений».

очевиден из рисунка. Описанные выше таблицы в основном применяются при подготовительных работах для статистического контроля качества.

2.3. Двумерное распределение

Во всех приведенных выше рассуждениях величину измеренного значения единственной случайной переменной обсуждали в зависимости от частоты ее появления. Однако иногда значение измерения или продукт характеризуется двумя взаимосвязанными случайными величинами. Эти две случайные величины x и y могут быть заданы в одинаковых или разных единицах.

[2.11]. Примером первого случая является измерение почернения линий, даваемых основным металлом и металлом — примесью при количественном эмиссионно-спектральном анализе. Примером второго являются данные о механической прочности и о содержании (%) основных легирующих элементов для характеристики сорта стали.

Как значения x , так и значения y подвержены случайным колебаниям. В рамках этих случайных ошибок возможны любые комбинации x и y для исследуемой пробы. Если требуется охарактеризовать цифровой материал, полученный в одном опыте, при помощи ступенчатой диаграммы, то необходимо прибегнуть к трудному для представления трехмерному изображению. Оси переменных x и y лежат в этом случае в основании фигуры, а частоты представляются вертикально в пространстве. Из-за трудности такого представления отдельные точки изображают на двумерной плоскости xy и судят о плотности точек. Максимум поверхности находится там, где обнаруживается наибольшая плотность точек в двумерном изображении. Вообще, все измеренные значения лежат внутри эллипса или окружности. Подобные распределения, в которых рассматривается частота двух взаимосвязанных случайных величин, называют *двумерными распределениями*. Двумерные распределения характеризуются также средним значением и рассеянием. Эти характеристики следует вычислять отдельно для обеих случайных переменных x и y . Точка $M(x, y)$ лежит в месте теоретически ожидаемого максимума частоты. Общее рассеяние s составляет сумму квадратов двух единичных рас-

сеяний (суммируются таким образом дисперсии). Подробности можно найти у Смирнова и Дунина-Барковского.

[2.12]. Для контроля качества в пробах стали сорта GS 50 определяется содержание углерода x (C, %) и механическая прочность y (кг/мм²). В течение одного квартала были получены следующие результаты *:

x	y	x	y	x	y	x	y
0,30	58,9	0,35	53,5	0,37	60,2	0,29	57,2
0,33	61,4	0,32	59,3	0,33	54,4	0,30	55,5
0,37	61,2	0,39	58,2	0,34	54,5	0,33	55,5
0,36	57,2	0,30	53,8	0,33	56,2	0,32	51,8
0,31	54,8	0,32	56,6	0,30	57,6	0,32	53,9
0,29	53,7	0,32	56,2	0,34	59,6	0,38	55,7
0,34	57,4	0,38	60,1	0,36	60,5	0,37	55,8
0,39	57,0	0,37	58,7	0,33	57,5	0,34	58,7
0,37	54,0	0,38	58,7	0,34	57,0	0,35	58,0
0,38	57,5	0,33	61,4	0,36	55,0	0,36	56,0

Из этих 40 пар значений (рис. 2.8) создается двумерное распределение частоты. Средние значения и средние квадратичные ошибки подсчитывают отдельно для обеих измеренных величин. Получают

$$\bar{x} = 0,34\% \text{ C}$$

$$\bar{y} = 57,0 \text{ кг/мм}^2$$

Вследствие этого средняя точка распределения M лежит при значениях $\bar{x} = 0,34$, $\bar{y} = 57,0$. В качестве средней квадратичной ошибки находят

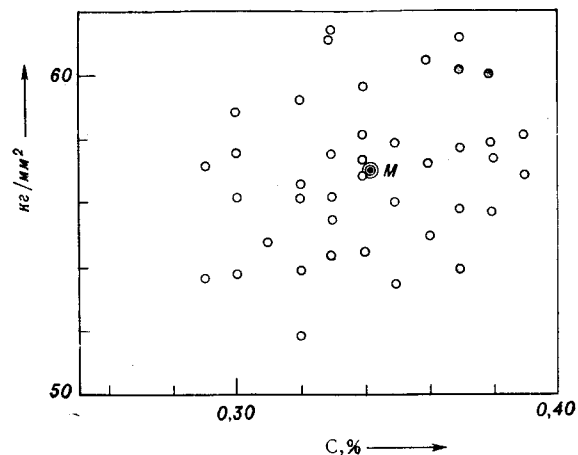
$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 0,03\% \text{ C}$$

$$s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = 2,4 \text{ кг/мм}^2$$

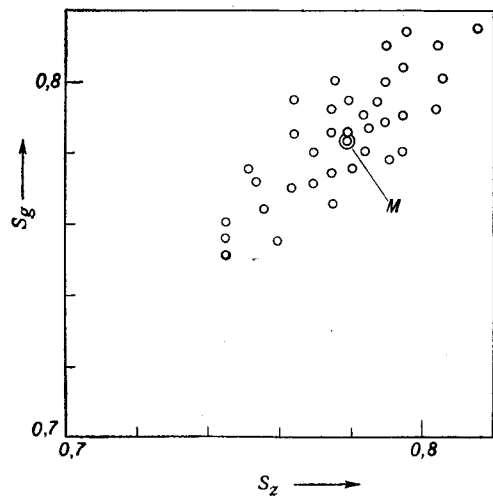
Из вида двумерного распределения можно сделать обратные выводы о степени связи обеих случайных переменных x и y .

[2.13]. В количественном эмиссионно-спектральном анализе почернение S_z , измеренное для анализируемой линии, относят к почернению линии основного металла S_g . Этим уменьшают случайные

* Эти данные предоставил автору Шульц (DAMW, Магдебург).



Р и с. 2.8. Двумерное распределение частот, характеризующее качество стали марки GS 50.

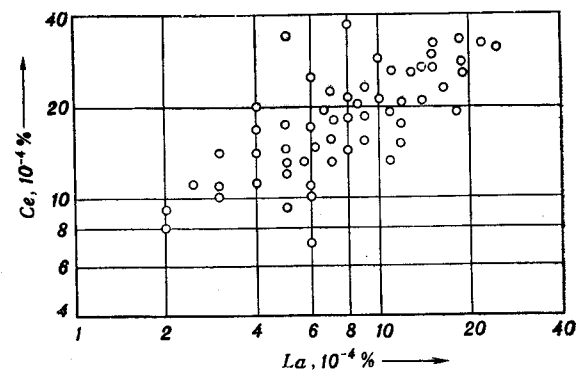


Р и с. 2.9. Двумерное распределение частот почернений, получаемых при количественном эмиссионно-спектральном анализе,

колебания, вызываемые, например, такими факторами, как неправильное положение источника излучения. Случайные колебания исключаются тем лучше, чем сильнее изменения почернений двух спектральных линий зависят друг от друга (Холдт [6]). На рис. 2.9 графически показаны почернения S_z и S_g , полученные при методических исследованиях. Связь между двумя величинами ясно выражена вытянутым эллиптическим распределением частот.

При двумерных распределениях может возникнуть необходимость в изменении масштаба. При этом особенно часто применяют логарифмическое преобразование.

[2.14]. Для характеристики месторождений полевого шпата особенно важно определение следов церия и лантана [7]. Двумерное логарифмическое распределение частот этих элементов для рассмотренных месторождений иллюстрирует рис. 2.10. При дальнейших



Р и с. 2.10. Двумерное логарифмическое распределение частот, полученное при определении содержания церия и лантана в полевом шпате.

исследованиях было установлено, что абсолютное содержание церия и лантана меняется от месторождения к месторождению. Однако во всех месторождениях оказывается одинаковое рассеяние и одинаковая степень корреляции содержания этих элементов. Отсюда можно сделать вывод о существовании сходных условий при образовании рассмотренных месторождений.

Графическое двумерное представление распределения частот позволяет легко оценить систематические ошибки при проведении совместных исследований в разных лабораториях. В этом случае необходимо иметь два до некоторой степени однородных вещества, которые незначи-

тельно отличаются по содержанию определяемых элементов. Каждая участвующая лаборатория анализирует обе пробы по два раза. При этом следует придавать особое значение независимости измеренных значений (ср. стр. 107). Если пробы обозначают x и y и нумеруют лаборатории от 1 до m , то получают m пар значений: $\bar{x}_1\bar{y}_1$; $\bar{x}_2\bar{y}_2$; ... $\bar{x}_m\bar{y}_m$. Эти m пар значений переносят в систему координат с равным делением абсциссы и ординаты. Через точку со средним значением M (\bar{x} , \bar{y}) проводят новые координатные оси, параллельные первым осям. При этом $\bar{x} = \sum \bar{x}_i/m$ и $\bar{y} = \sum \bar{y}_i/m$. Если имеют место только случайные ошибки, то положительные и отрицательные отклонения от средней точки M обладают равной вероятностью. При этом в каждом из четырех квадрантов находят одинаковое число точек. Однако если в данных одной или нескольких лабораторий становится заметной систематическая положительная или отрицательная ошибка, то эта ошибка сказывается как на результатах анализа пробы x , так и пробы y . Таким образом появляется избыток точек в первом и третьем квадрантах.

[2.15]. При совместном участии четырнадцати лабораторий калийной промышленности были исследованы две пробы*. На рис. 2.11

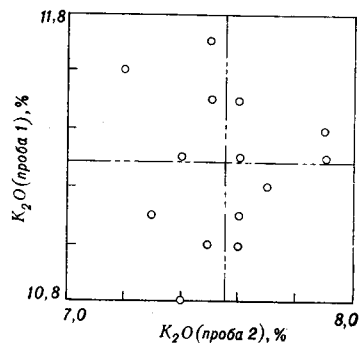


Рис. 2.11. Двумерное распределение частот, полученное при определении калия в двух пробах.

* Эти данные предоставлены автору доктором Шубертом (KFI, Зондсхаузен).

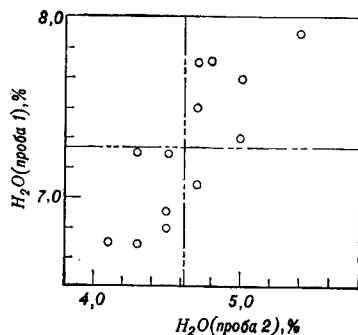


Рис. 2.12. Двумерное распределение частот, полученное при определении влажности в двух пробах.

приведены результаты определения калия, на рис. 2.12 — результаты определения влажности пробы.

При определении калия точки распределяются удивительно равномерно по четырем квадрантам. Напротив, при определении влажности подобного регулярного распределения точек не обнаружено. Точки разбросаны вдоль прямой, проходящей примерно под 45° к оси абсцисс. Это указывает на наличие систематической ошибки, которая в отдельных лабораториях имеет одинаковые знаки, но различные величины. В результате этого исследования можно высказать предположение о наличии отрицательной ошибки в данных большинства лабораторий.

Для большинства совместных исследований получают распределения, точки которых лежат внутри эллипса. Его большая ось определяется биссектрисой угла первого и соответственно третьего квадрантов. Чем сильнее сказываются систематические ошибки по сравнению со случайной, тем длиннее и уже становится эллипс. Плотное распределение отдельных точек вдоль биссектрисы угла является доказательством того, что метод анализа неудовлетворителен. С другой стороны, может возникнуть подозрение, что некоторые лаборатории применили отличающиеся методы, если большинство точек рассеяно внутри довольно широкого эллипса и только некоторые из них, находящиеся в первом или третьем квадрантах, лежат далеко вне эллипса.

ЛИТЕРАТУРА

1. Ahrens L. H., Geochim. cosmochim. Acta, 5, 49 (1954).
2. Daeves K., Beckel A., Grobzahlenmethodik und Häufigkeitsanalyse, Verlag Chemie, Weinheim, 1958.
3. Doerffel K., Z. anal. Chem., 184, 81 (1961).
4. Doerffel K., Geyer R., Wiss. Z. Techn. Hochschule Chem. Leuna-Merseburg, 6, 251 (1964).
5. Gaddum J. H., Nature (Paris), 156, 463 (1945).
6. Holdt G., Appl. Spectroscopy, 14, 64 (1960).
7. Leeder O., Diss., Freiberg, 1965.
8. Stevens S. S., Science, 121, 113 (1956).
9. Youden W. J., Anal. Chem., 32, 12, 23A (1960).

3. Теоретические распределения

Обсужденные в гл. 2 функции распределения сопровождаются упорядоченной систематизацией измерений и их графическим изображением. При этом, если случайные ошибки действительно малы, всегда обнаруживается похожая картина. Это позволяет предположить, что в основе подобных распределений лежат определенные математические закономерности. Некоторые из этих закономерностей для случая генеральной совокупности и выборки изложены ниже*.

3.1. Нормальное распределение

Допустим, что имеется очень много наблюдаемых значений ($n \rightarrow \infty$). На эти значения влияет несколько случайных причин, их влияние аддитивно и оно мало по сравнению со значением измерения. При очень небольшой ширине класса ($d \rightarrow 0$) распределение частот одной величины можно описать следующей функцией:

$$y = h(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.1)$$

Эта функция распределения называется *нормальным* или *гауссовым распределением*. Уравнение (3.1) описывает плотность вероятности этого распределения, μ и σ являются какими-нибудь действительными числами, их называют *параметрами* распределения. Если даны μ и σ , то y является только функцией x . Положение и форма кривой полностью определяются значениями обоих параметров μ и σ . Максимум кривой лежит в точке $x = \mu$,

точки перегиба находятся при $x_1 = \mu - \sigma$ и $x_2 = \mu + \sigma$ (рис. 3.1). Кривая достигает значения $y = 0$ при $x = \pm \infty$. Однако значениями ординаты при $x = \mu \pm 3\sigma$ практически уже можно пренебречь. Рис. 3.2 показывает три равных по площади гауссовых кривых с одинаковым средним значением μ , но с различной величиной средней квадратичной ошибки σ .

При уменьшающейся средней квадратичной ошибке σ кривые становятся все более узкими и игольобразными.

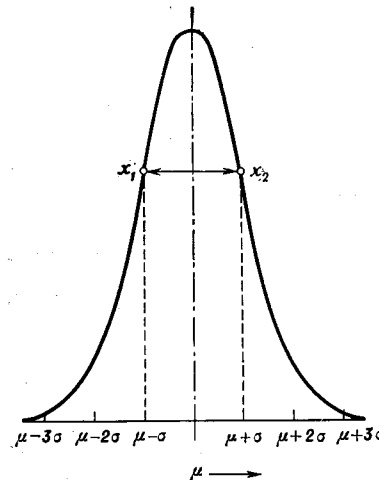


Рис. 3.1. Геометрическая интерпретация средней квадратичной ошибки.

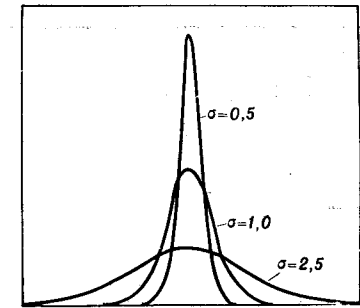


Рис. 3.2. Кривые нормальных распределений при различной средней квадратичной ошибке.

Большинство результатов измерений в обычных методах анализа следует нормальному распределению. Исключениями являются только счетные методы анализа (ср. разд. 3.2), а также, при известных условиях, методы, при которых оцениваются какие-нибудь биологические процессы (например, определение числа микробов в питьевой воде). Для ряда методов анализа (анализ следов вещества, полуколичественные методы) заранее нельзя предположить, что имеется нормальное распределение (при использовании линейной шкалы) [2, 4].

Для построения гауссовой кривой прежде всего при данной средней квадратичной ошибке определяют пик

* Для понимания математических положений в приложении приведены основные математические работы.

ординаты $y_{\max} = \frac{1}{\sigma} \sqrt{2\pi}$ при $x = \mu$. Дальнейшие значения ординаты получают из табл. 3.1.

Таблица 3.1

Значения ординат для гауссовой кривой

x	μ	$\mu \pm 0,5\sigma$	$\mu \pm \sigma$	$\mu \pm 1,5\sigma$	$\mu \pm 2\sigma$	$\mu \pm 3\sigma$
y	y_{\max}	$\frac{7}{8} y_{\max}$	$\frac{5}{8} y_{\max}$	$\frac{2,5}{8} y_{\max}$	$\frac{1}{7} y_{\max}$	$\frac{1}{80} y_{\max}$

Для практического применения нормируют уравнение 3.1, полагая $\frac{x-\mu}{\sigma} = u$. Тогда получают $\sigma_u = 1$, $\mu_u = 0$ и

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (3.2)$$

Значения ординаты, нормированной функции распределения в зависимости от u , можно найти в табл. 12.1.

В одномерном случае плотность вероятности можно представить в виде кривой на плоскости. На ось абсцисс наносят значения независимой переменной x , на ось ординат — полученные значения y . Подобным же образом можно интерпретировать нормальное распределение двух величин (ср. разд. 2.3), используя пространственное представление. Оба значения случайных переменных x_1 и x_2 наносят на координатные оси в плоскости основания, а затем значения y наносят на уходящую в пространство ось. При этом оказывается, что объемная картина с эллиптическим основанием имеет максимум, лежащий в точке с координатами $x_1 = \mu_1$ и $x_2 = \mu_2$. Положение эллипса определяется зависимостью или независимостью x_1 и x_2 друг от друга (рис. 3.3, ср. также стр. 38).

Функции распределения можно построить также пользуясь средними значениями $\bar{x}_1 \dots \bar{x}_m$, полученными из n_j параллельных определений. Каждую такую серию

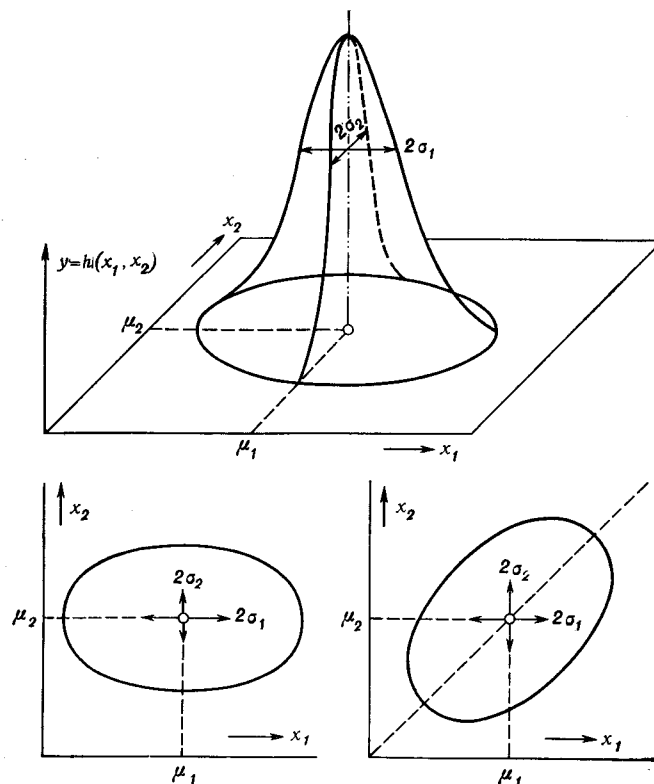


Рис. 3.3. Поверхность двумерного нормального распределения и положение ее основания при независимости (слева) и зависимости (справа) между x_1 и x_2 .

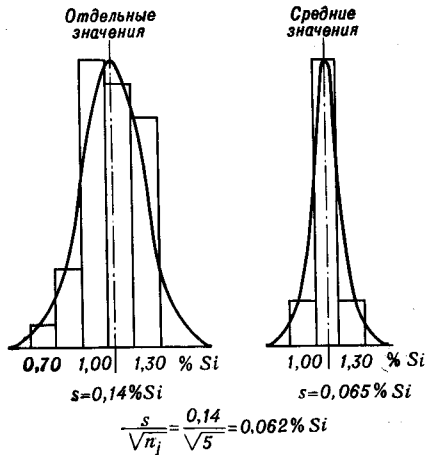
измерений объемом n_j можно рассматривать как выборку из одной и той же генеральной совокупности. Математически можно показать, что общее среднее этих выборок \bar{x} равно среднему значению μ генеральной совокупности. Следовательно, при одномерном распределении

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} = \mu$$

Но средняя квадратичная ошибка σ_M меньше, чем σ генеральной совокупности

$$\sigma_M = \frac{\sigma}{\sqrt{n_j}} \quad (3.3)$$

Функция распределения, построенная для средних значений, более иглообразна, чем для единичных (рис. 3.4),



Р и с. 3.4. Функция распределения частот для отдельных значений (слева) и для средних значений (справа) при определении кремния ($n_j = 5$ параллельных определений).

так как при получении средних значений сглажены отдельные высокие и низкие значения результатов единичных измерений*.

Для практического применения особенно важно, чтобы средние значения, полученные по крайней мере из $n_j = 5$ измерений, в общем достаточно хорошо следуют нормальному распределению даже тогда, когда единичные значения не распределены нормально. Это тем более справедливо, чем больше проведено число параллельных определений n_j . По уравнению (3.3) из большого числа средних значений можно также оценить среднюю квад-

* Уравнение (3.3) пригодно только тогда, когда единичные значения, относящиеся к единичным средним значениям $x_1 \dots x_m$, распределены беспорядочно. Уравнение (3.3) не выполняется, когда единичные значения образуют группы (ср. рис. 2.1).

ратичную ошибку для распределения единичных значений с тем, чтобы использовать эту величину в других целях (например, для оценки рассеяния единичных значений).

Интегрированием функции распределения для нормированного распределения одной величины [уравнение (3.2)] в пределах $-\infty \dots +\infty$ получают значение площади F , заключенной между гауссовой кривой и осью абсцисс:

$$F = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du \quad (3.4)$$

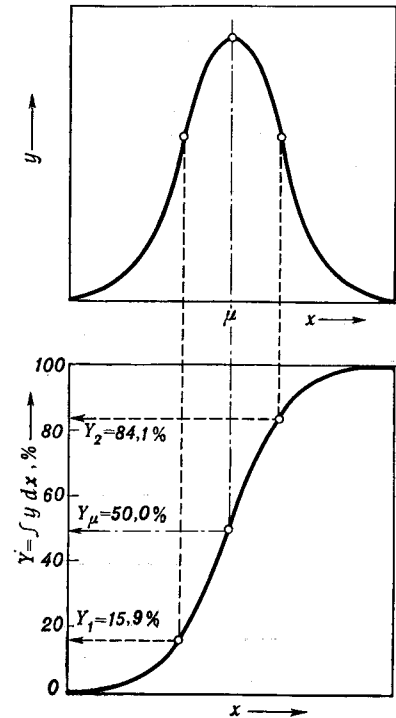
Получающуюся при интегрировании площадь равна единице (соответственно принимают за 100%). При переменном верхнем пределе интегрирования x получается

$$Y = F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du \quad (3.5)$$

Это выражение называют *гауссовым интегралом ошибок*.

Графическое представление этой функции в сопоставлении с колоколообразной кривой показано на рис. 3.5. Максимум колоколообразной кривой соответствует точке перегиба $Y = 0,5$ (или 50%) на интегральной кривой.

Две точки перегиба гауссовой кривой лежат при значениях интегральной кривой $Y_1 = 0,159$ (15,9%) и $Y_2 = 0,841$ (84,1%). Интегральную кривую можно представить прямой, если воспользоваться так называемой вероятностной



Р и с. 3.5. Гауссова кривая и относящаяся к ней интегральная кривая.

бумагой *. Эта прямая тем круче, чем меньше случайная ошибка (если ординату делят только до $Y = 0,50$ (50%), то получают частотную сетку, использованную в примере [2.2]).

Вероятностная бумага позволяет быстро проверить гипотезу о том, что найденное эмпирическое распределение частот взято из нормально распределенной генеральной совокупности. Найденные значения упорядочивают соответственно их величине по классам и подсчитывают [по уравнению (3.5)] процентную долю Y_i всех значений, лежащих ниже границы x_i . При нормальном распределении пары значений (x_i, Y_i) в области $10\% < Y_i < 90\%$ рассеяны вдоль некоторой прямой.

Пользуясь вероятностной бумагой, можно быстро и просто определить параметры μ и σ нормального распределения. Среднее значение μ находят при абсциссе, соответствующей $Y = 50\%$. Средняя квадратичная ошибка получается из полуразницы значений абсциссы, соответствующих значениям ординаты $Y_2 = 84,1\%$ и $Y_1 = 15,9\%$ (ср. стр. 53).

[3.1]. Необходимо установить, соответствуют ли нормальному распределению результаты, найденные в примере [2.1]. По уравнению (3.5) получают следующую таблицу накопленных частот:

Предел x_i , % Al	Частота абсолютная	Накопленная частота	
		абсолютная	процентная (Y_i)
0,008	2	2	3,3
0,010	4	6	10,0
0,012	16	22	36,7
0,014	18	40	66,7
0,016	14	54	90,0
0,018	5	59	98,3
0,020	1	60	100,0

* Поставляется фирмой «Schäfers Feinpapiere», Плауэн (Фогтланд); № 500 — для линейной шкалы на оси абсциссы, № 485 — для логарифмической шкалы.

Соотнесенные значения пар (x_i, Y_i) наносят на вероятностную бумагу (рис. 3.6). Так как отдельные точки очень мало отклоняются от прямой, нет никаких оснований отбросить предположение, что существует нормальное распределение. Среднее содержание пробы дает значение абсциссы, соответствующее $Y = 50\%$

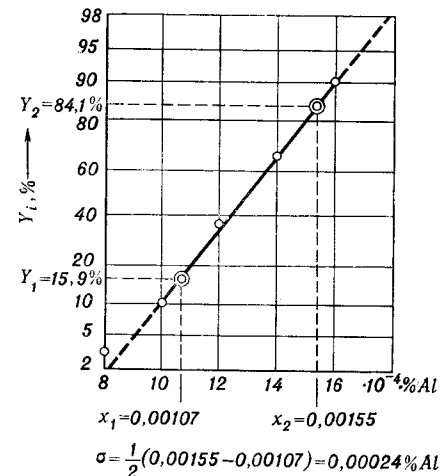


Рис. 3.6. Проверка нормального распределения на вероятностной бумаге.

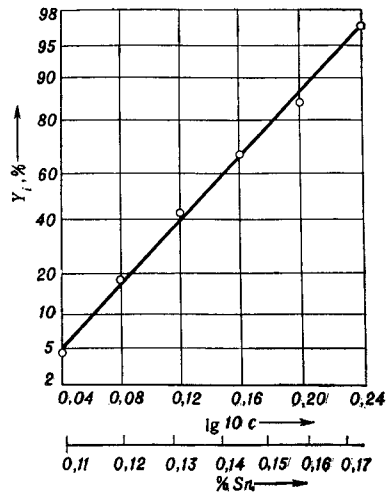
и $\mu = 0,0130\% \text{ Al}$. Среднюю квадратичную ошибку находят из полуразности значений абсциссы, соответствующих $Y_2 = 84,1$ и $Y_1 = 15,9\%$.

$$\sigma = \frac{1}{2} (0,0155 - 0,0107) = 0,0024\%$$

Описанные методы следует применять только тогда, когда имеется по меньшей мере 30 измерений. Отдельные точки могут только немного рассеиваться вдоль прямой. В сомнительных или трудных случаях следует вернуться к математической проверке (ср. разд. 7.6). Если при проверке на вероятностной бумаге не получается прямой, то это может свидетельствовать о неподходящем выборе шкалы измерений (например, возможно существование логарифмически нормального распределения).

[3.2]. Необходимо установить, соответствуют ли найденные в примере [2.4] результаты определения олова нормальному распределению. На основании графика (рис. 2.5) нужно ожидать логарифмического нормального распределения. Поэтому для логариф-

мов значений анализа подсчитывают процентную накопленную частоту аналогично примеру [3.1]. На вероятностной бумаге делят абсциссу в логарифмическом масштабе соответственно образованным классам. Отдельные точки мало отклоняются от прямой (рис. 3.7),



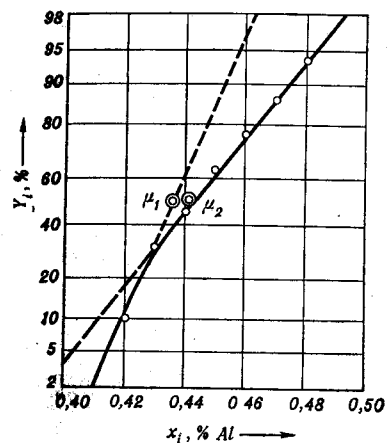
Р и с. 3.7. Проверка логарифмического распределения на вероятностной бумаге.

и нет никакого основания отбросить предположение о том, что после логарифмического преобразования получается нормальное распределение.

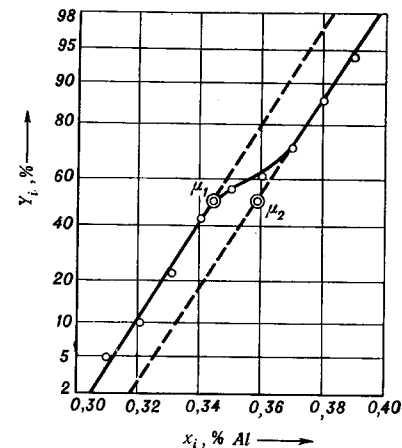
Если имеющиеся значения измерений распределяются в широкой области (несколько десятков процентов), то работа облегчается применением вероятностной бумаги с логарифмическим делением оси качества. Однако эта функциональная бумага неприменима, если значения лежат в узкой области, как в примере [3.2]. Тогда длину на абсциссе следует сжать.

На вероятностной бумаге часто обнаруживается распределение с двумя пиками из-за того, что расположение точек позволяет провести две прямые с различными значениями, соответствующими $Y = 50\%$. Прямые пересякаются в случае, если средние квадратичные ошибки обеих составных частей отличаются ($\sigma_1 \neq \sigma_2$), они параллельны, если $\sigma_1 = \sigma_2$.

[3.3]. В качестве примера первого случая служит представление задачи [2.2] на вероятностной бумаге, приведенное на рис. 3.8. Рис. 3.9 относится к совместному определению алюминия в магнелиевом сплаве.



Р и с. 3.8. Распределение с двумя максимумами при $\mu_1 \neq \mu_2$ и $\sigma_1 \neq \sigma_2$.



Р и с. 3.9. Распределение с двумя максимумами при $\mu_1 \neq \mu_2$ и $\sigma_1 = \sigma_2$.

Как и при построении частотных диаграмм (ср. стр. 48), в этом случае следует принимать во внимание только ясно обнаруженные явления. Незначительные отклонения от прямой почти никогда не бывают достаточно убедительными. Доказательство существования смешанного распределения часто легче получить, пользуясь вероятностной бумагой, чем при построении обычных частотных диаграмм, так как легче оценивать рассеяние точек относительно прямой. В смешанных распределениях с $\mu_1 = \mu_2$ и $\sigma_1 \neq \sigma_2$ лучше всего для такого анализа пригодна частотная диаграмма.

Если гауссов интеграл подсчитывают в пределах

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-u}^{+u} e^{-\frac{u^2}{2}} du \quad (3.6)$$

то получают долю P от общей площади $F = 1,000$ (рис. 3.10). Эта часть площади представляет собой вероят-

ность P появления значения измерения, лежащего в области $-\mu\sigma \dots +\mu\sigma$. Вероятность того, что значение

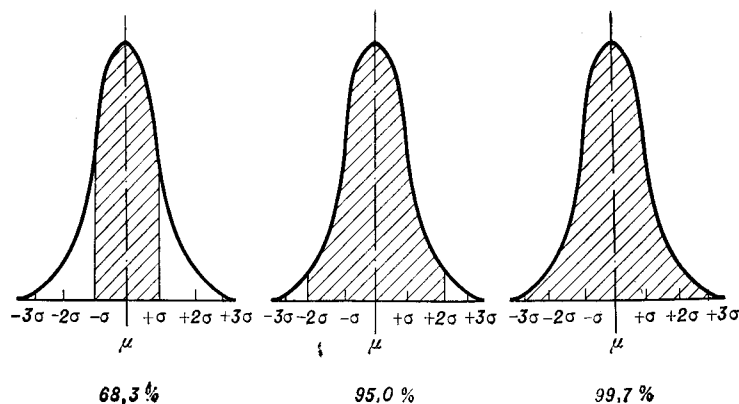


Рис. 3.10. Интегрирование в пределах $\mu \pm u(P)\sigma$. Заштрихованная площадь составляет соответственно 68,3, 95,0 и 99,7% всей площади.

измерения находится вне указанных пределов интегрирования равна $\alpha = 1 - P$. Часть площади P также выражают в процентах от всей площади и называют *статистической надежностью*.

Чем дальше раздвинуты пределы интегрирования $\pm u$, тем больше будет площадь P и тем больше значений будет лежать внутри и меньше вне пределов $\pm u\sigma$ (рис. 3.10). Из табл. 3.2 следует, что при достаточном числе измерений:

Таблица 3.2

Некоторые значения гауссового интеграла при интегрировании в пределах $-u \dots +u$

u	P	$\alpha = 1 - P$	u	P	$\alpha = 1 - P$
0,500	0,383	0,617	1,96	0,95	0,05
0,675	0,500	0,500	2,58	0,99	0,01
1,000	0,683	0,317	3,00	0,9973	0,0027
1,640	0,900	0,100	4,00	0,9999	0,0001

в области $-\sigma \dots +\sigma$ лежит 68,3% (почти две трети) всех результатов. Около 15% значений меньше $-\sigma$ и около 15% больше $+\sigma$;

внутри пределов $-1,96\sigma \dots +1,96\sigma$ находятся 95% всех значений; 2,5% лежат ниже $-1,96\sigma$ и 2,5% выше $+1,96\sigma$;

интервал $-3\sigma \dots +3\sigma$ охватывает 99,73%, следовательно, практически все значения. Итого, только 0,27% значений находятся вне этих пределов.

Часть площади P , подсчитанная по уравнению (3.6), определяется выбранными пределами интегрирования. Выбор одной величины определяет другие. Чтобы пояснить эту связь, в дальнейшем пределы интегрирования обозначим $u(P)$.

При помощи табл. 3.2 легко объяснить графическое определение средней квадратичной ошибки, приведенное на стр. 49. Площадь под гауссовой кривой в области $-\sigma \dots +\sigma$ составляет 68,3% всей площади. В графе накопленных частот находим значения абсцисс $-\sigma$ и $+\sigma$; им соответствуют ординаты $Y_1 = 50 - \frac{68,3}{2} =$

$$= 15,9\% \text{ и } Y_2 = 50 + \frac{68,3}{2} = 84,1\%.$$

Вероятности, приведенные в табл. 3.2, естественно пригодны, только если предположить, что выполняется нормальное распределение. Связь между пределами интегрирования и частью площади P можно получить для любых распределений, однако при этом значительно меньше значений будет лежать внутри пределов интегрирования. При любом распределении частот с максимумом вблизи арифметического среднего находят, скажем, в области $\pm 2\sigma$ только приблизительно 89%, в области $\pm 3\sigma$ только $\sim 95\%$ всех значений.

Средняя квадратичная ошибка средних значений для некоррелированных значений дается уравнением (3.3) в виде

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n_j}}.$$

При этом n_j означает число параллельных определений, по которым получают каждое из средних значений. Раз-

ности между выборочным средним и средним значением генеральной совокупности μ лежат примерно в P случаях в пределах $-u(P)\sigma_m$ и $+u(P)\sigma_m$. Итак,

$$-u(P)\frac{\sigma}{\sqrt{n_j}} < \mu - \bar{x} < +u(P)\frac{\sigma}{\sqrt{n_j}} \quad (3.7)$$

Прибавляя \bar{x} , получают

$$\bar{x} - u(P)\frac{\sigma}{\sqrt{n_j}} < \mu < \bar{x} + u(P)\frac{\sigma}{\sqrt{n_j}} \quad (3.8)$$

При очень частом повторении серии измерений можно ожидать, что 100 $P\%$ всех этих выборок должно находиться внутри интервала $\bar{x} \pm u(P)\frac{\sigma}{\sqrt{n_j}}$. В отсутствие систематической ошибки аналитически установленное среднее значение отличается от истинного содержания пробы меньше чем на $u(P)\frac{\sigma}{\sqrt{n_j}}$. Величина

$$u(P)\frac{\sigma}{\sqrt{n_j}} = \Delta\bar{x} \quad (3.9)$$

называется *доверительным интервалом* среднего значения \bar{x} ; тем самым задается частота, с которой появляются ошибки величины $\Delta\bar{x}$. В известной степени его можно использовать, чтобы охарактеризовать ошибку, которую следует ожидать для отдельного среднего значения. Следует подчеркнуть, однако, что при этом речь идет о статистической величине с вероятностью P , а не об особой ошибке, принадлежащей особому среднему значению.

При определении доверительного интервала следует установить, обе границы — верхняя и нижняя — интересны для анализа или только одна из них (рис. 3.11). Если доверительный интервал используют в виде равенства (3.9) как меру разброса для среднего значения выборки, то, естественно, интересны обе границы. Тогда говорят о *двусторонней границе* с вероятностью P .

При такой постановке вопроса ординатами $\bar{x} \pm u(P)\frac{\sigma}{\sqrt{n_j}}$ отсекаются справа и слева площади, равные $\frac{1}{2}(1 - P) =$

$= \frac{\alpha}{2}$. Часто устанавливают только одностороннее требование, например, что содержание примесей продукта не должно превышать некоторой верхней границы. В этом

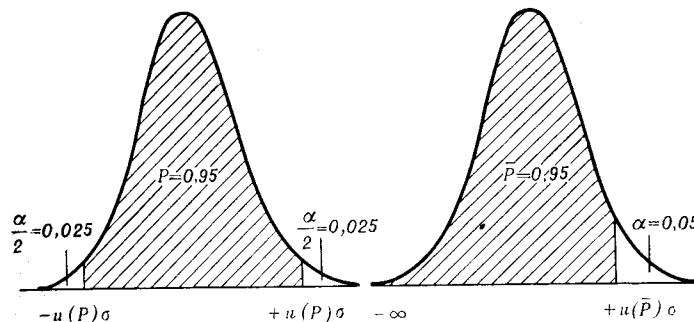


Рис. 3.11. Границы доверительного интервала при двусторонней (слева) и односторонней (справа) постановке задачи.

случае говорят об *односторонней границе* с соответствующей вероятностью \bar{P} . Она дается частью площади, соответствующей значениям от $x = -\infty$ до $x = u(\bar{P})\frac{\sigma}{\sqrt{n_j}}$ (рис. 3.11).

При установлении односторонней границы ордината отсекает справа или слева, в точке $\bar{x} + u(\bar{P})\frac{\sigma}{\sqrt{n_j}}$ или

Таблица 3.3

Значение P и \bar{P} для дву- и соответственно односторонней границы

P	\bar{P}
0,90	0,95
0,95	0,975
0,98	0,990
0,99	0,995
0,999	0,9995

соответственно в точке $\bar{x} - u(\bar{P}) \frac{\sigma}{\sqrt{n_j}}$, площадь, равную $1 - \bar{P} = \alpha$. Между вероятностями для односторонней границы \bar{P} и двусторонней P существует соотношение

$$\bar{P} = 0,5 + \frac{P}{2} \quad (3.10)$$

Более полное сопоставление соотношения величин P и \bar{P} дается в табл. 3.3.

3.2. Распределение Пуассона

Ряд методов современной аналитической химии дает результаты в виде целых величин. Примером этого является подсчет импульсов в радиохимии, подсчет квантов в рентгеноспектральном анализе, подсчет структурных элементов при исследовании шлифов и прочее. Все эти методы характеризуются одним свойством — число возможных событий очень велико (например, число распадающихся ядер атомов), а число фактически происходящих событий, напротив, очень мало (распад отдельных ядер). Вследствие редкости этих событий в наблюдаемом интервале времени состав пробы меняется несущественно. Если многократно повторяют один и тот же опыт, то зависимость между величиной измеренного значения и частотой ее появления можно представить следующим образом:

$$y = \frac{\mu^x e^{-\mu}}{x!} \quad (3.14)$$

Этот вид распределения носит название *распределения Пуассона*. Поразительно то, что распределение Пуассона характеризуется только одним параметром — средним значением μ . Между средним значением μ и средней квадратичной ошибкой σ существует зависимость

$$\sigma = \sqrt{\mu} \quad (3.12)$$

В отличие от нормального распределения распределение Пуассона является дискретным. Для малых значений μ оно обладает значительной асимметрией (рис. 3.12). Асимметрия очень быстро уменьшается с ростом μ , и форма кривой распределения приближается к нормальному распределению со средним значением μ и средней квадратич-

ной ошибкой $\sigma = \sqrt{\mu}$. Для практических целей достаточно удовлетворительное приближение к нормальному распределению наступает при $\bar{x} > 15$. Тогда, соответственно табл. 3.2, 68,3% всех значений лежат в пределах $\mu - \sqrt{\mu}$ и $\mu + \sqrt{\mu}$.

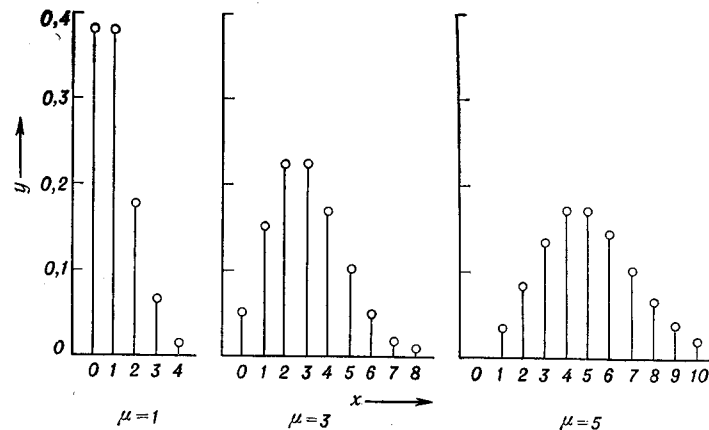


Рис. 3.12. Распределения Пуассона для различных значений арифметического среднего μ .

Возможности приближения к нормальным распределениям позволяют применять вероятностную бумагу для проверки гипотезы о выполнении распределения Пуассона. В этом случае накопленные частоты дают прямую, проходящую через точки

$$P_m (x_m = \mu; Y_m = 50\%)$$

$$P_1 (x_1 = \mu - \sqrt{\mu}; Y_1 = 15,9\%)$$

$$P_2 (x_2 = \mu + \sqrt{\mu}; Y_2 = 84,1\%)$$

Для практического выполнения этой проверки прежде всего строят сглаживающую прямую, пользуясь накопленными частотами и соответствующими им значениями содержания вещества. Из отдельных измерений определяют значение арифметического среднего \bar{x} и подсчитывают по равенству (3.12) координаты точек P_1 и P_2 . Прямая, проходящая через эти точки, должна почти совпадать с ранее построенной прямой.

[3.4]. При помощи счетчика измерили число импульсов, вызванных α -излучением. При составлении функции распределения частот получились следующие процентные соотношения отдельных классов:

Верхняя граница класса x_i , импульсы	Частота, %	Накопленная частота Y_i , %	Верхняя граница класса x_i , импульсы	Частота, %	Накопленная частота Y_i , %
3810	5	5	4010	19	87
3850	7	12	4050	8	95
3890	9	21	4090	3	98
3930	23	44	4130	1	99
3970	24	68	4170	1	100

Пары значений (x_i, Y_i) распределения суммарной частоты наносят на вероятностную бумагу и сравнивают с прямой рис. 3.13.

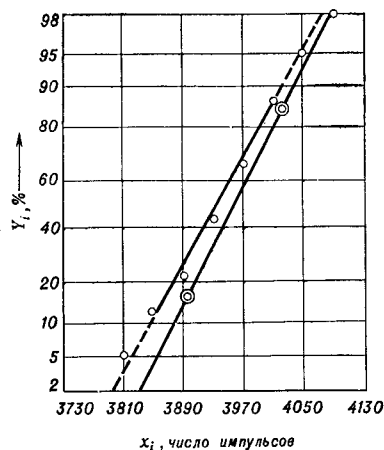


Рис. 3.13. Проверка распределения Пуассона на вероятностной бумаге.

Арифметическое среднее, полученное из ста результатов по уравнению (2.1), лежит при $x = 3958$ импульсов. Отсюда, пользуясь уравнением (3.12), получают обе точки P_1 и P_2 , относящиеся к те-

оретическому распределению. Значения их абсцисс равны $x_1 = 3958 - \sqrt{3958} = 3895$ и $x_2 = 3958 + \sqrt{3958} = 4021$, а соответствующие значения ординат: $Y_1 = 15,9\%$ и $Y_2 = 84,1\%$. Прямая, проведенная через точки P_1 и P_2 , почти совпадает с прямой из уравнения. Поэтому можно принять распределение Пуассона.

Если из графической проверки нельзя сделать достаточно точного вывода, то обращаются к описанному в дальнейшем математическому способу проверки (ср. разд. 7.6).

3.3. Специальные распределения

3.3.1. t -РАСПРЕДЕЛЕНИЕ

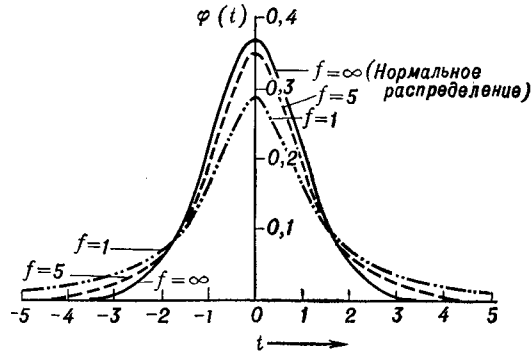
Нормальное распределение, описанное в разд. 3.1, подходит только для случая очень большого числа измерений. При малом числе измерений распределение может более или менее отклоняться от нормального. В математической статистике эта дополнительная ненадежность устраняется модифицированным симметричным распределением — t -распределением. Максимумы частоты нормального и t -распределения лежат при одном и том же значении абсциссы. Однако в отличие от нормального распределения высота и ширина кривых нормированного t -распределения зависят от степеней свободы f соответствующей средней квадратичной ошибки. Чем меньше число степеней свободы, тем более пологий ход имеет кривая при одной и той же средней квадратичной ошибке (рис. 3.14). При $f \rightarrow \infty$ t -распределение переходит в нормальное распределение. Соответственно этому для хода кривой, зависящего от f , пределы интегрирования при заданной вероятности P все больше удаляются от среднего значения с уменьшением числа степеней свободы f . Так, для $P = 0,95$ измеренные значения x больше не лежат в области $\mu - 1,96s \dots \mu + 1,96s$. Этот интервал становится тем шире, чем меньше измерений было проведено (рис. 3.15). Пределы интегрирования t -распределения в зависимости от вероятности P и степени свободы f для нормированного по $s = 1$ распределения приведены в табл. 12.3.

3.3.2. F -РАСПРЕДЕЛЕНИЕ

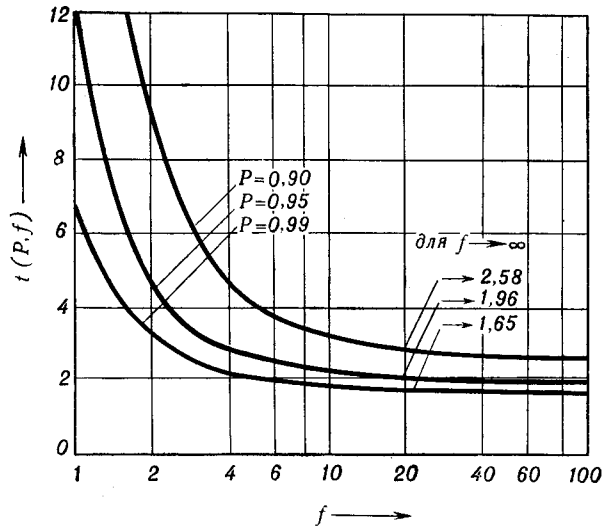
Из нормально распределенной генеральной совокупности взято две выборки объемом n_1 и n_2 . Подсчитывают дисперсии s_1^2 и s_2^2 со степенями свободы $f_1 = n_1 - 1$ и $f_2 = n_2 - 1$ и составляют отношение

$$F = \frac{s_1^2}{s_2^2}$$

$$(F > 1)$$

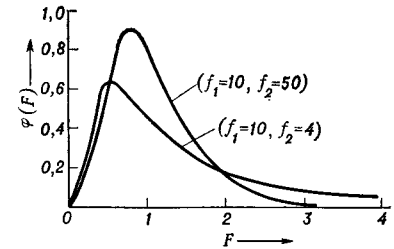


Р и с. 3.14. t -Распределение для $f = 1$ (---) и $f = 5$ (---) наряду с нормальным распределением (—).



Р и с. 3.15. Пределы интегрирования $t(P, f)$ t -распределения в зависимости от степени свободы f .

Кривая, полученная из функции распределения для всех возможных значений F , в отличие от ранее рассмотренных функций распределения носит односторонний характер; она проводится в первом квадранте между



Р и с. 3.16. F -Распределение для $(f_1 = 10; f_2 = 4)$ и $(f_1 = 10; f_2 = 50)$.

$F = 0$ и $F = \infty$ (рис. 3.16). Кривые обладают обратной симметрией при замене F на $1/F$, поскольку одновременно f_1 замещается на f_2 . При интегрировании функции распределения в пределах $0 \dots F_p$ ($F_p < \infty$) получают \bar{P} -часть всей площади под кривой. Она обозначает вероятность того, что найденное значение $F = s_1^2/s_2^2$ лежит между нулем и F_p . Эти пределы интегрирования $F(\bar{P}, f_1, f_2)$ для $\bar{P} = 0,95$ и $\bar{P} = 0,99$ в зависимости от степени свободы f_1 и f_2 даны в табл. 12.5 (ср. стр. 232). Интерполяцию не приведенных значений проводят при $f_1 = 24$ и $f_2 = 120$, задавая F как функцию $1/f$ (ср. пример [7.1]).

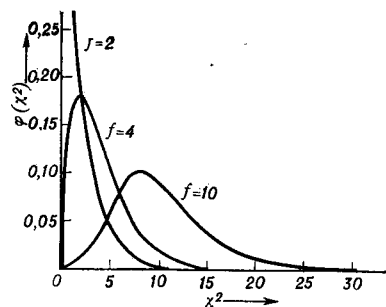
3.3.3. χ^2 -РАСПРЕДЕЛЕНИЕ

Пусть дано n независимых случайных величин x_1, x_2, \dots, x_n . Если имеет место нормальное распределение, то можно получить случайную величину

$$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 = (n-1) \frac{s^2}{\sigma^2} = \chi^2$$

с числом степеней свободы $f = n - 1$.

Функция распределения для χ^2 располагается в первом квадранте в области от $\chi^2 = 0$ до $\chi^2 = \infty$. Ее вид зависит в сильной степени от числа степеней свободы f



Р и с. 3.17. χ^2 -Распределение для $f = 2$, $f = 4$ и $f = 10$ степеней свободы.

