А.В. КОЗАЧЕК

ОСНОВЫ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ

♦ ИЗДАТЕЛЬСТВО ТГТУ ♦

УДК 504.75(075.8)
ББК Б1я73
К59

Р е ц е н з е н т ы:

Доктор технических наук, профессор заведующий лабораторией
Государственного научного учреждения
"Всероссийский научно-исследовательский и проектно-технологический институт по использованию техники и нефтепродуктов в
сельском хозяйстве" Российской академии сельскохозяйственных наук
С.А. Нагорнов

Кандидат физико-математических наук, доцент кафедры высшей
математики Тамбовского государственного технического университета
В.В. Васильев

Кандидат химических наук, доцент кафедры химии
Тамбовского государственного технического университета
Н.А. Абакумова

Козачек, А.В.

К59 Основы инженерных исследований в экологии : учебное пособие / А.В. Козачек. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2007. – 76 с. – 100 экз. – ISBN 978-5-8265-0658-5.

Составлено на основе требований Государственного образовательного стандарта высшего профессионального образования к уровню подготовки дипломированного специалиста по направлению "Защита окружающей среды" (специальность "Инженерная защита окружающей среды"). Изложены существующие виды и практические методы исследований и экспериментов, основы математического моделирования экологических и промышленных систем, различные способы обработки экспериментальных данных, полученных в ходе инженерно-экологического эксперимента.

Предназначено для подготовки студентов, бакалавров и магистров по инженерно-экологическим специальностям.

УДК 504.75(075.8)
ББК Б1я73

ISBN 978-5-8265-0658-5

Министерство образования и науки Российской Федерации

ГОУ ВПО "Тамбовский государственный технический университет"

А.В. Козачек

ОСНОВЫ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ

Утверждено Ученым советом университета в качестве
учебного пособия для студентов, бакалавров и магистров по
инженерно-экологическим специальностям

Тамбов
Издательство ТГТУ
2007

Учебное издание

КОЗАЧЕК Артёмий Владимирович

ОСНОВЫ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ

Учебное пособие

Редактор З.Г. Ч е р н о в а
Инженер по компьютерному макетированию М.Н. Р ы ж к о в а
Корректор О.М. Я р ц е в а

Подписано в печать 18.11.2007.
Формат 60 × 84/16. 4,42 усл. печ. л. Тираж 100 экз. Заказ № 718

Издательско-полиграфический центр
Тамбовского государственного технического университета
392000, Тамбов, Советская, 106, к. 14

Содержание

ВВЕДЕНИЕ. ЦЕЛЬ И ЗАДАЧИ КУРСА "ОСНОВЫ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ"

Полная и достоверная статистическая информация является тем необходимым основанием, на котором базируется процесс управления охраной окружающей среды. Принятие управленческих решений в области охраны окружающей среды на всех уровнях – от общегосударственного или регионального и до уровня фирмы или человека – невозможно без должного статистического обеспечения. Именно статистические данные позволяют определить объемы загрязнения воздуха, воды и почвы, выявить основные тенденции влияния загрязняющих веществ на здоровье населения, оценить уровень воздействия отдельных отраслей промышленности на окружающую среду, проанализировать состояние экосистем и ландшафтов, исследовать уровень существования организмов при антропогенном воздействии человека и т.д. Статистические данные получаются в процессе проведения соответствующего исследования.

Целью курса "Основы инженерных исследований в экологии" является изучение способов исследований явлений и процессов, происходящих в техносфере и окружающей среде.

На основе вышесказанного можно выделить **задачи практической реализации положений курса** "Основы инженерных исследований в экологии", а именно:

– максимальное сокращение сроков перехода от лабораторных исследований в практическую деятельность;
– выявление оптимального варианта осуществления процесса;
– получение наиболее достоверных результатов исследования;
– разработка методик обработки экспериментальных данных и планирования экспериментов.

1. СУЩНОСТЬ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ

1.1. ПОНЯТИЕ ИНЖЕНЕРНОГО ИССЛЕДОВАНИЯ В ЭКОЛОГИИ

Исследование – это изучение любых объектов путем проведения над ним эксперимента.

Соответственно, **инженерное исследование в экологии** – это изучение любых природных и антропогенных объектов природо-промышлен-
ных систем путем проведения над ним эксперимента.

Природо-промышленная система (ППС) – это множество объектов отраслей промышленного и сельскохозяйственного производства и объектов природной среды, образующих единую технико-экономическую и экологическую структуру рассматриваемого региона, упорядоченно взаимодействующих друг с другом в процессах обмена информацией, потребления материально-энергетических ресурсов и переработки отходов.

В состав природо-промышленной системы входят:

1) промышленная (антропогенная) подсистема (техносфера):

а) компоненты промышленной подсистемы:

– объекты (сфера) общественного потребления, являющиеся источниками выделения загрязняющих веществ, энергии и других факторов воздействия на окружающую среду (жилой сектор, транспорт, торговые и обслуживающие организации и т.д.);

– объекты (сфера) промышленного производства, являющиеся источниками выделения загрязняющих веществ, энергии и других факторов воздействия на окружающую среду (предприятия промышленности и сельского хозяйства, вокзалы, аэродромы, электростанции и т.д.);

– объекты (сфера) защиты окружающей среды (очистные сооружения, системы размещения, переработки или захоронения отходов и т.д.);

б) внутренние и внешние связи в промышленной подсистеме (потоки первичных природных ресурсов, идущих на общественные, промышленные и эколого-технологические нужды, потоки промышленных потребительских продуктов, потоки перерабатываемых и неперерабатываемых общественных и промышленных отходов и т.д.);

2) природная подсистема (окружающая среда) – экосфера с флорой и фауной (компоненты и связи атмосферы, гидросферы и литосферы).

Природо-промышленные системы являются объектом профессиональной деятельности инженера-эколога.

1.2. ЭЛЕМЕНТЫ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ

1.2.1. Цель инженерного исследования в экологии

Инженерные исследования в экологии чаще всего преследуют практическую **цель** – получение достоверной информации для выявления закономерностей развития явлений и процессов, происходящих в природо-промышленной системе. Задача наблюдения предопределяет его программу и формы организации. Неясно поставленная цель может привести к тому, что в процессе наблюдения будут собраны ненужные данные или, наоборот, не будут получены сведения, необходимые для анализа.

1.2.2. Объект инженерного исследования в экологии

При подготовке инженерного исследования кроме цели следует точно определить, что именно подлежит обследованию, т.е. установить объект инженерного исследования в экологии.

Под **объектом инженерного исследования в экологии** понимается некоторая статистическая совокупность, в которой проистекают исследуемые явления и процессы, происходящие в природо-промышленной системе. Объектом инженерного исследования в экологии может быть совокупность физических лиц (население отдельного региона, страны; лица, занятые на предприятиях отрасли), физические единицы (станки, машины, жилые дома), юридические лица (предприятия, фермерские хозяйства, коммерческие банки, учебные заведения), природные объекты (растительность, животный мир, недра).

Чтобы определить объект наблюдения, необходимо установить границы изучаемой совокупности. Для этого следует указать важнейшие признаки, отличающие его от других сходных объектов. Например, прежде чем проводить обследование влияния промышленных предприятий на окружающую их среду, следует определить отрасли промышленности, регионы, подлежащие наблюдению и т.д.

1.2.3. Единица инженерного исследования в экологии

Всякий объект инженерного исследования в экологии состоит из отдельных элементов – единиц инженерного исследования в экологии.

Единицей инженерного исследования в экологии (в зарубежной литературе используется термин "единица наблюдения", "элементарная единица") называют составной элемент объекта, являющийся носителем признаков, подлежащих регистрации. Например, при исследовании воздействия промышленных предприятий на окружающую их среду единицей наблюдения может быть человек, но может быть и все население, проживающее в населенных пунктах, расположенных рядом с предприятием.

По каждой единице инженерного исследования собираются необходимые сведения – статистические данные. Под **статистическими данными** (информацией) понимают совокупность количественных характеристик экологических явлений и процессов, полученных в результате статистического наблюдения, их обработки или соответствующих расчетов.

1.2.4. Программа инженерного исследования в экологии

Всякое явление обладает множеством различных признаков. Собирать информацию по всем признакам нецелесообразно, а часто и невозможно. Поэтому необходимо отобрать те признаки, которые являются существенными, основными для характеристики объекта, исходя из цели исследования. Для определения состава регистрируемых признаков разрабатывают программу инженерного исследования в экологии.

Программа инженерного исследования в экологии – это перечень признаков (или вопросов), подлежащих регистрации в процессе наблюдения. От того, насколько хорошо разработана программа статистического наблюдения, во многом зависит качество собранной информации.

Чтобы составить правильно программу инженерного исследования в экологии, исследователь должен:

– ясно представлять задачи обследования конкретного явления или процесса;

– определить состав используемых в анализе методов, необходимые группировки;

– на основе этого выявить те признаки, которые можно определить при проведении исследовательской работы.

Обычно программа выражается в форме вопросов переписного (опросного) листа.

К программе статистического инженерного исследования в экологии предъявляются следующие требования:

1) программа должна содержать существенные признаки, непосредственно характеризующие изучаемое явление, его тип, основные черты, свойства; не следует включать в программу признаки, имеющие второстепенное значение по отношению к цели обследования или значения которых заведомо будут недостоверны или отсутствовать (например, при незаинтересованности предприятий в представлении информации по используемому очистному оборудованию, так как она является предметом коммерческой тайны);

2) вопросы программы должны быть точными и не двусмысленными, иначе полученный ответ может содержать неверную информацию, а также легкими для понимания во избежание лишних трудностей при получении ответов;

3) при разработке программы следует не только определить состав вопросов, но и их последовательность; логичный порядок в последовательности вопросов (признаков) поможет получить достоверные сведения о явлениях и процессах.

1.2.5. Время инженерного исследования в экологии

Выбор времени инженерного исследования в экологии заключается в решении двух вопросов:

– установление критического момента (даты) или интервала времени;

– определение срока (периода) наблюдения.

Под **критическим моментом (датой) исследования** понимаются конкретные день года, час дня, по состоянию на который должна быть проведена регистрация признаков по каждой единице исследуемой совокупности. Критический момент устанавливается с целью получения сопоставимых статистических данных. Если же надо проанализировать изменение значений статистических данных, например в отчетном месяце по сравнению с предыдущим месяцем, то устанавливается не критический момент, а интервал времени, за который следует получить статистические данные. Выбор критического момента или интервала времени определяется, прежде всего, целью исследования.

Срок (период) исследования – это время, в течение которого происходит заполнение статистических формуляров, т.е. время, необходимое для проведения массового сбора данных. Этот срок определяется исходя из объема работы (числа регистрируемых признаков и единиц в обследуемой совокупности), численности персонала, занятого сбором информации. Следует учитывать, что отдаление периода наблюдения от критического момента или интервала времени может привести к снижению достоверности получаемых сведений.

1.3. ЭТАПЫ ПРОВЕДЕНИЯ ИНЖЕНЕРНЫХ ИССЛЕДОВАНИЙ В ЭКОЛОГИИ

В случае моделирования или экспериментального изучения процесса при неполной информации о его механизме обработка полученных в ходе исследования экспериментальных данных проводится в три этапа:

1) моделирование процесса или проведение основного реального эксперимента;

2) проведение дополнительных реальных экспериментов для определения различных величин в математической модели или для получения недостающих данных в реальном эксперименте;

3) обработка полученных экспериментальных данных (определение средних, вычисление отклонений и т.д.);

4) проверка гипотезы о достоверности полученных экспериментальных данных;

5) составление эмпирического уравнения, выражающего функциональную зависимость между факторами и выходными переменными.

Рассмотрим особенности перечисленных этапов.

1.4. Основы моделирования процессов в инженерных исследованиях в экологии

В настоящее время важнейшим средством повышения эффективности инженерных исследований при решении задач защиты окружающей среды является метод математического моделирования.

При наличии полной информации о механизме какого-либо процесса (термодинамике, кинетике, гидродинамике и т.д.) составляют детерминированную математическую модель процесса, представляющую собой систему дифференциальных уравнений в обыкновенных или в частных производных. Для определения неизвестных констант, входящих в систему дифференциальных уравнений, и проверки адекватности математической модели процесса проводится эксперимент.

При неполной информации о механизме процесса проводится функциональное изучение объекта: в ходе эксперимента фиксируют входные и выходные параметры объекта (рис. 1.1).

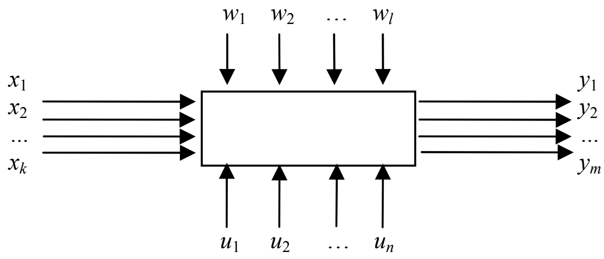

Рис. 1.1. Объект моделирования

Входные (основные) параметры x_1, x_2, …, x_k определяют условия эксперимента и регулируются экспериментатором в процессе проведения эксперимента.

В качестве выходных величин y_1, y_2, …, y_m рассматривают любой технологический или экономический показатель процесса. Используя при обработке опытных данных принципы регрессионного и корреляционного анализа, можно найти зависимость между переменными и определить условия оптимума.

Параметры w_1, w_2, …, w_l – это нерегулируемые внутренние характеристики объекта (например, коррозия металлических частей, засорение трубопроводных систем и т.д.). Также их еще называют "шумом" объекта.

Параметры u_1, u_2, …, u_n – нерегулируемые внешние характеристики окружающей объект среды (например, температура воздуха, давление и т.д.).

Любые нерегулируемые внутренние и внешние характеристики изменяются во времени случайным образом и, следовательно, являются случайными процессами. За время наблюдения случайный процесс принимает тот или иной конкретный вид, заранее неизвестный, называемый реализацией случайного процесса. Случайный процесс можно представить в виде бесконечного множества случайных величин, фиксируя значения случайного процесса через определенные интервалы времени τ_1, τ_2, …, τ_r.

Различают стационарные (рис. 1.2, а) и нестационарные (рис. 1.2, б) случайные процессы.

Стационарные случайные процессы протекают во времени приблизительно однородно и имеют вид случайных колебаний вокруг некоторого среднего значения, причем ни средняя амплитуда, ни характер этих колебаний не обнаруживают существенных изменений с течением времени.

Нестационарные случайные процессы имеют определенную тенденцию развития во времени, характеристики такого процесса зависят от начала отсчета. В данном случае принципиально невозможно получить модель процесса в виде алгебраического уравнения с постоянными коэффициентами (например, в процессе катализа, когда характеристики катализа резко ухудшаются).

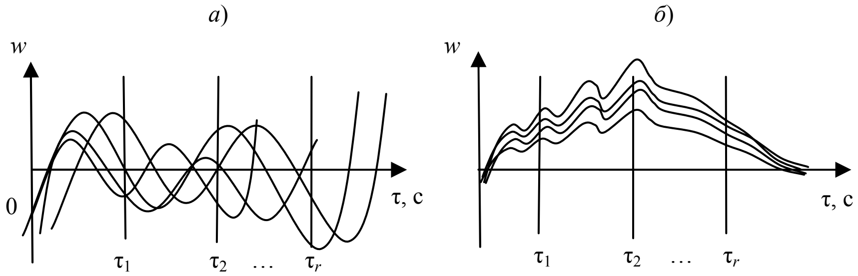

Рис. 1.2. Виды случайных процессов:
а – стационарный; б – нестационарный

Случайным будет любой параметр, не вошедший в число входных параметров, даже если он хорошо изучен. В зависимости от постановки задачи и технических возможностей некоторые измеряемые параметры относят к случайным нерегулируемым, что, однако, уменьшает точность математической модели.

Все изменяемые параметры модели делятся на зависимые и независимые переменные.

К независимым переменным относятся входные (x_1, x_2, …, x_k), нерегулируемые внутренние (w_1, w_2, …, w_l) и нерегулируемые внешние (u_1, u_2, …, u_n) параметры. Независимые переменные принято называть факторами.

К зависимым переменным относятся все входные параметры (y_1, y_2, …, y_m), так как они зависят от остальных, независимых параметров:

$$y_m = f(x_1, x_2, ..., x_m;\, w_1, w_2, ..., w_l;\, u_1, u_2, ..., u_n)\,. \qquad (1.1)$$

Зависимую переменную, характеризующую результаты эксперимента, называют функцией отклика, или целевой функцией, а при решении задач оптимизации – параметром оптимизации:

$$y_m = f(x_1,\ x_2, ...,\ x_m)\ . \qquad (1.2)$$

Принято называть координатное пространство с координатами x_1, x_2, …, x_k факторным пространством, а геометрическое изображение функции отклика в факторном пространстве – поверхностью отклика.

Зависимые и независимые переменные могут быть дискретными и непрерывными. Непрерывные переменные (например, температура, давление, концентрация и т.д.) могут принимать любые значения внутри выбранного интервала. Дискретные переменные принимают только некоторые отдельные значения в заданном интервале.

Совокупность (ряд) значений некоторой величины – это определенное количество измеренных значений одной и той же величины, расположенных затем в определенном или произвольном порядке (например, 3,15; 3,16; 3,14; 3,17; 3,15; 3,14 и т.д.).

Моделирование процессов в экологии проводится для выявления принципов, лежащих в их основе, и для предсказания последствий изменения тех или иных независимых переменных. Принципы (научные выводы и заключения), формулируемые на базе исследования моделей, получили название гипотез – пробных утверждений о наличии или отсутствии тех или иных причинно-следственных связей между определенными процессами и явлениями в экологии.

Истинность или ложность выдвинутой гипотезы проверяется путем сопоставления ее с реальными фактами экологической действительности, выявления ее соответствия (или несоответствия) этим фактам или определение вероятности существования такого соответствия. Способ проверки гипотезы на истинность называется верификацией. Способ проверки гипотезы на ложность называется фальсификацией. Сумма проверенных (истинных, верифицированных) гипотез образует теорию.

