

М.М. КАНЕ

АНАЛИЗ ИСХОДНЫХ ДАННЫХ ПРИ СТАТИСТИЧЕСКОЙ ОБРАБОТКЕ РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ

Допущено

Министерством образования Республики Беларусь
в качестве учебного пособия для студентов
учреждений высшего образования
по специальности магистратуры
«Инновационные технологии в машиностроении»



МИНСК
«ВЫШЭЙШАЯ ШКОЛА»
2024

УДК 621:001.89(075.8)
ББК 34.4(я73)
К19

Рецензенты: кафедра технологии металлов учреждения образования «Белорусский государственный аграрный технический университет» (профессор кафедры доктор технических наук, профессор *Л.М. Акулович*); начальник научно-технического Центра технологии машиностроения и технологического оборудования государственного научного учреждения «Объединенный институт машиностроения» Национальной академии наук Беларуси доктор технических наук, профессор *В.Л. Басинюк*

Все права на данное издание защищены. Воспроизведение всей книги или любой ее части не может быть осуществлено без разрешения издательства.

ISBN 978-985-06-3600-3

© Кане М.М., 2024
© Оформление. УП «Издательство
“Вышэйшая школа”», 2024

ПРЕДИСЛОВИЕ

Современная экономика, в том числе машиностроение, имеет инновационный характер. Развитие экономики в настоящее время обеспечивается не за счет наращивания объемов производимой продукции, а за счет повышения эффективности производства. Этот рост достигается путем создания новой техники, совершенствованием средств и методов ее производства. В основе этих процессов лежат фундаментальные и прикладные научные исследования во всех отраслях экономики.

При выполнении прикладных исследований в области машиностроения широко используются статистические методы планирования исследований и обработки их результатов. Применение этих методов позволяет получить объективные оценки изучаемых закономерностей и свойств исследуемых объектов, оценить достоверность результатов, ускорить проведение исследований.

При изучении статистических методов научных исследований основное внимание традиционно уделяется планированию экспериментов и обработке их результатов. Под планированием эксперимента здесь имеется в виду разработка алгоритма проведения исследования, выбор матрицы эксперимента, регламентирующей значения переменных в различных опытах и порядок их проведения. При этом не рассматриваются такие важные этапы планирования эксперимента, как выбор наиболее важных выходных характеристик (параметров оптимизации) и входных параметров (факторов) изучаемого процесса или объекта, предварительный анализ экспериментальных данных (исключение резко выделяющихся результатов, анализ законов распределения, оценка стационарности и эргодичности изучаемых процессов, оценка точности измерений, ее учет при записи результатов и др.). В некоторых работах эти методы называют методами предпланирования эксперимента.

Основной задачей данной учебной дисциплины является изучение методов подготовки к выполнению экспериментов (методов предпланирования экспериментов) с целью повышения эффективности экспериментов. В данном учебном пособии основное внимание уделено рассмотрению указанных методов предпланирования эксперимента. Приведено также краткое описание наиболее распространенных методов планирования многофакторного эксперимента с помощью полного факторного эксперимента (ПФЭ) и моделирования процессов с помощью корреляционно-регрессионного анализа (КРА). Эта информация позволит будущим магистрам по специальности «Инновационные технологии в машиностроении» повысить эффективность научных исследований, что будет способствовать ускорению инновационного развития машиностроения Республики Беларусь.

Автор

ПОНЯТИЯ И ПРИНЦИПЫ ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

1.1. Основные понятия и определения теории планирования эксперимента

Как и в любом сформировавшемся научном направлении, в теории планирования эксперимента выработалась определенная система основополагающих понятий и терминов. Приведем наиболее важные из них с необходимыми комментариями и в интерпретации, принятой большинством специалистов данной области.

Объект исследования (ОИ) есть носитель некоторых неизвестных и подлежащих изучению свойств и качеств. Предполагается, что исследователь обладает той или иной априорной информацией об объекте, хотя бы примерным перечнем переменных, влияющих на свойства объекта.

Как правило, любой ОИ можно представить в виде «черного ящика» с определенным количеством входов и выходов (рис. 1.1). Выделяют:

1) входные контролируемые и управляемые переменные, которыми исследователь может варьировать по своему усмотрению: вектор

$$X = \|x_1, x_2, \dots, x_n\|;$$

2) входные контролируемые, но неуправляемые переменные: вектор $Z = \|z_1, z_2, \dots, z_k\|$;

3) неуправляемые и неконтролируемые переменные: вектор $E = \|e_1, e_2, \dots, e_f\|$;

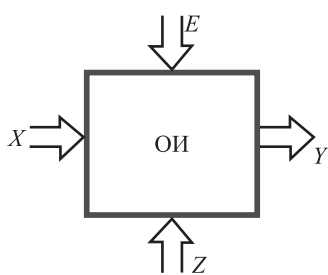


Рис. 1.1. Структурная схема объекта исследования

4) выходные показатели (характеристика исследуемых свойств или качеств объекта): вектор $Y = \|y_1, y_2, \dots, y_e\|$.