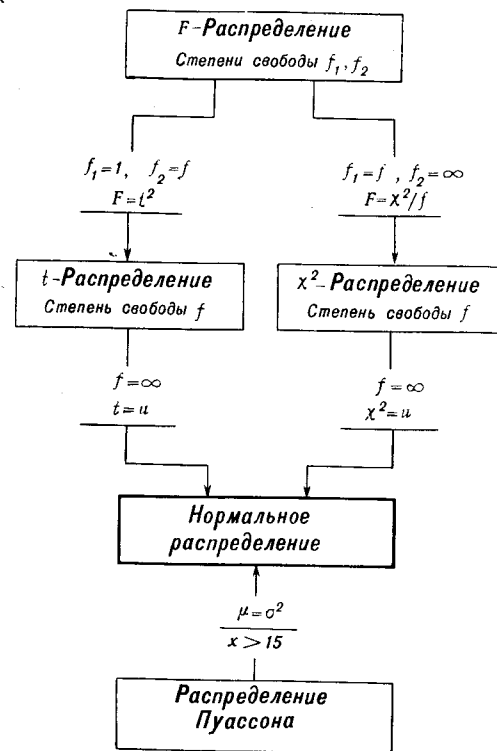
(рис. 3.17). Для малого числа степеней свободы кривая асимметрична, но с увеличением f асимметрия уменьшается, и при большом числе степеней свободы получают гауссову кривую с $\mu > 0$. Интегрирование функции распределения в пределах от 0 до χ_p^2 ($\chi_p^2 < \infty$) дает часть \bar{P} общей площади под кривой. Эта часть площади соответствует

вероятности того, что значение $\chi^2 = \sum_1^f x_i^2$, полученное из f независимых наблюдений, попадает в интервал $(0, \dots, \chi_p^2)$. Для практического использования пределы интегрирования χ^2 -распределения $\chi^2(\bar{P}, f)$ для $\bar{P} = 0,95$ и $\bar{P} = 0,99$ в зависимости от степени свободы f приведены в табл. 12.4.

3.4. Связь между отдельными распределениями

На первый взгляд может показаться, что все рассмотренные здесь теоретические распределения абсолютно различны и не взаимосвязаны. Однако ранее уже было показано, что это не так. Например, было установлено (ср. стр. 57), что распределение Пуассона может приближаться к нормальному, если выполняется условие $\bar{x} > 15$. Далее было указано, что t -распределение тоже переходит в нормальное при $f \rightarrow \infty$. Подобные связи существуют также между другими рассмотренными распределениями, что схематически приведено на рис. 3.18. Известно, что специальные распределения (t -, F - и χ^2 -распределения) получаются из нормального распределения для

определенно поставленной задачи и при ограниченном числе степеней свободы.



Р и с. 3.18. Связь между отдельными теоретическими распределениями.

ЛИТЕРАТУРА

1. Clancey V. J., Nature (Paris), 159, № 4036, 339, 340 (1947).
2. Ehrlich G., Gerbatsch R., Jaetsch K., Scholze H., Zur Genauigkeit spektralanalytischer Spurenbestimmungen, Vortrag IX Coll. Spectr. Int., Lyon, 1961.
3. Oertel A. C., Austral. J. Appl. Sci., 7, №. 2, 133, 141 (1956).
4. Schlecht H., Contribution to Geochemistry, Geol. Survey Bull. 992.

4. Закон распространения ошибок

Случайная ошибка метода анализа чаще всего состоит из нескольких частных ошибок. Чтобы общая ошибка при анализе была мала, следует найти оптимальные условия измерения. Их можно установить по законам распространения ошибок. При этом должны быть известны формулы, связывающие измеряемые величины и частные ошибки различных стадий процесса измерения. Предполагают, что все величины измерений независимы друг от друга (ср. стр. 38).

Если между n некоррелированными переменными существует функциональная связь вида

$$y = \varphi(x_1, x_2, \dots, x_n) \quad (4.1)$$

то общую ошибку σ_y приближенно можно выразить в виде

$$\sigma_y^2 = \left(\frac{\partial \varphi}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial \varphi}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{\partial \varphi}{\partial x_n}\right)^2 \sigma_{x_n}^2 \quad (4.2)$$

Суммарная ошибка из уравнения (4.2) σ_y больше, чем ошибка отдельного члена. Случайная ошибка результата увеличивается за счет суммарного действия нескольких содержащих ошибки некоррелированных частных значений. Поэтому бессмысленно давать вычисленную величину с большим числом десятичных знаков, в то время как лежащие в основе частные величины обладают только значащими десятичными знаками. Случаи, наиболее часто встречающиеся в аналитической химии, могут быть связаны с четырьмя основными видами уравнений. Если при этом рассматривать только две переменные x_1 и x_2 , то для распространения ошибок существуют следующие основные соотношения:

Расчетная формула	Общая ошибка
$\left. \begin{aligned} y &= x_1 + x_2 \\ y &= x_1 - x_2 \end{aligned} \right\}$	$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 \quad (4.3a)$
$\left. \begin{aligned} y &= x_1 \cdot x_2 \\ y &= x_1/x_2 \end{aligned} \right\}$	$\left(\frac{\sigma_y}{y}\right)^2 = \left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2 \quad (4.3b)$

При возникновении ошибок всегда суммируются соответствующие дисперсии, а именно при суммах или разностях — дисперсии абсолютных ошибок, при произведениях или частных — дисперсии относительных ошибок. Поэтому при обсуждении ошибок суммы или разности предпочитают пользоваться абсолютной ошибкой, при обсуждении ошибок произведения или частного — относительной ошибкой.

В аналитической химии особенно часто приходится иметь дело с разностью измеряемых величин (например, при взвешивании). Из уравнения (4.3a) следует, что при этом приходится пользоваться абсолютной ошибкой. Она увеличивается лишь незначительно и не зависит от величины разности. Часто полученную разность подставляют в формулы, содержащие произведение или частное от деления величин [ср. уравнение (4.4)]. В этом случае аналитик будет интересоваться относительной ошибкой разности $\sigma_{x_1-x_2}/x_1 - x_2$. Она тем больше, чем меньше разность $x_1 - x_2$; эта величина становится очень большой, когда x_1 и x_2 примерно равны. Поэтому следует по возможности избегать получения разности измеряемых величин, когда эти величины близки по своим значениям.

[4.1]. При определении хлора были получены следующие результаты измерений:

Тигель с AgCl	$X = 8,3453 \text{ г}$
Пустой тигель	$X_0 = 8,0875 \text{ г}$
AgCl	$y = X - X_0 = 0,2578 \text{ г}$

Ошибку взвешивания можно установить при обычной процедуре проведения весового анализа. Для обоих взвешиваний она одинакова и составляет $\sigma_X = 0,0002 \text{ г}$. Из уравнения (4.3a) получают абсолютную ошибку разности $\sigma_y = \sigma_X \sqrt{2} \sim 0,0003 \text{ г}$. Это ненамного больше, чем ошибка одного взвешивания. Относительная

ошибка отдельного взвешивания составляет $\sigma_x/X \sim 0,00003 \sim 0,003\%$. Разность $y = X - X_0$, напротив, оказывается связанной с заметно более высокой относительной ошибкой: $\sigma_y/y \sim 0,001$ (0,1%). Несмотря на высокую точность отдельных взвешиваний, разность можно определить только со сравнительно большой относительной ошибкой 0,1%.

При умножении содержащего ошибку числа на свободное от ошибки число (стехиометрический фактор пересчета) увеличивается только абсолютная ошибка, относительная же ошибка остается неизменной. Это также относится и к делению.

[4.2]. При гравиметрическом определении никеля диацетилглиоксимат никеля весил 0,3124 г, что соответствует $0,2250 \cdot 0,3124 = 0,0703$ г никеля. При весовой ошибке $\sigma_y = 0,0003$ г абсолютная ошибка определения никеля составляет для диацетилглиоксимата никеля 0,0003 г и соответственно $0,2250 \cdot 0,0003 = 0,00007$ г никеля; относительная ошибка составляет в обоих случаях $0,001 \sim 0,1\%$.

Закон распространения ошибок следует в дальнейшем обсудить в применении к различным методам аналитической химии*. Правда, рассмотрение ошибок этого вида при химических методах анализа дает представление только о точности измерений. Ошибки измерений всегда слишком малы по сравнению с ошибками, обусловленными нарушениями в ходе химических реакций. Несмотря на это, такое обсуждение ошибок нельзя отбросить, так как оно позволяет аналитику установить оптимальные условия проведения измерений. Кроме того, при специальном способе измерений можно получить представление о минимально возможной ошибке.

4.1. Гравиметрия

При гравиметрическом методе результат (процентное содержание вещества в исследуемой пробе) получают из следующего соотношения:

$$p, \% = \frac{100ka}{e} \quad (4.4)$$

* При создании оптимальных условий измерений часто удобнее опираться на ошибку, полученную путем вычислений [4]. Однако среднюю квадратичную ошибку точнее можно определить экспериментально. Поэтому описываемый здесь способ оценки ошибок является весьма действенным.

где p — содержание вещества в пробе; k — стехиометрический фактор пересчета; a — вес остатка; e — навеска пробы.

Итак, результат можно получить непосредственно из измерений, относящихся к анализируемой пробе, не прибегая к калибровке. Поэтому гравиметрию часто причисляют к «абсолютным методам». При этом методе анализа полученные массы a и e для веса остатка и навески получают почти всегда из измерений разности (ср. пример [4.1]). Обычно получают вес остатка доведением пробы до «постоянного веса» n_j раз. В качестве ошибки измерения при определении содержания получают

$$\frac{\sigma_p}{p} = \sqrt{\left(\frac{\sigma_e}{e}\right)^2 + \frac{1}{n_j} \left(\frac{\sigma_a}{a}\right)^2} \quad (4.5)$$

Поэтому ошибку определения содержания можно получить небольшой при малой ошибке измерений и больших значениях измеряемой величины. Уменьшение ошибки измерения ограничено аппаратными возможностями и нецелесообразностью затрат. Навеску обычно взвешивают один раз, а остаток — много раз. Поэтому рекомендуется пользоваться навеской достаточного веса и работать с малой ошибкой взвешивания (например, при работе в макромасштабе полумикровесов).

Увеличению значения измеряемой величины препятствуют главным образом причины, обусловленные методом анализа; например, в этих условиях нельзя хорошо переработать большее количество осадка. Поэтому вес остатка обычно не должен превышать 200 мг. В большинстве проводимых гравиметрических анализов и навеска, и вес остатка являются величинами одного порядка ($e \sim a$).

Ошибкой фактора пересчета k в общем случае можно пренебречь, если работают с четырьмя значащими десятичными знаками. Поэтому фактор пересчета k не фигурирует в уравнении (4.5). Несмотря на это, он косвенно влияет на общую ошибку, так как определяет величину навески, задаваемую весом остатка, ограниченного $a_{\max} \sim 200$ мг. Если навеска и вес остатка приблизительно равны, то большой фактор пересчета может благоприятно влиять на общую ошибку.

[4.3]. В магнезите (MgCO_3 , с $p \sim 25\%$) магний определяют вначале взвешиванием в форме 8-оксихинолята магния ($k_1 = 0,07780$), а затем в форме дифосфата ($k_2 = 0,2185$). Вес остатка составляет в обоих случаях $a = 200$ мг. При этом навеска в первом случае $e_1 \sim 60$ мг, во втором — $e_2 \sim 175$ мг. По уравнению (4.3б) для оксихинолятнго метода получают ошибку измерения

$$\left(\frac{\sigma_p}{p}\right)^2 = \left(\frac{0,3}{60}\right)^2 + \left(\frac{0,3}{200}\right)^2 = 0,00000273$$

$$\frac{\sigma_p}{p} = 0,0052 = 0,52\% \text{ (отн.)}$$

Для дифосфатного метода получают

$$\left(\frac{\sigma_p}{p}\right)^2 = \left(\frac{0,3}{175}\right)^2 + \left(\frac{0,3}{200}\right)^2 = 0,00000517$$

$$\frac{\sigma_p}{p} = 0,0023 = 0,23\% \text{ (отн.)}$$

Несмотря на «неблагоприятнейший» фактор пересчета, ошибка измерения во втором случае только вдвое меньше величины ошибки первого метода.

В некоторых сравнительно редких случаях вес остатка значительно меньше, чем навеска. Это имеет место при определении малых содержаний гравиметрическим анализом, например при определении фосфора в стали, докимастическом определении благородных металлов и т. д. Определяющей для ошибки содержания является в таких случаях чаще всего ошибка веса остатка (малое значение измеряемой величины). В противоположность методу, при котором навеска и остаток близки по весу, общая ошибка становится здесь относительно высокой. Хотя эта ошибка вносит малый относительный вклад при определении низких весовых содержаний, все же надо по возможности избегать применения этих методов, так как при малом количестве осадка довольно заметную роль играют загрязнения. Поэтому гравиметрию применяют как основной метод при определении средних и высоких содержаний. Гравиметрическое определение малых содержаний в большинстве случаев требует специальной техники анализа.

4.2. Объемный анализ

При объемном анализе процентное содержание исследуемой пробы получают по уравнению

$$p, \% = \frac{100kfv}{e} \quad (4.6)$$

где v — расход раствора; f — титр раствора.

При сопоставлении с уравнением (4.4) видно, что в уравнении (4.6) появляется титр раствора f . Его следует определять экспериментально, и поэтому объемный анализ в известном смысле можно рассматривать как метод с калибровкой.

Ошибку определения получают из уравнения (4.6) по закону распространения ошибок [уравнение (4.3б)]:

$$\frac{\sigma_p}{p} = \sqrt{\left(\frac{\sigma_e}{e}\right)^2 + \left(\frac{\sigma_v}{v}\right)^2 + \left(\frac{\sigma_f}{f}\right)^2} \quad (4.7)$$

В противоположность гравиметрии ошибка измерения объема состоит из нескольких частных ошибок. Наиболее существенными являются ошибки отсчета, ошибки капли, ошибки стекания. При точных измерениях также следует учитывать разницу калибровочной и рабочей температур для используемых приборов, так же как и различие плотности анализируемой пробы и титрованного вещества [5]. В общем случае можно считать ошибку измерения объема $\sigma_v \sim 0,05$ мл при работе с бюреткой на 50 мл.

Высокие значения измеряемых величин, требуемые в соответствии с равенством (4.7), ограничены возможностями используемых бюреток. При бюретке на 50 мл оптимальный расход равен 30—40 мл. Соответственно следует соразмерять величину навески e . Если приходится иметь дело с малым расходом раствора, то выбирают возможно более длинную и узкую бюретку, но титровать в этом случае следует медленно, чтобы ошибка стекания оставалась небольшой [14].

Фактор нормальности f можно найти либо установкой титра по веществу уже известного значения, либо точным взвешиванием действующего реагента. В каждом случае он получается комбинацией определения массы

и объема по формуле

$$f = \frac{e_E}{kv_E} \quad (4.8)$$

Отсюда находят ошибку «калибровки» метода

$$\frac{\sigma_f}{f} = \sqrt{\left(\frac{\sigma_{e_E}}{e_E}\right)^2 + \left(\frac{\sigma_{v_E}}{v_E}\right)^2} \quad (4.9)$$

Ошибка при установке титра должна быть мала по сравнению с двумя другими ошибками, появляющимися при анализе. Следовательно, должно иметь место неравенство

$$\frac{\sigma_f}{f} \ll \sqrt{\left(\frac{\sigma_v}{v}\right)^2 + \left(\frac{\sigma_e}{e}\right)^2} \quad (4.10)$$

Это требование выполняется, если $\sigma_f/f < 0,001$ или 0,1% (отн.) [13]. Только при этом условии объемный анализ может быть успешным.

[4.4]. При объемном определении содержания железа в красном железнике ($p \sim 90\%$ Fe_2O_3) расходуется $v \sim 30$ мл 0,1 н. раствора перманганата калия, если навеска e составляет ~ 250 мг. В предположении, что $\sigma_f/f = 0,001$, в качестве ошибки измерения для определения содержания по уравнению [4.7] получают

$$\frac{\sigma_f}{f} = \sqrt{\left(\frac{0,3}{250}\right)^2 + \left(\frac{0,05}{30}\right)^2} + 0,001^2 = 0,0023 = 0,23\% \text{ (отн.)}$$

Получающаяся здесь ошибка измерения по величине близка к ошибке гравиметрического анализа (ср. пример [4.3]).

Величина ошибки σ_f/f сильно зависит от вида и способа установления титра. Если используют заданный титр вещества, то навеска и расход объема жидкости при калибровке и анализе близки по величине и, следовательно, $e_E \sim e$ и $v_E \sim v$. Вследствие этого получают

$$\frac{\sigma_f}{f} = \sqrt{\left(\frac{\sigma_{e_E}}{e_E}\right)^2 + \left(\frac{\sigma_{v_E}}{v_E}\right)^2} \sim \sqrt{\left(\frac{\sigma_e}{e}\right)^2 + \left(\frac{\sigma_v}{v}\right)^2}$$

Таким образом, условие (4.10) не выполняется. Однако можно удовлетворить условию (4.10), если калибровку вести при более высоких требованиях к технике измерений. Так, следовало бы, например, предусмотреть по мень-

шей мере три параллельных определения ($n_j = 3$), вследствие чего ошибка уменьшилась бы на величину $1/\sqrt{n_j}$ (ср. разд. 3.1). Навеску e_E следует брать на полумикровесах, объем v_E в данном случае увеличивается при использовании больших бюреток. Также желательно использовать при титровании вещество с высоким эквивалентным весом.

[4.5]. При установке титра 0,1 н. раствора перманганата калия используют $e_E = 210$ мг оксалата натрия, а для титрования требуется $v_E \sim 30$ мл раствора. По уравнению (4.10) получают ошибку титра

$$\frac{\sigma_f}{f} = \sqrt{\left(\frac{0,3}{210}\right)^2 + \left(\frac{0,05}{30}\right)^2} = 0,0022 = 0,22\% \text{ (отн.)}$$

Требование $\sigma_f/f < 0,001$ также не выполняется. Однако, если навеска взята на полумикровесах ($\sigma_{e_E} \sim 0,1$ мг) и для титрования употребили объем $v_E = 40$ мл, то ошибка заметно уменьшается при трех параллельных определениях ($n_j = 3$). Получают

$$\frac{\sigma_f}{f} = \frac{1}{\sqrt{3}} \sqrt{\left(\frac{0,1}{280}\right)^2 + \left(\frac{0,05}{40}\right)^2} = 0,00075 = 0,075\% \text{ (отн.)}$$

При этом хорошо выполняется требование $\sigma_f/f < 0,001$.

Приведенные здесь выводы легко распространяются на другие калибровочные методы. В каждом случае необходимо стремиться к тому, чтобы ошибка при калибровке была мала по сравнению с ошибкой при последующем анализе. Это следует осуществлять чаще всего только за счет повышенных требований к технике измерений при калибровке. Для ряда растворов титр можно определить, устанавливая с высокой точностью массу действующего реагента и общий объем раствора. При достаточном больших количествах жидкости (например, 1 л) $e_E \gg e$ и $v_E \gg v$, и поэтому требование (4.10) выполняется. Такой способ установки титра также возможен, если используемый реагент приготовлен в определенной форме и с требуемой чистотой и если раствор сохраняет титр постоянным. При этом предполагается, что истинное значение не может меняться при изготовлении раствора. Реагентами, удовлетворяющими этим условиям, являются, например, бихромат калия, бромат калия или этилендиаминтетрауксусная кислота.

[4.6]. Для приготовления точно 0,1 н. раствора бихромата калия взвешивают $\epsilon_E = 4,9033$ г $K_2Cr_2O_7$ и наполняют мерную колбу до объема $v_E = 1000$ мл. При $\sigma_{v_E} \sim 0,10$ мл (табл. 4.1) получают ошибку измерения

$$\frac{\sigma_f}{f} = \sqrt{\left(\frac{0,3}{4903,3}\right)^2 + \left(\frac{0,10}{1000}\right)^2} = 0,00012 = 0,012\%$$

Максимально допустимая ошибка при установлении титра $\sigma_f/f < 0,001$ заметно больше полученной, хотя здесь не было особых затрат.

Таблица 4.1

Допуск для мерной посуды [10]

Мерные колбы					
Объем, мл	2000	1000	500	250	100
Максимально допустимая абсолютная ошибка $\pm 3\sigma$, мл	0,5	0,25	0,14	0,08	0,08
Максимально допустимая относительная ошибка $\pm 3\sigma$, %	0,025	0,025	0,028	0,032	0,080
Шплетки					
Объем, мл	100	50	25	10	2
Максимально допустимая абсолютная ошибка, $\pm 3\sigma$, мл	0,07	0,05	0,025	0,020	0,006
Максимально допустимая относительная ошибка, $\pm 3\sigma$, %	0,07	0,10	0,10	0,20	0,30

На основании всех этих рассуждений следует считать объемный анализ методом определения высоких и средних содержаний. По этой характеристике он совпадает с гравиметрией. В противоположность гравиметрии он дает, однако, результат лишь после предварительного экспериментального определения истинного значения. Только если ошибка при этой калибровке пренебрежимо мала, объемный анализ обладает достаточной точностью для характеристики исследуемого высокого содержания пробы.

4.3. Фотометрия

Основанием всех фотометрических измерений является закон Бугера — Ламберта — Бера. Если в исследуемой системе имеется только единственное светопоглощающее

вещество, то справедливо соотношение *

$$E = \epsilon(\lambda) cl \quad (4.11)$$

где E — экстинкция; $\epsilon(\lambda)$ — коэффициент экстинкции, зависящий от длины волны; c — концентрация; l — толщина стенок кюветы.

Используемые здесь значения экстинкции E вычисляют по

$$E = \ln \frac{I_0}{I} = \varphi(I_0, I). \quad (4.12)$$

При этом I_0 — интенсивность света при отсутствии пробы; I — интенсивность при наличии пробы на пути светового луча. Ошибку определения экстинкции получают по уравнению (4.2):

$$\sigma_E^2 = \left(\frac{\partial \varphi(I_0, I)}{\partial I_0}\right)^2 \sigma_{I_0}^2 + \left(\frac{\partial \varphi(I_0, I)}{\partial I}\right)^2 \sigma_I^2 = \frac{\sigma_{I_0}^2}{I_0^2} + \frac{\sigma_I^2}{I^2}$$

Отсюда при $\sigma_{I_0} = \sigma_I$ относительная ошибка

$$\frac{\sigma_E}{E} = \frac{\sigma_I}{E} \sqrt{\frac{1}{I_0^2} + \frac{1}{I^2}} \quad (4.13)$$

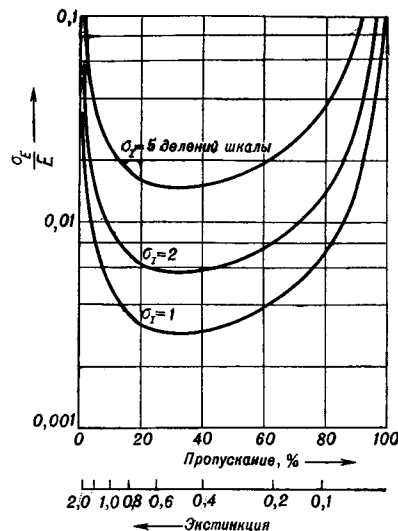
Как при гравиметрии, здесь также получают точный результат вследствие малого значения ошибки σ_I . При хорошей аппаратуре ее можно снизить до $\sigma_I = 2$ делениям шкалы (при шкале $I_0 = 1000$).

Влияние соотношения I_0/I на ошибку определения экстинкции показано на рис. 4.1 **. Из рисунка видно, что ошибка будет минимальной, если $I = 0,37 I_0$; в этом случае для $\sigma_I = 2$ делениям шкалы имеем $\sigma_E/E \sim 0,006$

* Уравнение (4.11) справедливо только для случая истинных растворов. Если поглощающее вещество находится в коллоидной форме, то первичный поток заметно ослабляется за счет рассеяния Тиндалля. При этом следует принимать в расчет радиус частичек в шестой степени [15]. Поэтому неизбежные отклонения в размерах частиц особенно сильно сказываются на результате. Отсюда следует, что при подобных методах анализа априори следует ожидать сильного разброса результатов измерений; точно такая же картина будет иметь место и в случае использования цветообразующих веществ в истинных растворах.

** При работе с аппаратурой с несколькими областями чувствительности кривая имеет смысл в каждой отдельной области.

или 0,6% (отн.). Минимум лежит в пологой части кривой, поэтому для анализа можно использовать область $0,05 I_0 < I < 0,7 I_0$, соответственно $1,3 > E > 0,2$. При этом получаются малые значения I (незначительная прозрачность) при относительно точном результате, в то время



Р и с. 4.1. Ошибка измерения при определении экстинкции в зависимости от соотношения I/I_0 при $I_0 = 1000$ делениям шкалы.

как при большой прозрачности ошибка очень резко возрастает. Поэтому фотометрический анализ всегда ненадежен у нижней границы выбранной области концентраций. Из уравнения (4.11) следует, что существует линейная зависимость между экстинкцией и концентрацией

$$c = kE \quad (4.14)$$

При этом коэффициент пропорциональности k можно найти экспериментально калибровкой, но в отличие от гравиметрии он также содержит ошибку. Его определяют, измеряя экстинкцию пробы известного содержания. Так как объектом измерения является концентрация, то особое внимание обращают на ошибки измерения экстинкции и навески, так же как и на ошибки при измерении объема (мерной колбы) и на заметные ошибки разбавления. С уче-

том всего этого для k ошибку получают в виде

$$\left(\frac{\sigma_k}{k}\right)^2 = \left(\frac{\sigma_E}{E}\right)^2 + \left(\frac{\sigma_e}{e}\right)^2 + \sum \left(\frac{\sigma_v}{v}\right)^2 \quad (4.15)$$

Ошибки экстинкции σ_E и навески σ_e можно уменьшить уже описанным способом. Для уменьшения всех ошибок, связанных с измерением объема, рекомендуется использовать большие мерные колбы и большие пипетки (табл. 4.1). Калибровку проводят m раз независимо друг от друга и получают отдельные значения. Вследствие этого ошибка уменьшается на величину $1/\sqrt{m}$ *.

[4.7]. Для калибровки фотометрического определения меди берут навеску 2,000 г $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ и мерную колбу объемом $v_0 = 1$ л. Из этого раствора пипеткой отбирают $v_1 = 5, 10, 20$ и 50 мл, добавляют цветообразующий реагент ($v_2 = 100$ мл) и измеряют экстинкцию. Для калибровочной пробы $v_1 = 10$ мл раствора ошибку калибровки следует оценивать из уравнения (4.15). Ошибку объема σ_v берут из табл. 4.1 и получают

$$\begin{aligned} \left(\frac{\sigma_k}{k}\right)^2 &= 0,006^2 + \left(\frac{0,3}{2000}\right)^2 + \left(\frac{0,08}{1000}\right)^2 + \left(\frac{0,007}{10}\right)^2 + \left(\frac{0,027}{100}\right)^2 \\ &= 0,006^2 + 0,00015^2 + 0,00008^2 + 0,0007^2 + 0,00027^2 \end{aligned}$$

Находят, что на ошибку калибровки σ_k/k значительно влияет ошибка измерения экстинкции. При использовании четырех калибровочных проб ($m = 4$) получается

$$\frac{\sigma_k}{k} \sim \frac{\sigma_E}{E \sqrt{4}} = \frac{\sigma_E}{2E}$$

Из уравнения (4.14) и закона распространения ошибок [уравнение (4.36)] следует:

$$\left(\frac{\sigma_c}{c}\right)^2 = \left(\frac{\sigma_E}{E}\right)^2 + \left(\frac{\sigma_E}{2E}\right)^2$$

$$\frac{\sigma_c}{c} = \frac{\sigma_E}{2E} \sqrt{5}$$

Для измерения в оптимальной области ($1,3 > E > 0,2$) $\sigma_c/c \sim 0,7\%$ (отн.)

Фотометрический анализ часто применяют для определения крайне низких концентраций. При этом часто мешают загрязнения использованных реагентов. Тогда при достаточной чувствительности прибора вместо кон-

* О калибровке при помощи вычисления регрессии см. гл. 9.

центрации $c = 0$ измеряют значения экстинкции, отличной от нуля $E > 0$, т. е. получают значение *холостого опыта*. Если исследуют большее число проб с концентрацией $c = 0$, то получают статистически различные значения холостого опыта. Поэтому значение холостого опыта принимают за случайную величину. Генеральная совокупность всех возможных значений холостого опыта характеризуется двумя параметрами — средним значением μ_B и средней квадратичной ошибкой σ_B . Все значения экстинкции, полученные фотометрическим методом, можно только тогда считать реальными величинами, когда они лежат выше уровня значений холостого опыта. Это относится абсолютно ко всем методам анализа [12]. Значение измерения x_u может отличаться от уровня значений холостого опыта с вероятностью \bar{P} только тогда, когда

$$x_u \geq \mu_B + u(\bar{P}) \sigma_B \quad (4.16)$$

Определяемая из уравнения (4.16) величина называется *границей обнаружения*. Она указывает, какое наименьшее измеренное значение следует еще считать надежным и является поэтому важнейшей характеристикой эффективности метода анализа. Определению σ_B следует уделить особое внимание [12].

Для экспериментального определения значения холостого опыта (например, «холостой» экстинкции) в литературе указан ряд приемов [1, 12].

1. «Холостую» экстинкцию получают из очень большого числа холостых опытов. При каждом анализе принимают во внимание эту общую «холостую» экстинкцию. Границу обнаружения получают тогда в соответствии с уравнением (4.16). Если при анализе проводят n_A параллельных определений, то приближенно получается (если $n_B \gg n_A$), что

$$\bar{x}_u = \mu_B + \frac{u(\bar{P}) \sigma_B}{\sqrt{n_A}} \quad (4.16a)$$

Увеличением числа параллельных определений при анализе можно улучшить границу обнаружения.

2. Для каждой пробы измеряют соответствующее специальное значение холостого опыта. Его вычитают из

результата измерения, полученного для пробы. Если значения холостого опыта и значения измерения анализируемой пробы получены из n_j параллельных определений, то из уравнений (4.3а) и (3.3) находят границу обнаружения

$$\bar{x}_u = \mu_B + u(\bar{P}) \sigma \sqrt{\frac{2}{n_j}} \quad (4.16б)$$

Здесь, используя среднее значение, повышают границу обнаружения.

3. Для проб с различной, но известной концентрацией c определяют экстинкцию E и графически или расчетным путем получают функцию $E = \varphi(c)$. Отрезок ординаты E_0 для концентрации $c_0 = 0$ представляет собой искомое значение холостого опыта. Определение границы обнаружения производят по уравнению регрессии (ср. гл. 9), получая результаты, аналогичные уравнению (4.16).

По первому из названных выше методов анализа результаты холостого опыта должны иметь близкие значения при переходе от пробы к пробе. Разброс значений холостого опыта должен соответствовать случайной ошибке метода анализа вблизи границы обнаружения. При выполнении этих условий можно особенно точно определить значение холостого опыта первым из описанных выше методов, так как тогда может быть использовано очень большое число анализов. (Для практических целей проводят $n_B \gg 25$ анализов.) Второй способ применим всегда, так как каждое значение анализа сочетается со «своим» значением холостого опыта. Недостатком в этом случае является то, что значение холостого опыта получают из сравнительно малого числа определений, что часто оказывается ненадежным. Поэтому эти методы используют только тогда, когда значения холостого опыта столь сильно разбросаны при переходе от пробы к пробе, что нельзя дать общего значения холостого опыта, так как это делается при первом способе. Как способ 1, так и способ 2 дают границу обнаружения в размерности измеряемой величины (например, экстинкции) *.

* Для получения результатов в единицах концентрации используют уравнение регрессии. Подробности см. в гл. 9.

В противоположность этому при использовании уравнения регрессии (способ 3) получают границу обнаружения непосредственно в единицах концентрации. При этом принимают во внимание все неопределенности, появляющиеся при калибровке. Кроме того, здесь нет необходимости пользоваться крайне ненадежным определением малых величин. Правда, этот способ получения границы обнаружения допустим только тогда, когда пробы лежат вблизи концентрации $c_0 = 0$.

Приведенная в примере [4.7] ошибка измерения для фотометрии больше, чем ошибка измерения при гравиметрическом или объемном методах (ср. пример [4.3] или [4.4]). Поэтому фотометрию применяют главным образом для определения малых содержаний, так как в этой области большая ошибка не имеет такого значения, как при анализе высоких содержаний. Фотометрический анализ тем чувствительнее, чем сильнее окрашены соответствующие соединения. Отсюда, однако, не следует, что надо отвергать попытки использовать фотометрию для определения средних и высоких содержаний, особенно там, где другие методы требуют значительных затрат (например, разделения) и потому могут быть ненадежными. Правда, при фотометрическом определении основных составных частей необходима специальная техника анализа.

4.4. Косвенные методы

При всех косвенных методах искомую величину измерения получают лишь из комбинации нескольких измеренных значений. Известны косвенные методы для всех широко распространенных методов определений — гравиметрии, объемного анализа, фотометрии и т. д. В простейшем случае искомую величину получают из разности двух значений анализа (например, метод обратного титрования). Если нужно определить совместно два или несколько элементов без разделения, то в смеси из n компонентов проводят n различных анализов. Из значений измерений составляют систему с n неизвестными, решение которой дает искомое содержание отдельных компонентов.

Особенно часто используют методы с обратным титрованием избытка раствора. Содержание, полученное обрат-

ным титрованием, находят по уравнению

$$p, \% = \frac{100(f_1 k_1 v_1 - f_2 k_2 v_2)}{e} \quad (4.17)$$

При этом величины v_1 и v_2 получают из разности двух показаний объема. Следует считать, что оба фактора пересчета примерно равны по величине ($k_1 \sim k_2$) и что $f_1 = f_2 = 1,000$. Тогда по уравнению (4.3а) для ошибки разницы объема $\Delta v = v_1 - v_2$ получают

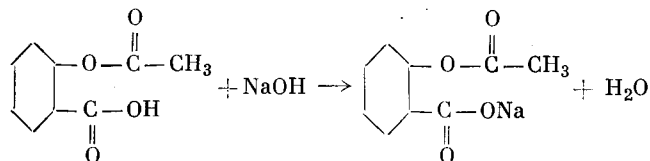
$$\sigma_{\Delta v} = \sqrt{\sigma_{v_1}^2 + \sigma_{v_2}^2} = \sigma_v \sqrt{2}$$

Если ошибка установки титра мала по сравнению с ошибкой определения объема и навески, то для ошибки определения содержания имеем

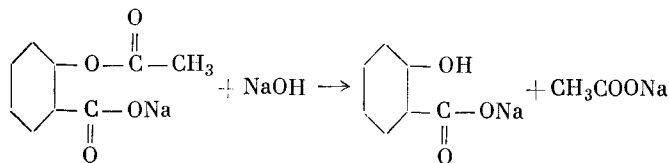
$$\frac{\sigma_p}{p} = \sqrt{\left(\frac{\sigma_e}{e}\right)^2 + 2\left(\frac{\sigma_v}{\Delta v}\right)^2} \quad (4.18)$$

Общая ошибка незначительно возрастает по сравнению с ошибкой прямых измерений, так как ошибка объема умножается на величину $\sqrt{2}$. Чтобы сохранить малую ошибку, необходимо стремиться к возможно большей разности объемов Δv .

[4.8]. При объемно-аналитическом определении содержания ацетилсалициловой кислоты в препарате [16] прежде всего нейтрализуют щелочью ацетилсалициловую кислоту



Затем добавляют избыток щелочи ($= v_1$) и омыляют при кипячении в соответствии с уравнением



Избыток щелочи оттитровывают кислотой ($= v_2$). При взятой навеске $e \sim 400$ мг и при содержании ацетилсалициловой кислоты

литические данные (искомое содержание пробы) находят, считая дискретные величины (например, импульсы). Этот подсчет повторяют несколько раз, и из отдельных числовых значений находят среднее значение \bar{x} . Чтобы эти величины можно было сравнить между собой, их чаще всего относят к единице времени, например к минуте. Если во время T минут измерили всего x импульсов, то для частоты импульсов получают

$$v = \frac{\bar{x}}{T} \text{ имп./мин} \quad (4.20)$$

Для определения частоты импульсов можно либо получить число импульсов в заданное время, либо установить число импульсов и затем измерить необходимое время. Первый метод называется *выбором постоянного времени*, второй — *выбором постоянного числа импульсов*.

Последовательность дискретных величин чаще всего следует распределению Пуассона (ср. разд. 3.2). При помощи соотношения $\sigma \sim \sqrt{\bar{x}}$ [уравнение (3.12)] можно легко оценить относительную ошибку подсчитанных событий. Находят

$$\frac{\sigma}{\bar{x}} = \frac{\sqrt{\bar{x}}}{\bar{x}} = \frac{1}{\sqrt{\bar{x}}} \quad (4.21)$$

[4.9]. Число импульсов радиоактивного препарата следует вычислить с точностью $0,01 = 1\%$ (отн.). Для среднего значения \bar{x} , полученного из серии измерений по уравнению (4.21), получают

$$\frac{\sigma}{\bar{x}} = 0,01 = \frac{1}{\sqrt{\bar{x}}}$$

Отсюда следует

$$\bar{x} = 10\,000$$

Следовательно, необходимо по меньшей мере 10^4 импульсов, чтобы добиться требуемой точности 1% (отн.).

Необходимую в уравнении (4.20) частоту импульсов v получают из разности частот для препарата плюс под-

ложка и для подложки, и следовательно,

$$v = v_P - v_U \quad (4.22)$$

где v_P — частота импульсов препарата с подложкой; v_U — частота импульсов подложки.

Если обе величины v_P и v_U определяют в заданное время T (выбор времени), то

$$v = \frac{\bar{x}_P}{T} - \frac{\bar{x}_U}{T} = \frac{\bar{x}}{T} \quad (4.23)$$

Так как оба значения получены за одно и то же время счета T , число импульсов можно использовать непосредственно как измеряемую величину. Уравнение (4.22) переходит в

$$\bar{x} = \bar{x}_P - \bar{x}_U \quad (4.24)$$

где \bar{x}_P — число импульсов препарата с подложкой; \bar{x}_U — число импульсов подложки.

Из уравнения (4.24) по закону распространения ошибок и с учетом $\sigma \sim \sqrt{\bar{x}}$ [уравнение (3.12)] следует

$$\sigma_{\bar{x}} = \sqrt{\sigma_P^2 + \sigma_U^2} = \sqrt{\bar{x}_P + \bar{x}_U} \quad (4.25)$$

Частоту импульсов v_U можно задать как дробную часть величины v_P :

$$v_U = \xi v_P \quad (\xi < 1) \quad (4.26)$$

Если выбран метод постоянного времени, то имеет место

$$\bar{x}_U = \xi \bar{x}_P$$

Это соотношение подставляют в уравнение (4.25) и получают относительную ошибку для определения числа импульсов метода постоянного времени

$$\frac{\sigma_{\bar{x}}}{\bar{x}} = \frac{\sqrt{(1+\xi)}}{\sqrt{\bar{x}_P(1-\xi)}} \quad (4.27)$$

При выборе числа импульсов задано $\bar{x}_P = \bar{x}_U = \bar{x}^*$. Для получения этого числа импульсов измеряют необходимое время T_P и T_U и, пользуясь уравнением (4.20), находят

частоту импульсов. Уравнение (4.22) приводит к

$$v = v_P - v_U = \frac{\bar{x}_P}{T_P} - \frac{\bar{x}_U}{T_U} = \bar{x}^* \left(\frac{1}{T_P} - \frac{1}{T_U} \right) \quad (4.28)$$

В уравнении (4.28) $\frac{1}{T_P} \neq \frac{1}{T_U}$, поэтому расчет ошибки σ_v ведется по уравнению (4.2). Зная, что $\sigma = \sqrt{\bar{x}}$, получают

$$\sigma_v = \sqrt{\frac{\sigma_P^2}{T_P^2} + \frac{\sigma_U^2}{T_U^2}} = \sqrt{\frac{\bar{x}_P}{T_P^2} + \frac{\bar{x}_U}{T_U^2}} \quad (4.29)$$

Снова выражают частоту импульсов v_U , как дробную часть v_P [аналогично уравнению (4.26)]. Если выбран метод постоянного числа импульсов, то имеет место соотношение

$$\frac{1}{T_U} = \xi \frac{1}{T_P}$$

Подставляя это в уравнение (4.29), при $\bar{x}_P = \bar{x}_U = \bar{x}^*$ находим относительную ошибку для частоты импульсов при методе постоянного числа импульсов

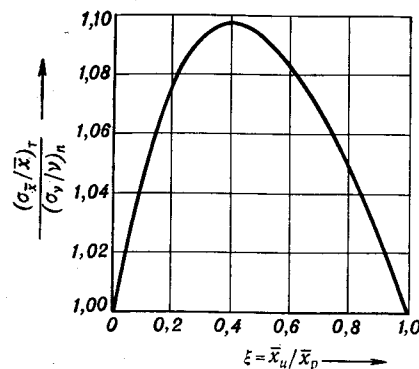
$$\frac{\sigma_v}{v} = \frac{\sqrt{(1+\xi^2)}}{\sqrt{\bar{x}^*(1-\xi)}} \quad (4.30)$$

Для сравнения относительных ошибок метода постоянного времени и метода постоянного числа импульсов находят их отношения, используя уравнения (4.27) и (4.30). Если обозначить ошибку метода постоянного времени индексом T , а ошибку метода постоянного числа импульсов индексом n , то при $(\bar{x}_P)_T = (\bar{x}^*)_n$ получим

$$\frac{(\sigma_{\bar{x}}/\bar{x})_T}{(\sigma_v/v)_n} = \sqrt{\frac{1+\xi}{1+\xi^2}} \quad (4.31)$$

Это частное определяется только соотношением числа импульсов, полученных для препарата с подложкой и для одной подложки. Ошибка метода постоянного времени почти всегда больше, чем ошибка метода постоянного числа импульсов, если в обоих методах подсчитывается примерно одинаковое число импульсов как при измерении активности подложек, так и активности препаратов с под-

ложками. За экономию времени (более быстрый счет импульсов подложки) приходится расплачиваться: увеличивается случайная ошибка. Графическое представление уравнения (4.31) показывает, что повышенная случайная ошибка (рис. 4.2) метода постоянного времени зависит от



Р и с. 4.2. Отношение ошибок двух методов измерения (постоянного времени и постоянного числа импульсов) в зависимости от соотношения числа импульсов для подложки и препарата [3].

соотношения измеренных частот импульсов v_P и v_U . В самом неблагоприятном случае ($\xi \sim 0,4$) при $(\bar{x}_P)_T = (\bar{x}^*)_n$ метод постоянного времени дает случайную ошибку примерно на 10% (отн.) больше, чем метод постоянного числа импульсов.

При количественном анализе предполагают, что существует зависимость между частотой импульсов v и концентрацией пробы c . При калибровке по пробе известной концентрации находят содержание анализируемой пробы в соответствии с

$$c = \frac{c_E v_A}{v_E} \quad (4.32)$$

где v_A , v_E — частоты импульсов анализируемой и калибровочной проб по уравнению (4.22); c_E — содержание калибровочной пробы.

Если для подсчета импульсов используют метод постоянного времени, то одновременно имеет место

$$c = \frac{c_E \bar{x}_A}{\bar{x}_E} \quad (4.33)$$

Закон распространения ошибок¹ дает

$$\frac{\sigma_c}{c} = \sqrt{\left(\frac{\sigma_{cE}}{cE}\right)^2 + \left(\frac{\sigma_{x_A}}{x_A}\right)^2 + \left(\frac{\sigma_{x_E}}{x_E}\right)^2}$$

В предположении, что $\bar{x}_A \sim \bar{x}_E = \bar{x}$ и при помощи уравнения (3.12) получается

$$\frac{\sigma_c}{c} = \sqrt{\left(\frac{\sigma_{cE}^2}{cE}\right)^2 + \frac{1}{x_E} + \frac{1}{x_A}} \quad (4.34)$$

Общая ошибка определения концентрации состоит из ошибки в оценке калибровочной концентрации и ошибки при вычислении \bar{x}_A и \bar{x}_E . Общую ошибку можно уменьшить точной калибровкой и высокими значениями отсчитываемых величин. Однако нецелесообразно стремиться к уменьшению общей ошибки только высокими значениями счета, если одновременно не уменьшить ошибку при определении калибровочной концентрации c_E .

[4.10]. Для методических исследований явлений соосаждения [18] сульфат бария осаждали в присутствии радиоактивного железа. При этом пользовались раствором, который состоял из 22,9 мг $^{59}\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ (= 4,73 мг $^{59}\text{Fe}^{3+}$), взвешенного на микровесах и растворенного в 50 мл. Частота импульсов, полученная из пяти отдельных измерений для этого раствора, $\nu_E = 122\,004$ имп/мин. После осаждения в том же объеме раствора была измерена частота импульсов $\nu_A = 103\,410$ имп/мин. В соответствии с уравнением (4.32) оставшаяся концентрация железа была равна

$$c = \frac{4,73 \cdot 103\,410}{122\,004} = 4,01 \text{ мг } ^{59}\text{Fe}^{3+}/50 \text{ мл}$$

При вычислении ошибки по (4.34) следовало иметь в виду, что величина c_E получается из определений массы и объема. Поэтому

$$\begin{aligned} \frac{\sigma_c}{c} &= \sqrt{\left(\frac{0,03}{22,9}\right)^2 + \left(\frac{0,04}{50}\right)^2 + \frac{1}{122\,004} + \frac{1}{103\,410}} = \\ &= \sqrt{1,71 \cdot 10^{-6} + 6,4 \cdot 10^{-7} + 8,2 \cdot 10^{-6} + 9,8 \cdot 10^{-6}} = \\ &= 0,0045 = 0,45\% \text{ (отн.)} \end{aligned}$$

Подходящим выбором условий удалось достигнуть того, что относительная ошибка при каждом измерении имела примерно одинаковую величину. Общая ошибка лежит в тех границах для случайной ошибки, которые обычно имеют место в методах, нуждающихся в калибровке.

Счетные методы анализа, особенно методы радиохимии, имеют широкое применение. Их можно использовать как

для определения малых и очень малых содержаний, так и для анализа основных составных частей. Эти широкие возможности применения требуют, чтобы ошибка измерения для этого метода подробно обсуждалась в каждом случае для нахождения соответствующих оптимальных условий измерения.

4.6. Отбор проб

Целью отбора и приготовления проб является отбор малого количества вещества из очень обширного материала, подлежащего анализу. Проба только тогда взята правильно, когда она представляет весь материал, подлежащий проверке. При количественном анализе это значит, что процентный состав взятого материала должен быть идентичен процентному составу анализируемой пробы. В то время как аналитик проявляет большое внимание при проведении анализов, безусловный отбор проб иногда осуществляется недостаточно четко. Анализы в таких случаях оказываются заведомо ложными. Поэтому необходимо исследовать, какие общие закономерности надо соблюдать, чтобы аналитик получал пробу, соответствующую действительному содержанию вещества.

Из жидкостей или газов отбор проб в общем прост. Случайные гравитационные нарушения можно легко учесть. Сложнее дело обстоит с отбором проб твердых веществ. Выбранный для анализа материал чаще всего существует в форме гетерогенной смеси разнообразных компонентов. Если отказаться от особых случаев, как, например, исследование месторождений, то интересующие нас компоненты распределены в общем объеме пробы случайно. Вследствие зернистости материала при отборе пробы рискуют получить слишком большую или слишком малую часть того или иного компонента. Поэтому при повторяющемся отборе отдельные пробы оказываются несколько различного состава. Вызываемую этим ошибку отбора пробы σ_P можно оценить при известных идеальных условиях. Для смеси двух компонентов, например руды и жильной породы, согласно [2], справедливо уравнение

$$\frac{\sigma_P}{\bar{x}} = \frac{q d_1 d_2}{100 \bar{d}^2} \sqrt{\frac{a^3}{e \bar{x}}} (100 d_1 - \bar{x} d) \quad (4.35)$$

где \bar{x} — среднее содержание руды в смеси (%); d_1, d_2 — плотность руды и жильной породы соответственно; \bar{d} — плотность пробы; q — содержание металла в чистой руде (%); e — использованное для анализа количество вещества (g); \bar{a} — средняя длина ребра частичек (см).

При составлении этого уравнения принимают, что все частицы пробы имеют равный объем. Поскольку это не так, следует брать объем частиц \bar{a}^3 такой, чтобы масса всех частиц $< \bar{a}$ составляла $\sim 75\%$. По уравнению (4.35) ошибка отбора пробы растет с увеличением размера частиц и с увеличением содержания металла в чистой руде, а также с уменьшением содержания руды и с уменьшением используемого количества пробы. Если ошибка отбора пробы должна лежать в заданных пределах, то пробу следует брать тем большего объема, чем более грубозернистым является материал.

Кроме уравнения (4.35), приведенного в работе [2], в литературе имеются указания на подобные опыты по оценке ошибки отбора проб. Особенно достойна упоминания диаграмма отбора проб, данная Ги [9]. Она позволяет рассчитать следующие величины:

1) необходимый минимальный вес пробы, если дан диаметр больших частиц руды и если задана ошибка отбора пробы;

2) необходимый размер зерен, до которого руда должна быть в дальнейшем измельчена, чтобы не была превышена допустимая ошибка отбора пробы;

3) относительная ошибка, с которой следует считаться при отборе пробы, если известны количество пробы и наибольший диаметр зерен данного образца руды.

Подход к решению этих задач основан на различных очень сильно идеализированных предположениях о содержании зерен и распределении величин зерен. Поэтому в общем они не приводят к согласованным значениям для вычисления ошибки отбора пробы. В связи с этим целесообразно рекомендовать обратный путь, а именно производить отбор исходя из практического опыта и статистически контролировать результаты такого отбора проб.

Ошибку отбора пробы σ_p находят при помощи дисперсионного анализа (см. гл. 8), пользуясь специально постав-

ленными для этого экспериментами [5]. По предложению Томлинсона [17] отбор пробы можно считать безукоризненным, если ошибка отбора составляет примерно четыре пятых общей ошибки.

Отобранную пробу обрабатывают в соответствии с требованиями, предъявляемыми при анализе. Для этого зернистость и объем пробы уменьшают измельчением и последующим делением общего объема пробы на части. При этом остающиеся количества пробы на отдельных стадиях измельчения должны соотноситься, как кубы размеров зерен (размеры отверстий сита). Чем меньше отобранный объем пробы, тем тоньше должно быть растерто вещество. Наконец, из конечной пробы берут навеску того или иного веса для анализа. При этом возникает та же проблема отбора пробы из первоначально имеющегося материала. Зернение анализируемой пробы должно быть поэтому достаточно тонким, чтобы гарантировать правильность пробы. Это требование является особенно критическим при всех методах, которые имеют дело с малыми навесками твердых проб (например, микроанализ, спектральный анализ и т. д.). Неблагоприятно для анализа, если оба компонента имеют очень разную плотность или если один из компонентов значительно преобладает.

[4.11]. Альтенбергскую оловянную руду можно принять за смесь жильной породы и руды в форме касситерита. Из уравнения (4.35) следует, что эта руда не очень пригодна для спектрохимического анализа с малой навеской. Разница в плотности между рудой ($d_1 = 6,9 \text{ г/см}^3$) и жильной породой ($d_2 = 2,7 \text{ г/см}^3$) относительно велика, а среднее содержание руды ($\bar{x} = 0,4\%$) и использованное количество вещества ($e \sim 0,005 \text{ г}$) оказываются низкими. В соответствии с этим ошибку отбора пробы по уравнению (4.35) можно сохранить в требуемых пределах только тогда, когда пробу тонко измельчают. Как следует из рис. 4.3, средняя длина ребра зерен должна быть равна примерно $0,0005 \text{ см}$, если порядок величины ошибки отбора пробы должен лежать около 1% . Выполнить это требование в данном случае не просто, так как касситерит тверд и в сравнении с породой мало хрупок.

