Министерство Науки и высшего образования Республики Казахстан

Некоммерческое акционерное общество
«Алматинский университет энергетики и связи имени Гумарбека Даукеева»

Л. К. Ибраева

ТЕОРИЯ И ТЕХНИКА ИНЖЕНЕРНОГО ЭКСПЕРИМЕНТА

Учебное пособие

Алматы 2022

УДК 62(075.8)
ББК 30 я73
И15

Рецензенты:
доктор технических наук, профессор АО «МУИТ»
Б.К. Синчев,
кандидат физико-математических наук, профессор КазНИТУ
им. К.И. Сатпаева
А.Н.Молдагулова,
кандидат физико-математических наук, профессор НАО АУЭС
им. Г. Даукеева
А.К. Искакова

Рекомендовано к печати Ученым советом Алматинского университета энергетики и связи им. Г. Даукеева (Протокол №7 от 12.11.22г.). Печатается по дополнительному плану издания ведомственной литературы АУЭС на 2022 год, позиция 64.

Ибраева Л.К.
И15 Теория и техника инжереного эксперимента: Учебное пособие (для магистрантов образовательной программы 7М07105 – «Автоматизация и управление») / Л.К. Ибраева – Алматы: АУЭС, 2022 – 117 с. табл. – 9, илл. – 10, библиогр. – 15.

ISBN 978-601-358-055-5

В учебном пособии рассматриваются общие вопросы построения регрессионных моделей объектов управления методами планирования экспериментов. Изложены основные предпосылки регрессионного анализа, расчет коэффициентов экспериментальной модели и статистической анализ результатов. Сформулированы основные критерии оптимальности планов регрессионного анализа, проведение экспериментов методами полного факторного и дробного факторного эксперимента, планы второго порядка, а также поисковые методы оптимизации.

Учебное пособие предназначено для магистрантов, обучающихся по образовательной программе 7М07105 – «Автоматизация и управление».

УДК 62(075.8)
ББК 30 я73

ISBN 978-601-358-055-5

©АУЭС 2022
Ибраева Л.К., 2022

Введение

При проектировании, разработке и создании сложных объектов, к которым относятся и объекты управления, требуются знания о количественных и качественных закономерностях, свойственных рассматриваемым объектам. Осуществить практическую проверку тех или иных закономерностей, присущих сложным объектам, не представляется часто возможным по ряду соображений. Кроме того, это потребовало бы больших материальных затрат и затрат времени. В связи с этим приобретает большое значение изучение свойств и закономерностей рассматриваемых сложных объектов на базе *методов моделирования.* Модель – упрощенная система, отражающая отдельные, ограниченные в нужном направлении характеристики рассматриваемого процесса. Важнейшим видом моделирования является *математическое* моделирование – формализованное описание системы с помощью математических соотношений или алгоритмов. Математическая модель позволяет изучать только те параметры оригинала, которые имеют математическое описание и связаны математическими соотношениями в уравнениях, относящихся как к математической модели, так и к оригиналу. При этом физика исследуемого процесса не сохраняется, моделирование здесь основано на способности одних и тех же уравнений описывать различные по своей природе явления и выявлять различные функциональные связи отдельных сторон поведения объекта без полного описания его поведения. То есть, математическая модель реального объекта есть некоторый математический объект, поставленный в соответствие данному физическому объекту.

В зависимости от источника информации, используемого при построении математической модели объекта управления, различают аналитические и статистические (эмпирические) модели.

Аналитические модели представляются в виде сложных систем уравнений (алгебраических, дифференциальных, интегральных или дифференциально-интегральных), позволяющих очень точно описать процессы, протекающие в объекте, и допускающих экстраполяцию в точки факторного пространства, в которых невозможно непосредственное наблюдение этих процессов.

При отсутствии или ограниченном объеме теоретических сведений о моделируемом объекте, когда неизвестен даже ориентировочный вид соотношений, описывающих его свойства, уравнения математического описания могут представлять собой систему эмпирических зависимостей, полученных в результате обследования действующего объекта. То есть статистические модели получают в результате статистической обработки экспериментальной информации, собранной на исследуемом объекте. Эти модели имеют относительно простую структуру и часто представляются в виде полиномов. Область их применения ограничивается ближайшей окрестностью рабочих точек, в которых проводятся эксперименты. Во многих случаях

построение таких моделей можно выполнить при сравнительно небольших затратах времени и средств.

Для получения статистических моделей предназначены методы планирования экспериментов. Теория планирования эксперимента формулирует приемы и способы оптимальной организации экспериментирования при исследовании объектов самой различной физической природы. Математическая теория, лежащая в основе построения различных эмпирических (статистических) моделей – теория вероятностей, изучающая закономерности массовых случайных явлений.

Основная особенность любой экспериментальной модели – то, что подобная модель не может точно описать поведение объекта в любом конкретном опыте. Статистическая модель описывает поведение объекта в среднем, характеризуя неслучайные свойства объекта, которые в полной мере могут проявиться лишь при многократном повторении опытов в неизменных условиях.

1 Статистическое моделирование объектов управления

1.1 Характеристика видов связей между рядами наблюдений

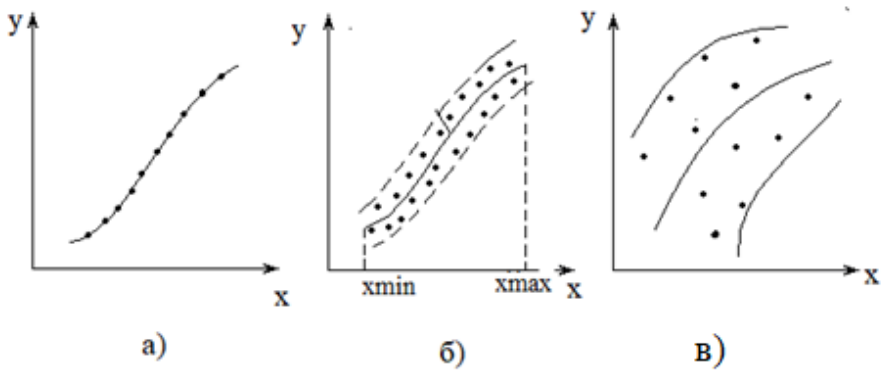
Под экспериментом понимается система операций и воздействий на объект, предназначенных для получения информации об объекте, явлении на основе результатов измерений. Важнейшей задачей методов обработки полученной в ходе эксперимента информации является задача построения математической модели изучаемого явления, процесса, объекта, которую будем представлять в виде некоторого оператора: $y = f(\overline{x})$. Ее можно использовать и при анализе процессов и при проектировании объектов. Другой задачей обработки полученной в ходе эксперимента информации является задача оптимизации, то есть нахождение такой комбинации влияющих независимых переменных, при которой выбранный показатель оптимальности принимает экстремальное значение.

При анализе погрешностей эксперимента и обработке результатов широко используется аппарат математической статистики и теории вероятностей. При рассмотрении основных положений теории вероятностей и математической статистики, определении параметров распределения исходят из предположения, что осуществляется достаточно большое, в пределе – бесконечное число испытаний N, что практически осуществить невозможно. Поэтому обычно используется *выборочный* метод, который позволяют оценить параметры *генеральной совокупности* по выборке ограниченного объема. Необходимость выполнения большого количества измерений вызвана тем, что в объекте присутствуют случайные величины, которые являются функциями других величин. Целью проведения экспериментов является установление вида функциональной зависимости $y = f(\overline{x})$. Для этого должны одновременно определяться как значения вектора $\overline{x}$, так и соответствующие им значения y, а

задачей эксперимента является установление математической модели исследуемой зависимости. Фактически речь идет об установлении *связи* между двумя рядами наблюдений (измерений).

Связи в общем случае являются достаточно многообразными и сложными. Обычно выделяют следующие виды связей:

Функциональные связи (зависимости). Это такие связи, когда при изменении величины $\overline{x}$ другая величина y изменяется так, что каждому значению x_i соответствует совершенно определенное (однозначное) значение y_i (рисунок 1.1, *а*). Таким образом, если выбрать все условия эксперимента абсолютно одинаковыми, то, повторяя испытания, получим одну и ту же зависимость, то есть кривые идеально совпадут для всех испытаний. К сожалению, таких условий в реальности не встречается. На практике не удается поддерживать постоянство условий (например, колебания физико-химических свойств теплоносителя при моделировании процессов тепломассопереноса). При этом влияние каждого случайного фактора в отдельности может быть мало, однако в совокупности они существенно могут повлиять на результаты эксперимента. В этом случае говорят о стохастической (вероятностной) связи между переменными.



а) функциональная связь, все точки лежат на линии; б) связь достаточно тесная, точки группируются возле линии регрессии, но не все они лежат на ней; в) связь слабая

Рисунок 1.1 – Виды связей между рядами наблюдений

Стохастичность связи состоит в том, что одна случайная переменная y реагирует на изменение другой $\overline{x}$ изменением своего *закона распределения* (рисунок 1.1, *б*).

Таким образом, зависимая переменная принимает не одно конкретное значение, а некоторое из множества значений. При повторении испытаний получаются другие значения функции отклика, и одному и тому же значению $\overline{x}$ в интервале $[x_{min}; x_{max}]$ в различных реализациях будут соответствовать различные значения y. Искомая зависимость $\hat{y} = f(x)$ может быть найдена лишь в результате совместной обработки полученных значений $\overline{x}$ и y.

На рисунке 1.1, *б* – это кривая зависимости, проходящая по центру полосы экспериментальных точек (математическому ожиданию), которые

могут и не лежать на искомой кривой $\hat{y} = f(x)$, а занимают некоторую полосу вокруг нее. Эти отклонения вызваны погрешностями измерений, неполнотой модели и учитываемых факторов, случайным характером самих исследуемых процессов и другими причинами.

При анализе стохастических связей можно выделить следующие основные типы зависимостей между переменными.

1. Зависимости между одной *случайной* переменной $\overline{x}$ от другой *случайной* переменной y и их условными средними значениями называются *корреляционными*. Условное среднее $\bar{y}_i$ – это среднее арифметическое для реализации случайной величины y при условии, что случайная величина $\overline{x}$ принимает значение $\overline{x_i}$

2. Зависимость случайной переменной y от неслучайной переменной $\overline{x}$ или зависимость математического ожидания M_y случайной величины y от *детерминированного* значения $\overline{x}$ называется *регрессионной*. Приведенная зависимость характеризует влияние изменений величины $\overline{x}$ на среднее значение величины y.

Стохастические зависимости характеризуются *формой, теснотой связи и численными значениями коэффициентов уравнения регрессии*.

Форма связи устанавливает вид функциональной зависимости $\hat{y} = f(x)$ и характеризуется *уравнением регрессии*. Задача выбора вида функциональной зависимости – задача не формализуемая, так как одна и та же кривая на данном участке примерно с одинаковой точностью может быть описана самыми различными аналитическими выражениями. Отсюда следует важный практический вывод: принятие решения о выборе той или иной математической модели остается за исследователем. Только экспериментатор знает, для чего будет в дальнейшем использоваться эта модель, на основе каких понятий будут интерпретироваться ее параметры.

Под *теснотой связи* понимается степень близости стохастической зависимости к функциональной, то есть это показатель тесноты группирования экспериментальных данных относительно принятого уравнения модели (рисунок 1.1, *б*).

Теория планирования эксперимента формулирует приемы и способы оптимальной организации экспериментирования при исследовании объектов различной физической природы

Эксперимент – это система операций и воздействий на объект, предназначенных для получения информации об объекте на основе результатов измерений. *Опыт* – воспроизведение исследуемого объекта в строго определенных условиях при возможности регистрации результатов (элементарная часть экспериментов).

Планирование эксперимента – рациональная организация измерений и наблюдений. Планирование эксперимента состоит в процедуре выбора числа и условий проведения опытов, необходимых и достаточных для исследования объекта с заданной точностью.

Планирование эксперимента *обеспечивает*:

- одновременное изменение всех переменных по специальным правилам;
- использование математического аппарата, формализующего многие действия экспериментатора;
- выбор четкой стратегии, позволяющей принимать обоснованные решения после каждой серии экспериментов;
- минимизацию числа опытов, ресурсов (финансовых, временных, материальных, человеческих).

Цель планирования эксперимента – нахождение таких условий и правил проведения опытов, при которых удается получить надежную и достоверную информацию об объекте с наименьшей затратой труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности.

1.2 Классификация видов экспериментальных исследований

По форме представления результатов выделяются следующие виды экспериментов:

- *качественный эксперимент*. Устанавливается факт существования каких-либо явлений, но количественных характеристик при этом не дается. Любой эксперимент, каким бы сложным он ни казался, заканчивается представлением результатов, формулировкой выводов, выдачей рекомендаций. Эта информация может быть представлена в виде графиков, чертежей, таблиц, формул, статистических данных или словесных описаний. Качественный эксперимент, как правило, предусматривает именно словесное описание. Однако словесное описание – самый неэффективный способ представления результатов, поскольку не позволяет дать количественные рекомендации, анализировать свойства объекта в иных условиях, решать задачи его управления. В инженерной практике основное содержание эксперимента должно представляться числом или количественными зависимостями;
- *количественный эксперимент*. Позволяет не только фиксировать существование того или иного явления, но и устанавливать количественные взаимосвязи между факторами, определяющими протекание процесса, а также устанавливать математическую модель влияния этих факторов на то или иное явление.

По условиям проведения различают:

- *лабораторный эксперимент*. В лаборатории меньше влияние случайных погрешностей, обеспечивается большая «стерильность» условий проведения опытов, осуществляется в большинстве случаев и более тщательная подготовка, одним словом, выше «культура эксперимента». Как правило, в лабораторных условиях экспериментатор может воспроизвести опыт «одинаково» значительно лучше, чем в промышленности. Это означает, что при прочих равных условиях для установления некоторого факта на заводе потребуется выполнить значительно больше опытов, чем в лаборатории. Другое важное отличие – различные ограничения на возможности

варьирования факторами. Когда в лаборатории исследуется химическая реакция, температуру по желанию можно менять в широких пределах, а в металлургических печах, напротив, если ее и можно менять, то в значительно более узком диапазоне и с большей осторожностью;

- *в промышленных экспериментах* эти условия обеспечить значительно сложнее. Усложняются измерения и сбор информации, значительно больше влияние различного рода помех на организацию и проведение эксперимента, измерительные приборы, поэтому особенно необходимо использовать специальные методы. Требуется по возможно меньшему числу измерений получить наиболее достоверные результаты. Заметим, что в современной математической статистике имеются специальные методы, которые при том же количестве измерений позволяют повысить точность или даже при их уменьшении получить более представительную информацию.

По виду воздействий на объект различают:

- *активный эксперимент* – это такой эксперимент, когда исследователь целенаправленно изменяет условия функционирования процесса и наблюдает результаты. Такой эксперимент можно планировать;

- *пассивный эксперимент* – это эксперимент, в котором исследователь регистрирует условия функционирования процесса в каждом опыте, но не задает его значение. В дальнейшем он обрабатывает результаты такой регистрации и пытается давать рекомендации.

1.3. Погрешности результатов исследований

Результаты опытов обычно не являются точными. По различным причинам результаты любых двух параллельных опытов отличаются друг от друга, за исключением случайных совпадений. Под точностью эксперимента понимают его качество, отражающее близость полученных результатов к истинному значению искомой величины. Точность эксперимента тем выше, чем меньше его погрешность. Виды погрешностей приведены ниже.

Абсолютная погрешность – это разность Δx между результатом эксперимента x и истинным значением искомой величины x^*:

$$\Delta = |x - x^*|.$$

Относительная погрешность:

$$\Delta x^* = \frac{|x - x^*|}{x^*} \cdot 100\% = \frac{\Delta}{x^*} \cdot 100\%.$$

Следует заметить, что истинное значение величины, определяемой в результате эксперимента, всегда остается неизвестным, поэтому и погрешности эксперимента могут быть оценены лишь приближенно.

Приведенной погрешностью называют отношение абсолютной погрешности Δx к нормирующему значению $x_н$, выраженному в процентах:

$$\gamma = \frac{\Delta x}{x_{\text{н}}} \cdot 100\%.$$

В качестве нормирующего значения используют условно принятое значение измеряемой величины, выраженное в тех же единицах, в качестве которых, как правило, используют абсолютные значения разности верхнего и нижнего пределов шкалы.

При проведении эксперимента его погрешности по причинам возникновения принято условно разделять на систематические, случайные и грубые (промахи).

Систематической называется погрешность, которая при повторных экспериментах остается постоянной или изменяется закономерно. Наличие систематических погрешностей может быть обнаружено путем анализа условий измерения одного и того же значения измеряемой величины разными методами или приборами. Примером переменной систематической погрешности может быть погрешность от измерения (закономерного) напряжения источника питания, если результат измерения зависит от напряжения (например, потенциометр). Систематические погрешности нельзя уменьшить увеличением числа параллельных опытов. Должны устраняться вызывающие их причины. Общим методом выявления причин систематических погрешностей является калибровка (поверка), которая представляет собой поверку прибора во всем диапазоне измеряемой величины с помощью известного эталона. Прибор может давать очень малый разброс показаний, но результат будет неверным вследствие наличия систематической ошибки.

Можно выделить следующие источники систематических погрешностей:

– *инструментальными* (приборными или аппаратурными) погрешностями средств измерений называются такие, которые принадлежат данному средству измерений, они могут быть определены при его испытаниях и занесены в его паспорт. Принято различать *основную погрешность* средств измерений, то есть погрешность в условиях, принятых за нормальные, и *дополнительную погрешность*, вызванную отклонением влияющих параметров за пределы области нормальных значений (вибрации, влажности среды, инерцией и т. п.);

– *методические* погрешности – это погрешности, которые не могут быть приписаны данному прибору, не смогут быть указаны в его паспорте, то есть связаны не с самим прибором, а с методикой проведения измерений. Очень часто причиной возникновения методической погрешности является то, что, организуя измерения, измеряют или вынуждены измерять не ту величину, которую в принципе требуется измерять, а некоторую другую, близкую, но не равную ей.

Отличительной особенностью методических погрешностей является то, что они могут быть определены лишь путем создания математической модели исследуемого объекта и не смогут быть найдены сколь угодно тщательным исследованием лишь самого измерительного прибора:

- *субъективные* погрешности, обусловленные особенностями исследователя.

Следует иметь в виду, что полностью исключить систематические погрешности невозможно, так как методы и средства, с помощью которых обнаруживаются и оцениваются систематические погрешности, сами имеют свои погрешности.

Случайной называется погрешность, обусловленная действием ряда причин, меняющихся случайным образом от эксперимента к эксперименту. Значение этой погрешности не может быть определено в каждом эксперименте, и на нее невозможно оказать влияние. В то же время в результате большого числа экспериментов могут быть выявлены некоторые закономерности, присущие этому типу погрешностей. К случайным относятся непостоянные погрешности, причины возникновения которых неизвестны. Эти погрешности, как правило, вызываются сложной совокупностью изменяющихся факторов, обычно неизвестные экспериментатору и трудно поддающиеся анализу. Таким образом, случайные погрешности представляют собой беспорядочные флуктуации показаний прибора относительно истинного значения измеряемой величины. Для исследования случайных погрешностей, возникающих при проведении эксперимента, широко используются математическая статистика и теория вероятностей.

Грубые погрешности (промахи) возникают вследствие непредвиденного изменения условий эксперимента, качества измерений, поломок прибора, неправильной записи в рабочих журналах, механических ударах прибора, неправильном отчете показаний прибора, отключении источника питания и т.п. Результат, содержащий грубую ошибку, резко отличается по величине от остальных измерений. Такие результаты должны быть исключены из рассмотрения до математической обработки результатов эксперимента.

1.4 Контрольные вопросы

1.4.1 Дайте определение случайной величины.

1.4.2 Дайте определения следующим понятиям: закон распределения вероятностей, интегральная функция распределения вероятностей, дифференциальная функция распределения.

1.4.3 Что такое генеральная совокупность? Дайте определение выборки.

1.4.4 Назначение выборочного метода?

1.4.5 Дайте определение корреляционной связи.

1.4.6 Что такое регрессионная зависимость?

1.4.7 Приведите классификацию видов экспериментальных исследований.

1.4.8 Дайте определение случайной и систематической погрешности.
1.4.9 Назовите источники систематических погрешностей.
1.4.10 Чем обусловлено появление случайных погрешностей измерения?

2 Основные понятия теории планирования экспериментов

2.1 Определения

На этапе формулирования цели исследований используем кибернетический подход, в основу которого положена идея «черного ящика». Термин «черный ящик» означает модель мало или плохо изученных свойств реальных объектов, в которой входные независимые переменные воздействуют на каждую из выходных зависимых переменных. При этом существует вполне определенная связь между значениями входных переменных и значениями выходных переменных объекта (Рисунок 2.1). Считается, что исследователь может наблюдать входы и выходы «черного ящика» и по результатам наблюдений определять зависимость между входами и выходами. Входы могут меняться детерминировано (контролируемые, управляемые) или случайно (контролируемые, но не управляемые).

Комплекс параметров $\bar{x} = (x_1, \dots, x_n)$ называется основным, он определяет условия эксперимента. В качестве случайных параметров рассматриваются обычно параметры, которые по тем или иным причинам невозможно учесть (или очень трудно). Такое подразделение на основные и случайные параметры условно. Случайным будет любой параметр, не вошедший в основной комплекс, даже если он хорошо изучен.

Входные переменные $x_1 \dots x_n$ называются *факторами*.

Выходная переменная – зависимая *y* называется *откликом* (реакцией). То есть, отклик (реакция на факторы) – наблюдаемая случайная переменная, по предположению, зависящая от факторов.

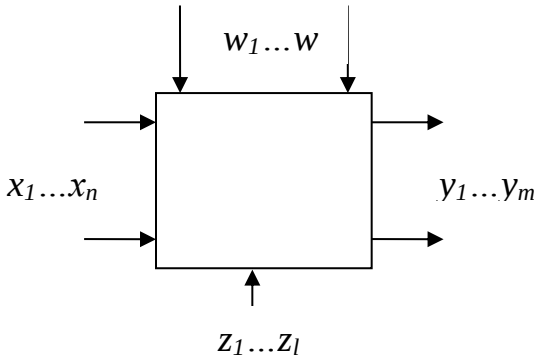


Рисунок 2.1 – Структурная схема объекта исследования

На этой схеме приняты следующие обозначения:

$\bar{x} = (x_1, \dots, x_n)$ – входные контролируемые и управляемые переменные объекта;

$\bar{z} = z_1, \ldots, z_l)$ – входные контролируемые, но не управляемые переменные;

$\bar{w} = (w_1, \ldots, w_k)$ – неконтролируемые, случайным образом изменяющиеся переменные, «шум» объекта;

$\bar{y} = (y_1, \ldots, y_m$ – наблюдаемые выходы объекта.

Обычно переменные $\bar{z}$ и $\bar{w}$ объединяются в один вектор $\bar{e}$, под влиянием которого выход объекта может меняться достаточно систематически (ситуация, когда имеет место «дрейф объекта») или случайным образом – в этом случае говорят о наличии помехи, случайного шумового поля. При этом, как правило, считается, что вероятностные свойства шумового поля не меняются во времени.

По физическому происхождению к переменным $\bar{e}$ прежде всего относятся ошибки измерительных приборов или методов анализа, а также неконтролируемые изменения сырья в промышленных объектах или изменение свойств вследствие старения и износа установок, влияние внешней среды и т.д., включая воздействие тех переменных, которые в принципе могли бы контролироваться, но не включены в число исследуемых переменных (вследствие трудностей их измерения, по ошибке ими незнанию). К переменным группы $\bar{e}$ должны быть отнесены и такие, как квалификация исследователя, особенности используемого экспериментального оборудования и методологии измерений, уровень метрологического обеспечения, то есть переменные, действие которых в данном конкретном эксперименте могут и не проявиться, но которые могут сказаться на стадии сопоставления результатов различных экспериментаторов, лабораторий и т.д.

Большинство методов планирования экспериментов развито для случая, когда размерность вектора $\bar{y}$ равна 1 и влияние переменных $\bar{e}$ (шумового поля) можно представить в виде аддитивной составляющей измеряемого значения y, то есть схему объекта представить в виде эквивалентной схемы, где $\tilde{y}$ – выход объекта без ошибок (рисунок 2.2).

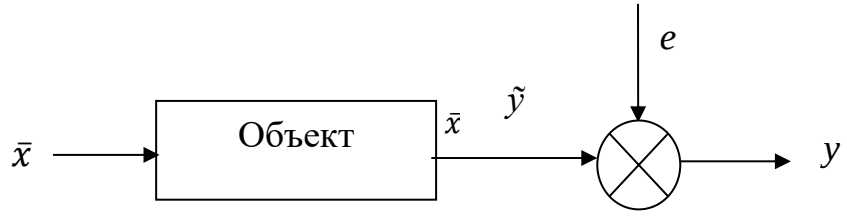


Рисунок 2.2 – Эквивалентная схема объекта исследования с аддитивной помехой e

Объектами исследования в практике планирования экспериментов могут быть: реальные физические объекты, физические модели реальных объектов, математические модели реальных объектов (в двух последних случаях в помеху $\bar{e}$ войдут также составляющие, обусловленные неточностью физического или математического моделирования объекта).

Отклик должен быть определен количественно. Однако могут встречаться и качественные признаки *y*. В этом случае возможно применение рангового подхода. Пример рангового подхода – оценка на экзамене, когда одним числом оценивается сложный комплекс полученных сведений о знаниях студента.

Независимые переменные $x_1, x_2, \ldots, x_n$ – факторы, также должны иметь количественную оценку. Если используются качественные факторы, то каждому их уровню должно быть присвоено какое-либо число. Важно выбирать в качестве факторов лишь независимые переменные, т. е. только те, которые можно изменять, не затрагивая другие факторы. Факторы должны быть однозначными. Для построения эффективной математической модели целесообразно провести предварительный анализ значимости факторов (степени влияния на функцию), их ранжирование и исключить малозначащие факторы.

Фактор – некоторая переменная величина, принимающая в каждый момент времени определенное значение из своей области определения и отражающая внешнее воздействие на объект.

Значения, которые факторы принимают в эксперименте, называются уровнями факторов.

Уровень фактора – конкретное значение фактора из его области определения при экспериментальном исследовании объекта.

В процессе эксперимента определенное сочетание конкретных значений входных факторов соответствует одному состоянию объекта. На основе анализа задачи исследования и априорной информации об исследуемом объекте для каждого входного фактора выделяется область, в пределах которой он будет изменяться во время эксперимента. Сочетание таких областей по всем входным факторам будем называть факторным пространством эксперимента.

Факторное пространство – это множество факторов, значения которых исследователь может контролировать в ходе подготовки и проведения эксперимента.

Целью эксперимента является нахождение функции *y*, при этом предполагается, что значение отклика складывается из двух составляющих:

$$y = f(x_1, x_2, \ldots, x_n) + e(x_1, x_2, \ldots, x_n), \qquad (2.1)$$

где $f(x_1 \ldots x_n)$ – функция отклика (неслучайная функция факторов);

$e\ (x_1 \ldots x_n)$ – ошибка эксперимента (случайная величина);

$x_1 \ldots x_n$ – определенное сочетание уровней факторов из факторного пространства.

Очевидно, что *y* является случайной переменной, так как зависит от случайной величины $e\ (x_1 \ldots x_n)$.

Дисперсия $D[y]$, которая характеризует точность измерения, равна дисперсии ошибки опыта $D[y]=D[e]$.

Функция отклика – зависимость математического ожидания отклика от факторов.

Поверхность отклика – геометрическое представление функции отклика.

Надо отметить, что выбор структуры модели (2.1) выполняется на начальном этапе, до проведения экспериментов. На этом этапе выполняется *априорное ранжирование факторов* – метод выбора наиболее важных факторов, основанный на экспертной оценке.

В теории планирования экспериментов рассматриваются вопросы, связанные с активным экспериментом.

План эксперимента – совокупность данных, определяющих число, условия и порядок проведения опытов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям. То есть это целенаправленное управление, которое реализуется в условиях неполного знания механизма изучаемого явления.

Область действия – область возможных значений факторов $\bar{x}$ при эксперименте.

Область планирования – область значений $\bar{x}$, в которых находятся точки, отвечающие условиям проведения опытов используемого плана эксперимента (это часть области действия или совпадает с ней).

2.2 Выбор уровней и интервалов варьирования факторов. Понятие воспроизводимости эксперимента

Планирование эксперимента начинают с выбора основного уровня каждого входного фактора.

Основной уровень – уровень, который представляет наибольший интерес в данный момент – вектор $\bar{x}_0$. В его окрестности и располагаются все точки плана. Вектор $\bar{x}_0$задает центр области планирования (центр эксперимента).

Построение плана эксперимента – это выбор точек (уровней входных факторов) относительно основного (иногда называемого нулевым) уровня. Для определения других уровней входных факторов вводится *интервал варьирования* каждого входного фактора. Чтобы определить верхний уровень входного фактора, следует интервал варьирования прибавить к нулевому уровню данного фактора, а чтобы определить нижний уровень – вычесть интервал варьирования из нулевого уровня. На интервал варьирования накладываются ограничения естественного характера снизу и сверху.

К интервалу варьирования входного фактора предъявляются следующие требования:

- он не может быть менее ошибки, с которой измеряется данный фактор, иначе уровни фактора будут неразличимы;

- он не может быть слишком большим, т. е. нижние и верхние уровни не должны покидать области определения фактора и области проведения эксперимента.

Факторы могут отличаться по числу уровней. Диапазоны изменения факторов задают область определения *y*. Диапазон изменения уровней факторов определяется исходя из конкретных условий эксперимента.

Если принять, что каждому фактору соответствует координатная ось, то полученное пространство будет факторным пространством, которое при *n=2* представляет собой прямоугольник, при *n=3* – куб, при *n >3* – гиперкуб.

Область планирования может задаваться *интервалами* возможного изменения факторов $x_{imin} \le x_i \le x_{imax}$,*i=1,…,n.* Характерные точки области планирования – x_{imin}, x_{imax}, основной уровень – x_{i0}.

Обозначим интервал (шаг) варьирования фактора x_i : Δx_i.

Тогда $x_{imax} = x_{i0} + \Delta x_i$; $x_{imin} = x_{i0} - \Delta x_i$.

Обычно

$$x_{i0} = \frac{x_{imin} + x_{imax}.}{2}$$

Интервал (шаг) варьирования фактора x_i:

$$\Delta x_{i0} = \frac{x_{imax} - x_{imin}.}{2}$$

Размах варьирования фактора – разность между максимальным и минимальным натуральными значениями фактора в данном плане. Тогда интервал варьирования фактора равен половине размаха варьирования фактора.

Обычно при первичном планировании эксперимента количество уровней по всем входным факторам выбирают одинаковым. Тогда количество опытов в эксперименте (N_{exp}) может быть определено по формуле:

$$N_{exp} = p_{exp}^k,$$

где p_{exp} – число уровней каждого входного фактора;

k – число входных факторов, исследуемых в эксперименте

Если влияние некоторого фактора на наблюдаемую переменную изменяется при изменении значения некоторого другого фактора, говорят, что между факторами существует *взаимодействие*. *Эффект взаимодействия факторов* – показатель зависимости изменения эффекта одного фактора от уровней других факторов.

Суть анализа заключается в разложении общей вариации случайной величины на независимые слагаемые – *эффекты*, каждый из которых характеризует влияние того или иного фактора (*главный эффект*) или их

взаимодействия (*эффект взаимодействия*). Слагаемое, которое характеризует влияние отдельного фактора, называется *главным эффектом*, которое характеризует взаимодействие факторов – *эффект взаимодействия.*

Факторы – размерные величины. Размерности факторов могут быть разными и числа, выражающие ~~их~~ значения факторов, могут иметь различный порядок. Поэтому для того, чтобы построить матрицу плана эксперимента и упростить обработку полученных результатов, необходима предварительная *нормализация* факторов, представляющая собой линейное преобразование факторного пространства. Такое преобразование состоит в выборе нового начала координат (центрирование) и изменении масштаба по координатным осям (масштабирование). С этой целью делают преобразование начала координат факторов и переходят к нормированному (стандартному) масштабу:

$$x_i = \frac{x_i - \tilde{x}_{i0}}{\Delta\tilde{x}_i}, i = 1,2,\ldots,n,$$

где x_i – нормированное значение;

$\tilde{x}_i$ – натуральное значение;

$\tilde{x}_{i0}$ – основной уровень;

$\Delta\tilde{x}_i$ – интервал варьирования.