1.5. Эксперимент в инженерных исследованиях в экологии

1.5.1. Система эксперимента в инженерных исследованиях в экологии

Как было сказано ранее, исследование – это изучение любых объектов путем проведения над ним эксперимента. При этом **система эксперимента** включает в себя следующие элементы:

1) экспериментальное оборудование – это оборудование, на котором непосредственно проводятся эксперименты (например, лабораторные стенды, испытательные стенды и т.д.);

2) измерительное оборудование – это оборудование, на котором регистрируются значения каких-либо параметров объекта в ходе проведения над ним эксперимента (например, амперметры, вольтметры, хронометры и т.д.);

3) методики планирования, проведения эксперимента и обработки данных эксперимента – это совокупность правил, формул и т.д., применяемых в определенной последовательности (например, методика испытаний на герметичность, методика расчета средних и т.д.);

4) средства отображения результатов эксперимента (например, экраны, записывающие бумажные ленты, перфокарты и т.д.);

5) средства обработки экспериментальных данных (например, калькуляторы, компьютеры и т.д.).

Выделяют следующие **этапы подготовки эксперимента**:

1) выявление параметров объекта, которые нужно менять или фиксировать в ходе эксперимента;

2) определение порядка измеряемых величин (десятки, сотни, тысячи и т.д.);

3) ранжирование (расстановка в определенном порядке) отдельных переменных (изменяемых в ходе эксперимента параметров объекта) и выяснение степени их влияния на процесс;

4) планирование эксперимента.

Функции экспериментатора в системе эксперимента:

– определение исходной информации для поведения эксперимента;
– определение направлений экспериментирования;
– проведение эксперимента;
– внесение изменений в ходе процесса экспериментирования;
– контроль правильности хода процесса экспериментирования;
– контроль достоверности получаемой количественной информации в ходе эксперимента;
– ручная обработка экспериментальной информации;
– разработка программ автоматизированной обработки экспериментальной информации и их реализация.

1.5.2. Методы постановки эксперимента в инженерных исследованиях в экологии

Существует два **метода постановки эксперимента**:

1) традиционный метод постановки эксперимента (метод однофакторного анализа) состоит в изменении одного какого-либо параметра при сохранении постоянными всех других параметров, влияющих на процесс; при такой постановке эксперимента требуется проведение очень большого количества опытов;

2) оптимальный метод постановки эксперимента (метод многофакторного анализа) предполагает одновременное изменение всех параметров, влияющих на процесс, что позволяет сразу установить степень взаимодействия параметров и значительно сократить общее число опытов.

Различают следующие **виды эксперимента**:

1) пассивный эксперимент – включает постановку эксперимента методом однофакторного анализа, а также сбор исходного статистического материала в режиме нормальной эксплуатации на реальном объекте; обработка экспериментальных данных при этом проводится методами классического регрессионного и корреляционного анализа;

2) активный эксперимент – включает постановку эксперимента методом многофакторного анализа по заранее составленному плану (планирование эксперимента); при этом план эксперимента выбирается в зависимости от априорной (однозначной) информации об объекте и от постановки задачи эксперимента, и на каждом этапе исследования выбирается оптимальная стратегия эксперимента.

В настоящее время пассивный эксперимент, несмотря на недостатки (проведение большого количества опытов) широко применяется в реальных условиях, поскольку при этом информацию о свойствах объекта получают без нарушений режима протекания процесса. Активный эксперимент применяется, в основном, в лабораторных и полузаводских условиях.

2. ОБРАБОТКА И ПРОВЕРКА ДОСТОВЕРНОСТИ ПОЛУЧЕННЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

2.1. ОБРАБОТКА ПОЛУЧЕННЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

2.1.1. Среднее значение величины

Величины независимых переменных (например, $x_1, x_2, \ldots, x_k$) определяют некоторое свойство совокупностей. При этом средней величиной будет такое значение x, при замене на которое отдельных значений $x_1, x_2, \ldots, x_k$ это свойство совокупности не изменяется. Из этого следует, что средние значения величин могут определяться различными способами, выбор которых обусловлен связью между усредняемыми величинами и тем свойством, которое они определяют.

Рассмотрим важнейшие средние и их свойства.

1. **Средняя арифметическая.** В общем виде средняя арифметическая взвешенная значений $x_1, x_2, \ldots, x_k$, имеющих веса или частоты $n_1, n_2, \ldots, n_k$, определяются равенством

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n_1 + n_2 + \ldots + n_k} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}. \quad (2.1)$$

Весом или частотой измерения может быть любая величина, характеризующая условие данного измерения, но не номер измерения.

Если вес или частота n_i каждого значения x_i одинаковы (n_i = idem), то $\bar{x}$ является простой средней арифметической:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_k}{k} = \frac{\sum_{i=1}^{k} x_i}{k}. \quad (2.2)$$

Среднюю взвешенную всегда можно представить в виде простой средней арифметической, приняв в уравнении (2.1) $\sum_{i=1}^{k} n_i = n$ и обозначив все n значений x, расположенных в любом порядке, через $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$. При этом получается

$$\bar{x} = \frac{x^{(1)} + x^{(2)} + \ldots + x^{(n)}}{n} = \frac{\sum_{i=1}^{n} x^{(i)}}{n}. \quad (2.3)$$

Из равенства (2.3) можно выразить

$$\sum_{i=1}^{n} x^{(i)} - n\bar{x} = 0, \quad (2.4)$$

откуда

$$\sum_{i=1}^{n} (x^{(i)} - \bar{x}) = 0, \quad (2.5)$$

т.е. сумма отклонений значений x от их средней арифметической равна 0.

Если каждое из отклонений наблюдалось n_i раз, то

$$\sum_{i=1}^{n} n_i (x_i - \bar{x}) = 0. \quad (2.6)$$

Важным свойством средней арифметической является следующая сумма квадратов отклонений значений x от их средней арифметической $\bar{x}$, которая меньше суммы квадратов отклонений их от любой другой величины a, т.е.

$$\sum_{i=1}^{n}(x^{(i)}-\bar{x})^2 < \sum_{i=1}^{n}(x^{(i)}-a)^2, \quad (2.7)$$

или

$$\sum_{i=1}^{n} n_i(x_i-\bar{x})^2 < \sum_{i=1}^{n} n_i(x_i-a)^2. \quad (2.8)$$

В тех случаях, когда наблюдению подвергается постоянная величина, средняя арифметическая является приближенным к истинному значению этой величины.

Если из большой совокупности случайных значений величины x сделать произвольную выборку части этой совокупности (например, из ста значений выбрать двадцать), то средняя арифметическая $\bar{x}$ значений попавших в выборку, приблизительно будет равна средней арифметической всех значений совокупности. Вычисление таким образом средней арифметической позволяет существенно экономить время и силы исследователя.

При наличии очень большой совокупности случайных значений измеряемой величины вычисление средней взвешенной по формуле (2.1) становится весьма громоздкой операцией. В этом случае проще пользоваться равенством

$$\bar{x} = a + \frac{\sum_{i-1}^{n} n_i(x_i-a)}{n}, \quad (2.9)$$

где a – произвольно выбранное число (величину a необходимо брать приблизительно равной средней арифметической, оцениваемой приближенно, "на глаз", без вычислений).

2. **Медиана.** Медианой (Me) называется такое среднее значение, которое делит совокупность значений величин x_i на две равные по количеству членов части, причем в одной из них все значения x_i меньше медианы, а в другой – больше.

Если расположить все члены совокупности в ряд в возрастающем порядке, то при нечетном числе членов, т.е. при

$$i = 2m + 1, \quad (2.10)$$

медианой будет значение среднего члена ряда, т.е.

$$Me = x_{m+1}. \quad (2.11)$$

Если число членов ряда четное, т.е.

$$i = 2m, \quad (2.12)$$

то за медиану принимается среднее арифметическое двух значений x_m и x_{m+1}, находящихся в середине ряда, т.е.

$$Me = \frac{x_m + x_{m+1}}{2}. \quad (2.13)$$

Если в данном ряде члены ряда, достаточно удаленные от медианы, подвергаются малым изменениям, то медиана при этом не меняется, в то время как средняя арифметическая изменится. Поэтому, если, как это часто бывает, значения x_i, находящиеся на концах ряда, не точны, то в качестве средней величины лучше пользоваться медианой.

3. **Мода.** Модой (Мо) называется наиболее вероятное значение случайной величины, либо то значение случайной величины, частота появления которого наибольшая.

Мода применяется для характеристики наиболее часто встречающихся значений в совокупности случайных величин.

4. **Средняя логарифмическая.** Многие естественные процессы, протекающие без участия человека, подчиняются логарифмическому закону.
В этих случаях кривая распределения величин имеет логарифмический характер, например:

$$x = x_0 e^{-kr}, \quad (2.14)$$

где k – некоторая константа (например, константа скорости реакции); r – время наблюдения, с; x_0 – начальное значение величины x.

Графически это можно изобразить, как показано на рис. 2.1.

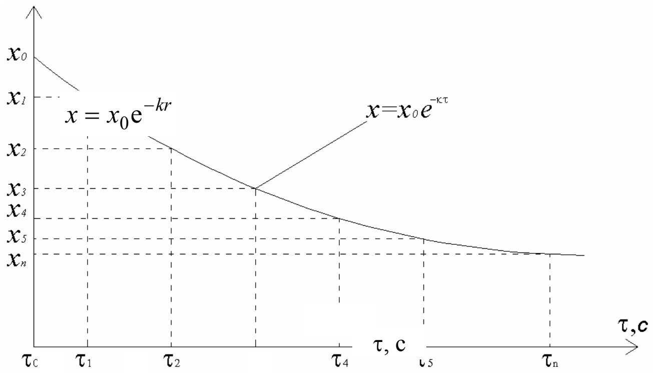

Рис. 2.1. Процесс, происходящий по логарифмическому закону

Тогда средняя логарифмическая двух величин x_1 и x_n есть отношение их разности к разности их натуральных логарифмов:

$$x_{\text{ср. лог}} = \frac{x_1 - x_n}{\ln x_1 - \ln x_n} = \frac{x_1 - x_n}{\ln \frac{x_1}{x_n}} = \frac{x_1 - x_n}{2{,}3 \ln \frac{x_1}{x_n}} . \quad (2.15)$$

Когда значения двух величин x_1 и x_n мало отличаются друг от друга (на практике, при $\frac{x_1}{x_n} < 2$), то средняя логарифмическая без большей погрешности (менее 4,4 %) может быть заменена средней арифметической, причем ошибка тем меньше, чем меньше разница между x_1 и x_n (т.е., при $\frac{x_1}{x_n} \to 1$). Средняя логарифмическая всегда меньше средней арифметической.

5. **Средняя квадратическая.** Средней квадратической n положительных или отрицательных величин x_1, x_2, …, x_n называется положительное значение квадратного корня и суммы квадратов из суммы квадратов этих величин, деленной на их число n:

$$x_{\text{ср. кв}} = +\sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} . \quad (2.16)$$

Если значения x_1, x_2, …, x_n имеют веса или частоты β, то определяется средняя квадратическая взвешенная:

$$x_{\text{ср. кв}} = +\sqrt{\frac{x_1^2\beta_1 + x_2^2\beta_2 + \ldots + x_n^2\beta_n}{\beta_1 + \beta_2 + \ldots + \beta_n}} . \quad (2.17)$$

6. **Средняя геометрическая (средняя пропорциональная).** Средней геометрической n положительных величин x_1, x_2, …, x_n называется положительное значение корня n-й степени из их произведения:

$$x_{\text{ср. геом}} = +\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} , \quad (2.18а)$$

или, после логарифмирования формулы (2.18а),

$$\lg x_{\text{ср. геом}} = \frac{\lg x_1 + \lg x_2 + \ldots + \lg x_n}{n} . \quad (2.18б)$$

Средняя геометрическая двух положительных неравных величин всегда меньше их средней арифметической.

7. **Средняя гармоническая.** Средней гармонической n положительных величин x_1, x_2, …, x_n называется величина H, обратное значение которой равно среднему арифметическому обратных значений величин x_1, x_2, …, x_n, т.е.

$$\frac{1}{H} = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}{n} , \quad (2.19а)$$

откуда

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} . \quad (2.19б)$$

Средняя гармоническая используется в тех случаях, когда приходится иметь дело с величиной, зависящей от обратных значений частных величин.

Если значения $x_1, x_2, \ldots, x_n$ имеют веса или частоты β, то определяется средняя гармоническая взвешенная:

$$H = \frac{\beta_1 + \beta_2 + \ldots + \beta_n}{\dfrac{\beta_1}{x_1} + \dfrac{\beta_2}{x_2} + \ldots + \dfrac{\beta_n}{x_n}}. \qquad (2.20)$$

8. **Средняя хронологическая.** Простая средняя хронологическая определяется в случаях разброса исследуемых значений в большом периоде времени и является частным случаем простой средней арифметической. Расчет по формуле средней хронологической ведется следующим образом:

$$x_{\text{ср. хрон}} = \frac{\dfrac{x_1^{\text{н}} + x_1^{\text{к}}}{2} + \dfrac{x_2^{\text{н}} + x_2^{\text{к}}}{2} + \ldots + \dfrac{x_n^{\text{н}} + x_n^{\text{к}}}{2}}{n}, \qquad (2.21а)$$

где $x_1^{\text{н}}$ – значение переменной на начало 1-го промежутка времени (например, первого месяца в году); $x_1^{\text{к}}$ – значение переменной на конец 1-го промежутка времени (например, первого месяца в году); $x_2^{\text{н}}$ – значение переменной на начало 2-го промежутка времени (например, второго месяца в году); $x_2^{\text{к}}$ – значение переменной на конец 2-го промежутка времени (например, второго месяца в году); $x_n^{\text{н}}$ – значение переменной на начало n-го промежутка времени (например, последнего месяца в году); $x_n^{\text{к}}$ – значение переменной на конец n-го промежутка времени (например, последнего месяца в году); n – число промежутков времени во временном периоде (например, число месяцев в году).

Средние хронологические взвешенные значений $x_1, x_2, \ldots, x_n$, имеющих веса или частоты в разные промежутки времени, определяются равенствами:

$$x_{\text{ср. хрон1}} = \frac{\dfrac{x_1^{\text{н}} + x_1^{\text{к}}}{2} a_1 + \dfrac{x_2^{\text{н}} + x_2^{\text{к}}}{2} a_2 + \ldots + \dfrac{x_n^{\text{н}} + x_n^{\text{к}}}{2} a_n}{a_1 + a_2 + \ldots + a_n}, \qquad (2.21б)$$

$$x_{\text{ср. хрон2}} = \frac{\dfrac{x_1^{\text{н}}\beta_1^{\text{н}} + x_1^{\text{к}}\beta_1^{\text{к}}}{\beta_1^{\text{н}} + \beta_1^{\text{к}}} + \dfrac{x_2^{\text{н}}\beta_2^{\text{н}} + x_2^{\text{к}}\beta_2^{\text{к}}}{\beta_2^{\text{н}} + \beta_2^{\text{к}}} + \ldots + \dfrac{x_n^{\text{н}}\beta_n^{\text{н}} + x_n^{\text{к}}\beta_n^{\text{к}}}{\beta_n^{\text{н}} + \beta_n^{\text{к}}}}{n}, \qquad (2.21в)$$

$$x_{\text{ср. хрон3}} = \frac{\dfrac{x_1^{\text{н}}\beta_1^{\text{н}} + x_1^{\text{к}}\beta_1^{\text{к}}}{\beta_1^{\text{н}} + \beta_1^{\text{к}}} a_1 + \dfrac{x_2^{\text{н}}\beta_2^{\text{н}} + x_2^{\text{к}}\beta_2^{\text{к}}}{\beta_2^{\text{н}} + \beta_2^{\text{к}}} a_2 + \ldots + \dfrac{x_n^{\text{н}}\beta_n^{\text{н}} + x_n^{\text{к}}\beta_n^{\text{к}}}{\beta_n^{\text{н}} + \beta_n^{\text{к}}} a_n}{a_1 + a_2 + \ldots + a_n}, \qquad (2.21г)$$

где a_1 – вес или частота измерения переменной в 1-м промежутке времени (например, в первом месяце в году); a_2 – вес или частота измерения переменной в 2-м промежутке времени (например, во втором месяце в году); a_n – вес или частота измерения переменной в n-м промежутке времени (например, в последнем месяце в году); $\beta_1^{\text{н}}$ – вес или частота измерения переменной на начало 1-го промежутка времени (например, первого месяца в году); $\beta_1^{\text{к}}$ – вес или частота измерения переменной на конец 1-го промежутка времени (например, первого месяца в году); $\beta_2^{\text{н}}$ – вес или частота измерения переменной на начало 2-го промежутка времени (например, второго месяца в году); $\beta_2^{\text{к}}$ – вес или частота измерения переменной на конец 2-го промежутка времени (например, второго месяца в году); $\beta_n^{\text{н}}$ – вес или частота измерения переменной на начало n-го промежутка времени (например, последнего месяца в году); $\beta_n^{\text{к}}$ – вес или частота измерения переменной на конец n-го промежутка времени (например, последнего месяца в году).

Средняя хронологическая динамическая определяется для значений переменных, часть из которых измерена в начале периода времени, а другие, измеренные впоследствии в различные промежутки периода времени, имеют и положительные, и отрицательные значения:

$$x_{\text{ср. хрон}} = x_{\text{перв}}^{\text{н}} + \sum_{i=1}^{n} \frac{i}{n} x_i^n - \sum_{j=1}^{m} \frac{j}{m} \left| x_j^m \right| , \qquad (2.22)$$

где $x_{\text{перв}}^{\text{н}}$ – значение переменной, измеренное в начале всего периода времени (например, в начале первого месяца года); x_i^n – положительные значения переменных, измеренных в различные промежутки периода времени (например, в различные месяцы года); i – номера полных промежутков времени (например, месяцев), в которых переменная x имела положительные значения; n – число полных промежутков времени (например, месяцев) с момента измерения, в которых переменная x имела положительные значения; $\left|x_j^m\right|$ – модули отрицательных значений переменных, измеренных в различные промежутки периода времени (например, в различные месяцы года); j – номера полных промежутков времени (например, месяцев), в которых переменная x имела отрицательные значения; m – число полных промежутков времени (например, месяцев) с момента измерения, в которых переменная x имела отрицательные значения.

Средние хронологические используются особенно часто в технико-экономических расчетах процессов защиты окружающей среды, например, при расчете средней стоимости введенных в действие и выбывших в течение большого промежутка времени сооружений очистки воздуха и воды на предприятии.

2.1.2. Математическое ожидание дискретной случайной величины при конечном числе ее значений

Рассмотрим вопрос о значениях, которые может принимать некоторая величина x в зависимости от случайных, т.е. не поддающихся учету, причин. При этом каждое значение x_i, полученного в результате единичного испытания, является случайной величиной, вероятность появления которой p_i.