Следующее из уравнения (4.37) требование согласуется с такими методами отбора проб, в которых для неомогенных материалов предпочитают иметь дело с множеством малых проб, а не с малым числом больших. При этом отдельные пробы должны иметь один и тот же объем,

Вопросы правильности отбора проб играют более или менее важную роль при всех методах анализа, при которых

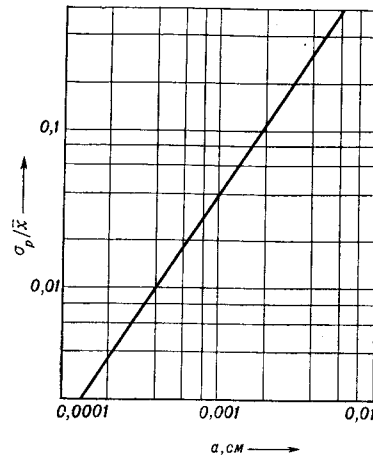


Рис. 4.3. Ошибка отбора пробы при спектрохимическом анализе альтенбергской оловянной руды в зависимости от длины ребра зерна.

твердые пробы годятся непосредственно для исследования. Сильно негетогенные пробы (например, залежи металлов) можно надежно анализировать только тогда, когда раство-

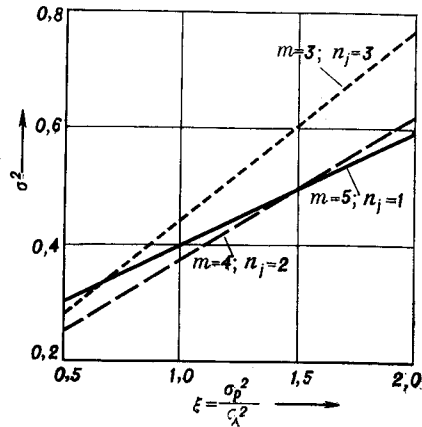


Рис. 4.4. Влияние ошибки отбора пробы и ошибки анализа на общую ошибку [5].

ряют достаточно большую навеску и исследуют гомогенный раствор.

Для каждого метода анализа важно знать суммарную общую ошибку, складывающуюся из ошибки отбора пробы и ошибки анализа. Особенно рекомендуется исследовать, какой метод анализа необходимо применить, чтобы общую ошибку σ свести к возможно меньшей.

Если берут m проб одного и того же вещества и каждую из этих проб анализируют n_j раз, то дисперсия для полученного среднего значения будет

$$\sigma^2 = \frac{\sigma_P^2}{m} + \frac{\sigma_A^2}{mn_j} \quad (4.36)$$

где σ_P — ошибка отбора пробы; σ_A — ошибка анализа.

Ошибку отбора пробы можно выразить как кратное ошибки анализа. Тогда

$$\sigma_P^2 = \xi \sigma_A^2$$

Из этого получают

$$\sigma^2 = \frac{\sigma_A^2}{m} \xi + \frac{\sigma_A^2}{mn_j} = A\xi + B \quad (4.37)$$

Общая дисперсия σ^2 также линейно зависит от отношения частных дисперсий σ_P^2/σ_A^2 . При заданном числе анализов $n = mn_j$ величина σ^2 мала, если m сделать возможно большим. Другими словами, это значит, что следует брать возможно большее число проб. Чтобы при этом издержки производства сохранять в приемлемых пределах, можно сократить число параллельных определений пробы.

[4.12]. Вместо того чтобы провести по три определения трех проб одного и того же вещества ($m = 3$; $n_j = 3$; $n = 9$), целесообразнее взять четыре пробы и анализировать каждую только по два раза ($m = 4$; $n_j = 2$; $n = 8$). Из рис. 4.4 следует, что, несмотря на сокращение работ, получают меньшую случайную ошибку. При одинаковой величине ошибок отбора проб и ошибок анализа можно уменьшить общую ошибку анализа и при $n = 5$, взяв пять проб ($m = 5$) и анализируя каждую из них только по одному разу. Несмотря на значительно меньшую затрату времени, воспроизводимость не ухудшается. Правда, следует учитывать, что при однократном анализе какой-либо пробы появляется опасность не выявленной грубой ошибки.

Все показанные здесь общие закономерности применяют в каждом частном случае с учетом поставленной зада-

чи. Нужно иметь в виду природу исследуемого вещества и требуемую информацию. Так, надо различать, хотя бы проверить с помощью анализа атмосферу рабочего места или в целях контроля качества отбирают пробы из газового потока. Только учет общих и частных закономерностей приведет к тому, что результат анализа исследуемого вещества, полученный с затратой большого труда, будет отражать действительное положение вещей.

ЛИТЕРАТУРА

1. Agterdenbos J., Z. anal. Chem., 157, 161 (1957).
2. Baule B., Benedetti-Pichler A., Z. anal. Chem., 74, 442 (1928).
3. Birks L. S., Brown D. M., Anal. Chem., 34, 240 (1962).
4. Doerffel K., Z. anal. Chem., 157, 197 (1957).
5. Doerffel K., Z. anal. Chem., 185, 1 (1962).
6. Erdey L., Svehla G., Z. anal. Chem., 176, 8 (1960).
7. Fuchs P., Angew. Chem., 54, 512 (1941).
8. Görlich P., Z. anal. Chem., 179, 266 (1961).
9. Gy P., Erzmetall, 9, 237 (1956).
10. Jander G., Jahr W., Maßanalyse, Verlag Walter de Gruyter and Co.
11. Johnson R. A., Graham C. B., Jr., Anal. chim. Acta, 12, 408 (1955).
12. Kaiser H., Z. anal. Chem., 209, 1 (1965).
13. Кольтгоф И. М., Стенгер В. А., Объемный анализ, Госхимиздат, т. I, 1950, т. II, 1952.
14. Linder L., Haslwantner F., Angew. Chem., 42, 821 (1949).
15. Lipatow S. M., Physikalische Chemie der Kolloide, Berlin, Akademie-Verlag, 1953.
16. Medicus M., Poethge W., Maßanalyse, Dresden, Verlag Th. Steinkopff, 1952.
17. Tomlison R. S., Sampling in Wilson C. L., Wilson D. W., Comprehensive Analytical Chemistry, Elsevier Publishing Comp. 1960.
18. Winzer A., Diss. Technik Hochschule für Chemie, Merseburg, 1966.
19. Zöllner H., Ber. dtsch. keram. Ges., 28, 492 (1951).

5. Случайные ошибки методов анализа

Для решения специальных задач аналитик должен выбрать подходящий метод анализа. Наряду с оценками затрат времени, необходимого оборудования, стоимости и т. д. часто решающую роль играет вопрос о случайной ошибке метода. Опытный аналитик может обычно качественно охарактеризовать эффективность выбранных методов. Он, например, знает, что при объемно-аналитическом определении цинка очень часто появляется большая ошибка, если титрование заканчивают обычными методами, а не потенциометрически. Однако точные данные о случайной ошибке не может дать ни большой опыт, ни общая оценка метода.

Описанные ранее границы разброса, особенно средняя квадратичная ошибка (ср. разд. 2.2.4), позволяют найти определенную характеристику случайной ошибки метода анализа. Поэтому необходимо исследовать, как можно высчитать эти величины в специальных условиях аналитической химии (малое число параллельных определений проб различного содержания). Далее, интересно обсудить вопрос об устойчивости полученных данных, о возможности их обобщения и рассмотреть условия, при которых должны производиться измерения.

5.1. Вычисление средней квадратичной ошибки

Для вычисления средней квадратичной ошибки необходим набор экспериментально полученных значений. Следует предположить, что на них влияет только случайная ошибка метода анализа и что не имеет места негомогенность проб и пренебрежимо малы случайные помехи, обусловленные особенностями работы лаборатории и другими факторами*. Тогда разброс значений внутри распределе-

* Учет негомогенности проб возможен при помощи простого дисперсионного анализа (ср. гл. 8). Влияние особенностей работы лабораторий и лаборантов (персональный фактор) можно определить по Морану [5], используя предложенную им детальную схему эксперимента (ср. также [2]).

ния частот определяется только случайной ошибкой метода анализа; ее можно характеризовать, задавая параметр σ — среднюю квадратичную ошибку. На практике аналитик никогда не располагает необходимым числом измерений. Поэтому вместо средней квадратичной ошибки σ он получает только оценку s . Расчет средней квадратичной ошибки по уравнению (2.5) чаще всего приводит к затруднениям, так как обычно для одной пробы редко проводят больше трех параллельных определений. Однако имеется возможность использовать результаты многократного анализа проб различного содержания. Из их частных средних квадратичных ошибок s_j путем усреднения вычисляют общую среднюю квадратичную ошибку s . Если имеется m проб и если для каждой пробы проводится n_j параллельных определений, то используют следующую схему:

Номер пробы	Номер измерения			
	1	2	...	i ... n_j
1	x_{11}	x_{12}	...	x_{1i} ...
2	x_{21}	x_{22}	...	x_{2i} ...
.
.
j	x_{j1}	x_{j2}	...	x_{ji} ...
.
.
m

Средняя квадратичная ошибка s (соответственно дисперсия s^2) получается из уравнения

$$s^2 = \frac{s_1^2(n_1-1) + s_2^2(n_2-1) + \dots + s_m^2(n_m-1)}{(n_1-1) + (n_2-1) + \dots + (n_m-1)} = \frac{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2 + \dots + \sum (x_{ji} - \bar{x}_j)^2}{n-m} \quad (5.1)$$

$$s = \sqrt{\frac{\sum_1^m \sum_1^{n_j} (x_{ji} - \bar{x}_j)^2}{n-m}}$$

со степенями свободы $f = n - m$.

В качестве оценки параметра σ всегда используют только среднюю квадратичную ошибку s . Здесь предполагают, что все s_j не должны значимо различаться (возможность проверки см. в разд. 7.2) и что средняя квадратичная ошибка чаще всего не зависит или зависит очень слабо от величины измеряемого значения (ср. стр. 109).

[5.1]. Содержание марганца в пяти разных пробах стали было определено по способу Простера и Смита. Из результатов получена средняя квадратичная ошибка метода. При вычислениях используют описанное выше преобразование (ср. пример [2.7]). В данном случае значение случайной ошибки вычисляют для каждой пробы отдельно, поэтому для разных проб можно использовать различные преобразования. При преобразовании следует только сохранять один и тот же порядок величин. Имеем

Значения анализа, % Mn				Преобразование		Преобразованные значения			
0,31	0,30	0,29	0,32	$X_{1i} = 100$	$x_{1i} - 30$	+1	0	-1	+2
0,59	0,57	0,58	0,57	$X_{2i} = 100$	$x_{2i} - 58$	+1	-1	0	-1
0,71	0,69	0,71	0,71	$X_{3i} = 100$	$x_{3i} - 70$	+1	-1	+1	+1
0,92	0,92	0,95	0,95	$X_{4i} = 100$	$x_{4i} - 93$	-1	-1	+2	+2
1,18	1,17	1,21	1,19	$X_{5i} = 100$	$x_{5i} - 119$	-1	-2	+2	0

При подсчете отдельных сумм квадратов по уравнению (2.6а) получается

$$\sum (X_{1i} - \bar{X}_1)^2 = 1^2 + 0^2 + 1^2 + 2^2 - 2^2/4 = 5$$

$$\sum (X_{2i} - \bar{X}_2)^2 = 1^2 + 1^2 + 0^2 + 1^2 - 1^2/4 = 3$$

$$\sum (X_{3i} - \bar{X}_3)^2 = 1^2 + 1^2 + 1^2 + 1^2 - 2^2/4 = 3$$

$$\sum (X_{4i} - \bar{X}_4)^2 = 1^2 + 1^2 + 2^2 + 2^2 - 2^2/4 = 9$$

$$\sum (X_{5i} - \bar{X}_5)^2 = 1^2 + 2^2 + 2^2 + 0^2 - 1^2/4 = 9$$

$$\sum \sum (X_{ji} - \bar{X}_j)^2 = 29$$

При $n = 20$ (общее число определений) и $m = 5$ (число проб) получится

$$S = \sqrt{\frac{29}{20-5}} = 1,4$$

После обратного преобразования (которое не учитывает смещение начала отсчета $X_{ji} - X_j$) получают $s = 0,014 \sim 0,01\%$ Мп (абс.) при $f = 15$ степеням свободы.

Обычно следует проводить два параллельных определения, получая для каждой пробы два значения. Если x' и x'' два результата, относящиеся к одной пробе, то для суммы квадратов можно написать

$$QS = \left[x' - \frac{(x' + x'')}{2} \right]^2 + \left[x'' - \frac{(x' + x'')}{2} \right]^2 = \frac{1}{2} (x' - x'')^2$$

Если исходят из уравнения (5.1), то при m пробах и $n = 2m$ анализов средняя квадратичная ошибка получается

$$s = \sqrt{\frac{1}{2} \frac{\sum (x' - x'')^2}{n - m}} = \sqrt{\frac{\sum (x' - x'')^2}{2m}} \quad (5.2)$$

при $f = m$ степеней свободы.

Возможна также проверка того, что ошибки, вычисленные для различных проб, не отличаются значимо.

[5.2]. При фотометрическом определении хрома в стали были проведены двукратные определения десяти проб с разным содержанием. Из найденных значений x' и x'' (данных в % Cr) среднюю квадратичную ошибку подсчитывают по следующей схеме:

Проба	x'	x''	$x' - x''$	$(x' - x'')^2$
1	3,77	3,75	0,02	0,0004
2	2,52	2,55	0,03	0,0009
3	2,46	2,48	0,02	0,0004
4	3,25	3,20	0,05	0,0025
5	1,82	1,85	0,03	0,0009
6	2,05	2,10	0,05	0,0025
7	0,88	0,90	0,02	0,0004
8	1,04	1,02	0,02	0,0004
9	1,10	1,13	0,03	0,0009
10	1,52	1,48	0,04	0,0016

$$\sum (x' - x'')^2 = 0,0109$$

Отсюда находят среднюю квадратичную ошибку по уравнению (5.2)

$$s = \sqrt{\frac{0,0109}{20}} = 0,023 \sim 0,02\% \text{ Cr (абс.)}$$

при $f = 10$ степеням свободы.

При наличии логарифмически-нормального распределения среднюю квадратичную ошибку подсчитывают для логарифмов значений измерений. Часто так подбирают метод анализа, что автоматически происходит потенцирование (например, логарифмическим делением концентрационной оси калибровочной кривой). В этих случаях для статистической оценки результатов необходимо опять перейти к логарифмам. При этом используют преимущественно четырех-, реже трехзначные таблицы логарифмов. Среднюю квадратичную ошибку подсчитывают для значений логарифмов уже описанным способом. Эта логарифмическая квадратичная ошибка s_{lg} представляет собой оценку параметра σ_{lg} в логарифмически нормальной генеральной совокупности. В практических целях эту среднюю квадратичную ошибку можно использовать только очень ограниченно, так как она не обладает достаточной наглядностью. При потенцировании получают асимметричное распределение частот (рис. 2.5), параметр которого σ нельзя оценить по тем значениям, для которых вычислялась s_{lg} . Поэтому среднюю квадратичную ошибку s_{lg} используют отдельно для возрастающих и убывающих значений. При этом $+s_{lg} = \lg s$ и $-s_{lg} = \lg 1/s$. Ошибка для высоких содержаний всегда больше, чем для низких, однако практически это заметно лишь при величине ошибки более 10% (отн.). Ошибка дается в виде относительной ошибки.

[5.3]. При спектрохимическом анализе бедных оловянных руд для четырех разных проб получены следующие результаты (% Sn):

Проба 1	Проба 2	Проба 3	Проба 4
0,095	0,14	0,38	0,80
0,120	0,18	0,44	0,70
0,080	0,16	0,31	0,85
0,107	0,21	0,36	0,95

Учитывая логарифмически-нормальное распределение (ср. пример [2.4]), эти значения преобразуют в логарифмы. После этого подсчитывают среднюю квадратичную ошибку по формуле (5.1)

подобно примеру [5.1]. При составлении разности $\sum (x_{ji} - \bar{x}_j)^2$ отрицательные значения характеристик логарифмов дают нуль, и их можно сразу не принимать во внимание. Мантиссы преобразуют уже описанным ранее способом (ср. пример [2.7]) так, что внутри каждой серии $\sum x_{ji}/4 = 0$. Получают следующую схему:

Проба	lg x	Преобразования мантиссы	Преобразованные значения	Сумма квадратов $\sum (X_{ji} - \bar{X}_j)^2$
1	0,978—2	$X_{1i} = 1000x_{1i} - 997$	-19	16 945
	0,079—1		+82	
	0,903—2		-94	
	0,029—1		+32	
2	0,146—1	$X_{2i} = 1000x_{2i} - 232$	-86	16 809
	0,255—1		+23	
	0,204—1		-28	
	0,322—1		+90	
3	0,580—1	$X_{3i} = 1000x_{3i} - 568$	+12	11 842
	0,643—1		+75	
	0,491—1		-77	
	0,556—1		-12	
4	0,903—1	$X_{4i} = 1000x_{4i} - 914$	-11	9 203
	0,845—1		-69	
	0,929—1		+15	
	0,978—1		+64	

$$\sum \sum (X_{ji} - \bar{X}_j)^2 = 54\ 799$$

При $n = 16$ (общее число всех определений) и $m = 4$ (число проб) получают

$$S = \sqrt{\frac{54\ 799}{16-4}} = 67,6$$

После обратного преобразования получают логарифмическую среднюю квадратичную ошибку $s_{lg} = 0,068$. Потенцирование приводит к $+0,068 = \lg 1,17$ и $-0,068 = \lg 1/1,17 = \lg 0,86$. Относительная средняя квадратичная ошибка охватывает интервал 0,86...1,17 (или -14%...+17%) при степенях свободы $f = 12$.

Приближенная оценка средней квадратичной ошибки σ возможна при использовании размаха варьирования R . По уравнению (2.7)

$$R = x_{\max} - x_{\min}$$

Таблица 5.1

Значения $d(n_j)$ в зависимости от числа параллельных определений и числа проб с соответствующим числом степеней свободы f [1]

n_j	2		3		4		5	
	$d(n_j)$	f	$d(n_j)$	f	$d(n_j)$	f	$d(n_j)$	f
1	1,41	1,0	1,91	2,0	2,24	2,9	2,48	3,8
2	1,28	1,9	1,81	3,8	2,15	5,7	2,40	7,5
3	1,23	2,8	1,77	5,7	2,12	8,4	2,38	11,1
4	1,21	3,7	1,75	7,5	2,11	11,2	2,37	14,7
5	1,19	4,6	1,74	9,3	2,10	13,9	2,36	18,4
6	1,18	5,5	1,73	11,1	2,09	16,6	2,36	22,0
7	1,17	6,4	1,72	12,9	2,08	19,4	2,35	25,6
8	1,16	7,2	1,71	14,8	2,08	22,1	2,35	29,3
9	1,15	8,1	1,70	16,6	2,07	24,9	2,34	32,9
10	1,14	9,0	1,69	18,4	2,07	27,6	2,34	36,5
$n_j > 10$	1,13	0,88	1,69	1,82	2,06	2,74	2,33	3,62

$d(n_j)$	f	6		7		8		9		10	
		$d(n_j)$	f	$d(n_j)$	f	$d(n_j)$	f	$d(n_j)$	f	$d(n_j)$	f
2,67	4,7	2,83	5,5	2,96	6,3	3,08	7,0	3,18	7,7		
2,60	9,2	2,77	10,8	2,91	12,3	3,02	13,8	3,13	15,1		
2,58	13,6	2,75	16,0	2,89	18,3	3,01	20,5	3,11	22,6		
2,57	18,1	2,74	21,3	2,88	24,4	3,00	27,3	3,10	30,1		
2,56	22,6	2,73	26,6	2,87	30,4	2,99	34,0	3,10	37,5		
2,56	27,1	2,73	31,9	2,87	36,4	2,99	40,8	3,10	45,0		
2,56	31,5	2,73	37,1	2,87	42,5	2,98	47,6	3,09	52,4		
2,55	36,0	2,72	42,4	2,86	48,5	2,98	54,3	3,09	59,8		
2,55	40,5	2,72	47,7	2,86	54,5	2,98	61,1	3,09	67,3		
2,55	44,9	2,72	52,9	2,86	60,6	2,98	67,8	3,09	74,8		
2,53	4,4	2,70	5,27	2,85	6,03	2,97	6,76	3,08	7,45		

Если для m разных проб проведено одинаковое число многократных определений n_j , то можно усреднить полученный размах варьирования:

$$\bar{R} = \sum R_j / m \quad (5.3)$$

(если $n_j = \text{const}$)

Если можно предположить нормальное распределение, то между средним значением размаха варьирования \bar{R} и средней квадратичной ошибкой σ существует приближенное соотношение

$$\sigma = \frac{\bar{R}}{d(n_j)} \quad (5.4)$$

Числовые значения для $d(n_j)$ надо брать из табл. 5.1. Число степеней свободы f для этой приближенно вычисленной средней квадратичной ошибки зависит от числа параллельных определений n_j и числа исследуемых проб m . Табл. 5.1 показывает, что f здесь всегда меньше, чем при расчете средней квадратичной ошибки по уравнению (5.1). Это уменьшение особенно значительно для $n_j > 6$ параллельных определений (ср. разд. 2.2.2).

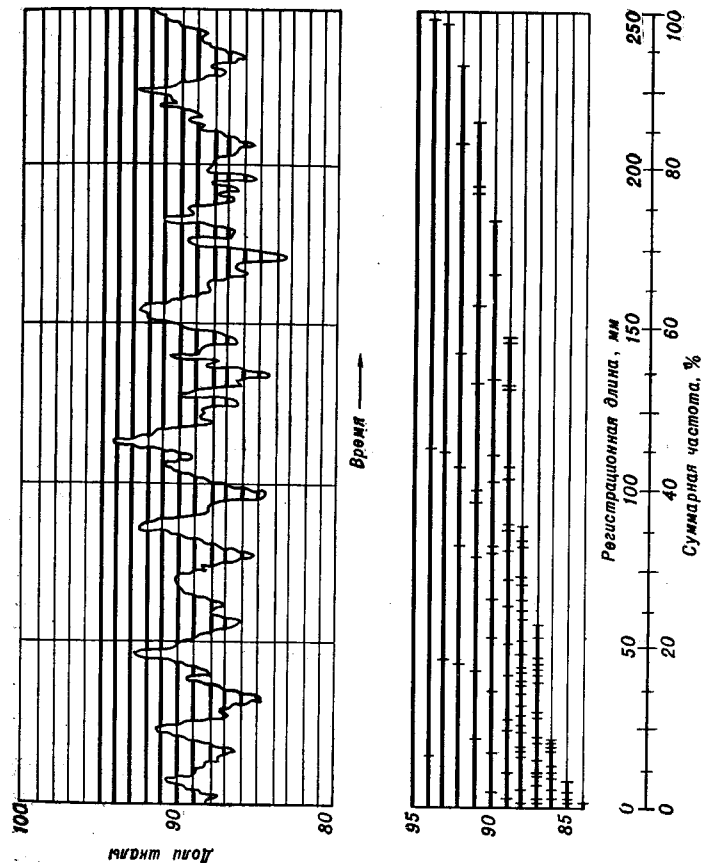
[5.4]. Из значений примера [5.1] нужно приблизительно оценить среднюю квадратичную ошибку по размаху варьирования. Получают

Значения анализа, % Mn				Размах варьирования R_j
0,31	0,30	0,29	0,32	0,03
0,59	0,57	0,58	0,57	0,02
0,71	0,69	0,71	0,71	0,02
0,92	0,92	0,95	0,95	0,03
1,18	1,17	1,21	1,19	0,04

Средний размах варьирования $\bar{R} = 0,03$

Фактор $d(n_j)$ в уравнении (5.4) берут из табл. 5.1 для $n_j = 4$ и $m = 5$ и получают $d(n_j) = 2,10$.
При этом

$$\sigma \sim \frac{0,03}{2,10} = 0,014 \sim 0,01\% \text{ Mn (абс.)}$$



Р и с. 5.4. Зависимость измеряемой величины от времени. Измеренные значения (вверху) и значения суммарной частоты (%) (внизу) для отдельных классов.

Число степеней свободы, найденное из табл. 5.1, равно $f = 13,9$; оно меньше, чем при подсчете средней квадратичной ошибки из сумм квадратов (ср. пример [5.1]).

При обычном аналитическом исследовании во все возрастающей степени используют наглядные способы. При этом обсуждают пригодность аппаратуры, в том числе уровень ее фона, т. е. оцениваются те изменения измеряемых значений, которые происходят во времени за счет аппа-

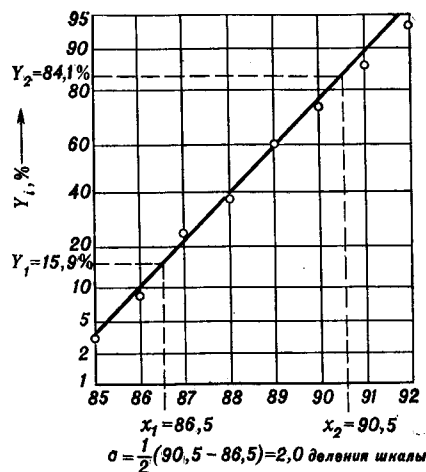


Рис. 5.2. Определение средней квадратичной ошибки по суммарным частотам из рис. 5.1.

ратуры. Для рассмотрения этого вопроса регистрируют спорную величину измерения в течение долгого времени. Если регистрация была произведена с достаточной чувствительностью, то при этом получают кривую того вида, который приведен на рис. 5.1. Из подобной регистрограммы можно получить среднюю квадратичную ошибку фона простым графическим методом.

Для этого прежде всего делят ось измеренных значений на достаточно большое число классов (около 10). Затем определяют длину диаграммы по отдельным границам классов сверху над кривой. Эти величины соответствуют суммируемым частотам функции распределения (ср. разд. 3.1). Их находят простейшим способом, обозначая два раза границы класса на полосе кальки. Сдвигом этой полосы в направлении оси времени суммируют все отрезки для соответствующих границ класса, которые лежат выше хода

кривой (рис. 5.1 внизу). Сумму отрезков выражают в процентах по отношению ко всей длине диаграммы и получают распределение суммы частот. Отдельные пары значений (верхняя граница класса и соответствующая сумма частот) заносят на вероятностную бумагу (рис. 5.2). В пределах $10\% < Y_i < 90\%$ находят прямую и получают σ (а если хотят, и μ) уже описанным методом (ср. пример [3.1]).

[5.5]. Рис. 5.1 показывает изменение значения холостого опыта во времени при спектрометрическом определении цинка в стабилизированной электродуге. Для определения средней квадратичной ошибки значения холостого опыта ось значений измерений (ординату) делят на отдельные классы. При этом в данном случае можно использовать деления диаграммной бумаги. Измеряют длину границ классов над зарегистрированной кривой. При этом находят, например, что границе класса $x = 84$ делениям шкалы соответствует длина 2 мм под кривой. Это равно 0,8% общей длины диаграммы (250 мм). Граница верхнего соседнего класса $x = 85$ делениям шкалы отрезает над кривой четыре отдельных отрезка общей длиной 8,5 мм (~3,4%). Таким образом можно установить следующее распределение сумм частот:

Верхняя граница x_i , деления шкалы	Общая длина над кривой		Верхняя граница x_i , деления шкалы	Общая длина над кривой	
	мм	Y_i , %		мм	Y_i , %
< 84	2,0	0,8	< 90	187,0	74,8
< 85	8,5	3,4	< 91	217,0	86,8
< 86	22,3	8,9	< 92	234,5	93,8
< 87	58,0	23,2	< 93	247,0	98,8
< 88	89,5	35,8	< 94	249,5	99,8
< 89	148,0	59,2	< 95	250,0	100,0

Пары значений (x_i, Y_i) наносят на вероятностную бумагу и получают в качестве средней квадратичной ошибки (рис. 5.2) $\sigma = 2,0$ деления шкалы (абс.).

При выборе регистрограммы следует позаботиться о том, чтобы она была отображена удачно, т. е. чтобы регистрация проходила за достаточно большое время. Иначе нормальное распределение может не выполняться, и, следовательно, будут получаться ложные данные о воспроизводимости аппаратуры. Чтобы сохранить ошибку отсчета в возможно

более узких границах, следует в достаточной мере растянуть масштаб абсциссы и ординаты. Подробности о других методах обработки измерений, представленных регистраграммами, можно найти у Смирнова и Дунина-Барковского.

5.2. Примеры

Средняя квадратичная ошибка характеризует случайную ошибку метода анализа общепринятым и ясным образом. Поэтому ее постоянно используют для оценки метода анализа. Неопределенные термины (например, «ошибка метода анализа», или «точность метода анализа», или «средняя ошибка» и т. д.) должны исчезнуть из литературы, так как они не удовлетворяют объективным требованиям и легко приводят к ложным толкованиям.

Каждую подсчитанную среднюю квадратичную ошибку можно представить как случайную величину, т. е. при повторении опыта всегда получают различные числовые значения s . Поэтому возникает вопрос, как можно установить доверительный интервал для оценки s аналогично доверительному интервалу среднего значения \bar{x} . Если обозначить верхнюю границу такого интервала s_0 , а нижнюю s_u , то F -распределение дает следующие соотношения (ср. разд. 3.3.2):

$$\frac{s_0^2}{s^2} = F(\bar{P}, f_1 = \infty, f_2 = f)$$

$$\frac{s_u^2}{s^2} = \frac{1}{F(\bar{P}, f_1 = f, f_2 = \infty)}$$

При этом f — число степеней свободы соответствующей оценки s . Если обозначить

$$\sqrt{F(\bar{P}, f_1 = \infty, f_2 = f)} = \kappa_0$$

$$\sqrt{1/F(\bar{P}, f_1 = f, f_2 = \infty)} = \kappa_u$$

то получим

$$s_0 = \kappa_0 s \quad (5.5)$$

$$s_u = \kappa_u s$$

Для вероятностей $\bar{P} = 0,95$ и $\bar{P} = 0,99$ значения κ_0 и κ_u приведены в табл. 5.2.

Таблица 5.2

Значения κ_0 и κ_u для подсчета доверительного интервала средней квадратичной ошибки

f	$\bar{P} = 0,95$		$\bar{P} = 0,99$	
	κ_0	κ_u	κ_0	κ_u
1	15,9	0,51	80	0,39
2	4,42	0,58	10	0,47
3	2,92	0,62	5,11	0,51
4	2,37	0,65	3,67	0,55
5	2,09	0,67	3,00	0,58
6	1,92	0,69	2,62	0,60
7	1,80	0,71	2,38	0,62
8	1,71	0,72	2,20	0,63
9	1,65	0,73	2,08	0,65
10	1,59	0,74	1,98	0,66
11	1,55	0,75	1,90	0,67
12	1,52	0,76	1,83	0,68
15	1,44	0,77	1,69	0,70
20	1,36	0,80	1,56	0,73
25	1,31	0,81	1,47	0,75
30	1,27	0,83	1,42	0,77
40	1,23	0,85	1,34	0,79
60	1,18	0,87	1,27	0,82
120	1,12	0,90	1,17	0,87

[5.6]. В качестве случайной ошибки определения марганца по Простеру и Смитсу в примере [5.1] была подсчитана средняя квадратичная ошибка $s = 0,014\%$ Мп при $f = 15$. Из уравнения (5.5) и табл. 5.2 при $\bar{P} = 0,95$ получают соответствующий доверительный интервал:

$$s_0 = 1,44s = 0,020\% \text{ Мп}$$

$$s_u = 0,77s = 0,011\% \text{ Мп}$$

При повторном определении s можно ожидать, что в 95% все выборочные значения s будут лежать выше 0,011 и ниже 0,020% Мп.

Чтобы найденная средняя квадратичная ошибка полностью характеризовала случайную ошибку метода анализа, должен выполняться ряд предположений. Как уже

было отмечено в начале этой главы, обязательно следует исключить негомогенность проб, ошибки, связанные с особенностями работы лаборатории, и персональные ошибки. Каждый анализ, который должен служить основанием для подсчета s , следует выполнить целиком, включая все повторяющиеся операции.

Особенно важно, чтобы отдельные параллельные определения не зависели друг от друга. Поэтому их проводят в разное время. Например, если надо провести для трех проб А, В и С по три параллельных определения, то порядок этого проведения такой [6]:

1-й день	2-й день	3-й день
$A_1B_1C_1$	$B_2C_2A_2$	$C_3A_3B_3$

Проведенные вместе параллельные определения не всегда можно считать нормально распределенными величинами. Они приводят к островершинному распределению частот (рис. 2.6, а) и дают заниженную ошибку метода. Их можно использовать только при оценке грубой ошибки.

Для подсчета средней квадратичной ошибки всегда используют неокругленные результаты анализа с ненадежным последним знаком после запятой. Это правило целиком справедливо для всех статистических расчетов. Преждевременное округление снизу или сверху может совершенно исказить значение ошибки метода.

При определении средней квадратичной ошибки следует стремиться к тому, чтобы за счет правильной постановки опыта полученная оценка s лежала возможно ближе к значению σ генеральной совокупности. Особое влияние на это оказывает число степеней свободы, относящееся к средней квадратичной ошибке. При степенях свободы $f > 30$ можно считать, что требование $s \sim \sigma$ для практических целей выполняется (табл. 5.2). В основе определения s должно лежать достаточно много значений измерений. В каждом случае следовало бы наряду со средней квадратичной ошибкой указывать соответствующее число степеней свободы. Только с этим дополнением средняя квадратичная ошибка может быть использована в дальнейшем.

Средняя квадратичная ошибка часто зависит от величины измеренного значения и состава пробы (табл. 5.3).

Таблица 5.3

Средняя квадратичная ошибка методов анализа стали (u — нелегированная сталь, h — легированная сталь с $Cr < 2\%$ и $Ni < 2\%$; определять только С, Mn, Si, P, S; s — низколегированная сталь с $Cr < 2\%$ и $Ni < 2\%$; h — высоколегированная сталь)

Элемент	Метод	Род пробы	Содержание x , %	Средняя квадратичная ошибка воспроизводимости, % (ср. стр. 156)	Степени свободы	Доверительный интервал, % ($P = 0,95$)
Углерод	Газовометрический	Армко	0,015—0,04	$s_w = 0,008$	40	$\pm 0,001$
		Сталь (u , s , h)	0,1—1,1	$s_w = 10^{-3}(2x + 3,8)$	100	$\pm 0,001$
Кремний	Гравиметрический	Чугун	1,5—3,0	$s_w = 0,011$	56	$\pm 0,001$
		Сталь (u , s)	0,2—1,3	$s_w = 10^{-3}(3,6x + 5,3)$	60	$\pm 0,001$
Марганец	Объемный	»	0,2—1,5	$s_w = 10^{-3}(2x + 5,6)$	100	$\pm 0,001$
		Сталь (u)	0,006—0,05	$s_w = 0,0006$	150	$\pm 0,0001$
Сера	Метод фиксации	»	0,007—0,02	$s_w = 0,0005$	88	$\pm 0,0001$
		Сталь (s)	0,01—0,06	$s_w = 10^{-4}(120x + 3,8)$	100	$\pm 0,0001$
Фосфор	Гравиметрический	»	0,03—0,1	$s_w = 0,003$	80	$\pm 0,001$
		Сталь (s , h)	0,1—1,0	$s_w = 0,006$	84	$\pm 0,001$
Хром	Фотометрический	»	1,0—3,0	$s_w = 0,5\%$ (отн.)	100	$\pm 0,1\%$ (отн.)
		Сталь (s , h)	3,0—10,0	$s_w = 0,4\%$ (отн.)	80	$\pm 0,1\%$ (отн.)
Никель	Гравиметрический	»	10,0—30,0	$s_w = 0,3\%$ (отн.)	42	$\pm 0,1\%$ (отн.)
		»	0,1—2,0	$s_w = 10^{-3}(7,5x + 5,5)$	60	$\pm 0,001$
Медь	Фотометрический	»	2,0—10,0	$s_w = 0,5\%$ (отн.)	100	$\pm 0,1\%$ (отн.)
		Сталь (u , s , h)	0,1—0,3	$s_w = 0,005$	200	$\pm 0,001$

Средняя квадратичная ошибка, полученная для заданного содержания и специального состава пробы, не может быть безоговорочно обобщена. Соответствующими измерениями и последующим применением статистического метода проверки (ср. гл. 7) следует доказать применимость полученного значения квадратичной ошибки.

В качестве оценки параметра функции распределения средняя квадратичная ошибка чаще всего дается в виде абсолютной ошибки. Однако если относительная ошибка меньше зависит от величины измеренных значений, то используют этот вид ошибки. При этом часто применяют коэффициент вариации $V = s/x$. Во избежание недоразумений следует использованный вид ошибок обозначать добавлением сокращений «абс.» для абсолютной ошибки и «проц.» для выражения ошибки в процентах.

[5.7]. Для методических исследований были получены средние квадратичные ошибки для определения калия по перхлоратному методу и при использовании пламенной фотометрии [4]. Исследование солей калия различного содержания дало следующую картину:

Содержание K_2O , %	Средняя квадратичная ошибка			
	перхлоратный метод		пламенная фотометрия	
	абс.	проц.	абс.	проц.
1,5	0,10	6,8	0,02	1,3
15	0,18	1,2	0,15	1,0
40	0,18	0,45	0,45	1,1
50	0,19	0,38	0,57	1,1
90	0,20	0,22	—	—

При перхлоратном методе абсолютная ошибка остается постоянной, а при пламеннофотометрическом методе (как при многих физических методах) постоянна относительная ошибка. Поэтому в первом случае используют абсолютную ошибку, а во втором — относительную. Дальнейшее рассмотрение показало, что перхлоратный метод имеет преимущество при точном определении высоких содержаний, а пламенная фотометрия дает оптимальные результаты при низких и средних содержаниях.

Таблица 5.4

Средние квадратичные ошибки (абс.) при анализе магниевого сплава

(Значения взяты из монографии [7])

Элемент	Метод	Содержание, %	Средняя квадратичная ошибка, %
Al	Гравиметрический в виде оксината	0,5	0,005
		8	0,025
As	Титрование оксихинолята	8	0,035
		0,01	0,00015
Cu	Иодометрическое титрование	0,05	0,0025
		0,2	0,005
	Электролитический	0,17	0,0015
		0,02	0,0005
Fe	Иодометрическое титрование	0,005	0,0001
		0,005	0,0001
K	Фотометрический с диэтилдитиокарбамамом	0,005	0,00015
		0,005	0,00015
Mn	Пламеннофотометрический	1	0,01
		0,005	0,0005
	Титрование перманганатом с солью Мора	0,25	0,004
		1,3	0,01
Na	Фотометрический в виде перманганата	0,05	0,0025
		0,005	0,0001
Ni	Пламеннофотометрический	0,05	0,0025
		0,005	0,0001
P	Фотометрический в виде диацетилглиоксимата никеля	0,01	0,0005
		0,01	0,0005
Si	Фотометрический в виде фосфорномолибденовой сини	0,2	0,0065
		0,15	0,0002
Th	Гравиметрический	0,5	0,007
		0,5	0,007
Zn	Фотометрический с молибденовой синью	2	0,09
		5	0,022
	Потенциометрическое титрование (индикатор — дифенилбензидин)	2	0,01
		0,02	0,001
Zr	Фотометрический (дитизон)	1	0,01
		0,5	0,01
	Гравиметрический в виде двуокиси циркония	40	0,1
		0,5	0,008
	Фотометрический с ализарином S	0,5	0,008

Во многих случаях такой ясной картины, как в примере [5.7], не получается. Поэтому следует указать зависимость средней квадратичной ошибки от величины измеренного значения. Существуют разные формы представления этой

зависимости. Если пользуются табличными данными, то следует обратить внимание на возможность линейной интерполяции. Нередко записывают функции, показывающие зависимость ошибки от содержания или от измеряемой величины. Этот вид представления особенно легко допускает интерполяцию промежуточных значений. Однако необходимо учитывать, что подобную функциональную зависимость следует понимать только как эмпирическую зависимость, а не как закон природы.

Значения средних квадратичных ошибок при анализе стали приведены в табл. 5.3. Указанные там данные являются результатами долголетней и очень тщательной исследовательской работы «Стандартный анализ проб стали и ферросплавов» управления по проведению измерений и проверки изделий в Магдебурге. Все эти данные были получены на основании статистической оценки большого числа анализов, выполненных аналитиками на различных предприятиях в течение нескольких лет. Эти значения справедливы только для отмеченных содержаний и типов сплавов. Однако они дают определенные ориентировочные указания об ожидаемой случайной ошибке при исследовании проб другого рода.

Значения средних квадратичных ошибок при анализе магниевых сплавов дает табл. 5.4. Эти сведения взяты из сборника стандартов анализа манчестерской «Magnesium Electron Ltd.»

ЛИТЕРАТУРА

1. David H. A., *Biometrika*, 38, 393 (1951).
2. Doerffel K., *Z. anal. Chem.*, 185, 1 (1962).
3. Doerffel K., Schulze M., *Neue Hütte*, 9, 690 (1964).
4. Knopf A., *Mitteilungen für die Kaliindustrie*, 4 (1961).
5. Moran R. F., *Anal. Chem.*, 19, 961 (1947).
6. Wood E. C., *Analyt. chim. Acta*, 2, 441 (1948).
7. *Chemical and Spectrochemical Analysis of Magnesium and its Alloys*. Magnesium Electron Ltd., Manchester.

6. Обсуждение данных анализа

Целью количественного анализа является получение информации о количественном составе исследуемого материала. Чтобы избежать недоразумений при оценке полученного результата анализа, следует указать соответствующую ошибку (ср. разд. 2.2). Ошибку физических измерений нельзя переносить непосредственно на методы аналитической химии, так как в аналитической химии ошибки измерения чаще всего играют подчиненную роль по отношению ко всем нарушениям хода химических реакций. Для характеристики возникающей ошибки может служить доверительный интервал, объясненный в разд. 3.1 [уравнение (3.9)]. Расчет этой величины для специальных условий химического анализа и примеры ее применения для описания качества анализируемых продуктов будут изложены в данной главе.

6.1. Вычисление доверительного интервала

В общем случае аналитик может строить свои доказательства только на очень ограниченном числе результатов. Поэтому для расчета доверительного интервала приходится возвращаться к нормальному распределению, используя связанное с ним t -распределение. По аналогии с уравнением (3.8) можно вычислить доверительный интервал среднего значения \bar{x} . Тогда среднее значение задают в виде

$$\bar{x} \pm \frac{t(P, f) s}{\sqrt{n_j}} = \bar{x} \pm \Delta \bar{x} \quad (6.1)$$

При этом следует предварительно выбрать вероятность P . Необходимые значения $t(P, f)$ можно взять из табл. 12.3. Из уравнений (6.1) следует, что при очень боль-

ших повторениях выборки в $100P\%$ случаев истинное значение пробы μ лежит внутри данной границы $\pm \Delta \bar{x}$. Поэтому доверительный интервал используют как границу для среднего значения \bar{x} . Кроме того, уравнение (6.1) дает границы, внутри которых лежит истинное значение μ , совместимое с найденным средним для выборки \bar{x} .

[6.1]. При анализе железной руды были найдены следующие значения ($\% \text{Fe}_2\text{O}_3$):

38,71
38,90
38,62
38,74

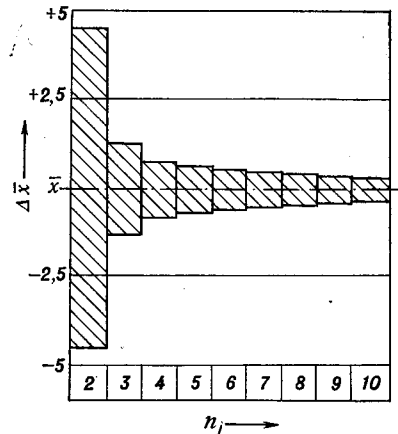
Среднее $\frac{38,74}{4}$

По уравнению (2.5) находят среднюю квадратичную ошибку $s = 0,12\% \text{Fe}_2\text{O}_3$ при $f = 3$. Из уравнения (6.1) доверительный интервал для среднего значения при $P = 0,95$ равен

$$\Delta \bar{x} = \frac{3,18 \cdot 0,12}{\sqrt{4}} = 0,19\% \text{Fe}_2\text{O}_3$$

С учетом этого результаты анализа: $(38,74 \pm 0,19)\% \text{Fe}_2\text{O}_3$ (для $P = 0,95$).

Рассчитанный по уравнению (6.1) доверительный интервал в значительной степени зависит от числа параллель-



Р и с. 6.1. Доверительный интервал для среднего значения при $s = 1$ и $P = 0,95$ в зависимости от числа параллельных определений n_j .

ных определений. Из рис. 6.1 видно, что при переходе от двух к трем или четырем параллельным определениям

точность данных значительно увеличивается. Однако это преимущество с дальнейшим ростом числа параллельных определений становится незначительным по отношению к рабочим затратам. Поэтому более четырех параллельных определений проводят только в особых случаях. Если оценка s средней квадратичной ошибки известна из более ранних исследований, то ее также можно применять для определения доверительного интервала. Чаще всего такие оценки уже известных средних квадратичных ошибок обладают большим числом степеней свободы. Вследствие этого величина $t(P, f)$ уменьшается; кроме того, уменьшается доверительный интервал, и для средней квадратичной ошибки (ср. стр. 106). Следовательно, получают значительно более надежные данные.

[6.2]. При гравиметрическом определении никеля в стали были найдены значения 4,64; 4,67 и 4,65% Ni. Среднюю квадратичную ошибку находят в табл. 5.3: $s = 0,5\% \text{Ni}$ (отн.) = $0,023\% \text{Ni}$ (абс.) при $f = 100$ степеней свободы. Из табл. 12.3 получают $t(P = 0,95; f = 100) \sim 2,0$, а отсюда доверительный интервал

$$\Delta \bar{x} = \frac{2,0 \cdot 0,023}{\sqrt{3}} = 0,027\% \text{Ni (абс.)}$$

Содержание исследуемой пробы может составлять $(4,65 \pm 0,027)\% \text{Ni}$, т. е. округленно $(4,65 \pm 0,03)\% \text{Ni}$ для вероятности $P = 0,95$. Если бы доверительный интервал определяли только из трех параллельных определений по уравнению (6.1), то получили бы

$$\Delta \bar{x} = \frac{3,18 \cdot 0,023}{\sqrt{3}} = 0,042\% \text{Ni}$$

Среднее значение в этом случае можно было бы установить со значительно меньшей точностью.

При известном значении средней квадратичной ошибки s можно также оценить интервал для отдельного значения (т. е. $n_j = 1$), внутри которого при вероятности $100P\%$ следует ожидать значение μ . Тогда отдельное значение можно записать в форме

$$\bar{x} \pm t(P, f)s = x \pm \Delta x \quad (6.2)$$

Величину Δx по стандарту TGL 0-51849 для $P = 0,997$ называют областью разброса.

[6.3] При определении марганца по Простеру и Смигу в примере [5.1] была получена средняя квадратичная ошибка $s = 0,014\%$ Мп при $f = 15$. Отсюда по уравнению (6.2) при $P = 0,95$ находят $\Delta x = t$ ($P = 0,95; f = 15$) $s = 2,13 \cdot 0,014 = 0,030\%$ Мп. Вследствие этого можно видеть результат анализа в виде $(x \pm 0,03)\%$ Мп (при $P = 0,95$).

При логарифмическо-нормальном распределении доверительная область сильно отличается по обе стороны от данного значения. Если $\bar{x}_{1g} = \lg \bar{x}$ и $s_{1g} = \lg s$, то $\Delta \bar{x}_{1g} = \pm t(P, f) s_{1g} / \sqrt{n_j}$. Если необходимо вернуться от логарифмов к обычным значениям измерений, то получают $\bar{x}_{1g} \pm \Delta \bar{x}_{1g} = \lg \bar{x} \pm \pm \lg \Delta x$. Это равнозначно $x \Delta x$ и соответственно $x / \Delta x$. Следует иметь в виду, что доверительный интервал в этом случае задается относительной ошибкой.

[6.4]. Логарифмическая средняя квадратичная ошибка для спектрохимического определения олова была дана в примере [5.3] как $s_{1g} = 0,068$ при $f = 12$. Доверительный интервал среднего значения из четырех параллельных определений ($n_j = 4$) для $P = 0,95$ равен $t(P = 0,95; f = 12) s_{1g} / \sqrt{n_j} = 2,18 \cdot 0,068 / \sqrt{4} = 0,074$. При потенцировании получают $\Delta x_o = 1,18$ и $\Delta x_u = 0,85$. Среднее значение из четырех параллельных определений является, следовательно, неопределенным в пределах $0,85 x < x < 1,18 x$.

Данные об ошибке должны получиться в другой форме, если результаты измерений следуют распределению Пуассона. Если полученное среднее значение достаточно велико ($x > 15$), то распределение Пуассона приближается к нормальному распределению (ср. разд. 3.2). В качестве доверительного интервала для отдельного числового результата получают

$$\mu \pm u(P) \sigma = \mu \pm u(P) \sqrt{\mu} \quad (6.3)$$

Это соотношение справедливо только тогда, когда μ известно или заменено расчетным значением. Однако при практических измерениях это часто не имеет места. Тогда из полученного результата x нужно определить границы, внутри которых следует ожидать средние значения μ ; для всех тех генеральных совокупностей, которые могут согласовываться с вероятностью P для найденных числовых значений x . Из уравнения (6.3) следует

$$-u(P) \sqrt{\mu} < \mu - x < +u(P) \sqrt{\mu}$$

Прибавляют x к каждой части этого неравенства и решают его относительно μ . Это приводит к следующему соотношению:

$$x + \frac{u^2(P)}{2} - \frac{u(P)}{2} \sqrt{u^2(P) + 4x} < \mu < x + \frac{u^2(P)}{2} + \frac{u(P)}{2} \sqrt{u^2(P) + 4x} \quad (6.4)$$

Истинное значение исследуемой пробы лежит внутри этих границ — при частом повторении выборки с вероятностью $100 P\%$ оно не будет выходить за эти границы. При чаще всего используемой вероятности $P = 0,95$ можно подставить $u(P) \sim 2$. Если $x > 15$, то $u^2(P = 0,95) \ll \ll 4x$, а соотношение (6.4) можно приближенно записать в виде

$$x + 2 - 2\sqrt{x} < \mu < x + 2 + 2\sqrt{x} \quad (6.5)$$

$(P = 0,95; x > 15)$

Из этого соотношения можно найти доверительный интервал для числовых результатов.

[6.5]. Для радиоактивного препарата было сосчитано 360 импульсов ($x = 360$). По уравнению (6.5) доверительный интервал получается

$$360 + 2 - 2\sqrt{360} < \mu < 360 + 2 + 2\sqrt{360}$$

$$324 < \mu < 400$$

При повторении измерения в тех же условиях следует ожидать, что значение μ будет лежать между 324 и 400 импульсами с вероятностью $P = 0,95$.

В соответствии с разд. 3.2 приближение распределения Пуассона нормальным распределением возможно только при $\bar{x} > 15$. При малых числовых результатах доверительный интервал уже нельзя определить при помощи уравнения (6.5). Следует отделить верхнюю и нижнюю границы доверительного интервала. Их находят из

$$\Delta x_o = x + \frac{\chi^2(\bar{P}, f)}{2} \quad (6.6a)$$

с числом степеней свободы $f = 2(x + 1)$

$$\Delta x_u = x - \frac{x}{F(\bar{P}, f_1, f_2)} \quad (6.6b)$$

с числом степеней свободы $f_1 = \infty, f_2 = 2x$

Числовые значения для $\chi^2(\bar{P}, f_j)$ следует брать из табл. 12.4 (пределы интегрирования для χ^2 -распределения). Значения для $F(\bar{P}, f_1, f_2)$ дает табл. 12.5 (пределы интегрирования для F -распределения). Следует иметь в виду, что Δx_0 и Δx_u дают односторонние границы доверительного интервала по уравнению (6.6) (рис. 3.11). Для получения двусторонних границ, которыми чаще всего интересуются в аналитических данных, следует произвести пересчет по $P = 2(\bar{P} - 1)$ [уравнение (3.10)].