Здесь начало координат – центр эксперимента, единица измерения – шаг варьирования факторов. Такое преобразование приводит к *безразмерным* величинам и упрощает математические выкладки.

Основная цель теории планирования эксперимента – построение математической модели исследуемого в эксперименте процесса. Чтобы воспользоваться методами теории планирования эксперимента, нужно сначала убедиться, что эксперимент воспроизводим.

Под *воспроизводимыми* экспериментами понимаются такие, в процессе которых в любой момент времени объект исследования и измерительное оборудование можно вернуть в исходное состояние и эксперимент повторить на одинаковом уровне случайных ошибок.

Для компенсации влияния случайных погрешностей каждый эксперимент рекомендуется повторить n раз. Эксперименты, повторенные несколько раз при одних и тех же значениях факторов, называют *параллельными*. Под *дублированием* понимают постановку параллельных экспериментов. Обычно число n параллельных экспериментов принимают равным 2-3, иногда – 4-5. При проведении исследований приходится иметь дело с тремя вариантами дублирования экспериментов: 1) с равномерным дублированием экспериментов; 2) с неравномерным дублированием экспериментов; 3) без дублирования экспериментов. При *равномерном дублировании* все строки матрицы планирования имеют одинаковые числа параллельных экспериментов. В случае *неравномерного дублирования* числа параллельных экспериментов неодинаковы. При отсутствии дублирования параллельные эксперименты не проводятся. Наиболее предпочтительным из

трех вариантов дублирования является первый. При этом варианте эксперимент отличается повышенной точностью, а математическая обработка экспериментальных данных – простотой. Характер дублирования влияет на содержание математической обработки результатов наблюдений.

Для проверки воспроизводимости эксперимента проводят несколько серий параллельных опытов.

Если опыты невоспроизводимы, то можно попытаться достигнуть воспроизводимости выявлением и устранением источников нестабильности эксперимента, а также использованием более точных методов и средств измерений. Если же никакими способами невозможно достигнуть воспроизводимости, то математические методы планирования к такому эксперименту применять нельзя.

В одинаковых условиях эксперимента, чем больше число параллельных опытов, тем более надежны полученные оценки дисперсии воспроизводимости. Однако с увеличением числа параллельных опытов возрастает число опытов плана. Поэтому, если обеспечивается требуемая достоверность оценок дисперсии воспроизводимости, то выбирается минимальное число параллельных опытов, которое обычно равно 2 или 3.

План эксперимента может быть задан либо матрицей плана, либо матрицей спектра плана вместе с матрицей дублирования.

Матрица плана – стандартная форм записи условий проведения эксперимента в виде матрицы, строки – опыты, столбцы – факторы. *(i,j)*-ый элемент – уровень *j*-го фактора в *i*-том опыте. Может иметь совпадающие строки.

Матрица спектра плана – матрица, состоящая из всех строк и матрицы плана, отличающихся уровнем хотя бы одного фактора. Все строки этой матрицы различны.

Матрица дублирования – квадратная матрица, диагональные элементы которой равны числу параллельных опытов в соответствующих точках спектра плана.

Под дублированием понимается не серия измерений в одном опыте, а полное повторение каждого опыта.

Таким образом, *план эксперимента* – совокупность точек $x_1, x_2, \ldots, x_N$, где $\bar{x}$ – вектор входа, N – число наблюдений с числом $m_1, m_2, \ldots, m_N$ наблюдений в каждой точке, то есть матрица спектра плана:

$$X = \begin{pmatrix} x_{11}x_{12}\ldots x_{1n} \\ \ldots\ldots\ldots\ldots\ldots \\ x_{N1}x_{N2}\ldots x_{Nn} \end{pmatrix}$$

совместно с матрицей дублирования экспериментов:

$$\begin{pmatrix} m_1 & 0 \ldots\ldots\ldots 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ 0 \ldots\ldots\ldots\ldots & m_N \end{pmatrix},$$

причем:

$$\sum_{g=1}^{N} m_g = N.$$

Для того чтобы в известной мере компенсировать систематические погрешности эксперимента, используют прием, называемый *рандомизацией*. Он заключается в том, что опыты проводят в случайной последовательности, которая устанавливается с помощью таблицы случайных чисел.

Параллельные опыты – рандомизированные во времени опыты, в которых уровни всех факторов сохраняются неизменными.

Пусть, например, требуется рандомизировать во времени 6 опытов, обозначенных цифрами I, II, …, VI. Поставим им в соответствие любые 6 последовательных чисел, взятых в любой строке или любом столбце таблицы случайных чисел. Если при этом встретятся повторяющиеся числа, то их следует отбросить. Например, могут быть получены следующие пары: I–60; II–12; III–05; IV–15; V–34; VI–30. Расположив случайные числа в порядке возрастания (или убывания), получим искомую последовательность реализации опытов: III, II, IV, VI, V, I (или I, V, VI, IV, II, III).

Эксперименты и экспериментаторы могут отличаться друг от друга, но фактически планирование, проведение и анализ всех экспериментов осуществляется в одинаковой последовательности. Хотя объекты исследований различны, однако методы экспериментальных исследований имеют много общего:

– каким бы простым эксперимент ни был, вначале составляется план его проведения;

– стремятся уменьшить число переменных, поскольку это упрощает работу и делает ее более экономичной;

– стараются исключить влияние внешних случайных факторов;

– необходимо контролировать ход эксперимента;

– обращается внимание на точность измерительных приборов и точность получения данных;

– в процессе любого эксперимента необходимо анализировать получаемые результаты и давать их интерпретацию, поскольку без этого решающего этапа весь процесс не имеет смысла.

Планирование экспериментов преследует 2 основные цели:

- сокращение общего объёма испытаний при соблюдении требований к достоверности и точности их результатов;

- повышение информативности каждого из экспериментов в отдельности.

2.3 Контрольные вопросы

2.3.1 В чем отличие количественного и качественного экспериментов?

2.3.2 В чем заключаются принципиальные отличия активного эксперимента от пассивного?

2.3.3 Что такое факторное пространство?

2.3.4 Перечислите требования к управляемым факторам.

2.3.5 Что такое функция отклика? Почему эта функция случайная?

2.3.6 Дайте определение уровня фактора.

2.3.7 Какие требования предъявляются к интервалу варьирования фактора?

2.3.8 Дайте определение матрицы плана, матрицы спектра плана и матрицы дублирования.

2.3.9 Что такое параллельные опыты?

2.3.10 Дайте определение воспроизводимого эксперимента.

3. Статистическое оценивание

3.1 Свойства оценок параметров случайных величин

При решении многих прикладных задач, связанных с обработкой экспериментальных данных, исследователю приходится обрабатывать результаты, имеющие заведомо вероятностный характер. Однако необходимые характеристики соответствующих случайных величин часто неизвестны и должны находиться по экспериментальным значениям. Подобное статистическое описание результатов наблюдений, построение и проверка различных математических моделей, использующих понятие вероятности, составляют основное содержание математической статистики. Методы обработки экспериментальных данных опираются на базовые понятия математической статистики. К их числу относятся понятия *генеральной совокупности, выборки, эмпирической функции распределения.*

Генеральная совокупность – совокупность всех мыслимых результатов наблюдений над случайной величиной, которые могут быть в принципе проведены при данных условиях. То есть под *генеральной совокупностью* понимают все возможные значения параметра, которые могут быть зарегистрированы в ходе неограниченного по времени наблюдения за объектом. Такая совокупность состоит *из бесконечного множества* элементов. В результате наблюдения за объектом формируется ограниченная по объему совокупность значений параметра $x_1, x_2, \dots, x_n$.

Выборка – это конечный набор значений случайной величины, получаемый в ходе наблюдения. Число элементов выборки называется ее *объемом*.

Для того чтобы по результатам наблюдения можно было делать какие-либо выводы, выборка должна быть *репрезентативной* (представительной), то

есть правильно представлять пропорции генеральной совокупности. Это требование выполняется, если объем выборки достаточно велик, а каждый элемент генеральной совокупности имеет одинаковую вероятность попасть в выборку.

Выборка называется *репрезентативной* (представительной), если она дает достаточное представление об особенностях генеральной совокупности.

Если о генеральной совокупности ничего неизвестно, то единственной гарантией репрезентативности может служить случайный отбор, так называемая *рандомизация* (приведение к случайности). При таком выборе каждая возможная выборка фиксированного объема имеет одну и ту же вероятность выбора, а последовательные наблюдения независимы.

Назначение статистических методов заключается в том, чтобы по выборке ограниченного объема N, то есть по некоторой части генеральной совокупности, высказать обоснованное суждение о ее свойствах в целом. Подобное суждение может быть сформировано путем *оценивания* выборочных параметров (характеристик) генеральной совокупности с помощью каких-либо подходящих *функций от наблюдений*.

Функция $G = g(x_1, x_2, \dots, x_n)$ от элементов выборки из рассматриваемой генеральной совокупности называется *статистикой*. Фиксированная статистика, подходящая для оценивания того или иного параметра генеральной совокупности, называется *статистической оценкой* (выборочным значением) данного параметра.

Произвольная статистика сама является случайной величиной и как любая случайная величина может быть описана с помощью плотности вероятности или числовых характеристик. В общем случае она зависит от параметров генеральной совокупности и объема выборки N. Любая оценка параметра генеральной совокупности является случайной величиной. Свойства (вероятностные) оценки параметра могут быть определены также с помощью выборочной функции распределения и ее характеристик.

Для параметров генеральной совокупности и их оценок вводятся разные обозначения: оценка параметра θ обозначается $\hat{\theta}$. Если параметры генеральной совокупности обозначаются греческими буквами (например, дисперсия σ_x^2, коэффициент корреляции ρ и т. д.), соответствующие выборочные значения обозначаются латинскими буквами (S_x^2, r и т. д.). Для оценки математического ожидания m_x общепринято специальное обозначение $\bar{x}$.

Значение параметра, вычисленное по ограниченному объему экспериментальных данных, является случайной величиной, т. е. значение такой величины от выборки к выборке может меняться заранее непредвиденным образом. Следовательно, в результате обработки экспериментальных данных определяется не значение параметра θ, а только лишь его приближенное значение $\hat{\theta}$ – статистическая *оценка* параметра.

Получить статистическую оценку параметра теоретического распределения означает найти функцию от имеющихся результатов наблюдения, которая и даст приближенное значение искомого параметра.

Различают два вида оценок – *точечные и интервальные*.

Точечными называют такие оценки, которые характеризуются одним числом. При малых объемах выборки точечные оценки могут значительно отличаться от истинных значений параметров, поэтому их применяют при большом объеме выборки.

Интервальные оценки задаются двумя числами, определяющими вероятный диапазон возможного значения параметра. Эти оценки применяются для малых и для больших выборок.

Рассмотрим вначале точечные оценки.

Для оценивания одного и того же параметра можно использовать различные оценки (статистики). Чтобы выбрать наилучшую из них, необходимо сформулировать некоторые требования к свойствам оценок, желательные с точки зрения практики.

Качество оценок характеризуется такими свойствами, как *состоятельность, несмещенность, эффективность и достаточность*.)

Оценка $\hat{\theta}$ параметра θ называется *состоятельной*, если с увеличением объёма выборки значение оценки стремится к значению параметра генеральной совокупности по вероятности, то есть:

$$\hat{\theta} \to \theta, \text{когда } N \to \infty.$$

Это означает, что с ростом N выборочное распределение все в большей степени концентрируется вокруг θ и точность оценки безгранично возрастает.

В частности, для состоятельной оценки выполняется условие:

$$\lim_{N\to\infty} lim\, D\,\{\hat{\theta}(X_1, X_2, \ldots, X_N)\} \to 0,$$

Оценка $\hat{\theta}$ параметра θ называется *несмещенной*, если для любого объема N выборки математическое ожидание оценки $\hat{\theta}$ равно оцениваемому параметру θ:

$$m_{\hat{\theta}} = M\{\hat{\theta}) = \theta.$$

Удовлетворение требования несмещенности устраняет систематическую погрешность оценки параметра, которая зависит от объема выборки N и в случае состоятельности стремится к нулю при $N\to\infty$. Несмещенность характеризует отсутствие систематических (в среднем) отклонений оценки от параметра при любом конечном, в том числе и малом, объеме выборки. Использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, приводит к систематическим ошибкам. Но даже несмещенная оценка может быть удалена от истинного значения.

Несмещенная оценка $\hat{\theta}_{эф}$ называется *эффективной*, если среди всех оценок параметра она обладает наименьшей дисперсией:

$$D\{\hat{\theta}_{эф}\} = \sigma^2_{\hat{\theta}_{эф}} = M\{(\hat{\theta} - \theta)^2\} = min.$$

Эффективность характеризует разброс случайных значений оценки около истинного значения параметра. Среди всех оценок следует выбрать ту, значения которой теснее сконцентрированы около оцениваемого параметра.

Все сказанное относится к равноточным измерениям, т. е. к измерениям, которые содержат только случайную погрешность, подчиняющуюся нормальному закону распределения.

Сущность задачи точечного оценивания параметров. Точечная оценка предполагает нахождение единственной числовой величины, которая и принимается за значение параметра. Такую оценку целесообразно определять в тех случаях, когда объем экспериментальных данных достаточно велик. Причем не существует единого понятия о достаточном объеме экспериментальных данных, его значение зависит от вида оцениваемого параметра (к этому вопросу вернёмся при изучении методов интервальной оценки параметров, а предварительно будем считать достаточной выборку, содержащую не менее чем 10 значений). При малом объеме экспериментальных данных точечные оценки могут значительно отличаться от истинных значений параметров, что делает их непригодными для использования.

Задача точечной оценки параметров в типовом варианте постановки состоит в следующем: имеется выборка наблюдений ($x_1, x_2, \ldots, x_N$) за случайной величиной X. Выборка должна быть представительной. Объем выборки N фиксирован. Известен вид закона распределения величины X, например, в форме плотности распределения $f(x,\theta)$, где θ – неизвестный (в общем случае векторный) параметр распределения. Параметр является неслучайной величиной. Требуется найти оценку $\hat{\theta}$ параметра θ закона распределения.

Существует несколько методов решения задачи точечной оценки параметров, самыми распространёнными из которых являются методы *максимального (наибольшего) правдоподобия, моментов и квантилей.*

3. 2 Метод максимального правдоподобия

Свойства оценок различных параметров θ во многом определяются видом закона распределения исследуемой генеральной совокупности, а также способом оценивания. Известны общие методы, позволяющие во многих случаях решить задачу нахождения по экспериментальным данным наилучших оценок неизвестных параметров. Метод максимального правдоподобия является наиболее важным общим методом оценивания неизвестных параметров по экспериментальным данным.

Метод был предложен Р. Фишером в 1912 г. Метод основан на исследовании вероятности получения выборки наблюдений $(x_1, x_2, \ldots, x_n)$ за некоторым количественным признаком X. Значения выборки случайные,

поэтому их рассматривают как независимые случайные величины, имеющие одинаковые распределения. Суть метода максимального правдоподобия заключается в свойстве случайной величины реализовывать в эксперименте в основном те свои значения, вероятность которых максимальна. В этом случае совместная плотность распределения вероятностей N случайных величин должна быть максимальной.

Пусть $x_1, x_2, \ldots, x_N$ – выборка из генеральной совокупности с плотностью вероятности $f_x(x;\theta)$, зависящей от постоянного неизвестного параметра θ. Элементы выборки статистически независимы, поэтому можем записать выборочную плотность вероятности для случайной выборки *N:*

$$f_{x_1,x_2,\ldots,x_N}(x_1, x_2, \ldots, x_N;\theta) = f_{x_1}(x_1;\theta) f_{x_2}(x_2;\theta) \cdot \ldots \cdot f_{x_N}(x_N;\theta). \qquad (3.1)$$

Тогда плотность вероятности, полученная при подстановке выборочных значений $x_1, x_2, \ldots, x_N$ в плотность вероятности (3.1), называется *функцией правдоподобия* $L(\theta)$ для параметра θ:

$$L(\theta) = f_{x_1}(x_1;\theta) f_{x_2}(x_2;\theta) \cdot \ldots \cdot f_{x_N}(x_N;\theta).$$

До эксперимента функция правдоподобия выражает предпочтительность различных значений θ при конкретной выборке. После того, как произведен эксперимент, значения выборки известны – это некоторые числа, а параметр θ неизвестен. Метод максимального правдоподобия состоит в том, что в качестве оценки $\hat{\theta}$ неизвестного параметра θ выбирается его значение, максимизирующее функцию правдоподобия, Во многих случаях максимум этой функции находится из уравнения:

$$\frac{dL(\theta)}{d\theta} = 0.$$

Иногда удобнее находить максимум не самой функции $L(\theta)$, а ее логарифма $l(\theta) = ln\, L\,(\theta)$. Тогда получаем уравнение:

$$\frac{dl(\theta)}{d\theta} = \frac{d}{d\theta}\left[\sum_{i=1}^{N} ln f_{x_i}(x_i;\theta)\right] = 0.$$

Два последних уравнения называются *уравнениями правдоподобия*. Для определения оценки используется процедура максимизации функции правдоподобия, так как эта функция выражает предпочтительность различных значений θ при конкретной выборке. Такая оценка дает предпочтительные значений, так как при этом вероятность получения выборки максимальна.

В случае n неизвестных параметров $\theta_1, \theta_2, \dots, \theta_n$ их оценки находятся аналогично путем определения максимума функции правдоподобия $L(\theta_1, \theta_2, \dots, \theta_n)$ или ее логарифма одновременно по всем n параметрам, от которых они зависят. Соответствующие оценки будут получаться в результате решения системы n уравнений такого же вида, как два последних уравнения (производные будут частные).

Главным достоинством метода максимального правдоподобия является то, что он позволяет сравнительно просто отыскать оценки, обладающие хорошими свойствами:

- если существует эффективная оценка, метод максимального правдоподобия дает именно такую оценку и более точную оценку найти нельзя;
- оценка метода максимального правдоподобия при довольно слабых ограничениях состоятельна, а также асимптотически нормальна.

Таким образом, нахождение оценок максимального правдоподобия включает следующие этапы: построение функции правдоподобия (ее натурального логарифма); дифференцирование функции по искомым параметрам и составление системы уравнений; решение системы уравнений для нахождения оценок; определение второй производной функции, проверку ее знака в точке оптимума первой производной и формирование выводов.

Пример 1. Время до выхода из строя прибора определяется показательным законом:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0; \\ 0, \quad x < 0. \end{cases}$$

По выборке с элементами 2, 3, 5, 6, 8 определить оценку параметра λ.

Решение: Функция правдоподобия имеет вид:

$$L(2,3,5,6,8,\lambda) = \lambda e^{-\lambda 2} \cdot \lambda e^{-\lambda 3} \cdot \lambda e^{-\lambda 5} \cdot \lambda e^{-\lambda 6} \cdot \lambda e^{-\lambda 8} = \lambda^5 e^{-\lambda 24}.$$

Тогда:

$$lnL(2.3.5.6.8) = 5\, ln\, \lambda - 24\, \lambda.$$

Уравнение правдоподобия:

$$\frac{\partial lnL}{\partial \lambda} = \frac{5}{\lambda} - 24 = 0.$$

Отсюда получаем:

$$\hat{\lambda} = \frac{5}{24}.$$

Метод максимального правдоподобия позволяет получить состоятельные, эффективные (если таковые существуют, то полученное

решение даст эффективные оценки), асимптотически нормально распределенные оценки. Этот метод может давать как смещенные, так и несмещенные оценки. Смещение удается устранить введением поправок. Метод особенно полезен при малых выборках. Если функция максимального правдоподобия имеет несколько максимумов, то из них выбирают глобальный.

3.3 Метод моментов

Метод предложен К. Пирсоном в 1894 г. Будем считать, что вид функции распределения изучаемого количественного признака известен, но параметры этого распределения неизвестны. Нужно оценить эти параметры по выборке. Сущность метода моментов заключается в том, что по конкретной выборке $x_1, x_2, \ldots, x_N$ генеральной совокупности, распределение которой известно с точностью до параметров $\theta_1, \ldots, \theta_m$, выбирается столько эмпирических моментов, сколько требуется оценить неизвестных параметров распределения. Желательно применять моменты младших порядков, так как погрешности вычисления оценок резко возрастают с увеличением порядка момента. Вычисленные по экспериментальным данным оценки моментов приравниваются к теоретическим моментам; параметры распределения определяются через моменты, и составляются уравнения, выражающие зависимость параметров от моментов, в результате получается система уравнений. Решение этой системы дает оценки параметров распределения генеральной совокупности.

Пример 2. Время до выхода из строя прибора определяется показательным распределением:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0; \\ 0, \quad x < 0. \end{cases}$$

По выборке 2, 3, 5, 6, 8 найдём точечную оценку неизвестного параметра распределения λ. Так как для показательного распределения нужно оценить единственный параметр λ, то для его оценивания нужно одно уравнение. Приравняем, например, первые начальные моменты – теоретический и выборочный.

Первый начальный теоретический момент равен:

$$m_1 = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \lambda.$$

Первый начальный выборочный момент вычисляется по формуле:

$$\hat{m}_x = \sum_{i=1}^{N} \frac{x_i}{N}.$$

Приравнивая их, получим:

$$\frac{1}{\lambda} = \sum_{i=1}^{N} \frac{x_i}{N}.$$

Отсюда:

$$\vec{\lambda} = \frac{1}{\sum_{i=1}^{N} \frac{x_i}{N}}.$$

3.4 Точечные оценки параметров нормального распределения

Числовые характеристики случайных величин играют важную роль в математической статистике. Но наиболее полно случайная величина задается с помощью закона распределения вероятностей. Наиболее часто для описания свойств различных случайных величин используется нормальное распределение, определяемое выражением:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{(x-m_x)^2}{2\sigma_x^2}},$$

где m_x – математическое ожидание;

σ_x^2 – дисперсия случайной величины.

Теоретическим обоснованием роли нормального распределения является *центральная предельная теорема*. Когда есть основания рассматривать случайную величину как сумму большого числа независимых случайных воздействий, влияние каждого из которого ничтожно мало, то даже если распределения составляющих воздействий произвольны, согласно центральной предельной теореме, можно ожидать, что исследуемая случайная величина будет распределена по нормальному закону. Отсюда, конечно, не следует, что любая случайная величина будет распределена по нормальному закону. Нормальное распределение есть один из типов распределения, оно достаточно хорошо описывает многие явления, встречающиеся в природе, и имеет большое практическое применение. Во многих задачах нормальное распределение обладает тем преимуществом, что оно характеризуется удобными математическими свойствами. Поэтому большинство методов математической статистики построено в предположении, что исследуемая величина подчиняется нормальному распределению, хотя на практике это предположение всегда требует специальной проверки.

Понятия состоятельности, несмещенности, эффективности используются в основном для точечных оценок. Рассмотрим оценки и их свойства для ряда важнейших характеристик.

Пусть X – нормально распределенная случайная величина с известной дисперсией σ_x^2.

1. Требуется по N наблюдениям оценить ее математическое ожидание m_x.

Функция правдоподобия будет иметь следующий вид:

$$L(m_x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_1-m_x)^2}{2\sigma_x^2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_N-m_x)^2}{2\sigma_x^2}} = \frac{1}{(\sqrt{2\pi}\sigma_x)^N} e^{-\frac{\sum_{i=1}^{N}(x_i-m_x)^2}{2\sigma_x^2}}.$$

Логарифмическая функция правдоподобия равна:

$$l(m_x) = -\frac{N}{2} ln\, 2\,\sigma_x - N\, ln\,\sigma_x - \frac{1}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x)^2 = 0.$$

Максимум этой функции находится из уравнения:

$$\frac{dl(m_x)}{dm_x} = -\frac{2}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x) = 0.$$

Отсюда получаем оценку математического ожидания:

$$\widehat{m}_x = \bar{x} = \sum_{i=1}^{N}\frac{x_i}{N}.$$

Чтобы описать изменчивость этой функции от выборки к выборке при фиксированном N, рассмотрим случайную величину $\bar{X}$:

$$\bar{X} = \frac{1}{N}(X_1 + X_2 + \ldots + X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Известно, что линейная комбинация нормально распределенных случайных величин также распределена нормально. Поэтому можно записать:

$$M\{\bar{X}\} = \frac{1}{N}\sum_{i=1}^{N} M\{X_i\} = \frac{Nm_x}{N} = m_x;$$

$$D\{X\} = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_x^2 = \frac{N\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}.$$

Отсюда следует, что выборочное среднее $\bar{x}$ является состоятельной несмещенной оценкой m_x. Для выборки из нормальной генеральной совокупности эта оценка также и эффективная.

Отметим, что при установлении свойств оценок существенным является вид закона распределения исследуемой генеральной совокупности. Например, для равномерного распределения эффективной оценкой m_x служит следующая статистика:

$$\widehat{m}_x = \frac{x_{max} + x_{min}}{2}.$$

2. Получим оценку дисперсии σ_x^2 для нормальной совокупности при известном математическом ожидании m_x.

Запишем логарифмическую функцию правдоподобия для σ_x^2:

$$l(\sigma_x^2) = -\frac{N}{2} ln\, 2\,\pi - N\, ln\, \sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x)^2 = 0.$$

Тогда уравнение правдоподобия будет иметь вид:

$$\frac{dl(\sigma_x^2)}{d\sigma_x^2} = -\frac{N}{2\sigma_x^2} + \frac{1}{2\sigma_x^4}\sum_{i=1}^{N}(x_i - m_x)^2 = 0,$$

откуда следует, что оценка дисперсии σ_x^2 случайной величины X при известном математическом ожидании выражается формулой:

$$\hat{\sigma}_x^2 = S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_x)^2. \qquad (3.2)$$

Оценка S_x^2 является состоятельной, несмещенной и для нормальной совокупности эффективной оценкой дисперсии σ_x^2.

3. Получим оценку дисперсии в случае, когда математическое ожидание m_x неизвестно. То есть подлежат оцениванию два параметра - m_x и σ_x^2. Функция правдоподобия будет зависеть от двух параметров:

$$l(m_x, \sigma_x^2) = -\frac{N}{2} ln\, 2\,\pi - N\, ln\, \sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x)^2 = 0.$$

Оценки метода максимального правдоподобия являются решением системы уравнений:

$$\begin{cases} \dfrac{\partial l(m_x, \sigma_x^2)}{\partial m_x} = -\dfrac{1}{2\sigma_x^2}\sum\limits_{i=1}^{N}(x_i - m_x) = 0\,; \\ \dfrac{\partial l(m_x, \sigma_x^2)}{\partial \sigma_x^2} = -\dfrac{N}{2\sigma_x^2} - \dfrac{1}{2\sigma_x^4}\sum\limits_{i=1}^{N}(x_i - m_x) = 0\,. \end{cases}$$

Отсюда получим:

$$\widehat{m}_x = \bar{x} = \sum_{i=1}^{N}\frac{x_i}{N};$$

$$\hat{\sigma}_x^2 = S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{m}_x)^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2.$$

Такая оценка дисперсии является состоятельной. Однако она является смещенной, так как:

$$M\{\hat{\sigma}_x^2\} = \frac{N-1}{N}\sigma_x^2 = \sigma_x^2\left(1 - \frac{1}{N}\right).$$

И хотя $M\{\hat{\sigma}_x^2\} \to \sigma_x^2$, когда $N \to \infty$, то есть оценка асимптотически несмещенная, при малых N смещение оценки может быть весьма велико. На практике предпочитают использовать несмещенную оценку σ_x^2 вида:

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2. \qquad (3.3)$$

Эта оценка состоятельная, но для нормальной совокупности эффективна асимптотически.

Итак, в любой задаче, связанной с выполнением измерений, возможны два способа получения оценки значения σ_x^2.

При использовании первого способа снимается последовательность показаний прибора и путем сравнения полученных результатов с известным или калиброванным значением измеряемой величины находится последовательность отклонений. Затем полученная последовательность отклонений используется для вычисления среднего квадратичного отклонения по формуле (3.2).

Второй способ получения оценки значения σ_x^2 состоит в определении среднего арифметического $\overline{x}$, так как в этом случае действительное (точное) значение измеряемой величины неизвестно. В этом случае целесообразно использовать другую формулу для определения оценки дисперсии (3.3).

Деление на (*N-1*) производится по той причине, что наилучшая оценка, получаемая путем усреднения массива *X*, будет отличаться от точного значения на некоторую величину, если рассматривается выборка, а не вся генеральная совокупность. В этом случае сумма квадратов отклонений:

$$\sum_{i=1}^{N}(x_i - \bar{x})^2$$

будет несколько меньше, чем при использовании истинного среднего:

$$\sum_{i=1}^{N}(x_i - m_x)^2.$$

При делении на (*N-1*) вместо *N* эта погрешность будет частично скорректирована. В некоторых руководствах по математической статистике рекомендуется при вычислении выборочного среднеквадратичного отклонения всегда делить на (*N-1*), хотя иногда этого делать не следует. Нужно делить на (*N-1*) лишь в тех случаях, когда истинное значение не было получено независимым способом.

При установлении свойств оценок существенным является вид закона распределения исследуемой генеральной совокупности. В частности, для другого часто используемого вида распределения – равномерного, которое определяется следующим образом:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{если } a \le x \le b, \\ 0, & \text{если } x < a, x > b; \end{cases}$$

эффективные оценки параметров (математического ожидания и дисперсии) определяются по формулам:

$$m_x = \frac{a+b}{2},$$
$$\sigma_x^2 = \frac{(b-a)^2}{12}.$$

Остановимся еще на следующих определениях.

Каждая величина, зависящая от элементов выборки и входящая в формулу выборочной дисперсии, называется *связью*.

Разность между объемом выборки и числом связей называется *числом степеней свободы*. То есть число степеней свободы учитывает в статистических ситуациях связи, ограничивающие свободу изменения

случайных величин. Так, при вычислении выборочной дисперсии (3.3) наблюдается одна связь, определяемая значением средней арифметической величины $\overline{x}$, поэтому число степеней свободы будет равно $m=N-1$, а для дисперсии (3.2) число степеней свободы равно числу испытаний $m=N$, так как m_x определено независимым способом.

Понятие о степени свободы поясним на примере решения системы линейных алгебраических уравнений. Допустим, что мы имеем систему из n линейных алгебраических уравнений с n неизвестными $x_1, x_2, \ldots, x_n$. Очевидно, решение такой системы (при линейной независимости уравнений) будет ограниченным, то есть такая система не будет иметь ни одной степени свободы. Но если для n неизвестных переменных мы имеем k уравнений, то такая система уравнений будет иметь число степеней свободы $m=n-k$.

Математическое ожидание и дисперсия генеральной совокупности оцениваются средним и выборочной дисперсией тем точнее, чем больше объем выборки. При этом среднее характеризует результат измерений, а дисперсия – точность этого результата (*дисперсия воспроизводимости*). Если выполнить m параллельных опытов и получить выборку $x_1, x_2, \ldots, x_N$, то дисперсия воспроизводимости:

$$S^2_{воспр} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1},$$

а ошибка опыта (ошибка воспроизводимости):

$$S_{воспр} = \sqrt{S^2_{воспр}}.$$

3.5 Интервальные оценки

Точечные оценки не дают информации о степени близости $\hat{\theta}$ к соответствующему теоретическому параметру θ. Поэтому более информативный способ оценивания неизвестных параметров заключается не в определении единичного точечного значения, а в построении интервала, где с заданной степенью достоверности будет находиться оцениваемый параметр, то есть в построении интервальной оценки параметра θ.

Интервальный метод оценивания параметров распределения случайных величин заключается в определении интервала (а не единичного значения), в котором с заданной степенью достоверности будет заключено значение оцениваемого параметра. Интервальная оценка характеризуется двумя числами – концами интервала, внутри которого предположительно находится истинное значение параметра. Иначе говоря, вместо отдельной точки для оцениваемого параметра можно установить интервал значений, одна из точек которого является своего рода «лучшей» оценкой. Интервальные оценки являются более

полными и надежными по сравнению с точечными оценками. Они применяются как для больших, так и для малых выборок. Совокупность методов определения промежутка, в котором лежит значение параметра θ, получила название методов интервального оценивания.