Зависимость между значением случайной величины x_i и ее вероятностью p_i называется распределением этой величины.

Допустим, что при очень большом количестве n испытаний дискретная случайная величина x принимает конечное число значений $x_1, x_2, \dots, x_n$ соответственно $m_1, m_2, \dots, m_n$ раз. Тогда среднее значение $\bar{x}$ равно

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_n x_n}{m_1 + m_2 + \dots + m_n} = \frac{m_1 x_1 + m_2 x_2 + \dots + m_n x_n}{n} =$$

$$= \frac{m_1}{n} x_1 + \frac{m_2}{n} x_2 + \dots + \frac{m_n}{n} x_n . \qquad (2.23)$$

Когда n велико, относительные частоты $\frac{m_1}{n}, \frac{m_2}{n}, \dots, \frac{m_n}{n}$ приблизительно равны вероятностям $p_1, p_2, \dots, p_n$ появления значений $x_1, x_2, \dots, x_n$. Поэтому при большем числе испытаний среднее значение $\bar{x}$ можно определить как

$$\bar{x} \approx p_1 x_1 + p_2 x_2 + \dots + p_n x_n = \sum_{i=1}^{n} p_i x_i \; . \qquad (2.24)$$

Здесь величина $\sum_{i=1}^{n} p_i x_i$ называется вероятным значением случайной величины x или ее математическим ожиданием M(x):

$$M(x) = \sum_{i=1}^{n} p_i x_i \approx \bar{x} \; . \qquad (2.25)$$

Таким образом, математическое ожидание M(x) является теоретической величиной, около которой колеблются средние значения $\bar{x}$ случайной величины x при большем числе испытаний n.

Основные свойства математического ожидания (при A = const):

1) математическое ожидание постоянной величины равно самой постоянной величине:

$$M(A) = A; \qquad (2.26)$$

2) математическое ожидание произведения случайной величины на постоянный множитель равно произведению математического ожидания случайной величины на этот множитель:

$$M(A\,x) = A\;M(x); \qquad (2.27)$$

3) математическое ожидание суммы случайных величин равно сумме их математических ожиданий:

$$M(x + y + z) = M(x) + M(y) + M(z); \qquad (2.28)$$

4) математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий:

$$M(xyz) = M(x)\,M(y)\,M(z). \qquad (2.29)$$

Математическое ожидание квадрата отклонения случайной величины от ее математического ожидания M(x) называется дисперсией σ^2 случайной величины x:

$$\sigma^2 = M(x - M(x))^2 = M(x^2) - M^2(x). \tag{2.30}$$

Дисперсия является мерой рассеяния значений x около их математического ожидания.

Если появление некоторого события в каждом испытании имеет вероятность p, то математическое ожидание частоты m этого события при n испытаниях равно

$$M(m) = np. \tag{2.31}$$

Из формул (2.30) и (2.31) следует, что дисперсия частоты m появления случайной величины равна

$$\sigma_m^2 = np\,(1 - p), \tag{2.32}$$

а для редких событий, когда n велико, а p очень мало, ее значение определится по формуле

$$\sigma_m^2 = np. \tag{2.33}$$

Для бесповторной выборки

$$\sigma_m^2 = n_1 p\,(1-p)\frac{n-n_1}{n}, \tag{2.34}$$

где n_1 – количество испытаний, попавших в выборку.

Дисперсия суммы независимых случайных величин равна сумме их дисперсий:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2. \tag{2.35}$$

Положительное значение квадратного корня из дисперсии называется средним квадратическим отклонением или стандартом:

$$\sigma = +\sqrt{M(x^2) - M^2(x)}. \tag{2.36}$$

2.1.3. Математическое ожидание непрерывной случайной величины при бесконечном числе ее значений

Обозначим через X некоторую непрерывную случайную величину, которая может принимать любые числовые значения из промежутка [a; b]. Пусть x есть некоторое число из этого промежутка. Вероятность dp того, что величина X принимает значения, заключенные между x и x + dx пропорциональна dx (при бесконечно малом dx) и зависит от x, т.е.

$$dp = \varphi(x)\,dx. \tag{2.37}$$

Здесь функция $\varphi(x)$ называется плотностью распределения вероятностей случайной величины X, а произведение $\varphi(x)\,dx$ – элементом вероятности.

Кривая $y = \varphi(x)$ называется кривой распределения вероятностей данной случайной величины X (рис. 2.2).

Вероятность того, что случайная величина X примет значения в промежутке $(x_1; x_2)$, равна площади, ограниченной кривой абсцисс и двумя ординатами, проведенных в точках $x = x_1$ и $x = x_2$, т.е.

$$p\,(x_1 < X < x_2) = \int_{x_1}^{x_2} \varphi(x)\,dx. \tag{2.38}$$

Следовательно, при разделении отрезка [a; b] на элементарные отрезки длиной Δx, вероятность того, что случайная величина x примет какие-либо значения из отрезка $[x; x + \Delta x]$, равна

$$p(x) = \varphi(x)\,\Delta x, \tag{2.39}$$

а математическое ожидание случайной величины (с учетом формулы (2.25)) будет равно

$$M(x) = x\,p(x) = \int_a^b x\,\varphi(x)\,dx = \lim_{\Delta x \to 0} \sum_x x\,\varphi(x)\,\Delta x \approx \sum_x x\,\varphi(x)\,\Delta x\,. \quad (2.40)$$

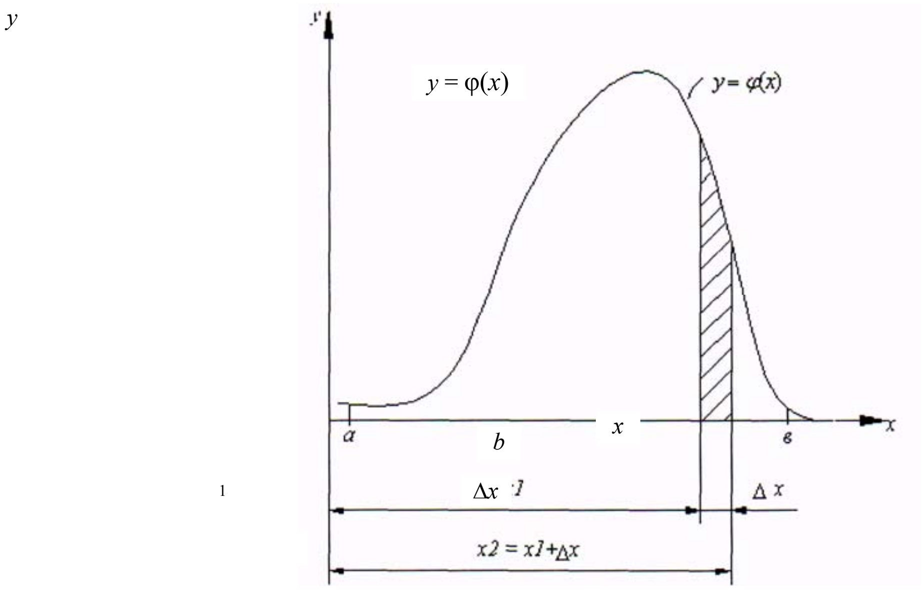

Рис. 2.2. Вид кривой распределения вероятностей данной случайной величины

2.1.4. Нормальное распределение

Нормальный закон распределения имеет чрезвычайно широкое распространение в процессах защиты окружающей среды.

Если плотность распределения вероятностей случайной величины имеет вид

$$\varphi(x) = \frac{h}{\sqrt{\pi}}\, e^{-h^2 (x - x_i)^2}\,, \quad (2.41)$$

то говорят, что случайная величина распределена нормально.

Кривую распределения φ(x) называют кривой нормального распределения или кривой Гаусса (рис. 2.3).

Математическое ожидание случайной величины, подчиняющийся нормальному закону распределения, определится по формуле

$$M(x) = x_i. \quad (2.42)$$

Дисперсия случайной величины, подчиняющейся нормальному закону распределения, будет равна

$$\sigma^2 = \frac{1}{2h^2}\,, \quad (2.43)$$

где h – мера точности (см. следующий параграф).

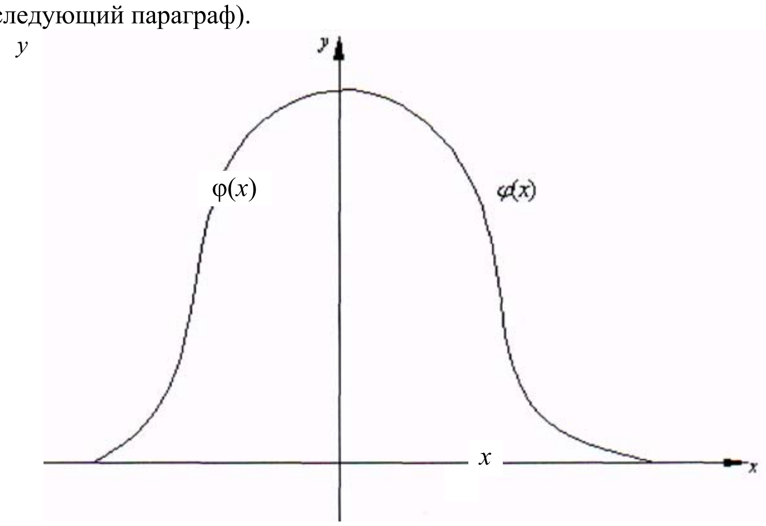

Рис. 2.3. Кривая нормального распределения

Ее среднее квадратическое отклонение определится как

$$\sigma = \frac{1}{h\sqrt{2}}\,. \tag{2.44}$$

С учетом формул (2.42) – (2.44) уравнение (2.41) кривой нормального распределения может быть записано в виде

$$y = \frac{1}{\delta\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-x_i)^2}{2\sigma^2}}\,. \tag{2.45}$$

2.1.5. Закон распределения ошибок

Результаты экспериментального измерения величин никогда не бывают полностью точными, а всегда имеют некоторые погрешности или ошибки.

Ошибки, возникающие по причинам, которые поддаются учету и устранению, называются систематическими.

Ошибки, имеющие место в результате большого количества случайных, не поддающихся учету причин, называются случайными ошибками.

Случайные ошибки подчиняются нормальному закону распределения.

$$\varphi(x) = \frac{1}{\delta\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-x_i)^2}{2\sigma^2}} = \frac{h}{\sqrt{\pi}}\,\mathrm{e}^{-h^2(x-x_i)^2}\,. \tag{2.46}$$

При достаточно больших значениях |x| функция (2.46) равна нулю.

Параметр h в формулах (2.41), (2.43), (2.44), (2.46) характеризует точность измерений, так как от него зависит характер группировки ошибок вблизи нуля. Например, вероятность появления ошибки при h = 2 вдвое, а при h = 3 втрое больше, чем при h = 1. По этой причине параметр h называется мерой точности.

Меру точности можно оценить по формуле

$$h = \sqrt{\frac{n}{2\sum_{i=1}^{n}(x-x_i)^2}} = \sqrt{\frac{n-1}{2\sum_{i=1}^{n}(\bar{x}-x_i)^2}} = \frac{1}{\sigma\sqrt{2}}\,, \tag{2.47}$$

где x – истинное значение измеряемой случайной величины или ошибки (обычно неизвестное, но заменяемое, как правило, какой-либо средней $\bar{x}$).

Из формулы (2.47) можно получить:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x}-x_i)^2}{n-1}}\,, \tag{2.48}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(\bar{x}-x_i)^2}{n-1}\,. \tag{2.49}$$

Формула (2.47) определяет то значение меры точности h, при котором вероятность получения данной системы случайных ошибок будет наибольшей.

Вероятность того, что ошибки отдельных наблюдений не превосходят по абсолютному значению заданной величины r, т.е. заключается в пределах от – r до + r, равна

$$p(|x-\bar{x}| < r) = \Phi(hr) = \Phi\left(\frac{r}{\sigma\sqrt{2}}\right), \tag{2.50}$$

где Φ – функция Лапласа.

Если в формуле (2.50) принять r = 3σ, то

$$p(|x-\bar{x}| < 3\sigma) = \Phi\left(\frac{3\sigma}{\sigma\sqrt{2}}\right) = \Phi(2{,}13) = 0{,}997\,, \tag{2.51}$$

т.е. вероятность того, что случайная ошибка по абсолютной величине будет меньше числа r = 3σ, равна 99,7 %.

Вероятность того, что случайная ошибка по абсолютной величине превзойдет число r = 3 σ, будет равна

$$p(|x-\bar{x}|>3\sigma)=1-0{,}997=0{,}003\,. \qquad (2.52)$$

Эта вероятность (формула (2.52)) ничтожно мала. Следовательно, практически все ошибки измерения заключены между значениями – 3σ и + 3σ. Данное правило называется правилом трех сигм.

Однако, следует учитывать, что ошибки, большие по абсолютному значению трех сигм, возможны, хотя и встречаются крайне редко – в среднем в трех случаях на тысячу. На этом основании число

$$\Delta = 3\sigma \qquad (2.53)$$

называют наибольшей возможной ошибкой.

Вероятной ошибкой измерений ρ называется такая величина, для которой с одинаковой вероятностью можно ожидать, что действительные ошибки по абсолютной величине окажутся как меньше ее, так и больше. Иными словами, при большом числе наблюдений, приблизительно половина отклонений $|x - x_i|$ окажется меньше ρ, а другая половина отклонений – больше ρ.

В соответствии с этим определением можно в формуле (2.50) принять $p=\frac{1}{2}$ и $r=\rho$, и после преобразований найти по специальным таблицам значение полученной функции

$$h\,\rho = 0{,}477, \qquad (2.54)$$

откуда

$$\rho=\frac{0{,}477}{h}\,. \qquad (2.55)$$

Подставляя (2.47) в (2.55), получим

$$\rho = 0{,}477\sqrt{\frac{2\sum_{i=1}^{n}\left(\bar{x}-x_i\right)^2}{n-1}} = 0{,}675\sqrt{\frac{\sum_{i=1}^{n}\left(\bar{x}-x_i\right)^2}{n-1}} = 0{,}675\sigma\,. \qquad (2.56)$$

2.1.6. Наивероятнейшее значение измеряемой величины и ее точность

Пусть при измерении некоторой величины, неизвестное истинное значение которой есть x, получен следующий ряд значений:

$$x_1,\ x_2,\ \ldots,\ x_n, \qquad (2.57)$$

тогда ошибки этих значений соответственно равны

$$x - x_1,\ x - x_2,\ \ldots,\ x - x_n, \qquad (2.58)$$

Вероятность того, что при измерении сделана одна ошибка $x - x_i$, равна (по формулам (2.38), (2.46))

$$p(x)=\int_{x-x_i}^{x-x_i+dx}\frac{h}{\sqrt{\pi}}\,e^{-h^2(x-x_i)^2}\,dx\,. \qquad (2.59)$$

Соответственно, вероятность того, что при n измерениях будут сделаны ошибки $x - x_1$, $x - x_2$, …, $x - x_n$, определится формулой

$$p(x_n)=\int_{x-x_n}^{x-x_n+dx}\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2(x-x_1)^2}\ e^{-h^2(x-x_2)^2}\ e^{-h^2(x-x_n)^2}\ (dx)^n =$$

$$=\int_{x-x_n}^{x-x_n+dx}\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2\sum_{i=1}^{n}(x-x_i)^2}\ (dx)^n\,. \qquad (2.60)$$

Наивероятнейшее значение измеряемой величины x определится при max $(p(x_n))$, т.е. в том случае, если показатель степени у числа "e" в формуле (2.60) будет иметь минимум.

Для нахождения значения $\min\sum_{i=1}^{n}(x-x_i)^2$ необходимо приравнять нулю первую производную данного выражения

$$2\sum_{i=1}^{n}\left(x - x_i\right) = 0\,, \tag{2.61}$$

откуда

$$x = \frac{x_1 + x_2 + \ldots + x_n}{n}\,, \tag{2.62}$$

т.е. наивероятнейшим значением измеряемой величины является простая средняя арифметическая (без учета частот или весов) или средняя арифметическая взвешенная (с учетом весов).

Поскольку результаты измерений x_1, x_2, …, x_n представляют собой случайные величины, то их средняя арифметическая $\bar{x}$ является также случайной величиной. Эта случайная величина распределена нормально.

Пусть h есть мера точности отдельного измерения. Обозначим через H меру точности средней арифметической. Тогда можно показать, что:

$$H^2 = nh^2, \tag{2.63}$$

$$H = h\sqrt{n}\,. \tag{2.64}$$

Подставляя (2.64) в (2.47) получим

$$H = \sqrt{\frac{n(n-1)}{2\sum_{i=1}^{n}\left(\bar{x} - x_i\right)}} > h\,. \tag{2.65}$$

Отсюда следует, что если, например, число измерений увеличить в четыре раза, то точность средней арифметической увеличится вдвое.

Вероятность того, что средняя арифметическая отличается от истинного значения на величину, меньшую r, определится по формуле (2.50):

$$p\left(\left| x - \bar{x} \right| < r\right) = \Phi(Hr) = \Phi\left(\frac{r\sqrt{n}}{\sigma\sqrt{2}}\right), \tag{2.66}$$

Средняя квадратическая ошибка σ_0 средней арифметической определится по формуле (2.48):

$$\sigma_0 = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x} - x_i)^2}{n\cdot(n-1)}} = \frac{\sigma}{\sqrt{n}}\,, \tag{2.67}$$

где σ – средняя квадратическая ошибка отдельного измерения.

Наибольшая возможная ошибка средней арифметической определится по формуле (2.53):

$$\Delta_0 = 3\,\sigma_0. \tag{2.68}$$

Вероятная ошибка средней арифметической определится по формуле (2.56):

$$\rho_0 = 0{,}675\,\sigma_0. \tag{2.69}$$

При записи средней арифметической $\bar{x}$ принято указывать ее среднюю квадратическую ошибку $\pm\,\sigma_0$.

2.1.7. Равноточные и неравноточные наблюдения

Пусть даны два ряда измерений, причем все измерения и первого, и второго ряда произведены с одинаковой точностью, т.е. измерения равноточны. Если каждый ряд содержит одинаковое число измерений, то результаты обработки рядов будут равноточны. Если же число измерений в рядах неодинаковое, то результаты обработки рядов будут неравноточны. Это следует, в том числе, и из формулы (2.67), куда входит число измерений n.