[6.6]. При структурных исследованиях графита на некотором элементе поверхности под микроскопом были подсчитаны поры определенного размера ($x = 11$). С вероятностью $P = 0,90$ по уравнению (6.6) и табл. 12.4 и, следовательно, табл. 12.5 доверительные интервалы Δx_0 и Δx_u получаются

$$\Delta x_0 = 11 + \frac{\chi^2(\bar{P}=0,95; f=24)}{2} = 11 + \frac{36,42}{2} \sim 29$$

$$\Delta x_u = 11 - \frac{11}{F(\bar{P}=0,95; f_1=\infty, f_2=22)} = 11 - \frac{11}{1,78} \sim 5$$

В 90% всех случаев следует ожидать от 5 до 29 пор на одинаковых элементах поверхности.

Как показывает пример [6.6], точность данных довольно незначительна при малом числе результатов. В таких случаях разумно ограничиваться выбором значения вероятности $P = 0,90$ или, в лучшем случае, $P = 0,95$. При более высоких значениях P доверительный интервал слишком растягивается и потому представляется малопримемым.

При исследовании пробы аналитик должен проводить два, три, а порой и четыре параллельных определения. Полученные значения в общем расходятся. Для оценки значений целесообразно иметь критерий допустимой разницы $x_k - x_l$ между параллельными определениями. При известной средней квадратичной ошибке σ справедливо, что

$$x_k - x_l < D(P, n_j) \sigma \quad (6.7)$$

Фактор $D(P, n_j)$ вычислен по Пирсону. Для $P = 0,95$ он приведен в табл. 6.1 для обычно используемых значений $n_j = 2, 3, 4$.

[6.7]. При гравиметрическом определении никеля были найдены значения 4,64; 4,65 и 4,67% Ni. Необходимо проверить, согласуется ли полученная разность 0,03% Ni со случайной ошибкой. Соответствующую среднюю квадратичную ошибку находят в табл. 5.3 в виде $s = 0,5\%$ Ni (отн.) $\sim 0,023\%$ Ni (абс.) при $f = 100$ степеней свободы. Вследствие столь большого числа степеней свободы можно считать $s \sim \sigma$ и получить из уравнения (6.7) с $D(P = 0,95; n_j = 3) = 3,31$

$$x_k - x_l = 0,03\% \text{ Ni} < 3,31 \cdot 0,023 = 0,076\% \text{ Ni.}$$

Разница между тремя значениями с вероятностью $P = 0,95$ согласуется со случайной ошибкой, и найденные значения можно объединить.

При очень большом числе измерений они, конечно, могут в отдельных случаях переходить границу, заданную уравнением (6.7). Если появляются большие различия, чем те, которые обычно встречаются, то целесообразна дополнительная проверка условий опыта (измерительных приборов, метода анализа, а также персонала).

Доверительный интервал с данной вероятностью P дает в общепринятой форме четкие сведения об ошибке результатов анализа. Поэтому его следовало бы всегда применять вместо довольно туманных терминов (например, «граница ошибки», «ошибка значения анализа» и т. д.). Доверительный интервал указывает, с какой вероятностью следует ожидать ошибки вычисленной величины Δx . Однако он не представляет собой специального вида ошибки для специального значения анализа. Возможность того, что отдельное значение несет более высокую ошибку чем Δx , остается с риском $\alpha = 1 - P$. Поэтому числовое значение доверительного интервала всегда должно дополняться указанием вероятности P . Выбор ее является предметом взаимного соглашения. Обычно для расчетов доверительного интервала пользуются $P = 0,95$. Для других внутрипроизводственных данных иногда достаточно $P = 0,90$. Ответственные решения требуют более высокой надежности (например, $P = 0,99$). В фармакологии и близких ей областях особенно целесообразно сохранять высокую надежность $P = 0,99$ или даже $P = 0,999$, когда ошибка практически полностью исключена. В физике час-

Таблица 6.1

Факторы для $P = 0,95$ по Пирсону

n_j	$D(P = 0,95; n_j)$
2	2,77
3	3,31
4	3,65

то довольствуются просто указанием средней квадратичной ошибки и мирятся с высоким риском $\alpha = 1 - 0,683 = 0,317$ появления больших отклонений. Это справедливо только для достаточно большого числа степеней свободы ($f > 10$). При наличии малого числа измерений риск заметно увеличивается. Следует помнить, что при таком способе представления результатов измерений задается доверительный интервал с $u(P) = 1$ и $t(P, f) = 1$, а вовсе не средняя квадратичная ошибка.

Доверительный интервал можно задавать как абсолютной ошибкой с представлением в тех единицах, в которых выражается результат анализа, так и относительной ошибкой, выраженной в процентах от результата анализа. В тех методах анализа, где ошибка остается постоянной для заданной области концентраций, предпочитают первый способ представления; при методах анализа с постоянной относительной ошибкой предпочитают второй способ. Так как результаты анализа чаще всего выражают в процентах, следует уяснить, идет ли речь об абсолютной или об относительной ошибке. Чаще всего это указывается сокращенно: (абс.), (отн.) или (проц.). Величиной ошибки определяется возможное значение отдельного измерения и среднего из нескольких измерений. Результаты измерения и ошибка должны быть выражены числами с одинаковым числом знаков после запятой. Поэтому их округляют в одинаковой степени. Для округления значения справедливы правила стандарта TGL 0-1333 (ср. гл. 10).

6.2. Статистическая оценка качества

Качество многих продуктов выявляется при проведении химико-аналитических исследований. Поскольку проверяемый материал можно считать гомогенным, рассеивание результатов анализа задается ошибкой анализа. Она должна указываться при оценке качества и должна соответствующим образом учитываться.

Чаще всего поставщик и потребитель заключают соглашение о качестве продукта. Продукт будет признан потребителем безукоризненным только тогда, когда аналитически установленное качество T лучше допустимо установ-

ленной нормы T_0 . Если анализ показал плохое качество, то товар не принимают. Вследствие случайной ошибки, появившейся при анализе, как потребитель, так и постав-

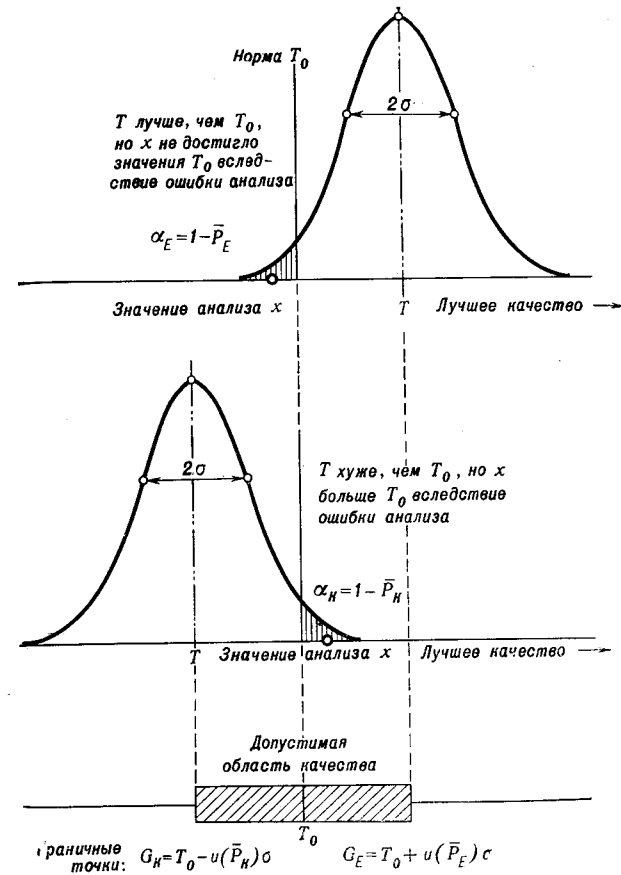


Рис. 6.2. Значение анализа и характеристика качества.

щик при соглашении идут на определенный риск (рис. 6.2). Производитель должен считаться с тем, что хороший продукт может быть ошибочно забракован. Когда качество продукта T выше согласованной нормы T_0 , значение ана-

лиза случайно может оказаться ниже T_0 . Тогда продукт отвергают напрасно (*риск производителя*). Потребитель должен учитывать, что ошибочно может быть принято плохое изделие. Когда качество продукта T ниже допустимой нормы T_0 , значение анализа случайно может оказаться выше T_0 . Продукт принимают ошибочно (*риск потребителя*). При известной случайной ошибке метода анализа σ и при согласованном риске изготовителя и потребителя можно указать границы, внутри которых может колебаться качество продукта. Если $\bar{\alpha}_E = 1 - \bar{P}_E$ риск производителя, а $\bar{\alpha}_K = 1 - \bar{P}_K$ риск потребителя, то получаются следующие границы для качества:

$$G_E = T_0 + \frac{u(\bar{P}_E)\sigma}{\sqrt{n_j}} \quad \text{и} \quad G_K = T_0 - \frac{u(\bar{P}_K)\sigma}{\sqrt{n_j}} \quad (6.8)$$

где n_j — число параллельных определений.

В случае если T_0 представляет минимальную границу (например, содержание удобрения), производителю следует считаться с тем, что анализ не менее чем в $100(1 - \bar{P}_E)\%$ всех случаев дает слишком низкое значение. Тогда продукт будет ошибочно отвергнут, хотя его качество может лежать в согласованных пределах вплоть до $T_0 + \frac{u(\bar{P}_E)\sigma}{\sqrt{n_j}}$. Равным образом существует риск для потребителя, что в $100(1 - \bar{P}_K)\%$ всех случаев значения анализа будут слишком высоки. Продукт будет ошибочно принят, хотя его качество лежит между T_0 и $T_0 - \frac{u(\bar{P}_K)\sigma}{\sqrt{n_j}}$, следовательно, ниже установленной нормы.

[6.8.] Для калийного удобрения с 38% окиси калия требуется, чтобы среднее из двух параллельных определений ($n_j = 2$) лежало по меньшей мере при $T_0 = 38,0\%$ K_2O . Продукты, которые дали при анализе более низкое содержание, могли быть отвергнуты потребителем. Если риск обеих сторон был $\bar{\alpha} = 1 - \bar{P} = 0,05$, то находят обе граничные точки G_E и G_K с $u(\bar{P} = 0,95) = 1,65$ и $\sigma = 0,18\%$ K_2O (ср. пример [5.7]) в виде

$$G_E = 38,0 + \frac{1,65 \cdot 0,18}{\sqrt{2}} = 38,2\% \text{ } K_2O$$

$$G_K = 38,0 - \frac{1,65 \cdot 0,18}{\sqrt{2}} = 37,8\% \text{ } K_2O$$

Не менее чем в 5% всех случаев существует возможность того, что анализ из двух параллельных определений даст значение, которое лежит ниже 38,0% K_2O , хотя проба содержит вплоть до 38,2% K_2O , и что на основании этих данных потребитель откажется от продукта. Равным образом потребитель идет на такой же риск, принимая продукт на основании полученного содержания 38,0% K_2O , хотя содержание K_2O лежит в интервале 38,0 и 37,8%.

В некоторых случаях производитель должен гарантировать качество продукта. Тогда следует договориться, что качество товара T только в $\bar{\alpha} = 100(1 - \bar{P})\%$ всех случаев окажется хуже, чем гарантируемое качество T_g . При выполнении подобной гарантии производитель опять должен учитывать случайную ошибку использованного метода анализа. Он может выдать только такой продукт, качество которого, задаваемое доверительным интервалом, вычисленным исходя из ошибки анализа σ , лучше, чем требуется по норме.

Таким образом:

Потребитель требует: гарантированное качество T_g

Производитель поставляет как минимум: качество $T_g + u(\bar{P})\sigma$ (6.9)

При n_j параллельных определениях доверительный интервал еще делится на $\sqrt{n_j}$. Эта граница качества будет точна в $100\bar{P}\%$ всех случаев. Однако потребитель при $100\bar{\alpha} = 100(1 - \bar{P})\%$ всех поставок должен считаться с тем, что товар не соответствует гарантированному качеству. В каждом случае надежность предоставленной гарантии должна быть четко оговорена.

[6.9.] Содержание продажной железной руды должно быть гарантировано $T_g = 22,00\%$ Fe_2O_3 с вероятностью $\bar{P} = 0,95$. Средняя квадратичная ошибка определения железа для этих содержаний дана $\sigma = 0,05\%$ Fe_2O_3 (справочник для металлургической лаборатории). Если в каждой пробе проводят три параллельных определения, то поставщик может отправлять только такую руду, содержание которой по меньшей мере

$$T = 22,00 + \frac{1,65 \cdot 0,05}{\sqrt{3}} = 22,05\% \text{ } Fe_2O_3$$

Для производителя гарантии качества представляют в ряде случаев определенную трудность. Чтобы действительно соблюсти гарантию, поставщик должен постоянно сдавать «слишком хороший» продукт. Поэтому справедливо, если поставщик установит более высокие финансовые требования для подобных гарантий качества.

Продажный продукт можно рассортировать по качеству, поставляя его с гарантиями различной степени надежности. Можно, например, дать гарантию с $\bar{P} = 0,99$, если установленное значение качества T лучше, чем $T_g + 2,33\sigma/\sqrt{n_j}$. Если при проверке качества получают данные T

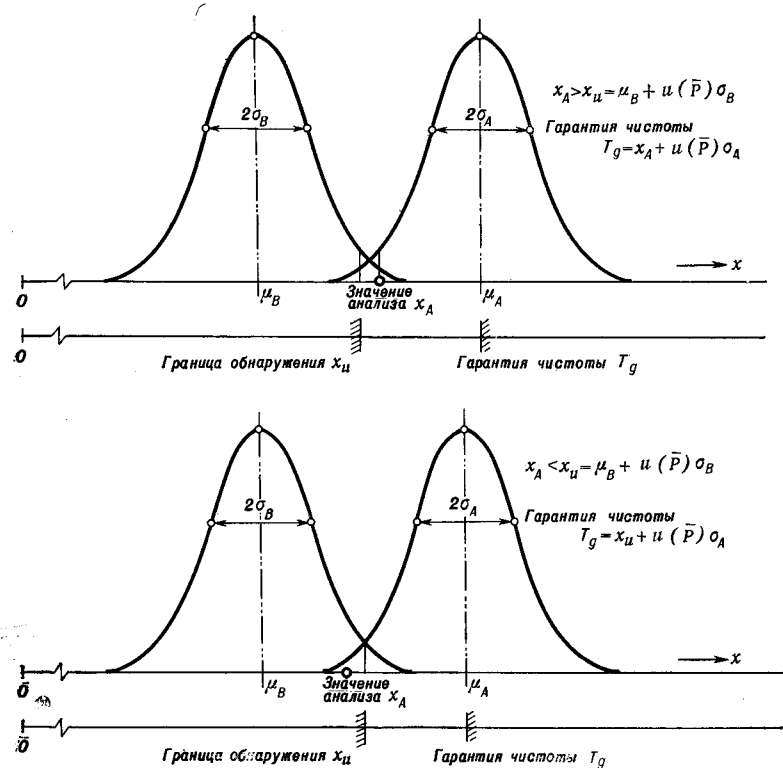
$$T_g + 1,65\sigma/\sqrt{n_j} < T < T_g + 2,33\sigma/\sqrt{n_j}$$

то гарантия возможна с $\bar{P} = 0,95$. Соответствующим образом можно задать пределы для других вероятностей (например, $\bar{P} = 0,90$ или $\bar{P} = 0,80$). Качество продукта следует оценивать тем выше, чем выше надежность данной гарантии. Вследствие этого между поставщиком и потребителем возможно согласование цен. Подобное дифференцирование качества заставляет поставщика добиваться особенно высокого качества продукта, а также позволяет ему разумно использовать партии товара низкого качества. Для потребителя это дифференцирование качества удобно тем, что позволяет подбирать необходимый продукт с учетом его применения.

Подобные обстоятельства имеют место и при получении технических данных о гигиене и технике безопасности. Предписываемая нормой верхняя граница G_H сохраняется с надежностью \bar{P} только тогда, когда значение анализа лежит за нижней границей доверительного интервала. При переходе этой границы следует считаться с возможностью превышения нормы, даже если значение анализа все еще лежит ниже G_H .

Предоставление гарантии качества описанным путем возможно тогда, когда значение измерения лежит достаточно далеко от границы обнаружения (ср. разд. 4.3). При измерениях в непосредственной близости от границы обнаружения возникают дополнительные трудности измерения. Если граница обнаружения установлена с вероят-

ностью \bar{P} , то в $100(1 - \bar{P}) = 100\alpha\%$ случаев существует возможность того, что результаты измерения лежат выше границы обнаружения, хотя исследуемая проба не содержала соответствующего элемента (ср. рис. 6.3). На осно-



Р и с. 6.3. Значение анализа и гарантия чистоты.

вании этих данных поставщик слишком плохо оценивает свой продукт и рискует, что вследствие этого случайно получившегося высокого «холостого» значения потребитель не примет продукт. Более опасен обратный случай. Допустим, что содержание исследуемой пробы лежит немного выше границы обнаружения. Если результат анализа случайно окажется ниже границы обнаружения, то про-

дукт на основании этого случайно полученного значения будет ошибочно принят, как очень хороший. Потребитель на основании этого ошибочно благоприятного анализа рискует получить продукт, который для него непригоден. Для поставщика такой анализ означает опасность последующей рекламации и потери репутации продукта.

При учете риска поставщика на основании результата измерения $x > x_u = \mu_B + u(\bar{P})\sigma_B$ можно дать гарантию, что исследуемая проба содержит такое количество примесей, которое соответствует значению измерения $x + u(P)\sigma_A$ (рис. 6.3). При этом σ_B и σ_A являются средними квадратичными ошибками холостого опыта и соответственно анализа. Это утверждение согласуется с уравнением (6.9). Если анализ дает результат $x < x_u = \mu_B + u(\bar{P})\sigma_B$, то это во всех $100\bar{P}\%$ случаев нельзя считать реальным содержанием вещества в пробе. Однако поскольку граница обнаружения дает наименьшее реально поддающееся учету значение, на основании результата $x < x_u = \mu_B + u(\bar{P})\sigma_B$ можно лишь гарантировать, что соответствующая примесь с вероятностью \bar{P} содержится по крайней мере в той части проб, которая соответствует значению измерения $x_u + u(\bar{P})\sigma_A = \mu_B + u(\bar{P})\sigma_B + u(\bar{P})\sigma_A$. Нельзя гарантировать, что содержание примеси $p = 0$, основываясь на том, что измеренное значение лежит ниже границы обнаружения ($x < x_u$).

Для упрощения дальнейшего изложения следует признать, что средние квадратичные ошибки анализа и холостого опыта равны по величине, т. е. $\sigma_A \sim \sigma_B = \sigma$. Это всегда так, если значения холостого опыта и анализа не очень отличаются по величине. Если поставщик и потребитель идут на одинаковый риск ($\bar{\alpha}_K = \bar{\alpha}_E = \bar{\alpha}$), то при n_j параллельных определениях в анализе ($n_B \gg n_A$) при помощи используемого метода получают условную границу гарантии

$$T_g = \mu_B + \frac{2u(\bar{P})\sigma}{\sqrt{n_A}} \quad (6.10)$$

При значении $x < x_u$ содержание p , относящееся к величине измерения T_g , также следует гарантировать как максимальное загрязнение.

Переход от значений x к содержанию p можно производить по уравнению регрессии (ср. разд. 9.4). На практике используют всего $u = 3,00$ и получают по уравнению (3.10) для риска потребителя и поставщика очень небольшое значение $\bar{\alpha} = 0,0015$. В 0,15% всех случаев возникает необходимость пересортировки продукта.

Случайные отклонения качества не всегда обусловлены случайной ошибкой измерения и метода анализа. Например, на качество часто могут сильно влиять неомогенности проб и ошибки отбора проб. Возможные границы качества рассматриваемого продукта можно правильно оценить только тогда, когда средняя квадратичная ошибка σ включает все эти причины ошибок. Поэтому она должна вычисляться из доброкачественных данных, полученных в достаточно большой промежуток времени. Тогда средняя квадратичная ошибка дает полную картину случайных отклонений качества продукта. Ее следует рассматривать как величину, характеризующую как метод измерений, так и метод анализа.

Часто к задачам аналитика относится непрерывное наблюдение за качеством готового продукта. В химической промышленности такое наблюдение проводят чаще всего как 100%-ный контроль, т. е. анализируют каждую шихту, каждый вагон и т. д. Столь тщательный контроль требует большого штата сотрудников и часто ведет к большому числу более или менее формальных анализов. При переходе от 100%-го к выборочному контролю часто можно получить существенную экономию [2]. При этом методе контроля выбирают ограниченное число проб из всего продукта в определенном интервале и проверяют на данной выборке интересующее состояние качества. Выборочный контроль продолжают до тех пор, пока уровень качества не оказывается ниже нормы. Если обнаруживают какую-нибудь выборку, которая не соответствует требованиям, то немедленно прибегают к 100%-ному контролю. Этот контроль продолжают, пока при установленном числе проб требования качества снова не будут удовлетворены.

При выборочном контроле необходимо гарантировать, что риск получения ошибочного результата из-за пропущенной пробы остается минимальным и что известны

точные границы, внутри которых необходимо наблюдать случайные колебания качества. Поэтому следует предположить, что интересующий уровень качества уже был определен достаточно надежно. Следовательно, известно среднее значение μ_T генеральной совокупности и средняя квадратичная ошибка σ_T , в которую должны быть включены все отклонения процесса производства, включая и сезонные отклонения. Далее принимают, что процесс производства протекает без частых грубых нарушений. При выполнении этих предположений можно задать вероятность \bar{P} того, что отдельное значение качества T не больше, чем предписываемое нормой верхнее граничное значение T_0 . Из уравнения (3.7) получаем

$$\frac{T_0 - \mu_T}{\sigma_T} = u(\bar{P}) \quad (6.11)$$

Соответствующую вероятность \bar{P} (односторонняя поставка задачи) берут из табл. 12.2.

При переходе от 100%-ного к выборочному контролю при n пробах исследуют только одну выборку, остальные $n - 1$ результатов остаются непроверенными. Вследствие этого всегда просматривают только $\frac{100}{n} = k\%$ всего контролируемого продукта.

При таком выборочном контроле естественно есть риск, что при $(100 - k)\%$ непроверенных проб результаты дадут отличное от нормы качество. Этот риск тем меньше, чем ближе друг к другу лежат взятые для проверки выборки. Для k -процентного выборочного контроля риск появления отличных от нормы результатов можно рассчитать по

$$\bar{\alpha}_K = (1 - \bar{P}^{n-1}) \quad (6.12)$$

где $n - 1$ — число непопавших в выборку проб. Данные будут достаточно надежны тогда, когда риск $\bar{\alpha}_K$ принимает очень небольшое значение. Если, например, $\bar{\alpha}_K \leq 0,003$, то это равнозначно тому, что самое большее в трех из тысячи случаев аномальный продукт не замечен контролем. Для практических целей этого более чем достаточно.

В соответствии с сформулированными выше предположениями для выборочного контроля отдельные значения качества должны принадлежать генеральной совокупности со средним значением μ_T и средней квадратичной ошибкой σ_T . Это условие выполняется с вероятностью P , пока отдельные известные значения качества беспорядочно рассеяны внутри границ

$$\mu_T \pm u(P) \sigma_T \quad (6.13)$$

Эти границы называют верхней и нижней контрольными границами. Для них разумно выбрать $P = 0,95$. Текущий контроль выполнения условий, заданных выражением (6.13), проводят графически. Измеренные по выборкам значения, характеризующие качества, упорядочивают во времени на особом бланке — контрольной карте. Данные располагаются в виде последовательности точек, наблюдая которую можно сделать заключения о регулярности процесса производства и о правомочности выборочного контроля. Пока точки беспорядочно рассеяны внутри контрольных границ, выполняется условие (6.13) и можно продолжать выборочный контроль. Однократное нарушение контрольной границы означает, что полученное значение качества с вероятностью P больше не принадлежит генеральной совокупности со средним значением μ_T и средней квадратичной ошибкой σ_T .

Предположения, на которых основан выборочный контроль, больше не выполняются, поэтому немедленно возвращаются к 100%-ному контролю и выясняют причины отклонения. Этот 100%-ный контроль продолжают до тех пор, пока отдельные точки опять не будут беспорядочно рассеяны внутри контрольной границы в течение продолжительного времени.

Если точки неоднократно приближаются к одной или другой контрольной границе, то это является признаком начинающихся изменений либо контролируемого продукта, либо применяемого метода анализа. Тогда целесообразно принять предупредительные меры. Подозрение о систематических изменениях возникает уже тогда, когда значения лежат еще внутри контрольных границ, но чрезмерно часто по одну и ту же сторону от средней линии.

Это особенно справедливо, если

Из 7	следующих	одно за	другим	значений	7
» 11	»	»	»	»	10
» 14	»	»	»	»	12
» 17	»	»	»	»	14
» 20	»	»	»	»	16

значений можно найти выше, либо ниже средней линии. Особое преимущество техники контрольных карт заключается в том, что можно уловить медленно начинающиеся изменения уже тогда, когда они еще не повлияли на результат решающим образом. Несколько типичных контрольных карт показано на рис. 6.4.

Аналитик должен пользоваться контрольными картами непосредственно на рабочем месте. Иногда может возникнуть опасность ошибочной записи. Однако подобные ошибочно заполненные контрольные карты очень легко распознать: если точки, которые должны были бы лежать за пределами контрольных границ, все время заносят внутрь границ, то образуются скопления точек на определенной контрольной границе, что соответствует появлению второго максимума в распределении частот.

[6.10]. Контроль за содержанием железа в технической соляной кислоте должен был производиться 10%-ным выборочным контролем с риском $\bar{\alpha}_{10} = 0,002$ [2]. Результаты предварительного изучения 420 анализов следовали нормальному распределению с $\mu_T = 0,01694\%$ Fe и $\sigma_T = 0,00398\%$ Fe. Положенные нормы допускали максимальное содержание $T_0 = 0,03\%$ Fe. По уравнению (6.11) получают

$$u(\bar{P}) = \frac{0,03000 - 0,01694}{0,00398} = 3,22$$

Из табл. 12.2 получается $Y(x = 3,22) = 0,999359$, а из этого следует, что риск $\bar{\alpha}_{10}$ при 10%-ном выборочном контроле [уравнение (6.12)]

$$\bar{\alpha}_{10} = 1 - (0,999359)^9 = 0,005754$$

Этот риск больше, чем требуемый. Поэтому следовало бы выбрать другую плотность выборок. Для 33%-ного выборочного контроля получается

$$\bar{\alpha}_{33} = 1 - (0,999359)^9 = 0,001282$$

$\bar{\alpha}_h$ значительно меньше тогда, когда при текущем контроле чистоты анализируют каждую третью пробу. Контрольные границы для контрольных карт подсчитывают по уравнению (6.13):

$$\mu \pm u(P = 0,95) \sigma_T = 0,01694 \pm 1,96 \cdot 0,00398$$

Из этого находят контрольные границы: верхнюю 0,0248% Fe и нижнюю 0,00914% Fe. Пока значения выборок беспорядочно рассеяны внутри этих границ, можно использовать 33%-ный выборочный контроль.

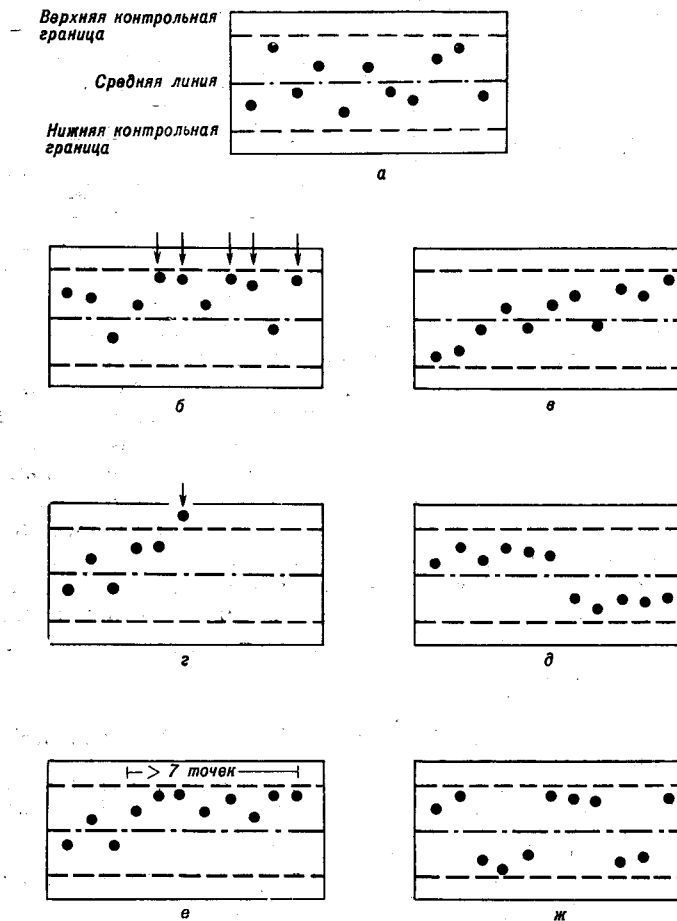
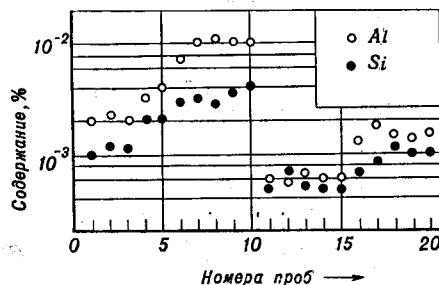


Рис. 6.4. Типы контрольных карт.

а—процесс находится под контролем; б—процесс вскоре выйдет из-под контроля; в—медленное изменение во времени; г—процесс выходит из-под контроля; д—разрыв; е—подозрение о более, чем случайном изменении; ж—периодичность.

Целый ряд продуктов характеризуется несколькими показателями качества. Эти показатели качества могут коррелировать из-за особенностей производственного процесса или в силу того, что они зависят от одних и тех



Р и с. 6.5. Корреляция примесей алюминия и кремния в гидрате окиси натрия.

же исходных веществ. Если изменяется одна из измеряемых величин, то одновременно изменяется и другая (рис. 6.5). Подобное корреляционное соотношение можно с успехом использовать при контроле качества и простым способом контролировать надежность полученных результатов анализа. При достаточно тесной корреляции можно даже выборочно измерять лишь один из показателей качества. Пока он рассеивается в нормальных границах, на основании корреляции следует ожидать нормального поведения и для других показателей качества. Как только проверяемый показатель приближается к критическим границам, следует провести измерение других показателей качества.

Л И Т Е Р А Т У Р А

1. Blankenburg H. J., Jagusch K. H., Z. angew. Geologie, 10, 640 (1964).
2. Felix M., Lemarie M., Chem. Techn., 16, 359 (1964).
3. Jilek M., Likar O., Wiss. Z. Techn. Universität Dresden, 11, 1253 (1962).
4. Kaiser H., Z. anal. Chem., 205, 1 (1965).
5. Schaafsma A. N., Willemze F. G., Moderne Qualitätskontrolle. Philips Techn. Bibl., 1955.
6. Schindowsky E., Schürz O., Statistische Qualitätskontrolle, Berlin, VEB Verlag Technik, 1959.
7. Strauch H., Statistische Güteüberwachung, München, Carl-Hanser-Verlag, 1956.
8. TGL 0-1333: Runden von Meßwerten.
TGL 44449: Begriffe der statistischen Qualitätskontrolle.
TGL 0-51849: Prüfung metallischer Werkstoffe.

7. Статистические методы проверки

Статистические методы проверки дают объективную интерпретацию результатам анализа. Они дают объективный ответ на вопрос, существует ли разница между средними значениями, найденными двумя аналитиками? При этом проверяется *статистическая гипотеза* о принадлежности результатов измерений к одной генеральной совокупности. По результатам, полученным для двух выборок, вычисляют значение некоторой контрольной величины λ и определяют область Λ , внутри которой следует ожидать λ с определенной вероятностью P . Если контрольная величина λ лежит вне области Λ , то выбранная гипотеза отбрасывается. Разница между полученными величинами называется *значимой* или *статистически значимой*. Однако эта разница представляет собой недостаточно надежную меру для оценки различия в тех генеральных совокупностях, к которым относятся результаты измерений. Из статистически значимой, например, разницы для двух средних значений $\bar{x}_1 - \bar{x}_2 = \Delta\bar{x}_{12}$ нельзя сделать вывод, что соответствующие совокупности отличаются на величину $\Delta\bar{x}_{12}$. Если контрольная величина λ находится внутри области Λ , то проверяемая гипотеза принимается. Однако из этого не следует, что гипотеза безусловно подтвердилась. Можно только сказать, что результаты измерений не противоречат проверяемой гипотезе. В этом случае говорят, что различие оказалось *незначимым*. Если установлена статистическая незначимость разности двух величин, то отсюда еще нельзя сделать вывод о равенстве этих величин. Вопрос о том, как такую «незначимую» разницу следует интерпретировать, нужно решить при полном понимании статистических методов проверки гипотез (см. Смирнов и Дунин-Барковский, а также [6]).

Решение об отбрасывании или принятии статистической гипотезы проводят на основании выборочных измерений. Поэтому не следует исключать возможность ошибки. Если, например, с вероятностью P отбрасывают гипотезу о том, что два средних значения x_1 и x_2 принадлежат к одной и той же генеральной совокупности, то из этого можно сделать вывод о различии этих средних значений. Вероятность того, что оба средних значения все же принадлежат к одной и той же генеральной совокупности, будет $\alpha = 1 - P$. Следовательно, можно ввести риск α того, что при использовании критерия $\lambda > \Lambda$ будет отброшена гипотеза, которая в действительности справедлива. Такое ошибочное заключение, возможное во всех $100\alpha\%$ случаев, называют *ошибкой первого рода*. Напротив, может случиться, что, когда $\lambda < \Lambda$, проверяемая гипотеза принимается, хотя она не соответствует действительности. Это ошибочное заключение называют *ошибкой второго рода*.

Допустимый процент возможных ошибок первого рода является вопросом взаимной договоренности. Ложные решения, например, при экспертизе, могут иметь более серьезные последствия, чем ошибочно декларированная чистота реактива. Поэтому в первом случае следует выбирать более высокую надежность и, следовательно, допускать меньшее число возможных ошибок первого рода, нежели второго.

В общем случае часто придерживаются следующих трех правил:

1. Проверяемая гипотеза отбрасывается, если ошибка первого рода может появиться в менее чем $100\alpha = 1\%$ всех случаев (т. е. $P \geq 0,99$). Тогда рассматриваемая разница является значимой.

2. Проверяемая гипотеза принимается, если ошибка первого рода возможна в более чем $100\alpha = 5\%$ всех случаев (т. е. $P \leq 0,95$). Тогда рассматриваемая разница является незначимой.

3. Отбрасываемую гипотезу следует дополнительно обсудить, если число возможных ошибок первого рода лежит в интервале 5 и 1% ($0,95 < P < 0,99$). Рассматриваемую разницу интерпретируют как спорную. Ситуация может проясниться с привлечением последующих изме-

рений. Если по каким-то причинам результатов измерений оказывается недостаточно, то полученные данные следует интерпретировать самым неблагоприятным образом.

В дальнейшем пользуются этими тремя правилами. Однако необходимо было бы еще раз отметить, что выбор α является делом обоюдной договоренности и что другие значения наряду с общеупотребительными также могут быть узаконены. Так, при решении многих внутрипроизводственных вопросов принимают, что проверяемая гипотеза может быть отброшена при числе возможных ошибок первого рода $100\alpha = 10\%$. При этом в особых случаях возникает необходимость практически ограничивать возможность ошибочного решения (например, оценка токсического действия фармацевтического препарата). Тогда проверяемую гипотезу впервые отклоняют при пренебрежимо малом числе возможных ошибок первого рода, например $100\alpha = 0,1\%$.

Результаты статистических методов проверки иногда неудобны для аналитика. Часто они дают незначимую ($P < 0,95$) или спорную ($0,95 < P < 0,99$) разницу, когда различие уже было установлено на основании субъективных суждений. В подобных случаях часто помогают дополнительные измерения: чем больше результатов имеется в распоряжении, тем меньшие разницы могут оказаться надежными. Ни в коем случае нельзя соблазняться заменой точных данных сомнительными на основании интуитивной оценки.

Все описанные в дальнейшем способы проверки сделаны при определенных предположениях о распределении результатов измерений. Если в каком-либо частном случае вид функции распределения не известен, то результаты проверки следует рассматривать с осторожностью, если контрольная величина λ лежит на границах ожидаемой области Λ .

7.1. Сравнение двух средних квадратичных ошибок (F -критерий)

Пусть необходимо сравнить две разные по величине оценки средних квадратичных отклонений s_1 и s_2 со степенями свободы f_1 и f_2 соответственно. Надо решить,

лежит ли разница между s_1 и s_2 в границах возможных случайных колебаний (ср. разд. 5.2), т. е. можно ли оба значения s_1 и s_2 рассматривать как оценку одной и той же дисперсии σ^2 генеральной совокупности. Проверяемая гипотеза, следовательно, будет $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Если это предположение выполняется, то отношение s_1^2/s_2^2 следует F -распределению (ср. разд. 3.3.2) со степенями свободы f_1 и f_2 . Поэтому получают

$$F = \frac{s_1^2}{s_2^2} \quad (7.1)$$

Значение этого отношения всегда должно быть больше единицы, так как большая из двух средних квадратичных ошибок стоит в числителе дроби. (При наличии логарифмического нормального распределения в уравнение (7.1) подставляют средние квадратичные ошибки логарифмов измеряемых величин s_{1g} .) Проверяемую гипотезу $\sigma_1^2 = \sigma_2^2 = \sigma^2$ следует отбросить, если $F > F(\bar{P}, f_1, f_2)$. Между значениями s_1 и s_2 тогда существует значимая разница, так что $\sigma_1^2 > \sigma_2^2$. Найденные значения для средних квадратичных ошибок не противоречат проверяемой гипотезе, если $F < F(\bar{P}, f_1, f_2)$; наблюдаемое различие тогда рассматривают как незначимое. Числовые значения для $F(\bar{P}, f_1, f_2)$ приведены в табл. 12.5. Промежуточные значения интерполируют способом, описанным на стр. 61.

[7.1]. Для методических исследований необходимо было сравнить воспроизводимость двух методов измерения при пламенно-фотометрическом определении натрия. Найденные средние квадратичные ошибки (в относительных процентах) дают следующую картину:

Метод	Средняя квадратичная ошибка	Степень свободы
Первый	$s_1 = 4,3\%$	$f_1 = 11$
Второй	$s_2 = 2,1\%$	$f_2 = 11$

Из уравнения (7.1) получают $F = 4,3^2/2,1^2 = 4,19$. Для $f_1 = 11$ степеней свободы в табл. 12.5 нет числового значения. Для интерполяции графически представляют табулированные значения для $F(\bar{P}, f_1, f_2 = 11)$ в зависимости от $1/f_1$ и находят $F(\bar{P} = 0,95; f_1 = f_2 = 11) = 2,82$ и соответственно $F(\bar{P} = 0,99; f_1 = f_2 = 11) = 4,46$ (рис. 7.1). По существующим правилам (см. стр. 134) не следует принимать решения, так как $F(\bar{P} = 0,95; f_1 = f_2 = 11) <$

$< F < F(\bar{P} = 0,99; f_1 = f_2 = 11)$. Поэтому для метода с меньшей случайной ошибкой — второго метода — были проведены дальнейшие исследования. Была получена средняя квадратичная ошибка $s_2' = 2,4\%$ с $f_2 = 24$ степени свободы. Из уравнения (7.1) имеем $F = 4,3^2/2,4^2 = 3,21$. Интерполяцией, аналогично приведенной на рис. 7.1, находим $F(\bar{P} = 0,99; f_1 = 11; f_2 = 24) = 3,09$. Так

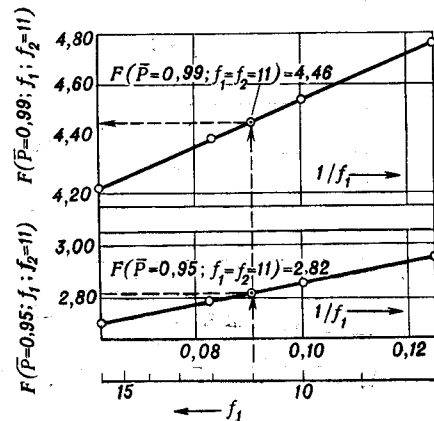


Рис. 7.1. Графическая интерполяция $F(\bar{P}, f_1, f_2)$.

как $F > F(\bar{P} = 0,99; f_1 = 11; f_2 = 24)$, различие в воспроизводимости между обоими методами измерения оказывается установленным с риском менее допустимой ошибки первого рода $100\alpha = 1\%$. Следовательно, можно считать, что второй метод измерений имеет меньшую случайную ошибку.

Имеющаяся между обоими методами анализа небольшая разница в воспроизводимости вначале не была признана значимой из-за малого числа измерений. Лишь при большем числе степеней свободы для меньшей квадратичной ошибки эта разница могла быть надежно установлена (с увеличением числа наблюдений повышается определенность суждений). На это обстоятельство особенно надо обращать внимание, когда отношение двух квадратичных ошибок s_1/s_2 окажется столь неблагоприятным, как это было в первой серии опытов.

Для практических целей эту проверку особенно просто можно провести графическим методом, если обе средние квадратичные ошибки имеют одно и то же число степеней свободы, т. е. $f_1 = f_2 = f$. Соответствующая номограмма показана на рис. 7.2. Находят соотношение $s_1/s_2 = \sqrt{F}$ и отыскивают на графике точку $\Pi(f, \sqrt{F})$. По положению этой точки по отношению к обоим кривым можно

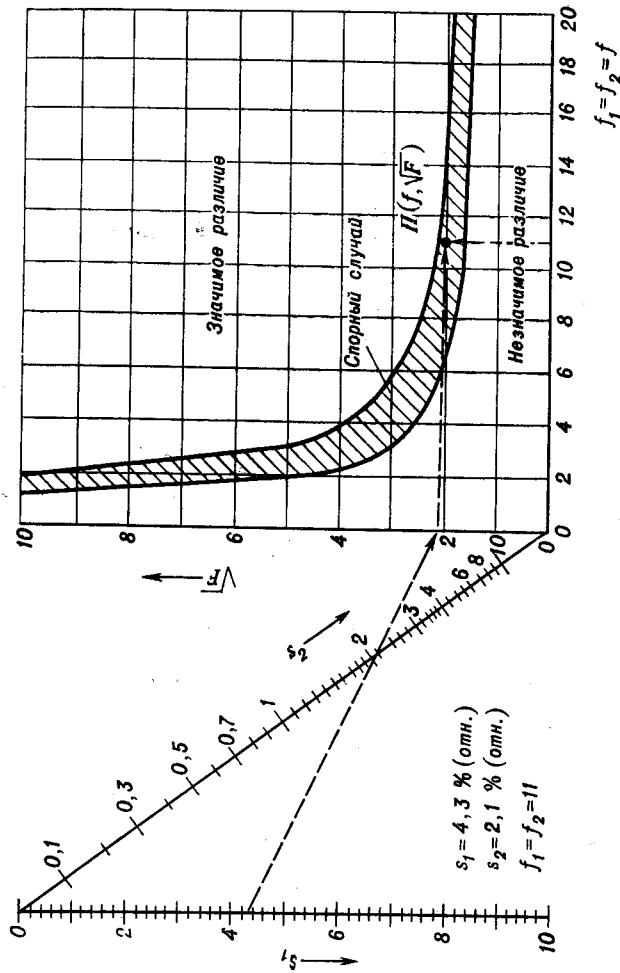


Рис. 7.2. Номограмма для проверки гипотезы с F-критерием.

оценить проверяемую гипотезу. На рис. 7.2 эта графическая проверка проведена со значениями s_1 и s_2 , взятыми из примера [7.1].

Из рис. 7.2 видно, сколь высокое значение должно иметь отношение s_1/s_2 , чтобы можно было оценить различия в квадратичных ошибках ($1\% < 100\bar{\alpha} < 5\%$). При двух сериях измерений с $f_1 = f_2 = 3$ необходимость дополнительной проверки впервые появляется, когда $s_1/s_2 = 3$, а при $f_1 = f_2 = 12$, когда $s_1/s_2 = \sqrt{3}$. Для разницы, значимой в смысле правила трех сигм ($100\alpha < 1\%$), следует, что в первом случае ($f_1 = f_2 = 3$) одна из квадратичных ошибок должна быть в пять раз больше другой, во втором случае ($f_1 = f_2 = 12$) — примерно в два раза больше. Случайную ошибку метода анализа с достаточной точностью можно оценить только из больших серий измерений. Значимость различия особенно сильно зависит от f_2 . Поэтому при подобном сравнении для меньшей средней квадратичной ошибки следует предусмотреть много степеней свободы (ср. пример [7.1]).

Из обширных ранних исследований или из табличных данных может быть известна квадратичная ошибка σ_0 . Интересно узнать, согласуется ли с σ_0 большая величина s со степенями свободы f , найденная при текущих исследованиях. Для этого нужно проверить, есть ли различие между σ_0 и σ для той генеральной совокупности, к которой принадлежит квадратичная ошибка s . Таким образом, проверяемая гипотеза сводится к $\sigma^2 = \sigma_0^2$. Эта гипотеза отбрасывается, если

$$\frac{s^2}{\sigma_0^2} > \frac{\chi^2(\bar{P}; f)}{f} \quad (7.2a)$$

Отбрасывание гипотезы $\sigma^2 = \sigma_0^2$ означает, что квадратичная ошибка генеральной совокупности, соответствующая оценке s , больше чем квадратичная ошибка σ_0 . Если, напротив, существует предположение, что оценке σ соответствует такой средней квадратичной ошибке σ , которая меньше, чем σ_0 , то проверяется та же гипотеза $\sigma^2 = \sigma_0^2$. Эта гипотеза будет отброшена, если

$$\frac{s^2}{\sigma_0^2} < \frac{\chi^2(1-\bar{P}; f)}{f} \quad (7.2b)$$

Если выполняется неравенство (7.26), то s принадлежит генеральной совокупности, средняя квадратичная ошибка которой σ значительно меньше средней квадратичной ошибки σ_0 . Величины $\chi^2(\bar{P}; f)$ ($\bar{P} = 0,95$ или $0,99$), так же как и $\chi^2(1 - \bar{P}; f)$ ($1 - \bar{P} = 0,05$ или $0,01$), необходимые для вычислений, следует брать из табл. 12.4.

[7.2]. По справочнику, составленному для металлургических лабораторий, средняя квадратичная ошибка потенциометрического определения хрома $\sigma_0 = 0,017\%$ для содержания 3% Cr. При исследовании получается несколько более высокое значение $s = 0,024\%$ Cr с $f = 6$. Следует выяснить, имеет ли место в действительности завышение случайной ошибки. Из неравенства (7.2а) с $\chi^2(\bar{P} = 0,95; f = 6) = 12,6$ получается

$$\frac{0,024^2}{0,017^2} = 1,99 < \frac{12,6}{6} = 2,10$$

Следовательно, нет никакого основания отбросить гипотезу $\sigma^2 = \sigma_0^2$. Поэтому нельзя утверждать, что в действительности имеет место завышение ошибки.

7.2. Сравнение нескольких средних квадратичных ошибок (критерий Бартлета)

Имеется m различных по величине, независимых друг от друга оценок средних квадратичных ошибок s_1, s_2, \dots, s_m с соответственно f_1, f_2, \dots, f_m степенями свободы. При этом предполагают, что число степеней свободы в каждой серии больше двух. Необходимо выяснить, следует ли интерпретировать различие между m отдельными средними квадратичными отклонениями только как случайные, т. е. можно ли принять гипотезу о том, что это оценки средней квадратичной ошибки σ одной генеральной совокупности. Следовательно, проверяемая гипотеза будет: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2 = \sigma^2$.

По Бартлету [1] для проверки составляют выражение, приближенно распределенное, как χ^2

$$\chi^2 = 2,303 (f_g \lg s^2 - \sum f_j \lg s_j^2) \quad (7.3)$$

где f_g — общее число степеней свободы; s — средняя квадратичная ошибка, подсчитанная по уравнению (5.1); f_j — число степеней свободы j -й группы ($f_j > 2$); s_j — средняя квадратичная ошибка j -й группы,

Найденная таким образом величина χ^2 сопоставляется с $\chi^2(\bar{P}, f)$ из табл. 12.4. Если имеется m серий измерений, то $\chi^2(\bar{P}, f)$ берется для числа степеней свободы $f = m - 1$.

Проверяемая гипотеза отбрасывается с ошибкой первого рода $100\alpha\% = 100(1 - \bar{P})\%$, если $\chi^2 > \chi^2(\bar{P}, f)$. Это значит, что несколько имеющихся оценок s_j^2 принадлежат генеральным совокупностям, средние квадратичные ошибки которых σ_j^2 больше, чем σ^2 .

Значение, вычисленное по соотношению (7.3), всегда довольно сильно завышено. Если оно превышает значение $\chi^2(\bar{P}, f)$ лишь немного, то χ^2 исправляют в соответствии с соотношением

$$\chi^{*2} = \frac{\chi^2}{C} \quad (7.4)$$

и снова сравнивают. Константу C получают из соотношения

$$C = \frac{\left(\sum \frac{1}{f_j}\right) - \frac{1}{f_g}}{3(m-1)} + 1 \quad (7.5)$$

Лишь когда $\chi^{*2} > \chi^2(\bar{P}, f)$, разницу между отдельными средними квадратичными ошибками следует считать значимой.

[7.3]. Средняя квадратичная ошибка газовольнометрического определения углерода была получена из данных анализа четырех проб близкого содержания, но разного состава. Необходимо проверить, можно ли установить различие между средними квадратичными ошибками.