Интервальной оценкой параметра θ называется интервал, границы которого l_1 и l_2 являются функциями выборочных значений $x_1, x_2, \ldots, x_N$ и который с заданной вероятностью p накрывает оцениваемый параметр θ:

$$P\{l_1 \leq \theta \leq l_2\} = p.$$

Интервал (l_1, l_2) называется *доверительным* интервалом, его границы l_1 и l_2, являющиеся случайными величинами (они определяются на основе выборочных данных) – соответственно, *нижним и верхним доверительными пределами,* вероятность p *– доверительной вероятностью,* а величина $q = 1\text{-}p$ - *уровнем значимости,* используемым при построении доверительного интервала.

Таким образом, *доверительный интервал* – это отрезок, центром которого является точечная оценка числовой характеристики, включающий истинное значение данной числовой характеристики с заданной *доверительной* вероятностью. То есть, доверительный интервал является мерой точности оценки, а доверительная вероятность характеризует достоверность оценки. Размер доверительного интервала зависит от того, каким значением доверительной вероятности задается экспериментатор. Чем выше доверительная вероятность, тем шире должен быть интервал, чтобы с заданной вероятностью включать в себя истинное значение числовой характеристики. Практически часто выбирают значение доверительной вероятности p = 0,95, полагая, что это значение достаточно велико, чтобы считать, что доверительный интервал «практически всегда» накрывает истинное значение. Только иногда, в случае ответственных и очень ответственных исследований, полагают p = 0,99 и 0,999 соответственно.

Ширина доверительного интервала не характеризует высокого качества оценки $\hat{\theta}$, если узкому доверительному интервалу соответствует невысокая доверительная вероятность. Доверительную вероятность и доверительный интервал всегда следует рассматривать лишь в совокупности. При фиксированной ширине доверительного интервала для состоятельных и несмещенных оценок доверительная вероятность p будет возрастать по мере увеличения объема выборки. При заданном объеме выборки чем шире доверительный интервал, тем больше доверительная вероятность. То есть при определенном объеме выборки нельзя повысить p без увеличения ширины доверительного интервала или невозможно уменьшить ширину этого интервала, не уменьшая доверительной вероятности p. Выбор p производится исследователем; приемлемый для данного исследования уровень достоверности в общем случае выбирается достаточно произвольно, хотя и

зависит от необходимой надежности результатов. Чаще всего используется значение p=0.95 (то есть уровень значимости 0.5).

Общая процедура получения интервальной оценки:

1. Записывается вероятностное утверждение относительно некоторой случайной функции вида:

$$P\{\delta_1 \le g(\theta, \hat{\theta}) \le \delta_2\} = \int_{l_1}^{l_2} f(g)dg,$$

где $f(g)$ – функция плотности вероятности некоторой подходящей статистики g.

При этом значения δ_1 и δ_2 берутся с учетом дополнительных условий:

$$P\{g(\theta, \hat{\theta}) < \delta_1\} = P\{g(\theta, \hat{\theta}) > \delta_2\} = (1 - p)/2 = q/2.$$

Такая функция несет информацию о степени близости величин. Необходимо, чтобы закон распределения упомянутой функции был известен.

2. Вероятностное утверждение преобразуется к виду, при котором границы доверительного интервала числовой характеристики представляются в явном виде, то есть так, чтобы в окончательном виде оцениваемый параметр оказался заключенным между величинами, найденными по выборке. Это и будут границы доверительного интервала.

Статистика $g(\theta, \hat{\theta})$ выбирается таким образом, чтобы она допускала подобное преобразование и имела известную, лучше табулированную функцию плотности вероятности $f(g)$. Последнее обстоятельство существенно упрощает определение значений δ_1 и δ_2.

Для определения доверительных вероятностей используется нормированная функция Лапласа; для облегчения расчетов эта функция представлена таблицами. Отклонение истинного значения случайной величины от математического ожидания не превосходит утроенного среднего квадратичного отклонения с вероятностью 0,997. Это свойство в математической статистике носит название «правило трех сигм».

Для анализа вероятностных свойств оценок $\hat{\theta}$ некоторого параметра θ в предположении, что θ известно, используются *выборочные распределения*. В частности, в такой ситуации оказывается возможным до извлечения выборки объема N указать границы, в которых с заданной вероятностью p окажется выборочное распределение $\hat{\theta}$. Иными словами, выборочные распределения на основании знания общих свойств генеральной совокупности дают возможность судить об отдельной частной выборке и оценках, полученных на ее основе. Однако реально параметры генеральной совокупности неизвестны и больший практический интерес представляет решение обратной задачи: по известной выборке вынести обоснованное суждение о свойствах генеральной

совокупности, то есть использовать выборочную оценку $\hat{\theta}$ для получения вероятностных утверждений относительно параметра θ.

Доверительные интервалы, найденные с помощью так называемых выборочных распределений, которые будут рассмотрены в следующем параграфе. Напомним, что эти интервалы имеют место только тогда, когда соответствующие случайные величины являются нормально распределенными, а наблюдения – независимыми.

3.6 Выборочные распределения

В математической статистике широкое применение находит ряд теоретических законов распределения. Как уже говорилось, наиболее важным из них является *нормальное распределение*. Различные вероятностные утверждения относительно значений точечных оценок могут быть сформулированы с помощью *выборочных распределений*. Это распределения *хи-квадрат, Стьюдента, Фишера,* а также интеграл вероятностей. Для указанных законов функции распределения аналитически не представимы. Значения функций определяются по таблицам или с использованием стандартных процедур пакетов прикладных программ. Указанные таблицы обычно построены в целях удобства проверки статистических гипотез в ущерб теории распределений – они содержат не значения функций распределения, а критические значения аргумента.

Распределение выборочного среднего при известной дисперсии.

В математической статистике доказано, что если случайная величина X подчиняется нормальному распределению с параметрами m_x, σ_x^2, то случайная величина:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

имеет нормальное распределение с параметрами $m_x, \frac{\sigma_x^2}{N}$.

Отсюда ясно, что распределение статистики:

$$G = \frac{\overline{x} - m_x}{\sigma_x\sqrt{N}}$$

есть *нормированное нормальное распределение* (U-распределение). Таблицу этого распределения используют, чтобы предсказать до извлечения выборки вероятность того, что выборочное среднее $\bar{x}$ окажется в некотором интервале с фиксированными границами, если только известны m_x и σ_x^2 исходного распределения.

Можно также решить другую задачу: определить интервал, содержащий с заданной вероятностью выборочное среднее.

Пример 3.1. Случайная нормально распределенная величина имеет следующие параметры: $m_x = 5, \sigma_x^2 = 25$. Пусть намечается извлечь выборку объема *N=16*. Найти интервал, в котором с вероятностью *p*=0,95 будет заключено $\bar{x}$.

Очевидно, нужно определить такие границы l_1 и l_2, что:

$$P\{l_1 \le \frac{\bar{x} - m_x}{\sigma_x/\sqrt{N}} \le l_2\} = p.$$

Этому условию удовлетворяет бесконечное множество границ l_1 и l_2, соответствующих интервалам различной ширины. Наименьший среди них симметричен относительно 0, т. е. $l_2 = -l_1 - l$. Тогда имеем:

$$P\left\{-U_{\alpha=q/2} \le \frac{\bar{x} - m_x}{\frac{\sigma_x}{\sqrt{N}}} \le U_{\alpha=\frac{q}{2}}\right\} = 0{,}95.$$

Для нахождения границы l следует использовать таблицу нормированного нормального распределения: $l = U_{\alpha=0.025} = 1{,}96$. Таким образом:

$$P\{-1{,}96 \le \frac{\bar{x} - m_x}{\sigma_x/\sqrt{N}} \le 1{,}96\} = 0{,}95,$$

а после подстановки значений m_x, σ_x^2 и *N* и простейших преобразований получаем:

$$P\{2{,}55 \le \bar{x} \le 7{,}45_2\} = 0{,}95.$$

Интерпретация этого соотношения: при многократном извлечении выборок объема $N = 16$. из данной генеральной совокупности в среднем 95% выборок будут иметь выборочное среднее $\bar{x}$, содержащееся в интервале от 2,55 до 7,45.

Выборочное распределение оценки дисперсии (χ^2- распределение).

Если X_i^*, $i = 1{,}2, \dots, \nu$ – независимые нормированные нормально распределенные случайные величины с математическим ожиданием $m_{X^*} = 0$ и дисперсией $\sigma_{X^*}^2 = 1$, то сумма квадратов этих величин $\chi^2 = (X_1^*)^2 + (X_2^*)^2 + \dots + (X_\nu^*)^2$ подчиняется распределению χ_ν^2 («хи-квадрат распределение с ν степенями свободы»).

С ростом значения ν распределение χ_ν^2 приближается к нормальному с $M\{\chi_\nu^2\} = \nu, D\{\chi_\nu^2\} = 2\nu$. Для удобства использования χ^2-распределения существуют многочисленные таблицы.

Для $\nu = 1 \div 30$ и ряда вероятностей α в таблице указаны такие значения $\chi_{\nu,\alpha}^2$, для которых вероятность $P\chi^2 > \chi_{\nu,\alpha}^2 = \alpha$. Вследствие несимметрии χ_ν^2-

распределения таблица включает как малые значения α (0.01, 0.05, 0.10), так и значения, близкие к 1 (0.99, 0.95, 0.90). Если υ>30, для определения значений $\chi^2_{\nu,\alpha}$ применяют соотношение, полученное на основе таблиц нормального распределения:

$$\chi^2_{\nu,\alpha} \cong \frac{1}{2}(\sqrt{2\nu - 1} + U_\alpha)^2.$$

$\chi^2_{\nu,\alpha}$-распределение применяется для описания вероятностных свойств выборочных дисперсий: нахождение интервала с заданной вероятностью p содержащего конкретное выборочное значение S^2_x, если только известна σ^2_x.

Соответствующая случайная величина S^2_X имеет вид:

- когда математическое ожидание m_x известно:

$$S^2_X = \frac{1}{N}\sum_{i=1}^{N}(X_i - m_X)^2; \quad (3.4)$$

- когда m_x неизвестно:

$$S^2_X = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2. \quad (3.5)$$

Если X_i – независимые случайные величины, подчиняющиеся нормальному распределению с параметрами m_x, σ^2_x, а m_x известно, то статистика:

$$G = N\frac{S^2_X}{\sigma^2_X}$$

имеет χ^2_N-распределение. В самом деле:

$$G = N\frac{S^2_X}{\sigma^2_X} = \sum_{i=1}^{N}(\frac{X_i - m_x}{\sigma_X})^2 = \sum_{i=1}^{N}(X_i^2)^2 = \chi^2_N. \quad (3.6)$$

Если m_X неизвестно, то целесообразно рассмотреть статистику:

$$G = \frac{(N-1)S^2_X}{\sigma^2_X} = \sum_{i=1}^{N}(\frac{X_i - \bar{X}}{\sigma_X})^2. \quad (3.7)$$

Можно показать, что при неизвестном m_x статистика G, задаваемая этим соотношением, подчиняется χ^2_{N-1}-распределению с числом степеней свободы $\nu = N - 1$.

Отметим, что в общем случае число степеней свободы, соответствующее той или иной оценке дисперсии (или некоторой сумме квадратов), определяется как количество независимых наблюдений, по которым вычисляется данная оценка дисперсии минус число линейных связей между теми же наблюдаемыми

значениями. Для суммы квадратов $\sum_{i=1}^{N}(X_i - \bar{X})^2$ из последнего выражения имеется одна линейная связь между X_i, а именно:

$$\sum_{i=1}^{N}(X_i - \bar{X}) = \sum_{i=1}^{N}(X_i - \sum_{i=1}^{N}\frac{X_i}{N}) = 0,$$

и поэтому в данном случае $\nu = N - 1$.

Статистики G вида (3.6) и (3.7) могут применяться для нахождения интервала с заданной вероятностью p, содержащего конкретное значение S_x^2, если только априори известна дисперсия σ_x^2.

Пример 3.2. Предполагается извлечь выборку объема *N=10* из нормальной совокупности с неизвестным математическим ожиданием m_x и дисперсией $\sigma_x^2 = 18$. Найти интервал, в котором с вероятностью p=0,9 будет находиться выборочное значение S_x^2.

Используя тот факт, что статистика $G = \frac{(N-1)S_X^2}{\sigma_X^2}$ распределена как χ_{N-1}^2, границы l_1 и l_2 в соотношении:

$$P\{l_1 \leq \frac{(N-1)S_X^2}{\sigma_X^2} \leq l_2\} = p$$

можно найти с помощью χ_ν^2-распределения для $\nu = N - 1$:

$$P\{\chi_{v,\alpha=0,95} \leq \frac{(N-1)S_X^2}{\sigma_X^2} \leq \chi_{v,\alpha=0,05} = 0,9.$$

Конечно, подобные границы могут быть найдены бесконечным числом способов. Однако на практике принято выбирать их таким образом, чтобы вероятности попадания в области $\chi_{v,\alpha=0,95} < l_1$ и $\chi_{v,\alpha=0,05} > l_2$ были бы одинаковыми (в данном случае по 0,05). После преобразований из последнего выражения получим:

$$P\left\{\frac{\chi_{v,\alpha=0,95}}{N-1}\sigma_X^2 \leq S_X^2 \leq \frac{\chi_{v,\alpha=0,05}}{N-1}\sigma_X^2\}\right\} = 0,9.$$

Используя таблицу $\chi_{v,\alpha}^2$-распределения, получим:

$$\chi_{9,\alpha=0,95}^2 = 3,325; \chi_{9,\alpha=0,05}^2 = 16,916;$$

$$P\{6,65 \leq S_X^2 \leq 33,84\} - 0,9.$$

Таким образом, при извлечении большого числа выборок объема *N=10* из рассматриваемой генеральной совокупности следует ожидать, что в среднем 9 выборочных значений дисперсий из 10 будут находиться в интервале (6,65;33,84).

Распределение выборочного среднего в случае неизвестной дисперсии (t-распределение Стьюдента).

Если U – нормированная нормальная случайная величина, χ_ν^2 – случайная величина, имеющая χ^2 – распределение с ν степенями свободы, причем U и χ_ν^2 независимы, то отношение:

$$t_\nu = \frac{U}{\sqrt{\chi_\nu^2/\nu}}$$

будет подчиняться t-распределению Стьюдента с ν степенями свободы. Кривая плотности вероятности t-распределения симметрична относительно 0 и с ростом ν приближается к кривой нормированного нормального распределения. При небольших ν распределение Стьюдента заметно отличается от нормального.

Для разных ν в таблице t-распределения приведены значения $t_{\nu,\alpha}$такие, что $P\{t_\nu > t_{\nu,\alpha}\} = \alpha$. В силу симметрии $P\{t_\nu \le -t_{\nu,\alpha}\} = \alpha$, поэтому в таблице даны лишь $t_{\nu,\alpha} > 0$. Строка таблицы $\nu = \infty$ соответствует нормальному распределению, причем уже для $\nu > 60$ t-распределение с очень высокой точностью аппроксимируется нормированным нормальным распределением.

Распределение Стьюдента применяется для выяснения вероятностных свойств выборочного математического ожидания $\bar{x}$, когда дисперсия генеральной совокупности σ_x^2 неизвестна.

В самом деле, случайная величина $G = \frac{\bar{x}-m_X}{\sigma_X/\sqrt{N}}$ имеет нормированное нормальное распределение, а $\frac{(N-1)S_X^2}{\sigma_X^2}$ подчиняется χ_{N-1}^2-распределению; можно показать также, что эти случайные величины независимы. Тогда отношение:

$$t = \frac{\bar{X}-m_X}{\sigma_X\sqrt{N}} : \sqrt{\frac{(N-1)S_X^2}{\sigma_X^2(N-1)}} = \frac{\bar{X}-m_X}{S_X\sqrt{N}}$$

имеет, согласно определению, t-распределение с $\nu = N-1$ степенями свободы.

Статистика $G = \frac{\bar{x}-m_X}{S_X\sqrt{N}}$ может использоваться до извлечения выборки для высказывания различного рода вероятностных утверждений относительно

значений выборочного среднего $\bar{x}$. В частности, для построения интервала, в который с вероятностью *p* попадает $\bar{x}$.

Пример 3.3. Планируется извлечь выборку объема *N=16* из нормальной совокупности с $m_x = 5$ и неизвестной дисперсией σ_X^2. Определить интервал, в котором с вероятностью *p*=0,95 окажется $\bar{x}$.

Границы интервала l_1 и l_2 задаются условием:

$$P\left\{l_1 \le \frac{\bar{x} - m_X}{S_X\sqrt{N}} \le l_2\right\} = p.$$

Для симметричного относительно нуля интервала, имеющего наименьшую ширину из всех подобных интервалов, имеем $l = l_2 = -l_1 = t_{\nu,\alpha=1-p/2}$. С помощью таблицы *t*-распределения для $\nu = N - 1$=15 и $\alpha = 0{,}025$ находим $t_{15,\alpha=0{,}025} = 2{,}13$.

После подстановки численных значений в последнюю формулу и преобразований получим:

$$P\{5\text{-}0{,}5335\cdot S_X \le \bar{x} \le 5+0{,}5335\cdot S_X\}=0{,}95.$$

Таким образом, при извлечении множества выборок объема $N = 16$ в среднем в 95% случаев значение $\bar{x}$ окажется в интервале $(5 - 0{,}5335S_X, 5 + 0{,}5335S_X)$. В данном варианте до извлечения выборки точное значение границ интервала для $\bar{x}$ определить невозможно, поскольку остается неизвестным выборочное значение S_X.

Распределение отношения двух выборочных дисперсий (F-распределение Фишера).

Пусть $\chi^2_{\nu_1}$ и $\chi^2_{\nu_2}$ – две независимые случайные величины, имеющие χ^2 - распределения с числом степеней свободы ν_1, ν_2соответственно. Тогда распределение, которому подчиняется отношение:

$$F_{\nu_1,\nu_2} = \frac{\chi^2_{\nu_1}}{\nu_1} : \frac{\chi^2_{\nu_2}}{\nu_2}$$

называется *F-распределением Фишера* с ν_1, ν_2 степенями свободы, где ν_1 – число степеней свободы числителя, ν_2 – число степеней свободы знаменателя. Плотность *F*-распределения – функция с единственным максимумом, несимметричная при $\nu_2 > 2$. Она концентрируется около 1 при больших ν_1 и ν_2. В этой таблице этого распределения-для ряда ν_1 и ν_2 указаны такие значения случайной величины $F_{\nu_1,\nu_2,\alpha}$, что $P\{F > F_{\nu_1,\nu_2,\alpha}\} = \alpha$. Таблицы обычно даются

лишь для $F \geq 1$. Но ее можно использовать и для $F < 1$, так как $F_{\nu_1,\nu_2,1-\alpha} = \frac{1}{F_{\nu_1,\nu_2,\alpha}}$.

В качестве частных случаев можно записать:

- при $\nu = \infty$, $F^2_{\nu_1,\infty} = \frac{1}{\nu_1}\chi^2_{\nu_1}$;

- при $\nu = 1$, $F_{1,\nu_1} = t^2_{\nu_2}$, где t_{ν_2} – случайная величина, подчиняющаяся t-распределению с числом степеней свободы ν_2.

F-распределение служит для сравнения оценок дисперсий $S^2_{X_1}$ и $S^2_{X_2}$, полученных по двум независимым выборкам объема N_1 и N_2. При извлечении подобных выборок из двух нормальных совокупностей с параметрами $m_{X_1}, \sigma^2_{X_1}$ и $m_{X_2}, \sigma^2_{X_2}$ статистики $\nu_1 \cdot \frac{S^2_{X_1}}{\sigma^2_{X_1}}$ и $\nu_2 \cdot \frac{S^2_{X_2}}{\sigma^2_{X_2}}$, подчиняются χ^2-распределениям с $\nu_1 = N_1 - 1$ и $\nu_2 = N_2 - 1$ степенями свободы (если, конечно, m_{X_1} и m_{X_2} неизвестны; при известных m_{X_1} и m_{X_2}, $\nu_1 = N_1$, $\nu_2 = N_2$). Поэтому отношение:

$$F_{\nu_1,\nu_2} = \frac{S^2_{x_1}}{S^2_{x_2}} : \frac{\sigma^2_{x_2}}{\sigma^2_{x_1}} \tag{3.8}$$

имеет F_{ν_1,ν_2}-распределение. На практике теоретические дисперсии $\sigma^2_{X_1}$ и $\sigma^2_{X_2}$ неизвестны, но если предположить, что $\sigma^2_{X_1} = \sigma^2_{X_2}$, то статистика $G = \frac{S^2_{X_1}}{S^2_{X_2}}$ будет подчиняться F-распределению с ν_1 и ν_2 степенями свободы. В том же случае, когда $\sigma^2_{X_1} \neq \sigma^2_{X_2}$, статистика G будет распределена как $\frac{\sigma^2_{x_1}}{\sigma^2_{x_2}} F_{\nu_1,\nu_2}$ и, следовательно, функция плотности такого распределения концентрируется не около 1, а около значения $\frac{\sigma^2_{X_1}}{\sigma^2_{X_2}}$.

F-распределение может применяться для высказывания тех или иных вероятностных утверждений об отношении выборочных дисперсий $\frac{S^2_{X_1}}{S^2_{X_2}}$ до извлечения выборки, например, для определения интервала, в котором с заданной вероятностью p будет заключено данное отношение.

Пример 3.4. Пусть необходимо извлечь выборку объема $N_1 = 21$ из нормальной совокупности с $\sigma^2_{X_1} = 5$ и неизвестным m_{X_1} и выборку объема $N_2 = 10$ из другой нормальной совокупности с $\sigma^2_{X_2} = 15$ и неизвестным m_{X_2}. Найти интервал, содержащий с вероятностью p=0.9 отношение выборочных дисперсий.

Из (3.8) следует, что величина $3 \cdot \frac{S^2_{X_1}}{S^2_{X_2}}$ имеет F-распределение с $\nu_1 = 20$ и $\nu_2 = 9$ степенями свободы. Тогда границы l_1 и l_2 интервала, для которого:

$$P\left\{l_1 \le 3\frac{S_{X_1}^2}{S_{X_2}^2} \le l_2\right\} = p = 0.9,$$

могут быть найдены с помощью таблицы *F*-распределения.

Если, как обычно, принять, что вероятности событий $F < l_1$ и $F > l_2$ одинаковы (здесь по 0.05), то:

$$l_2 = F_{20;9;\alpha=0.05} = 2.94; l_1 = \frac{1}{F_{20;9;\alpha=0,95}} F_{20;9;\alpha=0.05} = 0.41.$$

Тогда $P\left\{0.14 \le \frac{S_{X_1}^2}{S_{X_2}^2} \le 0.98\right\} = 0.9$. Таким образом, с вероятностью *p*=0.9 можно ожидать, что отношение $\frac{S_{X_1}^2}{S_{X_2}^2}$ в данном случае попадает в интервал от 0.14 до 0.98.

Таким образом, выборочные распределения используются для анализа вероятностных свойств различного рода оценок некоторого параметра в предположении, что этот параметр генеральной совокупности известен. Можно решить обратную задачу: по отдельной выборке вывести обоснованное суждение о свойствах параметров генеральной совокупности, то есть использовать выборочную оценку для получения вероятностных утверждений относительно параметра. В частности, можно получить интервальную оценку параметра генеральной совокупности.

Сравнение нескольких дисперсий нормальных генеральных совокупностей по выборкам одинакового объема (К-критерий Кохрена).

Пусть генеральные совокупности *X₁, X₂, X₃, …, Xₘ* распределены нормально. Из этих совокупностей извлечено m выборок одинакового объема N и по ним найдены выборочные дисперсии $S_1^2 + S_2^2 + S_3^2 + \cdots + S_m^2$ с одинаковым числом степеней свободы $\nu = n - 1$. Если значение ν для всех выборок одинаково, то для сравнения дисперсий можно использовать критерий Кохрена, который основан на статистике распределения отношения максимальной эмпирической дисперсии к сумме всех дисперсий:

$$K = \frac{S_{max}^2}{S_1^2 + S_2^2 + . S_3^2 + \cdots + S_m^2}.$$

Распределение случайной величины *K* зависит только от числа степеней свободы ν и количества выборок m.

Критическую область строят правостороннюю, исходя из требования, чтобы вероятность попадания *K*-критерия в эту область в предположении справедливости нулевой гипотезы $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_m^2$ была равна принятому уровню значимости α:

$$P[K > K_\alpha(\nu; m)] = \alpha.$$

Критическую точку $K_\alpha(\nu;\ m)$ находят по таблицам распределения Кохрена. Если рассчитанное по результатам экспериментов значение $K_{расч} > K_\alpha(\nu;\ m)$, то нуль-гипотезу отвергают при уровне значимости α.

3.7 Планирование эксперимента при построении интервальных оценок

Планирование эксперимента сводится к определению необходимого объема выборки *N* с тем, чтобы при фиксированной доверительной вероятности *p* была бы достигнута заданная точность оценивания параметров. В качестве характеристики этой точности используется относительная величина $\varepsilon = \frac{L}{2\sigma_x}$.

При определении математического ожидания имеем:

$$L = (\hat{l}_2 - \hat{l}_1) = \frac{2U_{\alpha=q/2}\sigma_x}{\sqrt{N}}.$$

Отсюда $\varepsilon = \frac{U_{\alpha=q/2}\sigma_x}{\sqrt{N}}$ и, задаваясь предельно допустимой относительной погрешностью $\varepsilon_{доп}$, получим:

$$N \geq \frac{U^2{}_{\alpha=\frac{q}{2}}}{\varepsilon^2_{доп}}. \qquad (3.9)$$

То есть объем выборки *N* определяется вероятностью *p* и относительным (по отношению к среднеквадратичному отклонению) значением доверительного интервала, то есть требуемой точностью определения измеряемой величины.

Пример 3.5. Пусть требуется оценить m_x с допустимой погрешностью $\varepsilon_{\text{доп}} = 0{,}5; q = 0.01$. Тогда $U_{\alpha=0.005} = 2.58$.

Подставив числовые значения в формулу (3.9), найдем $N \geq 26.62$.

Значит, необходимо проделать 27 опытов.

При оценивании среднеквадратического отклонения используется уравнение для определения $\nu = N - 1$ (доверительный интервал для σ_x оценивает среднеквадратическое отклонение с относительной погрешностью, не превышающей 100ε%):

$$\chi^2_{\upsilon,\alpha=q/2} / \chi^2_{\upsilon,\alpha=1-q/2} \leq (1 + \varepsilon_{доп})^2.$$

Решение ищется методом подбора с помощью таблицы χ^2-распределения: находится значение ν, удовлетворяющее приведенному неравенству, а затем $N = \nu + 1$.

3.8 Контрольные вопросы

3.8.1 Дайте определение генеральной совокупности и выборки.

3.8.2 Что такое репрезентативная выборка?

3.8.3 Дайте определение статистической оценки параметра распределения случайной величины.

3.8.4 Какие требования с точки зрения качества предъявляются к оценкам параметра?

3.8.5 Объясните метод максимального правдоподобия.

3.8.6 Объясните метод моментов.

3.8.7 Роль нормального распределения при описании свойства случайных величин.

3.8.8 Теоретические формулы определения параметров генеральной совокупности.

3.8.9 Точечные и интервальные оценки параметров.

3.8.10 Планирование эксперимента при определении необходимого объема выборки при оценивании математического ожидания и дисперсии.

4 Статистические гипотезы

4.1 Проверка статистических гипотез

Напомним, что при исследовании некоторого признака генеральной совокупности вначале извлекается репрезентативная выборка из этой совокупности; на основании изучения этой выборки делается вывод обо всей совокупности (*выборочный* метод). При статистическом оценивании параметров генеральной совокупности для обоснования выбора той или иной оценки неизвестного параметра распределения необходимо высказывать определенные предположения. Например, что распределение исследуемой величины не противоречит нормальному закону распределения; что при расчете оценок не обнаружено присутствие среди них грубых ошибок (резко выделяющихся значений). С еще большим количеством различных предположений приходится сталкиваться, когда необходимо не только определять случайные величины, но и сравнивать их между собой, и тем более, когда по результатам эксперимента строится функция отклика. Проверка и решение этих задач осуществляется с помощью статистических гипотез.

Статистическая гипотеза – определенное предположение относительно свойств генеральной совокупности, из которой извлекается выборка, формулируемое на основе выборки (предположение о виде неизвестного распределения или о параметрах известных распределений).

Нулевой гипотезой называют выдвинутую гипотезу (проверяемое предположение) и обозначают H_0. Чаще всего она формулируется о том, что оценки, полученные для различных выборок, принадлежат одной генеральной совокупности, то есть разница между оценками фактически равна нулю и

возникла в силу случайности отбора элементов и ограниченности объема выборки. Если проверяется какое-то качественное свойство генеральной совокупности, проверку гипотезы стараются привести к количественному виду.

Гипотезу, противопоставляемую нулевой, называют *альтернативной гипотезой* и обозначают H_1.

Если нулевая гипотеза будет отвергнута, то будет иметь место альтернативная гипотеза.

Правильность этих гипотез проверяется путем вычисления некоторых числовых характеристик по данным наблюдений (измерений) и сравнения их с теми, которые должны быть при условии, что проверяемая гипотеза истинна, а наблюдаемые отклонения объясняются случайными колебаниями в выборках и их ограниченностью. Такие характеристики называют критериями проверки статистических гипотез.

Критерий статистической гипотезы есть правило, позволяющее отвергнуть или принять данную гипотезу на основании выборки. Значение критерия, которое вычислено по выборке, называются эмпирическим значением (расчетным). При построении такого правила используются некоторые функции результатов наблюдения, они называются *статистиками для проверки гипотез*.

Все возможные значения подобных статистик для проверки той или иной гипотезы делятся на две группы: область принятия гипотезы и критическая область.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых нулевую гипотезу принимают.

Критическая область состоит из всех значений статистики, при которых принимается решение отвергнуть принимаемую гипотезу как ложную.

Критическими точками (границами) называют точки, отделяющие критическую область от области принятия гипотезы.

Проверка гипотезы сводится к выяснению, попадает или нет значение используемой статистики в критическую область: если не попадает – гипотеза принимается как не противоречащая результатам наблюдения, если попадает – гипотеза отвергается.

Таким образом, статистические гипотезы носят вероятностный характер. Это говорит о том, что в ряде случаев можно ошибиться. Для количественной характеристики степени ошибки используется показатель q, который называется *уровнем значимости*. Произведение $q \cdot 100\%$ показывает, в скольких случаях из 100 можно ошибаться, то есть долю отвергаемых гипотез. Тогда величина $p=1-q$ является *доверительной вероятностью (*надежностью*)*. Уровень значимости – мера наших требований к ответу: чем больших гарантий мы будем требовать, тем менее определенным станет ответ. Ответ одновременно очень точный и очень надежный, как правило, стоит очень дорого. Опыт использования статистики в разнообразных ситуациях в течение нескольких десятилетий показал, что обычно в практических ситуациях

определяющим значением q является 0,05. Такое значение, называемое иногда 5%-ным уровнем риска, соответствует вероятности верного ответа, то есть его надежности при проверке гипотез $p=1-0,05=0,95$ или 95%. При этом говорят, что в среднем только в 5 случаев из 100 возможна ошибка. Конечно, никогда нельзя дать 100%-й гарантии. Подобное желание приводит к необходимости выполнять бесконечное много опытов, что абсурдно.

Поскольку проверка гипотез ведется при ограниченной информации (по выборке), то могут возникнуть *ошибки двух родов*.

Если будет отвергнута правильная гипотеза, то совершается *ошибка первого рода*. Вероятность ошибки первого рода – *уровень значимости* q.

Если будет принята неверная гипотеза, то совершается *ошибка второго рода.* Вероятность ошибки второго рода обозначают β. Она связана с понятием мощности.

Мощность критерия есть вероятность отбросить неверную гипотезу и равна $1-\beta$. Чем больше мощность критерия, тем меньше вероятность совершить ошибку второго рода. Так, при контроле качества продукции величина q показывает риск производства, отвергая правильную гипотезу, эта величина бракует годную продукцию. Значение β характеризует риск потребителя, допуская ложную гипотезу, эта величина принимает и использует фактически непригодную продукцию.