Пусть даны два ряда измерений, но измерения каждого ряда произведены приборами разной точности. Тогда результаты обработки рядов по предыдущему разделу 2.7 будут неравноточны, даже в том случае, если бы число измерений в каждом ряду было бы одинаковым. Это следует из формулы (2.67), куда входит число σ.

Сущность обработки неравноточных рядов заключается в том, что после введения некоторых коэффициентов, являющихся весами, обработку неравноточных рядов производят так же, как и равноточных по предыдущему разделу 2.7.

Допустим, что значение какого-либо параметра определялось четырьмя группами измерений, что показано в табл. 2.1, причем для каждой группы измерений определены простые арифметические средние по формуле (2.2).

2.1. Обработка неравноточных рядов измерений

Номер группы, j / Номер измерения, i	1	2	3	4
1	107,10	107,57	107,51	107,42
2	107,68	107,45	107,57	107,00
3	107,45	107,07	107,16	–
4	107,62	107,35	107,48	–
5	107,68	107,17	–	–
6	107,08	107,46	–	–
7	107,44	–	–	–
8	107,28	–	–	–
Количество измерений в каждой группе, n_j	8	6	4	2
Сумма величин измерений в каждой группе, $\sum_{i=1}^{n_j} x_{i,j}$	859,33	644,07	429,72	214,42
Простая средняя арифметическая в каждой группе, $\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{i,j}}{n_j}$	107,416	107,345	107,430	107,210

В данном случае результаты обработки измерений (средние арифметические $\bar{x}_j$) будут неравноточны для каждой группы. Здесь возможны следующие случаи:

1. Так как все измерения в табл. 2.1 равноточны, то для нахождения общей простой арифметической средней нужно сложить данные всех n измерений и сумму разделить на n, т.е.

$$\bar{x} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{i,j}}{n}, \qquad (2.70)$$

где общее количество измерений по всем группам составляет

$$n = \sum_{j=1}^{k} n_j . \qquad (2.71)$$

Здесь k – число групп измерений.

Подставляя формулу (2.71) в формулу (2.70), получим

$$\bar{x} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{i,j}}{\sum_{j=1}^{k} n_j}. \qquad (2.72)$$

Подставляя значения табл. 2.1 в формулу (2.72), получим

$$\bar{x} = \frac{107{,}10+107{,}68+107{,}45+107{,}62+107{,}68}{8+6+4+2} +$$

$$+\frac{107{,}08+107{,}44+107{,}28+107{,}57+107{,}45+107{,}07}{8+6+4+2} +$$

$$+\frac{107{,}35+107{,}17+107{,}46+107{,}51+107{,}57}{8+6+4+2} +$$

$$+\frac{107{,}16+107{,}48+107{,}42+107{,}00}{8+6+4+2} = \frac{2147{,}54}{20} = 107{,}377 .$$

2. Предположим теперь, что задаются только простые арифметические средние каждой группы $\bar{x}_j$ и число измерений в ней n_j. В этом случае обработку результатов следует проводить путем введения весов измерений.

Весом измерений в данном случае можно считать количество измерений в каждой группе n_j, так как вес – это степень доверия к результатам наблюдения, а эта степень, очевидно, тем больше, чем больше измерений в группе.

Таким образом, общая средняя арифметическая, называемая в данном случае общей средней арифметической взвешенной или взвешенной средней, будет определяться по формуле

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k} = \frac{\sum_{j=1}^{k} n_j\bar{x}_j}{\sum_{j=1}^{k} n_j}. \quad (2.73).$$

Подставляя значения табл. 2.1 в формулу (2.73), получим

$$\bar{x} = \frac{8\cdot 107{,}416 + 6\cdot 107{,}345 + 4\cdot 107{,}430 + 2\cdot 107{,}210}{8+6+4+2} = 107{,}377.$$

3. Рассмотрим случай, когда число измерений неизвестно, но заданы средние квадратические ошибки результатов измерений. Тогда из формулы (2.67) можно определить число измерений

$$n = \frac{\sigma_{i,j}^2}{\sigma_{0j}^2}. \quad (2.74).$$

Если отдельные измерения обладают одинаковой точностью ($\sigma_{i,j} = \mathrm{idem}$), то средним арифметическим отдельных групп измерений $\bar{x}_j$ следует приписывать веса γ_j, обратно пропорциональные квадратам их средних квадратических ошибок σ_{0j}^2. Так как средняя квадратическая ошибка σ_0, вероятная ошибка ρ_0 и наибольшая возможная ошибка Δ_0 прямо пропорциональны друг другу, то в качестве весов γ_j средних арифметических $\bar{x}_j$ можно взять числа, обратно пропорциональные квадратам любых этих ошибок:

$$\gamma_1 : \gamma_2 : \ldots : \gamma_j = \frac{1}{\sigma_{01}^2} : \frac{1}{\sigma_{02}^2} : \ldots : \frac{1}{\sigma_{0j}^2} = \frac{1}{\rho_{01}^2} : \frac{1}{\rho_{02}^2} : \ldots : \frac{1}{\rho_{0j}^2} = \frac{1}{\Delta_{01}^2} : \frac{1}{\Delta_{02}^2} : \ldots : \frac{1}{\Delta_{0j}^2}, \quad (2.75)$$

откуда можно, например, принять

$$\gamma_1 = \frac{1}{\sigma_{01}^2}; \quad \gamma_2 = \frac{1}{\sigma_{02}^2}; \quad \ldots; \quad \gamma_j = \frac{1}{\sigma_{0j}^2}, \quad (2.76)$$

или, например

$$\gamma_1 = \frac{\sigma_{01}^2}{\sigma_{01}^2} = 1; \quad \gamma_2 = \frac{\sigma_{01}^2}{\sigma_{02}^2}; \quad \ldots; \quad \gamma_j = \frac{\sigma_{01}^2}{\sigma_{0j}^2}. \quad (2.77)$$

2.1.8. Среднее значение и дисперсия функции нескольких независимых случайных величин

Предположим, что $x_1, x_2, \ldots, x_n$ – независимые случайные величины, средние значения которых соответственно равны $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$, а дисперсии этих средних значений равны $\sigma_{01}^2, \sigma_{02}^2, \sigma_{0n}^2$. Пусть

$$z = k_1x_1 + k_2x_2 + \ldots + k_nx_n = f(x_1, x_2, \ldots, x_n) \quad (2.78)$$

– некоторая линейная функция этих величин, которая также будет некоторой случайной величиной.

Тогда среднее значение данной функции будет равно

$$\bar{z} = k_1\bar{x}_1 + k_2\bar{x}_2 + \ldots + k_n\bar{x}_n = f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n), \quad (2.79)$$

а дисперсия величины $\bar{z}$ определится по формуле:

$$\sigma_z^2 = k_1^2\sigma_{01}^2 + k_2^2\sigma_{02}^2 + \ldots + k_n^2\sigma_{0n}^2 = f(\sigma_{01}^2, \sigma_{02}^2, \ldots, \sigma_{0n}^2)\,. \quad (2.80)$$

Существуют следующие частные случаи формул (2.79) и (2.80).

1. Пусть

$$z = x_1 + x_2 + \ldots + x_n\,. \quad (2.81)$$

Тогда

$$\bar{z} = \bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_n\,, \quad (2.82)$$

$$\sigma_z^2 = \sigma_{01}^2 + \sigma_{02}^2 + \ldots + \sigma_{0n}^2\,. \quad (2.83)$$

2. Если все величины $x_1, x_2, \ldots, x_n$ обладают одной и той же дисперсией σ^2, то дисперсия их суммы (формула (2.81)) будет равна

$$\sigma_z^2 = n\sigma^2\,, \quad (2.84)$$

а среднее квадратическое отклонение составит

$$\sigma_z = \sigma\sqrt{n}\,. \quad (2.85)$$

3. Пусть z есть средняя арифметическая n случайных величин $x_1, x_2, \ldots, x_n$:

$$z = \frac{x_1 + x_2 + \ldots + x_n}{n}\,. \quad (2.86)$$

Тогда

$$\sigma_z^2 = \frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}{n}\,, \quad (2.87)$$

$$\sigma_z = \frac{1}{n}\sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}\,. \quad (2.88)$$

4. Если $\sigma_1 = \sigma_2 = \ldots = \sigma_n = \sigma,$ то

$$\sigma_z^2 = \frac{\sigma^2}{n}\,, \quad (2.89)$$

$$\sigma_z = \frac{\sigma}{\sqrt{n}}\,. \quad (2.90)$$

Из формулы (2.90) в том числе следует, что если $x_1, x_2, \ldots, x_n$ – результаты измерений какой-либо величины, то точность средней арифметической в $\sqrt{n}$ раз больше точности отдельных измерений.

Предположим теперь, что z есть нелинейная функция нескольких случайных величин, например

$$z = \varphi(x, y)\,. \quad (2.91)$$

Пусть $\bar{x}, \bar{y}$ – средние значения величин x и y, соответственно, а σ_{01}^2 и σ_{02}^2 – дисперсии этих средних значений.

Тогда можно разложить функцию (2.91) в ряд Тейлора по степеням x – a и y – b (где a и b – некоторые числа):

$$z = \varphi(x,y) = \varphi(a,b) + \frac{1}{1!}\left[(x-a)\frac{\partial\varphi}{\partial x} + (y-b)\frac{\partial\varphi}{\partial y}\right] +$$

$$+ \frac{1}{2!}\left[(x-a)^2\frac{\partial^2\varphi}{\partial x^2} + 2(x-a)(y-b)\frac{\partial^2\varphi}{\partial x\partial y} + (y-b)^2\frac{\partial^2\varphi}{\partial y^2}\right] +$$

$$+ \ldots + \frac{1}{n!}\left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]^n\varphi + \ldots\,. \quad (2.92)$$

Примем в формуле (2.92) $a = \bar{x}$, $b = \bar{y}$ и ограничимся первыми двумя членами этого ряда. Тогда

$$z = \varphi(x, y) = \varphi(\bar{x}, \bar{y}) + (x - \bar{x})\frac{\partial\varphi}{\partial\bar{x}} + (y - \bar{y})\frac{\partial\varphi}{\partial\bar{y}}, \tag{2.93}$$

где $\frac{\partial\varphi}{\partial\bar{x}}, \frac{\partial\varphi}{\partial\bar{y}}$ – частные производные функции φ, вычисленные при $x = \bar{x}$ и $y = \bar{y}$.

Формула (2.93) дает уже линейную зависимость между z и x, y.

Из формулы (2.93) с учетом формул (2.79) и (2.80) можно получить для нелинейных функций:

$$\bar{z} = \varphi(\bar{x}, \bar{y}), \tag{2.94}$$

$$\sigma_z^2 = \left(\frac{\partial\varphi}{\partial\bar{x}}\right)^2 \sigma_{01}^2 + \left(\frac{\partial\varphi}{\partial\bar{y}}\right)^2 \sigma_{02}^2. \tag{2.95}$$

Формулы (2.94), (2.95) приближенные, так как в их основе лежит приближенная формула (2.93).

2.1.9. Порядок обработки серии измерений

Обработку серии измерений следует проводить по этапам в следующем порядке:

1) определить среднюю арифметическую измерений $\bar{x}$ (по формулам (2.1) – (2.22));

2) найти среднюю квадратическую ошибку (отклонение) отдельного измерения σ (по формулам (2.36), (2.44) или (2.48));

3) определить наибольшую возможную ошибку Δ отдельного измерения (по формуле (2.53));

4) проверить, нет ли среди результатов измерений таких x_i, которые отличались бы от средней арифметической $\bar{x}$ более чем на Δ;

5) если таковые x_i оказались в наличии, то их следует отбросить и начать обработку сначала (с этапа 1);

6) повторять таким образом этапы 1 – 5 до тех пор, пока не останется таких результатов измерений x_i, которые не будут отличаться от средней арифметической более чем на Δ;

7) определить среднюю квадратическую ошибку σ_0 средней арифметической (по формуле (2.67)).

Остальные характеристики (r_0, Δ_0, h, H) находятся только в случае необходимости.

2.2. Проверка достоверности полученных экспериментальных данных

2.2.1. Критерий Пирсона

В практике инженерной защиты окружающей среды часто приходится встречаться с задачами следующего рода. Некоторое измерение производится несколько раз, причем известна теоретическая частота появления некоторого события при этом измерении. Однако на практике фактическая частота оказалась отличной от теоретической. Необходимо установить, можно ли объяснить имевшее место расхождение между частотами случайными причинами или это расхождение существенно и вызвано каким-либо реальным изменением (например, ухудшением работы очистных аппаратов, что приведет к увеличению интенсивности загрязнения).

Мерой расхождения между фактической и теоретической частотой появления некоторого события при измерении является критерий значимости Пирсона:

$$\chi^2 = \sum_{i,j=1}^{k}\left(\frac{\left(\Phi_j^{(i)} - \mathrm{E}_j^{(i)}\right)^2}{\mathrm{E}_j^{(i)}}\right), \tag{2.96}$$

где $\Phi_j^{(i)}$ – фактически полученное значение частоты появления некоторого события для каждого исхода измерения; $\mathrm{E}_j^{(i)}$ – теоретически ожидаемое значение частоты появления некоторого события для каждого измерения; k – количество исходов измерения.

Суммирование в формуле (2.96) производится по всем исходам измерения.

Не рекомендуется применять критерий Пирсона в тех случаях, когда какое-либо E < 5.

После нахождения численного значения критерия Пирсона необходимо по специальным таблицам определить вероятность того, что в силу случайных причин критерий Пирсона примет значение, равное или большее того, которое найдено из измерения. Если эта вероятность окажется малой, то это будет означать, что критерий Пирсона достиг своего расчетного значения не в силу случайных причин и расхождение между теоретической и фактической частотой велико, действительно существует и объясняется реальными причинами.

Таблица значений критерия Пирсона составлена по двум аргументам: вероятности p и числу степеней свободы f.

Под числом степеней свободы понимают число классов, значения которых можно задать произвольно. Иначе говоря, число степеней свободы f есть общее число классов в системе с минус число ограничений l, наложенных на изучаемую систему:

$$f = c - l. \tag{2.97}$$

Вопрос о том, какую вероятность p значения критерия Пирсона нужно считать малой, зависит от характера рассматриваемой задачи в каждом конкретном случае. Часто при решении подобных вопросов вероятность считают малой, если она меньше 0,05 (пятипроцентный уровень значимости). Однако следует иметь в виду, что при этом в одном из каждых 20 случаев мы будем утверждать наличие эффекта, не существующего в действительности. Если такой процент ошибки считается слишком большим, то принимают более высокий уровень значимости, например, 1 %, 5 %.

2.2.2. Критерий Фишера

Для сравнения точности двух рядов измерений (равноточны или неравноточны), для проверки устойчивости технологического процесса (например, переработки отходов) и других, используется критерий Фишера, являющийся отношением выборочных дисперсий двух рядов измерений:

$$F = \frac{\sigma_1^2}{\sigma_2^2}. \tag{2.98}$$

При сравнении двух дисперсий обычно в числителе критерия Фишера (формула 2.98) содержится бóльшая дисперсия.

Значения критерия Фишера зависят только от степеней свободы каждой из двух дисперсий и сведены в специальные таблицы.

После получения численного значения критерия Фишера оно анализируется аналогично значению критерия Пирсона (см. предыдущий раздел), но по таблице значений критерия Фишера.

2.2.3. Критерий Стьюдента

Часто необходимо определить, являются ли два средних арифметических $\bar{x}_1$ и $\bar{x}_2$ оценками одного математического ожидания, т.е. все измерения, лежащие в основе обоих средних арифметических, принадлежат одной и той же совокупности измерений. Для этого применяют критерий Стьюдента:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\sigma_e^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \tag{2.99}$$

где $\bar{x}_1$, $\bar{x}_2$ – средние арифметические двух групп (серий) измерений; n_1, n_2 – число измерений в каждой группе.

Средняя взвешенная дисперсия определяется по формуле

$$\sigma_e^2 = \frac{\sum_{j=1}^{k} f_j \sigma_j^2}{\sum_{j=1}^{k} f_j}, \tag{2.100}$$

где f_j – число степеней свободы в каждой j-й группе измерений; σ_j^2 – дисперсия каждой j-й группы измерений; j – номер группы измерений;
k – число групп измерений.

После получения численного значения критерия Стьюдента, оно анализируется аналогично критерию Пирсона (см. предыдущий раздел), но по таблице значений критерия Стьюдента.

Частным случаем рассмотренной гипотезы является сравнение среднего арифметического с постоянной величиной. Такая задача встречается, например, когда какую-то физико-химическую характеристику, вычисленную теоретически по формуле, необходимо проверить экспериментально. В этом случае исследователь будет иметь только одну выборочную дисперсию и одно среднее арифметическое из экспериментальных данных. Тогда критерий Стьюдента определится по формуле

$$t = \frac{|\bar{x} - a|\sqrt{n}}{\sqrt{\sigma^2}} = \frac{|\bar{x} - a|}{\sigma_0} > t_{p=f(f_e)}, \tag{2.101}$$

где t_p – критерий Стьюдента при определенной вероятности $p = f(f_e)$;
a – некая постоянная (расчетная) величина.

Среднее число степеней свободы определяется по формуле

$$f_e = \sum_{j=1}^{k} f_j = \sum_{j=1}^{k} \left(n_j - 1\right), \tag{2.102}$$

где n_j – число параллельных опытов в j-й группе или выборке измерений.

При выполнении неравенства (2.101) можно утверждать, что значение расчетной величины a не подтверждается экспериментальным путем.

В практике экологических экспериментов иногда появляется необходимость проверки однородности двух средних арифметических при отсутствии однородности дисперсий. Решение этой задачи можно осуществить лишь приближенно и при равных объемах выборок (количествах измерений в каждой группе) $n_1 = n_2 = n$. В этом случае критерий Стьюдента вычисляют по формуле:

$$t \approx \frac{\left|\bar{x}_1 - \bar{x}_2\right|\sqrt{n}}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{n}}} > t_{p=f(f_a)}. \tag{2.103}$$

Здесь для выбора критического значения $t_{p=f(f_a)}$ число степеней свободы находят по формуле

$$f_a = \frac{n-1}{g^2 + (1-g)^2}, \tag{2.104}$$

где

$$g = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.105}$$

Найденное по формуле (2.104) число степеней свободы может оказаться дробным, тогда его округляют.