Проба	Содержание	Средняя квадратичная ошибка	Степень свободы	Тип сплава
1	1,03	0,005% С	24	Cr 1,4%, Si 1,2%, Cr 1,2%, Ферромарганец Нелегированный
2	1,23	0,007% С	32	
3	1,30	0,010% С	28	
4	1,38	0,008% С	32	

Среднюю квадратичную ошибку преобразуют, применяя соотношение $S_j = 1000s_j$, и для вычислений по уравнению (7.3) пользуются следующей схемой:

s_j	s_j^2	f_j	$f_j s_j^2$	$\lg S_j^2$	$f_j \lg S_j^2$
5	25	24	600	1,3979	33,5496
7	49	32	1568	1,6902	54,0864
10	100	28	2800	2,0000	56,0000
8	64	32	2048	1,8062	57,7984
		116	7016		201,4344

$$S^2 = \frac{7016}{116} = 60,48$$

$$\lg S^2 = 1,7816$$

$$116 \lg S^2 = 206,6656$$

$$\chi^2 = 2,303 (206,6656 - 201,4344) = 12,0475$$

Из табл. 12.4 для $f = m - 1 = 3$ находят $\chi^2 (\bar{P} = 0,99; f = 3) = 11,3$. Так как вычисленная величина χ^2 лишь немного превышает табличное значение, повторяют проверку с χ^{*2} . Из уравнения (7.5) получают значение константы C

$$C = \frac{\frac{1}{24} + \frac{1}{32} + \frac{1}{28} + \frac{1}{32} + \frac{1}{116}}{3(4-1)} + 1 = 1,0146$$

По уравнению (7.4) получают

$$\chi^{*2} = \frac{12,0475}{1,0146} = 11,8740 \sim 11,87$$

Проверка с χ^{*2} ничего не изменила в первоначальном результате; между четырьмя средними квадратичными ошибками существует значимое различие. Возникает подозрение, что это различие вызвано пробой ферромарганца (проба 3) с более высокой средней квадратичной ошибкой $s_3 = 0,010\%$ С. Поэтому повторяют проверку без s_3 . При этом получается $\chi^2 = 5,63$ и $\chi^2 (\bar{P} = 0,95; f = 2) = 5,99$. Теперь между тремя средними квадратичными ошибками s_1, s_2 и s_4 нет никакого различия*.

* Критерием Бартлета надо пользоваться с большой осторожностью. Он очень чувствителен даже к небольшому отклонению от нормального распределения. — *Прим. ред.*

7.3. Сравнение двух средних значений (t -критерий)

Пусть даны два средних значения \bar{x}_1 и \bar{x}_2 , которые получены из двух независимых друг от друга серий измерений с n_1 и n_2 . Оба средних значения различаются на незначительную величину. Следует проверить, объясняется ли это различие только случайной ошибкой, т. е. принадлежат ли оба средних значения генеральной совокупности с одним и тем же средним значением μ . Проверяемая гипотеза будет: $\mu_1 = \mu_2 = \mu$. Предварительно необходимо исследовать, существует ли разница между средними квадратичными ошибками обеих серий s_1 и s_2 .

Если F -критерий (ср. разд. 7.1) дает значимое различие, то эти два средних значения нельзя сравнивать между собой. Для проверки прежде всего по закону распространения ошибок вычисляют среднюю квадратичную ошибку для разности двух средних значений из n_1 и n_2 измерений. Из уравнений (4.3а) и (3.3) следует

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

В соответствии с уравнением (5.1) можно также записать

$$s_{\bar{x}_1 - \bar{x}_2} = s \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (7.6)$$

со степенями свободы $f = n_1 + n_2 - 2$.

Разности $|\bar{x}_1 - \bar{x}_2|$ являются случайными величинами и при обычно имеющемся малом числе измерений следуют t -распределению (ср. разд. 3.3.1). Для того чтобы оценить вероятность появления статистически значимой разности $|\bar{x}_1 - \bar{x}_2|$, эту величину нормируют делением на $s_{\bar{x}_1 - \bar{x}_2}$ и получают

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (7.7)$$

Вычисленную по уравнению (7.7) величину сравнивают с пределами интегрирования t -распределения $t(P, f)$ (табл. 12.3). Проверяемую гипотезу $\mu_1 = \mu_2 = \mu$

следует отбросить с ошибкой первого рода $100\alpha = 100(1 - P)\%$, если $t > t(P, f)$. Между двумя средними значениями x_1 и x_2 существует в этом случае значимое различие. Различие между двумя средними значениями будет незначимым, если $t < t(P, f)$.

[7.4]. Две производственные группы должны были микроаналитически определить азот в данной пробе (цинхонин). Были найдены следующие значения (% N):

I группа	II группа
9,29	9,53
9,38	9,48
9,35	9,61
9,43	9,68
Среднее $\frac{9,36_3}{}$	$\frac{9,57_5}{}$

Необходимо проверить, существует ли различие между обоими средними значениями. Результаты измерений преобразуют, пользуясь соотношением $X_i = 100x_i - 952$ и получают

I	II
-23	+1
-14	-4
-17	+9
-9	+16
$\Sigma X_{1i} = -63$	$\Sigma X_{2i} = +22$
$\bar{X}_1 = -15,8$	$\bar{X}_2 = +5,50$

Для сравнения ошибок обеих серий вычисляют отдельно средние квадратичные ошибки по уравнению (2.5):

$$QS_1 = 23^2 + 14^2 + 17^2 + 9^2 - \frac{63^2}{4} = 102,75$$

$$S_1^2 = 34,25; S_1 = 5,85; f_1 = 3$$

$$QS_2 = 1^2 + 4^2 + 9^2 + 16^2 - \frac{22^2}{4} = 233,00$$

$$S_2^2 = 77,66; S_2 = 8,80; f_2 = 3$$

Проверка с помощью F -критерия [уравнение (7.1)] приводит к

$$F = \frac{77,66}{34,25} = 2,27$$

$$F(\bar{P} = 0,95; f_1 = f_2 = 3) = 9,28$$

Так как $F < F(\bar{P}, f_1, f_2)$, между обоими средними квадратичными ошибками не существует значимого различия. Поэтому соответственно уравнению (7.6) получают

$$S = \sqrt{\frac{34,25 + 77,66}{4 + 4 - 2}} = 4,32$$

Оба средних значения проверяют по уравнению (7.7)

$$t = \frac{|(-15,8) - (+5,5)|}{4,32} \sqrt{\frac{4 \cdot 4}{4 + 4}} = 6,97$$

Из табл. 12.3 находят $t(P = 0,99; f = 6) = 3,71$. Так как $t > t(P, f)$, между обоими средними значениями оказывается значимое различие. Средние значения различаются сильнее, чем это допускает случайная ошибка внутри обеих серий анализа. По крайней мере в одной из двух групп должна появляться систематическая ошибка.

Бывают случаи, когда необходимо проверить отклонение среднего значения \bar{x} от безошибочного числа μ_0 (например, теоретически выведенной величины или теоретически высчитанного содержания). Тогда проверяемой гипотезой будет $\mu = \mu_0$ и уравнение (7.7) в этом случае переходит в

$$t = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n_j} \quad (7.8)$$

где μ_0 — теоретически выведенное значение; n_j — число параллельных определений; s — средняя квадратичная ошибка в соответствии с уравнением (2.5) и с $f = n_j - 1$ степеней свободы.

В этом случае проверка также производится сопоставлением вычисленного по уравнению (7.8) значения t с приведенными в таблице пределами интегрирования.

[7.5]. Теоретическое содержание азота в соединении, исследованном в примере [7.4] (цинхонине), составляет $\mu_0 = 9,517\%$ N. При помощи использованного в примере [7.4] преобразования получается $M_0 = 100 \mu_0 - 952 = -0,3$. По уравнению (7.8) вычисляют:

$$t_1 = \frac{|-15,8 - (-0,3)|}{5,85} \sqrt{4} = 5,30$$

$$t_2 = \frac{|+5,5 - (-0,3)|}{8,80} \sqrt{4} = 1,32$$

$$t(P = 0,95; f = 3) = 3,18$$

$$t(P = 0,99; f = 3) = 5,84$$

Так как $t_1 > t(P = 0,95; f)$, следует предположить, что первой группой было найдено ошибочное значение. У второй группы отклонение от теоретического значения могло быть только случайным, так как $t_2 < t(P = 0,95; f)$

Пример [7.5] оказался особенно благоприятным для обнаружения ошибочной серии анализов, так как там было известно теоретически определяемое содержание исследуемого соединения. Если подобная проверка невозможна, решение следует принять после проведения третьей, независимо полученной серии анализов.

Описанный метод прежде всего пригоден для проверки разности между двумя средними значениями тогда, когда можно предположить, что имеет место нормальное и соответственно t -распределение. Однако ранее было показано (см. стр. 46), что средние значения для $n_j > 5$ параллельных определений часто все еще приближенно следуют нормальному распределению, если даже для отдельных значений это не выполняется. Если подлежащие проверке средние значения \bar{x}_1 и \bar{x}_2 получены из достаточно большого числа отдельных измерений, то можно применять t -распределение, если неизвестны функции распределения отдельных значений.

Применению t -критерия (особенно для неточных наблюдений) часто мешают трудности, связанные с расчетом средней квадратичной ошибки s . Поэтому в литературе для подобных приближенных оценок применяют метод проверки, который использует в качестве оценки для случайной ошибки размах варьирования R (*псевдо- t -критерий*).

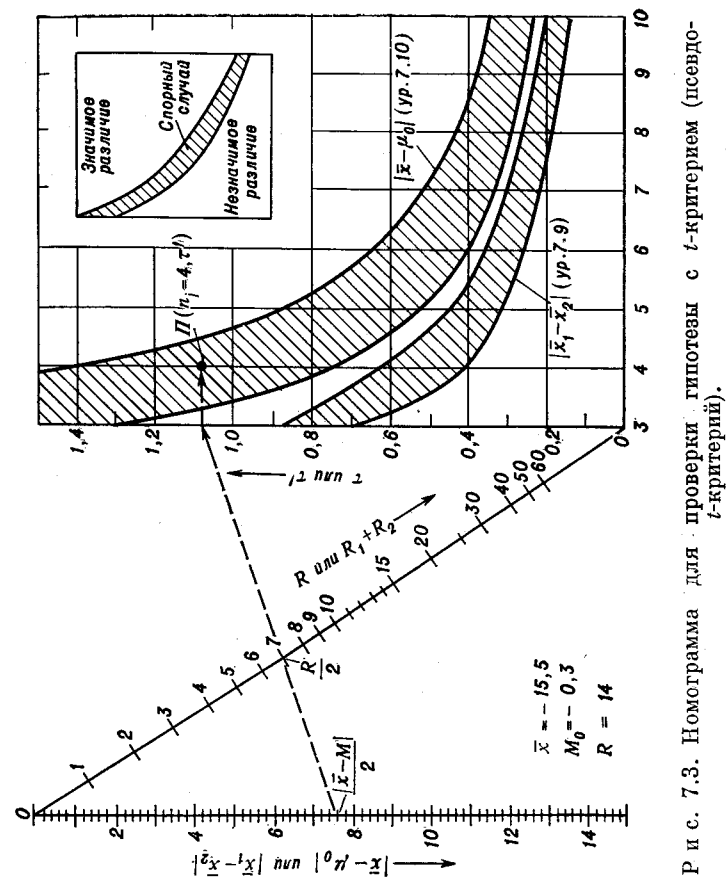
Когда два средних значения взяты из серий измерений равного объема и, следовательно $n_1 = n_2 = n_j$, по аналогии с уравнением (7.7) подсчитывают

$$\tau = \frac{|\bar{x}_1 - \bar{x}_2|}{R_1 + R_2} \quad (7.9)$$

и сравнивают τ таким же образом с теоретически найденной величиной $\tau(P, n_j)$. Если нужно проверить наличие разности между средним значением и теоретическим значением μ_0 , то при n_j параллельных определениях получают по аналогии с уравнением (7.8)

$$\tau' = \frac{|\bar{x} - \mu_0|}{R} \quad (7.10)$$

и сравнивают τ с τ' (P, n_j). Эти два метода проверки можно особенно просто выполнять при помощи номо-



Р и с. 7.3. Номограмма для проверки гипотезы с t -критерием (псевдо- t -критерий).

граммы, которая приведена на рис. 7.3. На вертикальной оси номограммы откладывают значения τ или τ' и затем отыскивают в номограмме точку $\Pi [n_j; \tau (P, n_j)]$ и (или) соответственно $\Pi [n_j; \tau' (P, n_j)]$. По положению этой точки по отношению к обеим кривым на номограмме оценивают значимость разности средних значений. На рис. 7.3 показана эта графическая проверка данных группы I из примера [7.5]. Здесь числовые значения не укладываются на вертикальной оси, поэтому как для разности $\bar{x} - \mu_0$, так и для размаха варьирования R числовые значения были уменьшены вдвое.

7.4. Сравнение частот

Различие между двумя отсчетами x_1 и x_2 , которые следуют распределению Пуассона, можно аналогичным образом интерпретировать как разность между двумя средними значениями. В предположении, что $x_1 > 15$ и $x_2 > 15$, имеющееся распределение Пуассона можно приближать нормальным распределением (ср. разд. 3.2). Проверяют, принадлежат ли обе вычисленные частоты двум генеральным совокупностям с одним и тем же параметром x , следовательно, справедливо ли утверждение, что $x_1 = x_2 = x$. Дальше предполагают, что x_1 и x_2 представляют абсолютные значения, полученные за два интервала времени T_1 и T_2 , а не относительные величины (скажем, величины, отнесенные к единице отсчета — минуте). При этом предположении получают нормально распределенную величину

$$u = \frac{|x_1 T_2 - x_2 T_1|}{\sqrt{T_1 T_2 (x_1 + x_2)}} \quad (7.11)$$

Сравнением с табличным $u(P)$ можно обычным путем оценить значимость разности между обоими числовыми результатами.

[7.6]. Найденная активность двух α -препаратов $\nu_1 = 17$ и $\nu_2 = 13$ *имп/мин*. Первый результат подсчитан счетчиком за 6 *мин*, второй — за 7 *мин*. Результаты опыта можно представить следующей схемой:

Проба	Частота импульсов	Время счета	Число импульсов
1	$\nu_1 = 17$	$T_1 = 6$	$x_1 = 102$
2	$\nu_2 = 13$	$T_2 = 7$	$x_2 = 91$

В соответствии с уравнением (7.11) получают

$$u = \frac{102 \cdot 7 - 91 \cdot 6}{\sqrt{6 \cdot 7 (102 + 91)}} = 1,87$$

Из табл. 12.2 находят $u(P = 0,95) = 1,96$. При этом разность между обоими результатами недостаточно велика, чтобы можно было на ее основании считать различие значимым.

Если приближение распределения Пуассона нормальным распределением невозможно из-за малого числа результатов ($x_1 < 15$; $x_2 < 15$), то можно провести проверку при помощи следующего F -критерия. Если T_1 и T_2 являются двумя интервалами счета, то получают

$$F = \frac{T_2(2x_1 + 1)}{T_1(2x_2 + 1)} \quad (7.12)$$

При этом предполагают, что должно выполняться $T_2(2x_1 + 1) > T_1(2x_2 + 1)$, т. е. значение дроби должно быть больше единицы. Подсчитанные частные сравнивают с пределами интегрирования F -распределения $F(\bar{P}, f_1, f_2)$ при $f_1 = 2x_1 + 1$ и $f_2 = 2x_2 + 1$ (табл. 12.5). Различие считается значимым, если $F > F(\bar{P}, f_1, f_2)$.

[7.7]. Обогащение цирконом ($ZrSiO_4$) фракции тяжелых минералов определяется подсчетом флуоресцирующих зерен циркона. Было подсчитано 500 вспышек. В двух пробах было получено $x_1 = 15$ и $x_2 = 9$ флуоресцирующих зерен. При применении уравнения (7.12) нашли

$$F = \frac{500(2 \cdot 15 + 1)}{500(2 \cdot 9 + 1)} = 1,63$$

Из табл. 12.5а интерполяцией находят $F(\bar{P} = 0,95; f_1 = 31; f_2 = 19) = 2,07$. Так как $F = 1,63 < F(\bar{P} = 0,95; f_1; f_2)$, то нет значимого различия между обоими результатами.

Как уже было показано ранее (ср. пример [6.6]), и здесь при небольшом числе значений критерий сравнения реагирует только на большие расхождения. Поэтому, безусловно, требуется применение статистических методов, если из результатов хотят получить обоснованные выводы.

7.5. Определение грубых ошибок

При многократном повторении одного измерения, значение этого измерения иногда особенно сильно отклоняется в одну или другую сторону без достаточного

основания. Тогда следует решить, идет ли речь лишь о случайном, особенно большом отклонении или о настоящей «грубой ошибке», которая может быть сглажена при дальнейшей обработке числовых данных или которую лучше исключить, анализируя повторяющиеся измерения. В аналитической химии чаще всего речь идет о сериях с малым числом измерений, поэтому определение грубых ошибок лучше оценивать при помощи размаха варьирования. Для этого составляют отношение

$$Q = \frac{|x_1 - x_2|}{R} \quad (7.13)$$

где x_1 — подозрительно выделяющееся значение; x_2 — соседнее значение; R — размах варьирования.

Вычисленную величину Q сопоставляют с табличным значением $Q(\bar{P}, n_j)$ (табл. 7.1). Наличие грубой ошибки действительно доказано, если $Q > Q(\bar{P}, n_j)$.

Таблица 7.1

Числовые значения
для $Q(P, n_j)$ [4]

n_j	$\bar{P} = 0,90$	$\bar{P} = 0,95$	$\bar{P} = 0,99$
3	0,89	0,94	0,99
4	0,68	0,77	0,89
5	0,56	0,64	0,76
6	0,48	0,56	0,70
7	0,43	0,51	0,64
8	0,40	0,48	0,58

[7.8]. При определении графита в сером чугуне получены следующие упорядоченные по величине значения (% графита):

2,86
2,89
2,90
2,91
2,99

Значение $x_5 = 2,99\%$ подозрительно велико. В соответствии с уравнением (7.13) составляют

$$Q = \frac{2,99 - 2,91}{2,99 - 2,86} = 0,62$$

Из табл. 7.1 находят $Q(\bar{P} = 0,95; n_j = 5) = 0,64$. Так как $Q < Q(\bar{P}, n_j)$, можно считать, что подозрительное значение не оказывается грубой ошибкой. При последующей оценке его следует принимать во внимание наряду с другими результатами.

В качестве другого метода проверки для контроля грубой ошибки можно применять просто метод номо-

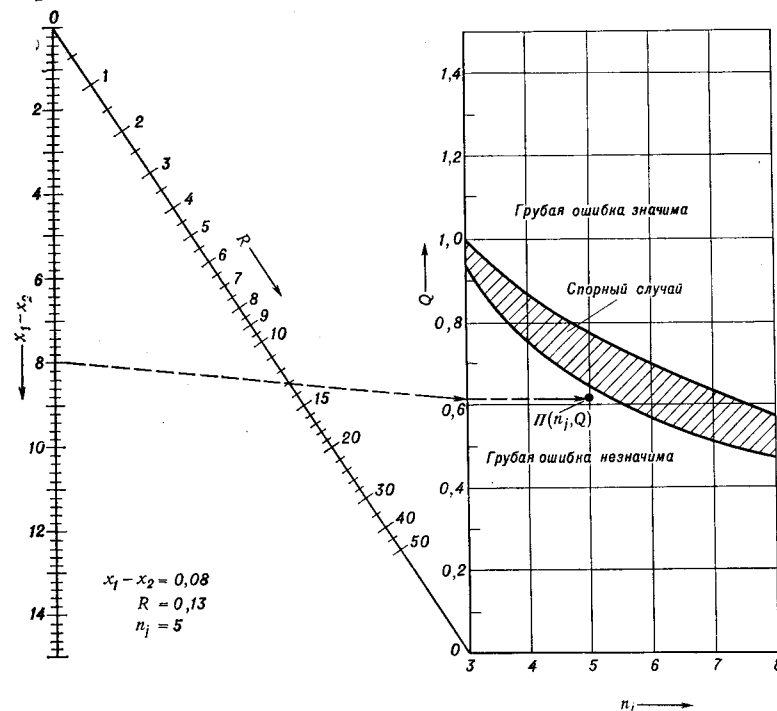


Рис. 7.4. Номограмма для выявления грубой ошибки.

граммы. На рис. 7.4. показано применение этого метода для числовых значений примера [7.8].

Описанный здесь метод выявления грубой ошибки будет недостаточно чувствителен, если имеется большее число измерений, так как в этом методе проверки используется только подозрительное значение и два других значения ряда измерений. При больших же сериях измерений эффективнее определение грубых ошибок, описан-

ное Графом и Хеннингом [5], которое применимо для значений $4 < n < 1000$. Для проверки гипотезы о наличии грубой ошибки вычисляют арифметическое среднее и среднюю квадратичную ошибку из полученных измерений, не включая подозрительно выделяющегося значения. Если имеется более десяти измерений, то пропущенное значение является доказанной грубой ошибкой, если оно удалено от среднего значения более чем на $4s^*$.

7.6. Исследование эмпирических распределений **

Все приведенные здесь статистические методы предполагают выполнение определенных теоретических функций распределения — нормального распределения или распределения Пуассона. Ранее уже было показано (см. гл. 3), каким образом можно сравнивать полученные эмпирические функции распределения с их теоретической моделью, однако тогда не было меры для оценки адекватности этого приближения. При помощи статистических методов проверки теперь можно дать объективную оценку ситуации.

Для проведения подобной проверки прежде всего выдвигают предположения, что между эмпирическим и теоретическим распределениями не существует никакого различия. Выборку из n значений делят на m классов, причем должно быть $m \sim \sqrt{n}$. Для каждого такого класса определяют абсолютную частоту h содержащихся в нем значений измеряемой величины и сопоставляют ее с частотой h_t , теоретически ожидаемой в соответствии с моделью. Для разных теоретических распределений ее табулируют для случая $\sigma = 1$. Поэтому прежде всего для расчета теоретической частоты нормируют ширину класса по $\frac{x - \mu}{\sigma}$. Для этих нормированных значений в соответствующей таблице (табл. 12.1) находят относящиеся к ним значения ординаты. Для имеющихся n измерений, при ширине класса d и средней квадратичной ошибке

* Оценка грубых ошибок — это весьма деликатная задача. Ей посвящена большая литература. См., например, обзор Н. Г. Микешиной, Заводская лабор., № 3, 310—318 (1966). — Прим. ред.

** Вопрос выбора подходящей функции распределения требует специального внимания. Этому целиком посвящена находящаяся в печати монография Г. Хана и С. Шапиро «Статистические модели в инженерных задачах», «Мир», М., 1969. — Прим. ред.

вычисляют теоретически ожидаемую абсолютную частоту h_t заполнения отдельных классов. Из найденных и подсчитанных частот составляют выражение

$$\chi^2 = \sum \frac{(h - h_t)^2}{h_t} \quad (7.14)$$

Если теоретически найденное значение h_t для отдельных классов достаточно большое ($h_t > 5$, также ср. [3]), то найденное значение будет следовать χ^2 -распределению с числом степеней свободы $r = m - k$. При этом k задается числом параметров, необходимых для характеристики выборки. Для нормального распределения $k = 3$ (среднее значение \bar{x} , средняя квадратичная ошибка s и объем выборки n). Для распределения Пуассона $k = 2$ (среднее значение \bar{x} и объем выборки n). Необходимое для отдельных классов значение $h_t > 5$ можно получить, объединяя несколько ниже расположенных классов.

Если при проверке получается, что $\chi^2 > \chi^2(\bar{P}, f)$, то проверяемая гипотеза отбрасывается, т. е. между эмпирическим и теоретическим распределениями существует значимая разница. Различие незначимо, если $\chi^2 < \chi^2(\bar{P}, f)$.

Вычисление теоретических частот и величины χ^2 ведут по схеме, указанной в примере [7.9]. Значения ординат нормального распределения $\phi(u)$ следует брать из табл. 12.1. Соответствующие значения для распределения Пуассона можно брать из статистических таблиц, если не хотят применять метод, описанный ниже,

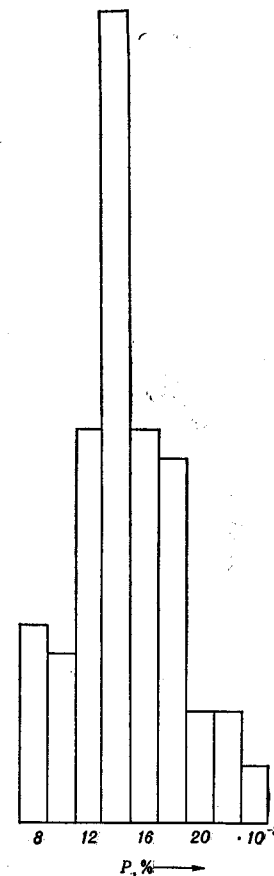


Рис. 7.5. Эмпирическое распределение частот при определении фосфора.

[7.9]. При проведении совместного опыта результаты измерений дали заостренное распределение частот с возможно несколько асимметричным максимумом (рис. 7.5). Поэтому следует проверить, согласуются ли эмпирически найденное распределение с нормальным.

По результатам измерений вычисляют арифметическое среднее \bar{x} и среднюю квадратичную ошибку s . Имеется достаточное число измерений, поэтому можно обе эти оценки приравнять соответствующим величинам генеральной совокупности μ и σ . Для подсчета величины снова надлежащим образом преобразуют. Получается:

	Экспериментально найденные	Преобразованные в соответствии с $X_i = 10\,000 x_i$
Среднее (μ)	0,0147% P	147
Средняя квадратичная ошибка (σ)	0,0037% P	37
Ширина класса (d)	0,0020% P	20

С этими преобразованными значениями проводят проверку по указанной ниже схеме. После объединения трех последних классов число классов составит $m = 7$. Число степеней свободы $f = 7 - 3 = 4$. Из табл. 12.4 находят $\chi^2(\bar{P} = 0,95; f = 4) = 9,49$. Так как $\chi^2 > \chi^2(P, f)$, между имеющимся эмпирическим распределением и нормальным распределением существует значимое различие.

X	h	$u = \frac{X - \mu}{\sigma}$	$\Phi(u)$ (табл. 12.1)	$h_t = \frac{nd}{\sigma}$	$\Phi(u) \frac{(h - h_t)^2}{h_t}$
80	7	1,81	0,0775	3,8959	2,4732
100	6	1,27	0,1781	8,9531	0,9741
120	14	0,73	0,3056	15,3626	1,2086
140	29	0,19	0,3918	19,6959	4,3951
160	14	0,35	0,3752	18,8614	1,2530
180	13	0,89	0,2685	13,4976	0,0183
200	4	1,43	0,1435	7,2138	10,9539
220	4	1,97	0,0573	2,8805	
240	2	2,51	0,0171	0,8596	

$n = 93$

91,2203

$\chi^2 = 10,4054$
 $\sim 10,41$

Проверку различия между эмпирическим распределением и распределением Пуассона можно проводить соответствующим образом. Проще можно проводить эту проверку, если имеется большое число исследуемых проб ($m > 20$).

Из имеющихся m результатов вычисляют арифметическое среднее \bar{x} и по уравнению (2.5) среднюю квадратичную ошибку s со степенями свободы $f = m - 1$. Эту среднюю квадратичную ошибку сопоставляют со средней квадратичной ошибкой, полученной теоретически из $\sigma = \sqrt{\bar{x}}$. Сравнение обеих средних квадратичных ошибок проводят при помощи F -критерия [уравнение (7.1)]. В результате получают

$$F = \frac{s^2}{\bar{x}} \quad (7.15)$$

($F > 1$), и сравнивают как обычно с $F(\bar{P}, f_1, f_2)$ при $f_1 = m - 1$ и $f_2 = \infty$. Предположение о распределении Пуассона следует отклонить, если $F > F(\bar{P}, f_1, f_2)$.

[7.10]. Из $m = 100$ примера [3.4] находят $\bar{x} = 3958$ *имп* [по уравнению (2.1)]; $s = 71$ *имп* [по уравнению (2.5)]. Отсюда получается

$$F = \frac{71^2}{3958} = 1,27$$

Из табл. 12.5а интерполяцией находят $F(\bar{P} = 0,95; f_1 = 99; f_2 = \infty)$. Так как $F < F(\bar{P}, f_1, f_2)$, то нет значимого отличия от распределения Пуассона.

ЛИТЕРАТУРА

- Bartlett M. S., Proc. Roy. Soc., A 160, 268 (1937).
- Bauer E. L., A Statistical Manual for Chemists, N.Y. & London, Academic Press, 1960.
- Cochran W. G., Biometrics, 10, 417 (1954).
- Dean R. B., Dixon W. J., Anal. Chem., 23, 636 (1951).
- Graf U., Henning H.-J., Mitteilungsblatt math. Statistik, 4, 1 (1952).
- Köb H., Textil.-Praxis, 33, 231, 367 (1960).
- Lord E., Biometrika (London), 34, 66 (1947).
- Ostle B., Statistics in Research, Amsterdam — N.Y., Verlag Feffer & Simons Inc.
- Handbuch für das Eisenhüttenlaboratorium, Berlin-Göttingen-Heidelberg, Springer-Verlag, 1955.

8. Простой дисперсионный анализ

Рассмотренные до сих пор вопросы касались определенных частных случаев. Так, при подсчете и применении средней квадратичной ошибки или доверительного интервала предполагалось, что мог быть лишь один-единственный источник ошибок, задаваемый методом анализа. Сравнение средних значений посредством t -критерия ограничивалось случаем только двух серий измерений. Решение этой проблемы на неоднородном числовом материале, при котором появляется более чем одна причина ошибок (например, ошибка отбора пробы и ошибка анализа), а также сравнение более чем двух средних значений возможно при помощи *простого дисперсионного анализа*. Его применение предполагает нормальное распределение цифровых данных, отдельные значения которых получаются независимо одно от другого.

Дисперсионный анализ чувствителен по отношению к отклонениям от нормального распределения. Поэтому дисперсионный анализ можно применять только к подходящим образом преобразованным величинам (см. книгу Вебера).

8.1. Несколько источников случайной ошибки

Случайная ошибка метода анализа характеризуется средней квадратичной ошибкой. Эту величину определяют из ряда повторяющихся независимых измерений на гомогенном материале пробы. Предполагается, что величина этой ошибки не изменится, если повторить опыты при одинаковых условиях в любой лаборатории при тех же предположениях. На этом основании эту величину называют *средней квадратичной ошибкой воспроизводимости* s_W (TGL 0-51849).

В целях сравнения часто следует анализировать одну и ту же гомогенную пробу в разных лабораториях. При

этом в каждой лаборатории необходимо проводить ряд параллельных определений. Из-за незначительного различия в приемах работы результаты, полученные в отдельных лабораториях, имеют небольшие систематические отклонения. Это можно увидеть, например, рассматривая распределение частот в примере [2.1]. Здесь результаты отдельных лабораторий лежат очень близко, однако на диаграмме ясно видны отличающиеся одна от другой группы. Эти систематические ошибки изменяются от лаборатории к лаборатории, и вследствие этого они оказываются дополнительной причиной неустойчивости, увеличивающей случайную ошибку метода (ср. гл. 1). Эта суммарная случайная ошибка образуется от совместного действия межлабораторной ошибки и ошибки воспроизводимости. В стандарте TGL 0-51849 она называется *ошибкой сопоставимости* s_V^* . Если в каждой участвующей лаборатории проведено n_j параллельных определений, то среднюю квадратичную ошибку сопоставимости получают из соотношения

$$s_V^2 = s_W^2 + n_j s_L^2 \quad (8.1)$$

s_L — межлабораторная ошибка.

Для одновременного вычисления s_W и s_V используют простой дисперсионный анализ. Имеющийся цифровой материал распределяют на m отдельных групп, соответственно их поступлению из m отдельных лабораторий. Случайные ошибки, определяемые внутри этих групп, должны быть равны по величине. Это определяют при помощи критерия Бартлета (ср. разд. 7.2). Если обнаруживается статистически значимое различие в ошибках воспроизводимости, то данные следует объединять в группы с одинаковой воспроизводимостью. Величины, необходимые для дисперсионного анализа (суммы квадратов, степени

* По стандарту TGL 0-51849 устанавливают такие условия при изучении воспроизводимости, чтобы один и тот же наблюдатель повторял измерения с одним и тем же прибором и с частью одной и той же пробы при возможно одних и тех же условиях. При изучении сопоставимости различные наблюдатели в разных лабораториях проводят измерения подлежащей исследованию пробы на приборах заданного типа. Понятие сопоставимости не следует употреблять чисто механически, в каждом случае нужно точно определить условия сравнения.

свободы, дисперсии), подсчитывают по следующей схеме (с целью упрощения обозначения s_V , s_W и s_L заменены на s_1 , s_2 и s^*):

Причины	Сумма квадратов	Степень свободы	Дисперсия	Компоненты дисперсий
Разброс между m -группами	$QS_1 = n_j \sum (\bar{x}_j - \bar{\bar{x}})^2$	$f_1 = m - 1$	$s_1^2 = \frac{QS_1}{f_1}$	$s_1^2 = s_2^2 + n_j s^{*2}$
Разброс внутри m -групп	$QS_2 = \sum \sum (x_{ji} - \bar{x}_j)^2$	$f_2 = n - m$	$s_2^2 = \frac{QS_2}{f_2}$	—
Общий разброс	$QS = QS_1 + QS_2 = \sum (x_{ji} - \bar{\bar{x}})^2$	$f = f_1 + f_2 = n - 1$	—	—

Суммы квадратов «внутри m -групп» и «общую» вычисляют по уравнению (2.6а). Суммы квадратов «между m -группами» вычисляют для средних значений по группам \bar{x}_j по формуле

$$\sum n_j (\bar{x}_j - \bar{\bar{x}})^2 = \sum_1^m \frac{(\sum_1^{n_j} x_{ji})^2}{n_j} - \frac{(\sum_1^m \sum_1^{n_j} x_{ji})^2}{n}. \quad (8.2)$$

Прежде всего проверяют гипотезу о том, что между s_1^2 и s_2^2 нет значимого различия (нуль-гипотеза $\sigma_1^2 = \sigma_2^2$). Это равнозначно тому, что в уравнении (8.1) величина $s_L^2 = 0$. Для проверки нуль-гипотезы составляют отношения в соответствии с уравнением (7.1)

$$F = \frac{s_1^2}{s_2^2},$$

При этом дисперсия «между сериями» (s_1^2) всегда стоит в числителе дроби. Нуль-гипотеза принимается, если

$F < F(\bar{P}, f_1, f_2)$. В этом случае материал считают однородным и объединяют суммы квадратов обеих составных частей ошибки и этим увеличивают число степеней свободы. Если нуль-гипотезу следует отбросить [$F > F(\bar{P}, f_1, f_2)$], то между s_1 и s_2 оказывается значимое различие; тогда компонента дисперсии s^{*2} отличается от нуля и цифровой материал следует считать неоднородным. По уравнению (8.1) можно вычислить отдельные компоненты дисперсии, если опыт был «симметричным», т. е. все серии состояли из одинакового числа измерений. Доверительные границы для s_1 и s_2 можно подсчитать по уравнению (5.5).

Если необходимо указать доверительный интервал для среднего значения \bar{x} , полученного из неоднородного цифрового материала, то в основе должна лежать обусловленная негомогенностью средняя квадратичная ошибка s_1 . Получают

$$\Delta \bar{x} = \frac{t(P, f_1) s_1}{\sqrt{n}} \quad (8.3)$$

Аналогично в подобном случае получают доверительный интервал среднего значения серии из n_j параллельных определений

$$\Delta \bar{x}_j = \frac{t(P, f_2) s_2}{\sqrt{n_j}} \quad (8.4)$$

Таким образом оказывается, что точность данных в сильной степени зависит от числа участвующих лабораторий. При $f < 4$ величина $t(P, f)$ особенно быстро возрастает, а точность данных уменьшается (рис. 3.15). Поэтому совместное определение необходимо проводить по меньшей мере пятью лабораториями. Зато число проводимых в каждой лаборатории параллельных определений меньше влияет на величину доверительного интервала.

В общем случае следует иметь не менее трех и не более пяти параллельных определений. При планировании опыта надо предусмотреть, чтобы параллельные определения вели независимо друг от друга (лучше всего, в различные дни, см. разд. 5.2). Большое число параллельных опре-

делений $n_j = 5$ следует проводить только при ответственных анализах или тогда, когда есть основания ожидать отклонения от нормального распределения. Целесообразно в каждой лаборатории проводить одинаковое число параллельных определений.

[8.1]. Содержание кремния в пробе ферросилиция определяли совместным опытом. Из измеренных значений надо найти среднее значение пробы \bar{x} , а также ошибку воспроизводимости и ошибку сопоставимости. Были получены следующие результаты (%Si):

Лаборатории

A	B	C	D	E	F	G
45,09	45,20	45,37	45,23	45,40	45,63	44,93
45,19	45,27	45,45	45,26	45,40	45,65	44,95
45,22	45,30	45,48	45,31	45,45	45,73	44,95
45,25	45,40	45,60	45,39	45,60	45,85	45,14
45,31	45,43	45,62	45,44	45,60	45,85	45,17

Дальнейший подсчет ведут по следующей схеме:

1. Преобразование в соответствии с соотношением $X_i = 100x_i - 4535$

Лаборатории

A	B	C	D	E	F	G
-26	-15	+2	-12	+5	+28	-42
-16	-8	+10	-9	+5	+30	-40
-13	-5	+13	-4	+10	+38	-40
-10	+5	+25	+4	+25	+50	-21
-4	+8	+27	+9	+25	+50	-18
Суммы:						
-69	-15	+77	-12	+70	+196	-161
Общая сумма:	+86					

2. Проверка по χ^2 -критерию (ср. разд. 7.2).

$$\begin{aligned}
 QS_A &= 26^2 + 16^2 + 13^2 + 10^2 + 4^2 - 69^2/5 = 264,8; & s_A^2 &= 66,2 \\
 QS_B &= 15^2 + 8^2 + 5^2 + 5^2 + 8^2 - 15^2/5 = 358,0; & s_B^2 &= 89,5 \\
 QS_C &= 2^2 + 10^2 + 13^2 + 25^2 + 27^2 - 77^2/5 = 441,2; & s_C^2 &= 110,3 \\
 QS_D &= 12^2 + 9^2 + 4^2 + 4^2 + 9^2 - 12^2/5 = 309,2; & s_D^2 &= 77,3 \\
 QS_E &= 5^2 + 5^2 + 10^2 + 25^2 + 25^2 - 70^2/5 = 420,0; & s_E^2 &= 105,0 \\
 QS_F &= 28^2 + 30^2 + 38^2 + 50^2 + 50^2 - 196^2/5 = 444,8; & s_F^2 &= 111,2 \\
 QS_G &= 42^2 + 40^2 + 40^2 + 21^2 + 18^2 - 161^2/5 = 544,8; & s_G^2 &= 136,2
 \end{aligned}$$

s_j^2	f_j	$f_j s_j^2$	$\lg s_j^2$	$f_j \lg s_j^2$
66,2	4	264,8	1,8209	7,2836
89,5	4	358,0	1,9518	7,8072
110,3	4	441,2	2,0426	8,1704
77,3	4	309,2	1,8882	7,5528
105,0	4	420,0	2,0212	8,0848
111,2	4	444,8	2,0461	8,1844
136,2	4	544,8	2,1342	8,5368
	28	2782,8		55,6200

$$S^2 = \frac{2782,8}{28} = 99,39$$

$$\lg S^2 = 1,9973$$

$$28 \lg S^2 = 55,9244$$

$$\chi^2 = 2,303 (55,9244 - 55,6200) = 0,7010 \sim 0,70$$

$\chi^2 (\bar{P} = 0,95; f = 6) = 12,6$; так как $\chi^2 < \chi^2 (\bar{P}, f)$, то для случайных ошибок в отдельных сериях нет значимого различия. Поэтому отдельные серии можно сравнивать.

3. Вычисление сумм квадратов

а) Разброс между значениями семи лабораторий ($m = 7$) [ср. уравнение (8.2)]

$$\begin{aligned}
 QS_1 &= \frac{69^2 + 15^2 + 77^2 + 12^2 + 70^2 + 196^2 + 161^2}{5} - \frac{86^2}{35} = \\
 &= 15\,847,9 \quad (\text{с } f_1 = m - 1 = 6)
 \end{aligned}$$

б) Разброс значений внутри лабораторий

$$QS_2 = QS_A + QS_B + \dots + QS_G = 2782,8 \quad (\text{с } f_2 = n - m = 28)$$

в) Общий разброс [ср. уравнения (2.5) и (2.6)]

$$\begin{aligned}
 QS &= 26^2 + 16^2 + 13^2 + 10^2 + 4^2 + 15^2 + \dots + 18^2 - \frac{86^2}{35} = \\
 &= 18\,630,7 \quad (\text{с } f = n - 1 = 34)
 \end{aligned}$$

4. Результаты вычислений

Причина	Сумма квадратов	Степени свободы	Дисперсия
Разброс между значениями лабораторий	15 847,9	6	2641,3
Разброс значений внутри лабораторий	2 782,8	28	99,4
Общий разброс	18 630,7	34	—

5. Расчет отдельных средних квадратичных ошибок

Прежде всего проверяют нуль-гипотезу $\sigma_1^2 = \sigma_2^2$

$$F = \frac{2641,3}{99,4} = 26,57$$

$$F(\bar{P}=0,99; f_1=6; f_2=28) = 3,53$$

Так как $F > F(\bar{P}, f_1, f_2)$, нуль-гипотезу следует отбросить для ошибки первого рода $100\alpha < 1\%$. Дисперсии s_1^2 и s_2^2 следует считать различающимися. Из них получают средние квадратичные ошибки воспроизводимости и сопоставимости.

$$S_V = \sqrt{S_1^2} = \sqrt{2641,3} = 51,4 \text{ соответственно } s_V = 0,51\% \text{ Si}$$

$$S_W = \sqrt{S_2^2} = \sqrt{99,4} = 10,0 \text{ соответственно } s_W = 0,10\% \text{ Si}$$

По уравнению (8.1) часть s_V , обусловленная лабораторией, получается

$$S_L^2 = \frac{2641,3 - 99,4}{5} = 508,4$$

$$S_L = 23 \text{ соответственно } s_L = 0,23 \text{ Si}$$

6. Среднее значение и доверительный интервал

Среднее значение из всех измерений $\bar{x} = 45,37\% \text{ Si}$
Соответствующий доверительный интервал при сравнимых условиях опыта находят по уравнению (8.3). Получают

$$\Delta \bar{x} = \frac{t(P=0,95; f_1=6) s_1}{\sqrt{n}} = \frac{2,45 \cdot 0,51}{\sqrt{35}} = 0,21\% \text{ Si}$$

Из этого содержание исследуемой пробы с вероятностью $P = 0,95$ получается $(45,37 \pm 0,21)\% \text{ Si}$.

Сравнение средних квадратичных ошибок воспроизводимости и сопоставимости приведено в табл. 8.1. Известно, что между обеими величинами s_W и s_V нет простой зависимости. Поэтому s_V и s_W в каждом случае необходимо определять экспериментально.

Таблица 8.1

Средние квадратичные ошибки воспроизводимости и сопоставимости при анализе стали

Элемент	Содержание, %	$s_W, \%$	$s_V, \%$	Элемент	Содержание, %	$s_W, \%$	$s_V, \%$
C	0,1	0,004	0,010	P	0,001	0,0005	0,0024
	0,4	0,005	0,014		0,003	0,0008	0,0024
	1,0	0,006	0,026		0,010	0,0006	0,0019
Mn	0,3	0,006	0,015	S	0,5	0,006	0,009
	1,0	0,008	0,030		10	0,04	0,12
Si	0,3	0,006	0,021	Cu	0,2	0,005	0,015
	1,0	0,009	0,028				

Знание средней квадратичной ошибки сопоставимости особенно важно для оценки произведенного продукта, так как при этом следует учитывать разброс результатов для партнеров, участвующих в анализе.

Иногда среднюю квадратичную ошибку сопоставимости нельзя получить описанным способом из-за дефектов в работе партнеров совместного опыта. При участии только двух лабораторий это можно сделать, когда каждая лаборатория проведет определения на m пробах. Если x_i' — значение одной лаборатории, x_i'' — значение второй лаборатории, то получают среднюю квадратичную ошибку сопоставимости по уравнению (5.2).

8.2. Устранение ошибок

Общая ошибка метода анализа чаще всего состоит из отдельных частных ошибок. Их суммируют по закону распространения ошибок (ср. разд. 4). Знание этих частных ошибок важно, например, при разработке нового метода анализа, так как можно улучшить ход анализе на стадиях с большей ошибкой.

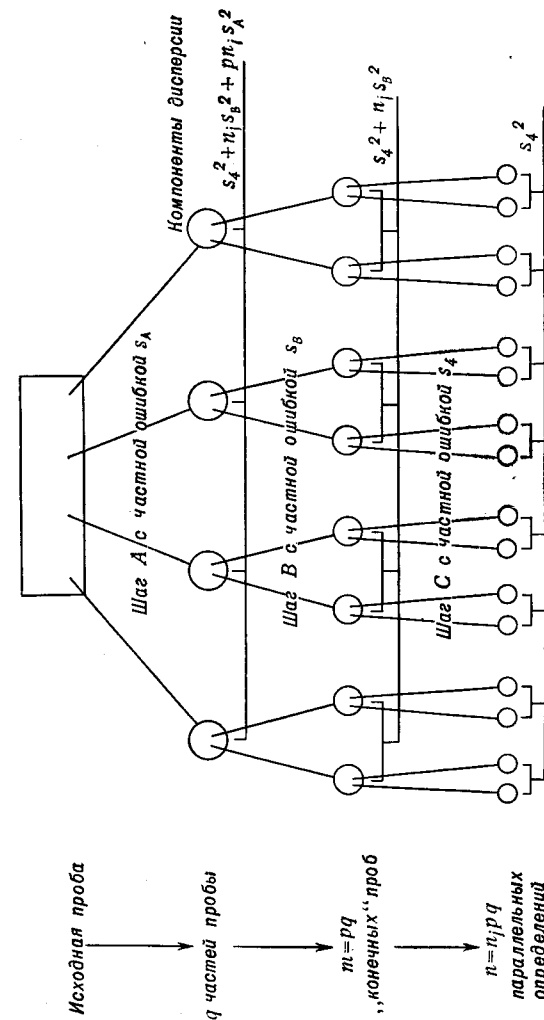
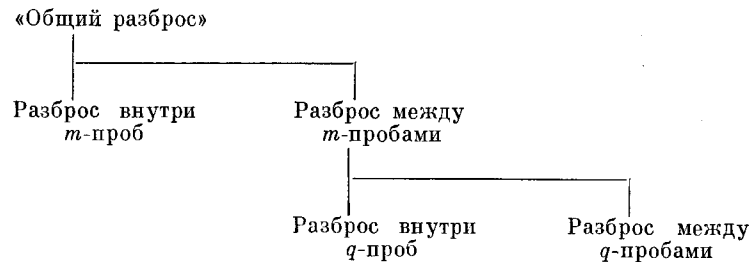
Выявление и устранение двух частных ошибок можно провести при помощи простого дисперсионного анализа [2]. Естественно, результаты получаются интереснее, если можно учесть больше, чем две причины ошибок. Если, например, из общей ошибки аналитического определения хотят выделить еще две частные ошибки, то пользуются схемой опыта, показанной на рис. 8.1. Исходной пробой является гомогенная проба достаточного объема. На ней проводят нужные испытания (шаг А). После этого ее разбивают на q частей. На каждой такой q -пробе проводят второе испытание (шаг В) и делают вслед за тем каждую q -пробу еще p раз. В результате имеют $m = pq$ «конечных» проб. На каждой m конечной пробе проводят n_j параллельных определений (шаг С).

Для определения частных ошибок обеих операций (компонентов дисперсий s_A^2 и s_B^2) необходимо расширить данную на стр. 158 схему простого дисперсионного анализа — отдельные группы разделить еще на подгруппы (рис. 8.1). Простой дисперсионный анализ с подгруппами ведут в две стадии:

1. Прежде всего проводят обычный простой дисперсионный анализ. При этом принадлежность m конечных проб к различным q -пробам оставляют без внимания. Согласно данной на стр. 158 схеме, выделяют разброс «между m -пробами», разброс «внутри m -пробы» и «общий» разброс.

2. На следующей стадии подсчета сумма квадратов для разброса «между m -пробами» разбивается на суммы квадратов «между q -пробами» и «внутри q -проб».

Таким образом проводят дважды последовательно простой дисперсионный анализ. Ход расчета можно представить следующей схемой:



Р и с. 8.1. Схема эксперимента для разложения суммарной ошибки на отдельные составляющие при совместном действии двух частных ошибок и остаточной ошибки опыта.

Расчет величин, необходимых для дисперсионного анализа в подгруппах (суммы квадратов, степени свободы, дисперсии), проводят по схеме, приведенной ниже.

Суммы квадратов «внутри m -групп» и «общую» снова подсчитывают по уравнению (2.6а). Для расчета остальных сумм квадратов в основу кладут подходящие суммы аналогично уравнению (8.2) (вместо средних значений). Для суммы квадратов «между q -пробами» получают

$$pn_j \sum (\bar{x}_k - \bar{x})^2 = \sum_1^q \frac{(\sum_1^p \sum_1^{n_j} x_{kji})^2}{pn_j} - \frac{(\sum_1^q \sum_1^p \sum_1^{n_j} x_{kji})^2}{qpn_j} \quad (8.5)$$

Причина	Сумма квадратов	Степень свободы	Дисперсия	Компоненты дисперсии
Разброс между m -пробами	$QS_1 = n_j \sum (\bar{x}_{kj} - \bar{x})^2$	$f_1 = m - 1 = pq - 1$	—	—
Разброс между q -пробами	$QS_2 = pn_j \sum (\bar{x}_k - \bar{x})^2$	$f_2 = q - 1$	$s_2^2 = \frac{QS_2}{f_2}$	$s_4^2 + n_j s_B^2 + pn_j s_A^2$
Разброс внутри q -проб	$QS_3 = n_j \sum (\bar{x}_{kj} - \bar{x}_k)^2 = QS_1 - QS_2$	$f_3 = q(p - 1) = f_1 - f_2$	$s_3^2 = \frac{QS_3}{f_3}$	$s_4^2 + n_j s_B^2$
Разброс внутри m -проб (ошибка опыта)	$QS_4 = \sum \sum \sum (x_{kji} - \bar{x}_{kj})^2$	$f_4 = n - m = pq(n_j - 1)$	$s_4^2 = \frac{QS_4}{f_4}$	s_4^2
Общий разброс	$QS = \sum (x_{kji} - \bar{x})^2 = QS_2 + QS_3 + QS_4$	$f = n - 1 = pqn_j - 1 = f_2 + f_3 + f_4$	—	—

Сумму квадратов «между m -пробами» вычисляют по уравнению (8.2). Сумма квадратов «внутри q -проб» получается из разности.

Искомую частную ошибку (оценка компонентов дисперсии) находят по следующей схеме:

$$\left. \begin{aligned} &\text{Ошибка конечного определения } s_4^2 \\ &\text{Ошибка шага B} \\ &\text{Ошибка шага A} \end{aligned} \right\} \begin{aligned} &s_B^2 = \frac{s_3^2 - s_4^2}{n_j} \\ &s_A^2 = \frac{s_2^2 - s_3^2}{n_j p} \end{aligned} \quad (8.6)$$

Ранее указанным путем можно вычислить доверительный интервал для компонент дисперсии (ср. разд. 5.2).

[8.2]. Определение вязкости пластмассы дало сильный разброс значений. Поэтому необходимо было проверить, на какой стадии метода появляется ошибка [8]. Во внимание приняли следующие факторы:

- 1) неоднородность проб;
- 2) вымывание минеральных добавок;
- 3) измерение вязкости.