Ситуации, возникающие при статистической проверке гипотез представлены в таблицах 4.1–4.2.

Таблица 4.1 – Возможные варианты принятия решения о правильности гипотез

Гипотеза	Объективно верна	Объективна неверна
Принимается	Правильное решение	Ошибка 2-го рода
Отвергается	Ошибка 1-го рода	Правильное решение

Нахождение критериев, позволяющих получить минимальную суммарную вероятность ошибок первого и второго рода – сложная задача. Поэтому выбирают определенное значение уровня значимости и стремятся найти критерий, минимизирующий β. На практике применяются значения q, равные 0,01; 0,05 (наиболее часто), 0,1.

При фиксированном объеме выборки и заданном уровне значимости из двух критериев для проверки одной и той же гипотезы предпочтительнее будет тот, который обладает, в каком-то смысле, большей мощностью (например, кривая функции мощности этого критерия лежит выше, чем аналогичная кривая сопоставляемого критерия).

Таблица 4.2 – Ошибки при проверке гипотез

	Решение	
	Принять H_0	Принять H_1
Справедлива H_0	Правильное с вероятностью 1-α	Ошибочное с вероятностью α
Справедлива H_1	Ошибочное с вероятностью β	Правильное с вероятностью 1- β

При заданном объеме выборки вероятность совершения ошибки первого рода можно уменьшить, снижая уровень значимости q. Однако при этом увеличивается вероятность ошибки второго рода β (снижается мощность критерия). Единственный способ уменьшить обе вероятности состоит в увеличении объема выборки (плотность распределения оценки параметра при этом становится более «узкой»). При выборе критической области руководствуются правилом Неймана – Пирсона: следует так выбирать критическую область, чтобы вероятность q была мала, если гипотеза верна, и велика – в противном случае. Однако выбор конкретного значения q относительно произволен. Употребительные значения лежат в пределах от 0,001 и до 0,2.

При выборе уровня значимости необходимо учитывать мощность критерия при альтернативной гипотезе. Иногда большая мощность критерия оказывается существеннее малого уровня значимости, и его значение выбирают относительно большим, например 0,2. Такой выбор оправдан, если последствия ошибок второго рода более существенны, чем ошибок первого рода. Например, если отвергнуто правильное решение «продолжить работу пользователей с текущими паролями», то ошибка первого рода приведет к некоторой задержке в нормальном функционировании системы, связанной со сменой паролей. Если же принято решение не менять пароли, несмотря на опасность несанкционированного доступа посторонних лиц к информации, то эта ошибка повлечет более серьезные последствия.

4.2 Критерии значимости

Критерии, с помощью которых проверяются гипотезы об абсолютных значениях параметров или о соотношениях между ними для генеральных совокупностей с известной (с точностью до параметров) функцией распределения вероятностей, называются *критериями значимости.* При проверке подобных гипотез, если распределение генеральной совокупности определяется целой группой параметров, а проверяемая гипотеза касается лишь их части, остальные параметры полагаются известными или же вычисляются по данным выборки.

С помощью критериев значимости выясняется

а) не противоречат ли результаты наблюдений предположению о том, что некоторые из параметров генеральной совокупности равны определенным числам;

б) если имеются две или большее количество выборок, существуют ли объективные основания считать, что они являются представителями одной и той же генеральной совокупности, или же экспериментальные данные говорят о наличии неслучайных различий между выборками, свидетельствующих о том, что они извлечены из разных генеральных совокупностей.

Пусть некоторая оценка $\hat{\theta}$ вычислена по выборке объема N. Пусть есть причины считать, что истинное значение оцениваемого параметра θ, то есть его значение в генеральной совокупности равно θ_0. Это проверяемое предположение – *нулевая гипотеза* обозначается H_0: $\theta = \theta_0$.

Даже если нулевая гипотеза справедлива, выборочное значение не будет точно совпадать с θ_0, так как $\hat{\theta}$ является всего лишь одним из конкретных значений случайной величины $\hat{\theta}$, порожденной всевозможными выборками объема N. Насколько сильно $\hat{\theta}$ должно отличаться от θ_0, чтобы в достаточной мере обоснованно можно было бы отвергнуть нулевую гипотезу? Необходимо знать, каковы значения вероятностей попадания выборочных величин $\hat{\theta}$ в различные зоны из области возможных значений $\hat{\theta}$, если нулевая гипотеза верна. То есть надо знать функцию плотности распределения вероятности $f_{\hat{\theta}}(\hat{\theta})$, построенную теоретически в предположении, что нулевая гипотеза верна. Эта функция $f_{\hat{\theta}}(\hat{\theta})$ – выборочное распределение статистики. Если она получена, то с ее помощью легко найти такую зону, случайное попадание в которую выборочного значения $\hat{\theta}$ при справедливой нулевой гипотезе маловероятно.

Пусть вероятность такого попадания равна достаточно малому числу q. Тогда, если $\hat{\theta}$ все-таки попадет в эту зону, имеются серьёзные основания подвергнуть сомнению справедливость нулевой гипотезы. Подобную зону малой вероятности можно использовать в качестве критической области, отвергая нулевую гипотезу, если $\hat{\theta}$ попадает в нее. Конечно, при этом сохраняется возможность совершить ошибку первого рода, хотя вероятность данной ошибки, равная q, по условию мала и определяется выбором уровня значимости.

Вид критической области зависит от характера *альтернативной гипотезы* H_1, то есть той гипотезы, которая противопоставляется нулевой и в пользу которой склоняется исследователь, отвергая проверяемую нулевую гипотезу.

Если нулевой гипотезе H_0: $\theta = \theta_0$ противопоставляется альтернативная гипотеза H_1: $\theta \neq \theta_0$, то критерий для проверки H_0 носит название двустороннего, а его критическая область состоит из двух частей (рисунок 4.1).

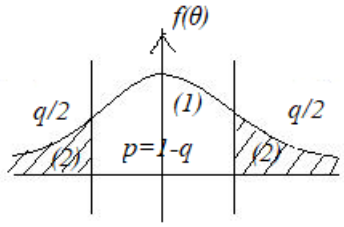


Рисунок 4.1 – Область принятия гипотезы (1) и критическая область (2) для двустороннего критерия

Обычно границы критической области выбираются таким образом, чтобы вероятность попадания в левую и правую части критической области была бы одинакова (равна *q/2*). Тогда двусторонняя критическая область определяется неравенством $|k| > k_{kp}$.

Если же при нулевой гипотезе H_0:$\theta=\theta_0$ альтернативная гипотеза формулируется как H_1: $\theta>\theta_0$ или H_1: $\theta<\theta_0$, то соответствующие критерии называются *односторонними* (правосторонний и левосторонний) и их критические области содержат всего одну зону (рисунок 4.2).

$k < k_{kp}$ $k > k_{kp}$

а) б)

Рисунок 4.2 – Область принятия (1) гипотезы и критическая область для одностороннего критерия

а) H_1: $\theta<\theta_0$я, б) H_1: $\theta>\theta_0$

Основные этапы построения критерия значимости:

1) выдвижение нулевой гипотезы H_0;

2) определение альтернативной гипотезы H_1;

3) установление подходящего уровня значимости q (вероятности ошибки первого рода);

4) выбор подходящей статистики g для проверки гипотезы. Эта статистика должна, по возможности, удовлетворять таким требованиям: а) просто вычисляться; б) иметь известное (лучше табулированное), выборочное распределение; в) быть наилучшей (например, в смысле получения наиболее мощного критерия). Для проверки многих гипотез подобные статистики известны и хорошо изучены (см. п. 3.6);

5) определение выборочного распределения используемой статистики в предположении, что гипотеза H_0 верна;

6) установление критической области для проверки гипотезы с учетом альтернативной гипотезы H_1 и уровня значимости q; для этого задаемся

уровнем значимости q и критические точки ищем, исходя из следующих соотношений:

- для правосторонней критической области $P(k > k_{kp}) = q(k_{kp} > 0)$;

- для левосторонней критической области $P(k < k_{kp}) = q(k_{kp} < 0)$;

- для двусторонней симметричной области $P(k > k_{kp}) = \frac{q}{2}(k_{kp} > 0)$; $P(k < -k_{kp}) = \frac{q}{2}$;

7) получение случайной выборки объема *N*, вычисление значения статистики и принятие решения: если статистика попала в область принятия решения, то гипотеза H_0 – справедлива; если же статистика оказалась в критической области, то H_0 отвергается (H_0 будет отвергнута в *100%* случаев, когда она верна).

Пример 4.1. Значения статического коэффициента передачи тока для некоторого прибора являются случайными, подчиняющимися нормальному распределению величинами с параметром $\sigma_x^2 = 25$. Номинальное значение параметра равно 50.

Требуется, измерив значения коэффициента для 20 приборов, выяснить, обеспечивается ли равенство коэффициентов номинальному значению, если нежелательным является его отклонение в меньшую сторону.

Последовательность проверки:

1) формулируем нулевую гипотезу:

$$H_0: m_x = 50;$$

2) в качестве альтернативной выбирается гипотеза:

$$H_1: m_x < 50$$

(так как нежелательно отклонение коэффициента в меньшую сторону);

3) выберем уровень значимости $q = 0.05$;

4) так как известна дисперсия исследуемой величины, наилучшая для использования статистика:

$$g = \frac{\overline{x} - m_x}{\sigma_x/\sqrt{N}};$$

5) статистика g имеет нормированное нормальное распределение;

6) граница критической области одностороннего критерия находится по таблице нормированного нормального распределения и равна:

$$-U_{\alpha=q} = -U_{\alpha=0.05} = -1{,}64$$

(берем со знаком минус, так как критическая область размещена слева);

7) проведено 20 независимых измерений коэффициента; получено среднее $\bar{x}$, рассчитано выборочное значение статистики g=-1,43.

Так как выборочное значение статистики не попадает в критическую область, следует принять нулевую гипотезу, считая, что результаты эксперимента не противоречат высказанному предположению.

Критерии для проверки различных гипотез, касающихся значений параметров генеральной совокупности, приводятся в литературе по математической статистике. Их можно использовать, если выполняются следующие условия:

- выборки являются случайными, с независимыми наблюдениями;
- исследуемые совокупности имеют нормальное распределение.

Эти критерии не исчерпывают все возможные их разновидности.

4.3. Отсев грубых погрешностей

Как известно из теории вероятности, при нормальном законе распределения случайная величина не должна отличаться от своего математического ожидания с вероятностями: 90% на 1,64σ_x; 95% на 1,96 σ_x и т.п. Но это теоретические расчеты, в предположении, что известно математическое ожидание m_x, среднеквадратичное отклонение σ_x, а результаты измерений подчиняются нормальному закону распределения при числе испытаний N. Таким образом, в случае нормального закона распределения, при достаточно большом числе наблюдений (как показывает практический опыт, более 30), если какое-либо из измеренных значений x_{max} отличается от его математического ожидания более чем на 1,96·σ_x, то с вероятностью 95% его можно отбросить.

На практике, как правило, число измерений конечно и в большинстве случаев не превышает 15–30. При таком малом числе наблюдений мы можем определить только оценки математического ожидания m_x и среднеквадратичного отклонения σ_x.

Процедура отсева грубых погрешностей измерений заключается в следующем:

1. По результатам наблюдений (измерений) и объему выборки n рассчитываются оценки математического ожидания $\overline{x}$ и дисперсии S_x^2.

2. Из всего ряда наблюдений выбирается наблюдение (измерение), имеющее наибольшее отклонение от среднеарифметического значения x_{max}.

3. Формулируется нуль-гипотеза H_0: отклонение x_{max} от $\overline{x}$ несущественно с доверительной вероятностью p (уровнем значимости q).

4. Для оценки этой гипотезы рассчитывается максимальное относительное (по отношению к среднеквадратичному) отклонение:

$$t_{\text{эксп}} = \frac{|x_{max} - \bar{x}|}{S_x}, \qquad (4.1)$$

где x_{max} – «выскакивающее» значение.

5. В качестве критерия проверки статистической нуль-гипотезы используется теоретическое значение критерия Стьюдента *t*, зависящее от уровня значимости *q* или доверительной вероятности *p=1-q*. Значение $t_{q,m=N-1}$, представляет собой допустимое отклонение случайной величины, выраженное в долях оценки среднеквадратичного отклонения и учитывает ограниченность объема выборки (*N*) и заданную доверительную вероятность (*p*).

6. Если $t_{эксп}>t_{\alpha,m}$, то имеется достаточно оснований с вероятностью *p* исключить «выскакивающее» значение как грубую ошибку и отвергнуть нуль-гипотезу. В противном случае нуль гипотеза H_0 принимается и от отсева «выскакивающего» значения лучше воздержаться с вероятностью *p*.

Пример 4.1. Пирометром измеряется температура поверхности нагретого тела. Будем предполагать, что температура видимой поверхности нагретого тела во всех точках одинакова. Различными исследователями было проведено шесть измерений температуры и получены следующие их значения: ^{0}C: 925, 950, 975, 1000, 1025, 1050 (*N=6*).

Имеются ли среди этих измерений грубые погрешности? Предварительно вычислим оценки $\overline{x}$ и S_x:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 987.5^0C; \;\; S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{N-1}} = 46.8^0C.$$

В формуле для определения S_x использовали (*N-1*), так как истинное значение измеряемой температуры нам не известно. Заметим, что здесь это важно, поскольку выполнено мало измерений (всего *N=6*).

Выберем измерения, имеющие наибольшее отклонение от среднеарифметического значения.

Таких значений оказалось два: 925 ^{0}C и 1050 ^{0}C.

Предварительно вычислим:

$$t_{\text{эксп}} = \frac{1050 - 987{,}5}{46{,}8} = 1{,}34.$$

При *q=0,05* и *m=N-1=5* определяем $t_{0,05;5}$=2,57.

Так как $t_{эксп}<t_{\alpha,m}$, то от отсева выделяющихся наблюдений лучше воздержаться.

Особенности статистического вывода. Статистический вывод, как любой статистический метод, содержит определенные противоречия. Особенность статистического вывода такова, что при отклонении гипотезы можно оценить заранее вероятность возможной ошибки, когда отбрасывается правильная гипотеза. Но если гипотеза принята, это не значит, что она

подтверждена с заданной вероятностью. Она согласуется с результатами данного статистического эксперимента, хотя, может быть, при получении дополнительного экспериментального материала окажется, что ее следует отвергнуть.

Иными словами, методика проверки статистических гипотез позволяет отбросить ту или иную гипотезу как неправильную, но не позволяет доказать, что она верна. То есть статистические критерии указывают лишь на отсутствие опровержения со стороны данных опытов.

На практике можно предложить много разных гипотез, одинаково хорошо согласующихся с результатами наблюдений, и никогда нельзя гарантировать, что рассмотрены все возможные гипотезы. Формализованные процедуры проверки статистических гипотез дают лишь правила при решении вопроса об отбрасывании или принятии гипотез, причем результатам статистического анализа не надо приписывать слишком большую объективность, поскольку всегда используется и априорная, субъективная информация.

4.4 Контрольные вопросы

4.4.1 Дайте определение статистической гипотезы.
4.4.2 Что такое нулевая гипотеза? Как она формулируется?
4.4.3 Как формируется альтернативная гипотеза?
4.4.4 Поясните виды ошибок при проверке гипотез.
4.4.5 Объясните понятие «уровень значимости».
4.4.6 Что такое критерий значимости?
4.4.7 Приведите стандартные критерии для проверки значений параметров генеральной совокупности.
4.4.8 Как определяется область принятия гипотезы при различных альтернативных гипотезах?
4.4.9 Что такое критическая область? Как она определяется?
4.4.10 Приведите основные этапы проверки гипотез.

5 Построение регрессионных моделей объектов

5.1 Элементы регрессионного анализа

Статистические методы планирования активного эксперимента являются одним из эмпирических способов получения математического описания статики сложных объектов исследования, т. е. уравнения связи отклика объекта $y(\bar{x})$ и независимых управляемых нормированных входных переменных (факторов) $\bar{x} = (x_1; x_2; \dots, x_n)$.

При этом математическое описание представляется в виде некоторого полинома — отрезка ряда Тейлора, в который разлагается неизвестная зависимость в окрестности основной точки $\bar{x}_0$:

$$y(\bar{x}) = \varphi(x_1, x, \dots, x_n) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{\substack{i=1,l \\ i<l}}^{n} \beta_{il} x_i x_l + \sum_{i=1}^{n} \beta_{ii} x_i^2,$$

где

$\beta_i = \frac{\partial \varphi}{\partial x_i}\Big|_{\bar{x}=\bar{x}_0}$, $\beta_{il} = \frac{\partial^2 \varphi}{\partial x_i \partial x_l}\Big|_{\bar{x}=\bar{x}_0}$, $\beta_{ii} = \frac{\partial^2 \varphi}{\partial^2 x_i}\Big|_{\bar{x}=\bar{x}_0}$ – теоретические коэффициенты.

Вследствие наличия неуправляемых и даже неконтролируемых факторов изменение величины *y* носит случайный характер, поэтому функциональная зависимость $\varphi(\bar{x})$ не дает точной связи между управляемыми факторами и откликом y_g объекта в каждом g-м опыте, а лишь между управляемыми факторами и математическим ожиданием случайной величины *y*:

$$M\{y_g\} = \varphi(\bar{x}_g),$$

здесь $\bar{x}_g = (x_{1g}; x_{2g}; \dots, x_{ng})$ – g -ая точка пространства независимых управляемых факторов (факторного пространства).

Классическим статистическим методом построения таких моделей объектов исследования является регрессионный анализ. В таком случае по результатам эксперимента можно отыскать уравнение регрессии в форме некоторого полинома:

$$\hat{y} = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{\substack{i;l=1 \\ i<l}}^{n} b_{il} x_i x_l + \sum_{i=1}^{n} b_{ii} x_i^2 + \cdots,$$

где выборочные коэффициенты регрессии b_0, b_i, b_{il}, $b_{ii,}$ … являются лишь оценками для теоретических коэффициентов, соответственно $\beta_0, \beta_1, \dots, \beta_{ii}$, а $\hat{y}$ – оценкой для *M{y}*.

Основное назначение регрессионного анализа – получение по экспериментальным данным регрессионной зависимости моделей объектов исследования вида:

$$y = \varphi(\bar{x}) + e,$$

то есть получение формульной зависимости, связывающей значение выходной переменной с факторами $\bar{x}$ при наличии аддитивной помехи *e* случайного характера.

Наиболее распространены модели, линейные по параметрам, то есть:

$$y = \varphi(\bar{x}) + e = \sum_{j=0}^{d} \beta_j f_j(\bar{x}) dx = \bar{f}^T(\bar{x})\bar{\beta} + e,$$

где $\bar{x} = \{x_1, \ldots, x_n\}$ – вектор факторов;

$\bar{\beta} = \{\beta_1, \ldots, \beta_n\}$ – вектор неизвестных коэффициентов регрессии;

$f(\bar{x}) = \{f_0(\bar{x}), f_1(\bar{x}), \ldots, f_d(\bar{x})\}$ – вектор базисных функций.

Система базисных функций должна быть выбрана до проведения эксперимента и последующих расчетов. Выбор этой системы производится обычно исходя из априорной информации о характере зависимости $\varphi(\bar{x})$ и обеспечения максимальной простоты для проведения расчетов. Теоретически считается, что вид модели известен и требуется по экспериментальным данным с помощью N опытов найти неизвестные $\bar{\beta}$. Реально же адекватную (соответствующую объектам исследования) модель приходится подбирать методом проб и ошибок. Наиболее часто в качестве базисных функций используются системы ортогональных полиномов (полиномы Лагерра, полиномы Лежандра), система тригонометрических функций, полиномиальные переменные. Важно отметить, что по известным измерениям факторов всегда можно рассчитать и значения базисных функций. Практически чаще применяются полиномиальные регрессионные модели. Тогда задача подбора адекватной модели сводится к выбору порядка полинома, начиная с простейшего (полинома первого порядка), с постепенным увеличением порядка. Обычно, чем больше число факторов, тем меньше максимальный порядок подбираемой модели.

Для построения таких моделей вводятся некоторые предпосылки относительно статистических свойств случайной составляющей e:

1. Результаты наблюдений $y_1,...y_N$ отклика в N точках факторного пространства представляют собой нормально распределенные случайные величины, то есть на них воздействуют нормально распределенные случайные помехи e с нулевым математическим ожиданием $M\{e\}=0$ и постоянной дисперсией $\sigma_e^2=const$. Постоянство дисперсии означает, что интенсивность шума не меняется в различных наблюдениях. Данное требование необходимо для корректного выполнения всех статистических процедур в регрессионном анализе.

2. Значения помехи в различных наблюдениях некоррелированы. Для обеспечения этого требования используется рандомизация опытов, когда порядок их реализации выбирается случайным.

3. Дисперсии $\sigma^2\{y_g\}$ $g=1,..., N$ равны. Иными словами, получаемые в результате многократных повторных наблюдений над величиной y_g в точке x_g выборочные оценки $S_g^2\{y\}$ однородны, а дисперсия $\sigma^2\{y_g\}$ не зависит от математического ожидания $M\{y_g\}$, то есть не отличается от дисперсии $\sigma^2\{y_g\}$, полученной при повторных наблюдениях в любых других точках факторного пространства (воспроизводимость с равной точностью). Это означает постоянство дисперсии помехи, равной дисперсии воспроизводимости

σ_e^2=*const*. Это значит, что интенсивность шума не меняется при изменении факторов.

При выполнении измерений в различных условиях возникает задача сравнения точности измерений. При этом следует подчеркнуть, что экспериментальные данные можно сравнить только тогда, когда их дисперсии однородны. Это означает, как уже отмечалось, принадлежность экспериментальных данных к одной и той же генеральной совокупности. Напомним, что однородность дисперсий свидетельствует о том, что среди сравниваемых дисперсий нет таких, которые с заданной надежностью превышали бы все остальные, то есть была бы большая ошибка.

4. Независимые управляемые факторы $x_1,...x_n$ измеряются с пренебрежимо малыми ошибками по сравнению с ошибками в определении *y*. То есть вклад, вносимый случайными ошибками измерений факторов $\bar{x}$ в дисперсию воспроизводимости, должен быть пренебрежимо мал по сравнению с действием других неконтролируемых факторов, образующих *e*.

Это предположение можно проверить лишь после получения регрессионной модели, так как степень влияния ошибки измерения некоторого фактора зависит не только от величины ошибки, но и от того, насколько сильно сам фактор влияет на *y*. Чем сильнее его влияние, тем в большей степени скажется и ошибка в его измерении. Поэтому эта предпосылка просто постулируется и лишь затем, если найденная регрессионная модель окажется неработоспособной, начинают анализировать, не связано ли это с наличием существенных ошибок в измерении x.

5. Вектор-столбцы базисных функций должны быть линейно-независимыми, то есть:

$$\sum_{g=1}^{N}[af_j(\bar{x}_g) + bf_i(\bar{x}_g)] \neq 0 \text{ при } \forall\, a \text{ и } b \neq 0.$$

Выполнение этого требования необходимо для получения раздельных оценок всех коэффициентов регрессии β, входящих в регрессионную модель. В активном эксперименте это обеспечивается соответствующим выбором плана эксперимента. В пассивном эксперименте его выполнение должно быть проверено в ходе расчетов. Строгая линейная зависимость вряд ли будет иметь место, но может наблюдаться сильная коррелированность столбцов. Это обстоятельство хоть и не препятствует оцениванию всех коэффициентов регрессии, но может повлечь за собой появление больших ошибок вычислительного характера. Рассматриваемое условие ограничивает и общее количество коэффициентов, входящих в регрессионную модель: оно не должно быть больше числа опытов, то есть $d+1 \leq N$.

Регрессионный анализ включат в себя:

а) оценивание неизвестных коэффициентов β регрессионной модели, а также дисперсии воспроизводимости σ_e^2;

б) статистический анализ результатов (выяснение точности оценивания отдельных коэффициентов регрессии, регрессионной модели в целом, проверки ее адекватности и работоспособности).

5.2 Оценки параметров регрессионной модели

При изучении множественной регрессии не существует графической интерпретации многофакторного пространства. При проведении экспериментов в такой ситуации исследователь записывает показания приборов о состоянии функции отклика и всех факторов, от которых она зависит. То есть исходным материалом для получения точечных оценок параметров регрессионной модели является матрица N наблюдений над значениями факторов $x_1,...x_n$ и отклика y:

$$\bar{x} = \begin{Vmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ x_N \end{Vmatrix} = \begin{Vmatrix} x_{11} \dots x_{1n} \\ x_{21} \dots x_{2n} \\ \dots\dots\dots\dots \\ x_{N1} \dots x_{Nn} \end{Vmatrix}; \quad y = \begin{Vmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{Vmatrix},$$

где x_{ij} – значение j-го фактора в i-м опыте;

x_{ij} – значение выходного параметра для i-го опыта.

Численные значения всех базисных функций в каждом из наблюдений тоже записываются в матричной форме:

$$\bar{F} = \begin{Vmatrix} \bar{f}_1(\bar{x}) \\ \bar{f}_2(\bar{x}) \\ \dots \\ \bar{f}_g(\bar{x}) \\ \dots \\ \bar{f}_N(\bar{x}) \end{Vmatrix} = \begin{Vmatrix} f_{10}\ f_{11}(\bar{x}) \dots f_{1j}(\bar{x}) \dots f_{1d}(\bar{x}) \\ f_{20}\ f_{21}(\bar{x}) \dots f_{2j}(\bar{x}) \dots f_{2d}(\bar{x}) \\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots \\ f_{g0}\ f_{g1}(\bar{x}) \dots f_{gj}(\bar{x}) \dots f_{gd}(\bar{x}) \\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots \\ f_{N0}\ f_{N1}(\bar{x}) \dots f_{Nj}(\bar{x}) \dots f_{Nd}(\bar{x}) \end{Vmatrix}.$$

Используя эту информацию, необходимо оценить неизвестные коэффициенты $\beta_0, \dots \beta_d$, то есть найти соответствующие значения *оценок* этих коэффициентов $\bar{b} = \{b_0, \dots b_d\}$. Эти оценки – случайные величины, так как результаты измерений сопровождаются случайными ошибками. Выбор критерия зависит от характера помех, то есть от их статистических свойств. При нормальном распределении помех наибольшую точность дает среднеквадратичный критерий, и задача решается просто (метод называется *методом наименьших квадратов*). Согласно этому методу значения $\bar{b}$ находятся из условия минимума суммы квадратов отклонений измеренных значений отклика от полученных с помощью модели, то есть путем минимизации следующей суммы:

$$S = \sum_{g=1}^{N} [y_g - \varphi(x_i, \bar{\beta})]^2 \rightarrow min$$

или

$$S = \sum_{g=1}^{N} [y_g - (b_0 f_{g0} + b_1 f_{g1} + \cdots + b_d f_{gd}]^2 \rightarrow min.$$

В последнем выражении коэффициенты b_i рассматриваются как переменные величины, подгоняемые под данный экспериментальный материал, а не как их конкретные значения. Величины b_i, при которых указанная сумма квадратов достигает минимума, принимаются за оценки коэффициентов регрессии.

Минимизация суммы квадратов производится обычным способом: путем приравнивания нулю частных производных по всем b_0, b_1, ..., b_d:

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2 \sum_{g=1}^{N} [y_g - (b_0 f_{g0} + b_1 f_{g1} + \ldots + b_d f_{gd})] \cdot f_{g0} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{\partial S}{\partial bd} = -2 \sum_{g=1}^{N} [y_g - (b_0 f_{g0} + b_1 f_{g1} + \ldots + b_d f_{gd})] \cdot f_{gd} = 0. \end{cases}$$

После преобразований получается система линейных алгебраических уравнений:

$$\begin{cases} b_0 \sum_{g=1}^{N} f_{g0}^2 + b_1 \sum_{g=1}^{N} f_{g0} f_{g1} + \cdots + b_d \sum_{g=1}^{N} f_{g0} f_{gd} = \sum_{g=1}^{N} y_g f_{g0}, \\ b_0 \sum_{g=1}^{N} f_{g0} f_{g1} + b_1 \sum_{g=1}^{N} f_{g1}^2 + \cdots + b_d \sum_{g=1}^{N} f_{g1} f_{gd} = \sum_{g=1}^{N} y_g f_{g1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_0 \sum_{g=1}^{N} f_{g0} f_{gd} + b_1 \sum_{g=1}^{N} f_{g1} f_{gd} + \cdots + b_d \sum_{g=1}^{N} f_{gd}^2 = \sum_{g=1}^{N} y_g f_{gd} \end{cases}$$

решая которую, можно получить искомые оценки b_0, b_1,...b_d. Эта система называется *системой нормальных уравнений*.

Система нормальных уравнений в матричной форме имеет вид:

$$(\bar{F}^T \cdot \bar{F}) \cdot \bar{b} = (\bar{F}^T \cdot y).$$

Матрица $\bar{F}^T \cdot \bar{F} = \overline{\Phi}$ – матрица известных коэффициентов этой системы:

$$\bar{F}^T \cdot \bar{F} = \bar{\Phi} = \left\| \begin{array}{cccc} \sum_{g=1}^{N} f_{g0}^2 & \sum_{g=1}^{N} f_{g0} f_{g1} & \cdots & \sum_{g=1}^{N} f_{g0} f_{gd} \\ \sum_{g=1}^{N} f_{g0} f_{g1} & \sum_{g=1}^{N} f_{g1}^2 & \cdots & \sum_{g=1}^{N} f_{g1} f_{gd} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{g=1}^{N} f_{g0} f_{gd} & \sum_{g=1}^{N} f_{g0} f_{gd} & \cdots & \sum_{g=1}^{N} f_{gd}^2 \end{array} \right\|.$$

Эта матрица называется *информационной матрицей Фишера* и состоит из (d+1) строки и (d+1) столбца. На пересечении j-ой строки и k-го столбца этой матрицы стоит сумма:

$$\sum_{g=1}^{N} f_{gj} f_{gk}.$$

Эта матрица симметрична относительно главной диагонали (это упрощает составление необходимой системы нормальных уравнений для той или иной конкретной регрессионной модели).

Для существования единственного решения нормальной системы необходимо и достаточно, чтобы $det\|\Phi\| \neq 0$. Это условие сводится к условию линейной независимости вектор-столбцов матрицы $\bar{F}$ (одна из предпосылок регрессионного анализа).

Решить систему нормальных уравнений можно любым известным способом. В общем виде в матричной форме решение можно записать так:

$$\bar{b} = \varphi^{-1} \cdot (\bar{F}^T \cdot y),$$

где φ^{-1} – матрица, обратная матрице φ.

Пусть C_{jk} – элемент матрицы φ^{-1}: $C_{jk} = \varphi_{kj}/|\varphi|$, где φ_{kj} – алгебраическое дополнение элемента f_{kj} в матрице φ. Тогда:

$$b_j = \sum_{k=0}^{d} C_{jk} \left(\sum_{g=1}^{N} y_g \cdot f_{gk} \right), j = 0, .., d.$$

Из этого выражения видно, что значения коэффициентов зависят от количества членов уравнений регрессии. То есть увеличение или уменьшение числа членов уравнения влияет на значения коэффициентов всех включенных

в уравнение членов. Такая неопределенность в оценивании коэффициентов регрессии крайне затрудняет их физическую интерпретацию.

Если же матрица Фишера диагональная (столбцы матрицы $\bar{F}$ попарно ортогональны), то:

$$C_{jl} = \sum_{g=1}^{N} f_{gj} \cdot f_{gl} = 0, j \neq l,$$

система нормальных уравнений распадается на независимые уравнения, и коэффициенты получаются независимо:

$$b_j = \sum_{k=0}^{d} C_{jj} (\sum_{g=1}^{N} y_g f_{gk}) , j = \overline{0, d}.$$

Таким образом, для получения независимых оценок коэффициентов $\bar{b}$. столбцы матрицы $\bar{F}$ должны быть линейно независимы и ортогональны. Линейная независимость ограничивает общее количество коэффициентов, входящих в регрессионную модель: $d+1 \leq N$ (например, три коэффициента, два фактора – минимум четыре опыта); ортогональность позволяет получить независимые оценки коэффициентов (независимо от числа членов уравнения регрессии).

Примеры. Запишем формулы оценок коэффициентов регрессии для простейших регрессионных моделей.