С помощью критерия Стьюдента можно также характеризовать отклонение среднего арифметического $\bar{x}$ данной выборки, состоящей из n измерений, от истинного значения среднего арифметического $\bar{X}$ всей совокупности. В этом случае критерий Стьюдента определяется по формуле

$$t = \frac{\left|\bar{x} - \bar{X}\right|\sqrt{n}}{\sigma} = \frac{\left|\bar{x} - \bar{X}\right|}{\sigma_0}. \tag{2.106}$$

2.2.4. Критерий Аббе

В инженерно-экологической практике иногда приходится сталкиваться с тем, что со временем под влиянием неизвестных или еще не изученных факторов происходит смещение результатов измерений. В других случаях в некоторой группе измерений может существовать сильно отличающееся от остальных значение, которое оказывается следствием нарушения технологического режима либо неправильности проведения измерений (в этом случае данное измерение отбрасывают), а может и оказаться реально существующим проявлением некоторого реального неконтролируемого фактора, нарушающего стабильность процесса (в этом случае измерение оставляют). В таких случаях необходимо сделать вывод: отбросить данное измерение или оставить. Для этого используются критерий Аббе, определяемый по следующему алгоритму:

1) все результаты экспериментов располагают в ряд во времени или в соответствии с упорядоченным (например, по возрастанию) расположением исследуемого фактора;

2) вычисляют величину критерия Аббе по формуле

$$V = \frac{\sum_{i=1}^{n} \left(x_i - x_{i+1}\right)^2}{2 \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}; \tag{2.107}$$

3) выбирают критическое значение критерия Аббе $V_{кр}$ в соответствии с необходимым уровнем значимости p и числом опытов n по специальным таблицам;

4) сравнивают V и $V_{кр}$;

5) если $V < V_{кр}$, то гипотезу о равенстве математических ожиданий отвергают, т.е. принимают реальным существование смещения результатов экспериментов или во времени, или под воздействием неких реальных факторов (соответственно, измерение, значительно отличающееся от других, оставляют в выборке, а не отбрасывают).

При числе измерений $n > 60$ вместо критерия Аббе (формула (2.107)) используют критерий Стьюдента-Аббе:

$$T = -(1-V)\sqrt{\frac{2n+1}{2-(1-V)^2}}. \tag{2.108}$$

Критическое значение $T_{кр}$ находят по таблицам как критерий Стьюдента $T_{кр} = t$ при числе степеней свободы $f = \infty$.

2.2.5. Доверительные пределы

Любая статистическая экспериментальная характеристика является приближенной. Поэтому она может иметь какой-то определенный смысл лишь в том случае, когда указываются границы возможной погрешности оценки, или, иначе говоря, указывается интервал, о котором с известной вероятностью можно утверждать, что он покрывает оцениваемое нами, вообще говоря, постоянное значение параметра.

Если для оценки некоторого неизвестного параметра θ мы определим вместо одного какого-то значения два значения A и B таким образом, что здесь имеется вероятность (1 – α) того, что

$$A < \theta < B, \tag{2.109}$$

то A и B будут называться 100(1 – α) %-ными доверительными пределами. Так как вероятность того, что этот интервал не включает в себя θ, составляет α (в долях), то при обратном утверждении (формула (2.109)) мы рискуем ошибиться на 100α %. Следует отметить, что мы не утверждаем, что θ имеет вероятность (1 – α) для попадания в область между данными пределами A и B. Значение θ есть просто неизвестная постоянная, и поэтому мы не можем относительно нее сделать такого рода предположения.

Предположим, например, что необходимо по данным выборки оценить характеристику $\overline{X}$ истинной средней арифметической нормальной генеральной совокупности экспериментальных данных, среднее квадратичное отклонение которой считается неизвестным.

В этом случае некоторая величина $\overline{x}$ (средняя арифметическая данной выборки) подчинена нормальному закону с центром $\overline{X}$ и дисперсией σ_0^2. Следовательно, величина критерия Стьюдента t по формуле (2.106) есть нормированное отклонение нормально распределенной случайной величины $\overline{x}$ от истинной средней арифметической всей генеральной совокупности экспериментальных данных.

Пусть t_α есть значение критерия Стьюдента при некотором уровне значимости α. Тогда можно записать:

$$p\left(-t_\alpha < \frac{\left|\overline{x}-\overline{X}\right|\sqrt{n}}{\sigma} < +t_\alpha\right) = 1-\alpha\,, \tag{2.110}$$

или

$$p\left(\overline{x}-\frac{t_\alpha\sigma}{\sqrt{n}} < \overline{X} < \overline{x}+\frac{t_\alpha\sigma}{\sqrt{n}}\right) = 1-\alpha\,. \tag{2.111}$$

Иначе говоря, интервал между $\left(\overline{x}-\frac{t_\alpha\sigma}{\sqrt{n}}\right)$ и $\left(\overline{x}+\frac{t_\alpha\sigma}{\sqrt{n}}\right)$ есть 100 (1 – α) %-ный доверительный интервал для неизвестного среднего $\overline{x}$, если дисперсия σ^2 известна.

Рассмотренный доверительный интервал (формула (2.111)) дает два значения. Иногда приходится решать задачи определения вероятности (1 – α) для значений, которые только больше или только меньше, чем $\overline{X}$. Такие интервалы называются односторонними доверительными интервалами.

Предположим, что выбирается значение $t_{2\alpha}$ таким образом, что $p\left(|t| > t_{2\alpha}\right)$. Тогда вследствие симметрии нормального распределения будем иметь

$$p\left(t < t_{2\alpha}\right) = 1-\alpha \tag{2.112}$$

и

$$p\left(t > -t_{2\alpha}\right) = 1-\alpha\,. \tag{2.113}$$

Преобразуя формулы (2.112) и (2.113) с учетом формул (2.106), (2.110), (2.111), получим:

$$p\left(\overline{x}-\frac{t_{2\alpha}\sigma}{\sqrt{n}} < \overline{X}\right) = 1-\alpha \tag{2.114}$$

и

$$p\left(\overline{x}+\frac{t_{2\alpha}\sigma}{\sqrt{n}} > \overline{X}\right) = 1-\alpha\,. \tag{2.115}$$

Таким образом, интервалы со значениями больше, чем $\left(\overline{x}-\frac{t_{2\alpha}\sigma}{\sqrt{n}}\right)$, и меньше, чем $\left(\overline{x}+\frac{t_{2\alpha}\sigma}{\sqrt{n}}\right)$, являются искомыми односторонними 100 × × (1 – α) %-ными доверительными интервалами для неизвестной средней $\overline{X}$.

Доверительные интервалы могут быть применены к любой нормально распределенной переменной с известным квадратическим отклонением. Например, если мы имеем две выборки n_1 и n_2 из нормальной совокупности со средними $\bar{x}_1$ и $\bar{x}_2$ и дисперсиями σ_1^2 и σ_2^2, то величина

$$\bar{d} = \bar{x}_1 - \bar{x}_2 \tag{2.116}$$

будет нормально распределенной со средней

$$\delta = \bar{X}_1 - \bar{X}_2 \tag{2.117}$$

и с дисперсией

$$\sigma_\delta^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}. \tag{2.118}$$

Следовательно, если σ_1^2 и σ_2^2 известны, то двусторонним доверительным интервалом для δ будет

$$\bar{d} - t_\alpha \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) < \delta < \bar{d} + t_\alpha \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right). \tag{2.119}$$

2.2.6. Статистическая проверка гипотез

Получение экспериментальных данных часто имеет целью высказать некоторое суждение относительно совокупности по опытным данным, которые получены для отдельных образцов. Например, необходимо определить, дает ли новый технологический процесс переработки твердых отходов увеличение выхода полезного продукта по сравнению со старым. Использование экспериментальных данных для ответа на вопросы такого рода называется испытанием или проверкой гипотез.

Простейший случай испытания статистической гипотезы состоит в следующем. Мы принимаем, что рассматриваемая совокупность может быть описана некоторой функцией распределения, которая зависит от одного неизвестного параметра θ. Тогда на основании данных исследования образца $x_1, x_2, ..., x_n$, взятого из этой совокупности, мы желаем либо принять, либо опровергнуть гипотезу о том, что θ имеет некоторое частное значение θ_0.

Испытание данной статистической гипотезы в общем виде производится следующим образом:

1) по результатам измерений $x_1, x_2, ..., x_n$ вычисляются соответствующие статистические данные для отдельных образцов ($\bar{x}, \sigma$ и т.д.);

2) принимая, что гипотеза верна, определяется вероятность отклонения статистических величин от ожидаемого значения;

3) если полученная вероятность меньше некоторого малого определенного значения доверительной вероятности α, например, $\alpha = 0{,}05$, то гипотеза опровергается.

Например, если на основании испытания выборки с величинами $x_1, x_2, ..., x_n$ необходимо проверить гипотезу, что какая-либо неизвестная средняя $\bar{X}$ равна некоторому значению $\bar{X}_0$, то нужно определить вероятность отклонения экспериментальной средней $\bar{x}$ от гипотетической средней $\bar{X}_0$, т.е. $|\bar{x} - \bar{X}_0|$, и отклонить гипотезу о равенстве $\bar{X} = \bar{X}_0$, если эта вероятность мала. Такого рода испытания часто сводятся к испытанию значимости, и, если гипотеза отклоняется, то говорят, что истинная величина θ сильно отличается от гипотетического θ_0 при уровне значимости α.

Для того чтобы придти к определенному заключению хотя бы вероятностного характера, делается гипотетическое допущение о равенстве $\theta = \theta_0$. Такого рода вспомогательные гипотезы об отсутствии интересующего нас различия между параметрами часто называют "нулевыми гипотезами", так как мы никогда не утверждаем, что $\theta = \theta_0$, а только полагаем, что θ не отличается значимо от θ_0.

Необходимо отметить, что в каждом случае доверительные пределы для параметра θ (см. предыдущий раздел) определяют испытание гипотезы $\theta = \theta_0$, так как если θ_0 не попадает в область между этими пределами, то мы можем с 100α %-ным риском или при уровне значимости α заключить, что $\theta \neq \theta_0$.

При испытании гипотез могут быть ошибки первого или второго рода. Во-первых, отклонение гипотезы возможно тогда, когда она верна; вероятность такой ошибки определяется путем выбора уровня значимости α. Во-вторых, ошибка возможна и при утверждении неверной гипотезы.

Пусть, например, гипотеза, которая подлежит испытанию, состоит в том, что с изменением некоторого процесса переработки отходов в полезный продукт выход этого продукта не увеличивается. Тогда, принимая, что выход увеличивается, когда в действительности этого нет, исследователь совершает ошибку первого рода. Однако, утверждение о том, что выход продукта не увеличивается, когда фактически имеет место обратное явление, приводит к ошибке второго рода.

Рассмотрим подробнее задачу испытания гипотезы о том, что неизвестное среднее $\overline{X}$ равно некоторому значению $\overline{X}_0$ на основании результатов наблюдения $x_1, x_2, ..., x_n$, т.е.

$$\overline{X} = \overline{X}_0 . \qquad (2.120)$$

В качестве нормированного отклонения случайной величины $\bar{x}$ от истинной неизвестной средней арифметической $\overline{X}$ принимается величина критерия Стьюдента t по формуле (2.106).

Пусть t_α есть значение критерия Стьюдента при некотором уровне значимости α.

Предположим, что значение t выбрано таким образом, что

$$P(|t| > t_\alpha) = \alpha . \qquad (2.121)$$

Тогда, если

$$t_0 = \frac{|\bar{x} - \overline{X}|\sqrt{n}}{\sigma} > t_\alpha , \qquad (2.122)$$

то гипотезу по формуле (2.120) отклоняем, утверждая, что разность $(\overline{X} - \overline{X}_0)$ на самом деле существует и значима (т.е. $\overline{X} \neq \overline{X}_0$) при уровне значимости α.

Если же гипотеза по формуле (2.120) на самом деле верна, то должно выполняться условие

$$P(|t_0| > t_\alpha) = \alpha , \qquad (2.123)$$

причем α в формуле (2.123) есть вероятность отклонения гипотезы, когда она на самом деле достоверна, или вероятность утверждения, что разность $|\bar{x} - \overline{X}_0|$ значима, когда гипотеза недостоверна.

Теперь предположим, что гипотеза по формуле (2.120) неверна и что $\overline{X}$ в действительности равно некоторому другому значению $\overline{X}_1$, т.е.

$$\overline{X} \neq \overline{X}_0 , \qquad (2.124)$$

$$\overline{X} = \overline{X}_1 . \qquad (2.125)$$

Тогда необходимо исследовать вероятность того, что $|t_0| < t_\alpha$, или что гипотеза не будет отклонена, когда она неверна (вероятность появления ошибки второго рода).

В этом случае величина

$$t_1 = \frac{|\bar{x} - \overline{X}_1|\ \sqrt{n}}{\sigma} \qquad (2.126)$$

будет нормированным отклонением случайной экспериментальной величины $\bar{x}$ от истинной неизвестной средней арифметической $\overline{X}_1$.

Величину t_1 можно представить (с учетом формулы (2.106)) в виде

$$t_1 = \frac{|\bar{x} - \overline{X}_0|\sqrt{n}}{\sigma} = \frac{|\overline{X}_1 - \overline{X}_0|\sqrt{n}}{\sigma} = t_0 - D\sqrt{n} , \qquad (2.127)$$

где

$$D = \frac{\overline{X}_1 - \overline{X}_0}{\sigma} . \qquad (2.128)$$

Из формулы (2.127) можно получить

$$t_0 = t_1 + D\sqrt{n} . \qquad (2.129)$$

Тогда вероятность того, что гипотеза по формулам (2.124) и (2.125) не будет отклонена, она определится по формуле

$$P\left(\left|t_0\right| < t_\alpha\right) = P\left(-t_\alpha < t_0 < +t_\alpha\right) = P\left(-t_\alpha < t_1 + D\sqrt{n} < +t_\alpha\right) =$$

$$= P\left(-t_\alpha - D\sqrt{n} < t_1 < +t_\alpha - D\sqrt{n}\right). \tag{2.130}$$

Формула (2.130) используется в тех случаях, когда есть одна верная и одна неверная гипотезы, и нужно найти вероятность того, что эта неверная гипотеза не будет отклонена, т.е. найти вероятность появления ошибки второго рода. Естественно, такая вероятность должна быть как можно меньше.

С увеличением количества измерений n и при постоянной величине D величины $\left(-t_\alpha - D\sqrt{n}\right)$ и $\left(+t_\alpha - D\sqrt{n}\right)$ уменьшаются при положительных значениях D или возрастают при отрицательных значениях D, и, следовательно, вероятность того, что t_1 попадет в интервал между ними (т.е. вероятность того, что неверная гипотеза не будет отклонена) приближается к нулю.

С помощью формулы (2.130) также определяют вероятность того, что из двух гипотез верная гипотеза не будет отклонена, по формуле

$$P' = 1 - P. \tag{2.131}$$

Как правило, P' должно принимать значения больше 0,95.

По формулам (2.120) – (2.123), (2.131) гипотеза $\overline{X} = \overline{X}_0$ одинаково опровергалась как в случае $\overline{X} < \overline{X}_0$, так и в случае $\overline{X} > \overline{X}_0$. Однако, исследователь часто заинтересован в том, чтобы при возможности отклонить нулевую гипотезу по формуле (2.120), если $\overline{X} > \overline{X}_0$, но остается безразличным к тому, опровергается ли эта гипотеза, если $\overline{X} < \overline{X}_0$. Например, исследуя влияние изменения процесса переработки отходов, предназначенного для увеличения выхода полезного продукта, можно испытать гипотезу о том, что выход не изменяется. Тогда остается отклонить эту гипотезу в случае увеличения выхода продукта. Для практических целей уменьшение выхода продукта является менее пригодным, чем неизменный выход, и поэтому испытываемая нулевая гипотеза заключается в том, что выход не изменяется.

Естественно, при таких условиях нулевая гипотеза по формуле (2.120) отклоняется только тогда, когда действительное неабсолютное значение t_0 больше, чем $t_{2\alpha}$, причем $t_{2\alpha}$ выбирается таким образом, чтобы выполнялось равенство

$$P\left(\left|t\right| > t_{2\alpha}\right) = 2\alpha. \tag{2.132}$$

Соответственно, вероятность того, что не будет отклонена неверная гипотеза $\overline{X} = \overline{X}_0$ при фактических условиях $\overline{X} = \overline{X}_1$, $\overline{X} \neq \overline{X}_0$ (формулы (2.124), (2.125) определится по формуле

$$P\left(t_0 < t_{2\alpha}\right) = P\left(t_1 < t_{2\alpha} - D\sqrt{n}\right). \tag{2.133}$$

Рассмотренные выше случаи могут быть применены для любой переменной, распределенной по нормальному закону, Например, даны две выборки измерений с известными средними арифметическими $\overline{X}_1$, $\overline{X}_2$, дисперсиями σ_1^2, σ_2^2 и количеством измерений в каждой выборке n_1, n_2. Тогда можно испытать гипотезу (с учетом формул (2.103), (2.116) – (2.119)), что

$$\delta = 0, \tag{2.134}$$

или

$$\overline{X}_1 = \overline{X}_2 \tag{2.135}$$

путем расчета критерия Стьюдента по формуле

$$t = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\delta_\delta} = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} \tag{2.136}$$

с последующей обработкой полученного результата.

2.2.7. Быстрые методы обработки экспериментальных данных

Применение быстрых методов обработки экспериментальных данных позволяет резко снизить затраты на статистическую обработку опытных данных. При этом в большинстве случаев не теряется строгость подхода к решению задачи.

В основе быстрых методов лежит использование для расчетов разности между максимальным $x_{\max}$ и минимальным $x_{\min}$ значением случайной величины в выборке, называемой размахом R_n, т.е.:

$$R_n = x_{\max} - x_{\min}, \tag{2.137}$$

где n – объем выборки, равный числу параллельных измерений (измерений x_i одного и того же значения x некоторой величины при одних и тех же условиях).

Ниже рассмотрим отдельные задачи статистической обработки экспериментальных данных быстрыми методами.

1. Среднее квадратическое отклонение (ошибка) может быть вычислено по следующей формуле:

$$\sigma_n = \frac{R_n}{d_n}, \tag{2.138}$$

где d_n – табличный коэффициент, зависящий от объема выборки n.