Соответственно этим трем факторам было взято 10 достаточно больших q -проб ($a_1 \dots a_{10}$) (шаг А). Каждая из q -проб была разбита на две m -пробы ($p = 2$) (b_1 и b_2) (шаг В). Каждую m -пробу вымывали. После этого на каждой m -пробе ($pq = m = 20$) провели по два измерения вязкости (c_1 и c_2) (шаг С). В результате получили следующие значения измерения:

	a_1		a_2		a_3		a_4		a_5	
	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
c_1	59,8	60,0	65,0	64,5	65,0	65,5	62,5	60,9	59,8	56,0
c_2	61,2	65,0	65,8	64,5	65,2	63,5	61,9	61,5	60,9	57,2
	a_6		a_7		a_8		a_9		a_{10}	
	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
c_1	68,8	62,5	65,2	61,0	59,6	62,3	61,0	73,0	65,0	62,0
c_2	69,0	62,0	65,6	59,3	58,5	61,5	64,0	71,7	64,0	63,0

Дальнейший расчет на одинаковых ступенях вели как при простом дисперсионном анализе:

1. Преобразование по соотношению $X_i = 10x_i - 600$

	a_1		a_2		a_3		a_4		a_5	
	b_1	b	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
c_1	-2	0	50	45	50	55	25	9	-2	-40
c_2	+12	50	58	45	52	35	19	15	+9	-28
Σ	10	50	108	90	102	90	44	24	7	-68
$\Sigma\Sigma$	60		198		192		68		-61	
$\Sigma\Sigma\Sigma$										

2. Расчет сумм квадратов

а) Разброс между m -пробами ($m = 20$) [уравнение (8.2)]

$$QS_1 = \frac{10^2 + 50^2 + 108^2 + 90^2 + \dots + 50^2}{2} - \frac{1247^2}{40} = 45\,799 \quad (c \ f_1 = m - 1 = 19)$$

б) Разброс внутри m -проб [для расчета этой суммы квадратов при $n_j = 2$ параллельным определениям см. (5.2)]

$$QS_4 = \frac{1}{2} \{[-2 - (+12)]^2 + (0 - 50)^2 + (50 - 58)^2 + \dots + (20 - 30)^2\} = 2645 \quad (c \ f_4 = n - m = 20)$$

в) Общий разброс [уравнение (2.6а)]

$$QS = -2^2 + 12^2 + 0^2 + 50^2 + \dots + 20^2 + 30^2 - \frac{1247^2}{40} = 48\,444 \quad (c \ f = n - 1 = 39)$$

г) Разброс между q -пробами ($q = 10$) [уравнение (8.5)]

$$QS_2 = \frac{60^2 + 198^2 + 192^2 + \dots + 140^2}{4} - \frac{1247^2}{40} = 25\,683 \quad (c \ f_2 = q - 1 = 9)$$

a_6		a_7		a_8		a_9		a_{10}	
b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
88	25	52	10	-4	23	10	130	50	20
90	20	56	-7	-15	15	40	117	40	30
178	45	108	3	-19	38	50	247	90	50
223		111		19		297		140	
1247									

По приведенной на стр. 166 схеме находят сумму квадратов и число степеней свободы по разности значений, полученных в а) и г). Имеем:

$$QS_3 = QS_1 - QS_2 = 45\,799 - 25\,683 = 20\,116 \quad (c \ f_3 = f_1 - f_2 = 10)$$

3. Результаты вычислений

Причина	Сумма квадратов	Степени свободы	Дисперсия	Компоненты дисперсий
Разброс между m -пробами	45 799	19	—	—
Разброс между q -пробами	25 683	9	2854	$s_4^2 + 2s_B^2 + 4s_A^2$
Разброс внутри q -проб	20 116	10	2012	$s_4^2 + 2s_B^2$
Разброс внутри m -проб	2 645	20	132	s_4^2
Общий разброс	48 444	39	—	

4. Расчет компонент дисперсии

Для оценки компонент дисперсии по уравнению (8.6) получают следующие значения:

$$S_3^2 = 132$$

$$S_B^2 = \frac{2012 - 132}{2} = 940$$

$$S_A^2 = \frac{2854 - 2012}{4} = 211$$

После обратного преобразования получают:

$$s_4^2 = 13,2 \quad s_4 = 3,63$$

$$s_B^2 = 94,0 \quad s_B = 9,70$$

$$s_A^2 = 21,1 \quad s_A = 4,60$$

Отсюда следует, что частная ошибка s_B , вызванная вымыванием, оказывается наибольшей для трех рассмотренных стадий.

Многоступенчатые опыты описанного здесь вида являются эффективным вспомогательным средством в исследованиях, направленных на уменьшение случайной ошибки метода анализа. Если опыт должен дать достаточно достоверную информацию, то каждая частная дисперсия s_1^2, s_2^2, \dots должна иметь по меньшей мере десять степеней свободы. Поэтому особенно на первой стадии (шаг А) следует предусмотреть возможность достаточно широкого дробления пробы. Опыт должен быть симметричным, т. е. на каждой ступени следует проводить равное число определений. На каждой стадии необходима однородная проба (если не определяют ошибку отбора проб). Поэтому применяют преимущественно растворы, которые делят объемным путем (табл. 4.1). Наконец, следует учитывать, что исходная проба должна быть достаточно большой.

8.3. Сравнение нескольких средних значений

Для проверки различия между двумя средними ($m=2$) значениями \bar{x}_1 и \bar{x}_2 в разд. 7.3 разность $|\bar{x}_1 - \bar{x}_2|$ сопоставляли с ошибкой опыта внутри этих двух серий. Если ошибку опыта обозначить s_2 , а относящиеся к ней степени свободы f_2 , то из уравнения (7.6) для случая

$n_1 = n_2 = n_j$ получается

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_2} \sqrt{\frac{n_j}{2}}$$

при степенях свободы f_2 .

Возводя в квадрат, получают

$$t^2 = \frac{(\bar{x}_1 - \bar{x}_2)^2 \cdot n_j / 2}{s_2^2} = \frac{n_j \left[\bar{x}_1^2 + \bar{x}_2^2 - \frac{(\bar{x}_1 + \bar{x}_2)^2}{2} \right]}{s_2^2} = \frac{n_j \left[\sum \bar{x}_j^2 - \frac{(\sum \bar{x}_j)^2}{m} \right]}{s_2^2}$$

По (2.6а), числитель этого выражения соответствует

$$n_j \left[\bar{x}_j^2 - \frac{(\sum \bar{x}_j)^2}{m} \right] = n_j \sum (\bar{x}_j - \bar{\bar{x}})^2$$

сумме квадратов QS_1 при простом дисперсионном анализе для разброса «между сериями» (ср. стр. 158). Так как в данном случае $m = 2$ и $f_1 = 1$, эта величина одновременно представляет дисперсию s_1^2 . В разд. 3.4 было показано, что $t^2(P, f_2) = F(\bar{P}, f_1 = 1; f_2)$. Вследствие этого:

$$t^2(P, f_2) = \frac{s_1^2}{s_2^2} = F(\bar{P}; f_1 = 1; f_2)$$

Проверка среднего значения сводится к проверке различия между двумя дисперсиями s_1^2 и s_2^2 , следовательно, к задаче дисперсионного анализа. Иногда можно проверить разность между сколь угодно многими средними значениями. При этом проверяемой гипотезой является то, что генеральные совокупности, соответствующие средним значениям \bar{x}_j , должны иметь одно и то же среднее значение μ , следовательно, предполагается, что $\mu_1 = \mu_2 = \dots = \mu_m$. Для проверки этой гипотезы пользуются простым дисперсионным анализом. Если принимается нуль-гипотеза [$F = s_1^2/s_2^2 < F(\bar{P}, f_1, f_2)$], то между средними значениями нет значимой разницы и проверка на этом заканчивается. Если нуль-гипотеза не принимается [$F = s_1^2/s_2^2 > F(\bar{P}, f_1, f_2)$], то дополнительно проводят попарную проверку средних значений серий при помощи критерия Дункана. Для этого отдельные средние значения упорядочивают по убыванию и нуме-

руют: $p = 1, 2, 3 \dots m$. Разность между какими-либо двумя средними значениями \bar{x}_k и \bar{x}_l является значимой, если

$$q = \frac{|\bar{x}_k - \bar{x}_l|}{s_2} \sqrt{\frac{2n_k n_l}{n_k + n_l}} > q(P, p, f_2) \quad (8.7)$$

Числовые значения для $q(P, p, f_2)$ находят в табл. 12.6.

[8.3]. Калибровочная проба для эмиссионноспектрального анализа была приготовлена распиливанием однородной железной штанги на маленькие пластинки ($3 \times 3 \text{ см}^2$). Для контроля однородности каждую пластинку подвергали четырехкратному обыскриванию. Содержание хрома, найденное для каждой пятой пластинки, показало следующее: у третьей пробы один анализ вышал вследствие образования окислы. Необходимо проверить, можно ли считать однородной железную штангу, т. е. действительно ли нет значимой разницы между составом отдельных пластинок.

	1	2	3	4	5	6	7	8
	1,42	1,42	1,42	1,38	1,36	1,37	1,38	1,32
	1,42	1,39	1,38	1,41	1,37	1,34	1,37	1,33
	1,41	1,38	1,41	1,41	1,37	1,38	1,36	1,34
	1,44	1,38	—	1,42	1,39	1,34	1,37	1,32
Средние значения	1,423	1,393	1,403	1,405	1,373	1,358	1,370	1,328

1. Преобразование в соответствии с соотношением $X_i = 100x_i - 140$

Порядковый номер	1	2	3	4	5	6	7	8
	+2	+2	+2	-2	-4	-3	-2	-8
	+2	-1	-2	+1	-3	-6	-3	-7
	+1	-2	+1	+1	-3	-2	-4	-6
	+4	-2	—	+2	-1	-6	-3	-8
Суммы	+9	-3	+1	+2	-11	-17	-12	-29
Средние значения	+2,25	-0,75	+0,33	+0,50	-2,75	-4,25	-3,00	-7,25
Общая сумма				-60				

2. χ^2 -проверка

Результаты анализа для всех восьми пластинок были получены одним и тем же способом. Искровые пятна лежали плотно одно к другому. Условия проведения опыта не допускали появления неоднородной ошибки. Можно было не производить χ^2 -проверку.

3. Расчет сумм квадратов

а) Разброс «между пластинками» [уравнение (8.2)]

$$QS_1 = \frac{9^2}{4} + \frac{3^2}{4} + \frac{1^2}{3} + \frac{2^2}{4} + \frac{11^2}{4} + \frac{17^2}{4} + \frac{12^2}{4} + \frac{29^2}{4} - \frac{60^2}{31} = 256,45 \quad (\text{с } f_1 = m - 1 = 7)$$

б) Разброс «внутри пластинок»

$$QS_2 = 2^2 + 2^2 + 1^2 + 4^2 - \frac{9^2}{4} + 2^2 + \dots + 8^2 - \frac{29^2}{4} = 55,42 \quad (\text{с } f_2 = n - m = 23)$$

в) «Общий» разброс (уравнение [2.6а])

$$QS = 2^2 + 2^2 + 1^2 + 4^2 + 2^2 + \dots + 8^2 - \frac{60^2}{31} = 311,87 \quad (\text{с } f = f_1 + f_2 = n - 1 = 30)$$

4. Результаты вычислений

Причина	Суммы квадратов	Степени свободы	Дисперсия
Разброс между пластинками	256,45	7	36,64
Разброс внутри пластинок	55,42	23	2,41
Общий разброс	311,87	30	—

5. Проверка нуль-гипотезы

$$F = \frac{36,64}{2,41} = 15,20$$

Согласно табл. 12.56, $F(\bar{P} = 0,99; f_1 = 7; f_2 = 23) \sim 3,60$.

Так как $F > F(\bar{P}, f_1, f_2)$, нуль-гипотезу следует отбросить. В отдельных пластинках обнаружено значимое различие в содержании хрома.

6. Парная проверка

На основании F -проверки нуль-гипотезу отбрасывают и дополнительно проводят парную проверку отдельных средних значе-

ний. Для этого средние значения упорядочивают по убыванию и нумеруют от 1 до p . Получают:

Номер пластинки	1	4	3	2	5	7	6	8
Среднее значение \bar{x}_j	+2,25	+0,50	+0,33	-0,75	-2,75	-3,00	-4,25	-7,25
Число параллельных определений n_j	4	4	3	4	4	4	4	4
Порядковый номер p	1	2	3	4	5	6	7	8

Для проверки, например, $\bar{X}_1 = +2,25$ по сравнению с $\bar{X}_2 = -0,75$ получают по уравнению (8.7)

$$q = \frac{2,25 - (-0,75)}{\sqrt{2,41}} \sqrt{\frac{2 \cdot 4 \cdot 4}{4 + 4}} = 3,86$$

По табл. 12.6 $q(P = 0,95; p = 4; f_2 = 23) = 3,16$; для $q(P = 0,99; p = 4; f_2 = 23)$ находят значение 4,26. В смысле ранее приведенных правил (ср. гл. 7) не существует значимого различия, однако допуская неблагоприятнейшую возможность, признают различие в двух средних значениях.

При проверке $\bar{X}_4 = +0,50$ относительно $\bar{X}_3 = +0,33$ совершенно аналогично получают

$$q = \frac{0,50 - 0,33}{\sqrt{2,41}} \sqrt{\frac{2 \cdot 4 \cdot 3}{4 + 3}} = 0,20$$

Среднее значение $\bar{x}_4 = +0,50$ является теперь наибольшим значением в ряду средних значений. Поэтому оно получает номер $p = 1$. Вследствие этого при проверке \bar{X}_4 относительно следующего среднего значения \bar{X}_3 ($p = 2$) из табл. 12.6а получается $q(P = 0,95; p = 2; f_2 = 23) = 2,93$. Вследствие того, что $q < q(P, p, f_2)$, между обоими средними значениями нет значимой разницы.

Если эту проверку проводят для всех возможных комбинаций пластинок, то получают результаты, представляемые по следующей схеме (стр. 175). В качестве сокращения используют: 0 — нет значимого различия, + значимое различие с $P = 0,95$ и ++ значимое различие с $P = 0,99$.

Почти все комбинации с восьмой пластинкой имеют значимое различие с вероятностью $P = 0,99$. Оказывается, что содержание хрома в этой пластинке особенно сильно отличается от среднего содержания его в штанге (головная часть). Разности между двумя

средними значениями различаются тем более надежно, чем дальше пластинки удалены друг от друга. Из этого следует, что содержание хрома постоянно изменяется вдоль всей штанги.

Пластинка	2	3	4	5	6	7	8
1	+	0	0	++	++	++	++
2		0	0	0	++	++	++
3			0	+	++	++	++
4				+	++	++	++
5					0	0	++
6						0	+
7							++

Методом простого дисперсионного анализа можно проверять и сравнивать результаты анализа, полученные для содержания вещества в пробе. Простой дисперсионный анализ весьма удобен для оценки, например, результатов совместного опыта различных лабораторий, при проведении которого стремятся устанавливать состав пробы возможно точнее. Однако дисперсионный анализ непригоден для проверки самого аналитического метода при помощи совместного опыта. В этом случае безусловно необходимо отбирать несколько проб разного содержания в тщательно спланированном эксперименте. При анализе только одной пробы существует опасность, что обычная систематическая ошибка останется неизвестной — она может быть скомпенсирована действием случайных факторов. Для проведения и оценки совместных экспериментов при проверке метода анализа следовало бы обратиться к Манделю и Лейшоу [7] *.

ЛИТЕРАТУРА

1. Doerffel K., Z. anal. Chem., 184, 81 (1961).
2. Doerffel K., Z. anal. Chem., 185, 1 (1962).
3. Doerffel K., Chemia Analityczna, 8, 333 (1963).

* Подробное и вполне доступное изложение методов дисперсионного анализа имеется в книге Хикса «Основные принципы планирования эксперимента», «Мир», 1968. — Прим. ред.

4. Doerffel K., Schulze M., Neue Hütte, 9, 690 (1964).
5. Duncan D. B., Biometrics, 11, 1 (1955).
6. Inglot J., Kucharczyk J., Prace IH, 16, 327 (1964).
7. Mandel J., Lashof T. W., ASTM Bull., 239, № 7, 53 (1959).
8. Matthias R. H., Anal. Chem., 29, 1046 (1957).
9. Pierson R. A., Fay E. A., Anal. Chem., 31, Н. 12, 25A, 49A (1959).
10. TGL 0-51849: Prüfung metallischer Werkstoffe.

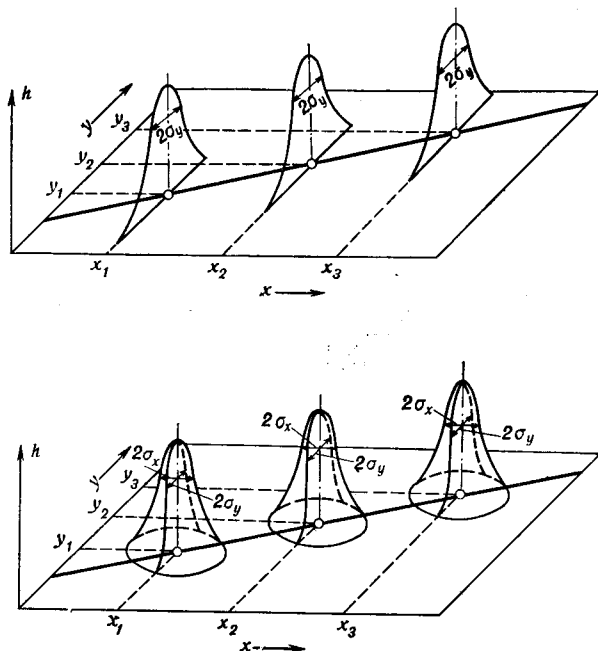
9. Статистика прямых линий (регрессионный и корреляционный анализ)

Как во всех точных науках, так и в аналитической химии необходимо отыскать и охарактеризовать связь между результатами измерений. Так, инструментальные методы анализа чаще всего требуют калибровки. Задачей аналитика является нахождение калибровочной функции из заданного содержания и измеренных значений и получение из этих данных сведений о точности метода анализа. Нередко константы калибровочной функции имеют тот или иной физико-химический смысл. Полученные для них числовые значения затем следует дополнить соответствующим доверительным интервалом.

Все эти проблемы можно решить при помощи регрессионного анализа. Этот метод применим всегда, если требуется лучше оценить известную заранее зависимость между двумя (или несколькими) переменными. При этом числовые значения независимых переменных x уже заданы перед опытом, а числовые значения зависимых переменных y получаются в ходе опыта. Ошибка значений x несравнимо меньше ошибки значений y . Для каждого значения независимой переменной x должна быть нормально распределена зависимая случайная переменная y (рис. 9.1, а). Поэтому для каждого значения x возможно много значений y , а из функции, найденной подсчетом регрессии, нельзя непосредственно получить обратную зависимость.

Не всегда известно заранее, имеется ли связь между двумя случайными переменными. Это необходимо проверить подсчетом корреляции. В противоположность регрессионному анализу при вычислении корреляции обе переменные связаны одинаковым образом с ошибками. Различные значения y могут принадлежать одному значению x , и наоборот. Поэтому каждая полученная пара значений принадлежит двумерному распределению (рис. 9.1, б).

В аналитической химии преимущественно имеют место линейные связи. Более сложные зависимости встречаются редко, и часто подходящими преобразованиями их можно



Р и с. 9.1. Задача регрессии (вверху) и задача корреляции (внизу).

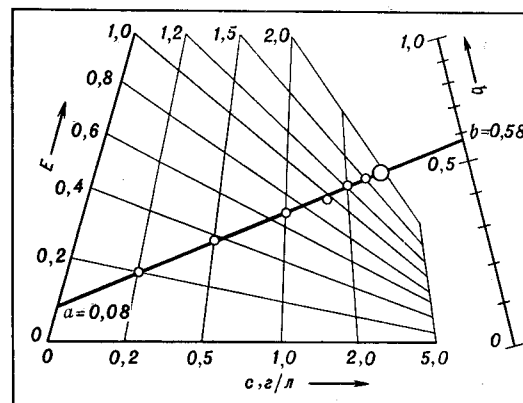
привести к линейному виду. Поэтому дальнейшее изложение будет ограничено изучением линейных зависимостей.

9.1. Определение констант

При измерении нашли m пар значений (x_i, y_i) ($m > 2$). Известно, что между обеими переменными существует линейная зависимость $Y = a + bx$ и константы этой функции a и b надо определить численно. При этом требуется, чтобы разность между измеренным y_i и значениями, вычис-

ленными из уравнения Y_i , была возможно меньше, т. е. ищут «оптимальнейшую» функцию. Для решения этой задачи используют графические способы приближения и счетные методы.

При графическом определении результаты измерений в виде точек наносят на координатную сетку. Через эти точки проводят прямую. При этом отдельные точки должны



Р и с. 9.2. Графическое сравнение значений для примера [9.1] в проекционно-искаженной сетке [3].

быть равномерно распределены выше и ниже прямой. Константу a находят как отрезок ординаты y при $x = 0$, величина b представляет собой тангенс угла наклона прямой. Непосредственное отыскание констант a и b позволяет проекционно-искаженная сетка * (рис. 9.2).

При сильном разбросе значений измерений графическое выравнивание описанным способом часто неосуществимо. Тогда довольствуются тем, что соединяют точки прямыми линиями попарно. Через середины полученных отрезков проводят прямую. Этот метод применяют до тех пор, пока не получат несколько точек, через которые можно провести сглаживающую прямую (рис. 9.3). Однако

* Проекционно-искаженная бумага по Фишеру. Поставляется фирмой «Schäfers Feinpapier», Плауэн (Фогтланд).

этот метод нельзя применять в проекционной сетке. В этом случае возвращаются к простому расчетно-графическому методу [13]. При большом числе «результатов» измерений можно вернуться к выравниванию, так называемым «канальным» методом по Вернеру [16]. Через

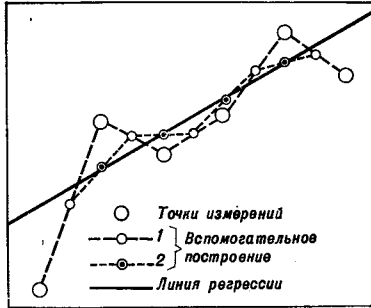


Рис. 9.3. Построение линии регрессии при сильно разбросанных значениях измерений.

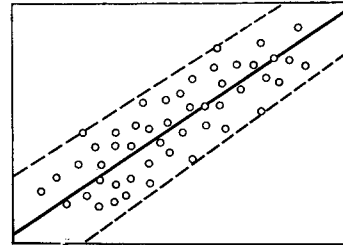


Рис. 9.4. Построение линии регрессии при очень многих точках измерений.

крайние верхние и нижние точки проводят прямые. В середине этого «канала» лежит сглаживающая прямая (рис. 9.4).

Константы a и b могут быть вычислены методом регрессионного анализа. Если имеется m взаимосвязанных пар значений (x_i, y_i) , то это можно записать следующим образом:

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\dots \\ y_m &= a + bx_m \end{aligned}$$

В левой части этих выражений стоят измеренные значения y_i , в правой части — вычисленные значения $Y_i = a + bx_i$. Разность между обеими величинами дает ошибку. Совпадение между измеренными и вычисленными значениями является наилучшим, если сумма квадратов

$$\sum (y_i - Y_i)^2 = \sum (y_i - a - bx_i)^2 \quad (9.1)$$

будет минимальной. Это выполняется, если

$$\frac{\partial}{\partial a} \left[\sum_1^m (y_i - a - bx_i)^2 \right] = -2 \sum_1^m (y_i - a - bx_i) = 0$$

$$\frac{\partial}{\partial b} \left[\sum_1^m (y_i - a - bx_i)^2 \right] = -2 \sum_1^m x_i (y_i - a - bx_i) = 0$$

Решая эти уравнения относительно a и b , получим

$$b = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m \sum x_i^2 - (\sum x_i)^2} \quad (9.2)$$

$$a = \frac{\sum y_i - b \sum x_i}{m} \quad (9.3)$$

Константы a и b являются случайными величинами, при помощи которых оцениваются теоретические параметры α и β . Можно найти доверительный интервал для a и b , так как это сделано для отдельного значения измерения [уравнение (3.7)]. Для этого прежде всего определяют дисперсию, характеризующую разброс измеренных значений (y_i) относительно вычисленных (Y_i)

$$s_0^2 = \frac{\sum_1^m (y_i - Y_i)^2}{m-2} \quad (9.4)$$

со степенями свободы $f = m - 2$.

Здесь имеется число степеней свободы $m - 2$, так как проведение прямых требует по меньшей мере двух точек. (Однако, если для каждой из m проб проводят по n_j параллельных определений, так что имеется $mn_j = n$ результатов измерений, то с уравнением (9.4), естественно, будет связано $f = n - 2$.) Сумму квадратов в уравнении (9.4) удобнее определять, пользуясь следующим выражением:

$$\sum (y_i - Y_i)^2 = s_0^2 (m - 2) = \sum y_i^2 - a \sum y_i - b \sum x_i y_i \quad (9.5)$$

При проведении вычислений по уравнению (9.5) следует обратить внимание на то, что расчеты нужно делать

при достаточно большом числе знаков после запятой, так как искомую сумму квадратов часто ищут для весьма близких значений. Поэтому на этой ступени даже незначительная ошибка в вычислениях приводит к большим погрешностям.

Дисперсии для констант a и b можно найти, пользуясь законом распространения ошибок. Получают

$$s_b^2 = \frac{s_0^2}{\sum (x_i - \bar{x})^2} = \frac{ms_0^2}{m \sum x_i^2 - (\sum x_i)^2} \quad (9.6)$$

$$s_a^2 = \frac{s_0^2 \sum x_i^2}{m \sum (x_i - \bar{x})^2} = \frac{s_0^2 \sum x_i^2}{m \sum x_i^2 - (\sum x_i)^2} = \frac{s_b^2}{m} \sum x_i^2 \quad (9.7)$$

со степенями свободы $f = m - 2$.

Для константы b дисперсия s_b^2 тем меньше, чем дальше значение x лежит от его среднего значения \bar{x} , т. е. чем шире была выбрана рассматриваемая область (например, область концентраций).

Доверительный интервал для b и для a находят, вычисляя ошибки при помощи равенств (9.6) и (9.7)

$$\Delta b = t(P, f) s_b \quad (9.8)$$

$$\Delta a = t(P, f) s_a \quad (9.9)$$

Зная Δb и Δa , определяют число необходимых знаков после запятой для b и a .

Полученную функцию $y = a + bx$ можно использовать, чтобы из заданных, следовательно, почти безошибочных значений x вычислить значения зависимых переменных. Для одного заданного значения x_k находят одно значение Y_k . Вследствие неизбежной ошибки при определении констант a и b необходимо рассматривать Y_k также как случайную величину. Зная частные ошибки s_a и s_b , находят доверительный интервал для вычисленного значения Y_k

$$\begin{aligned} \Delta Y_k &= t(P, f) \sqrt{s_0^2 \left[\frac{1}{m} + \frac{(x_k - \bar{x})^2}{(\sum (x_i - \bar{x})^2)} \right]} = \\ &= t(P, f) \sqrt{s_0^2 \left[\frac{1}{m} + \frac{m(x_k - \bar{x})^2}{m \sum x_i^2 - (\sum x_i)^2} \right]} \quad (9.10) \end{aligned}$$

Следовательно, доверительный интервал зависит от разности $(x_k - \bar{x})$ и становится тем больше, чем дальше x_k лежит от среднего значения \bar{x} .

Таким образом, экстраполяция даже при наличии линейной связи может сопровождаться очень большими ошибками. Поэтому к ней прибегают только в вынужденных случаях в малой области. При экстраполяции к значению $x = 0$ необходимо было бы попытаться получить дополнительные измерения вблизи $x = 0$, если это еще можно сделать с достаточной точностью. Подобные экстраполяции возможны, если фактор b не слишком большая величина.

[9.1.] Для получения градуировочного графика при фотометрическом определении бензола в этаноле в УФ-области были измерены экстинкции семи проб известного содержания. Полученные значения приведены в таблице.

Концентрация x_i , г бензола/л	Экстинкция, y_i
0,2	0,20
0,5	0,37
1,0	0,64
1,5	0,93
2,0	1,22
2,5	1,50
3,0	1,80

Для получения градуировочной прямой производят следующие вычисления:

прежде всего определяют

$$\sum x_i = 10,7 \quad \sum x_i^2 = 22,79 \quad (\sum x_i)^2 = 114,49$$

$$\sum y_i = 6,66 \quad \sum y_i^2 = 8,4298 \quad \sum x_i y_i = 13,850$$

$$m = 7$$

По уравнениям (9.2) и (9.3) находят

$$b = \frac{7 \cdot 13,850 - 10,7 \cdot 6,66}{7 \cdot 22,79 - 114,49} = 0,570337$$

$$a = \frac{6,66 - 0,570337 \cdot 10,7}{7} = 0,079628$$

Сумму квадратов для дисперсии, характеризующей отклонение измеренных значений относительно вычисленных, находят по соотношению (9.5)

$$\sum (y_i - Y_i)^2 = 8,429800 - 0,079628 \cdot 6,66 - 0,570337 \cdot 13,850 = 0,000311$$

$$s_0^2 = \frac{0,000311}{5} = 0,0000622$$

Дисперсии для констант b и a получают из равенств (9.6) и (9.7)

$$s_b^2 = \frac{7 \cdot 0,0000622}{7 \cdot 22,79 - 114,49} = 0,00000967$$

$$s_b = 0,00311 \quad (c f = 5)$$

$$s_a^2 = \frac{0,0000622 \cdot 22,79}{7 \cdot 22,79 - 114,49} = 0,00003147$$

$$s_a = 0,00561 \quad (c f = 5)$$

При $P = 0,95$ $t(P, f)$ будет равно 2,57. Отсюда $t(P, f) s_b = 0,00799$ и $t(P, f) s_a = 0,014$. Искомые константы градуировочной прямой при $P = 0,95$

$$b = 0,570 \pm 0,008$$

$$a = 0,079 \pm 0,014$$

При связи вида $y = bx$ вычисление регрессии значительно упрощается. Тогда получают

$$b = \frac{\sum x_i y_i}{\sum x_i^2} \quad (9.11)$$

$$s_0^2 = \frac{\sum (y_i - Y_i)^2}{m - 1} \quad (9.12)$$

с $f = m - 1$ степеней свободы (ср. стр. 181 внизу).

$$\sum (y_i - Y_i)^2 = s_0^2 (m - 1) = \sum y_i^2 - b \sum x_i y_i \quad (9.13)$$

$$s_b^2 = \frac{s_0^2}{\sum x_i^2} \quad (9.14)$$

с $f = m - 1$ степеней свободы.

$$\Delta b = t(P, f) s_b = t(P, f) \sqrt{\frac{s_0^2}{\sum x_i^2}} \quad (9.15)$$

9.2. Проверка констант

Если для прямой $y = a + bx$ при расчете констант для a получается очень небольшое значение, то возникает подозрение, что уравнение может иметь вид $y = b'x$. Это всегда возможно в том случае, когда обе величины s_0 и s'_0 отличаются друг от друга только случайно. Для проверки вычисляют суммы квадратов QS [равенство (9.5)]

и QS' [уравнение (9.13)] с $f = m - 2$ и соответственно $f' = m - 1$ степенями свободы. При этом сумма квадратов, относящаяся к $y = a + bx$, никогда не будет больше, чем подсчитанная для $y = b'x$. Соотношение между обеими суммами квадратов приведено ниже:

Функции	Сумма квадратов	Степень свободы	Дисперсия
$y = b'x$	QS'	$m - 1$	$s'^2 = QS'/m - 1$
$y = a + bx$	QS	$m - 2$	$s^2 = QS/m - 2$
Разность	$QS' - QS$	1	

Затем находят соотношение

$$F = \frac{QS' - QS}{s^2} \quad (9.16)$$

и сравнивают, как обычно, с $F(\bar{P}, f_1 = 1; f_2 = m - 2)$. Для перехода к уравнению $y = b'x$ имеются основания только тогда, когда $F < F(\bar{P}, f_1, f_2)$.

[9.2]. Для определения следов COS в синтетическом газе [8] измеряли экстинкцию ИК-абсорбционной полосы при 2054 см^{-1} . В области 0,003—0,05 об.% COS между концентрацией и экстинкцией существует связь, определяемая уравнением $y = a + bx$. Вследствие незначительности члена a следует проверить, возможно ли уравнение $y = b'x$. Имеются следующие значения:

$$y = 0,101 + 115,567x \quad y = 119,625x$$

$$s_0^2 (m - 2) = 0,4107 \quad s'_0{}^2 (m - 1) = 0,4682$$

$$m = 19$$

Результаты представляют следующей схемой:

	QS	f	s ²
$y = b'x$	0,4682	18	
$y = a + bx$	0,4107	17	0,0242
Разность	0,0575	1	

$$F = \frac{0,0575}{0,0242} = 2,38$$

Из табл. 12.5 получают $F(\bar{P} = 0,95; f_1 = 1; f_2 = 17) = 4,45$. Между обеими дисперсиями нет значимой разницы, следовательно, переход к уравнению $y = b'x$ допустим.

Различие между факторами для двух прямых вида $y = a + bx$ можно проверить так же, как проверяют различие двух средних (ср. разд. 7.3). В соответствии с равенством (9.5) объединяют полученные суммы квадратов и подсчитывают общую дисперсию

$$s_g^2 = \frac{QS_1 + QS_2}{m_1 + m_2 - 4} \quad (9.17)$$

Дисперсию разности $|b_1 - b_2|$ получают

$$s_d^2 = s_g^2 \left[\frac{1}{\sum (x_{1i} - \bar{x}_1)^2} + \frac{1}{\sum (x_{2i} - \bar{x}_2)^2} \right] =$$

$$= s_g^2 \left[\frac{1}{\sum x_{1i}^2 - (\sum x_{1i})^2/m_1} + \frac{1}{\sum x_{2i}^2 - (\sum x_{2i})^2/m_2} \right] \quad (9.18)$$

с $f = m_1 + m_2 - 4$ степеней свободы.

Для проверки значимости разности $|b_1 - b_2|$ образуют

$$t = \frac{|b_1 - b_2|}{s_d} \quad (9.19)$$

и сравнивают обычным образом с $t(P, f)$.

Для прямых $y = bx$ эти уравнения упрощаются:

$$s_g^2 = \frac{QS_1 + QS_2}{m_1 + m_2 - 2} \quad (9.17a)$$

$$s_d^2 = s_g^2 \left[\frac{1}{\sum x_{1i}^2} + \frac{1}{\sum x_{2i}^2} \right] \quad (9.18a)$$

[9.3]. Установка титра 0,1 н. раствора перманганата калия была проведена один раз по оксалату натрия и параллельно по окиси железа(III) по Бранду. Между значением x (мг оттитрованного вещества) и расходом y (пересчитанным на 1 мг оттитрованного вещества) существовала зависимость. Полученные результаты выглядели следующим образом:

Оксалат натрия		Окись железа(III)	
x	y	x	y
136,2	140,0	141,3	152,5
161,2	171,5	203,5	208,4
200,3	207,7	242,0	253,1
235,5	244,6	283,1	191,7
271,1	285,4	327,6	345,0
		362,0	370,0

Подсчитывают:

	Оксалат натрия	Окись железа(III)
$\sum x_i y_i$	223291,35	454750,12
$\sum x_i^2$	213611,43	438453,31
$\sum y_i^2$	233433,86	471760,31
m	5	6
b	1,045316	1,037169
$s_b^2(m-1)$	23,84	107,59

При помощи соотношений (9.17a) и (9.18a) получают

$$s_g^2 = \frac{23,84 + 107,59}{5 + 6 - 2} = 14,60$$

$$s_d^2 = 14,60 [(1/213611,43) + (1/438453,31)] = 0,001016473$$

$$s_d = 0,0319$$

$$t = \frac{1,0453 - 1,0372}{0,0319} = 0,25$$

С $f = 5 + 6 - 2 = 9$ степеней свободы $t(\bar{P} = 0,95; f = 9) = 2,26$. Между полученными значениями титра для раствора перманганата нет значимой разницы.

9.3. Проверка гипотезы линейности

Не всегда заранее можно утверждать, что имеет место линейная зависимость. Для проверки этой гипотезы для каждой из m заданных величин x_i нужно провести по n_j параллельных определений. Найденная при этом случайная ошибка не будет находиться в противоречии

с разбросом результатов измерений относительно уравнения прямой (если имеет место линейная зависимость). Проверку проводят простым дисперсионным анализом. При этом суммарное рассеяние раскладывают на разброс «между параллельными определениями» и разброс «внутри параллельных определений». Сумму квадратов «между параллельными определениями» можно дальше разбить следующим образом:

$$\sum_1^m n_j (\bar{y}_j - \bar{y})^2 = \sum_1^m n_j (\bar{y}_j - Y_j)^2 + \sum_1^m n_j (Y_j - \bar{y})^2$$

Первое слагаемое соответствует разбросу наблюдаемых средних значений \bar{y}_j относительно вычисленных Y_j . Второе слагаемое характеризует разброс значений регрессии Y_j относительно общего среднего значения \bar{y} . Получается следующее разбиение:

Источники рассеяния	Сумма квадратов	Степень свободы	Дисперсия
Разброс между параллельными определениями (из дисперсионного анализа)	$QS_1 = \sum_1^m n_j (\bar{y}_j - \bar{y})^2 =$ $= \sum \frac{(\sum y_{ji})^2}{n_j} - \frac{(\sum y_{ji})^2}{n}$	$f_1 = m - 1$	$s_1^2 = \frac{QS_1}{f_1}$
Разброс значений регрессии	$QS_{11} = \sum_1^m n_j (Y_j - \bar{y})^2 =$ $= \frac{\left(\sum x_j y_{ji} - \frac{\sum x_j \sum y_{ji}}{n} \right)^2}{\sum x_j^2 - \frac{(\sum x_j)^2}{n}}$	$f_{11} = 1$	
Разброс средних значений	$QS_{12} = \sum_1^m n_j (\bar{y}_j - Y_j)^2 =$ $= QS_1 - QS_{11}$	$f_{12} = m - 2$	$s_{12}^2 = \frac{QS_{12}}{f_{12}}$

Составляют отношение

$$F = \frac{\text{Дисперсия «разброса средних значений»}}{\text{Дисперсия «разброса внутри параллельных определений»}} = \frac{s_1^2}{s_2^2}$$

Дисперсию для разброса «внутри параллельных определений» находят методом дисперсионного анализа и сравнивают найденное значение F с табличным $F(\bar{P}, f_{12}, f_2)$. Если $F < F(\bar{P}, f_{12}, f_2)$, то наблюдения не находятся в противоречии с проверяемой гипотезой о наличии линейной регрессии. Однако это не значит, что прямая представляет единственно возможную линию регрессии.

[9.4]. Для построения градуировочного графика при спектрохимическом определении олова для заданного содержания олова ($=x_j$) были измерены следующие разности почернения ΔS :

% Sn lg 10 x_j	0,24 0,380	0,31 0,491	0,40 0,602	0,60 0,778	0,76 0,881	0,90 0,954
Измеренные ΔS	-0,148 -0,153 -0,158	-0,096 -0,080 -0,082	-0,028 -0,040 -0,035	+0,068 +0,072 +0,064	+0,122 +0,120 +0,125	+0,140 +0,148 +0,145

Следует проверить, имеется ли линейное соотношение между логарифмами концентрации и разностями почернения. Для этого прежде всего проводят простой дисперсионный анализ. Пользуясь преобразованием $Y = 1000\Delta S$, получают

Источник рассеяния	Сумма квадратов	Степень свободы	Дисперсия
Разброс между параллельными определениями	215335	5	43 067,0
Разброс внутри параллельных определений	352	12	29,3
Общий разброс	215 687	17	—

Разброс «между параллельными определениями» подвергают дальнейшему разложению. Пользуясь преобразованием $X_j = 1000 \lg 10x_j$, вычисляют

$$\begin{aligned} \sum X_j &= 12\,258 & \sum X_j^2 &= 9\,118\,338 \\ \sum_{m=6} X_{ji} &= 184 & \sum_{n=18} X_j Y_{ji} &= 532\,017 \end{aligned}$$

Отсюда получают сумму квадратов для разброса значений регрессий

$$QS_{11} = \frac{\left(532\,017 - \frac{12\,258 \cdot 184}{18}\right)^2}{9\,118\,338 - \frac{12\,258^2}{18}} = 214\,647$$

В результате имеем

Источники рассеяния	Сумма квадратов	Степень свободы	Дисперсия
Разброс между параллельными определениями	215 335	5	—
Разброс значений регрессии	214 647	1	214647,0
Разброс средних значений	688	4	172,0

При значении $s_2^2 = 29,3$, взятом из дисперсионного анализа, составляют

$$F = \frac{172,0}{29,3} = 5,87$$

При $\bar{P} = 0,99$ с $f_{12} = 4$ и $f_2 = 12$ $F(\bar{P}, f_{12}, f_2) = 5,41$. Измеренные значения не согласуются с гипотезой о линейности.

Если проверкой установлено, что гипотеза линейности не выполняется, то пытаются подходящим образом преобразовать результаты измерений. Во многих случаях можно провести логарифмическое преобразование. В полулогарифмической сетке прямой линией будет представлена экспоненциальная функция $y = ab^x$ или ее обратная функция в зависимости от того, по какой из осей откладываются логарифмы. На двойной логарифмической сетке

прямой будет представлена функция типа $y = ax^n$. В особых случаях можно также использовать другие преобразования (например, обратная величина температуры при измерениях давления пара). В целях упрощения всегда пытаются получить линейную связь посредством подходящего преобразования переменных. По возможности все преобразования следует обосновывать теоретически (например, логарифмическое преобразование оси времени при измерениях времени, ср. стр. 24). Если речь идет о чисто эмпирическом соотношении, то это надо всегда отмечать. При всех подобных преобразованиях следует учесть, что может быть нарушено нормальное распределение. Из-за этого могут возникнуть затруднения с представлением результатов. Если рассматриваемую связь не удается преобразованием привести к линейному виду, то ее можно приблизить полиномом второй степени $y = a + bx + cx^2$. Три константы a , b и c этого уравнения находят решением следующей системы уравнений:

$$\begin{aligned} \sum y_i &= ta + b \sum x_i + c \sum x_i^2 \\ \sum x_i y_i &= a \sum x_i + b \sum x_i^2 + c \sum x_i^3 \\ \sum x_i^2 y_i &= a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 \end{aligned}$$

Квадратичное приближение часто оказывается достаточным, особенно в ограниченной области.

9.4. Оценка вторичных аналитических методов

Методы, чаще всего применяемые в инструментальном анализе, нуждаются в градуировке. Для оценки этих методов особенно эффективен метод регрессии. Обрабатывая результаты анализа, полученные в опытах, проводимых для построения градуировочного графика, сразу получают оценку для случайной ошибки, для значения «холостого опыта» и для границы обнаружения.

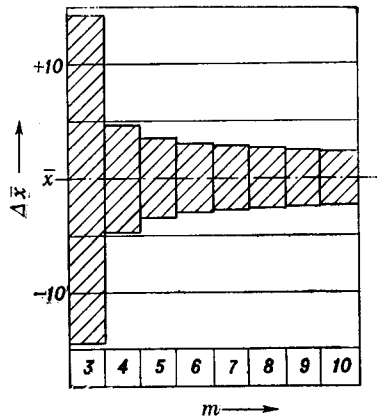
Если градуировочный график задается прямой $y = a + bx$, то ошибка метода анализа состоит из трех частных ошибок. Здесь имеет место ошибка градуировки (ошибки для констант a и b) и ошибка, которая появляется при определении значения измерения y_A . Эти три

частные ошибки суммируют по закону распространения ошибок. Если каждая из m эталонных проб анализируется без повторений, а проба с неизвестным содержанием анализируется n_j раз и из этих n_j параллельных определений образовано среднее \bar{y}_A , то среднюю квадратичную ошибку для метода с s_0 получают по соотношению (9.4)

$$s = \frac{1}{b} \sqrt{s_0^2 \left[\frac{1}{n_j} + \frac{1}{m} + \frac{(\bar{y}_A - \bar{y})^2}{b^2 \sum (x_i - \bar{x})^2} \right]} = \frac{1}{b} \sqrt{s_0^2 \left[\frac{1}{n_j} + \frac{1}{m} + \frac{m(\bar{y}_A - \bar{y})^2}{b^2 [m \sum x_i^2 - (\sum x_i)^2]} \right]} \quad (9.20)$$

с $f = m - 2$ степенями свободы.

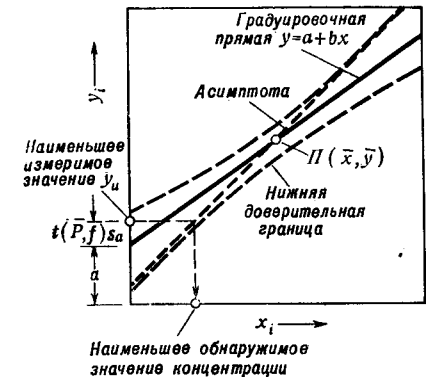
На ошибку, которой следует ожидать при анализе, судя по соотношению (9.20), влияет фактор b градуировочной прямой, число эталонных проб m и величина значе-



Р и с. 9.5. Доверительный интервал для вторичных методов анализа при $s = 1$ и $P = 0,95$ в зависимости от числа эталонных проб m .

ния y_A . Чем круче идет градуировочная кривая — следовательно, чем больше коэффициент регрессии b , — тем меньше ошибка анализа. Число эталонных проб входит в расчет с $f = m - 2$ степеней свободы. Поэтому для градуировки рекомендуется применять по меньшей мере пять проб (рис. 9.5). Вообще эталонные пробы равномерно рас-

пределяют в исследуемой области концентраций. Однако если ошибка измерения y зависит от величины значения измерения (например, в фотометрии), то в области высоких значений ошибки измерения следует предусмотреть дополнительные эталонные пробы. Средняя квадратичная ошибка



Р и с. 9.6. Доверительные границы и наименьшее обнаружимое значение концентрации при вторичных методах анализа.

далее определяется разностью $\bar{y}_A - \bar{y}$, причем $\bar{y} = \sum y_i / m$. Она будет наименьшей тогда, когда $\bar{y}_A = \bar{y}$. Она будет тем больше, чем больше среднее значение \bar{y}_A отличается от \bar{y} . Поэтому градуировочный график можно использовать только в той области концентраций, где лежат эталонные пробы. Эта область концентраций не должна быть слишком растянутой.

Пользуясь средней квадратичной ошибкой, обычным образом записывают доверительный интервал для найденного значения концентрации. Границы доверительного интервала в зависимости от значения концентрации x на рис. 9.6 даются двумя ветвями гиперболы. Они лежат симметрично относительно линии регрессии и прямой $y = \bar{y} - \frac{1}{b}(x - \bar{x})$, а также прямой, проходящей через точку $\Pi(\bar{x}, \bar{y})$, и перпендикулярной линии регрессии.

[9.5]. При фотометрическом определении бензола в этиловом спирте (ср. пример [9.1]) в УФ-области спектра была измерена экстинкция в трех параллельных определениях

1,52

1,55

1,53

Для среднего значения $\bar{y}_A = 1,53_3$ по градуировочному графику, построенному в примере [9.1], находят:

$$\bar{x}_A = \frac{1,533 - 0,079}{0,570} = 2,55 \text{ г бензола/л}$$

Согласно уравнению (9.20), при $P = 0,95$ получают доверительный интервал:

$$\Delta \bar{x}_A = \frac{2,57}{0,570} \sqrt{0,0000622 \left[\frac{1}{3} + \frac{1}{7} + \frac{7(1,533 - 6,66/7)^2}{0,570^2(7 \cdot 22,79 - 114,49)} \right]} = 0,028 \text{ г бензола/л}$$

Отсюда содержание исследуемой пробы равно $(2,55 \pm 0,03)$ г бензола на 1 л этилового спирта для $P = 0,95$. Если у второй пробы получается среднее значение экстинкции из трех параллельных определений $\bar{y}_A = 0,93$, то концентрация $\bar{x}_A = 1,39$ г бензола/л.

Так как в этом случае $\bar{y}_A \sim \bar{y}$, третий член под корнем в уравнении (9.20) становится малым, при $P = 0,95$ получают искомое содержание $1,39 \pm 0,02$ г бензола/л этилового спирта. В силу того что $\bar{y}_A \sim \bar{y}$, абсолютная ошибка значения анализа теперь становится меньше.

При некоторых методах анализа (например, пламенной спектрометрии) обычно получают логарифмы значений измерений. Оценку метода анализа проводят дальше таким же путем, причем в применяемые уравнения необходимо подставлять логарифмы значений измерений. Средняя квадратичная ошибка и доверительный интервал задают относительную ошибку, так как это уже говорилось выше (ср. стр. 115).

[9.6]. Для построения градуировочного графика при спектрометрическом определении Zn в стабилизированной вольтовой дуге [5] было взято восемь проб известного содержания. Откладывая по одной оси логарифмы измеренных значений, а по другой — известные концентрации, получили прямую. Необходимо определить случайную ошибку метода анализа. Были измерены следующие значения:

x_i , мг Zn/л	y_i , деления шкалы	$X_i = \lg x_i$	$Y_i = \lg y_i$
50	29	1,6990	1,4624
100	47	2,0000	1,6721
150	64	2,1761	1,8062
250	91	2,3979	1,9590
400	130	2,6021	2,1139
500	149	2,6990	2,1732
600	171	2,7782	2,2330
1000	231	3,0000	2,3636

Получили:

$$\begin{aligned} \sum X_i &= 19,3523 & \sum X_i^2 &= 48,145857 & (\sum X_i)^2 &= 374,511515 \\ \sum Y_i &= 15,7834 & \sum Y_i^2 &= 31,798837 & \sum X_i Y_i &= 39,117342 \\ & & m &= 8 & & \end{aligned}$$

и отсюда рассчитали:

$$b = \frac{8 \cdot 39,117342 - 19,3523 \cdot 15,7834}{8 \cdot 48,145857 - 374,511515} = 0,703276$$

$$a = \frac{15,7834 - 0,703276 \cdot 19,3523}{8} = 0,271674$$

$$s_y^2(m-2) = 31,798837 - 0,271674 \cdot 15,7834 - 0,703276 \cdot 39,117342 = 0,0006108$$

$$s_y^2 = 0,0001018$$

Для $\bar{y}_A = \bar{y}$ при $n_j = 3$ параллельных определениях из уравнения (9.20) получают логарифмическую среднюю квадратичную ошибку для значения измерения $\bar{y}_A = \bar{y}$.

$$s_{1g} = \frac{1}{0,703276} \sqrt{0,0001018 \left(\frac{1}{3} + \frac{1}{8} \right)} = 0,00971$$

Потенцирование приводит к $s = 1,023$. Средняя квадратичная ошибка в середине области измерения составляет 2,3% (отн.). Пользуясь логарифмической средней квадратичной ошибкой, находят доверительный интервал для $P = 0,95$ для среднего значения из $n_j = 3$ параллельных определений.

$$\Delta \bar{X} = 2,45 \cdot 0,00971 = 0,0237$$

Из этого получают $\Delta \bar{x} = 1,056$. Доверительный интервал для значения концентрации, соответствующей $\bar{Y}_A = \bar{Y}$, составляет $\pm 5,6\%$ (отн.) при $P = 0,95$.