1. Модель выбрана в виде $y = \varphi(\bar{x}, \beta) = \beta x$.

Для нее система базисных функций $f_i(x)$ и матрица числовых значений базисных функций F имеют вид:

$$f_0(x) = 0; f_1(x) = x; N = 2,$$
$$F = \begin{Vmatrix} f_{10} & f_{11} \\ f_{20} & f_{21} \end{Vmatrix} = \begin{Vmatrix} 0 & x_1 \\ 0 & x_2 \end{Vmatrix}.$$

Тогда система нормальных уравнений записывается в следующем виде:

$$\begin{cases} b_0 \cdot 0 + b_1 \cdot 0 = y_1 \cdot 0 + y_2 \cdot 0 = 0, \\ b_0 \cdot 0 + b_1 \cdot (x_1^2 + x_2^2) = y_1 \cdot x_1 + y_2 \cdot x_2. \end{cases}$$

Отсюда находим оценку неизвестного коэффициента модели:

$$b = (\sum_{i=1}^{2} x_i y_i) / (\sum_{i=1}^{2} x_i^2).$$

В общем случае при N опытах имеем:

$$b = (\sum_{i=1}^{N} x_i y_i)/(\sum_{i=1}^{N} x_i^2).$$

2. Для модели следующего вида $y = \varphi(\bar{x}, \beta) = \beta_0 + \beta_1 x$

система базисных функций: $f_0(x) = 1; f_1(x) = x;$

матрица $F = \begin{Vmatrix} f_{10} & f_{11} \\ f_{20} & f_{21} \end{Vmatrix} = \begin{Vmatrix} 1 & x_1 \\ 1 & x_2 \end{Vmatrix};$

а система нормальных уравнений в случае двух экспериментов имеет вид:

$$\begin{cases} b_0 \cdot 2 + b_1 \cdot (x_1 + x_2) = y_1 + y_2, \\ b_0 \cdot (x_1 + x_2) + b_1 \cdot (x_1^2 + x_2^2) = y_1 \cdot x_1 + y_2 \cdot x_2. \end{cases}$$

В общем случае:

$$\begin{cases} b_0 \cdot N + b_1 \cdot \sum_{g=1}^{N} x_g = \sum_{g=1}^{N} y_g, \\ b_0 \cdot \sum_{g=1}^{N} x_g + b_1 \cdot \sum_{g=1}^{N} x_g^2 = \sum_{g=1}^{N} x_g y_g. \end{cases}$$

Решив эту систему, можно записать выражения для b_0 и b_1 (выражения для коэффициентов получаются громоздкими, поэтому не приводятся).

3. Аналогично для модели $\varphi(\bar{x}, \beta) = \beta_0 + \beta_1(x_1 - \bar{x}),$

где $\bar{x} = \frac{1}{N}\sum_{g=1}^{N} x_g,$

базисные функции имеют вид:

$$f_0(x) = 1; f_1(x) = x_1 - \bar{x}.$$

Эта линейная модель эквивалентна рассмотренной выше, но в данном случае задача получения оценки коэффициентов решается более просто, и для коэффициентов модели справедливы формулы:

$$b_0 = (\sum_{g=1}^{N} y_g)/N = \bar{y};$$

$$b_1 = (\sum_{g=1}^{N} (x_g - \bar{x}) \cdot y_g) = (\sum_{g=1}^{N} (x_g - \bar{x})^2).$$

Таким образом, используя различные модификации базисных функций и не изменяя в конечном счете вида искомой регрессионной модели, можно упростить расчеты. Как показывает практика, ввод центрированных переменных $x - \bar{x}$ всегда при этом желателен.

Свойства оценок коэффициентов.

Приведем основные свойства оценок параметров регрессионной модели:

1. Математическое ожидание оценок $M\{b_j\} = \beta_j, j = \overline{0,d}$, то есть оценки b_j являются несмещенными.

2. Дисперсии оценок коэффициентов равны $\sigma_{b_j}^2 = \sigma_e^2 \cdot C_{jj}$ и минимальны в классе линейных несмещенных оценок (отсюда следует, что кроме оценок коэффициентов, необходимы оценки дисперсии воспроизводимости).

3. Оценки $b_0,\ldots,b_d$ подчиняются совместному $(d+1)$-мерному распределению с вектором математических ожиданий $M\{\bar{b}\} = \bar{\beta}$.

Оценим точность определения коэффициентов для выше рассмотренных моделей.

Для модели $\varphi(\bar{x}, \beta) = \beta x$ имеем:

$$\sigma_b^2 = \sigma_e^2 / (\sum_{g=1}^{N} x_g^2).$$

При заданном наборе значений фактора x по этой формуле можно определить дисперсию оценок коэффициентов в долях от σ_e^2. Но можно использовать это соотношение и на стадии планирования эксперимента для выбора таких значений фактора x, которые обеспечивали бы при заданном числе наблюдений наилучшую точность оценивания коэффициента регрессии β, например, оценивания с минимальной дисперсией σ_b^2.

Для модели

$$\varphi(\bar{x}, \beta) = \beta_0 + \beta_1(x_1 - \bar{x}), \text{где } \bar{x} = \frac{1}{N}\sum_{g=1}^{N} x_g$$

можно получить:

$$\sigma_{b_0}^2 = \frac{\sigma_e^2}{N};\ \sigma_{b_1}^2 = \sigma_e^2 / [\sum_{g=1}^{N} (x_g - \bar{x})^2)].$$

Эти соотношения можно использовать как при обработке результатов эксперимента для расчета дисперсий оценок коэффициентов, когда заданы численные значения фактора x во всех N опытах, так и на этапе планирования эксперимента для выбора этих значений с тем, чтобы обеспечить наибольшую точность оценивания:

- $\sigma_{b_0}^2$ не зависит от x, а лишь от числа опытов;

- $\sigma_{b_1}^2$ зависит от x – чем шире диапазон изменения x, тем меньшего значения дисперсии $\sigma_{b_1}^2$ можно добиться; поэтому предпочтительно размещение точек плана на границе диапазона.

5.3 Статистический анализ результатов построения регрессионной модели

Статистический анализ позволяет выяснить качество полученной регрессионной модели: точность, адекватность и работоспособность.

5.3.1 Оценки дисперсии воспроизводимости.

Кроме оценок коэффициентов регрессии b_j в общем случае необходимы оценки дисперсии воспроизводимости σ_e^2.

Если заранее известно, что регрессионная модель адекватна объекту исследования, то в качестве оценки дисперсии σ_e^2 можно использовать так называемую остаточную дисперсию:

$$S_{ост}^2 = \frac{1}{N-(d+1)} \sum_{g=1}^{N} (y_g - \hat{y}_g)^2,$$

где $\hat{y}_g$ – предсказанное по уравнению регрессии значение отклика в опыте с номером g:

$$\hat{y}_g = b_0 f_0(\bar{x}_g) + b_1 f_1(\bar{x}_g) + \ldots + b_d f_d(\bar{x}_g).$$

Действительно, для адекватной модели единственной причиной различий между фактическим значением y_g в g-ом опыте и предсказанным по уравнению $\hat{y}_g$ может быть влияние шума. Поэтому $S_{ост}^2$ является оценкой σ_e^2:

$$S_e^2 = S_{ост}^2; \nu = N-(d+1), \nu - \text{число степеней свободы.}$$

Напомним, что величина, зависящая от элементов выборки и входящая в формулу дисперсии, называется *связью*. Разность между объемом выборки N и числом связей называется *числом степеней свободы*.

Если же вид адекватной модели неизвестен, то проводим проверку воспроизводимости эксперимента, то есть проверку предпосылки регрессионного анализа об однородности выборочных дисперсий S_g^2. Иными словами, хотим проверить гипотезу о равенстве генеральных дисперсий $\sigma^2\{y_1\} = \sigma^2\{y_2\} = \ldots = \sigma^2\{y_N\}$ при опытах в точках $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N$.

Оценки дисперсий находятся по известной формуле:

$$S_g^2 = \frac{1}{m-1}\sum_{k=1}^{m}(y_{gk} - \bar{y}_g)^2\,, \nu_{вос} = m - 1.$$

Для проверки гипотезы об однородности дисперсий S_g^2 используется критерий Кохрена (табулированная функция), который основан на законе распределения отношения максимальной оценки дисперсии к сумме всех сравниваемых оценок дисперсий:

$$G_{\text{кр}} = \max\{(y)\}/(\sum_{g=1}^{N} S_g^2\,(y)), \nu_{1вос} = m - 1, \nu_{2вос} = N.$$

И если значение критерия G, вычисленное по данным эксперимента, меньше критического значения, полученного по таблице (при числе степеней свободы числителя ν_1, знаменателя – ν_2 и уровне значимости, например, q=5%), то гипотеза об однородности отвечает результатам эксперимента и наилучшая оценка генеральной дисперсии воспроизводимости $\sigma^2\{y\}$ равна:

$$S_{вос}^2\{y\} = \frac{1}{N}\sum_{g=1}^{N} S_g^2\,\{y\} = \frac{1}{N(m-1)}\sum_{g=1}^{N}\sum_{k=1}^{m}(y_{gk} - \bar{y}_g)^2$$

с числом степенней свободы $\nu = N(m-1)$.

Если бы для получения оценок S_e^2 использовали специальные опыты (L опытов), результаты которых не использовались для получения модели, то:

$$S_e^2\{y\} = \frac{1}{L-1}\sum_{l=1}^{L}(y_l - \bar{y})^2\,, \bar{y} = \frac{1}{L}\sum_{l=1}^{L} y_l.$$

Если проверка воспроизводимости дала отрицательный результат, то надо увеличить число параллельных опытов, так как эксперимент невоспроизводим относительно управляемых факторов вследствие наличия больших шумов относительно неуправляемых и неконтролируемых факторов.

Полученные в результате решения задачи оценки коэффициентов $\bar{b}$ (и предсказательные значения отклика $\hat{y}$) тем предпочтительнее, чем меньше интегральные характеристики точности, например дисперсия оценок коэффициентов регрессии $\sigma_{b_j}^2$. Так как метод обработки задан (метод наименьших квадратов), система базисных функций выбрана, то эти характеристики зависят лишь от наблюдаемых во всех опытах значениях входов. При пассивном эксперименте извлечь пользу из этой зависимости нельзя. Другое дело, когда эксперимент активный, можно предусмотреть, при

какой комбинации значений $\bar{x}$ следует проводить каждый отдельный опыт, то есть можно заранее составить *план* эксперимента.

Для рассматриваемых регрессионных моделей, линейных по параметрам, планы не зависят от результатов измерений отклика, поэтому могут быть найдены заранее.

5.3.2 Проверка значимости коэффициентов регрессии.

После определения оценок $\bar{b}$ коэффициентов регрессии нужно проверить гипотезы об их значимости. В результате проверки устанавливается, статистически значимы или незначимы отличия от нуля оценок коэффициентов регрессии. То есть выясняется, обусловлено ли отличие b_i от 0 чисто случайными обстоятельствами, влиянием помехи или же это отличие неслучайно и вызвано тем, что в теоретической модели присутствует соответствующий коэффициент регрессии $\beta_j \neq 0$. Проверка значимости коэффициентов регрессии чаще всего производится раздельно, поочередно для каждой оценки b_i.

Итак, для каждого коэффициента проверяется нуль-гипотеза $\beta = 0$.

Проверка значимости сводится к последовательной проверке гипотез вида:

$$\begin{cases} H_0: \beta_j = 0, \\ H_1: \beta_j \neq 0, j = 0, .., d. \end{cases}$$

Проверку таких гипотез производят с помощью *критерия Стьюдента*, эмпирическое значение которого:

$$t_j = \frac{|b_j|}{S_{b_j}},$$

где $S_{b_j}^2 = S_e^2 \cdot C_{jj} = \frac{1}{Nm} S_e^2, j = 0, .., d$ – дисперсия оценки коэффициента b_j;

N – число точек факторного пространства, в которых проводится эксперимент.

m – число параллельных опытов в этих точках.

Если найденная величина параметра t_j превышает значение $t_{кр}$, определенное из таблицы t – распределения для числа степеней свободы $\nu_{зн} = N$, при заданном уровне значимости q (обычно 5%), то проверяемая нулевая гипотеза отвергается и соответствующую оценку b_j коэффициента признают статистически значимой; коэффициент должен быть сохранен. Это означает, что отличие оценки от нуля, видимо, не случайно; коэффициент должен быть сохранен в уравнении регрессии.

В противном случае нулевую гипотезу не отвергают и оценку b_j считают статистически незначимой. Его отличие от нуля можно объяснить чисто случайными обстоятельствами. В этом случае надо принять решение: оставить или отбросить незначимый коэффициент. Формально этот коэффициент должен быть исключен из уравнения и его необходимо отбросить, но практически решение принимается после анализа, особенно если t_j близко к критическому. Дело в том, что причины незначимости того или иного коэффициента могут быть принципиально различными. Они могут свидетельствовать о том, что в уравнении регрессии соответствующий коэффициент отсутствует, то есть его оценка незначима, исключение из эмпирического уравнения правомерно. Однако незначимость коэффициента может быть следствием того, что влияние этого коэффициента в ходе данного конкретного эксперимента не проявилось на фоне помехи. Поэтому для принятия решения следует проанализировать, нет ли дополнительных доводов в пользу его сохранения в модели, особенно если значение критерия близко к критическому табличному значению. Уверенность специалистов, что данный фактор необходим, может быть доводом для того, чтобы его оставить. И если первоначальная система базисных функций выбрана правильно, то отбрасывание тех, чьи коэффициенты незначительно отличаются от нуля по *t*-критерию, ведет к смещению оценок остальных коэффициентов регрессии.

Статистическая незначимость оценки b_j коэффициента регрессии может быть обусловлена следующими причинами:

1) данный j-й фактор не имеет функциональной связи с откликом y, то есть $\beta_j = 0$;

2) уровень x_{j0} базового режима $\bar{x}_0$ находится в точке частного экстремума функции отклика по фактору x_j и тогда $\beta_j = \frac{\partial y}{\partial x_j} = 0$;

3) интервал варьирования Δx_j выбран малым;

4) вследствие влияния неуправляемых и неконтролируемых факторов велика ошибка воспроизводимости эксперимента.

Процедура поочередной проверки значимости оценок коэффициентов имеет недостатки. Если матрица φ недиагональна, то, во-первых, уменьшается мощность *t*-критерия и он может отбрасывать как незначимые те коэффициенты, которые при других условиях были бы оставлены. Может получиться так, что все коэффициенты окажутся незначимыми. Во-вторых, может появиться неоднозначность в окончательном виде регрессионной модели, если изменяется порядок проверки значимости коэффициентов.

Чтобы исключить это обстоятельство, поступают следующим образом:

- вычисляют статистики t_j для всех b_j;

- упорядочивают их по возрастанию $t_{(0)} = \min t_j \leq t_{(1)} \leq t_{(2)} \leq t_{(d)} = max\ t_j$;

- $t_{(0)}$ сравнивают с критическим: если $t_{(0)} > t_{\text{кр}}$, то принимают решение о значимости всех коэффициентов; если же $t_{(0)} < t_{\text{кр}}$, то соответствующий коэффициент b_j является незначимым.

Если, несмотря на незначимость коэффициента, его оставляют в уравнении, то с табличным сравнивают следующую статистику $t_{(1)}$ и весь ход проверки повторяется; если же коэффициент отбрасывают, то на этом завершают процедуру.

5.3.3 Проверка адекватности регрессионной модели.

Такая проверка модели проводится для того, чтобы выяснить, правильно или нет выбран вид уравнения модели. Чтобы проверить гипотезу об адекватности математического описания опытным данным, достаточно оценить отклонение предсказанной по полученному уравнению регрессии величины отклика $\hat{y}_g$ от результатов наблюдений $\bar{y}_g$ в одних и тех же g-х точках факторного пространства. Проверяется модель со значимыми коэффициентами. Эта проверка возможна, если получена оценка дисперсии шума S_e^2 и выполнено условие $N > d^*$ (d^* – число отобранных коэффициентов). Действительно, если $N=d^*=d+1$, то проверять адекватность нет смысла, так как в этом случае $y_g = \bar{y}_g, g = 1,..,N,$ то есть поверхность, соответствующая модели, проходит точно через все экспериментальные точки. В этом случае не остается степеней свободы для проверки гипотезы об адекватности математического описания.

Рассеяние результатов наблюдений вблизи уравнения регрессии, оценивающего истинную функцию отклика, можно охарактеризовать с помощью дисперсии адекватности:

$$S_{ад}^2 = \frac{m}{N-d}\sum_{g=1}^{N}(\bar{y}_g - \hat{y}_g)^2,$$

где d – число членов аппроксимирующего полинома.

Дисперсия адекватности определяется числом степеней свободы $\nu_{зн} = N - d$.

Идея проверки адекватности состоит в следующем. Для адекватной модели остаточная дисперсия (дисперсия адекватности) является оценкой дисперсии шума $\sigma_e^2 (S_{ост}^2 = S_{ад}^2 = S_e^2)$. Если же модель неадекватна, то $S_{ост}^2 = S_{ад}^2$ будет оценкой σ_e^2 плюс некоторая компонента, обусловленная неадекватностью. То есть надо $S_{ост}^2$ сопоставить с какой-либо оценкой S_e^2 дисперсии шума σ_e^2. При адекватной модели отличие между ними должно быть числом случайным; при неадекватной $S_{ост}^2$ должно быть значимо больше S_e^2. Это сопоставление проверяется с помощью *F-критерия Фишера*:

$$F = \frac{S^2_{ост}}{S^2_e}.$$

Критерий Фишера позволяет проверить гипотезу об однородности двух выборочных дисперсий $S^2_{ад}$ и S^2_e. Если вычисленное по результатам наблюдений эмпирическое значение критерия F меньше критического ($F \leq F_{кр}$), найденного по таблицами критерия Фишера, для соответствующих степеней свободы $\nu_1 = N - d, \nu_2 = \nu_{зн} = N(m-1)$ при заданном уровне значимости $q_{ад}$, то гипотезу об адекватности не отвергают. В противном случае гипотезу отвергают, и математическое описание признается неадекватным.

Если гипотеза об адекватности отвергается, необходимо переходить к более сложной форме математического описания либо, если это возможно, проводить эксперимент с меньшим интервалом варьирования Δx_j. Максимальная величина интервала варьирования определяется условием адекватного описания объекта в области варьирования. Если при больших интервалах варьирования математическая модель неадекватна, то возникают систематические ошибки в определении коэффициентов, для уменьшения которых требуется сузить область варьирования. Однако с уменьшением интервала варьирования появляется целый ряд новых трудностей: растет отношение помехи к полезному сигналу, что приводит к необходимости увеличивать число параллельных опытов для выделения полезного сигнала на фоне шума, то есть уменьшаются абсолютные значения оценок b_j коэффициентов, величины которых непосредственно зависят от Δx_j, и оценки коэффициентов могут стать статистически незначимыми.

Для выбора интервала варьирования проводят предварительные эксперименты. Интервал варьирования можно выбирать равным 0.05–0.3 от допустимого диапазона варьирования факторов, то есть область варьирования составляет примерно 10–60% от всего диапазона. Начальную точку варьирования (базовую точку) выбирают возможно ближе к центру области факторного пространства, в которой ищется математическое описание объекта.

5.3.4 Анализ работоспособности модели.

С помощью такого анализа выясняется практическая возможность использования регрессионной модели для решения определенной задачи. (предсказание значения отклика, управление объектом и т. д.) Этот анализ необходим, так как иногда даже адекватные модели оказываются бесполезными из-за низкой точности.

Анализ включает в себя несколько процедур:

а) исследование остатков – разности между наблюдаемыми значениями отклика и предсказанными по уравнению регрессии $l_g = y_g - y_g$.

Если регрессионная модель подобрана правильно, то есть адекватна функции отклика, то остатки являются проявлением действия шума. Так как о свойствах шума были сделаны определенные предположения, то теперь можно выяснить, не противоречат ли свойства остатков этим предположениям. Подобный анализ проводится графически (зависимость l_g от времени, y_g, x_g и

т.д.). Просматривая остатки, можно понять, какова погрешность предсказания в точке, где проводились измерения, устраивает ли величина погрешности, при каких комбинациях факторов погрешность больше и т.д.

Исследование остатков плодотворно тогда, когда число наблюдений больше количества коэффициентов. Если это не так, то для проверки качества модели желательно провести дополнительные эксперименты;

б) вычисление *коэффициента детерминации*; этот коэффициент показывает, какая доля из общего рассеяния экспериментальных значений отклика относительно своего среднего обусловлена регрессионной зависимостью:

$$R^2 = \left[\sum_{g=1}^{N}(y_g - \bar{y}_g)^2\right] / [\sum_{g=1}^{N}(y_g - \hat{y}_g)^2].$$

Этот коэффициент меняется следующим образом: $0 \leq R^2 \leq 1$.

При $R^2 = 0$ поверхность отклика *y* определяется полностью случайными возмущениями, влияние факторов не обнаружено; при $R^2=1$ регрессионная кривая проходит через все экспериментальные точки.

Очевидно, чем ближе R к единице, тем лучше уравнение регрессии описывает изменение *y*, а малое *R* всегда говорит о низкой точности уравнения. Этот вывод справедлив при $N>d+1$, то есть когда количество опытов достаточно велико по сравнению с числом параметров модели. Когда число параметров модели приближается к числу наблюдений, остаточная сумма квадратов уменьшается и может быть близка нулю. Можно указать некоторую нижнюю границу значений коэффициента детерминации, выше которого уравнение регрессии представляет достаточный практический интерес для целей предсказания или других применений. Если граничные значения коэффициента детерминации выбираются из условия, что ошибка предсказания по уравнению регрессии должна быть в два раза меньше, чем ошибка предсказания по среднему значению отклика (не принимая во внимание факторы), то тогда $R^2_{min}\approx 0,75$. Для $R^2<0,75$ уравнение регрессии вряд ли работоспособно.

5.4 Контрольные вопросы

5.4.1 Дайте определение регрессионной модели.

5.4.2 Что собой представляет система базисных функций?

5.4.3 Перечислите предпосылки регрессионного анализа относительно статистических свойств случайной составляющей *e*.

5.4.5 Как представляется исходная информация для определения оценок параметров регрессионной модели?

5.4.6 Поясните составляющие матрицы $\bar{F}$.

5.4.7 Как формируется система нормальных уравнений?

5.4.8 Дайте определение информационной матрицы Фишера.

5.4.9 Как определяются оценки параметров регрессионной модели?

5.4.10 Приведите свойства оценок коэффициентов модели.

5.4.11 Как проверить воспроизводимость эксперимента?

5.4.12 При каких условиях не соблюдается требование воспроизводимости эксперимента и как следует поступить в этом случае?

5.4.13 Объясните понятие «значимость» коэффициентов. Как проверяется значимость оценок коэффициентов регрессии?

5.4.14 При каких условиях оценки коэффициентов регрессии незначимы и как эти условия устранить?

5.4.15 Дайте определение адекватной экспериментальной модели. Как проверить адекватность математической модели?

5.4.16 При каких условиях не соблюдается требование адекватности математической модели и как следует поступить в этом случае?

5.4.17 Какая модель является работоспособной? Как проверить работоспособность модели?

6 Планы регрессионного анализа

6.1 Основные принципы планирования экспериментов

Полученные в результате решения задачи построения регрессионной модели оценки коэффициентов модели $\bar{b}$ и предсказательные значения отклика $\hat{y}$ тем предпочтительнее, чем меньше интегральные характеристики точности, например дисперсия оценок коэффициентов регрессии $\sigma_{b_j}^2$. Для рассматриваемых линейных по параметрам моделей при выполнении основных предпосылок регрессионного анализа эти характеристики зависят от вектора базисных функций, значений факторов во всех опытах, а также метода обработки данных. Так как метод обработки задан (метод наименьших квадратов), система базисных функций выбрана, то эти характеристики зависят лишь от наблюдаемых во всех опытах значениях входов. При пассивном эксперименте извлечь пользу из этой зависимости нельзя. Другое дело, когда эксперимент активный, можно предусмотреть, при какой комбинации значений $\bar{x}$ следует проводить каждый отдельный опыт, то есть можно заранее составить *план* эксперимента. Тогда сразу обнаруживается, что различные планы могут отличаться значительно с позиций числовых характеристик точности построенных на их основе регрессионных моделей. То есть возникает необходимость выбора оптимального в некотором смысле плана эксперимента.

Основные принципы планирования экспериментов направлены на повышение эффективности экспериментирования, то есть на получение максимума информации при минимуме опытов и состоят в следующем:

1. *Отказ от полного перебора возможных состояний*. Для получения исчерпывающей информации о свойствах функции отклика необходимо

бесконечное число опытов во всех точках области планирования. В противном случае всегда есть вероятность пропустить некоторую особенность поверхности отклика. Это практически нереально. Поэтому задается дискретная сетка значений факторов, выбирается фиксированное число уровней каждого фактора. В результате какие-то особенности функции отклика будут не обнаружены. Опасность такого рода ошибок всегда существует при экспериментировании. Вообще-то необходимо увеличение числа уровней и факторов, что приводит к росту общего числа опытов. В теории планирования экспериментов отказываются от полного перебора входных состояний. Но на практике обычно задаются выбором вида функции отклика и выбор числа уровней варьирования по каждому фактору связан с этой функцией. Эту функцию выбирают, используя следующий принцип.

2. *Принцип постепенного усложнения математической мо*дели (принцип последовательного планирования). В отсутствие информации нет смысла сразу строить сложную модель, надо начать с простейшей. Если же модель непригодна, необходим следующий этап: постановка новых дополнительных опытов, позволяющих получить более сложную модель и т.д. до тех пор, пока не будет получена модель, которая будет признана достаточно хорошей. Если обратиться к широко используемым полиномиальным моделям, то надо начинать с линейной модели; в случае неудачи перейти к построению квадратичных моделей и т.д.

Приведенный подход составляет часть общего принципа последовательного эксперимента: после каждого шага производится анализ результатов (проверка качества результатов), а затем принимается решение о дальнейшей деятельности. В теории планирования эксперимента для проверки используются различные статистические процедуры, вытекающие из вероятностных свойств шума.

3. *Принцип сопоставления с шумо*м. Если случайная помеха значительна, то разумно ли тратить силы на получение сложной модели? Например, расточительно подключать приборы высокого класса точности для измерения переменной, отягощенной большой случайной ошибкой. То есть точность получаемой модели должна быть сопоставима с интенсивностью случайной помехи. Чем меньше уровень помехи, тем более точной (и, следовательно, сложной) должна быть модель; и чем выше уровень помехи, тем в большей степени можно ожидать, что более простая и менее точная модель окажется работоспособной. Так как многие реальные объекты характеризуются высоким уровнем помех, получили большое распространение полиномиальные регрессионные модели. С теоретической точки зрения эти модели менее содержательны, чем теоретические модели типа дифференциальных уравнений. Но с практической точки зрения являются весьма эффективным, а иногда и единственным средством изучения сложных объектов.

4. *Принцип рандомизации* (принцип приведения к случайности). Этот принцип состоит в такой организации эксперимента, которая позволяет сделать случайными (рандомизировать) систематически действующие помехи, не

поддающиеся учету. То есть, не в силах учесть действие неслучайных переменных, искусственно создаем в эксперименте случайную ситуацию. Например, систематические ошибки возможны из-за каких-то погрешностей при настройке экспериментальной установки. И если в серии из восьми опытов проводить первые четыре опыта на нижнем уровне первого фактора, то погрешность будет одинаковой для всех четырех опытов (если опыты проводить подряд).

5. *Принцип оптимальности планирования эксперимента* – центральный в теории планирования экспериментов. План должен обладать некоторыми оптимальными свойствами с точки зрения определенного заранее выбранного критерия оптимальности. Критерии могут формулироваться по-разному, зависеть от типа решаемой задачи, но принцип «*меньше опытов – больше информации*» сохраняется.

Таким образом, необходим такой план, который соответствовал бы экстремальным значениям выбранного показателя эффективности плана. Эти показатели называются *критериями оптимальности* регрессионных планов. В настоящее время известно несколько десятков таких критериев. Наиболее распространенные из них основаны на использовании введенных ранее различных характеристик точности регрессионного уравнения.

6.2 Критерии оптимальности планов

Существует две группы критериев оптимальности: к первой группе относятся критерии, связанные с точностью оценивания коэффициентов регрессии; вторую группу составляют критерии, связанные с предсказательными свойствами эмпирического уравнения регрессии. Рассмотрим несколько критериев из обеих групп:

1 группа. Наиболее часто используемый критерий – *критерий D-оптимальности:* план ε^* называется *D-оптимальным,* если:

$$det\, \Phi\,(\varepsilon^*) = \max_{\varepsilon \in \{\varepsilon\}} det\, \Phi\,(\varepsilon),$$

где $\{\varepsilon\}$ – множество планов заданной области планирования.

D-оптимальный план максимизирует величину определителя матрицы Φ или, что то же самое – минимизирует обобщенную дисперсию оценок коэффициентов регрессии. С геометрической точки зрения это означает минимизацию объема эллипсоида рассеяния данных оценок.

Критерий ортогональности. План ε^* называется *ортогональным*, если матрица Φ (а, следовательно, и Φ^{-1}) диагональная, то есть столбцы матрицы $\bar{F}$ являются попарно ортогональными векторами. Доказано, что при заданных значениях диагональных элементов матрицы Φ дисперсии оценок коэффициентов регрессии минимальны для ортогонального плана ε^*. Оценки

коэффициентов регрессии получаются независимо, что облегчает их анализ, резко сокращает все вычисления.

2 группа. План называется *G-оптимальным*, если:

$$\max_{x \in H} \sigma^2\{\hat{y}(\bar{x}, \varepsilon^*)\} = \min_{\varepsilon \in \{\varepsilon\}} \max_{x \in H} \sigma^2\{\hat{y}(\bar{x}, \varepsilon)\}.$$

То есть *G*-оптимальный план минимизирует максимальную дисперсию предсказания выхода модели по уравнению регрессии в области *H*.

Критерий ротатабельности. План ε^* называется *ротатабельным*, если дисперсия предсказания отклика по уравнению регрессии, полученному с помощью этого плана, постоянна на фиксированном расстоянии от центра эксперимента, то есть:

$$\sigma^2\{y(\bar{x}, \varepsilon^*)\} = h(\rho),$$

где $h(\rho)$ – некоторая функция расстояния ρ от центра планирования:

$\rho = \sqrt{x_i^2}$.

Для ротатабельного планирования точность предсказания одна и та же на равном расстоянии от центра эксперимента, независимо от того, на каком расстоянии находятся точки $\bar{x}$, где вычисляются $y(\bar{x})$.

Кроме этих критериев желательны и некоторые другие свойства планов (насыщенность и композиционность).

План называется *насыщенным*, если число наблюдений равно числу неизвестных параметров регрессионной модели. Желательно, чтобы любой план был близок к насыщенному, особенно на начальном этапе, когда требуется получить хотя бы приближенное представление об объекте, но зато с минимальными затратами.

План называется *композиционным*, если в его спектр в качестве составной части входят точки спектра плана, который был реализован при построении более простой модели. То есть, если модель, полученная на предшествующем этапе, не устраивает и требуется ее усовершенствование, например от полинома *k*-й степени переходим к полиному (*k+1*)-й степени, то точки спектра первого этапа используются и в дальнейшем. Композиционность плана – важное условие эффективности практической реализации принципа постепенного усложнения модели.

План называется *центральным*, если его центр расположен в начале координат.

Желательны еще простота вычислений коэффициентов моделей, симметрия в расположении точек плана и минимальное число уровней варьирования по каждому из переменных.

Для рассматриваемых регрессионных моделей, линейных по параметрам, планы не зависят от результатов измерений отклика, поэтому могут быть найдены заранее.

6.3 Планы первого порядка

С точки зрения структуры модели объектов могут быть линейно-параметризованными и нелинейно-параметризованными, то есть линейными и нелинейными по параметрам. Мы рассматриваем линейно-параметризованные модели. Такие модели могут быть линейными по факторам, неполно-квадратичными по факторам и квадратичными по факторам. Для рассматриваемых регрессионных моделей, линейных по параметрам, планы не зависят от результатов измерения отклика, поэтому могут быть найдены заранее. На практике чаще применяются полиномиальные регрессионные модели.