2. Доверительный интервал для математического ожидания, минуя расчет дисперсии, можно вычислить по формуле

$$\delta_n = \pm R_n K_n, \tag{2.139}$$

где K_n – табличный коэффициент, зависящий от объема выборки n.

3. С помощью быстрого метода можно произвести сравнение средней арифметической $\bar{x}$ экспериментальных данных с неслучайной (истинной) величиной. В этом случае проверку гипотезы отсутствия значимого различия между значениями $\bar{x}$ и a можно выполнить с использованием формулы

$$t = \frac{|a - \bar{x}|}{R_n} < t_{n(\text{кр})}, \tag{2.140}$$

где $t_{n(\text{кр})}$ – модифицированный критерий Стьюдента (табличный), зависящий от объема выборки n.

Выполнение неравенства (2.140) указывает на отсутствие значимого различия между $\bar{x}$ и a.

4. Сравнение двух средних арифметических с помощью быстрого метода проводят в два этапа.

На первом этапе сравнивают две дисперсии σ_1^2 и σ_2^2 двух средних арифметических $\bar{x}_1$ и $\bar{x}_2$. Вычисления проводят по формуле

$$F_n = \frac{R_{n1}}{R_{n2}} < F_{n(\text{кр})}, \tag{2.141}$$

учитывая, что $R_{n1} > R_{n2}$. Здесь $F_{n(\text{кр})}$ – критическое значение модифицированного критерия Фишера (табличное), зависящее от объема выборки n.

Если неравенство (2.141) выполняется, то переходят ко второму этапу, предварительно вычислив средний размах по формуле

$$\bar{R}_n = \frac{R_{n1} + R_{n2}}{2}. \tag{2.142}$$

Затем определяют величину K_Δ (по таблице) и величину R_Δ по формуле

$$R_\Delta = \bar{R}_n K_\Delta. \tag{2.143}$$

Проверяют выполнение неравенства

$$|\bar{x}_1 - \bar{x}_2| < R_\Delta. \tag{2.144}$$

Если неравенство (2.144) выполняется, то делают вывод, что между средними арифметическими $\bar{x}_1$ и $\bar{x}_2$ нет значимого различия, а если неравенство (2.144) не выполняется, то сравнение двух средних арифметических проводят по формулам (2.99) или (2.103).

Однако, наличие быстрых методов не исключает необходимости применения обычных способов статистической обработки экспериментальных данных, так как промежуточные значения x_i в выборке (между $x_{\min}$ и $x_{\max}$) несут определенный объем информации, пренебрегать которым следует не во всех случаях. Здесь имеет место обычная картина, связанная с применением упрощенных методов: выигрыш в затрате времени на вычисления и проигрыш, связанный с получением более грубой оценки по причине потери части информации.

2.2.8. Анализ остатков

Остатки – это неожидаемые резко выделяющиеся отклонения каких-либо значений в отдельных точках ряда параллельных измерений.

Анализ остатков может дать дополнительную информацию об исследуемом процессе. Вывод о том, отбросить данное измерение или оставить, делается с помощью критерия Аббе (см. раздел 2.2.4). Непосредственный же анализ остатков проводят визуально посредством нанесения их на график.

Если при проведении процесса либо при измерении экспериментальных данных по процессу не происходило реальных нарушений, то остатки случайно распределяются в пределах доверительного интервала (рис. 2.4) с определенной вероятностью.

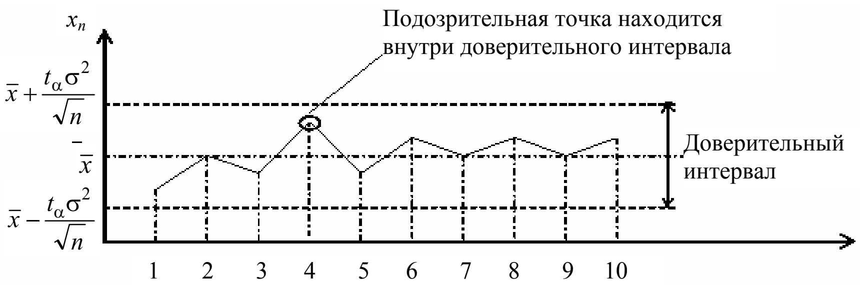

Рис. 2.4. Распределение остатков в пределах доверительного интервала

Если нарушения реально существуют, то наиболее часто на графиках распределения экспериментальных данных встречаются следующие виды остатков.

1. Выбросы – отдельные остатки, превосходящие доверительный интервал по абсолютной величине (рис. 2.5).

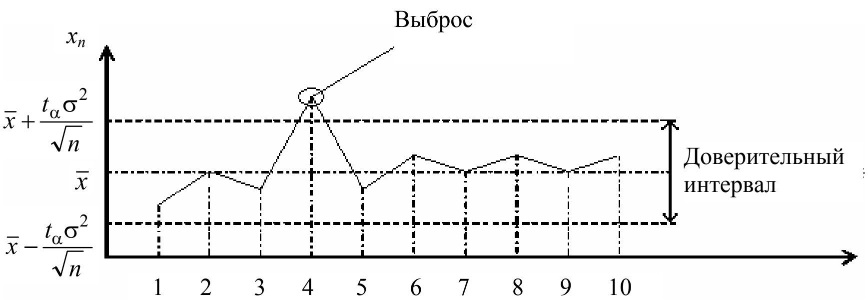

РИС. 2.5. ОСТАТКИ, ПРЕВОСХОДЯЩИЕ ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ПО АБСОЛЮТНОЙ ВЕЛИЧИНЕ

Существование выбросов может быть связано с нарушением режима ведения процесса, соответствующего данной точке. В этом случае необходима постановка дополнительных экспериментов в точках выбросов. Если нарушение режима ведения процесса не подтверждается дополнительными экспериментами, то точки выбросов могут оказаться особо интересными для исследователя, так как их наличие в данном случае может быть связано с неправильным проведением измерений, несоответствием вида разработанной математической модели, неправильной формой поверхности отклика или механизмом процесса.

2. Временной дрейф – смещение во времени экспериментальных измерений одной и той же величины при одних и тех же условиях, как правило, при проведении длительного эксперимента. Различные варианты временного дрейфа возможны также в том случае, когда эксперименты ставили на нескольких однотипных установках.

Временной дрейф имеет различные формы. Встречаются наклонный, квадратичный и скачкообразный дрейфы.

3. ПОСТРОЕНИЕ ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЕЙ МЕЖДУ ПОЛУЧЕННЫМИ ЭКСПЕРИМЕНТАЛЬНЫМИ ДАННЫМИ

3.1. ЭМПИРИЧЕСКИЕ ДАННЫЕ

Экспериментальное изучение какой-либо неизвестной закономерности

$$y = f(x) \tag{3.1}$$

дает результаты наблюдений в виде таблицы соответственных значений x_j и y_j, причем x_j есть среднее значение величин $x_{i,j}$, а y_j есть среднее значение величин $y_{i,j}$.

По значениям x_j и y_j можно построить кривую зависимости y от x. Эту же зависимость можно приближенно представить некоторой эмпирической формулой

$$y = \varphi(x)\ . \tag{3.2}$$

Очевидно, что выбор той или иной эмпирической формулы диктуется требованием наилучшего приближения $\varphi(x)$ к $f(x)$ в некотором доверительном интервале значений

$$\alpha \le x \le \beta\ . \tag{3.3}$$

Функцию $f(x)$ можно выразить различными эмпирическими формулами.

В некоторых задачах в качестве $\varphi(x)$ берут функцию, для которой в заданном интервале (3.3) наибольшее значение величины $|f(x) - \varphi(x)|$ будет меньше, чем при выборе любой другой эмпирической формулы.

Более удобно производить оценку приближения $\varphi(x)$ к $f(x)$ по методу наименьших квадратов. В этом случае функцией, дающей лучшее приближение, считается такая функция, для которой величина

$$S = \int_{\alpha}^{\beta} [f(x) - \varphi(x)]^2 dx \tag{3.4}$$

имеет наименьшее значение.

Так как обычно бывают известны значения функции лишь для отдельных значений x_j в заданном интервале, то искомую эмпирическую формулу (3.2) подчиняют требованию: сумма

$$S = \sum_{j=0}^{k} \left[f(x_j) - \varphi(x_j) \right]^2\ . \tag{3.5}$$

должна иметь наименьшее значение из всех возможных.

Перебирая различные варианты эмпирических формул, обычно можно добиться любой степени приближения и даже полного совпадения между опытными данными и формулой. Однако, необходимо отметить, что нет нужды стремиться к полному совпадению всех экспериментальных данных с эмпирической формулой, так как в силу наличия погрешности экспериментальных данных (отклонения их от истинного неизвестного значения) такое совпадение иногда даже уменьшает точность формулы.

3.2. ПРОВЕРКА ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ЭМПИРИЧЕСКОЙ ФОРМУЛЫ

В некоторых случаях выбор типа эмпирической формулы может быть произведен на основе теоретических представлений о характере изучаемой зависимости или об изменении измеряемых величин. В других случаях приходится подбирать формулу, сравнивая кривую, построенную по данным наблюдений, с типичными графиками формул, приведенных в справочниках. Иногда оказывается, что эмпирическая кривая похожа на несколько кривых, уравнения которых различны. С другой стороны, нередки случаи, когда та или иная эмпирическая формула достаточно точно выражает зависимость между заданными численными значениями величин, но типичный график этой формулы совершенно не похож на экспериментальную кривую – это может иметь место, когда экспериментальная кривая и график формулы построены для разных промежутков изменения аргумента.

Изменение численных значений коэффициентов, входящих в эмпирическую формулу, часто резко меняет вид ее графика. Выбор масштаба координатных осей отражается на форме построенной кривой, что также может привести к кажущемуся отличию экспериментальной кривой от графика вполне соответствующей ей формулы.

Поэтому, прежде чем определять численные значения коэффициентов в выбранной эмпирической формуле, необходимо проверить возможность ее использования методом выравнивания. Лишь после этого можно перейти к отысканию тех значений постоянных коэффициентов, которые дадут наилучшее приближение опытных и вычисляемых величин.

Метод выравнивания заключается в преобразовании функции $y = \varphi(x)$ таким образом, чтобы превратить ее в линейную функцию. Достигается это путем замены переменных x и y новыми переменными:

$$X = \psi(x, y) \tag{3.6}$$

и

$$Y = \xi(x, y), \tag{3.7}$$

которые выбираются так, чтобы получилось уравнение прямой линии

$$Y = A + BX. \tag{3.8}$$

Вычислив значения X_j и Y_j по заданным x_j и y_j, наносят их на диаграмму с прямоугольными координатами (X, Y). Если построенные таким образом точки располагаются вблизи прямой линии, то выбранная эмпирическая формула $y = \varphi(x)$ подходит для характеристики зависимости $y = f(x)$.

3.3. ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ, ВХОДЯЩИХ В ЭМПИРИЧЕСКУЮ ФОРМУЛУ

После того, как установлена возможность использования выбранной эмпирической формулы, необходимо определить численные значения входящих в формулу коэффициентов. Наилучшие результаты дает использование способа наименьших квадратов. Однако этот способ громоздок, и во многих случаях его можно заменить более простым способом средних, дающим менее точные, но вполне удовлетворительные результаты.

Способ средних заключается в следующем: использовав метод выравнивания и получив линейную зависимость (3.8), составляют условные уравнения

$$Y_j = A + BX_j, \tag{3.9}$$

число n которых равно числу имеющихся соответственных значений X_j и Y_j.

Условные уравнения (3.9) разбивают на две приблизительно равные группы, и уравнения, входящие в каждую из этих групп, складывают. Получают два уравнения:

$$\sum_{j=1}^{k} Y_j = kA + B\sum_{j=1}^{k} X_j, \tag{3.10}$$

и

$$\sum_{j=k+1}^{n} Y_j = (n-k)\,A + B\sum_{j=k+1}^{n} X_j, \tag{3.11}$$

из которых находят неизвестные коэффициенты A и B.

Группировку условных уравнений перед их суммированием можно произвести различными способами, причем каждый из них дает несколько отличающиеся значения коэффициентов.

Лучшим способом группировки будет тот, который приводит к решению, дающему наименьшую сумму квадратов отклонений вычисленных значений функции $\varphi(x)$ от опытных значений функции $f(x)$. Этот лучший способ может быть выбран только путем сравнения результатов вычислений по всем возможным способам группировки, что является очень длительным процессом. Поэтому обычно группируют уравнения (3.9) в последовательности экспериментальных данных, разбивая их на равные или приблизительно равные группы. Считается, что этот прием чаще всего дает наиболее удовлетворительные результаты, хотя теоретически обосновать этого нельзя.

Способ средних тем более надежен, чем больше имеется экспериментальных точек, числу которых соответствует число условных уравнений (3.9).

В инженерно-экологической практике используются различные виды эмпирических зависимостей.

3.4. ОСНОВНЫЕ ЭМПИРИЧЕСКИЕ ФОРМУЛЫ

3.4.1. Зависимость вида $y = ae^{bx}$

Рассмотрим зависимость вида

$$y = ae^{bx}. \tag{3.12}$$

Для проверки возможности применения формулы (3.12) используют метод выравнивания – логарифмируют данную формулу:

$$\lg y = \lg a + \frac{b}{2{,}303} x \,. \qquad (3.13)$$

Если при нанесении на график значений $\lg y$ в зависимости от значений x построенные точки располагаются приблизительно на прямой линии, то это указывает на то, что переменные x и y действительно связаны зависимостью вида (3.12).

Примеры графиков функций (3.12) и (3.13) показаны на рис. 3.1.

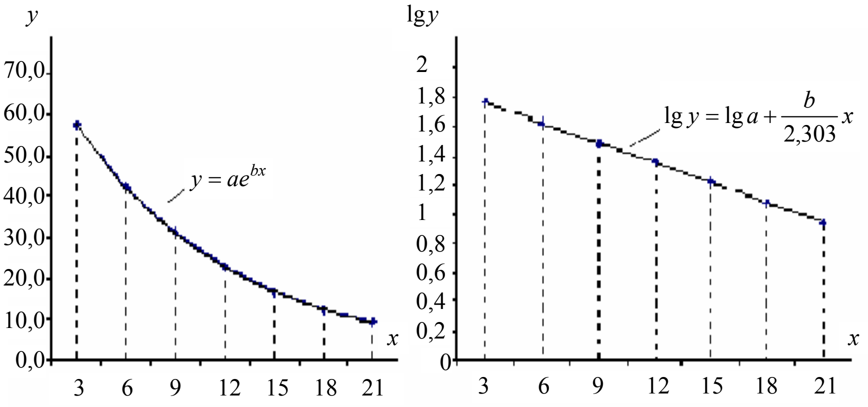

Рис. 3.1. Примеры графиков функций $y = ae^{bx}$ **и** $\lg y = \lg a + \frac{b}{2{,}303} x$

3.4.2. Зависимость вида $y = ax^b$

Рассмотрим зависимость вида

$$y = ax^b \,. \qquad (3.14)$$

Логарифмируя формулу (3.14), найдем

$$\lg y = \lg a + b \lg x \,. \qquad (3.15)$$

Рис. 3.2. Пример графика функции $y = ax^b$

Если при нанесении на график значений $\lg y$ в зависимости от значений $\lg x$ построенные точки располагаются приблизительно на прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.14).

Пример графика функции (3.14) показан на рис. 3.2.

3.4.3. Зависимость вида $y = ax^b + c$

Рассмотрим зависимость вида

$$y = ax^b + c. \tag{3.16}$$

Логарифмируя формулу (3.16), найдем

$$\lg(y - c) = \lg a + b \lg x. \tag{3.17}$$

Если при нанесении на график значений $\lg(y - c)$ в зависимости от значений $\lg x$ построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.16).

Для определения коэффициента c в формуле (3.17) отмечают крайние точки экспериментальной кривой (x_1, y_1) и (x_2, y_2) и находят из чертежа значение y_3 для

$$x_3 = \sqrt[2]{x_1 x_2}. \tag{3.18}$$

Так как координаты этих трех точек удовлетворяют экспериментальной кривой, то

$$y_1 - c = ax_1^b, \tag{3.19}$$

$$y_2 - c = ax_2^b, \tag{3.20}$$

$$y_3 - c = ax_3^b. \tag{3.21}$$

Возводя зависимость (3.18) в степень b и умножив ее на a, полу-чают

$$ax_3^b = \sqrt{ax_1^b \cdot ax_2^b}, \tag{3.22}$$

или, подставляя формулы (3.19) – (3.21) в формулу (3.22):

$$y_3 - c = \sqrt{(y_1 - c)\,(y_2 - c)}, \tag{3.23}$$

откуда

$$c = \frac{y_1 y_2 - y_3^2}{y_1 + y_2 - 2y_3}. \tag{3.24}$$

Пример графика функции (3.16) показан на рис. 3.3.

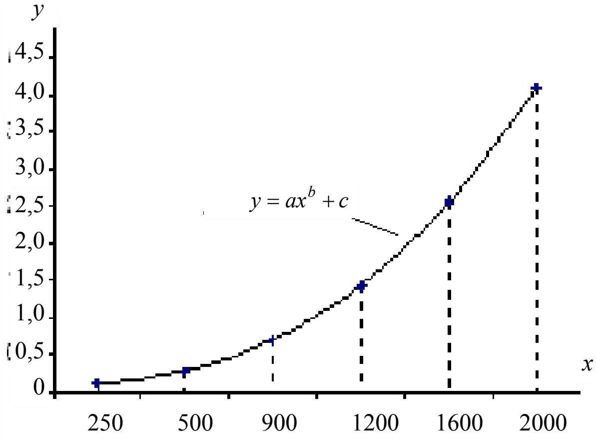

Рис. 3.3. Пример графика функции $y = ax^b + c$

3.4.4. Зависимость вида $y = a \cdot 10^{bx}$

Рассмотрим зависимость вида

$$y = a \cdot 10^{bx}. \tag{3.25}$$

Логарифмируя формулу (3.25), найдем

$$\lg y = \lg a + bx. \tag{3.26}$$

Если при нанесении на график значений $\lg y$ в зависимости от значений x построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.25).

Пример графика функции (3.25) показан на рис. 3.4.