Ранее уже было указано (ср. разд. 4.3), что при анализе граница обнаружения задается значением холостого опыта и его случайными отклонениями. Пользуясь уравнением регрессии, всегда можно найти нужное значение зависимой переменной y . Значение y для холостого опыта можно получить, экстраполируя прямую к значению концентрации $x = 0$. Если между значениями концентрации и значениями измерений существует линейная связь $y = a + bx$, то получают наименьшее значение измерения y_u в виде

$$y_u = a + t(\bar{P}, f) s_a. \quad (9.21)$$

Для характеристики метода анализа нужно знать то наименьшее значение концентрации x_u , которое может

быть обнаружено этим методом. При переходе от измеренных значений y к концентрации x необходимо принимать во внимание доверительный интервал линии регрессии (рис. 9.6). Так, следовало бы прямую $y_u = a + t(\bar{P}, f) s_a$ провести до пересечения с ветвью гиперболы, лежащей ниже линии регрессии. Получающееся значение абсциссы представляет наименьшее определенное значение концентрации с вероятностью \bar{P} (односторонняя постановка задачи). Однако на практике этот расчет довольно труден. Если ожидаемое наименьшее значение концентрации x_u лежит вблизи нулевой точки и достаточно удалено от x , вместо гиперболы приближенно можно использовать соответствующую асимптоту. Уравнение асимптоты

$$y = a + bx + t(\bar{P}, f) s_b (x - \bar{x}). \quad (9.22)$$

Подставляя наименьшее определенное значение измерения y_u , найденное из соотношения (9.21), получают

$$x_u = \frac{t(\bar{P}, f) [s_a + \bar{x}s_b]}{b + t(\bar{P}, f) s_b} \quad (9.23)$$

Интересно обратить внимание на то, что в соотношении (9.23) значение холостого опыта не оказывает никакого влияния на наименьшее определенное значение концентрации. Оно определяется величиной случайных ошибок s_a и s_b , наклоном b градуировочной прямой и средним значением \bar{x} . Число эталонных проб мало влияет на величину $t(\bar{P}, f)$. Требуемое малое среднее значение \bar{x} означает, что область концентраций не может быть слишком широкой. Условия эксперимента выбирают так, чтобы получить возможно больший коэффициент регрессии b (о графическом изображении градуировочной прямой, см. также гл. 10). Доверительные границы в большинстве случаев строят для $\bar{P} = 0,99$. Так как при этой вероятности на величину $t(\bar{P}, f)$ ниже $f = 10$ сравнительно сильно влияет число степеней свободы, для установления доверительной границы следовало бы взять по меньшей мере двенадцать эталонных проб.

Если при помощи вторичных методов анализа необходимо дать гарантию чистоты продукта, то в соответствии с разд. 6.2 следует также отдельно учитывать случай-

ную ошибку метода анализа. Для условия $y > y_u$ минимальную чистоту гарантирует выражение

$$T_g = \frac{y - a}{b} + t(\bar{P}, f) s \quad (9.24a)$$

причем s вычисляют по уравнению (9.20). В случае $y < y_u$ гарантированный предел получается в виде

$$T_g = x_u + t(\bar{P}, f) s \quad (9.24б)$$

На основании того, что $y < y_u$, ни в коем случае нельзя делать вывод, что содержание $x = 0$

[9.7]. Для определения бензола в спирте (см. пример [9.2]) в качестве минимально определяемого среднего значения для $\bar{P} = 0,99$ получают

$$y_u = 0,079 + 3,37 \cdot 0,0056 = 0,098$$

Из соотношения (9.23) для минимально определяемого значения концентрации получают

$$x_u = \frac{3,37 \left(0,0056 - \frac{10,7}{7} \cdot 0,0031 \right)}{0,570 + 3,37 \cdot 0,0031} = 0,060 \text{ г бензола/л}$$

Все эти положения о расчете доверительной границы справедливы только тогда, когда измеренные величины следуют нормальному распределению. При наличии логарифмического нормального распределения можно получить доверительную границу, когда значения холостого опыта измеряют достаточно часто и в соответствии с уравнением (4.16) подсчитывают величины μ_v и σ_v , а затем, пользуясь градуировочным графиком, пересчитывают их в масштабе концентраций. Так же это справедливо, если значение холостого опыта было получено на регистрограмме (см. пример [5.5]).

Если предполагаемая здесь линейная зависимость между значением измерения и концентрацией не выполняется, то должна иметь место параболическая зависимость. Подробности о построении и оценке градуировочных графиков изложены Манделем и Линнигом [14].

9.5. Определение систематических ошибок

Систематические ошибки влияют на результаты измерений таким образом, что они постоянно отклоняются

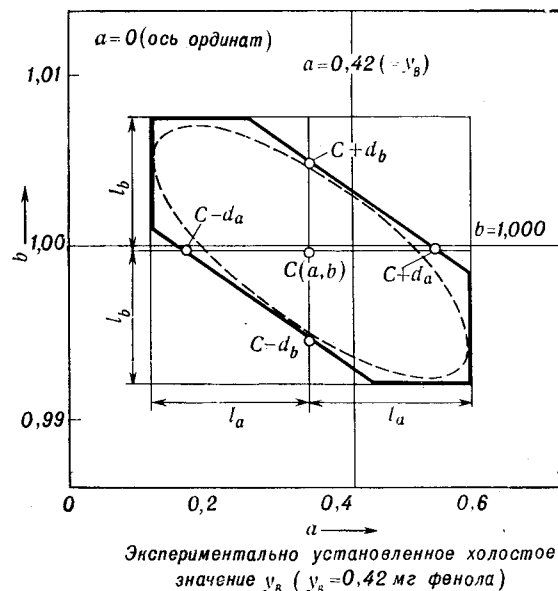
в одну сторону от истинного значения. Метод дает неверные значения [11]. Для оценки метода анализа важно знать вид появляющейся систематической ошибки, так как это позволит сделать вывод о причине ее возникновения и в дальнейшем, по возможности, улучшить метод.

Обстоятельная проверка систематических ошибок осуществляется методом регрессионного анализа. Найденные для каждой пробы содержания y_i сопоставляются с эталонными, безошибочно определенными содержаниями вещества в образце x_i . В отсутствие случайных и систематических ошибок следовало бы ожидать линейной регрессии с $a = 0$ и $b = 1,000$. Однако из-за случайных ошибок эти константы в большинстве случаев отклоняются от идеальных значений. Тогда следует проверить, сопоставимы ли разности $|a - 0|$ и $|b - 1|$ со случайной ошибкой, или они должны вызываться дополнительным влиянием систематических ошибок.

Математическое обсуждение этой проблемы показывает, что a и b всегда в слабой степени коррелированы отрицательно, поскольку они были подсчитаны из одной и той же серии данных [14]. Поэтому разности $|a - 0|$ и $|b - 1|$ следует рассматривать одновременно при проверке гипотезы о значимости систематических ошибок. Это приводит к двумерному распределению (см. разд. 2.3) со случайными переменными a и b . Строят контурный эллипс (основание двумерного распределения) с центром тяжести $C(a, b)$. Внутри этого эллипса должно быть $100P\%$ всех пар значений $\Pi(a_i, b_i)$, которые с вероятностью P принадлежат соответствующему двумерному распределению. Постоянная ошибка будет иметь место тогда, когда прямая $a = 0$ не пересекает этого эллипса. Это значит, что точка $\Pi(a = 0; b_i)$ не принадлежит рассматриваемому двумерному распределению. Аналогичная картина будет наблюдаться, если принять, что $b = 1$, когда в действительности имеет место линейно изменяющаяся ошибка. Если хотят исследовать, приводит ли экспериментально найденная корректировка (например, значение холостого опыта) к систематической ошибке, то в рассматриваемой системе координат строят соответствующую прямую. Для экспериментально найденного холостого значения y_B это будет прямая $a = y_B$.

Исправление не приводит к появлению систематической ошибки, если прямая пересекает контур эллипса.

Расчет констант a и b и определение суммы квадратов $s_0^2(m - 2)$ при сглаживании значений $y_i = f(x_i)$ производится при помощи соотношений, приведенных



Р и с. 9.7. Построение шестиугольника касательных и проверка систематических ошибок.

в разд. 9.1. Ради упрощения необходимый для проверки контурный эллипс заменяют тремя парами параллельных касательных. Для этого в рассматриваемой системе координат строят прямоугольник со сторонами

$$l_b = \sqrt{2F(\bar{P}; f_1 = 2; f_2 = m - 2) s_0^2} \quad (9.25a)$$

по оси ординат и

$$l_a = \sqrt{2F(\bar{P}; f_1 = 2; f_2 = m - 2) s_0^2} \quad (9.25b)$$

по оси абсцисс (рис. 9.7). Центр тяжести C этого прямоугольника лежит в точке $C(x = a, y = b)$. [s_0^2 и s_a^2 под-

считывают по уравнениям (9.6) и (9.7)]. При построении прямоугольника следует так выбирать масштаб, чтобы стороны l_a и l_b были примерно равны по величине. Для построения касательных центр тяжести принимают за точку пересечения осей. Затем наносят точки $C \pm d_b$ на вертикальную и $C \pm d_a$ на горизонтальную оси. При этом, если применить сокращенное обозначение $\sqrt{m \sum x_i^2} = z$, то получают

$$d_b = \sqrt{\frac{4zs_0^2 F(\bar{P}; f_1=2; f_2=m-2)}{(\sum x_i \sum x_i^2) + z}} \quad (9.26a)$$

$$d_a = d_b \frac{l_a}{l_b} \quad (9.26b)$$

Обе параллельные прямые, проходящие через точки $C + d_b$ и $C + d_a$ и $C - d_b$ и $C - d_a$, являются касательными к искомому эллипсу. Они включают доверительный интервал двумерного распределения при выбранной вероятности \bar{P} . Для проверки систематической ошибки строят прямые $a = 0$ (при наличии постоянной ошибки) и $b = 1,000$ (при наличии линейно изменяющейся ошибки). Систематическая ошибка существует тогда, когда прямая не пересекает эллипс или шестиугольник касательных.

[9.8]. При бромометрическом титровании фенола были получены следующие значения (мг фенола):

Дано (x_i)	Найдено (y_i)	Дано (x_i)	Найдено (y_i)
3,72	4,12	29,76	30,20
7,44	7,77	37,24	37,51
14,88	15,28	44,67	45,02
22,32	22,58	52,04	52,44

Значение холостого опыта для $x = 0$ равно $y_b = 0,42$ мг фенола.

Необходимо проверить, должно ли найденное значение холостого опыта указывать на статистически значимую постоянную ошибку и появляется ли, кроме того, еще линейно изменяющаяся ошибка. Далее интересно, можно ли экспериментально найденное

значение холостого опыта взять в качестве безупречной поправки для компенсации этой постоянной ошибки. Подсчитывают

$$\begin{aligned} \sum x_i &= 212,07 & \sum x_i^2 &= 7764,8345 & (\sum x_i)^2 &= 44\,973,6849 \\ \sum y_i &= 214,92 & \sum y_i^2 &= 7916,4762 & \sum x_i y_i &= 7840,1326 \\ & & m &= 8 & & \end{aligned}$$

Константы прямых [уравнения (9.2) и (9.3)]

$$b = \frac{8 \cdot 7840,1326 - 212,07 \cdot 214,92}{8 \cdot 7764,8345 - 44\,973,6849} = 0,999882$$

$$a = \frac{214,92 - 0,999882 \cdot 212,07}{8} = 0,359378$$

Разброс значений измерений вокруг прямой [уравнение (9.5)]

$$s_0^2(m-2) = 7916,476200 - 0,359378 \cdot 214,92 - 0,999882 \cdot 7840,1326 = 0,031216$$

$$s_0^2 = 0,005203$$

Разброс констант a и b [уравнения (9.6) и (9.7)]

$$s_b^2 = \frac{8 \cdot 0,005203}{8 \cdot 7764,8345 - 44\,973,6849} = 0,000002428$$

$$s_a^2 = \frac{0,005203 \cdot 7764,8345}{8 \cdot 7764,8345 - 44\,973,6849} = 0,002356$$

Построение касательных [уравнения (9.25) и (9.26) для $\bar{P} = 0,99$].

$$l_b = \sqrt{2 \cdot 10,92 \cdot 0,000002428} = 0,007283$$

$$l_a = \sqrt{2 \cdot 10,92 \cdot 0,002356} = 0,2268$$

$$z = \sqrt{8 \cdot 7764,8345} = 249,24$$

$$d_b = \sqrt{\frac{4 \cdot 249,24 \cdot 0,005203 \cdot 10,92}{212,07 \cdot 7764,8345 + 249,24}} = 0,005867$$

$$d_a = 0,005867 \frac{0,2268}{0,007283} = 0,1827$$

Графическая проверка (рис. 9.7.).

Ось ординат, которая соответствует прямой $a = 0$, лежит далеко вне доверительного интервала для $\bar{P} = 0,99$, т. е. найденное значение холостого опыта вызывается действием постоянной ошибки. Прямая $b = 1,000$ пересекает контур эллипса прямо в центре тяжести. Следовательно, линейно изменяющаяся ошибка не появляется. Прямая $y_b = 0,42$ пересекает контур эллипса. Экспериментально найденное значение холостого опыта представляет действительную и адекватную поправку, компенсирующую постоянную ошибку.

Проверка систематической ошибки значительно упрощается, если имеется только постоянная или изменяющаяся ошибка. Соответствующие сведения получают

из предварительной графической проверки. Для этого разности между заданными и найденными значениями наносят на график, упорядочивая их по величине x .

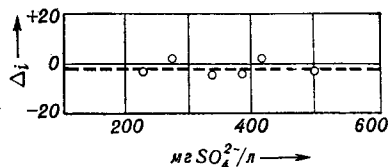
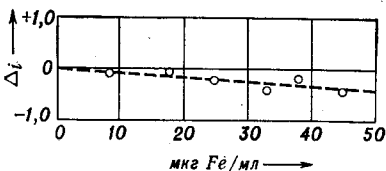


Рис. 9.8. Рассеяние результатов измерений при появлении постоянной ошибки.

Если эти точки рассеяны параллельно оси абсцисс (рис. 9.8), то следует ожидать только постоянной ошибки. Если получается восходящее или убывающее множество точек (рис. 9.9), то это указывает на линейно изменяющуюся ошибку.

Рис. 9.9. Рассеяние результатов измерений при появлении линейно изменяющейся ошибки.



Для проверки постоянной ошибки для каждой пары значений заданного и найденного значения получают разность $d_i = y_i - x_i$ (y_i — найденное значение, x_i — образец). Отклонение среднего значения $\bar{d} = \sum d_i / m$ от ожидаемого идеального значения $d_0 = 0$ проверяют по уравнению (7.8). Получается следующая схема расчета:

Проба	Заданное (x_i)	Найденное (y_i)	Разность
1	x_1	y_1	$d_1 = y_1 - x_1$
2	x_2	y_2	$d_2 = y_2 - x_2$
⋮	⋮	⋮	⋮
m	x_m	y_m	$d_m = y_m - x_m$
		Среднее значение	$\bar{d} = \sum d_i / m$

Проверяют среднее значение \bar{d} относительно идеального значения $d_0 = 0$ [соотношение (7.8)], вычисляя

$$t = \frac{|\bar{d}|}{s_d} \sqrt{m}$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{m - 1}}$$

со степенями свободы $f = m - 1$.

Найденное значение t обычным образом сравнивают с табличным значением $t(P, f)$. Постоянная ошибка будет тогда, когда $t > t(P, f)$.

[9.9]. Для проверки нового объемного определения сульфатов в воде были проанализированы объемным методом гравиметрически проверенные пробы воды [7]. Получены следующие значения:

Данные (x_i)	Найденные (y_i)	Разность $d_i = y_i - x_i$
226,3	223,3	-3,0
273,5	274,0	+0,5
338,5	333,5	-5,0
386,6	382,1	-4,5
416,0	416,4	+0,4
502,0	499,2	-2,8
Среднее значение $\bar{d} = -2,40$		

При построении графика зависимости разностей от содержания водорода в образце получается семейство точек, параллельное оси абсцисс (рис. 9.8). Следовательно, надо проверить только постоянную ошибку. По приведенной схеме расчета получают:

$$s_d = \sqrt{\frac{3,0^2 + 0,5^2 + \dots - 14,4^2/6}{6 - 1}} = 2,36$$

$$t = \frac{2,40}{2,36} \sqrt{6} = 2,49$$

При $P = 0,95$ и $f = m - 1 = 5$ $t(P, f) = 2,57$. Так как $t < t(P, f)$, в методе, несмотря на отклонение среднего значения \bar{d} от нуля, нет значимой постоянной ошибки.

Для определения линейно изменяющейся ошибки сравнивают найденные значения y_i с заданным содержа-

нием вещества в образце x_i . Здесь справедливо соотношение

$$y_i = bx_i$$

Отклонение коэффициента регрессии b от идеального значения $b_0 = 1,000$ проверяют по

$$t = \frac{|1-b|}{s_b}$$

s_b вычисляют по соотношению (9.14). Линейно изменяющаяся ошибка будет, если $t > t(P, f)$.

[9.10]. При проверке фотометрического метода определения железа было исследовано шесть проб известного содержания. Найдены следующие значения (мкг Fe/мл):

Заданные (x_i)	Найденные (y_i)	Разность $d_i = y_i - x_i$
8,5	8,4	-0,1
18,0	17,9	-0,1
25,0	24,8	-0,2
33,0	32,6	-0,4
38,0	37,8	-0,2
45,0	44,6	-0,4

При построении разностей в зависимости от содержания веществ в образце x получается семейство точек (рис. 9.9). Поэтому проверяют лишь линейно изменяющуюся ошибку. Вычисляют

$$\sum x_i^2 = 5579,25 \quad \sum x_i y_i = 5532,80$$

$$\sum y_i^2 = 5486,77 \quad m = 6$$

Из этого по соотношениям (9.11) и (9.14) получают

$$b = \frac{5532,80}{5579,25} = 0,991675$$

$$s_0^2(m-1) = 5486,77 - 0,991675 \cdot 5532,80 = 0,0306$$

$$s_0^2 = 0,00612$$

$$s_b = \sqrt{\frac{0,00612}{5579,25}} = 0,00105$$

$$t = \frac{1,000000 - 0,991675}{0,00105} = 7,93$$

Из табл. 12.3 получают $t(P = 0,99; f = 5) = 4,03$. Так как $t > t(P, f)$, имеется линейно изменяющаяся ошибка.

Во всех описанных выше методах для определения систематических ошибок предполагали, что содержание исследуемых проб известно. Это требование должно выполняться при испытании метода анализа, однако оно может не осуществляться в обычном производстве. Метод проверки, при котором можно производить контроль непосредственно по текущим анализам, не зная истинного содержания, предполагает следующие условия [2]:

1. Измеренная величина (например, мг осадка или мл титруемой жидкости) относится к известной навеске.
2. Между измеренной величиной x и навеской e существует пропорциональность ($x \sim e$).
3. Определяемый элемент можно прибавить к анализу в точно известном количестве.

Для определения постоянной ошибки исходят из параллельного определения двух разных по величине навесок *. Если анализы выполняются без ошибки, то в соответствии с предположением 1 имеет место соотношение

$$\frac{x_1}{e_1} = \frac{x_2}{e_2} \quad (9.27)$$

Когда появляется постоянная ошибка a , то

$$\begin{aligned} x'_1 &= x_1 + a \\ x'_2 &= x_2 + a \end{aligned} \quad (9.28)$$

Подстановкой в уравнение (9.27) и решением относительно a , получают

$$a = \frac{x'_2 e_1 - x'_1 e_2}{e_1 - e_2} \quad (9.29)$$

Особенно нагляден расчет для $e_1 = 2e_2$. Уравнение (9.29) переходит тогда в выражение

$$a = 2x'_2 - x'_1 \quad (9.30)$$

Для проверки линейно изменяющейся ошибки оба определения следует проводить на одинаковой навеске. При $e_2 = e_3$ будет также $x_2 = x_3 = x$. При одном из ана-

* Из параллельных определений с различной навеской можно найти значение холостого опыта, не прибегая к общепринятому методу — анализу «пустой» пробы.

лизов добавляют определяемый элемент в количестве z . Эта добавка должна быть одинакова по величине для всех проб, ее выбирают так, чтобы она почти удвоила концентрацию требуемого элемента. Если метод содержит линейно изменяющуюся ошибку b , то

$$x'_2 = bx$$

$$x'_3 = b(x + z) \quad (9.31)$$

Исключая неизвестное x и решая относительно b , получают

$$b = \frac{x'_3 - x'_2}{z} \quad (9.32)$$

Обе константы a и b определяют из ряда m проб и вычисляют среднее $\bar{a} = \sum a_i/m$ и $\bar{b} = \sum b_i/m$. Вследствие случайной ошибки чаще всего получают отклонения от ожидаемых идеальных значений $a_0 = 0$ и $b_0 = 1,000$. Для доказательства систематического характера ошибки надо сопоставить \bar{a} и \bar{b} с их возможными идеальными значениями. По соотношению (7.8) получают

$$t_a = \frac{|\bar{a}|}{s_a} \sqrt{m} \quad t_b = \frac{|1 - \bar{b}|}{s_b} \sqrt{m} \quad (9.33)$$

с $f = m - 1$ степеней свободы.

При этом

$$s_a = \sqrt{\frac{\sum (a_i - \bar{a})^2}{m-1}} \quad s_b = \sqrt{\frac{\sum (b_i - \bar{b})^2}{m-1}}$$

Статистически значимое систематическое отклонение будет признано, если $t_a > t(P, f)$ и $t_b > t(P, f)$. Отдельная проверка обоих видов ошибок (см. стр. 198) здесь возможна, так как постоянную и изменяющуюся ошибку подсчитывают не по одним и тем же результатам анализа.

[9.11.]. При определении серебра по Фольгарду в целях проверки были умышленно сделаны постоянная и изменяющаяся ошибки [2]. Были заданы $a = -0,1$ мл 0,1 н. раствора AgNO_3 и $b = 0,990$. Затем проанализированы четыре пробы; $e_1 : e_2 : e_3 = 2 : 1 : 1$ и $z = 10$ мл 0,1 н. раствора AgNO_3 . Градуировочную добавку

ввели к третьему анализу. При титровании были получены следующие значения:

Проба	Измеренные значения			a , мл	e
	x'_1	x'_2	x'_3		
1	15,71	7,79	17,72	-0,13	0,993
2	23,58	11,75	21,68	-0,08	0,993
3	29,52	14,73	24,66	-0,06	0,993
4	39,43	19,67	29,50	-0,09	0,983

$$\sum a_i = -0,36 \text{ мл} \quad \bar{a} = -0,090 \text{ мл}$$

$$\sum b_i = 3,962 \quad \bar{b} = 0,991$$

$$s_a = \sqrt{\frac{0,13^2 + 0,08^2 + 0,06^2 + 0,09^2 - 0,36^2/4}{4-1}} = 0,029$$

$$t_a = \frac{0,090}{0,029} \sqrt{4} = 6,21$$

Для вычисления s_b из отдельных значений b_i вычитают величину 0,990. Получают:

$$s_b = \sqrt{\frac{0,003^2 + 0,003^2 + 0,003^2 + 0,007^2 - 0,001^2/4}{4-1}} = 0,0050$$

$$t_b = \frac{1,000 - 0,991}{0,0050} \sqrt{4} = 3,60$$

При $f = 3$ степени свободы $t(P, f) = 3,18$ ($P = 0,095$) и $5,84$ ($P = 0,99$). Оказывается, что \bar{a} и \bar{b} отклоняются от идеальных значений более, чем это можно было бы ожидать для случайных величин. Однако систематический характер отклонения достаточно надежно установлен только для постоянной ошибки [$t_a > t(P = 0,99; f)$]. Для бесспорного доказательства наличия изменяющейся ошибки необходимо провести дальнейшие анализы.

Описанный способ расчета применим в тех методах анализа, которые построены по принципу навеска — раствор — измерение. При ходе анализа, в котором мешающие компоненты удаляют осаждением, расчет может оказаться неверным, если осадок имеет склонность захватывать определяемый элемент. Анализы нужно проводить с пробами, близкими по микросоставу. Отдельные анализы надо вести в возможно более случайной последовательности, но при совершенно одних и тех же усло-

виях. Необходимые навески ($e_1:e_2:e_3 = 2:1:1$) наиболее просто получают объемным делением проб (отбором аликвотных частей). При больших сериях анализов различные виды ошибок можно определять на разных пробах. Таким образом можно производить нужный контроль на обычных параллельных определениях без какой-либо дополнительной работы.

9.6. Линейная регрессия при трех и более переменных

Все приведенные ранее рассуждения относились к особому случаю, когда имелись только независимая переменная x и зависимая переменная y . Однако обычно на результаты наблюдений y влияет более одной переменной. Подобные зависимости встречаются в аналитической химии особенно часто при вторичных методах анализа. Так, модуль экстинкции E/l смеси из нескольких компонентов определяется концентрацией всех присутствующих светопоглощающих веществ [уравнение (4.19)]. Целью такого фотометрического многокомпонентного анализа является получение числовых значений для отдельных концентраций. Для смеси из m различных компонентов при этом надо решить систему по меньшей мере из m неидентичных и непротиворечащих уравнений. Результаты анализа становятся значительно надежнее, если берут больше чем m уравнений и производят сглаживание так, чтобы сумма квадратов ошибок стала минимальной.

Для дальнейшего рассуждения предположим, что имеется смесь, состоящая из трех веществ А, В и С. Для каждого из этих трех компонентов в используемой области концентраций выполняется закон Бугера — Ламберта — Бера, т. е. между экстинкцией y и концентрацией x существует пропорциональность. Тогда получают уравнение регрессии

$$y = b_A x_A + b_B x_B + b_C x_C \quad (9.34)$$

Коэффициенты b_A , b_B и b_C называют *частными коэффициентами регрессии*. В случае вторичного фотометрического метода анализа эти коэффициенты предварительно определяют на чистых веществах. Экстинкцию y можно было бы измерить при n разных длинах волн

($n > 3$). Тогда получается система уравнений:

$$\begin{aligned} y_1 &= b_{A_1} x_A + b_{B_1} x_B + b_{C_1} x_C \\ y_2 &= b_{A_2} x_A + b_{B_2} x_B + b_{C_2} x_C \\ &\vdots \\ y_i &= b_{A_i} x_A + b_{B_i} x_B + b_{C_i} x_C \\ &\vdots \\ y_n &= b_{A_n} x_A + b_{B_n} x_B + b_{C_n} x_C \end{aligned} \quad (9.35)$$

В левой части этой системы уравнений стоят измеренные значения (y_i), в правой — вычисляемые значения ($Y_i = b_{A_i} x_A + b_{B_i} x_B + b_{C_i} x_C$).

Разность между двумя значениями дает ошибку. Наилучшее соответствие между измеренными и вычисленными значениями получают тогда, когда соотношение

$$\sum (y_i - Y_i)^2 = \sum (y_i - b_{A_i} x_A - b_{B_i} x_B - b_{C_i} x_C)^2 \quad (9.36)$$

будет минимальным. Это условие можно сформулировать так же, как и в случае простой регрессии [уравнение (9.1)]. Для трех различных веществ получают систему из трех нормальных уравнений (для простоты индексы введены только для трех различных веществ):

$$\begin{aligned} x_A \sum b_A^2 + x_B \sum b_A b_B + x_C \sum b_A b_C - \sum b_A y &= 0 \\ x_A \sum b_A b_B + x_B \sum b_B^2 + x_C \sum b_B b_C - \sum b_B y &= 0 \\ x_A \sum b_A b_C + x_B \sum b_B b_C + x_C \sum b_C^2 - \sum b_C y &= 0 \end{aligned} \quad (9.37)$$

Решение этих нормальных уравнений чаще всего производят методом Гаусса. Общий вид формулы весьма сложен, в то время как математические вычисления легко доступны. Поэтому решение проще показать непосредственно на числовом примере.

[9.12]. Методом спектрофотометрии в УФ-области спектра надо было определить содержание изомеров в технической толуолсульфокислоте [6]. Измерения при шести разных волновых числах

($n = 6$) привели к следующим результатам (коэффициенты экстинкции были определены так, чтобы результат получился в процентах):

$1/\lambda$, см ⁻¹	Коэффициенты экстинкции изомеров			Анализ E/I (=y)
	орто b_A	мета b_B	пара b_C	
36 500	0,505	0,382	0,251	0,375
37 000	0,582	0,520	0,300	0,445
38 000	0,442	0,337	0,241	0,330
40 500	0,122	0,143	0,072	0,105
43 500	1,052	1,305	1,107	1,110
45 000	0,923	1,153	0,086	0,575

Вычисляют:

$$\begin{aligned} \sum b_A^2 &= 2,745350 & \sum b_A b_B &= 3,092289 & \sum b_A y &= 2,298880 \\ \sum b_B^2 &= 3,582776 & \sum b_A b_C &= 1,655783 & \sum b_B y &= 2,612400 \\ \sum b_C^2 &= 1,449111 & \sum b_B b_C &= 1,887188 & \sum b_C y &= 1,592935 \end{aligned}$$

В соответствии с уравнением (9.37) получают следующую систему нормальных уравнений:

$$2,745350x_A + 3,092289x_B + 1,655783x_C - 2,298880 = 0 \quad (I)$$

$$3,092289x_A + 3,582776x_B + 1,887188x_C - 2,612400 = 0 \quad (II)$$

$$1,655783x_A + 1,887188x_B + 1,449111x_C - 1,592935 = 0 \quad (III)$$

Уравнение (I) умножают на коэффициенты

$$-\frac{3,092289}{2,745350} = -1,12637332 \quad -\frac{1,655783}{2,745350} = -0,60312273$$

При этом получают:

$$-3,092289x_A - 3,4830718x_B - 1,8650298x_C + 2,5893971 = 0$$

$$-1,6557830x_A - 1,8650298x_B - 0,9986404x_C + 1,3865068 = 0$$

Первую строку прибавляют к уравнению (II), вторую к уравнению (III). В результате получают простые преобразованные уравнения.

$$0,0997042x_B + 0,0221582x_C - 0,0230029 = 0 \quad (IIa)$$

$$0,0221582x_B + 0,4504706x_C - 0,2064282 = 0 \quad (IIIa)$$

Уравнение (IIa) снова преобразуют, умножая на

$$-\frac{0,0221582}{0,0997042} = -0,2222394$$

При этом получают:

$$-0,0221582x_B - 0,0049244x_C + 0,0051122 = 0$$

Прибавляя к уравнению (IIIa), получают уравнение

$$0,4455462x_C - 0,2013160 = 0 \quad (IIIb)$$

При расчете трех неизвестных получают:

$$\text{из уравнения (IIIb)} \quad x_C = \frac{0,2013160}{0,4455462} = 0,452$$

$$\text{из уравнения (IIa)} \quad x_B = \frac{0,0230029 - 0,0221582 \cdot 0,452}{0,0997042} = 0,130$$

из уравнения (I) $x_A =$

$$= \frac{2,298880 - 3,092289 \cdot 0,130 - 1,655783 \cdot 0,452}{2,745350} = 0,418$$

Следовательно, проба состоит из 41,8% орто-, 13,0% мета- и 45,2% пара-изомеров.

Пример [9.12] показывает, что в этом случае затрачивается немного времени на дополнительные измерения и много усилий на расчет, так как расчеты приходится производить с достаточно большим числом знаков после запятой во избежание ошибок округления. Поэтому для подобных операций целесообразно воспользоваться вычислительной машиной. Лишь в этих условиях метод может оказаться действительно эффективным. При применении электронной машины можно без особого труда вычислить доверительный интервал для каждого результата. Формулы, лежащие в основе такого расчета, приведены Гроссманом [9].

9.7. Проверка взаимной зависимости двух переменных (вычисление корреляции)

Связь между двумя величинами x и y легко определить, если случайная ошибка достаточно мала. При большой случайной ошибке зависимость между двумя величинами может затухать, так как результаты измерения разбросаны внутри более или менее широкой области. Тогда говорят о *стохастической связи*, или, другими словами, что обе величины связаны корреляцией. Сила подобной связи характеризуется *коэффициентом*

том корреляции r . Он является безразмерной величиной, изменяющейся в области $-1 < r < +1$. При $r = +1$ имеет место строго линейная связь, причем с увеличением x растет и y . Если $r = -1$, то связь также строго линейная, но с увеличением одной из величин вторая уменьшается. В случае $r = 0$ x и y считают некоррелированными. Это наблюдается чаще всего тогда, когда x и y независимы друг от друга. Однако из того, что $r = 0$ нельзя сделать обратное заключение — о независимости x и y . Чем ближе r лежит к ± 1 , тем строже наблюдаемая линейная связь. Для обычной линейной корреляции коэффициенты корреляции получают по формуле

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[m \sum x_i^2 - (\sum x_i)^2][m \sum y_i^2 - (\sum y_i)^2]}} \quad (9.38)$$

Проверяется статистическая значимость отличия найденного значения коэффициента корреляции от нуля. Для этого его сопоставляют со значением $r(P, f)$, взятым из табл. 9.1, при степенях свободы $f = m - 2$. Связь имеет место, если $|r| > r(P, f)$.

Таблица 9.1

Граничные значения $r(P, f)$ для проверки статистической значимости коэффициентов корреляции

f	$P = 0,95$	$P = 0,99$	f	$P = 0,95$	$P = 0,99$	f	$P = 0,95$	$P = 0,99$
1	1,00	1,00	11	0,55	0,68	25	0,38	0,49
2	0,95	0,99	12	0,53	0,66	30	0,35	0,45
3	0,88	0,96	13	0,51	0,64	35	0,33	0,42
4	0,81	0,92	14	0,50	0,62	40	0,30	0,39
5	0,75	0,87	15	0,48	0,61	45	0,29	0,37
6	0,71	0,83	16	0,47	0,59	50	0,27	0,35
7	0,67	0,80	17	0,46	0,58	60	0,25	0,33
8	0,63	0,77	18	0,44	0,56	70	0,23	0,30
9	0,60	0,74	19	0,43	0,55	80	0,22	0,28
10	0,58	0,71	20	0,42	0,54	100	0,20	0,25

[9.13]. При применении метода калийбромидных таблеток в количественном ИК-спектральном анализе часто экстинкцию выбранной для анализа полосы поглощения относят к экстинкции добавленного стандартного вещества. Таким образом должны уменьшиться случайные ошибки, возникающие при приготовлении таблеток. Однако такой способ введения внутреннего стандарта только тогда эффективен, когда коррелированы экстинкции обеих полос поглощения — аналитической и стандартной (взятой для сравнения), т. е. когда случайные изменения, влияющие на одну из полос поглощения, ведут к подобным же изменениям поглощения в другой полосе.

Для определения остаточного содержания нитрильных групп в продуктах из полиакрилонитрила в качестве стандартного вещества использовали $K_4 [Fe(CN)_6]$. Для проверки корреляции на десяти пробах одного состава были получены следующие значения экстинкции для аналитической (E_A) и стандартной (E_S) полос поглощения [4].

E_A	E_S	Преобразованные значения	
		X	Y
0,428	0,626	-22	-22
0,460	0,650	+10	+2
0,427	0,633	-23	-15
0,445	0,670	-5	+22
0,460	0,660	+10	+12
0,470	0,651	+20	+3
0,456	0,668	+6	+20
0,460	0,650	+10	+2
0,448	0,643	-2	-5
0,450	0,630	0	-18

Для простоты расчета использовали преобразования $X = 1000 \times E_A - 450$ и $Y = 1000 E_S - 648$. Из преобразованных значений вычислили:

$$\begin{aligned} \sum X_i &= +4 & (\sum X_i)^2 &= 16 & \sum X_i^2 &= 1778 \\ \sum Y_i &= +1 & (\sum Y_i)^2 &= 1 & \sum Y_i^2 &= 2103 \\ & & \sum X_i Y_i &= 1069 & m &= 10 \end{aligned}$$

Пользуясь уравнением (9.38), получили

$$r = \frac{10 \cdot 1069 - 4 \cdot 1}{\sqrt{(10 \cdot 1778 - 16)(10 \cdot 2103 - 1)}} = 0,553$$

По табл. 9.1 нашли $|r| < r(P = 0,95; f = 8) = 0,63$. Между экстинкциями аналитической и стандартной полос поглощения не оказалось никакой значимой корреляции. Из имеющихся результатов нельзя сделать вывод об уменьшающем ошибку влиянии стандартного образца.

В целях сравнения часто возникает вопрос, существует ли различная степень связи между переменными x и y в двух разных сериях измерений. В этом случае необходимо проверить есть ли статистически значимое различие для коэффициентов корреляции r_1 и r_2 , полученных из m_1 и m_2 измерений. Для этого составляют выражение:

$$t_r = 1,1513 \sqrt{\frac{(m_1-3)(m_2-3)}{m_1+m_2-6}} \lg \frac{(1+r_1)(1-r_2)}{(1-r_1)(1+r_2)} \quad (9.39)$$

Затем вычисленное значение сравнивают с табличным значением $t(P, f)$ при $f = m_1 + m_2 - 4$. Различие оказывается значимым в том случае, если $t_r > t(P, f)$.

[9.14]. Для геохимических исследований необходимо было выяснить, существует ли связь между содержанием натрия и лития в воде. В первой серии опытов, состоящей из 10 проб воды, коэффициент корреляции составил $r_1 = 0,838$. При повторении опыта в другое время года значения, полученные на 15 пробах, привели к коэффициенту корреляции $r_2 = 0,738$. В соответствии с соотношением (9.39)

$$t_r = 1,1513 \sqrt{\frac{(10-3)(15-3)}{10+15-6}} \lg \frac{1,838 \cdot 0,262}{0,162 \cdot 1,738} = 0,564$$

Из табл. 12.3 получают $t(P = 0,95; f = 21) = 2,08$. Так как $t_r < t(P, f)$ для двух коэффициентов корреляции нет значимой разницы.

Вычислением корреляции можно также отыскать и оценить связи между весьма отдаленными величинами. Однако при этом существует опасность, что устанавливают корреляции, смысл которых весьма спорен. Проверка корреляции надежна только на таком материале, который по происхождению можно считать однородным. Отсутствие единообразия (например, два различно работающих аналитика) может замаскировать корреляцию или привести к ложной корреляции. Проверка корреляции на временных рядах из-за временной изменчивости одной из величин может при известных обстоятельствах

указать на связь, которой в действительности нет. Поэтому обоснованность корреляционных связей следует тщательно продумывать в каждом случае.

Л И Т Е Р А Т У Р А

1. Baumann M., Appl. Spectroscopy, **13**, 156 (1956).
2. Doerffel K., Z. anal. Chem., **157**, 241 (1957).
3. Doerffel K., Z. anal. Chem., **185**, 1 (1962).
4. Doerffel K., Wiss. Z. Techn. Hochschule Chemie, Leuna-Merseburg Jg., 1965.
5. Doerffel K., Lichtner J., Spectrochim. Acta, London, 1966.
6. Grotkopf P., Diplomarbeit TH Chemie Leuna-Merseburg, 1964.
7. Geyer R., Doerffel K., Z. anal. Chem., **158**, 419 (1957).
8. Geyer R., Doerffel K., H ö b o l d W., Z. anal. Chem., Jg., 1966.
9. Grossmann W., Grundzüge der Ausgleichsrechnung, Berlin-Göttingen-Heidelberg, Springer, 1954.
10. Hold G., Strassheim A., Appl. Spectroscopy, **14**, 64 (1960).
11. Kaiser H., Specker H., Z. anal. Chem., **149**, 46 (1956).
12. Kirchhof S., Jagusch K. H., Z. anal. Chem., **199**, 197 (1964).
13. Körwin H., Graphisches Rechnen, Leipzig, VEB Fachbuchverlag, 1949.
14. Mandel J., Linnig F. J., Anal. Chem., **29**, 743 (1957).
15. Maurice M., Z. anal. Chem., **158**, 271 (1958).
16. Wörner H., Z. wirtsch. Fertigung, **52**, 99 (1957).

10. Приемы вычислений

При оценке результатов анализа значительную роль играют арифметические вычисления. Возникает вопрос о затратах на вычисления, появляется также проблема разумного выбора техники вычислений и выбора счетных вспомогательных средств. Для этого ниже приведено несколько кратких указаний.

Затраты на вычисления всегда определяются случайной ошибкой имеющихся измерений. Результат вычислений никогда не может быть точнее, чем используемые измерения. Для ошибки вычисленных результатов справедливы правила распространения ошибок, данные в начале гл. 4. При линейных расчетах суммируется абсолютная ошибка отдельных значений, а при нелинейных — относительная ошибка. Расчет всегда проводят с сохранением одной лишней цифры по сравнению со случайной ошибкой конечного результата. Эту лишнюю цифру опускают лишь в конечном результате. Для необходимого при этом округления чисел справедливы правила стандарта TGL 0-1333.

1. При округлении последняя сохраняемая цифра остается без изменения, если за ней следуют цифры 0, 1, 2, 3 и 4 (например, 8,2738 → 8,27).

2. При округлении последнюю значащую цифру увеличивают на единицу, если за ней стоят цифры 9, 8, 7 и 6 (например, 8,2763 → 8,28).

3. Если перед округлением за значащей цифрой стоит цифра 5 хотя бы с одной последующей цифрой, отличной от нуля, то последнюю значащую цифру увеличивают на единицу (например, 4,35001 → 4,4). Однако если эта пятерка получилась уже после округления, то для дальнейшего округления необходимо принимать во внимание первоначальное число (например, 6,315 получено из 6,3149; поэтому при дальнейшем округлении 6,315 → 6,31).

4. Если перед округлением цифра, следующая за значащей цифрой, была точно 5, то последняя значащая цифра округляется до ближайшей четной (например, $1/16 = 0,0625000 \rightarrow 0,062$; $3^3/4 = 3,750 \rightarrow 3,8$).

При таком округлении получается малая ошибка. Ее можно сохранить в таких размерах, выбирая подходящим образом схему подсчета. Для вычисления сумм квадратов $\sum (x_i - \bar{x})^2$, например по равенству (2.6), обе формулы арифметически равнозначны.

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad (2.6a)$$

$$= \sum x_i^2 - n\bar{x}^2 \quad (2.6b)$$

Однако для практического расчета применяют первую формулу, так как в уравнении (2.6b) при подсчете среднего значения при известных условиях может появиться ошибка округления, которая затем увеличится на фактор n .

Для проведения расчета в распоряжении имеются разнообразные вспомогательные средства. Для большинства приведенных здесь вычислительных методов необходимы малые вычислительные машины*, обладающие достаточно большим запасом знаков. Выгодна возможность переноса в качестве множителя числа из счетчика результатов в умножающий механизм. Расчет всегда следует вести так, чтобы производить возможно меньшее число регулировок. При этом числовые таблицы (квадратов, обратных чисел) оказывают большую помощь. Часто целесообразно начинать расчет как бы с конца.

Вычисления по формуле (2.6a) производят, например, следующим образом:

1. Суммируя отдельные x_i , получают в качестве частного результата $\sum x_i$.

2. Частный результат $\sum x_i$ переносят в умножающий механизм.

3. Возводят в квадрат $\sum x_i$.

4. Снова переносят частный результат $(\sum x_i)^2$ в умножающий механизм.

* См., например, книгу Л. С. Хренов, «Малые вычислительные машины», ГИТТЛ, М., 1957. — Прим. ред.

5. Умножают на $1/n$ (по таблице обратных чисел).

6. Переносят в механизм суммирования с обратным знаком.

7. Суммируют отдельные значения x_i^2 (при помощи таблицы квадратов).

8. Получают окончательный результат.

При помощи такой схемы расчета избегают ошибки, которая бывает обусловлена неправильной записью или подстановкой промежуточных результатов.

Возможности современных вычислительных устройств легко могут привести к тому, что будут получены результаты с бессмысленно большим числом знаков. Поэтому особенно важно остановиться на нужном числе знаков.

Часто целесообразно перенести повторяющиеся расчеты или расчеты с очень большим числом знаков после запятой (например, многомерную линейную регрессию) на электронные вычислительные машины. При наличии заранее составленной программы вычисления, занесенной на перфоленты или перфокарты, можно значительно сэкономить время. При этом намного уменьшается риск появления обычных ошибок вычисления, особенно при расчетах со многими знаками после запятой.

Обычное вспомогательное средство химика — таблица логарифмов. В чаще всего используемой пятизначной таблице следует учитывать ошибку счета $s_p = 5 \cdot 10^{-3} \%$ [1]. Это равнозначно тому, что в каждом результате можно дать пять значащих цифр. Вычисления с логарифмическими таблицами можно облегчить, если производить сложение и вычитание, используя простые суммирующие устройства.

Вслед за таблицей логарифмов большой популярностью пользуется логарифмическая линейка, а за последнее время круглая счетная линейка. Чаще всего используют логарифмическую линейку с длиной шкалы $l = 250$ или 500 мм, а карманного формата с $l = 125$ мм. У круглой счетной линейки длина шкалы доходит до 900 мм. Ошибка логарифмической линейки составляет по Грюксу [1]

$$s_p = \frac{23}{l} \sqrt{n} \%, \quad (10.1)$$

где l — длина шкалы (мм), n — число отсчетов.

[10.1]. Простое выполнение умножения и деления в виде xy/z (простейший тип формулы для расчета результатов анализа из значений измерений) требует четырех отсчетов. При общеупотребительной длине шкалы $l = 250$ мм в соответствии с уравнением (10.1)

$$s_p = \frac{23}{250} \sqrt{4} \sim 0,2\%$$

Следовательно, можно ожидать точности расчета с тремя знаками. Этого достаточно для обычных серийных анализов.

Обычно используемая счетная линейка двойной длины ($l = 500$ мм) согласно соотношению (10.1) не дает никакого существенного выигрыша. Заметное уменьшение ошибки расчета по уравнению (10.1) (т. е. возможность получения еще одного знака после запятой) достигают лишь при десятикратной длине шкалы. Подобную шкалу можно осуществить в счетном цилиндре. Здесь шкала делится на ряд отдельных областей, наносимых на цилиндр. Такое устройство позволяет получить шкалу с эффективной длиной 10 м и больше. Практика использования счетного цилиндра совершенно аналогична счетной линейке. Однако из-за большей длины шкалы цилиндр дает намного более высокую точность. При эффективной длине шкалы 10 м по соотношению (10.1) для задачи одновременного умножения и деления xy/z ошибка будет

$$s_p = \frac{23}{10^4} \sqrt{4} \sim 5 \cdot 10^{-3}\%$$

Это соответствует ошибке пятизначных логарифмических таблиц. Счетный цилиндр объединяет преимущества логарифмической линейки (быстрый подсчет, составление пропорций и т. д.) с преимуществом логарифмических таблиц — малой ошибкой. На этом основании его следовало бы чаще применять в производственных лабораториях.

При оценке методов анализа, нуждающихся в градуировке, предпочитают использовать графические методы (градуировочные кривые). Ошибка счета тогда наименьшая, когда получают прямую под 45° . Этого условия (внутри определенных пределов) можно достигнуть выбором подходящего масштаба осей координат. Сложные градуировочные кривые в любом случае маложелательны.

Вспомогательные деления	COS, об. % ^a	при 20°			COS, 10 ⁻⁴ %	Вспомогательные деления
		COS, мг/м ³	сера, мг/м ³	сера, г/100 м ³		
	0	0	0	0	0	
		100				
		200	100	10		
	0.01	300			100	
		400	200	20		
		500			200	
	0.02	600	300	30		
		700			300	
	0.03	800	400	40		
		900			400	
	0.04	1000	500	50		
		1100			500	
	0.05	1200	600	60		
		1300			600	
		1400	700	70		
	0.06	1500	800	80		
		1600			800	
	0.07	1700	900	90		
		1800			900	
		1900	1000	100		
	0.08	2000	1100	110		
		2100			1100	
		2200			1200	
	0.09	2300	1200	120		
		2400			1300	
	0.10	2500	1300	130		
		2600			1400	
		2700	1400	140		
	0.11	2800	1500	150		
		2900			1500	
	0.12	3000	1600	160		
					1600	

^a Область измерений { 0,001 - 0,012 об. % COS
0,0001 - 0,0012 об. % COS
0,00001 - 0,00012 об. % COS

Значения { · 10⁻¹
· 10⁻²
· 10⁻³

Графический метод расчета применяют чаще всего, если хотят использовать определенный расчет. В качестве примера здесь приведен рис. 10.1, который является таблицей для пересчета результатов анализа в различную систему шкал. Другие примеры, особенно для комбинированных счетных таблиц, указаны в гл. 7. При составлении таких номограмм целесообразно прежде всего получить рисунок большого формата, а затем уменьшить его фотографически. Для составления счетных таблиц следует обратиться к соответствующим работам по номографии.

ЛИТЕРАТУРА

1. Grüss G., Differential- und Integralrechnung, 2. Auflage, Leipzig, Akademische Verl., 1953.
2. Höbold W., Diss., Merseburg, 1965.
3. Körwin H., Graphisches Rechnen, Leipzig, VEB Fachbuchverlag, 1949.
4. Neumann J. V., Die Rechenmaschine und das Gehirn, München und Wien, Verlag R. Oldenbourg, 1965.
5. Pohl F. A., Z. anal. Chem., 209, 19 (1965).
6. Willers F. A., Matematische Maschinen und Instrumente, Berlin, Akademie-Verlag, 1951.

Рис. 10.1. Диаграмма для пересчета концентраций [2]. Шкалы по отдельным столбцам составлены так, что соответствующие друг другу значения лежат на соединяющихся горизонтальных линиях. «Вспомогательные деления» облегчают работу с диаграммой.

11. Послесловие

Математическая статистика дает аналитику набор разнообразнейших методов для оценки результатов и методов анализа. Цель этой книги ясна тем аналитикам, которым приходится сталкиваться с проблемой выбора методов анализа. В конце изложения следовало бы еще раз рассмотреть вопрос об общих возможностях и пределах описанных методов.

В очень многих случаях аналитик прибегает к методам математической статистики, если речь идет об ошибке метода или результатов анализа. Вообще у аналитиков все в возрастающей степени наблюдается тенденция не только оперировать значениями анализа, но и тщательно их интерпретировать. Эта интерпретация также важна, как и упорядоченный отбор проб, ибо в ходе дальнейших работ результаты анализов все чаще используют не химики-аналитики. Методы математической статистики являются общепринятыми, а их справедливость общепризнана. Введение этих методов облегчает также взаимопонимание между аналитиками и теми, для кого предназначены анализы, и помогает избежать ошибочных заключений и предотвратить недоразумения.

Использование статистических методов для оценки ошибки и интерпретации результатов — это использование лишь простейших приемов. Математическая статистика приносит наибольшую пользу тогда, когда эксперимент ставится наилучшим образом. Эта проблема — проблема наилучшего планирования эксперимента так же решается при помощи математической статистики. Планирование эксперимента применяют, как для решений простейших вопросов, таких, например, какое наилучшее число параллельных определений для оценки среднего значения, так и для сложных задач, например для постановки совместного опыта. Поэтому математическую ста-

тистику не следует понимать, как некоторое вспомогательное средство обработки проведенных измерений, ее следует привлекать уже при планировании опыта. Она покажет, при каких условиях проведения эксперимента следует ожидать оптимального результата.

В некоторых случаях аналитическая проблема вообще разрешима лишь при помощи математической статистики. Примером этого является вторичный фотометрический анализ смеси нескольких компонентов. Лишь при помощи многомерной регрессии удастся проанализировать смесь весьма сложного состава с приемлемо малой ошибкой. Статистические методы в подобных случаях не просто средство планирования эксперимента или его оценки — они являются необходимым инструментом для решения определенной аналитической задачи.

Естественно, статистика также имеет границы. Она не может дать для специальных единичных случаев соответствующие специальные сведения. Статистика дает сведения лишь с заданной или оговоренной вероятностью и риском соответствующей ошибки. Неправильно проведенный анализ не может дать при помощи статистики надежный результат. Но там, где возможен осмысленный анализ, нужно применять статистику. Методы математической статистики не могут заменить аналитику критический склад ума, но, вероятно, они могут оказать эффективную помощь.