Планы первого порядка предназначены для экспериментального получения линейных регрессионных моделей следующего вида:

$$\varphi(X_1, \ldots, X_n) = B_0 + B_1 X_1 + \ldots + B_n X_n.$$

Для определения оценок неизвестных коэффициентов регрессии подобных моделей каждый из факторов должен варьироваться, по крайней мере, на двух уровнях. Примем $X_i^0 -$ базовый (основной) уровень, ΔX_i – шаг варьирования фактора X_i. Экспериментальные точки должны располагаться внутри или на границе гиперпараллелепипеда:

$$X_i^0 - \Delta X_i \le X_i \le X_i^0 + \Delta X_i, i = \overline{1, n}.$$

Для упрощения расчетов перейдем от обычного масштаба к нормализованному (стандартизованному):

$$x_i = \frac{X_i - X_i^0}{\Delta X_i}, i = \overline{1, n}.$$

Для нормализованных переменных имеем:

$$\varphi(\bar{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n;$$

$$\beta_i = B_i \Delta X_i, i = \overline{1, n}; \beta_0 = B_0 + \sum_{i=1}^{n} B_i X_i;$$

$$-1 \le x_i \le +1, i = \overline{1, n}.$$

Отсюда следует, что все точки плана, координаты которых представлены в стандартизованном масштабе, не должны выходить за пределы ограничивающего гиперкуба.

Известно несколько разновидностей планов первого порядка – однофакторный эксперимент, полный факторный эксперимент, дробный факторный эксперимент. Естественно, предполагается, что все основные предпосылки регрессионного анализа выполняются.

6.3.1 Однофакторный эксперимент.

Однофакторный эксперимент предусматривает поочередное варьирование каждого из факторов, в то время как все остальные факторы стабилизируются на некотором уровне. Для получения линейной модели достаточно двух уровней для каждого фактора. Будем считать, что каждый фактор принимает значения $x_{i\text{в}} = x_i^0 + \Delta x_i;\ x_{i\text{н}} = x_i^0 - \Delta x_i$, а все другие факторы в этот момент стабилизированы на своем базовом уровне $x_j^0, j \neq i$. Для нормализованных переменных $x_{iв} = +1, x_{iн} = -1, x_j = 0$.

Тогда матрица спектра выглядит следующим образом (рисунок 6.1, *а*) или в табличном виде (рисунок 6.1, *б*).

$$\left\|\begin{matrix} -1 & 0 & 0 \dots \dots \dots 0 \\ +1 & 0 & 0 \dots \dots \dots 0 \\ 0 & -1 & 0 \dots \dots \dots 0 \\ 0 & +1 & 0 \dots \dots \dots 0 \\ \dots & \dots & \dots \dots \dots \dots \dots \\ 0 & 0 & 0 \dots \dots -1 \\ 0 & 0 & 0 \dots \dots +1 \end{matrix}\right\|$$

а)

g	x_1	x_2	x_3	…	x_n
1	-1	0	…	…	0
2	+1	0	…	…	0
3	0	-1	0	…	0
4	0	+1	0	…	0
…	….	…	…	…	…
N-1	…	…	…	…	-1
N	…	…	…	…	+1

б)

Рисунок 6.1 – Матричный (а) и табличный (б) вид спектра плана однофакторного эксперимента

Геометрически точки плана располагаются в центре граней гиперкуба (в середине сторон ограничивающего квадрата для *n=2*).

Вектор-столбец базисных функций $\bar{f}^T(\bar{x}) = \|1, x_1, \dots, x_n\|$.

Вектор-столбцы матрицы численных значений базисных функций попарно-ортогональны. С учетом этого информационная матрица Фишера *Ф* и обратная ей *Ф*$^{-1}$ имеют вид, как на рисунке 6.2.

Отсюда получаем формулы для дисперсии оценок коэффициентов (напомним, что общая формула имеет вид $\sigma_{b_i}^2 = \sigma_e^2 \cdot C_{ii}, i = \overline{1,n}$, где C_{ii} – диагональный элемент матрицы *Ф*$^{-1}$):

$$\sigma_{b_0}^2 = \frac{\sigma_e^2}{2n}, \sigma_{b_i}^2 = \frac{\sigma_e^2}{2}, i = \overline{1,n}.$$

$$\Phi = \begin{Vmatrix} 2n & 0 & 0 \dots 0 \\ 0 & 2 & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots 2 \end{Vmatrix} \qquad \Phi^{-1} = \begin{Vmatrix} \frac{1}{2n} & 0 & 0 \dots 0 \\ 0 & \frac{1}{2} & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots \frac{1}{2} \end{Vmatrix}$$

а) б)

Рисунок 6.2 – Матрица нормальной системы уравнений (а) и обратная к ней (б)

Дисперсия предсказания по регрессионной модели находится из выражения:

$$\sigma^2\{\hat{y}(\bar{x}, \bar{b})\} = D\{b_0 + b_1 x_1 + \cdots + b_n x_n\} = \sigma_{b_0}^2 + \sum_{i=1}^{n} x_i \sigma_{b_i}^2 = \sigma_e^2 \left[\frac{1}{2n} + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right].$$

Для фиксированного расстояния от центра эксперимента $\rho = \sqrt{\sum_{i=1}^{n} x_i^2}$ эта дисперсия $\sigma_e^2 \left[\frac{1}{2n} + \frac{1}{2}\rho^2\right] = const$. То есть рассматриваемый план является ротатабельным.

С точки зрения других критериев оптимальности, однофакторный эксперимент неудовлетворителен. Поэтому не находит широкого применения.

6.3.2 Полный факторный эксперимент.

Полным факторным экспериментом (ПФЭ) называется эксперимент, реализующий все возможные неповторяющиеся комбинации *n* независимых управляемых факторов на всех уровнях их изменения. Если каждый фактор изменяется на двух уровнях, что достаточно для построения линейных регрессионных моделей, то общее число таких комбинаций равно $N = 2^n$. Такие планы обозначаются ПФЭ 2^n. 2^n – число элементов спектра плана.

Для *n*=2 матрица спектра плана ПФЭ 2^2 для нормализованных переменных имеет вид, приведенный на рисунке 6.3. Матрица плана ПФЭ 2^3 приведена на рисунке 6.4.

$$\left\|\begin{matrix}-1 & -1\\ +1 & -1\\ -1 & +1\\ +1 & +1\end{matrix}\right\|$$

g	x_1	x_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

а) б)

Рисунок 6.3 – Матричный (а) и табличный вид (б) спектра плана ПФЭ 2^2

Для записи спектра плана произвольного ПФЭ 2^n используют простое правило: сначала записывается столбец x_1, начиная с -1 в первой строке и поочередно изменяя знаки при переходе к очередной строке; для каждого последующего столбца частота изменения знаков в 2 раза меньше, чем в предыдущем столбце. Для свободного коэффициента регрессионной модели вносится фиктивная переменная x_0, значение которой во всех опытах равно +1.

Геометрически точки плана ПФЭ 2^n расположены в вершинах n-мерного гиперкуба (для n=2 – точки размещаются в вершинах ограничивающего квадрата, для n=3 – в вершинах куба).

$$\left\|\begin{matrix}-1 & -1 & -1\\ +1 & -1 & -1\\ -1 & +1 & -1\\ +1 & +1 & -1\\ -1 & -1 & +1\\ +1 & -1 & +1\\ -1 & +1 & +1\\ +1 & +1 & +1\end{matrix}\right\|$$

g	x_1	x_2	x_3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

а) б)

Рисунок 6.4 – Матричный (а) и табличный вид (б) спектра плана ПФЭ 2^3

ПФЭ 2^n позволяет в принципе оценить 2^n коэффициентов регрессионного при 2^n базисных функциях. Первые (n+1) базисных функций очевидны: $f_0(\bar{x}) = 1, f_1(\bar{x}) = x_1, \dots, f_n(\bar{x}) = x_n$. Они составляют линейную модель. Какие еще базисные функции могут входить в регрессионную модель? Напомним, что в число предпосылок регрессионного анализа входит условие линейной независимости столбцов численных значений этих базисных функций (столбцов матрицы $\bar{F}$). Так как в произвольный полином могут быть включены степени x_i или различные комбинации их произведений, а сами эти факторы в ПФЭ 2^n принимают значения, равные +1 или -1, то и столбец, отвечающий любой базисной функции, состоит из +1 и -1. Тогда условие линейной независимости можно сформулировать так: в матрице $\bar{F}$ не должно быть полностью совпадающих или полностью противоположных (по знаку)

столбцов. Отсюда следует, что в числе базисных функций не может быть функции вида x_i^k, где $k > 1$, целое, так как при четном k столбец состоит из +1, а при нечетном $x_i^k = x_i$. В то же время любые комбинации произведений факторов могут быть в числе базисных функций.

Например, для $n=2$ набор базисных функций:

$$f_0 = 1, f_1 = x_1, f_2 = x_2,$$

а также можно включить $f_3 = x_1 \cdot x_2$;

при $n=3$ система базисных функций имеет вид:

$$f_0 = 1, f_1 = x_1, f_2 = x_2, f_3 = x_3, f_4 = x_1x_2, f_5 = x_1x_3, f_6 = x_2x_3, f_7 = x_1x_2x_3.$$

Каждой из функций в этих наборах соответствует свой, отличный от других столбец (рисунки 6.4, 6.5). На этих рисунках прямоугольной рамкой выделен спектр плана.

g	$f_0=1$	$f_1=x_1$	$f_2=x_2$	$f_3=x_1x_2$
1	+1	-1	-1	+1
2	+1	+1	-1	-1
3	+1	-1	+1	-1
4	+1	+1	+1	+1

Рисунок 6.4 - Матрица плана ПФЭ 2^2

g	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7
1	+1	-1	-1	-1	+1	+1	+1	-1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	+1	-1	+1	-1	-1	-1
5	+1	-1	-1	+1	+1	-1	-1	+1
6	+1	+1	-1	+1	-1	+1	-1	-1
7	+1	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1	+1

Рисунок 6.5 – Матрица плана ПФЭ 2^3

Таким образом, с помощью ПФЭ 2^n возможно получение регрессионных моделей вида:

$$\varphi(\bar{x}) = \sum_{j=0}^{2^n-1} \beta_j f_j(\bar{x}) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1<i_2}}^{n} \beta_{i_1 i_2} x_{i_1} x_{i_2}$$

$$+ \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{\substack{i_3=1 \\ i_1<i_2<i_3}}^{n} \beta_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + + \ldots + \beta_{12\ldots n} x_1 x_2 \ldots x_n$$

Общее количество коэффициентов $1 + n + C_n^2 + C_n^3 + \cdots + C_n^n = 2^n$, то есть план ПФЭ является насыщенным.

Матрица $\bar{F}$ численных значений базисных функций обладает следующими свойствами:

1) свойство симметричности: алгебраическая сумма элементов вектор-столбцов базисных функций равна нулю (за исключением столбца, соответствующего свободному члену):

$$1) \sum_{g=1}^{N} f_j\left(x_g\right) = 0, j = 1, \ldots, d-1;$$

$$\sum_{g=1}^{N} f_0\left(x_g\right) = N;$$

2) свойство ортогональности столбцов матрицы значений базисных функций: скалярное произведение всех вектор-столбцов (сумма почленных произведений любых двух вектор-столбцов матрицы) равно нулю:

$$\sum_{g=1}^{N} f_j(\bar{x}_g) f_k(\bar{x}_g) = 0, j \neq k, k = 0, \ldots, N-1.$$

Благодаря этому свойству уменьшаются трудности, связанные с расчетом коэффициентов уравнения регрессии;

3) свойство нормирования: сумма квадратов элементов каждого столбца равна числу опытов:

$$\sum_{g=1}^{N} f_j^2\left(\bar{x}_g\right) = N.$$

Эти соотношения записываются в предположении, что дублирование опытов не производится. Отсюда находим:

$$\Phi = \begin{Vmatrix} N & 0 & 0 \dots 0 \\ 0 & N & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots N \end{Vmatrix} \qquad \Phi^{-1} = \begin{Vmatrix} \frac{1}{N} & 0 & 0 \dots 0 \\ 0 & \frac{1}{N} & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots \frac{1}{N} \end{Vmatrix},$$

то есть ПФЭ 2^n является ортогональным планом.

4) $\sigma_{b_j}^2 = \frac{\sigma_e^2}{N}$ – это означает, что все коэффициенты рассчитываются с равной точностью;

5) дисперсия предсказания отклика по линейной регрессионной модели, полученной с помощью ПФЭ 2^n:

$$\sigma^2 \hat{y}(\bar{x}, \bar{b}) = D b_0 + \sum_{i=1}^{n} b_i x_i =$$

$$= \sigma_{b_0}^2 + \sigma_{b_i}^2 \sum_{i=1}^{n} x_i^2 = \frac{\sigma_e^2}{N}\left(1 + \sum_{i=1}^{n} x_i^2\right) = \frac{\sigma_e^2}{N}(1 + \rho^2).$$

То есть, для фиксированного ρ дисперсия является постоянной, следовательно, план ПФЭ 2^n является ротатабельным. Он также является *A-, E, D-, G*-оптимальным. Следовательно, при достаточной простоте построения ПФЭ 2^n обладает хорошими свойствами.

6.3.3 Нелинейная регрессионная модель.

Линейные модели оказываются адекватными лишь при решении ограниченного круга задач. Если линейная модель оказалась неадекватной, то она достраивается до нелинейной. Одна из наиболее часто встречающихся причин нелинейности математических моделей заключается в том, что эффект (влияние на выходной параметр) одного фактора зависит от уровней, на которых находятся другие факторы. В этом случае говорят, что имеет место *взаимодействие между факторами.*

Пример: из практики бурения известно, что с увеличением осевой нагрузки на долото Gд и расхода бурового раствора Q в рабочей области изменения этих параметров режима механическая скорость бурения Vм увеличивается. Пусть при увеличении только Gд на 15% Vм увеличивается на 20%, а при увеличении только Q на 20% Vм растет на 10%. Логично ожидать, что при одновременном увеличении Gд и Q на те же величины Vм вырастет на 20+10= 30%. Однако фактически Vм может увеличиться, например, на 40%.

Причина кроется в особенностях физического процесса – в механизме влияния осевой нагрузки на долото и расхода бурового раствора на процесс разрушения горных пород на забое бурящейся скважины. Механизм этого явления сводится к следующему. С увеличением осевой нагрузки на долото

процесс разрушения забоя интенсифицируется, механическая скорость проходки растет. Это происходит до определенного предела – пока образующийся шлам своевременно уносится потоком жидкости, и долото работает по чистому забою. При дальнейшем увеличении осевой нагрузки на долото шлама на забое становится больше, прежнего расхода раствора уже недостаточно, скорость бурения увеличивается менее значительно. Если в этот момент расход бурового раствора значительно увеличить, то скорость бурения начнет увеличиваться с гораздо большим темпом. То есть эти два параметра режима бурения в данном случае усиливают действие друг друга. Факторы могут влиять и в сторону ослабления действия друг друга.

Определение порядка взаимодействия факторов.

При двух факторах x_1, x_2 может иметь место взаимодействие лишь между этими факторами, которое принято записывать в виде $x_1 \cdot x_2$. Говорят, что имеет место парное взаимодействие или взаимодействие 2-го порядка. Между тремя факторами x_1, x_2, x_3 число взаимодействий может быть равно четырем: три двойных взаимодействия $x_1 \cdot x_2, x_1 \cdot x_3, x_2 \cdot x_3$ и еще одно, которое называется тройным взаимодействием или взаимодействием 3-го порядка – $x_1 \cdot x_2 \cdot x_3$. С. ростом числа факторов число возможных взаимодействий растет. Это число находится по формуле комбинаторики – число сочетаний из k элементов по m:

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

где k – число факторов;

m – число элементов во взаимодействии.

Если известны механизмы таких взаимодействий между факторами, то они могут с успехом использоваться при решении многих задач в различных областях и техники, и жизнедеятельности людей.

ПФЭ позволяет количественно оценить эффект взаимодействия факторов независимо от того, известен его механизм или нет. Если механизм взаимодействия известен, то это позволяет целенаправленно проводить исследования, точнее выбирать уровни факторов и интервалы варьирования.

Для учета эффектов взаимодействий факторов, пользуясь правилом перемножения столбцов, получают столбец соответствующего произведения факторов, которое включают в математическую модель. При вычислении коэффициента, соответствующего эффекту взаимодействия, с новым вектор-столбцом поступают как с новым фактором. Очень важным является то, что при добавлении столбцов эффектов взаимодействий все рассмотренные оптимальные свойства матрицы планирования сохраняются. Столбцы независимых факторов задают собственно план эксперимента, по ним непосредственно определяют условия проведения опытов, а все остальные столбцы, в том числе столбец фиктивной переменной и вектор-столбец результатов опытов y служат для расчета.

Пользуясь таким планом, можно вычислить коэффициенты регрессии уравнения неполной степенной модели. Общее число всех возможных эффектов, включая b_0, линейные эффекты и эффекты взаимодействий всех порядков, равно числу опытов ПФЭ.

Ортогональность матрицы планирования позволяет получить независимые друг от друга оценки коэффициентов. Однако это утверждение справедливо лишь для случаев, если модель включает только линейные эффекты и эффекты взаимодействия. Между тем существенными могут оказаться коэффициенты при квадратах факторов, их кубах и т.д.

С практической точки зрения ПФЭ присущ существенный недостаток, связанный с быстрым ростом количества опытов при увеличении числа факторов n (например, при $n=10$, количество опытов $N=1024$). Вместе с тем, трудно представить функцию отклика, для которой квадратичные члены x_j^2 несущественны, но в то же время играют роль взаимодействия высокого порядка (вида $x_1 x_2 \dots x_n$). То есть эти модели для $n>5$ вряд ли находят практическое применение. Как правило, исследователь на первых порах ограничивается линейной моделью с дополнительными парными (в крайнем случае, тройными) взаимодействиями. В подобной ситуации ПФЭ становится избыточным. Возникает необходимость в построении более экономной с точки зрения числа опытов разновидности планов. Поэтому были разработаны дробные факторные эксперименты (ДФЭ).

6.3.4 Определение коэффициентов уравнения регрессии.

Учитывая свойства матрицы плана ПФЭ, оценки коэффициентов регрессионной модели при свободном и линейных членах получаются *независимо* и находятся по формулам:

$$b_0 = \frac{1}{N}\sum_{g=1}^{N} x_{0g}\bar{y}_g ;$$

$$b_i = \frac{1}{N}\sum_{g=1}^{N} x_{ig}\bar{y}_g \ (i = 1,2,\ \dots,\ n).$$

То есть любые коэффициенты уравнения регрессии определяются скалярным произведением столбца $\bar{y}$ на столбец соответствующего фактора.

Аналогичным образом определяются и коэффициенты при линейных взаимодействиях (двойных, тройных):

$$b_{il} = \frac{1}{N}\sum_{g=1}^{N} x_{ig}x_{lg}\bar{y}_g \ (i, l = 1,2,\ \dots,\ n; i \neq l).$$

$$b_{ilk} = \frac{1}{N}\sum_{g=1}^{N} x_{ig}x_{lg}x_{kg}\,\bar{y}_g\ (i,l,k = 1,2,\ \ldots,\ n; i \neq l \neq k).$$

Поскольку коэффициенты независимы, каждый коэффициент характеризует роль соответствующей переменной в процессе или силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает этот фактор. Если коэффициент имеет знак плюс, то с увеличением значения фактора отклик увеличивается, а если минус – уменьшается. После статистического анализа уравнения регрессии какие-то коэффициенты могут быть признаны незначимыми и факторы, имеющие коэффициенты, незначимо отличающиеся от нуля, выведены из состава уравнения, так как их влияние на параметры отклика будет отнесено к ошибке эксперимента. В этом случае, учитывая ортогональность плана, оставшиеся коэффициенты уравнения регрессии можно не пересчитывать. При отсутствии ортогональности плана эксперимента все коэффициенты необходимо пересчитывать заново.

6.3.5 Пример построения регрессионной модели методом ПФЭ.

Исследуется выход *y* некоторой химической реакции, который зависит от длительности реакции t и температуры T℃. Эту зависимость можно считать линейной в окрестности точки $t = 4$ ч.$, T = 220$℃.

Построим регрессионную модель этого процесса и проведем ее анализ. Для удобства обозначим длительность реакции t – x_1, температуру T℃ – x_2, выходная переменная – *y*.

Структура линейной модели имеет вид:

$$y = b_0 + b_1x_1 + b_2x_2,$$

число членов модели $d=3$.

Факторы и уровни их варьирования приведены в таблице:

Факторы	x_1	x_2
Основной уровень	4	220
Интервал варьирования	1	10
Верхний уровень	5	230
Нижний уровень	3	210

Нахождение модели методом планирования эксперимента ПФЭ 2^2 состоит из следующих этапов:

а) планирование эксперимента.

Обычно управляющие факторы составляют план эксперимента, остальные столбцы матрицы *F* получаются перемножением; в нашем случае матрица плана в нормализованном масштабе имеет вид:

x_1	x_2
-1	-1
+1	-1
-1	+1
+1	+1

б) проведение эксперимента.

Так как изменение выходной величины носит случайный характер, приходится в каждой точке плана проводить m параллельных опытов и рассчитывать среднее по результатам наблюдений:

$$\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i.$$

В каждой серии реализуется матрица планирования. Перед реализацией плана необходимо рандомизировать варианты варьирования, то есть определять случайную последовательность реализации вариантов матрицы планирования в каждой серии опытов.

Результаты экспериментов для $N=4$, $m=2$ приведены ниже (y^I, y^{II} – значения откликов в первой и второй серии экспериментов, $\bar{y}$ – среднее значение откликов):

g	x_1	x_2	y^I	y^{II}	$\bar{y}$
1	-1	-1	44,9	45,0	44,95
2	+1	-1	55,0	55,2	55,1
3	-1	+1	55,0	54,8	54,9
4	+1	+1	65,5	65,6	65,55

в) вычисление коэффициентов регрессии.

В силу ортогональности матрицы числовых значений базисных функций матрица Φ нормальной системы уравнений Фишера является диагональной, что позволяет сразу рассчитать коэффициенты модели. В результате получили модель:

$$y = 55.125 + 5.2x_1 + 5.1x_2.$$

Статистический анализ результатов.

г) проверка воспроизводимости выборочных дисперсий σ_i^2.

Это означает проверку предпосылки регрессионного анализа об однородности выборочных дисперсий σ_i^2.

Оценки дисперсий в каждой точке факторного пространства вычисляются по формуле:

$$S_g^2 = \frac{1}{m-1}\sum_{k=1}^{m}(y_{gk} - \bar{y}_g)^2, \nu_{вос} = m - 1.$$

Для проверки используется критерий Кохрена, который основан на законе распределения отношения максимальной оценки дисперсии к сумме всех сравниваемых оценок дисперсий:

$$G = \frac{max\{S_g^2\}}{\sum_{g=1}^{N} S_g^2\{y\}}; \nu_{1\text{вос}} = m - 1, \nu_{2\text{вос}} = N.$$

Если $G_{\text{расч}} < G_{\text{табл}}$ для $\nu_1 = m - 1$ и $\nu_2 = N$ и выбранного уровня значимости q, то гипотеза об однородности принимается. И наилучшая оценка генеральной дисперсии воспроизводимости $\sigma^2\{y\}$ равна:

$$S_{вос}^2\{y\} = \frac{1}{N}\sum_{g=1}^{N} S_g^2\{y\}$$

с числом степенней свободы $\nu = N(m-1)$.

Если проверка воспроизводимости дала отрицательный результат, то надо либо увеличить количество параллельных опытов, либо признать эксперимент невоспроизводимым относительно управляемых переменных вследствие наличия флуктуаций неуправляемых и неконтролируемых переменных, создающих на выходе большой уровень «шума».

Расчеты приведены в таблице:

g	x_1	x_2	y^I	y^{II}	$y_{ср}$	S_q^2
1	-1	-1	44,9	45,0	44,95	0,0025
2	+1	-1	55,0	55,2	55,1	0,02
3	-1	+1	55,0	54,8	54,9	0,02
4	+1	+1	65,5	65,6	65,55	0,005
						∑=0,0475

$$G_{rasch} = \frac{max\{S_g^2\}}{\sum_{g=1}^{N} S_g^2\{y\}};$$

$G_{\text{расч}} = \dfrac{0{,}02}{0{,}0475} = 0{,}421.$

$G_{\text{табл}} = 0{,}9$ $(\nu_1 = m - 1 = 1, \nu_2 = N = 4)$.

$G_{\text{расч}} < G_{\text{табл}}$ – эксперимент воспроизводим.

Оценка дисперсии воспроизводимости:

$$S_e^2\{y\} = \frac{1}{N}\sum_{g=1}^{N} S_g^2\{y\} = 0{,}105.$$

Если проверка воспроизводимости дает отрицательный результат, то надо либо увеличить количество параллельных опытов, либо признать эксперимент невоспроизводимым относительно управляемых переменных вследствие наличия флуктуаций неуправляемых и неконтролируемых переменных, создающих на выходе большой уровень «шума».

д) исследование значимости коэффициентов проводится по критерию Стьюдента. Имеем:

$$C_{jj} = 4,\ S_a^2 = \frac{S_e^2}{4}=0{,}026.$$

Здесь C_{jj} – диагональный элемент матрицы Фишера.

Расчетные значения критерия для всех коэффициентов:

$$g_i = \frac{b_i}{S_b^2};\ g_1 = 2094, g_2 = 197.6, g_1 = 193{,}8.$$

Табличное значение критерия: $t_{\text{табл}} -= 2{,}13.$

Так как $t_j > t_{\text{табл}}$, то нуль-гипотеза отвергается, все коэффициенты являются значимыми.

е) проверка адекватности модели.

Дисперсия адекватности вычисляется по формуле:

$$S_{ad}^2 = \frac{1}{N-d}\sum_{g=1}^{N}(\bar{y}_g - \hat{y}_g)^2,$$

где *N=4, d=3*;

$\hat{y}_g$ – значение отклика, рассчитанное по уравнению регрессии.

Расчеты приведены ниже в таблице:

y_{cp}	$\hat{y}_g$	$(y_{cp} - \hat{y}_g)^2$
44,95	44,825	0.0156
55,1	55,225	0.000625
54,9	55,025	0,0156
65,55	65,425	0.000625
		∑=0,0312

Значение дисперсии адекватности: $S_{ad}^2 = 0.0312.$

Проверка адекватности с помощью критерия Фишера:

$$F_{rasch} = \frac{S_{ad}^2}{S_e^2} = \frac{0{,}0312}{0{,}105} = 0{,}297.$$

$F_{\text{табл}} = 7{,}71$ при $\nu_1 = N - d^* = 1, \nu_2 = N(m-1) = 4,$

здесь d^* – число членов модели со значащими коэффициентами (в примере все коэффициенты значимы).

$F_{\text{расч}} > F_{\text{табл}}$, значит, модель адекватна;

ж) проверка работоспособности модели. Рассчитываем коэффициент детерминации по формуле:

$$R^2 = \left[\sum_{g=1}^{N}(y_g - \bar{y}_g)^2\right] / [\sum_{g=1}^{N}(y_g - \hat{y}_g)^2].$$

Расчеты приведены ниже в таблице:

y^I	y^{II}	y_{cp}	$\hat{y}_g$	$\sum_{g=1}^{N}(y_g - \bar{y}_g)^2$	$\sum_{g=1}^{N}(y_g - \hat{y}_g)^2$
44,9	45,0	44,95	44,825	0,005	0,03625
55,0	55,2	55,1	55,225	0,02	0,05125
55,0	54,8	54,9	55,025	0,02	0,05125
65,5	65,6	65,55	65,425	0,005	0,03625
				∑=0,05	∑=0,175

$R^2 = 0.2857 < 1.$

Малое R говорит о низкой точности уравнения, работоспособность модели низкая. Но этот вывод справедлив, когда количество опытов достаточно велико по сравнению с числом параметров модели. В нашем случае число параметров модели приближается к числу наблюдений. Для проверки качества модели желательно провести дополнительные эксперименты.

6.3.6 Дробный факторный эксперимент.

Во многих практических задачах идентификации влияние взаимодействий (произведений факторов) второго и высших порядков отсутствует или пренебрежимо мало. Например, при настройке автоматических систем регулирования энергоблоков в базовом режиме достаточно знать статистические характеристики регулируемых параметров лишь в линейном приближении. Кроме того, на первых порах часто нужно получить в первом приближении лишь линейную аппроксимацию изучаемого уравнения связи при минимальном количестве опытов. В других задачах, кроме основных факторов, необходимо учитывать лишь некоторые взаимодействия из множества возможных взаимодействий, входящих в планы ПФЭ. Поэтому неэффективно

использовать ПФЭ для оценивания коэффициентов лишь при линейных членах и некоторых парных произведениях из-за реализации большого числа вариантов варьирования (2^n), в особенности при большом числе факторов n. При линейном росте числа независимых факторов число вариантов варьирования растет по показательному закону, в результате чего на проверку гипотезы об адекватности остается излишне много степеней свободы.

Поэтому возникает проблема сокращения числа опытов. При этом задача заключается в том, чтобы не только уменьшить число опытов, но и получить достаточное количество информации об объекте исследования. Уменьшить число опытов можно за счет использования избыточности ПФЭ. Обычно число опытов в ПФЭ больше, чем количество значимых (или интересующих исследователя) коэффициентов. Например, неполная степенная двухфакторная модель имеет вид:

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

Предположим, из априорных сведений известно, что при выбранных интервалах варьирования поведение объекта с требуемой точностью описывается линейной моделью:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3.$$

В таких случаях для построения статистической модели объекта используют дробный факторный эксперимент, в процессе проведения которого реализуется *часть* ПФЭ (*дробная реплика*). Эксперимент, реализующий часть полного факторного эксперимента, называется *дробным факторным экспериментом* (ДФЭ). ДФЭ позволяет получить линейное приближение искомой регрессионной зависимости в некоторой небольшой окрестности точки базового режима при минимуме опытов.

Для вышеприведенного примера нелинейный член $b_{12} x_1 x_2$, характеризующий взаимодействие факторов, может не учитываться. Поэтому вектор-столбец $x_1 x_2$ может быть использован для введения в план ПФЭ какого-либо нового фактора, не увеличивая числа опытов, например фактора x_3. Эта матрица будет половиной более крупного плана. То есть для оценки влияния трех факторов можно воспользоваться половиной ПФЭ 2^3.

Если рассматривается шестифакторная модель и за счет малозначимых взаимодействий факторов, построенных из x_1, x_2, x_3, вводится дополнительно три новых фактора, например, x_4, x_5, x_6, то полученный новый план будет представлять 1/8 часть ПФЭ2 ^{6}. Какую именно восьмую часть представляет этот план, зависит от знаков столбцов x_4, x_5, x_6, определяемых знаками взаимодействий, вместо которых будут введены новые факторы x_4, x_5, x_6.

При большом числе факторов n для получения линейного приближения можно построить дробные реплики высокой степени дробности. Планы ДФЭ 2^{n-p} называются *репликами степени дробности p*; n – количество факторов. В

обозначении ДФЭ даже для конкретных *n* и *p* вычитание не производится, а сохраняются обе цифры, где первая характеризует число факторов, а вторая – степень дробности ДФЭ. Для вышеприведенных примеров для трехфакторной модели можно реализовать ДФЭ 2^{3-1} – 1/2 реплика от ПФЭ типа 2^3. Для шестифакторной модели – ДФЭ 2^{6-3} -1/8 реплика от ПФЭ типа 2^6

Процедура получения плана ДФЭ: чтобы ввести в исходный план ПФЭ новый фактор, не увеличивая число опытов, этому фактору необходимо присвоить вектор-столбец взаимодействия, построенного из исходных факторов и влиянием которого можно пренебречь. Значение нового фактора, введенного в исходный план, должно изменяться в соответствии со знаками этого столбца. То есть основная идея – сокращение числа опытов. Такая экономия достигается за счет определенной априорной информации о свойствах изучаемого объекта.

В эксперименте по плану ДФЭ пропускаются некоторые сочетания уровней факторов. Сокращение перебора уровней приводит к потере информации. Поэтому при дробном факторном эксперименте важно так спланировать эксперимент, чтобы терялась наименее существенная, при данной постановке задачи, информация. Особенно широко используется дробный факторный эксперимент, в котором теряется информация о взаимодействиях изучаемых факторов. Это правомерно в тех случаях, когда эффекты взаимодействия заведомо отсутствуют или настолько малы, что их можно не учитывать.