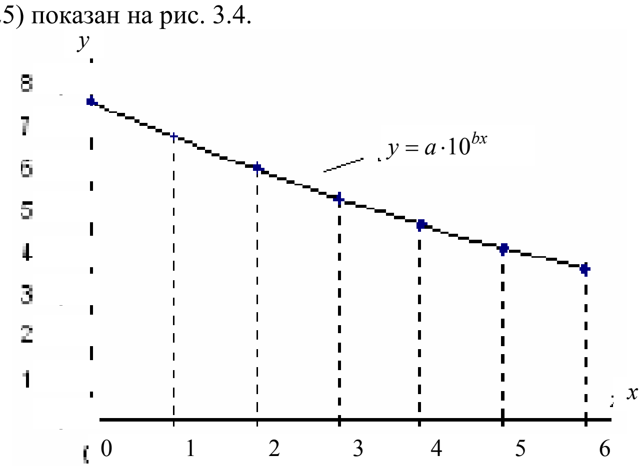

Рис. 3.4. Пример графика функции $y = a \cdot 10^{bx}$

3.4.5. Зависимость вида $y = 10^{a+bx}$

Рассмотрим зависимость вида

$$y = 10^{a+bx}. \tag{3.27}$$

Логарифмируя формулу (3.27), найдем

$$\lg y = a + bx. \tag{3.28}$$

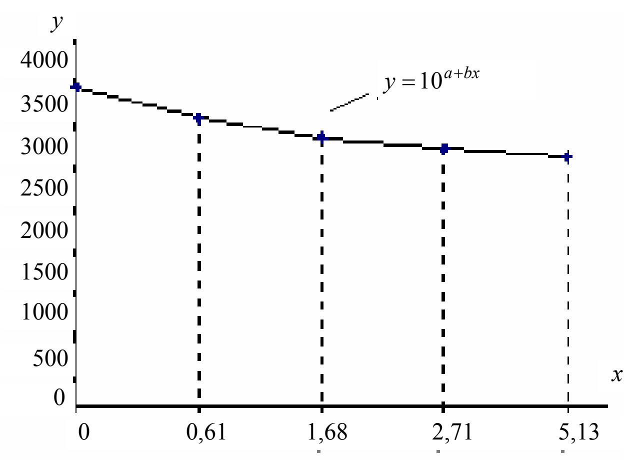

Рис. 3.5. Пример графика функции $y = 10^{a+bx}$

Если при нанесении на график значений $\lg y$ в зависимости от значений x построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.27).

Пример графика функции (3.27) показан на рис. 3.5.

3.4.6. Зависимость вида $y = 10^{a+bx} + c$

Рассмотрим зависимость вида

$$y = 10^{a+bx} + c\ . \tag{3.29}$$

Логарифмируя формулу (3.29), найдем

$$\lg(y - c) = a + bx\ . \tag{3.30}$$

Если при нанесении на график значений $\lg(y-c)$ в зависимости от значений x построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.29).

Для определения коэффициента c в формуле (3.30) выбирают на кривой экспериментальных данных крайние точки (x_1, y_1) и (x_2, y_2) и находят из чертежа значение y_3 для

$$x_3 = \frac{x_1 + x_2}{2}\ . \tag{3.31}$$

Так как все три данные точки лежат на экспериментальной кривой, то

$$y_1 - c = 10^{a+bx_1}\ , \tag{3.32}$$

$$y_2 - c = 10^{a+bx_2}\ , \tag{3.33}$$

$$y_3 - c = 10^{a+bx_3} \tag{3.34}$$

и, соответственно,

$$\lg(y_1 - c) = 10^{a+bx_1}\ , \tag{3.35}$$

$$\lg(y_2 - c) = 10^{a+bx_2}\ , \tag{3.36}$$

$$\lg(y_3 - c) = 10^{a+bx_3}\ . \tag{3.37}$$

Подставив формулу (3.31) в формулу (3.37) и произведя элементарные преобразования, получим

$$a + bx_3 = \frac{a+a}{2} + b\frac{x_1 + x_2}{2} = \frac{1}{2}\left[(a + bx_1) + (a + bx_2)\right]. \tag{3.38}$$

Подставив формулы (3.35) – (3.37) формулу (3.38), получим

$$\lg(y_3 - c) = \frac{1}{2}\left[\lg(y_1 - c) + \lg(y_2 - c)\right] = \frac{1}{2}\lg\left[(y_1 - c)\ (y_2 - c)\right], \tag{3.39}$$

откуда

$$y_3 - c = \sqrt{(y_1 - c)\ (y_2 - c)}\ , \tag{3.40}$$

откуда

$$c = \frac{y_1 y_2 - y_3}{y_1 + y_2 - 2y_3}. \quad (3.41)$$

Пример графика функции (3.29) показан на рис. 3.6.

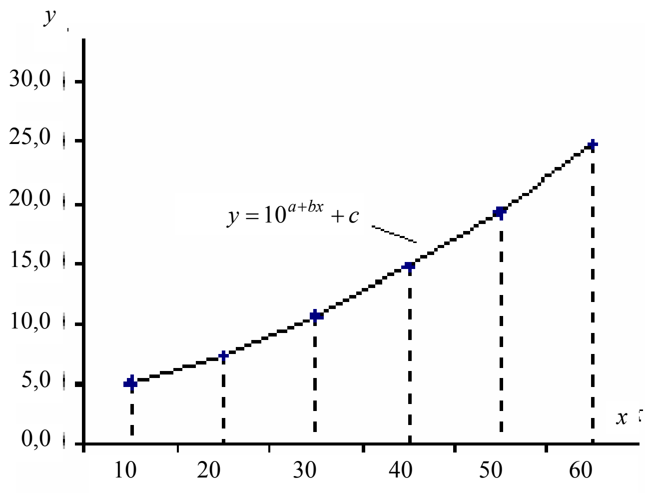

Рис. 3.6. Пример графика функции $y = 10^{a+bx} + c$

3.4.7. Зависимость вида $y = a + bx + cx^2$

Рассмотрим зависимость вида

$$y = a + bx + cx^2. \quad (3.42)$$

Произведем выравнивание формулы (3.42), для чего внесем в нее вместо x и y какие-либо их значения x_1 и y_1 из некоторой экспериментальной зависимости $y = f(x)$:

$$y_1 = a + bx_1 + cx_1^2. \quad (3.43)$$

Вычтем формулу (3.43) из формулы (3.42):

$$y - y_1 = b(x_1 - x_2) + c(x^2 - x_1^2). \quad (3.44)$$

Разделим уравнение (3.44) на $(x - x_1)$:

$$\frac{y - y_1}{x - x_1} = b + c(x + x_1) = b + cx + cx_1. \quad (3.45)$$

Если при нанесении на график значений $\frac{y - y_1}{x - x_1}$ в зависимости от значений x построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.42).

Пример графика функции (3.42) показан на рис. 3.7.

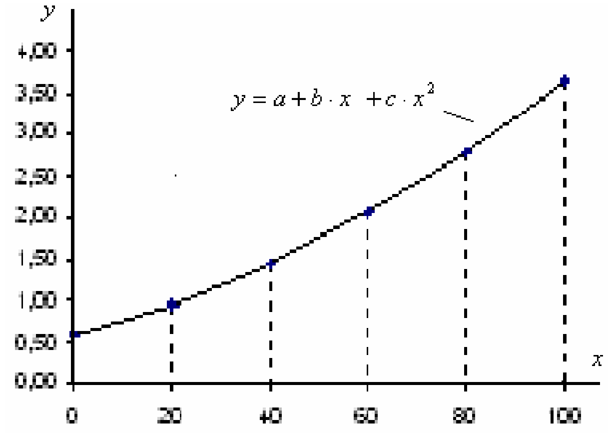

Рис. 3.7. Пример графика функции $y = a + bx + cx^2$

3.4.8. Зависимость вида $y = \dfrac{x - x_1}{a + bx} + y_1$

Рассмотрим зависимость вида

$$y = \frac{x - x_1}{a + bx} + y_1 . \tag{3.46}$$

Произведя элементарные преобразования с формулой (3.46), получим

$$\frac{x - x_1}{y - y_1} = a + bx . \tag{3.47}$$

Если при нанесении на график значений $\dfrac{y - y_1}{x - x_1}$ в зависимости от значений x построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.46).

Пример графика функции (3.46) показан на рис. 3.8.

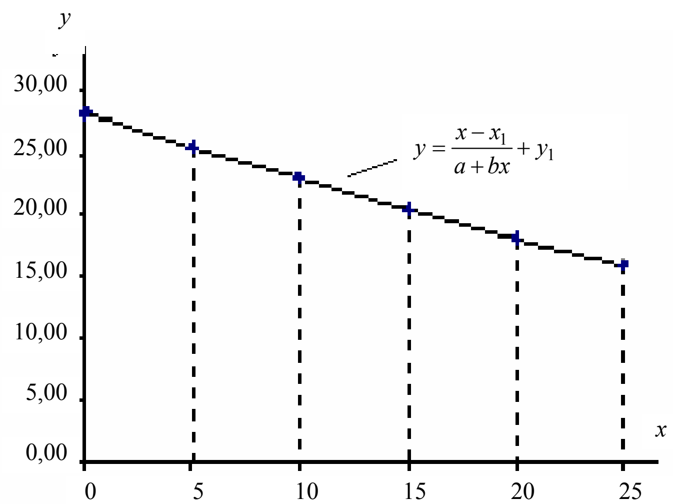

Рис. 3.8. Пример графика функции $y = \dfrac{x - x_1}{a + bx} + y_1$

3.4.9. Зависимость вида $y = \dfrac{x}{a + bx}$

Рассмотрим зависимость вида

$$y = \frac{x}{a + bx} . \tag{3.48}$$

Произведя элементарные преобразования с формулой (3.48), получим

$$\frac{1}{y} = a\frac{1}{x} + b . \tag{3.49}$$

Если при нанесении на график значений $\frac{1}{y}$ в зависимости от значений x построенные точки располагаются приблизительно на одной прямой линии, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.48).

Пример графика функции (3.48) показан на рис. 3.9.

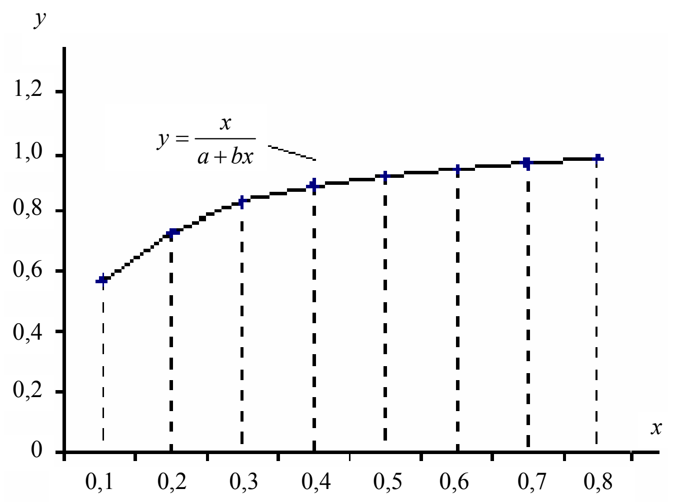

Рис. 3.9. Пример графика функции $y = \frac{x}{a+bx}$

3.4.10. Зависимость вида $y = a + bx + 10^{c+dx}$

Рассмотрим зависимость вида

$$y = a + bx + 10^{c+dx} . \tag{3.50}$$

Формула (3.50) отражает случай, когда значительная часть точек экспериментальной кривой y = f (x) располагается на прямой линии, в то время как остальные точки располагаются на кривой (рис. 3.10).

Пример графика функции (3.50) показан на рис. 3.10.

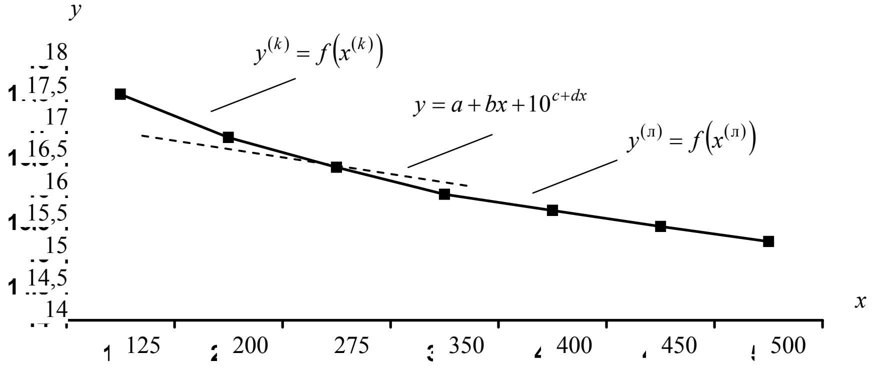

Рис. 3.10. Пример графика функции $y = a + bx + 10^{c+dx}$

В заданном случае формула (3.50) преобразуется к виду

$$y = y' + y'', \tag{3.51}$$

где

$$y' = a + bx, \tag{3.52}$$

$$y'' = 10^{c+dx}. \tag{3.53}$$

Коэффициенты a и b в формуле (3.52) находят сразу по способу средних по формулам (3.10) и (3.11), принимая $y'_j = y_j^{(л)}$ для точек, расположенных на прямой линии. Затем полученные значения a и b подставляют в формулу (3.51) для значений $x_i^{(k)}$, располагающихся на кривой, и находят значения $y_j^{(k)}$ на кривой. Формулу (3.53) преобразуют к виду по формуле (3.28):

$$\lg y'' = c + dx, \tag{3.54}$$

а затем находят коэффициенты c и d в формуле (3.54) по способу средних по формулам (3.10) и (3.11) , принимая

$$\lg y'' = \lg(y - y^{(k)}), \tag{3.55}$$

или

$$\lg y''_i = \lg(y_i - y_i^{(k)}). \tag{3.56}$$

Найденные в результате коэффициенты a, b, c, d подставляют в формулу (3.50), получая искомую эмпирическую формулу y = φ(x).

3.4.11. Зависимость вида $y = 10^{a+b\lg x+c\lg^2 x}$

Рассмотрим зависимость вида

$$y = 10^{a+b\lg x+c\lg^2 x}. \tag{3.57}$$

Произведем выравнивание формулы (3.57). Первоначально прологарифмируем формулу (3.57):

$$\lg y = a + b\lg x + c\lg^2 x. \tag{3.58}$$

Внесем в уравнение (3.58) вместо x и y координаты (x′, y′) какой-либо точки экспериментального графика:

$$\lg y' = a + b\lg x' + c\lg^2 x'. \tag{3.59}$$

Преобразуем формулу (3.59) к виду

$$\frac{\lg y - \lg y'}{\lg x - \lg x'} = b + c\lg x' + c\lg x. \tag{3.60}$$

Если при нанесении на график значений $\frac{\lg y - \lg y'}{\lg x - \lg x'}$ в зависимости от значений lg x построенные точки располагаются приблизительно на данной прямой, то это указывает на то, что экспериментальные переменные x и y действительно связаны зависимостью вида (3.57).

Коэффициенты a, b, c находят по способу средних по формулам (3.10) и (3.11) из формулы

$$\frac{\lg y - \lg y'}{\lg x - \lg x'} = a' + b'\lg x, \tag{3.61}$$

откуда сначала находят значения коэффициентов a′, b′, затем подставляют их в эту же формулу (3.61) и выражают из данной формулы значение y.
В результате полученная формула сразу будет включать коэффициенты a, b, c в числовом виде.

Пример графика функции (3.57) показан на рис. 3.11.

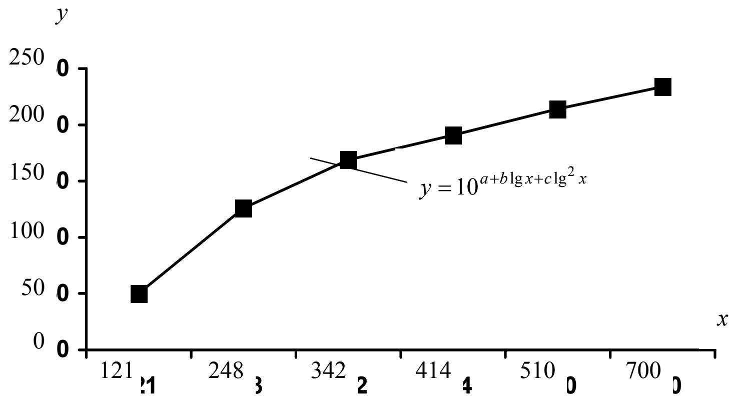

Рис. 3.11. Пример графика функции $y=10^{a+b\lg x+c\lg^2 x}$

3.4.12. Зависимость вида $\cos(57{,}29y)=f(x)$

Рассмотрим зависимость вида

$$\cos(57{,}29y)=f(x)\,. \tag{3.62}$$

Выравнивание формулы (3.62) производится путем приведения к виду

$$\cos(57{,}29y+\alpha)=a+bx\,. \tag{3.63}$$

Здесь значение α имеет размерность градусов угла и определяется методом подбора таким образом, чтобы получить прямую линию при изображении функции (3.63) на графике. Значение y в формулах (3.62), (3.63) имеет размерность радиан, а число 57,29 является коэффициентом перевода радианов в градусы угла.

Пример графика функции (3.62) показан на рис. 3.12.

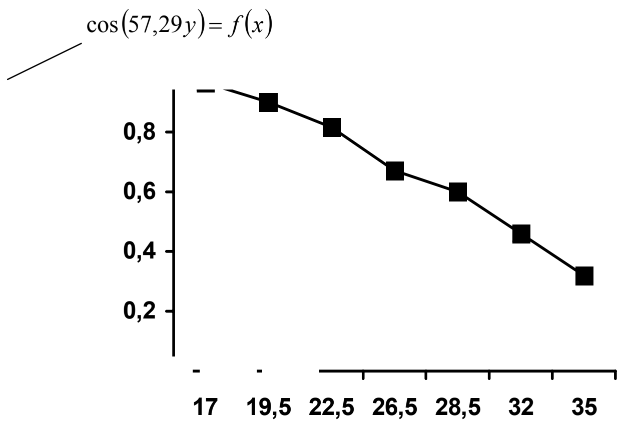

Рис. 3.12. Пример графика функции $\cos(57{,}29y)=f(x)$

3.4.13. Зависимость вида $y=\gamma(x)$

Рассмотрим зависимость вида

$$y=\gamma(x)\,. \tag{3.64}$$

В некоторых случаях в инженерно-технической практике исследование некой функции (3.64) показывает, что она была бы линейной, если бы значениям $y_1, y_2, \ldots, y_n$ соответствовали бы значения не $x_1, x_2, \ldots, x_n$, а x'_1, x'_2, …, x'_n причем $x' = f(x)$, т.е. если бы имела место линейная зависимость

$$y = a + bx' = a + b\,f(x)\,, \tag{3.65}$$

например,

$$y = a + b(x + cx^d)\,. \tag{3.66}$$

Здесь коэффициенты a и b находятся по способу средних по формулам (3.10) и (3.11). Затем определяется вид функции $x' = f(x)$ и находятся по способу средних коэффициенты c и d, входящие в функцию $x' = f(x)$.