Специальные теоретико-математические познания едва ли необходимы для применения статистики. Следует лишь понимать ход рассуждений, знать, как надо провести эксперимент, какие данные может дать статистика и где следует предполагать границы ее применимости. При такой точке зрения настоящая книга могла бы побудить аналитика критически применить положения статистики к разрешаемой проблеме для его собственного и для общего блага.

ЛИТЕРАТУРА

1. Analyse der Metalle, B. III, Probenahme, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1956.

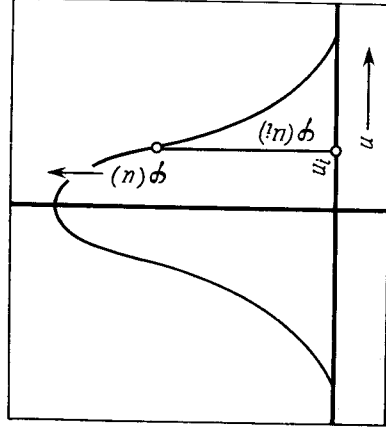
2. B a u l e B., Die Mathematik des Naturforschers und Ingenieurs, B. II, Ausgleichs- und Näherungsrechnung, 2. Auflage, Verlag S. Hirzel, Leipzig, 1948.
3. B e n n e t C. A., F r a n k l i n N. L., Statistical Analysis in Chemistry and Chemical Industry, Verlag Chapman & Hall, N.Y., 1954.
4. E c k s c h l a g e r K., Fehler bei chemischen Analysen, Akademische Verl., Leipzig, 1964.
5. Ф и ш е р Р. А., Статистические методы для исследователей, Госстатиздат, 1958.
6. G r a f U., H e n n i n g H.-J., Formeln und Tabellen der mathematischen Statistik, Verlag Springer, Berlin-Göttingen-Heidelberg, 1958.
7. G r ü s s G., Differential- und Integralrechnung, 2. Auflage, Akademische Verl., Leipzig, 1953.
8. K l e m m L., R i e h l H. J., S i e g e l H., T r o l l W., Statistische Kontrollmethoden in der Textilindustrie, 2. Auflage, VEB Fachbuchverl., Leipzig, 1964.
9. L i n d e r A., Statistische Methoden für Naturwissenschaftler, Mediziner und Ingenieure, Verlag Birkhäuser, Basel, 1951.
10. L i n d e r A., Planen und Auswerten von Versuchen, Verlag Birkhäuser, Basel, 1953.
11. L i n d e r A., Statistische Methoden für Naturwissenschaftler, Mediziner und Ingenieure, 3. Auflage, Verlag Birkhäuser, Basel, 1960.
12. L o r e n z P., Anschauungsunterricht in mathematischer Statistik, S. Hirzel, Leipzig, 1955.
13. Н а л и м о в В. В., Применение математической статистики при анализе вещества, Физматгиз, М., 1960.
14. Д у н и н - Б а р к о в с к и й И. В., С м и р н о в Н. В., Теория вероятностей и математическая статистика в технике, Гостехиздат, М., 1955*.
15. W e b e r E., Grundriß der biologischen Statistik für Naturwissenschaftler, Landwirte und Mediziner, 4. Auflage, VEB Gustav-Fischer-Verlag, Jena, 1961.
16. Y o u d e n W. J., Statistical Methods for Chemists, 2. Auflage, Wiley & Sons, N.Y., 1951.
17. Z i m m e r m a n n K. F., Tabellen, Formeln und Fachausdrücke zur Variationsstatistik, VEB Deutscher Verlag der Wissenschaften.

* См. также книгу тех же авторов «Курс теории вероятностей и математической статистики для технических приложений, «Наука», 1965. — Прим. ред.

12. Приложение (таблицы)

Значения ординат нормального распределения

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,39894	0,39892	0,39886	0,39876	0,39862	0,39844	0,39822	0,39767	0,39797	0,39733
0,1	0,39695	0,39654	0,39608	0,39559	0,39505	0,39448	0,39387	0,39322	0,39253	0,39181
0,2	0,39104	0,39024	0,38940	0,38853	0,38762	0,38667	0,38568	0,38466	0,38361	0,38251
0,3	0,38139	0,38023	0,37903	0,37780	0,37654	0,37524	0,37391	0,37255	0,37115	0,36973
0,4	0,36827	0,36678	0,36526	0,36371	0,36213	0,36053	0,35889	0,35723	0,35553	0,35381
0,5	0,35207	0,35029	0,34849	0,34667	0,34482	0,34294	0,34105	0,33912	0,33718	0,33521
0,6	0,33322	0,33121	0,32918	0,32713	0,32506	0,32297	0,32086	0,31874	0,31659	0,31443
0,7	0,31225	0,31006	0,30785	0,30563	0,30339	0,30114	0,29887	0,29659	0,29431	0,29200
0,8	0,28969	0,28737	0,28504	0,28269	0,28034	0,27798	0,27562	0,27324	0,27086	0,26848
0,9	0,26609	0,26369	0,26129	0,25888	0,25647	0,25406	0,25164	0,24923	0,24681	0,24439
1,0	0,24197	0,23955	0,23713	0,23471	0,23230	0,22988	0,22747	0,22506	0,22265	0,22025
1,1	0,21785	0,21546	0,21307	0,21069	0,20831	0,20594	0,20357	0,20121	0,19886	0,19652
1,2	0,19419	0,19186	0,18954	0,18724	0,18494	0,18265	0,18037	0,17810	0,17585	0,17360
1,3	0,17137	0,16915	0,16694	0,16474	0,16256	0,16038	0,15822	0,15608	0,15395	0,15183
1,4	0,14937	0,14764	0,14556	0,14350	0,14146	0,13943	0,13742	0,13542	0,13344	0,13147
1,5	0,12952	0,12758	0,12566	0,12376	0,12188	0,12051	0,11816	0,11632	0,11450	0,11270
1,6	0,11092	0,10915	0,10741	0,10567	0,10396	0,10226	0,10059	0,09893	0,09728	0,09566
1,7	0,09405	0,09246	0,09089	0,08933	0,08780	0,08628	0,08478	0,08329	0,08183	0,08038
1,8	0,07895	0,07754	0,07614	0,07477	0,07341	0,07206	0,07074	0,06943	0,06814	0,06687
1,9	0,06582	0,06438	0,06316	0,06195	0,06077	0,05959	0,05844	0,05730	0,05618	0,05508
2,0	0,05399	0,05292	0,05186	0,05082	0,04980	0,04879	0,04780	0,04632	0,04586	0,04491
2,1	0,04398	0,04307	0,04217	0,04128	0,04041	0,03955	0,03871	0,03788	0,03706	0,03626
2,2	0,03547	0,03470	0,03394	0,03319	0,03246	0,03174	0,03103	0,03034	0,02965	0,02898
2,3	0,02833	0,02768	0,02705	0,02643	0,02582	0,02522	0,02463	0,02406	0,02349	0,02294
2,4	0,02239	0,02186	0,02134	0,02083	0,02033	0,01984	0,01936	0,01889	0,01842	0,01797
2,5	0,01753	0,01709	0,01667	0,01625	0,01585	0,01545	0,01506	0,01468	0,01431	0,01394
2,6	0,01358	0,01323	0,01289	0,01256	0,01223	0,01191	0,01160	0,01130	0,01100	0,01071
2,7	0,01042	0,01014	0,00987	0,00961	0,00935	0,00909	0,00885	0,00861	0,00837	0,00814
2,8	0,00792	0,00770	0,00748	0,00727	0,00707	0,00687	0,00668	0,00649	0,00631	0,00613
2,9	0,00595	0,00578	0,00562	0,00545	0,00530	0,00514	0,00499	0,00485	0,00471	0,00457
3,0	0,00443	0,00327	0,00238	0,00172	0,00123	0,00087	0,00061	0,00042	0,00029	0,00020

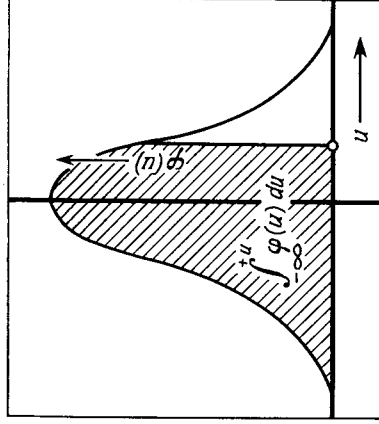


В табл. 12.1 даны значения кривой Гаусса.

Интеграл Гаусса

Площадь F под нормированной кривой Гаусса в пределах $-\infty \dots +u$. Пересчет на площадь F' в пределах $-u \dots +u$ следует по $F' = 2(F - 0,5)$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500000	0,503989	0,507978	0,511966	0,515953	0,519938	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563560	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621720	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,702944	0,707402	0,711840	0,716260	0,720661	0,725043	0,729405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802338	0,805106	0,807850	0,810570	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843732	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,879000	0,881000	0,882977
1,2	0,884930	0,886861	0,888768	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903200	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914656	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923642	0,925066	0,926471	0,927855	0,929219	0,930563	0,931889
1,5	0,933193	0,934478	0,935744	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950528	0,951543	0,952540	0,953521	0,954486
1,7	0,955434	0,956367	0,957284	0,958185	0,959070	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965620	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971283	0,972033	0,972761	0,973471	0,973810	0,974412	0,975002	0,975581	0,976148	0,976704
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987454	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990862	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995539	0,995731	0,995957	0,996185	0,996414	0,996643	0,996872	0,997100	0,997328
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3,0	0,998650	0,999032	0,999313	0,999517	0,999663	0,999767	0,999841	0,999892	0,999928	0,999952



В табл. 12.2 дана площадь под кривой Гаусса.

Таблица 12.3

Пределы интегрирования для t -распределения в зависимости от вероятности P (двусторонняя постановка задачи) и \bar{P} (односторонняя постановка задачи) и степени свободы f

f	$P = 0,50$	0,75	0,90	0,95	0,98	0,99
1	1,00	2,41	6,31	12,7	31,82	63,7
2	0,816	1,60	2,92	4,30	6,97	9,92
3	0,765	1,42	2,35	3,18	4,54	5,84
4	0,741	1,34	2,13	2,78	3,75	4,60
5	0,727	1,30	2,01	2,57	3,37	4,03
6	0,718	1,27	1,94	2,45	3,14	3,71
7	0,711	1,25	1,89	2,36	3,00	3,50
8	0,706	1,24	1,86	2,31	2,90	3,36
9	0,703	1,23	1,83	2,26	2,82	3,25
10	0,700	1,22	1,81	2,23	2,76	3,17
11	0,697	1,21	1,80	2,20	2,72	3,11
12	0,695	1,21	1,78	2,18	2,68	3,05
13	0,694	1,20	1,77	2,16	2,65	3,01
14	0,692	1,20	1,76	2,14	2,62	2,98
15	0,691	1,20	1,75	2,13	2,60	2,95
16	0,690	1,19	1,75	2,12	2,58	2,92
17	0,689	1,19	1,74	2,11	2,57	2,90
18	0,688	1,19	1,73	2,10	2,55	2,88
19	0,688	1,19	1,73	2,09	2,54	2,86
20	0,687	1,18	1,73	2,09	2,53	2,85
21	0,686	1,18	1,72	2,08	2,52	2,83
22	0,686	1,18	1,72	2,07	2,51	2,82
23	0,685	1,18	1,71	2,07	2,50	2,81
24	0,685	1,18	1,71	2,06	2,49	2,80
25	0,684	1,18	1,71	2,06	2,49	2,79
26	0,684	1,18	1,71	2,06	2,48	2,78
27	0,684	1,18	1,71	2,05	2,47	2,77
28	0,683	1,17	1,70	2,05	2,47	2,76
29	0,683	1,17	1,70	2,05	2,46	2,76
30	0,683	1,17	1,70	2,04	2,46	2,75
40	0,681	1,17	1,68	2,02	2,42	2,70
60	0,679	1,16	1,67	2,00	2,39	2,66
120	0,677	1,16	1,66	1,98	2,36	2,62
∞	0,674	1,15	1,64	1,96	2,33	2,58
f	$\bar{P} = 0,75$	0,875	0,95	0,975	0,99	0,995

Таблица 12.4

Пределы интегрирования для χ^2 -распределения в зависимости от вероятности \bar{P} и степени свободы f

f	$\bar{P} = 0,01$	0,05	0,10	0,50	0,90	0,95	0,99
1	0,03157	0,02393	0,0158	0,455	2,71	3,84	6,64
2	0,0201	0,103	0,211	1,39	4,61	5,99	9,21
3	0,115	0,352	0,584	2,37	6,25	7,81	11,3
4	0,297	0,711	1,06	3,36	7,78	9,49	13,3
5	0,554	1,15	1,61	4,35	9,24	11,1	15,1
6	0,872	1,64	2,20	5,35	10,6	12,6	16,8
7	1,24	2,17	2,83	6,35	12,0	14,1	18,5
8	1,65	2,73	3,49	7,34	13,4	15,5	20,1
9	2,09	3,33	4,17	8,34	14,7	16,9	21,7
10	2,56	3,94	4,87	9,34	16,0	18,3	23,2
11	3,05	4,57	5,58	10,3	17,3	19,7	24,7
12	3,57	5,23	6,30	11,3	18,5	21,0	26,2
13	4,11	5,89	7,04	12,3	19,8	22,4	27,7
14	4,66	6,57	7,79	13,3	21,1	23,7	29,1
15	5,23	7,26	8,55	14,3	22,3	25,0	30,6
16	5,81	7,96	9,31	15,3	23,5	26,3	32,0
17	6,41	8,67	10,1	16,3	24,8	27,6	33,4
18	7,01	9,39	10,9	17,3	26,0	28,9	34,8
19	7,63	10,1	11,7	18,3	27,2	30,1	36,2
20	8,26	10,9	12,4	19,3	28,4	31,4	37,6
21	8,90	11,6	13,2	20,3	29,6	32,7	38,9
22	9,54	12,3	14,0	21,3	30,8	33,9	40,3
23	10,2	13,1	14,8	22,3	32,0	35,2	41,6
24	10,9	13,8	15,7	23,3	33,2	36,4	43,0
25	11,5	14,6	16,5	24,3	34,4	37,7	44,3

Таблица 12.5

Пределы интегрирования для F -распределения в зависимости от степени свободы f_1 и f_2 . О возможности интерполяции см. разд. 3.3.2

а) для $\bar{P} = 0,95$

f_2	$f_1 = 1$	2	3	4	5	6	8
1	161	200	216	225	230	234	239
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94
f_2	$f_1 = 1$	2	3	4	5	6	8

Продолжение табл. 12.5

10	12	16	20	24	50	∞	f_2
242	244	246	248	249	252	254	1
19,39	19,41	19,43	19,44	19,45	19,47	19,50	2
8,78	8,74	8,69	8,66	8,64	8,58	8,53	3
5,96	5,91	5,84	5,80	5,77	5,70	5,63	4
4,74	4,68	4,60	4,56	4,53	4,44	4,36	5
4,06	4,00	3,92	3,87	3,84	3,75	3,67	6
3,63	3,57	3,49	3,44	3,41	3,32	3,23	7
3,34	3,28	3,20	3,15	3,12	3,03	2,93	8
3,13	3,07	2,98	2,93	2,90	2,80	2,71	9
2,97	2,91	2,82	2,77	2,74	2,64	2,54	10
2,86	2,79	2,70	2,65	2,61	2,50	2,40	11
2,76	2,69	2,60	2,54	2,50	2,40	2,30	12
2,67	2,60	2,51	2,46	2,42	2,32	2,21	13
2,60	2,53	2,44	2,39	2,35	2,24	2,13	14
2,55	2,48	2,39	2,33	2,29	2,18	2,07	15
2,49	2,42	2,33	2,28	2,24	2,13	2,01	16
2,45	2,38	2,29	2,23	2,19	2,08	1,96	17
2,41	2,34	2,25	2,19	2,15	2,04	1,92	18
2,38	2,31	2,21	2,15	2,11	2,00	1,88	19
2,35	2,28	2,18	2,12	2,08	1,96	1,84	20
2,32	2,25	2,15	2,09	2,05	1,93	1,81	21
2,30	2,23	2,13	2,07	2,03	1,91	1,78	22
2,28	2,20	2,10	2,05	2,00	1,88	1,76	23
2,26	2,18	2,09	2,02	1,98	1,86	1,73	24
2,24	2,16	2,06	2,00	1,96	1,84	1,71	25
2,22	2,15	2,05	1,99	1,95	1,82	1,69	26
2,20	2,13	2,03	1,97	1,93	1,80	1,67	27
2,19	2,12	2,02	1,96	1,91	1,78	1,65	28
2,18	2,10	2,00	1,94	1,90	1,77	1,64	29
2,16	2,09	1,99	1,93	1,89	1,76	1,62	30
2,07	2,00	1,90	1,84	1,79	1,66	1,51	40
1,99	1,92	1,81	1,75	1,70	1,60	1,39	60
1,90	1,83	1,72	1,65	1,61	1,45	1,25	120
1,83	1,75	1,63	1,57	1,52	1,35	1,00	∞
10	12	16	20	24	50	∞	f_2

Продолжение табл. 12.5

б) для $\bar{P} = 0,99$

f_2	$f_1 = 1$	2	3	4	5	6	8
1	4052	4999	5403	5625	5764	5859	5981
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82
120	6,85	4,79	3,95	3,48	3,17	2,96	2,66
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51
f_2	$f_1 = 1$	2	3	4	5	6	8

Продолжение табл. 12.5

10	12	16	20	24	50	∞	f_2
6056	6106	6169	6208	2634	6302	6366	1
99,40	99,42	99,44	99,45	99,46	99,48	99,50	2
27,23	27,05	26,83	26,65	26,60	26,35	26,12	3
14,54	14,37	14,15	14,02	13,93	13,69	13,46	4
10,05	9,89	9,68	9,55	9,47	9,24	9,02	5
7,87	7,72	7,52	7,39	7,31	7,09	6,88	6
6,62	6,47	6,27	6,15	6,07	5,85	5,65	7
5,82	5,67	5,48	5,36	5,28	5,06	4,86	8
5,26	5,11	4,92	4,80	4,73	4,51	4,31	9
4,85	4,71	4,52	4,41	4,33	4,12	3,91	10
4,54	4,40	4,21	4,10	4,02	3,80	3,60	11
4,30	4,16	3,98	3,86	3,78	3,56	3,36	12
4,10	3,96	3,78	3,67	3,59	3,37	3,16	13
3,94	3,80	3,62	3,51	3,43	3,21	3,00	14
3,80	3,67	3,48	3,36	3,29	3,07	2,87	15
3,69	3,55	3,37	3,25	3,18	2,96	2,75	16
3,59	3,45	3,27	3,16	3,08	2,86	2,65	17
3,51	3,37	3,19	3,07	3,00	2,78	2,57	18
3,43	3,30	3,12	3,00	2,92	2,70	2,49	19
3,37	3,23	3,05	2,94	2,86	2,63	2,42	20
3,31	3,17	2,99	2,88	2,80	2,58	2,36	21
3,26	3,12	2,94	2,83	2,75	2,53	2,31	22
3,21	3,07	2,89	2,78	2,70	2,48	2,26	23
3,17	3,03	2,85	2,74	2,66	2,44	2,21	24
3,13	2,99	2,81	2,70	2,62	2,40	2,17	25
3,09	2,96	2,77	2,66	2,58	2,36	2,13	26
3,06	2,93	2,74	2,63	2,55	2,33	2,10	27
3,03	2,90	2,71	2,60	2,52	2,30	2,06	28
3,00	2,87	2,68	2,57	2,49	2,27	2,03	29
2,98	2,84	2,66	2,55	2,47	2,24	2,01	30
2,80	2,66	2,49	2,37	2,29	2,05	1,80	40
2,63	2,50	2,32	2,20	2,12	1,87	1,60	60
2,47	2,34	2,15	2,03	1,95	1,68	1,38	120
2,23	2,18	1,99	1,87	1,79	1,52	1,00	∞
10	12	16	20	24	50	∞	f_2

Критические значения для критерия Дункана в зависимости от степени свободы F_2 и ранга p измеренных значений

F_2	$p=2$	3	4	5	6	7	8	9	10	12	14	16	18	20
1	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0
2	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09
3	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50
4	3,93	4,01	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02
5	3,64	3,74	3,79	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83
6	3,46	3,58	3,64	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68
7	3,35	3,47	3,54	3,58	3,60	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61
8	3,26	3,39	3,47	3,52	3,55	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56
9	3,20	3,34	3,41	3,47	3,50	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52	3,52
10	3,15	3,30	3,37	3,43	3,46	3,47	3,47	3,47	3,47	3,47	3,47	3,47	3,47	3,47
11	3,11	3,27	3,35	3,39	3,43	3,44	3,45	3,46	3,46	3,46	3,46	3,46	3,46	3,46
12	3,08	3,23	3,33	3,36	3,40	3,42	3,44	3,44	3,45	3,45	3,46	3,46	3,46	3,47
13	3,06	3,21	3,30	3,35	3,38	3,41	3,42	3,44	3,45	3,45	3,46	3,46	3,46	3,47
14	3,03	3,18	3,27	3,33	3,37	3,39	3,41	3,42	3,44	3,44	3,45	3,46	3,46	3,47
15	3,01	3,16	3,25	3,31	3,36	3,38	3,40	3,42	3,43	3,44	3,45	3,46	3,46	3,47
16	3,00	3,15	3,23	3,30	3,34	3,37	3,39	3,41	3,43	3,44	3,45	3,46	3,46	3,47
17	2,98	3,13	3,22	3,28	3,33	3,36	3,38	3,40	3,42	3,44	3,45	3,46	3,46	3,47
18	2,97	3,12	3,21	3,27	3,32	3,35	3,37	3,39	3,41	3,43	3,45	3,46	3,46	3,47
19	2,96	3,11	3,19	3,26	3,31	3,35	3,37	3,39	3,41	3,43	3,45	3,46	3,46	3,47
20	2,95	3,10	3,18	3,25	3,30	3,34	3,36	3,38	3,40	3,43	3,44	3,45	3,46	3,47
22	2,93	3,08	3,17	3,24	3,29	3,32	3,35	3,37	3,39	3,42	3,44	3,45	3,46	3,47
24	2,92	3,07	3,15	3,22	3,28	3,31	3,34	3,37	3,38	3,41	3,44	3,45	3,46	3,47
26	2,91	3,06	3,14	3,21	3,27	3,30	3,34	3,37	3,38	3,41	3,44	3,45	3,46	3,47
28	2,90	3,04	3,13	3,20	3,26	3,30	3,33	3,36	3,38	3,41	3,43	3,45	3,46	3,47
30	2,89	3,04	3,12	3,20	3,25	3,29	3,32	3,35	3,37	3,40	3,43	3,45	3,46	3,47
40	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	3,35	3,39	3,42	3,44	3,46	3,47
60	2,83	2,98	3,08	3,14	3,20	3,24	3,28	3,31	3,33	3,37	3,40	3,44	3,46	3,47
100	2,80	2,95	3,05	3,12	3,18	3,22	3,26	3,29	3,32	3,36	3,40	3,43	3,45	3,47
∞	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26	3,29	3,34	3,38	3,41	3,44	3,47

б) для $P = 0,99$

1	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0
2	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0
3	8,26	8,5	8,6	8,7	8,8	8,9	8,9	8,9	9,0	9,0	9,1	9,2	9,3	9,3
4	6,51	6,8	6,9	7,0	7,1	7,1	7,2	7,2	7,3	7,3	7,4	7,4	7,5	7,5
5	5,70	5,96	6,11	6,18	6,26	6,33	6,40	6,44	6,5	6,6	6,6	6,7	6,8	6,8
6	5,24	5,51	5,65	5,73	5,81	5,88	5,95	6,00	6,0	6,1	6,2	6,3	6,3	6,3
7	4,95	5,22	5,37	5,45	5,53	5,61	5,69	5,73	5,8	5,8	5,9	6,0	6,0	6,0
8	4,74	5,00	5,14	5,23	5,32	5,40	5,47	5,51	5,5	5,6	5,7	5,7	5,8	5,8
9	4,60	4,86	4,99	5,08	5,17	5,25	5,32	5,36	5,4	5,5	5,5	5,6	5,7	5,7
10	4,48	4,73	4,88	4,96	5,06	5,13	5,20	5,24	5,28	5,36	5,42	5,48	5,54	5,55
11	4,39	4,63	4,77	4,86	4,94	5,01	5,06	5,12	5,15	5,24	5,28	5,34	5,38	5,39
12	4,32	4,55	4,68	4,76	4,84	4,92	4,96	5,02	5,07	5,13	5,17	5,22	5,24	5,26
13	4,26	4,48	4,62	4,69	4,74	4,84	4,88	4,94	4,98	5,04	5,08	5,13	5,14	5,15
14	4,21	4,42	4,55	4,63	4,70	4,78	4,83	4,87	4,91	4,96	5,00	5,04	5,06	5,07
15	4,17	4,37	4,50	4,58	4,64	4,72	4,77	4,81	4,84	4,90	4,94	4,97	4,99	5,00
16	4,13	4,34	4,45	4,54	4,60	4,67	4,72	4,76	4,79	4,84	4,88	4,91	4,93	4,94
17	4,10	4,30	4,41	4,50	4,56	4,63	4,68	4,72	4,75	4,80	4,86	4,88	4,89	4,89
18	4,07	4,27	4,38	4,46	4,53	4,59	4,64	4,68	4,71	4,76	4,82	4,84	4,85	4,85
19	4,05	4,24	4,35	4,43	4,50	4,56	4,61	4,64	4,67	4,72	4,76	4,79	4,81	4,82
20	3,99	4,17	4,28	4,36	4,42	4,48	4,53	4,58	4,60	4,65	4,68	4,71	4,74	4,75
24	3,96	4,14	4,24	4,33	4,39	4,44	4,49	4,53	4,57	4,62	4,65	4,68	4,71	4,72
26	3,93	4,11	4,21	4,30	4,36	4,41	4,46	4,50	4,53	4,58	4,62	4,65	4,67	4,69
28	3,91	4,08	4,18	4,28	4,34	4,39	4,43	4,47	4,51	4,56	4,60	4,62	4,65	4,67
30	3,89	4,06	4,16	4,26	4,32	4,36	4,41	4,45	4,48	4,54	4,58	4,61	4,63	4,65
40	3,82	3,99	4,10	4,17	4,24	4,30	4,34	4,37	4,41	4,46	4,51	4,54	4,57	4,59
60	3,76	3,92	4,03	4,12	4,17	4,23	4,27	4,31	4,34	4,39	4,44	4,47	4,50	4,53
100	3,71	3,86	3,98	4,06	4,11	4,17	4,21	4,25	4,29	4,35	4,38	4,42	4,45	4,48
∞	3,64	3,80	3,90	3,98	4,04	4,09	4,14	4,17	4,20	4,26	4,31	4,34	4,38	4,41

Квадраты чисел от 1 до 1000

Таблица 12.7

	0	1	2	3	4	5	6	7	8	9
0	0	1	4	9	16	25	36	49	64	81
10	100	121	144	169	196	225	256	289	324	361
20	400	441	484	529	576	625	676	729	784	841
30	900	961	1 024	1 089	1 156	1 225	1 296	1 369	1 444	1 521
40	1 600	1 681	1 764	1 849	1 936	2 025	2 116	2 209	2 304	2 401
50	2 500	2 601	2 704	2 809	2 916	3 025	3 136	3 249	3 364	3 481
60	3 600	3 721	3 844	3 969	4 096	4 225	4 356	4 489	4 624	4 761
70	4 900	5 041	5 184	5 329	5 476	5 625	5 776	5 929	6 084	6 241
80	6 400	6 561	6 724	6 889	7 056	7 225	7 396	7 569	7 744	7 921
90	8 100	8 281	8 464	8 649	8 836	9 025	9 216	9 409	9 604	9 801
100	10 000	10 201	10 404	10 609	10 816	11 025	11 236	11 449	11 664	11 881
110	12 100	12 321	12 544	12 769	12 996	13 225	13 456	13 689	13 924	14 161
120	14 400	14 641	14 884	15 129	15 376	15 625	15 876	16 129	16 384	16 641
130	16 900	17 161	17 424	17 689	17 956	18 225	18 496	18 769	19 044	19 321
140	19 600	19 881	20 164	20 449	20 736	21 025	21 316	21 609	21 904	22 201
150	22 500	22 801	23 104	23 409	23 716	24 025	24 336	24 649	24 964	25 281
160	25 600	25 921	26 244	26 569	26 896	27 225	27 556	27 889	28 224	28 561
170	28 900	29 241	29 584	29 929	30 276	30 625	30 976	31 329	31 684	32 041
180	32 400	32 761	33 124	33 489	33 856	34 225	34 596	34 969	35 344	35 721
190	36 100	36 481	36 864	37 249	37 636	38 025	38 416	38 809	39 204	39 601
0	0	1	2	3	4	5	6	7	8	9

Продолжение табл. 12.7

	0	1	2	3	4	5	6	7	8	9
200	40 000	40 401	40 804	41 209	41 616	42 025	42 436	42 849	43 264	43 681
210	44 100	44 521	44 944	45 369	45 796	46 225	46 656	47 089	47 524	47 961
220	48 400	48 841	49 284	49 729	50 176	50 625	51 076	51 529	51 984	52 441
230	52 900	53 361	53 824	54 289	54 756	55 225	55 696	56 169	56 644	57 121
240	57 600	58 081	58 564	59 049	59 536	60 025	60 516	61 009	61 504	62 001
250	62 500	63 001	63 504	64 009	64 516	65 025	65 536	66 049	66 564	67 081
260	67 600	68 121	68 644	69 169	69 696	70 225	70 756	71 289	71 824	72 361
270	72 900	73 441	73 984	74 529	75 076	75 625	76 176	76 729	77 284	77 841
280	78 400	78 961	79 524	80 089	80 656	81 225	81 796	82 369	82 944	83 521
290	84 100	84 681	85 264	85 849	86 436	87 025	87 616	88 209	88 804	89 401
300	90 000	90 601	91 204	91 809	92 416	93 025	93 636	94 249	94 864	95 481
310	96 100	96 721	97 344	97 969	98 596	99 225	99 856	100 489	101 124	101 761
320	102 400	103 041	103 684	104 329	104 976	105 625	106 276	106 929	107 584	108 241
330	108 900	109 561	110 224	110 889	111 556	112 225	112 896	113 569	114 244	114 921
340	115 600	116 281	116 964	117 649	118 336	119 025	119 716	120 409	121 104	121 801
350	122 500	123 201	123 904	124 609	125 316	126 025	126 736	127 449	128 164	128 881
360	129 600	130 321	131 044	131 769	132 496	133 225	133 956	134 689	135 424	136 161
370	136 900	137 641	138 384	139 129	139 876	140 625	141 376	142 129	142 884	143 641
380	144 400	145 161	145 924	146 689	147 456	148 225	148 996	149 769	150 544	151 321
390	152 100	152 881	153 664	154 449	155 236	156 025	156 816	157 609	158 404	159 201
0	0	1	2	3	4	5	6	7	8	9

Продолжение табл. 12.7

	0	1	2	3	4	5	6	7	8	9
400	160 000	160 801	161 604	162 409	163 216	164 025	164 836	165 649	166 464	167 281
410	168 100	168 921	169 744	170 659	171 396	172 225	173 056	173 889	174 724	175 561
420	176 400	177 241	178 084	178 929	179 776	180 625	181 476	182 329	183 184	184 041
430	184 900	185 761	186 624	187 489	188 356	189 225	190 969	190 066	191 844	192 721
440	193 600	194 481	195 364	196 249	197 136	198 025	198 096	199 809	200 704	201 601
450	202 500	203 401	204 304	205 209	206 116	207 025	207 936	208 849	209 764	210 681
460	211 600	212 521	213 444	214 369	215 296	216 225	217 156	218 089	219 024	219 961
470	220 900	221 841	222 784	223 729	224 676	225 625	226 576	227 529	228 484	229 441
480	230 400	231 361	232 324	233 289	234 256	235 225	236 196	237 169	238 144	239 121
490	240 100	241 081	242 064	243 049	244 036	245 025	246 016	247 009	248 004	249 001
500	250 000	251 001	252 004	253 009	254 016	255 025	256 036	257 049	258 064	259 081
510	260 100	261 121	262 144	263 169	264 196	265 225	266 256	267 289	268 324	269 361
520	270 400	271 441	272 484	273 529	274 576	275 625	276 676	277 729	278 784	279 841
530	280 900	281 961	283 024	284 089	285 156	286 225	287 296	288 369	289 444	290 521
540	291 600	292 681	293 764	294 849	295 936	297 025	298 116	299 209	200 304	301 401
550	302 500	303 601	304 704	305 809	306 916	308 025	309 136	310 249	311 364	312 481
560	313 600	314 721	315 844	316 969	318 096	319 225	320 356	321 489	322 624	323 761
570	324 900	326 041	327 184	328 329	329 476	330 625	331 776	332 929	334 084	335 241
580	336 400	337 561	338 724	339 889	341 056	342 225	343 396	344 569	345 744	346 921
590	348 100	349 281	350 464	351 649	352 836	354 025	355 216	356 409	357 604	358 801
0	0	1	2	3	4	5	6	7	8	9

Продолжение табл. 12.7

	0	1	2	3	4	5	6	7	8	9
600	360 000	361 201	362 404	363 609	364 816	366 025	367 236	368 449	369 664	370 881
610	372 100	373 321	374 544	375 769	376 996	378 225	379 456	380 689	381 924	383 161
620	384 400	385 641	386 884	388 129	389 376	390 625	391 876	393 129	394 384	395 641
630	396 900	398 161	399 424	400 689	401 956	403 225	404 496	405 769	407 044	408 321
640	409 600	410 881	412 164	413 449	414 736	416 025	417 316	418 609	419 904	421 201
650	422 500	423 801	425 104	426 409	427 716	429 025	430 336	431 649	432 964	434 281
660	435 600	436 921	438 244	439 569	440 896	442 225	443 556	444 889	446 224	447 561
670	448 900	450 241	451 584	452 929	454 276	455 625	456 976	458 329	459 684	461 041
680	462 400	463 671	465 124	466 489	467 856	469 225	470 596	471 969	473 344	474 721
690	476 100	477 481	478 864	480 249	481 636	483 025	484 416	485 809	487 204	488 601
700	490 000	491 401	492 804	494 209	495 616	497 025	498 436	499 849	501 264	502 681
710	504 100	505 521	506 944	508 369	509 796	511 225	512 656	514 089	515 524	516 961
720	518 400	519 841	521 284	522 729	524 176	525 625	527 076	528 529	529 984	531 441
730	532 900	534 361	535 824	537 289	538 756	540 225	541 696	543 169	544 644	546 121
740	547 600	549 081	550 064	552 049	553 536	555 025	556 516	558 009	559 504	561 001
750	562 500	564 001	565 504	567 009	568 516	570 025	571 536	573 049	574 564	576 081
760	577 600	579 121	580 644	582 169	583 696	585 225	586 756	588 289	589 824	591 361
770	592 900	594 441	595 984	597 529	599 076	600 625	602 176	603 729	605 284	606 841
780	608 400	609 961	611 524	613 089	614 656	616 225	617 796	619 369	620 944	622 521
790	624 100	625 681	627 264	628 849	630 436	632 025	633 616	635 209	636 804	638 401
0	0	1	2	3	4	5	6	7	8	9

	0	1	2	3	4	5	6	7	8	9
800	640 000	641 601	643 204	644 809	646 416	648 025	649 636	651 249	652 864	654 481
810	656 100	657 721	659 344	660 969	662 596	664 225	665 856	667 489	669 124	670 761
820	672 400	674 041	675 684	677 329	678 976	680 625	682 276	683 929	685 584	687 241
830	688 900	690 561	692 224	693 889	695 556	697 225	698 896	700 569	702 244	703 921
840	705 600	707 281	708 964	710 649	712 336	714 025	715 716	717 409	719 014	720 801
850	722 500	724 201	725 904	727 609	729 316	731 025	732 736	734 449	736 164	737 881
860	739 600	741 321	743 044	744 769	746 496	748 225	749 956	751 689	753 424	755 161
870	756 900	758 641	760 384	762 129	763 876	765 625	767 376	769 129	770 884	772 641
880	774 400	776 161	777 924	779 689	781 456	783 225	784 996	786 769	788 544	790 321
890	792 100	793 881	795 664	797 449	799 236	801 025	802 816	804 609	806 404	808 201
900	810 000	811 801	813 604	815 409	817 216	819 025	820 836	822 649	824 464	826 281
910	828 100	829 921	831 744	833 569	835 396	837 225	839 056	840 889	842 724	844 561
920	846 400	848 241	850 084	851 929	853 776	855 625	857 476	859 329	861 184	863 041
930	864 900	866 761	868 624	870 489	872 356	874 225	876 096	877 969	879 844	881 721
940	883 600	885 481	887 364	889 249	891 136	893 025	894 916	896 809	898 704	900 601
950	902 500	904 401	906 304	908 209	910 116	912 025	913 936	915 849	917 764	919 681
960	921 600	923 521	925 444	927 369	929 296	931 225	933 156	935 089	937 024	938 961
970	940 900	942 841	944 784	946 729	948 676	950 625	952 576	954 529	956 484	958 441
980	960 400	962 361	964 324	966 289	968 256	970 225	972 196	974 169	976 144	978 121
990	980 100	982 081	984 064	986 049	988 036	990 025	992 016	994 009	996 004	998 001

Обратные величины от 1 до 1000

Делитель	0	1	2	3	4	5	6	7	8	9
1	—	90 909	83 333	76 923	71 429	66 667	62 500	58 824	55 536	52 632
2	50 000	47 619	45 455	43 478	41 667	40 000	38 462	37 037	35 714	34 483
3	33 333	32 258	31 250	30 303	29 412	28 574	27 778	27 027	26 316	25 641
4	25 000	24 390	23 810	23 256	22 727	22 222	21 739	21 277	20 833	20 408
5	20 000	19 608	19 231	18 868	18 519	18 182	17 857	17 544	17 241	16 949
6	16 667	16 393	16 129	15 873	15 625	15 385	15 152	14 925	14 706	14 493
7	14 286	14 085	13 889	13 699	13 514	13 333	13 158	12 987	12 821	12 658
8	12 500	12 346	12 195	12 048	11 905	11 765	11 628	11 494	11 364	11 236
9	11 111	10 989	10 870	10 753	10 638	10 526	10 417	10 309	10 204	10 101
10	10 000	9 901	9 804	9 709	9 615	9 524	9 434	9 346	9 259	9 174
11	9 091	9 009	8 929	8 850	8 772	8 696	8 621	8 547	8 475	8 403
12	8 333	8 264	8 197	8 130	8 065	8 000	7 937	7 874	7 813	7 752
13	7 692	7 634	7 576	7 519	7 463	7 407	7 353	7 299	7 246	7 194
14	7 143	7 092	7 042	6 993	6 944	6 897	6 849	6 803	6 757	6 711
15	6 667	6 623	6 579	6 536	6 494	6 452	6 410	6 369	6 329	6 289
16	6 250	6 211	6 173	6 135	6 098	6 061	6 024	5 988	5 952	5 917
17	5 882	5 848	5 814	5 780	5 747	5 714	5 682	5 650	5 618	5 587
18	5 556	5 525	5 495	5 464	5 435	5 405	5 376	5 348	5 319	5 291
19	5 263	5 236	5 208	5 181	5 155	5 129	5 102	5 076	5 051	5 025

Продолжение табл. 12.8

Делитель ↓	0	1	2	3	4	5	6	7	8	9
20	5 000	4 975	4 951	4 926	4 902	4 878	4 854	4 831	4 808	4 785
21	4 762	4 739	4 717	4 695	4 673	4 651	4 630	4 608	4 587	4 566
22	4 545	4 525	4 505	4 484	4 464	4 444	4 425	4 405	4 386	4 367
23	4 348	4 329	4 310	4 292	4 274	4 255	4 237	4 219	4 202	4 184
24	4 167	4 149	4 132	4 115	4 098	4 082	4 069	4 049	4 032	4 016
25	4 000	3 984	3 968	3 953	3 937	3 922	3 906	3 891	3 876	3 861
26	3 846	3 831	3 817	3 802	3 788	3 774	3 759	3 745	3 731	3 717
27	3 704	3 690	3 676	3 663	3 650	3 636	3 623	3 610	3 597	3 584
28	3 571	3 559	3 546	3 534	3 521	3 509	3 497	3 484	3 472	3 460
29	3 448	3 436	3 425	3 413	3 401	3 390	3 378	3 367	3 356	3 344
30	3 333	3 322	3 311	3 300	3 289	3 279	3 268	3 257	3 247	3 236
31	3 226	3 215	3 205	3 195	3 185	3 175	3 165	3 155	3 145	3 135
32	3 125	3 115	3 106	3 096	3 086	3 077	3 067	3 058	3 049	3 040
33	3 030	3 021	3 012	3 003	2 994	2 985	2 976	2 967	2 959	2 950
34	2 941	2 933	2 924	2 915	2 907	2 899	2 890	2 882	2 874	2 865
35	2 857	2 849	2 841	2 833	2 825	2 817	2 809	2 801	2 793	2 786
36	2 778	2 770	2 762	2 755	2 747	2 740	2 732	2 725	2 717	2 710
37	2 703	2 695	2 688	2 681	2 674	2 667	2 660	2 653	2 646	2 639
38	2 632	2 625	2 618	2 611	2 604	2 597	2 591	2 584	2 577	2 571
39	2 564	2 558	2 551	2 545	2 538	2 532	2 525	2 519	2 513	2 506
	0	1	2	3	4	5	6	7	8	9

Продолжение табл. 12.8

Делитель ↓	0	1	2	3	4	5	6	7	8	9
40	2 500	2 494	2 488	2 481	2 475	2 469	2 463	2 457	2 451	2 445
41	2 439	2 433	2 427	2 421	2 415	2 410	2 404	2 398	2 392	2 387
42	2 381	2 375	2 370	2 364	2 358	2 353	2 347	2 342	2 336	2 331
43	2 326	2 320	2 315	2 309	2 304	2 299	2 294	2 288	2 283	2 278
44	2 273	2 268	2 262	2 257	2 252	2 247	2 242	2 237	2 232	2 227
45	2 222	2 217	2 212	2 208	2 203	2 198	2 193	2 188	2 183	2 179
46	2 174	2 169	2 165	2 160	2 155	2 151	2 146	2 141	2 137	2 132
47	2 128	2 123	2 119	2 114	2 110	2 105	2 101	2 096	2 092	2 088
48	2 083	2 079	2 075	2 070	2 066	2 062	1 058	2 053	2 049	2 045
49	2 041	2 037	2 033	2 028	2 024	2 020	2 016	2 012	2 008	2 004
50	2 000	1 996	1 992	1 988	1 984	1 980	1 976	1 972	1 969	1 965
51	1 961	1 957	1 953	1 949	1 946	1 942	1 938	1 934	1 931	1 927
52	1 923	1 919	1 916	1 912	1 908	1 905	1 901	1 898	1 894	1 890
53	1 887	1 883	1 880	1 876	1 873	1 869	1 866	1 862	1 859	1 855
54	1 852	1 848	1 845	1 842	1 838	1 835	1 832	1 828	1 825	1 821
55	1 818	1 815	1 812	1 808	1 805	1 802	1 799	1 795	1 792	1 789
56	1 786	1 783	1 779	1 776	1 773	1 770	1 767	1 764	1 761	1 757
57	1 754	1 751	1 748	1 745	1 742	1 739	1 736	1 733	1 730	1 727
58	1 724	1 721	1 718	1 715	1 712	1 709	1 706	1 704	1 701	1 698
59	1 695	1 692	1 689	1 686	1 684	1 681	1 678	1 675	1 672	1 669
	0	1	2	3	4	5	6	7	8	9

Продолжение табл. 12.8

Делитель ↓	0	1	2	3	4	5	6	7	8	9
60	1 667	1 664	1 661	1 658	1 656	1 653	1 650	1 647	1 645	1 642
61	1 639	1 637	1 634	1 631	1 629	1 626	1 623	1 621	1 618	1 616
62	1 613	1 610	1 608	1 605	1 603	1 600	1 597	1 595	1 592	1 590
63	1 587	1 585	1 582	1 580	1 577	1 575	1 572	1 570	1 567	1 565
64	1 563	1 560	1 558	1 555	1 553	1 550	1 548	1 546	1 543	1 541
65	1 538	1 536	1 534	1 531	1 529	1 527	1 524	1 522	1 520	1 517
66	1 515	1 513	1 511	1 508	1 506	1 504	1 502	1 499	1 497	1 495
67	1 493	1 490	1 488	1 486	1 484	1 481	1 479	1 477	1 475	1 473
68	1 471	1 468	1 466	1 464	1 462	1 460	1 458	1 456	1 453	1 451
69	1 449	1 447	1 445	1 443	1 441	1 439	1 437	1 435	1 433	1 431
70	1 429	1 427	1 425	1 422	1 420	1 418	1 416	1 414	1 412	1 410
71	1 408	1 406	1 404	1 403	1 401	1 399	1 397	1 395	1 393	1 391
72	1 389	1 387	1 385	1 383	1 381	1 379	1 377	1 376	1 374	1 372
73	1 370	1 368	1 366	1 364	1 362	1 361	1 359	1 357	1 355	1 353
74	1 351	1 350	1 348	1 346	1 344	1 342	1 340	1 339	1 337	1 335
75	1 333	1 332	1 330	1 328	1 326	1 325	1 323	1 321	1 319	1 318
76	1 316	1 314	1 312	1 311	1 309	1 307	1 305	1 304	1 302	1 300
77	1 299	1 297	1 295	1 294	1 292	1 290	1 289	1 287	1 285	1 284
78	1 282	1 280	1 279	1 277	1 276	1 274	1 272	1 271	1 269	1 267
79	1 266	1 264	1 263	1 261	1 259	1 258	1 256	1 255	1 253	1 252
0	0	1	2	3	4	5	6	7	8	9

Продолжение табл. 12.8

Делитель ↓	0	1	2	3	4	5	6	7	8	9
80	1 250	1 248	1 247	1 245	1 244	1 242	1 241	1 239	1 238	1 236
81	1 235	1 233	1 232	1 230	1 229	1 227	1 225	1 224	1 222	1 221
82	1 220	1 218	1 217	1 215	1 214	1 212	1 211	1 209	1 208	1 206
83	1 205	1 203	1 202	1 200	1 199	1 198	1 196	1 195	1 193	1 192
84	1 190	1 189	1 188	1 186	1 185	1 183	1 182	1 181	1 179	1 178
85	1 176	1 175	1 174	1 172	1 171	1 170	1 168	1 167	1 166	1 164
86	1 163	1 161	1 160	1 159	1 157	1 156	1 155	1 153	1 152	1 151
87	1 149	1 148	1 147	1 145	1 144	1 143	1 142	1 140	1 139	1 138
88	1 136	1 135	1 134	1 133	1 131	1 130	1 129	1 127	1 126	1 125
89	1 124	1 122	1 121	1 120	1 119	1 117	1 116	1 115	1 114	1 112
90	1 111	1 110	1 109	1 107	1 106	1 105	1 104	1 103	1 101	1 100
91	1 099	1 098	1 096	1 095	1 094	1 093	1 092	1 091	1 089	1 088
92	1 087	1 086	1 085	1 083	1 082	1 081	1 080	1 079	1 078	1 076
93	1 075	1 074	1 073	1 072	1 071	1 070	1 068	1 067	1 066	1 065
94	1 064	1 063	1 062	1 060	1 059	1 058	1 057	1 056	1 055	1 054
95	1 053	1 052	1 050	1 049	1 048	1 047	1 046	1 045	1 044	1 043
96	1 042	1 041	1 040	1 038	1 037	1 036	1 035	1 034	1 033	1 032
97	1 031	1 030	1 029	1 028	1 027	1 026	1 025	1 024	1 022	1 021
98	1 020	1 019	1 018	1 017	1 016	1 015	1 014	1 013	1 012	1 011
99	1 010	1 009	1 008	1 007	1 006	1 005	1 004	1 003	1 002	1 001
0	0	1	2	3	4	5	6	7	8	9

СОДЕРЖАНИЕ

Предисловие	5
Предисловие автора	7
Использованные обозначения	9
1. Виды ошибок в аналитической химии	11
Литература	17
2. Эмпирические распределения частот	18
2.1. Обсуждение эмпирических распределений частот	18
2.2. Статистические характеристики	25
2.2.1. Средние значения	25
2.2.2. Границы разброса	29
2.2.3. Асимметрия и эксцесс	31
2.2.4. Табличное вычисление арифметического среднего и средней квадратичной ошибки	34
2.3. Двумерное распределение	36
Литература	41
3. Теоретические распределения	42
3.1. Нормальное распределение	42
3.2. Распределение Пуассона	56
3.3. Специальные распределения	59
3.3.1. t -Распределение	59
3.3.2. F -Распределение	59
3.3.3. χ^2 -Распределение	61
3.4. Связь между отдельными распределениями	62
Литература	63
4. Закон распространения ошибок	64
4.1. Гравиметрия	66
4.2. Объемный анализ	69
4.3. Фотометрия	72
4.4. Косвенные методы	78
4.5. Счетные методы	81
4.6. Отбор проб	87
Литература	92
5. Случайные ошибки методов анализа	93
5.1. Вычисление средней квадратичной ошибки	93

5.2. Примеры	104
Литература	110
6. Обсуждение данных анализа	111
6.1. Вычисление доверительного интервала	111
6.2. Статистическая оценка качества	118
Литература	130
7. Статистические методы проверки	131
7.1. Сравнение двух средних квадратичных ошибок (F -критерий)	133
7.2. Сравнение нескольких средних квадратичных ошибок (критерий Бартлета)	138
7.3. Сравнение двух средних значений (t -критерий)	141
7.4. Сравнение частот	146
7.5. Определение грубых ошибок	147
7.6. Исследование эмпирических распределений	150
Литература	153
8. Простой дисперсионный анализ	154
8.1. Несколько источников случайной ошибки	154
8.2. Устранение ошибок	161
8.3. Сравнение нескольких средних значений	168
Литература	173
9. Статистика прямых линий (регрессионный и корреляционный анализ)	175
9.1. Определение констант	176
9.2. Проверка констант	182
9.3. Проверка гипотезы линейности	185
9.4. Оценка вторичных аналитических методов	189
9.5. Определение систематических ошибок	195
9.6. Линейная регрессия при трех и более переменных	206
9.7. Проверка взаимной зависимости двух переменных (вычисление корреляции)	209
Литература	213
10. Приемы вычислений	214
Литература	219
11. Послесловие	220
Литература	221
12. Приложение (таблицы)	223

К. ДОЕРФЕЛЬ

**СТАТИСТИКА
В АНАЛИТИЧЕСКОЙ ХИМИИ**

Редактор *Г. М. Мануйлова*
Художник *К. П. Сиротов*
Художественный редактор *Р. В. Свириш*
Технический редактор *Н. А. Турсулова*
Корректор *О. Ф. Иванова*

Сдано в производство 7/VIII 1968 г.
Подписано к печати 31/I 1969 г.
Бумага № 1 84 × 108¹/₃₂ = 3,88 бум. л.
13,02 усл. печ. л. Уч.-изд. л. 11,83.
Изд. № 3/4486. Цена 1 р. 03 к. Зак. № 438
(Тем. план 1968 г. из-ва «Мир», пор. № 129)

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-Рижский пер., 2

Московская типография № 16
Главполиграфпрома Комитета по печати
при Совете Министров СССР
Москва, Трехпрудный пер., 9