То есть мы должны иметь дополнительную априорную информацию о свойствах изучаемого объекта. Эта информация должна быть сформирована в виде списка *существенных переменных* – факторов и их взаимодействий, представляющих собой интерес для дальнейшего исследования. То есть «существенные переменные» — это переменные, отобранные для возможного включения в модель. Конечно, в ходе дальнейшего исследования может оказаться, что часть этих переменных не влияет на отклик. В списке существенных переменных всегда присутствует фиктивная переменная x_0, которой соответствует коэффициент b_0. Список существенных переменных, или вид первоначальной модели, может быть получен в результате качественного изучения физической сути исследуемого явления, путем опроса специалистов, в результате постулирования определенных свойств объекта.

Качество результатов, полученных с помощью ДФЭ, зависит от правильности списка «существенных переменных». При анализе качества найденной модели может быть принято решение о построении более сложной и точной модели.

Матрица спектра плана ДФЭ 2^{n-p} является частью матрицы ПФЭ 2^n. Выбор 2^{n-p} строк, составляющих матрицу спектра ДФЭ 2^{n-p}, должен производиться так, чтобы в матрице численных значений базисных функций *F* сохранялось условие линейной независимости столбцов, то есть не было полностью совпадающих или полностью противоположных столбцов. Только

при выполнении этого требования можно раздельно оценить все коэффициенты уравнения регрессии.

Если список существенных переменных содержит (*d+1*) элемент, то для того, чтобы оценить раздельно коэффициенты регрессии, необходимо выполнение условия:

$$N = 2^{n-p} \geq d + 1.$$

Это условие позволяет при заданном числе *n* факторов и известном списке существенных переменных *(d+1)* найти минимальную степень дробности *p*.

Возникает вопрос, какой план ПФЭ следует выбрать в качестве исходного. Для того чтобы дробная реплика представляла собой ортогональный план, в качестве реплики следует брать ближайший полный факторный эксперимент. В большинстве случаев для инженерного анализа пригодны линейные модели. Число опытов для оценки коэффициентов линейной модели при n факторах должно быть не менее *n+1*. Кроме того, дополнительно хотя бы один опыт (одну степень свободы) необходимо иметь для проверки адекватности линейной модели. Поэтому минимальное число опытов ДФЭ должно быть не менее, чем *n+2*. Определив значение величины *n+2*, из ряда 8, 16, 32, … выбирают ближайшее большее. Оно укажет минимальное число опытов плана ПФЭ, который должен быть выбран в качестве исходного, и, следовательно, число опытов плана ДФЭ.

Например, требуется построить план ДФЭ для исследования влияния на выходной параметр технологического процесс пяти факторов. Информация о силе эффектов взаимодействия факторов отсутствует. Построим план ДФЭ для получения линейной модели. Так как. *n*+1 =7, то минимальное число опытов ПФЭ равно 8. Следовательно, исходный план ПФЭ 2^3, в качестве исходных выбираем любые три фактора. Строим для этих факторов план ПФЭ2^3. Для построения требуемого ДФЭ 2^{5-2} введем два фактора вместо малозначимых взаимодействий, построенных только из исходных факторов. Так как информация о силе эффектов взаимодействий отсутствует, то для фактора x_4 лучше выбрать взаимодействие высшего порядка в исходном ПФЭ3^5, например, $x_1x_2x_3$. Для фактора x_5 можно использовать одно из произведений x_1x_2, x_1x_3 или x_2x_3. Построенный план ДФЭ 2^{5-2} представляет собой ¼ часть плана ПФЭ 2^5. Если линейная модель, построенная по результатам опытов ДФЭ 2^{5-2}, окажется неадекватной, то необходимо ввести в модель наиболее влияющие взаимодействия факторов.

При планировании ДФЭ важным является вопрос, какие из взаимодействий следует выбрать для введения вместо них новых факторов. Если информация о силе влияния эффектов взаимодействий отсутствуют, то при введении в исходный план нового фактора выбирают вектор-столбец с большим порядком взаимодействия, так как обычно эффекты взаимодействия более высоких порядков менее значимы.

Например, надо решить трехфакторную задачу в линейном приближении, то есть получить модель $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$. Для решения задачи можно ограничиться четырьмя вариантами варьирования, если в планировании типа 2^2 произведение x_1x_2 приравнять третьему независимому фактору. Такое планирование позволяет найти свободный член b_0 и три оценки коэффициентов регрессии при линейных членах b_1, b_2, b_3 (из четырех опытов нельзя получить более четырех оценок коэффициентов регрессии). Итак, в ПФЭ 2^2 используем столбец x_1x_2 в качестве плана для x_3. Такой сокращенный план представлен ниже в таблице.

g	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+1	-1	-1	+1	+1	-1	-1	+1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+	+1	+	+1	+	+1	+1

Применение ДФЭ всегда связано со смешиванием, то есть с совместным оцениванием нескольких теоретических коэффициентов математической модели. В рассматриваемом примере, если коэффициенты регрессии при парных взаимодействиях отличны от нуля, каждый из найденных коэффициентов b_j служит оценкой двух теоретических коэффициентов регрессии:

$$b_0 \to \beta_0 + \beta_{123}; b_1 \to \beta_1 + \beta_{23}; b_2 \to \beta_2 + \beta_{13}; b_3 \to \beta_3 + \beta_{12}.$$

Действительно, указанные теоретические коэффициенты в таком планировании не могут быть оценены раздельно, поскольку столбцы матрицы планирования для линейных членов и парных произведений совпадают полностью (полностью коррелированны).

Как же конкретно должен производиться выбор эффекта взаимодействия. Приведенный выше формальный алгоритм (проверка столбцов матрицы *F*) является неприемлемым вследствие трудоемкости. Для практического применения используется следующая процедура:

1. Из числа $x_1, x_2, \ldots, x_n$ факторов выбираются k ведущих, k=n-p. Пусть, например, первые $x_1, x_2, \ldots, x_k$. Для них записывается план ПФЭ 2^k, то есть задается программа изменения каждого из ведущих факторов в ходе экспериментов.

2. Для оставшихся p=n-k факторов в качестве программы изменения в ходе эксперимента выбираются столбцы, соответствующие тем или иным произведениям ведущих факторов.

Соотношение, приводящее один из столбцов $x_{k+1}, x_{k+2}, \ldots, x_n$ к столбцу определенного произведения факторов $x_1, x_2, \ldots, x_k$, называется *генерирующим соотношением.*

Всего должно быть *p* генерирующих соотношений. Выбор их произволен, но нельзя использовать существенные переменные – факторы и те произведения ведущих факторов, которые имеются в списке существенных (иначе в матрице *F* будут совпадающие столбцы – для одного из факторов и для одного из взаимодействий). В рассматриваемом примере полуреплика плана ПФЭ 2^3 задавалась с помощью генерирующего соотношения $x_3 = x_1 x_2$. Генерирующее соотношение можно использовать со знаком «-».

3. Теперь необходимо проверить пригодность найденного спектра плана. Если в матрице *F* нет полностью совпадающих или полностью противоположных (в смысле знака) столбцов, то спектр плана пригоден. Если такие столбцы есть, то надо выбрать другое генерирующее соотношение или изменить набор ведущих факторов.

Процедура построения спектра плана ДФЭ, содержащего минимальное число точек и обеспечивающего получение необходимой регрессионной модели, требует при реализации значительных усилий. Тем более, что трудно заранее предсказать, существует ли в принципе решение. Существует формальный алгоритм, избавляющий от необходимости каждый раз записывать матрицу *F* и позволяющий получить системы совместных оценок. Для этого используется понятие *определяющего соотношения*.

Выражение определяющего соотношения (ОС) получается из генерирующего соотношения и имеет вид:

слева 1, справа – произведение левой и правой частей генерирующего соотношения.

Например, $x_3 = -x_1 x_2, OC: 1 = x_1 x_2 x_3$.

Число определяющих соотношений равно числу генерирующих соотношений.

Если определяющих соотношений несколько, то записывается *обобщенное определяющее соотношение* (ООС), которое равно почленному произведению всех определяющих соотношений.

Знание определяющего соотношения позволяет найти всю систему совместных оценок без изучения матрицы планирования ДФЭ. Соотношения, задающие эти оценки, можно найти, последовательно перемножая каждую переменную из списка существенных переменных на определяющее соотношение (с учетом того, что $x_f^{2k} = 1, x_f^{2k+1} = x_f$). Получаем d+1 равенство, каждое из которых содержит 2^p переменных (факторов и их взаимодействий). Действия всех переменных, входящих в определенное равенство, смешаны между собой. Им соответствуют одинаковые столбцы матрицы *F*.

Таким образом, для определения пригодности найденного плана ДФЭ требуется установить, не содержат ли какие-либо из указанных равенств хотя бы двух переменных из списка существенных переменных. Если нет, то план

пригоден для экспериментирования, если содержит, то он не позволяет оценить коэффициенты регрессионной модели и должен быть изменен.

Итак, этот алгоритм не только позволяет проверить пригодность спектра плана ДФЭ 2^{n-p}, но и заранее уяснить, какие из переменных, не включенных в список существенных, могут влиять на значения тех или иных коэффициентов регрессии. Если в ходе исследования выясняется, что список существенных переменных требуется откорректировать, то эти сведения дают возможность определить, какие коэффициенты регрессии определены неточно и представляют собой оценки действия нескольких переменных.

Отметим свойства планов ДФЭ 2^{n-p}. Эти планы принадлежат к группе ортогональных планов. По отношению к линейной модели они являются A-, D- и G-оптимальными, ротатабельными, если областью планирования эксперимента является гиперкуб с координатами вершин, принимающими значения ±1. Эти свойства, разумеется, имеют место только в тех случаях, когда возможно получение несмешанных оценок всех коэффициентов модели.

При использовании ДФЭ необходимо иметь четкое представление о так называемой разрешающей способности дробной реплики, то есть определить заранее, какие коэффициенты являются несмешанными оценками соответствующих генеральных коэффициентов. Тогда в зависимости от задачи получаются дробные реплики.

Пример 1: Пусть задан список существенных переменных:

$$x_0, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_2x_4.$$

Количество факторов – 4, в модель входят 8 базисных функций. Поэтому попытаемся использовать ДФЭ 2^{4-1}.

Выберем ведущими факторы $x_0, x_1, x_2. x_3$. Для фактора x_4 требуется указать генерирующее соотношение. Выберем в качестве генерирующего следующее произведение $x_4 = x_1x_2x_3$. То есть задаем программу изменения фактопа x_4, совпадающую с тем, как изменялся бы столбец взаимодействия $x_1x_2x_3$.

Определяющее соотношение будет иметь вид:

$$OC: 1 = x_1x_2x_3x_4.$$

Отсюда получаем систему смешивания:

$$x_0 = x_0x_1x_2x_3x_4,\ x_1 = x_2x_3x_4, x_2 = x_1x_3x_4, x_3 = x_1x_2x_4, x_4 = x_1x_2x_3,$$
$$x_1x_2 = x_3x_4, x_2x_3 = x_1x_4, x_2x_4 = x_1x_3.$$

План пригоден. Можно использовать генератор плана и с противоположным знаком: $x_4 = -x_1x_2x_3$.

Выясним, годится ли для этой модели генерирующее соотношение $x_4 = x_1x_3$?

Определяющее соотношение имеет вид: *ОС:* $1 = x_1x_3x_4$.

Система смешивания:

$$x_0 = x_0x_1x_3x_4,\ x_1 = x_3x_4, x_2 = x_1x_2x_3x_4, x_3 = x_1x_4, x_4 = x_1x_3,$$
$$x_1x_2 = x_2x_3x_4, x_2x_3 = x_1x_2x_4, x_2x_4 = x_1x_2x_3.$$

Рассматриваемый план можно использовать.

Генерирующие соотношения x_1x_2, x_2x_3, или x_2x_4 не можем использовать, так как они присутствуют в списке существенных.

Пример 2. Изменим немного список существенных переменных из предыдущего примера:

$$x_0, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_3x_4.$$

(Изменили последнее взаимодействие).

Вновь попытаемся использовать ДФЭ 2^{4-1}. В системе смешивания получим выражение $x_1x_2 = x_3x_4$. Полученный план непригоден для нахождения требуемой регрессионной модели.

Пример 3. Для модели со списком существенных переменных:

$$x_0, x_1, x_2, x_3, x_4, x_5, x_1x_2$$

попытаемся применить ДФЭ 2^{5-2}.

Необходимое условие $N= 2^{5-2}=8>d=1=7$ выполняется.

Выберем генераторы планов:

$$x_4 = x_1x_2x_3, x_5 = x_1x_3.$$
$$ООС: 1 = x_1x_2x_3x_4 \cdot x_1x_3x_5 = x_2x_4x_5.$$

Система смешивания:

$$x_0 = x_2x_4x_5,\ x_1 = x_1x_2x_4x_5, x_2 = x_4x_5, x_3 = x_2x_3x_4x_5, x_4 = x_2x_5,$$
$$x_5 = x_2x_4, x_1x_2 = x_1x_4x_5.$$

План пригоден.

Если к списку существенных этого примера добавим член x_4x_5, то получим систему смешивания:

$$x_0 = x_2x_4x_5,\ x_1 = x_1x_2x_4x_5, x_2 = x_4x_5, x_3 = x_2x_3x_4x_5, x_4 = x_2x_5,$$
$$x_5 = x_2x_4, x_1x_2 = x_1x_4x_5, x_4x_5 = x_2.$$

Отсюда следует, что в матрице численных значений базисных функций обнаружены совпадающие столбцы x_2 и x_4x_5. План непригоден, так как оценки коэффициентов при x_2 и x_4x_5 смешаны между собой.

Таким образом, этот алгоритм позволяет выяснить, какие из переменных, не включенных в список существенных, могут влиять на значения тех или иных коэффициентов регрессии.

6.3.7 Реализация экспериментальных планов.

Практическая реализация планов первого порядка содержит следующие этапы:

а) выбор центра плана, то есть значения базовых уровней по каждому фактору. Как ранее отмечалось, центр плана представляет собой уровень фактора, который представляет наибольший интерес в данный момент. В окрестности этой точки строится регрессионная модель;

б) один из наиболее ответственных этапов при подготовке к экспериментированию – определение интервалов варьирования по каждому фактору. На выбор этих интервалов влияет:

- информация об интенсивности помехи: при прочих равных условиях, чем сильнее влияние шума, тем шире нужно выбирать шаг варьирования для того, чтобы обеспечить достаточное отношение сигнал/шум;

- информация о степени нелинейности функции отклика: чем в большей степени можно ожидать, что функция отклика отличается от линейной, тем меньше должна быть величина интервала варьирования фактора;

- тип решаемой задачи. Величина интервала варьирования зависит от того, производится ли построение регрессионной модели для целей предсказания или же она будет использоваться в процедурах поиска экстремумов, где шаг варьирования должен быть намного меньше, чем в задачах предсказания. Это связано с тем, что в поисковых процедурах эксперимент применяется для изучения локальных свойств поверхности отклика вблизи центра эксперимента. Регрессионная модель, предназначенная для целей предсказания эксперимента, проводится в большей части области, представляющей интерес для исследования.

Поскольку исходная информация, с помощью которой осуществляется выбор интервалов варьирования, имеет ограниченный характер, то может появиться необходимость в корректировке первоначального варианта;

в) установить общее число опытов, то есть план эксперимента в целом; так как спектр плана строится в соответствии с вышеизложенными правилами, то в данном случае задача сводится к выбору элементов матрицы дублирования. Это связано с тем, что результаты наблюдений отклика носят случайный характер, приходится в каждой точке плана проводить не один, а m параллельных опытов, осреднение результатов которых дает возможность уменьшить погрешности оценки истинного значения отклика.

В планировании первого порядка обычно применяется равномерное дублирование опытов, то есть в матрице дублирования все диагональные элементы равны между собой и равны m. То есть каждая комбинация значений

факторов на определенной строке матрицы спектра плана повторяется в ходе эксперимента m раз. Общее число опытов при этом равно Nm.

Эксперимент реализуется отдельными сериями. Каждая серия включает в себя проведение N опытов, соответствующих всем строкам матрицы плана. Количество серий – m. В каждой серии порядок реализации опытов должен быть случайным, для чего применяется рандомизация: с помощью таблиц случайных чисел определяется случайная последовательность реализации экспериментов.; m-кратное проведение данной процедуры обеспечивает случайную последовательность реализации строк матрицы спектра плана в каждой серии опытов. Это позволяет в определенной степени исключить предвзятость и субъективизм исследователя, а также систематические ошибки, связанные, например, с разогревом или охлаждением агрегатов и др. факторов при исследовании процесса.

Конечным результатом проведения эксперимента является таблица значений выходного показателя – отклика объекта, которая является исходным материалом для последующей обработки.

Основные данные, позволяющие сравнить между собой различные планы первого порядка, приведены в таблице 6.1. Эти сведения могут оказаться полезными при выборе конкретного плана, отвечающего имеющимся практическим потребностям.

Таблица 6.1 – Основные характеристики планов первого порядка

План	Количество уровней варьирования	Число точек в спектре плана; близость к насыщенному	Критерии оптимальности, другие свойства	Возможный вид модели
Одно-факторный	3	2n; ненасыщенный при n>1 (2n>n+1)	Ротатабельный, ортогональный	Только линейная
ПФЭ 2^n	2	2^n; ненасыщенный при n>1 (2^n>n+1)	*A*-, *E*-, *D*-, *G*-оптимальный, ротатабельный, ортогональный	Линейная и неполная степенная
ДФЭ 2^{n-p}	2	2^{n-p}; насыщенный при 2^{n-p} = n+1; ненасыщенный при 2^{n-}>n+1 (более близок к насыщенному, чем ПФЭ)		В соответствии со списком существенных переменных

6.4 Контрольные вопросы

6.4.1 Приведите основные принципы планирования экспериментов.
6.4.2 Дайте определения критериям оптимальности планов.

6.4.3 Дайте определения планам первого порядка: однофакторный эксперимент, полный факторный эксперимент, дробный факторный эксперимент.

6.4.4 Свойства планов первого порядка.

6.4.5 Какие регрессионные модели можно получить с помощью планов первого порядка?

6.4.6 Как составляется матрица планирования ПФЭ 2^n? Объясните порядок проведения эксперимента методом ПФЭ.

6.4.7 Что такое дробная реплика? От чего зависит разрешающая способность дробных реплик?

6.4.8 Что такое генерирующее соотношение и как оно выбирается?

6.4.9 Как составляется матрица планирования ДФЭ?

6.4.10 Что такое определяющее соотношение и как с его помощью составляется система совместных оценок?

7. Планы второго порядка

7.1 Общие сведения

Планы второго порядка предназначены для получения регрессионной модели в виде полного квадратичного полинома (полинома второй степени):

$$y(x) = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} b_{ii} x_i^2 + \sum_{\substack{i,l=1 \\ i<l}}^{n} b_{il} x_i x_l. \qquad (7.1)$$

Такие планы используются, когда заранее известно, что объект обладает существенно нелинейными свойствами, либо когда использование планирования первого порядка не позволило получить адекватную регрессионную модель и появилась необходимость ее усложнить.

Планы второго порядка более сложные по структуре, требуют при своей реализации увеличенного количества опытов. Модель (7.1) содержит $k = 1 + 2n + C_n^2 = \frac{(n+1)(n+2)}{2}$ членов, что в $\frac{n+2}{2}$ раза больше, чем в линейной модели.

Следовательно, возрастает минимально необходимое количество точек в спектре плана. Для получения квадратичной зависимости каждый фактор должен изменяться, по крайней мере, на трех уровнях.

Все планы второго порядка делятся на две большие группы: симметричные и несимметричные.

Симметричным называется такой план, для которого выполняются следующие условия:

$$\sum_{g=1}^{N} x_{ig} = \Lambda_2; \sum_{g=1}^{N} x_{ig}^2 \, x_{jg} = \Lambda_{22}, i \neq j; \sum_{g=1}^{N} x_{ig}^4 = \Lambda_4;$$

$$\sum_{g=1}^{N} x_{ig} = \sum_{g=1}^{N} x_{ig}^3 = \sum_{g=1}^{N} x_{ig}\, x_{jg} = \sum_{g=1}^{N} x_{ig}^2\, x_{jg} \sum_{\substack{g=1 \\ i\neq j\neq k}}^{N} x_{ig}\, x_{jg} x_{kg} = 0$$

для любых *i,j,k=1,2,…n; i≠j, N* – число точек плана.

То есть для всех факторов x_i или указанных комбинаций нескольких факторов записанные суммы равны определенным константам, причем эти константы одни и те же независимо от конкретных значений номеров факторов. Для упрощения предполагается, что параллельные опыты не проводятся. Несимметричные планы такими свойствами не обладают.

Симметричные планы обладают большей упорядоченностью в расположении точек. Для них характерно подобие в программе изменения каждого из факторов, а в некоторых случаях и полная симметрия. Для симметричных планов справедливы достаточно простые соотношения для оценок коэффициентов регрессии, их выборочных дисперсий и ковариаций. Это связано с тем, что информационная матрица Фишера имеет своеобразную блочную структуру, причем большинство ее элементов равно нулю, а остальные равны Λ_2, Λ_{22}. Несимметричные планы во многом менее удобны, чем симметричные. Но зато более экономны по необходимому количеству опытов.

При использовании планов второго порядка весьма важным оказывается вопрос об области планирования эксперимента, то есть той части факторного пространства, где должны располагаться точки плана. Можно указать три варианта ее задания:

1) область планирования естественная, полученная путем достройки планов первого порядка. При этом новые, дополнительные точки, координаты которых определяются с учетом тех или иных критериев оптимальности, могут выходить за пределы области планирования на первом этапе, то есть за пределы гиперкуба $-1<=x_i<=+1$, $i=1,2,\ldots,n$. Размеры области планирования, получающиеся путем дополнения, зависят от вида критерия оптимальности и количества факторов n.

2) область планирования – гиперкуб $|x_{ig}| \leq 1, i = 1,\ldots,.n, g = 1,\ldots,N.$

3) область планирования – единичный гипершар $x_1^2 + x_2^2 + \ldots + x_n^2 \leq 1.$

Два последних варианта в отличие от первого предусматривают, что область планирования выбирается заранее, исходя от особенностей решаемой задачи. Область в виде гиперкуба более естественна с практической точки зрения.

При необходимости план всегда можно перенастроить для иной области планирования. Например, чтобы получить план, вписанный в гиперкуб, если имеется план, заданный в естественной области, для каждого фактора x_i находят максимальное по модулю значение и заменяют x_i на $\frac{x_i}{max\, x_{il}}$.

7.2 Симметричные планы второго порядка

Центральные композиционные (ЦКП), то есть строящиеся последовательно, планы состоят из трех блоков:

1) основу или ядро плана составляют точки ПФЭ 2^n или ДФЭ 2^{n-p};

2) так называемые «звездные» точки, расположенные на координатных осях на расстоянии $\pm\alpha$ от центра эксперимента; общее число таких точек $N_\alpha = 2n$;

3) опыты в центре плана, то есть нулевые (центральные) точки; число таких точек $N_0>=1$.

Общее число точек ЦКП: $N = N_ф + N_{ά} + N_0$.

Произвольный симметричный ЦКП приведен в таблице 7.1. Для примера в таблице 7.2 приведен план второго порядка для трех факторов (нормированных), которые варьируются на 5 уровнях: -α, -1, 0, +1, + α. Конкретные значения α и N_0 выбираются, исходя из тех или иных критериев оптимальности регрессионных экспериментов. В связи с этим рассматриваются ортогональные (ОЦКП) и ротатабельные (РЦКП) центральные композиционные планы.

Таблица 7.1 – Произвольный симметричный план ЦКП

Составляющие части ЦКП	g	x_1	x_2	…	x_n	Число точек
Ядро плана	1	-1	-1	..	-1	2^{n-p}
	2	+1	-1	…	-1	
	…	-1	+1	…	-1	
	2^{n-p}	…	…	…	…	
«Звездные» точки	2^{n-p} +1	-α				2n
	2^{n-p} +2	+ α				
	…		- α			
	,,,		+ α			
	…				- α	
	2^{n-p} +2n				+ α	
Центральные точки	2^{n-p} +2n+1	0			0	N_0
	…	…				
	2^{n-p} +2n+N_0	0			0	

Ортогональное центральное композиционное планирование. В случае ОЦКП критерием оптимальности плана является ортогональность столбцов матрицы планирования. В силу ортогональности все оценки коэффициентов определяются независимо друг от друга. Как правило, в ОЦКП $N_0=1$. По критерию ортогональности матрица Фишера должна быть диагональной, для чего необходимо принять меры для обеспечения попарной ортогональности столбцов, отвечающих свободному члену β_0 и квадратичным коэффициентам $\beta_{ii}, i=1,\ldots, n$, а также столбцов, отвечающих квадратичным членам между собой.

Таблица 7.2 – ЦКП второго порядка для трех факторов

g	x_0	x_1	x_2	x_3	x_1^2	x_2^2	x_3^3
1	+1	-1	-1	-1	+1	+1	+1
2	+1	+1	-1	-1			
…	…	…	…	…	…	…	…
8	+1	+1	+1	+1	+1	+1	+1
9	+1	-α	0	0	α^2	0	0
10	+1	+ α	0	0	α^2	0	0
11	+1	0	- α	0	0	α^2	0
12	+1	0	+ α	0	0	α^2	0
13	+1	0	0	- α	0	0	α^2
14	+1	0	0	+ α	0	0	α^2
15	+1	0	0	0	0	0	0

Ортогонализацию столбцов x_0 и x_i^2 производят с помощью преобразования:

$$\tilde{x}_i^2 = x_i^2 - \frac{1}{N}\sum_{g=1}^{N} x_{ig}^2 = x_i^2 - \bar{x}_i^2,$$

где N – общее число строк матрицы планирования.

В этом случае условие ортогональности выполняется, то есть:

$$\sum_{g=1}^{N} x_{0g}\,\tilde{x}_{ig}^2 = \sum_{g=1}^{N} x_{0g}\left[x_{ig}^2 - \frac{1}{N}\sum_{i=1}^{N} x_{ig}^2\right] = \sum_{i=1}^{N} x_{ig}^2 - \frac{1}{N}N\sum_{i=1}^{N} x_{ig}^2 = 0. \quad (7.2)$$

Иными словами, регрессионная модель ищется в виде:

$$\varphi(\bar{x}) = \beta_0^{'} + \sum_{i=1}^{n}\beta_i\, x_{i-}\sum_{i=1}^{n-1}\sum_{j=1}^{n}\beta_{ij}\, x_i x_j + \sum_{i=1}^{n}\beta_{ii}\,(x_i^2 - \bar{x}_i^2); \beta_0^{'} = \beta_0 - \sum_{i=1}^{n}\beta_{ij}\bar{x}_i^2.$$

Соотношение (7.2) справедливо при любом α. Но при произвольном α остаются неортогональными столбцы матрицы F, отвечающие различным квадратичным переменным. Поэтому в ОЦКП значение α выбирается как раз из условия ортогональности именно этих столбцов, то есть из следующего условия:

$$\sum_{g=1}^{N}\tilde{x}_{ig}^2\,\tilde{x}_{jg}^2 = \sum_{\substack{g=1\\ i<j}}^{N}(x_{ig}^2 - \bar{x}_i^2)(x_{jg}^2 - \bar{x}_j^2) = 0.$$

Отсюда выводится уравнение для расчета величины звездного плеча:

$$4\alpha^4 + 4N_ф\alpha^2 - N_ф(N_\alpha - N_0) = 0,$$

где α – величина звездного плеча;

N_ϕ – число точек ПФЭ (ДФЭ);

N_0 – число центральных точек;

N_α – число «звездных» точек.

Параметры ОЦКП для *n=2,3,4* приведены в таблице 7.3, в таблице 7.4 приведены базисные функции для двухфакторной модели.

Таблица 7.3 – Параметры ОЦКП для *n=2,3,4*

n	α	$N_{\acute{a}}$	N_0	N_ϕ	N
2	1,0	4	1	4	9
3	1,215	6	1	8	15
4	1,414	8	1	16	25

Таблица 7.4 – Базисные функции ОЦКП при *n=2*

g	x_0	x_1	x_2	x_1x_2	$(x_1^2 - \bar{x}_1^2)$	$(x_2^2 - \bar{x}_2^2)$
1	+1	-1	-1	+1	1/3	1/3
2	+1	+1	-1	-1	1/3	1/3
3	+1	-1	+1	-1	1/3	1/3
4	+1	+1	+1	+1	1/3	1/3
5	+1	-1	0	0	1/3	-2/3
6	+1	+1	0	0	1/3	-2/3
7	+1	0	-1	0	-2/3	1/3
8	+1	0	+1	0	-2/3	1/3
9	+1	0	0	0	-2/3	-2/3

Матрица Фишера ОЦКП имеет вид:

$$\Phi = \begin{Vmatrix} N & 0 & \dots & 0 & 0 & \dots & 0 & 0 \dots 0 \\ 0 & 2^{n-p} + 2\alpha^2 & \dots & 0 & 0 & \dots & 0 & 0 \dots 0 \\ & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 2^{n-p} + 2\alpha^2 & 0 & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 2^{n-p} & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \dots & 2^{n-p} & 0 \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 2\alpha^4 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 \dots & 0 & 0 \dots 2\alpha^4 \end{Vmatrix}$$

Обратная матрица:

$$\Phi = \left\| \begin{array}{ccccccccccc} \frac{1}{N} & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{2^{n-p}+2\alpha^2} & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{2^{n-p}+2\alpha^2} & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \frac{1}{2^{n-p}} & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \frac{1}{2^{n-p}} & 0 & \dots & 0 & & \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \frac{1}{2\alpha^4} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & \frac{1}{2\alpha^4} \end{array} \right\|$$

Дисперсии оценок коэффициентов вычисляются по следующим формулам:

$$\sigma^2_{b'_0} = \frac{\sigma_e^2}{N};\ \sigma^2_{b_i} = \frac{\sigma_e^2}{2^{n-p}+2\alpha^2}\ ;$$

$$\sigma^2_{b_{ij}} = \frac{\sigma_e^2}{2^{n-p}};\ \sigma^2_{b_{ii}} = \frac{\sigma_e^2}{2\alpha^4}\ ; i,j = 1,\dots,n.$$

Точность оценивания различных групп коэффициентов регрессии неоднородна (не как в ПФЭ): наиболее точно оценивается свободный член β'_0, несколько менее точно – линейные коэффициенты, еще хуже – коэффициенты β_{ij} при парных взаимодействиях, наименее точно β_{ii} – коэффициенты при квадратичных членах.

Для *n=2* имеем:

$$\sigma^2_{b'_0} = \frac{\sigma_e^2}{9};\ \sigma^2_{b_1} = \sigma^2_{b_2} = \frac{\sigma_e^2}{6}\ ; \sigma^2_{b_{12}} = \frac{\sigma_e^2}{4};\ \sigma^2_{b_{11}} = \sigma^2_{b_{22}} = \frac{\sigma_e^2}{2}.$$

Вычисление оценок коэффициентов регрессии, как обычно, производится методом наименьших квадратов, и в силу свойства ортогональности плана можно сразу написать для них выражения:

$$b'_0 = \frac{\sum_{g=1}^{N} \bar{y}_g}{N};\ b_i = \frac{\sum_{g=1}^{N} x_{ig}\bar{y}_g}{2^{n-p}+2\alpha^2};\ b_{ik} = \frac{\sum_{g=1}^{N} x_{ig}x_{kg}\bar{y}_g}{2^{n-p}}, i,k; = 1,2,\dots,n, i<k;$$

$$b_{ii} = \frac{\sum_{g=1}^{N} \tilde{x}_{ig}^2 \bar{y}_g}{\sum_{g=1}^{N} (\tilde{x}_{ig}^2)^2} = \frac{\sum_{g=1}^{N} [\tilde{x}_i^2 - \frac{1}{N}(2^{n-p} + 2\alpha^2]\bar{y}_g}{2\alpha^4};$$

$$b_0 = b_0' - N \sum_{i=1}^{n} b_{ii} \left(\bar{x}_i^2\right) = b_0' - \frac{2^{n-p}+2\alpha^2}{N} \sum_{i=1}^{n} b_{ii}.$$

ОЦКП не удовлетворяет критерию ротатабельности планирования, причем при фиксированном расстоянии от центра планирования дисперсии предсказания отклика по разным направлениям могут различаться весьма значительно – в два и более раз, что крайне нежелательно.