Пример графика функции (3.66) показан на рис. 3.13.

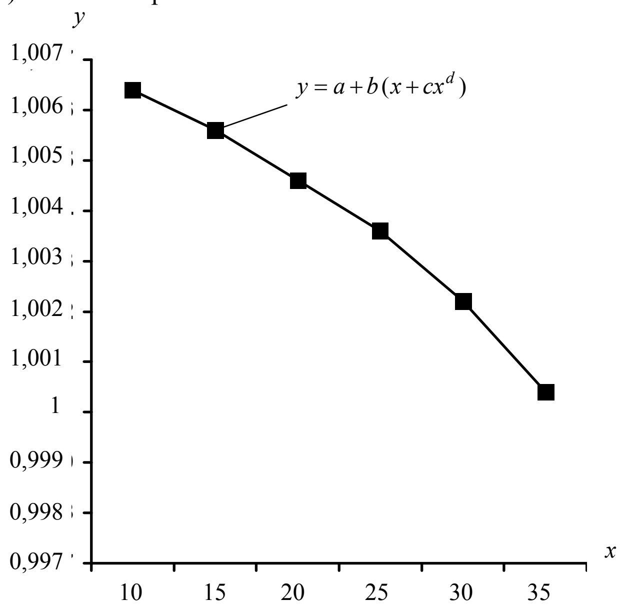

Рис. 3.13. Пример графика функции $y = a + b(x + cx^d)$

3.4.14. Зависимость вида $y = a + bx + \xi$

Рассмотрим зависимость вида:

$$y = a + bx + \xi\,, \tag{3.67}$$

где

$$\xi = \frac{\alpha}{e^{n(X-x)} + e^{n(x-X)}}\,. \tag{3.68}$$

Формула (3.67) отражает зависимость, образующую на графике прямую линию с выступом (рис. 3.14). Значение коэффициента n в формуле (3.68) определяют путем подбора.

В формуле (3.68) X есть значение x, отвечающее максимальному отклонению кривой y по формуле (3.67) от прямой линии y′, где

$$y' = a + bx\,. \tag{3.69}$$

Данное максимальное отклонение y′ от y определяется по формуле

$$\delta = y - y' = \frac{\alpha}{2}\,. \tag{3.70}$$

Соответственно, X равно некоторому x_j^{δ} при $\delta = \max[\delta_j]$.

Пример графика функции (3.67) показан на рис. 3.14.

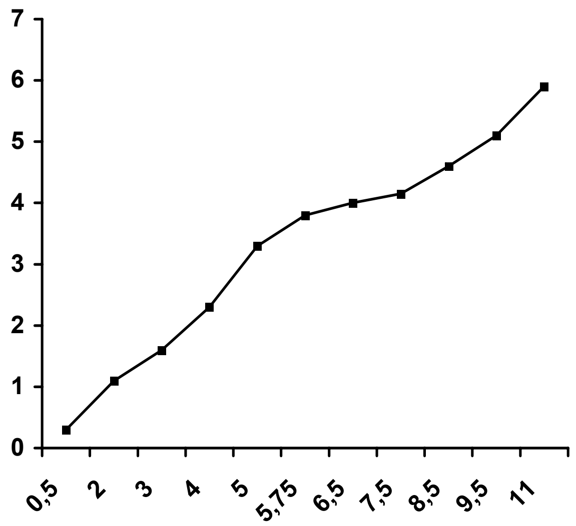

Рис. 3.14. Пример графика функции $y = a + bx + \dfrac{\alpha}{e^{n(X-x)} + e^{n(x-X)}}$

3.5. ИНТЕРПОЛЯЦИОННАЯ ФОРМУЛА ЛАГРАНЖА

Пусть функция y = f (x) задана таблицей. Часто оказывается необходимым вычислить значение функции y при значении x′, не помещенном в этой таблице. Эта задача называется интерполированием.

Интерполирование табличных значений, разность между которыми мала, производится, обычно, с помощью пропорций, как, например, при вычислении логарифмов.

Если же табличные разницы значительны и быстро изменяются, то интерполирование можно осуществить путем нахождения приближенного аналитического представления данной функции.

Рассмотрим интерполяционную формулу Лагранжа. Пусть при $x = a_1, a_2, \dots, a_n$ функция принимает, соответственно, значения $y_1, y_2, \dots, y_n$. Лагранж нашел выражение для многочлена степени (n – 1), который принимает те же значения при $x = a_1, a_2, \dots, a_n$, что и заданная функция.

Этот многочлен имеет следующий вид

$$y = y_1 \frac{(x-a_2)\,(x-a_3)\,\dots\,(x-a_n)}{(a_1-a_2)\,(a_1-a_3)\,\dots\,(a_1-a_n)} + $$

$$+\, y_2 \frac{(x-a_1)\,(x-a_3)\,\dots\,(x-a_n)}{(a_2-a_1)\,(a_2-a_3)\,\dots\,(a_2-a_n)} + \dots +$$

$$+\, y_n \frac{(x-a_1)\,(x-a_2)\,\dots\,(x-a_{n-1})}{(a_n-a_1)\,(a_n-a_2)\,\dots\,(a_n-a_{n-1})}\,. \qquad (3.71)$$

Для упрощения расчетов по этой формуле рекомендуется следующий прием вычислений.

1. Заполняется таблица (табл. 3.1). Левая часть таблицы позволяет находить числители, а правая – знаменатели отдельных членов формулы Лагранжа.

3.1. Члены формулы Лагранжа

Числитель					Знаменатель						
	a_1	a_2	a_3	a_4		a_1	a_2	a_3	a_4	+	–
y_1		$x - a_2$	$x - a_3$	$x - a_4$	a_1		$a_1 - a_2$	$a_1 - a_3$	$a_1 - a_4$		
y_2	$x - a_1$		$x - a_3$	$x - a_4$	a_2	$a_2 - a_1$		$a_2 - a_3$	$a_2 - a_4$		
y_3	$x - a_1$	$x - a_2$		$x - a_4$	a_3	$a_3 - a_1$	$a_3 - a_2$		$a_3 - a_4$		
y_4	$x - a_1$	$x - a_2$	$x - a_3$		a_4	$a_4 - a_1$	$a_4 - a_2$	$a_4 - a_3$			

2. Для нахождения коэффициента при y_1 в формуле Лагранжа нужно произведение выражений, помещенных в ячейки строки "y_1" левой части таблицы, разделить на произведение чисел, находящихся в строке "a_1" правой части таблицы. Аналогично находятся и все остальные коэффициенты.

3.6. СПЕЦИАЛЬНЫЕ МЕТОДЫ НАХОЖДЕНИЯ ЭМПИРИЧЕСКИХ ФОРМУЛ ДЛЯ ТРЕХ ПЕРЕМЕННЫХ

Некоторые вопросы в защите окружающей среды приводят нас к задаче о нахождении эмпирической зависимости, связывающей три переменные x, y, z.

Для определения будем считать x и y независимыми переменными, а z – функцией:

$$z = f(x, y). \tag{3.72}$$

Общий метод решения таких задач состоит в следующем. Считая x постоянным, свяжем z с y зависимостью

$$\Phi(z) = a + b\ F(y), \tag{3.73}$$

которую определим, применяя какие-либо из вышерассмотренных методов (разделы 3.3 – 3.5).

Числа a и b зависимости (3.73) являются функциями от x, которые нужно найти эмпирически.

3.7. ЭМПИРИЧЕСКИЕ ФОРМУЛЫ ПЕРИОДИЧЕСКОГО ХАРАКТЕРА

Нередко при обработке числового материала приходится встречаться с формулами, имеющими периодический характер.

Если Δx_p есть период функции $y = f(x)$, то

$$f(x + \Delta x_p) = f(x). \tag{3.74}$$

Так, например, если x есть угол (в радианах), то sin x, cos x имеют период

$$\Delta x_p = 2\pi, \tag{3.75}$$

поскольку

$$\sin(x + 2\pi) = \sin x \tag{3.76}$$

и

$$\cos(x + 2\pi) = \cos x. \tag{3.77}$$

Пусть y есть некоторая характеристика, численные значения которой повторяются через каждые 24 часа. Тогда, если x изображает время (в часах), то периодическая функция имеет период $\Delta x_p = 24$.

Обозначим через θ угол (в радианах). Тогда однозначная органическая периодическая функция от θ с периодом 2π может быть представлена бесконечным тригонометрическим рядом в следующем виде:

$$y = f(\theta) = y_m + a_1\cos\theta + a_2\cos(2\theta) + \ldots +$$

$$+ a_n\cos(n\theta) + b_1\sin\theta + b_2\sin(2\theta) + \ldots + b_n\sin(n\theta) + \ldots. \tag{3.78}$$

В соответствии с этим рядом эмпирическая формула периодического характера включает в себя конечное число членов его, определяемое необходимой степенью точности функции.

Если функция y = f (θ) известна, то постоянные ряда могут быть найдены следующим образом:

$$y_m = \frac{1}{2\pi}\int_0^{2\pi} y\, d\theta, \tag{3.79}$$

$$a_k = \frac{1}{\pi}\int_0^{2\pi} y\ \cos(k\theta)\, d\theta, \tag{3.80}$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} y\ \sin(k\theta)\, d\theta, \tag{3.81}$$

где k = 1…n.

Тригонометрический ряд может состоять только из косинусов или же только из синусов угла θ, выраженного в радианах. Для доказательства справедливости этого положения запишем

$$y = a_k \cos(k\theta) + b_k \sin(k\theta) =$$

$$= \sqrt{a_k^2 + b_k^2} \left[\frac{a_k}{\sqrt{a_k^2 + b_k^2}} \cos(k\theta) + \frac{b_k}{\sqrt{a_k^2 + b_k^2}} \sin(k\theta) \right]. \quad (3.82)$$

Обозначим:

$$\sqrt{a_k^2 + b_k^2} = C_k , \quad (3.83)$$

$$\Phi_k = \operatorname{arctg} \frac{b_k}{a_k} . \quad (3.84)$$

тогда

$$y = C_k \left[\cos\Phi_k \cos(k\theta) + \sin\Phi_k \sin(k\theta) \right] = C_k \cos(k\,\theta\Phi_k) . \quad (3.85)$$

Последняя формула является уравнением волны, называемой k-й гармоникой с амплитудой C_k и фазой Φ_k. Период этой волны составляет $\frac{2\pi}{k}$, так как здесь мы имеем

$$\cos\left[k\left(\theta + \frac{2\pi}{k} \right) - \Phi_k \right] = \cos(k\theta + 2\pi - \Phi_k) = \cos(k\theta - \Phi_k) . \quad (3.86)$$

Для k = 1, 2, 3, … графики функции (3.78) носят названия, соответственно, основной волны, второй, третий и т.д. гармоники.

Таким образом, тригонометрический ряд можно записать в виде

$$y = y_m + C_1 \cos(\theta - \Phi_1) + C_2 \cos(2\theta - \Phi_2) + \ldots + C_n \cos(n\theta - \Phi_n) + \ldots . \quad (3.87)$$

Если для периодической функции значения y известны при различных значениях x, то эмпирическая формула, удовлетворяющая экспериментальным данным, может быть найдена путем вычисления y_m, амплитуд и фаз основной и последующих гармоник C_1, Φ_1, C_2, Φ_2, …, C_n, Φ_n.

Переменная x может выражать не только значения угла, но и значения любых других величин. Наиболее часто эта переменная в периодической функции обозначает время.

Для того чтобы выразить значения угла через переменную x, необходимо определить период функции, т.е. Δx_p.

Имеем

$$\theta = mx . \quad (3.88)$$

Выразим θ в градусах, тогда

$$m = \frac{360}{\Delta x_p} , \quad (3.89)$$

или в радианах

$$m = \frac{2\pi}{\Delta x_p} . \quad (3.90)$$

Разделим x на n равных промежутков для всего периода и обозначим ординаты кривой через y_0, y_1, y_2, …, y_n, …, соответственно точкам x_0, x_1, x_2, …, x_n. Тогда y_m будет выражать среднее значение n ординат, а a_k и b_k – удвоенные средние значения произведений, образующихся при умножении каждой ординаты на $\cos(k\theta)$ или $\sin(k\theta)$:

$$y_m = \frac{1}{n} \sum y = \frac{1}{n} (y_0 + y_1 + \ldots + y_{n-1}) , \quad (3.91)$$

$$a_k = \frac{2}{n} \sum \left[y \cos(k\theta) \right] = \frac{2}{n} \left[y_0 \cos(k\theta_0) + y_1 \cos(k\theta_1) + \ldots + y_{n-1} \cos(k\theta_{n-1}) \right], \quad (3.92)$$

$$b_k = \frac{2}{n} \sum \left[y \sin(k\theta) \right] = \frac{2}{n} \left[y_0 \sin(k\theta_0) + y_1 \sin(k\theta_1) + \ldots + y_{n-1} \sin(k\theta_{n-1}) \right] . \quad (3.93)$$

ЗАКЛЮЧЕНИЕ

В первое столетие существования науки "Экология" ее считали гуманитарной общеобразовательной дисциплиной. Однако, в XX веке современное общество столкнулось с чисто техническими проблемами, такими как загрязнение атмосферы промышленными выбросами, загрязнение водоемов сточными водами и образование огромного количества отходов производственного происхождения. Для контроля и анализа состояния окружающей среды человечеству понадобилось использовать специальные методы исследования, позволяющие определить концентрации загрязняющих веществ в различных средах, изучать химические и физические процессы в природе, происходящие под воздействием человека, и прогнозировать их последствия. Эти методы были основаны на инженерных методах, ранее применявшихся для чисто технических целей, таких как методы постановки натурного эксперимента, методы математического и физического моделирования технологических процессов и методы статической обработки результатов измерений. Как оказалось, данные методы прекрасно подходят для использования их в экологических исследованиях. В результате симбиоза инженерной мысли и экологических наук появилась такая отрасль прикладной науки, как "Основы инженерных исследований в экологии".

В данном учебном пособии автор постарался раскрыть основные моменты и особенности применения инженерных методов в процессах исследования состояния окружающей среды. Были определены порядок и методы проведения инженерно-экологического эксперимента, изложены основы математического моделирования инженерно-экологических процессов. Большую часть пособия занял раздел, посвященный статической обработке данных, полученных в ходе инженерно-экологического эксперимента. Наконец, были изучены особенности составления эмпирических формул, что позволяет исследователю-экологу на основе полученных в ходе эксперимента и обработанных статическими методами экспериментальных данных разработать уравнение, определяющее порядок изменения выходных значений интересуемых переменных в зависимости от их входных характеристик.

Таким образом, эколог-исследователь получает мощную базу не только для получения и обработки экспериментальных данных, но и для прогнозирования состояния различных компонентов окружающей среды при изменении методов воздействия на нее человека.

СПИСОК ЛИТЕРАТУРЫ

1. Аленов, М.И. Распознавание природных сред, веществ и их загрязнений / М.И. Аленов, В.Г. Бирюков, В.Н. Иванов. – СПб. : Гидрометеоиздат, 2004. – 268 с.
2. Ахназарова, С.Л. Оптимизация эксперимента в химии и химической технологии : учебное пособие для химико-технологических вузов / С.Л. Ахназарова, В.В. Кафаров. – М. : Высшая школа, 1978. – 319 с.
3. Батунер, Л.М. Математические методы в химической технике / Л.М. Батунер, М.Е. Позин ; под общ. ред. проф. М.Е. Позина. – Изд. 3-е, перераб. и доп. – Л. : ГХИ, 1960. – 636 с.
4. Бертокс, П. Стратегия защиты окружающей среды от загрязнений / П. Бертокс, Д. Радд. – М. : Мир, 1980. – 608 с.
5. Горский, В.Г. Планирование промышленных экспериментов / В.Г. Горский, Ю.П. Адлер. – М. : Металлургия, 1974.
6. Козачек, А.В. Основы инженерных исследований в экологии : программа спецкурса / А.В. Козачек. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2005. – 12 с.
7. Информационный анализ и автоматизированное проектирование станций биохимической очистки : учеб. пособие / Е.Н. Малыгин, Н.С. Попов, В.А. Немтинов, С.Я. Егоров. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2004. – 120 с.
8. Энерго- и ресурсосберегающие технологии и оборудование защиты окружающей среды : учеб. пособие / Н.С. Попов, А.Г. Ткачев, З.А. Михалева, А.И. Попов, Е.А. Сергеева, А.В. Козачек. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2004. – 56 с.
9. Попов, Н.С. Моделирование и управление природо-промышленными системами / Н.С. Попов // Малоотходные и безотходные технологии – главный фактор охраны окружающей среды : тез. докл. Всесоюз. науч.-техн. совещания. – М., 1983. – Ч. 1. – С. 68–69.
10. Проектирование информационно-аналитической системы природопользования / Н.С. Попов, А.Г. Назаров, Н.П. Петрова, А.А. Алексеев // Информационные системы и процессы : сб. науч. тр. / под ред. проф. В.М. Тютюнника. – Тамбов–М.–СПб.–Баку–Вена : Изд-во "Нобелистика", 2003. – Вып. 1. – С. 136 – 152.
11. Пустыльник, Е.И. Статистические методы анализа и обработки наблюдений / Е.И. Пустыльник. – М. : Наука, 1968.
12. Рузинов, Л.П. Планирование эксперимента в химии и химической технологии / Л.П. Рузинов, Р.И. Слободчикова. – М. : Химия, 1980. – 280 с.
13. Теория систем в приложении к проблемам защиты окружающей среды / под ред. С. Ринальди. – Киев : Вища школа, 1981. – 264 с.
14. Тимонин, А.С. Инженерно-экологический справочник / А.С. Тимонин. – Калуга : Изд-во Н.Ф. Бочкаревой, 2003. – 2800 с. (в 3-х томах).
15. Шенк, Х. Теория инженерного эксперимента / Х. Шенк. – М. : Мир, 1972.