Ротатабельное центральное композиционное планирование. РЦКП строится исходя из требования критерия ротатабельности. Критерий оптимальности в этом случае $\sigma^2\{\hat{y}\} = const$ при *R*=const. Величину звездного плеча α вычисляют из соображений ротатабельности по формуле $\alpha = 2^{\frac{n-p}{4}}$.

Ротатабельные планы оптимальны и в том смысле, что они позволяют минимизировать систематические ошибки, связанные с неадекватностью представления результатов исследования полиномами второго порядка.

Число центральных точек N_0 не влияет на свойство ротатабельности и выбирается таким образом, чтобы обеспечивать равную точность предсказания выходной величины $\widehat{y}$ внутри области планирования.

Необходимые значения α_0 и N_0 приведены в таблице 7.5, в таблице 7.6 приведены базисные функции для двухфактороной модели.

Ротатабельный план не требует ортогонализации, поэтому никаких преобразований переменных при составлении матрицы планирования не производят. Поэтому записи матрицы Фишера и обратной к ней матрицы, дисперсии оценок коэффициентов в общем случае громоздки.

Таблица 7.5 – Параметры РЦКП для *n=2,3,4.*

n	α	$N_{\acute{a}}$	N_0	N_{ϕ}	*N*
2	1,414	4	5	4	13
3	1,682	6	6	8	20
4	2,000	8	7	16	31

Нахождение экстремальной точки с помощью полиномов второго порядка. Адекватные полиномы второго порядка используют для предсказания оптимальной точки (точки экстремума), так как градиентные методы оптимального поиска вблизи оптимума непригодны из-за ограниченной точности определения градиента (ограничен пробный шаг, ограничена точность измерения выходной величины).

Таблица 7.6 – Матрица планирования для двухфакторной задачи (*n=2, α=1,414; N_0=5).*

g	x_1	x_2	x_1x_2		x_1^2	x_2^2
1	+1	-1	+1		1	1
2	+1	+1	-1		1	1
3	+1	-1	-1		1	1
4	+1	+1	+1		1	1
5	-1,414	0	0		$(1,414)^2$	0
6	+1,414	0	0		$(1,414)^2$	0
7	0	-1,414	0		0	$(1,414)^2$
8	0	+1,414	0		0	$(1,414)^2$
9	0	0	0		0	0
10	0	0	0		0	0
11	0	0	0		0	0
12	0	0	0		0	0
13	0	0	0		0	0

Для нахождения оптимальной точки приравнивают нулю значения компонент вектора-градиента, вычисленных по формуле:

$$\frac{\partial y}{\partial x_i} = b_i + 2b_{ii}x_i = \sum_{k=1}^{n} b_{ik}x_k = 0, \;\; i,k = 1,2,\dots,n, i < k.$$

Решая систему из n линейных уравнений по числу n факторов, получают n координат оптимальной точки в относительных единицах (не следует забывать, что начало координат в этом случае находится в центре эксперимента, а уровни факторов нормированы). Затем по формуле для дисперсии $\sigma^2\{\hat{y}\}$ оценивают точность предсказания целевой функции $\hat{y}_g$.

Формула для точности предсказания значения выходной величины в заданной точке факторного пространства x_{ij} по полученному полиному:

$$\sigma^2\{\hat{y}_g\} = \sigma^2\{b_0\} + \sum_{i=1}^{n} x_{ig}^2\,\sigma^2\{b_i\} + \sum_{i=1}^{n} x_{ig}^2 x_{kig}^2\,\sigma^2\{b_{ik}\}, i < k.$$

7.3 Контрольные вопросы

7.3.1 В каких случаях используется планирование второго порядка?

7.3.2 В чем отличие центральных композиционных планов от планирования ПФЭ и ДФЭ?

7.3.3 В чем преимущество симметричных планов второго порядка?

7.3.4 Как определяется общее число точек ЦКП?

7.3.5 Что является критерием оптимальности плана при ОЦКП и РЦКП?

7.3.6 Как достигается ортогональность матрицы планирования при ОЦКП?

7.3.7 Как с помощью полученного математического описания находится точка экстремума?

7.3.8 Приведите матрицу планирования ОЦКП для двухфакторного и трехфакторного эксперимента.

7.3.9 Приведите матрицу планирования РЦКП для двух и трехфакторного эксперимента

7.3.10 Для чего необходимо проделывать параллельные опыты в одних и тех же точках факторного пространства?

8. Планирование экстремальных экспериментов

8.1 Задача оптимизации в экспериментальных исследованиях

Задача оптимизации – одна из наиболее важных и распространенных задач, встречающихся в практике научных и инженерных исследований как теоретического, так и прикладного характера. Во многих случаях основной целью является определение наилучших в некотором смысле условий, решений, значений параметров, уровней факторов. Такие оптимизационные задачи возникают, например:

1) при управлении различными технологическими процессами, агрегатами, где необходимо достижение максимума производства при наилучшем качестве и минимуме затрат;

2) при проектировании различных инженерных устройств, приборов, когда необходимо подобрать параметры таким образом, чтобы получить наилучшие эксплуатационные характеристики;

3) при создании новых образцов продукции с наилучшими свойствами (например, сплав с максимальной прочностью);

4) при определении регрессионной модели методом наименьших квадратов или численном построение плана экспериментов в соответствии с выбранным критерием возникают задачи чисто вычислительного характера.

Задача оптимизации формулируется следующим образом: найти значения управляемых факторов $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_n = x_n^*$ объекта исследования, при которых его отклик y (целевая функция, критерий оптимизации) достигает своего экстремального значения (минимума или максимума):

$$y^* = extr\ y(\overline{x}),$$

где $\overline{x} = (x_1, x_2, \ldots, x_n)$.

Экстремальные точки во многих случаях находятся с учетом определенных ограничений на $\bar{x}$.

Большинство объектов имеют ограничения, и в этих условиях применение методов поиска имеет свои особенности. Имеется два типа ограничений: факторные и функциональные.

Факторными называются ограничения, накладываемые на допустимые значения отдельных факторов:

$$x_{min} \le x_i \le x_{max}, i = 1, \dots, n. \quad (8.1)$$

Ограничения этого типа фактически всегда присутствуют при решении любой оптимизационной задачи.

Ограничения на значения совокупности нескольких факторов:

$$A_{jmin} \le \varphi_i(x_1, \dots, x_n \le A_{jmax}, \ j = 1,2, \dots, k. \quad (8.2)$$

Такие ограничения имеют место, если какие-либо комбинации значений факторов x_i оказываются недопустимыми. Например,

$$x_1 + x_2 + x_3 \le A_{max}, x_1 \cdot x_2 \le A_{min}, \sqrt{\sum x_i^2 \le A} \text{ и т.д.}$$

Приведенные типы ограничений полностью известны, заданы в аналитической форме.

Функциональные ограничения – это ограничения, накладываемые на целевые функции (при многооткликовом объекте исследования), характеризующие качественные или количественные стороны объекта:

$$y_{pmin} \le y_p \le y_{pmax,} \quad (8.3)$$

где p – номер отклика.

То есть требуется получить экстремальное значение одной функции отклика при ограничениях, наложенных на другие функции отклика (например, объем выпуска продукции должен быть не ниже плана, при этом процент примесей в готовой продукции не должен быть выше допустимого).

В общем случае неравенства (8.1–8.3) определяют в факторном пространстве область допустимых значений Ω. Если точка экстремума находится вне области Ω, то задача оптимизации формулируется как поиск условного экстремума, когда требуется найти максимум y на подмножестве $x \in \Omega$, то есть наилучшее значение на границе области ограничений.

В условиях ограничений задача поисковой оптимизации формулируется таким образом: с помощью целенаправленного поиска найти координаты точки максимума (минимума) целевой функции внутри области ограничений, то есть $\max \{y(x) | x \in \Omega$.

Все методы оптимизации можно разбить на два класса:

1) теоретические методы, когда задача оптимизации определена с математической точки зрения и допускает применение аналитических методов

оптимизации – дифференциальное или вариационное исчисление, линейное, динамическое, целочисленное программирование;

2) экспериментальные методы, используемые в условиях, когда функция отклика $y(\bar{x})$ неизвестна и имеется возможность измерить значения y при различных комбинациях величин $x_1, x_2, \ldots, x_n$. Такая ситуация может быть не только при исследовании физических объектов, но и в задачах теоретического плана, когда аналитические методы непригодны и используются численные методы (например, эксперименты проводятся на вычислительной машине).

Отличием методов оптимизации в задачах исследования реальных объектов от процедур чисто вычислительного плана является наличие неконтролируемых факторов – шума случайного характера.

Рассматриваемые методы оптимизации делятся на две группы: поисковые методы; методы, основанные на предварительном получении эмпирической модели объекта, описывающей его поведение в области оптимума.

В поисковых методах осуществляется последовательное локальное изучение поверхности отклика. В этом случае экстремальное значение достигается с помощью последовательных процедур, зависящих от алгоритма поиска. Вторая группа методов предусматривает изучение не локальных, а общих свойств поверхности отклика в определенной области факторного пространства, где предположительно находится оптимум. В качестве подобной модели могут использоваться регрессионные модели в виде полного степенного (квадратичного) полинома. Определение точки экстремума осуществляется с помощью известных теоретических методов (дифференциального исчисления, линейного программирования и т. п.). Для методов оптимизации данной группы наибольший интерес представляет проблема планирования эксперимента, обеспечивающего наивысшую точность в определении координат оптимума.

В дальнейшем для рассматриваемых методов предполагается:

а) объект исследования имеет единственный отклик, стационарный (отсутствует временной дрейф);

б) функция отклика имеет единственный экстремум (свойство унимодальности) в области исследования (то есть выделен промежуток локального экстремума);

в) в методах оптимизации, использующих регрессионные модели, выполняются предпосылки регрессионного анализа.

8.2 Одномерный поиск

Методы одномерного поиска применяются для решения простых однофакторных оптимизационных задач, могут входить в многомерные поисковые процедуры. Их рассмотрение полезно для лучшего понимания в дальнейшем методов многомерного поиска. Эти методы делятся на две группы: методы исключения и шаговые поисковые методы.

Суть методов исключения состоит в следующем. Предполагается, что точка экстремума достигается при каком-то значении фактора x из заранее известного интервала (x_{min}, x_{max}); этот интервал называется *интервалом неопределенности.* Требуется с помощью наименьшего количества опытов в максимальной степени сузить длину этого интервала, последовательно исключая из рассмотрения те его части, где нахождение экстремума невозможно. Предполагается, что функция отклика $y(x)$ – унимодальная и не имеет участков постоянства (монотонна); помеха отсутствует. В этих условиях, чтобы уменьшить длину исходного интервала неопределенности $L = |x_{min} - x_{max}|$, необходимы, как минимум, два опыта в некоторых точках $x_{(1)}, x_{(2)}$ таких, что $x_{min} < x_{(1)} < x_{(2)} < x_{max}$. Могут иметь место следующие варианты исходов подобного эксперимента:

1) $y(x_{(1)}) < y(x_{(2)})$; максимум будет находиться в точке $x^* > x_{(1)}$, и первоначальный интервал неопределенности превратится в новый $(x_{(1)}, x_{max})$;

2) $y(x_{(1)}) > y(x_{(2)})$; максимум располагается в точке $x^* < x_{(2)}$, исходный интервал надо заменить на $(x_{min}, x_{(2)})$;

3) $y(x_{(1)}) = y(x_{(2)})$; тогда максимум находится в интервале $(x_{(1)}, x_{(2)})$.

Существуют различные методы размещения указанных точек на каждом этапе экспериментирования. Для их сопоставления вводится показатель эффективности соответствующего плана эксперимента: отношение длины начального интервала неопределенности L и интервала L_N, получившегося после проведения N опытов: $E = \frac{L}{L_N}$.

Рассмотрим некоторые методы одномерного поиска.

Метод последовательной дихотомии. Этот метод предусматривает размещение на каждом этапе экспериментирования сразу двух новых точек, расположенных симметрично относительно середины интервала неопределенности на расстоянии ε друг от друга, где ε – малая величина. Допустимое значение ε – это минимальная разница между соседними наблюдениями, которая может быть обнаружена инструментально с помощью измерительных средств. Координаты первых двух точек равны:

$$x_{(1)} = \frac{(x_{max} + x_{min} - \varepsilon)}{2}; x_{(2)} = \frac{(x_{max} + x_{min} + \varepsilon)}{2}.$$

Координаты точек экспериментирования на следующих шагах определяются по аналогичным формулам с учетом новых границ получающихся интервалов неопределенности. Длина интервала неопределенности после k шагов равна:

$$L_N = \frac{L}{2^k} + \left(1 - \frac{1}{2^k}\right) \cdot \varepsilon; N = 2k.$$

Отсюда можно найти значение показателя эффективности метода.

Метод золотого сечения. В этом методе координаты первой точки находятся по формуле:

$$x_{(1)} = x_{min} + qL,$$

где $q = 0{,}382$.

В этом методе каждая новая точка делит очередной интервал на две части, причем отношение большей и меньшей частей равно отношению всего интервала к его большей части. Деление отрезка таким образом называется «золотым сечением». Эффективность этого метода:

$$E = 1/(1-q)^{N-1} = 1/(1-0{,}618)^{N-1}.$$

Недостаток рассмотренных методов: при случайных помехах теряется их работоспособность. В этих случаях надо использовать шаговые процедуры поиска.

Алгоритм с постоянным шагом. В окрестности исходной точки x_0 проводятся два опыта в точках:

$$x_{(1)} = x_0 - \frac{\lambda}{2}, x_{(2)} = x_0 + \frac{\lambda}{2}$$

и определяются $y(x_{(1)}), y(x_{(2)})$.

Сопоставляя эти значения, можно определить направление движения и выполнять шаговое движение в этом направлении, число шагов заранее не оговаривается. Процедура заканчивается, когда для трех последовательных шагов имеют место соотношения:

$$y(x_{(m-1)}) < y(x_{(m)}), y(x_{(m+1)}) < y(x_{(m)}).$$

Тогда точка x_m соответствует искомому максимальному значению.

Этот алгоритм простой, но точность определения максимума связана с λ. Эта величина не может быть слишком большой, но и чрезмерное его уменьшение может повлечь необходимость проведения большого числа опытов.

Алгоритм с переменным шагом. Идея состоит в том, что сначала с помощью больших шагов получают приближенную информацию об интервале, содержащем точку экстремума, а затем уточнить его, уменьшая шаг.

8.3 Методы многомерного поиска

Задача поиска экстремума многомерной функции целевого отклика встречается на практике чаще, чем одномерные задачи. Вместе с тем решение этих задач является трудоемким:

- размерность задачи количественно увеличивается, в связи с чем усложняются расчеты;
- требуемое свойство унимодальности функции отклика в многомерном случае менее вероятно;
- даже если функция отклика унимодальна, на поиск могут влиять локальные свойства поверхности отклика (так называемые «овраги», «гребни», «хребты»), которые не имеют аналогов в одномерном случае;
- трудность сопоставления различных алгоритмов поиска, поскольку для любого алгоритма можно подобрать такую функцию отклика, когда он окажется предпочтительнее других, а также построить другую функцию отклика, при которой этот алгоритм станет неработоспособным;
- повышенные требования к точности локализации точки экстремума.

В поисковых методах осуществляется последовательное локальное изучение поверхности отклика. В этом случае экстремальное значение достигается с помощью последовательных процедур, включающих в себя:

а) определение по результатам специально спланированных экспериментов направления движения из некоторой начальной точки, в окрестности которой проводится эксперимент. Это направление зависит от локальных свойств поверхности отклика вблизи данной точки и определяется таким образом, чтобы движение в найденном направлении приводило к значению отклика, более близкому к оптимальному, чем в предыдущей точке;

б) организацию движения в найденном направлении;

в) многократное повторение указанных этапов до достижения точки оптимума.

Методы поиска экстремума делятся на две большие группы:

- градиентные (метод простого градиента, метод крутого восхождения, метод сопряженного градиента;
- неградиентные методы: метод Гаусса – Зейделя, метод случайного поиска, симплексный метод.

Рассмотрим из каждой группы по одному методу, получившему наибольшее распространение. Это простые, «грубые» алгоритмы.

8.3.1 Метод Гаусса – Зейделя.

Этот метод является простым неградиентным методом. Неградиентные методы поиска отличаются большим разнообразием идей, положенных в их основу. Данные методы объединяет необходимость совершения пробных шагов с последующим движением в ту сторону, где результаты проб оказались

благоприятными. В методе Гаусса – Зейделя при поиске экстремума каждый из факторов изменяется поочередно.

Метод Гаусса – Зейделя предусматривает поочередное нахождение частных экстремумов целевой функции по каждому фактору x_i $(i=1,2,..,n)$. При этом на каждом i-м этапе стабилизируют $n-1$ факторов и варьируют только один i-й фактор.

Рассмотрим процедуру метода для двухфакторного примера.

Задачу поиска максимума методом Гаусса – Зейделя решают в несколько этапов, объединенных в циклы:

1 этап:

а) выбирают основную (начальную, базовую) точку по правилам, описанным выше;

б) выбирают интервал варьирования Δx_1 по фактору x_1. Очевидно, что ступень варьирования не должна быть слишком малой, иначе движение к экстремуму окажется замедленным. Кроме того, на интервале варьирования $\Delta x_i (i = 1,2,\dots,n)$ изменение целевой функции должно быть существенно большим, чем погрешность ее измерения δy (не менее, чем в 5–10 раз):

в) определяют координаты пробных точек M_1 и M_2:

$$\bar{x}(M_1), = (x_{10} + \Delta x_1; x_{20}\dots; x_{n0}), \bar{x}(M_2), = (x_{10} - \Delta x_1; x_{20}\dots; x_{n0});$$

г) в точках M_1 и M_2 ставят пробные опыты (для повышения точности результатов могут выполняться параллельные опыты), измеряют отклики $y(M_1), y(M_2)$;

д) сравнивают полученные отклики, и если

$$y(M_2) > y(M_1),$$

то совершают рабочее движение на один рабочий шаг Δx_1 по направлению $\overrightarrow{M_0 M_2}$ в точку M_3;

е) аналогичные шаги продолжают в том же направлении до тех пор, пока на каком-то k-ом шаге не окажется, что

$$y(M_k) < y(M_{k-1}),$$

то есть значение отклика в очередной k-ой рабочей точке станет уменьшаться – это и служит признаком достижения частного экстремума. За частный экстремум принимают $(k-1)$-ю точку с откликом $y(M_{k-1})$.

2 этап.

Его проводят в том же порядке, что и 1 этап, с той лишь разницей, что стабилизируют все факторы, кроме фактора x_2. За новую базовую точку принимают точку с координатами:

$$\bar{x}(M_{k-1}), = (x_{10} \pm \Delta x_1 \cdot (k-2); x_{20} \ldots; x_{n0}),$$

а x_2 варьируют на выбранную по аналогичным условиям величину интервала варьирования Δx_2. По достижении частного экстремума по фактору x_2 точку нового частного экстремума принимают за новую базовую точку.

Первый цикл продвижения к экстремуму заканчивается n-м этапом, на котором стабилизируют все факторы, кроме фактора x_n. Для него выбирают шаг варьирования Δx_nи совершают пробное, а затем рабочее движение до достижения частного экстремума по фактору x_n. Если экстремум не достигнут, то выполняют второй цикл поиска.

Второй цикл, как и первый, начинается с 1 этапа, на котором варьируют фактор x_1 при стабилизации остальных $x_i (i \neq 1)$, затем последовательно выполняют n этапов по каждому из n факторов.

Поисковое движение к экстремуму заканчивают по достижении такой точки факторного пространства, при движении из которой в любую сторону по всем n факторным осям x_i в положительном и отрицательном направлениях, значения отклика оказываются меньшими. Такую точку принимают за экстремум (максимум).

Метод Гаусса – Зейделя является простым для практической реализации, но ясно, что траектория поиска не будет кратчайшей. Кроме этого, может случиться ложная остановка процедуры поиска, если поисковая точка окажется на узком «гребне» функции отклика. Это обстоятельство существенно снижает область применимости этого метода.

Из других неградиентных методов можно упомянуть метод *случайного поиска, симплексный метод.*

8.3.2 Метод простого градиента.

Все градиентные методы основаны на предварительном определении градиента функции отклика $y = y(x_1, .., x_n)$. Как известно, градиентом непрерывной функции $y(\bar{x})$ называется вектор с координатами:

$$\frac{\partial y}{\partial x_i} |_{\bar{x}_0}, i = 1,2, \ldots, n,$$

где $\bar{x}^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ – точка, в которой берется значение градиента.

Градиентные методы имеют несколько разновидностей, различающихся правилами выбора ступеней варьирования и рабочих шагов на каждом этапе движения к экстремуму. Сущность стратегии всех этих разновидностей состоит в том, что на каждом этапе вокруг очередной базовой точки организуют пробные эксперименты, по результатам которых оценивают новое направление градиента, после чего в этом направлении совершают один рабочий шаг.

Итак, вектор-градиент в n-факторном пространстве определяется соотношением:

$$grady = \frac{\partial y}{\partial x_1}\bar{x}_1^0 + \frac{\partial y}{\partial x_2}\bar{x}_2^0 + \ldots + \frac{\partial y}{\partial x_n}\bar{x}_n^0,$$

где $\bar{x}_i^0 (i = 1,2,\ldots,n)$ – единичные направляющие векторы, расположенные вдоль факторных осей;

$\frac{\partial y}{\partial x_i}$ – частная производная целевой функции по i-му фактору.

Пробные опыты (по два в точках, расположенных на прямых, параллельных каждой факторной оси и проходящих через базовую точку) проводят с целью получить приближенные оценки частных производных.

Метод *простого градиента* осуществляется по следующей процедуре:

а) выбирают начальную (базовую) точку $\bar{x}_0 = (x_{10}, x_{20}, \ldots, x_{n0})$ – точку L_0;

б) выбирают интервал варьирования Δx_i по каждому из факторов $(x_i, i = 1,2,\ldots,n)$, пользуясь изложенными выше правилами;

в) определяют координаты пробных точек:

$$\bar{x}(L_1), = (x_{10} - \Delta x_1; x_{20}\ldots; x_{n0}), \bar{x}(L2), = (x_{10} + \Delta x_1; x_{20}\ldots; x_{n0}),$$

то есть варьируют один фактор x_1 при стабилизации остальных факторов на базовом уровне. Аналогично вычисляют координаты пробных точек вдоль направлений, параллельных остальным факторным осям $x_2, x_3, \ldots, x_n$. В пробных точках ставят опыты и получают значения целевой функции y;

г) по результатам пробных опытов вычисляют оценки составляющих вектор-градиента в точке L_0 для каждого i-го фактора:

$$\text{grad}y(L_0)|_{x_i} = \frac{\partial y}{\partial x_i} \approx \frac{\Delta y|_{x_i}}{2\Delta x_i} = \hat{a}_i.$$

В частности, для фактора x_1 по результатам опытов в точках L_1 и L_2 вычисление выполняют по формуле:

$$\text{grad}y(L_0)|_{x_1} \approx \frac{\Delta y|_{x_i}}{2\Delta x_i} = \frac{y(L_2) - y(L_1)}{x_1(L_2) - x_1(L_1)} = \hat{a}_i;$$

д) находят координаты рабочей точки на направлении градиента. Для этого выбирают параметр рабочего шага $\rho_{гр}$ и вычисляют координаты первой рабочей точки по всем факторным осям $x_i, i = 1,2,\ldots,n$:

$$x_{i1} = x_{i0} + \rho_{гр}\hat{a}_{i0};$$

е) первую рабочую точку принимают за новую базовую точку и вокруг нее организуют новые пробные опыты для оценивания нового направления градиента, после чего совершают новый рабочий шаг.

В общем случае в каждой *k*-й рабочей точке по результатам пробных опытов вокруг нее получают оценки составляющих градиента $\hat{a}_{ik}$ и совершают (*k+1)*-й рабочий шаг (*k=0,1,2…,*) в точку с координатами:

$$x_{i,k+1} = x_{ik} + \rho_{гр}\hat{a}_{ik};$$

ж) рабочее движение производят до тех пор, пока на очередном шаге все составляющие градиента не станут пренебрежимо малыми, то есть все $\hat{a}_{i,k+1} \approx 0, (i = 1,2,\dots,n)$. Для этого достаточно, чтобы выполнялось условие:

$$\rho_{гр}\hat{a}_{i,k+1} < 1.$$

Если по результатам пробных опытов в (*k+1)*-й рабочей точке выполняется это условие, то движение к экстремуму прекращают и эту рабочую точку принимают за точку экстремума.

Метод простого градиента в целом эффективен и позволяет двигаться к экстремуму по достаточно короткому пути. Но в большинстве случаев он требует значительного числа опытов.

Другие градиентные методы: метод крутого восхождения, метод сопряженных градиентов.

Следует отметить, что разные поисковые методы в разных условиях обладают различной помехоустойчивостью. Под помехоустойчивостью метода понимается его способность правильно оценивать направление рабочего движения, а также способность быстро и точно приводить рабочую точку в область экстремума несмотря на наличие помех. Метод Гаусса – Зейделя является достаточно помехоустойчивым. Помехоустойчивость метода простого градиента не слишком высока, причем при попытке ее повысить приходится увеличивать число экспериментов. Тем не менее, он применяется на практике достаточно часто, особенно когда уровень помех невысок.

Что касается задачи определения *условного* экстремума, то формальное применение рассмотренных методов поиска не приводит успеху. Методы должны быть модифицированы. Различные методы в разной степени удобны для решения задачи поиска условного экстремума. Например, Метод Гаусса – Зейделя хорошо приспособлен для работы при ограничениях типа (8.1); симплексный метод достаточно эффективен и в условиях ограничений, причем алгоритм поиска почти не меняется, так как любая новая вершина, не удовлетворяющая ограничениям, отбрасывается как наихудшая. В настоящее время при поиске экстремума в условиях ограничений наибольшее применение находят градиентные методы.

До тех пор, пока точка поиска находится внутри области Ω, все процедуры, характерные для этих методов, в принципе остаются без изменений. Единственным отличием является необходимость на каждом шаге при реализации каждого опыта осуществлять проверку выполнения заданных ограничений: аналитически для ограничений типа (8.1), (8.2), экспериментально – для ограничений типа (8.3). И только если окажется, что имеет место нарушение одного или нескольких ограничений, требуется изменить алгоритм определения направления движения последующего шага поиска.

8.4 Контрольные вопросы

8.4.1 Сформулируйте задачу оптимизации.

8.4.2 Отличие поисковых методов оптимизации от теоретических методов.

8.4.3 Что общего у всех методов экспериментального поиска экстремума?

8.4.4 Для чего нужны пробные опыты в поисковых методах?

8.4.5 В чем отличие базовой и рабочей точек в методах поиска экстремума?

8.4.6 Опишите одномерные поиски экстремума.

8.4.7 Что служит критерием достижения экстремума?

8.4.8 Опишите алгоритм поиска экстремума методом Гаусса – Зейделя.

8.4.9 В чем заключается основная идея и процедура метода простого градиента?

8.4.10 Какого вида ограничения могут быть в задачах поиска экстремума?

Список литературы

1 Берикашвили В. Ш. Статистическая обработка данных, планирование эксперимента и случайные процессы: учебное пособие для бакалавриата и магистратуры – Москва: Издательство Юрайт, 2019.

2 Третьяк Л. Н. Основы теории и практики обработки экспериментальных данных: учебное пособие для бакалавриата и магистратуры – Москва: Издательство Юрайт, 2017.

3 Ибраева Л.К. Теория и техника инженерного эксперимента. Конспект лекций для магистрантов специальности 6М070200 – Автоматизация и управление . – Алматы: АУЭС, 2015.

4 Сидняев Н. И. Теория планирования эксперимента и анализ статистических данных: учебное пособие для магистров. – Москва: Издательство Юрайт, 2014.

5 Ибраева Л.К., Тохтасынова Н.Р. Теория и техника инженерного эксперимента. Методические указания к выполнению лабораторных работ для магистрантов специальности 6М070200 – Автоматизация и управление. – Алматы: АУЭС, 2017.

6 Карпушкин С., Глебов А.О. Теория инженерного эксперимента. Учебное пособие. – Тамбов: ТГТУ, 2017.

7 Реброва, И. А. Теория планирования эксперимента: учебное пособие. – Омск: СибАДИ, 2016.

8 Адамбаев М.Д. Теория и практика технического эксперимента в электроэнергетике. 6М070200 – Автоматизация и управление. – Алматы: КазНТУ, 2013.

9 Семенов В. А. Теория вероятностей и математическая статистика: Учебное пособие — СПб.: Питер, 2013.

10 Григорьев-Голубев В.В. Теория вероятностей и математическая статистика. Руководство по решению задач. – СПб.: БХВ-Петербург, 2014.

11 Красовский Г.И., Филаретов Г.Ф. Планирование эксперимента. Изд-во БГУ им. В. И. Ленина, 1982. - https://www.studmed.ru/krasovskiy-gi-filaretov-gf-planirovanie-eksperimenta_1c0562109a1.html

12 https://www.matematicus.ru/tablitsy-po-teorii-veroyatnosti-i-statistiki

13 http://bek.sibadi.org/fulltext/esd104.pdf

14 https://www.chitai-gorod.ru/catalog/book/1053786/

15 http://elar.urfu.ru/bitstream/10995/39965/1/978-5-9904848-4-9_2015.pdf

Содержание

Введение 3
1 Статистическое моделирование объектов управления 4
1.1 Характеристика видов связей между рядами наблюдений 4
1.2 Классификация видов экспериментальных исследований 7
1.3 Погрешности результатов исследований 8
1.4 Контрольные вопросы 10
2 Основные понятия теории планирования экспериментов 11
2.1 Определения 11
2.2 Выбор уровней и интервалов варьирования факторов. Понятие воспроизводимости эксперимента 14
2.3 Контрольные вопросы 19
3 Статистическое оценивание 19
3.1 Свойства оценок параметров случайных величин 19
3. 2 Метод максимального правдоподобия 22
3.3 Метод моментов 25
3.4 Точечные оценки параметров нормального распределения 26
3.5 Интервальные оценки 31
3.6 Выборочные распределения 34
3.7 Планирование эксперимента при построении интервальных оценок 42
3.8 Контрольные вопросы 43
4 Статистические гипотезы 43
4.1 Проверка статистических гипотез 43
4.2 Критерии значимости 46
4.3 Отсев грубых погрешностей 50
4.4 Контрольные вопросы 52
5 Построение регрессионных моделей объектов 52
5.1 Элементы регрессионного анализа 52
5.2 Оценки параметров регрессионной модели 56
5.3 Статистический анализ результатов построения регрессионной модели 62
5.4 Контрольные вопросы 68
6 Планы регрессионного анализа 69
6.1 Основные принципы планирования экспериментов 69
6.2 Критерии оптимальности планов 71
6.3 Планы первого порядка 73
6.4 Контрольные вопросы 95
7 Планы второго порядка 96
7.1 Общие сведения 96
7.2 Симметричные планы второго порядка 98
7.3 Контрольные вопросы 103

8 Планирование экстремальных экспериментов 104
8.1Задача оптимизации в экспериментальных исследованиях 104
8.2 Одномерный поиск 106
8.3 Методы многомерного поиска 109
8.4 Контрольные вопросы 114
Список литературы 115
Содержание 116

Ибраева Лида Куандыковна

ТЕОРИЯ И ТЕХНИКА ИНЖЕНЕРНОГО ЭКСПЕРИМЕНТА

Учебное пособие

Редактор Жанабаева Е.Б.

Пописано в печать __. __. _____
Тираж 100 экз. Формат 60×84 1/16

Бумага типографская № 2
Объем 7,0 уч.-из. л.. Заказ № ___
Цена 3500 тенге

Копировально-множительное бюро
некоммерческого акционерного общества
«Алматинский университет энергетики и связи»
050013, Алматы, ул. Байтурсынова, 126/1