Переменные X и Z принято называть факторами. Факторы могут меняться во времени как детерминированные (чаще всего факторы X) или случайные (чаще всего факторы Z) переменные. Пространство контролируемых переменных образует *факторное пространство*.

Переменные группы E есть некоторые мешающие переменные, причем характер их влияния на Y может быть двояким. Если мысленно представить себе, что факторы X и Z стабилизированы во времени, то под влиянием переменных E показатели Y могут меняться достаточно систематическим (ситуация, когда говорят, что имеет место «дрейф объекта») или практически непредсказуемым образом. Тогда говорят о наличии случайной помехи, случайного шумового поля. При этом, как правило, полагают, что вероятностные свойства шумового поля не меняются во времени.

По физическому происхождению к переменным E прежде всего относятся ошибки измерительных приборов или методов анализа, а также неконтролируемые изменения сырья в промышленных объектах или изменения свойств вследствие старения или износа установок, влияние внешней среды, обслуживающего персонала и т.д., включая воздействие тех переменных, которые в принципе могли бы контролироваться экспериментатором, но не включены им в число исследуемых факторов (вследствие трудностей их измерения, по ошибке или незнанию). К переменным группы E должны быть отнесены и такие, как квалификация экспериментатора, особенности используемого экспериментального оборудования и методологии измерений, уровень метрологического обеспечения, т.е. переменные, действия 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 эксперимента развито для случая, когда размерность вектора Y равна 1 и влияние переменных E (шумового поля) можно представить в виде аддитивной составляющей измеряемого значения Y , т.е. в виде эквивалентной, приведенной к выходу объекта аддитивной помехи e (рис. 1.2).

Выходную переменную Y – зависимую переменную объекта – часто называют *откликом* или *параметром оптимизации*, зависимость отклика от рассматриваемых факторов – *функцией отклика*, а геометрическое представление функции отклика – *поверхностью отклика*.

Отметим, что объектами исследования в практике планирования экспериментов могут быть:

- реальные физические объекты (устройства, лабораторные и промышленные установки и т.д.) и процессы (технологические, экономические, биологические и др.);
- физические модели реальных объектов (модели информационных систем, электрических и магнитных полей и т.д.);

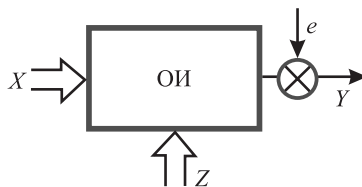


Рис. 1.2. Структурная схема объекта исследования с аддитивной помехой e

- математические модели реальных объектов (модели в виде совокупностей нелинейных дифференциальных или интегродифференциальных уравнений, наличие которых, тем не менее, не позволяет широко внедрить их в практику по причине чрезвычайной сложности, или же модели имитационного программного характера).

В двух последних случаях в помеху ϵ войдут также составляющие, обусловленные неточностью физического или математического моделирования объектов.

Математическая модель ОИ — это зависимость, содержательно отражающая те или иные свойства изучаемого объекта, в частности структуру и количественные связи, его характеризующие. Математические модели, используемые в планировании эксперимента, как правило, изображаются символами и называются *знаковыми моделями* (в отличие от алгоритмических или программных моделей). В дальнейшем будут рассматриваться только такие модели, конечное количественное представление которых получено на базе экспериментальных исследований, или, иными словами, экспериментальные математические модели (в отличие от теоретических моделей). При этом предполагается, что именно такого рода модели могут быть получены в ситуации, когда имеют место случайная помеха, шум ϵ , т.е. в ситуации, наиболее часто встречающейся в задачах экспериментального характера. Фактически методы планирования эксперимента и предназначены для получения статистических моделей, вследствие чего сами эти методы часто именуют *статистическими методами планирования эксперимента*.

Основной особенностью любой экспериментальной статистической модели является то, что подобная модель не может точно описать поведение объекта в любом конкретном опыте (скажем, указать точное значение Y , даже если функция отклика определена абсолютно точно, а векторы X и Z стабилизированы). 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. Исследователь не может предсказать точное значение Y в каждом опыте, но с помощью соответствующей статистической модели может указать, вокруг какого центра будут группироваться значения Y при данном сочетании значений факторов X и Z , если для этого сочетания опыты повторять многократно; какова степень разброса Y и т.д.

Из различных статистических моделей, пожалуй, наибольший практический интерес представляют *регрессионные модели* (модели регрессионного анализа). Регрессионная модель выражается соотношением

$$\begin{aligned}
 Y &= \varphi(x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_k; \Theta_0, \Theta_1, \Theta_2, \dots, \Theta_d) + e = \\
 &= \varphi((X, Z); \bar{\Theta}) + e,
 \end{aligned}
 \tag{1.1}$$

где $\Theta = \|\Theta_0, \Theta_1, \dots, \Theta_d\|$ – вектор параметров модели, причем вид функции φ полагают заданным, а параметры $\bar{\Theta}$ неизвестны и подлежат определению из эксперимента. Для произвольного опыта с номером g , очевидно, справедливо

$$Y_g = \varphi((X_g, Z_g); \bar{\Theta}) + e.
 \tag{1.2}$$

Подобные модели можно интерпретировать как разложение неизвестной функции отклика в ряд (в тригонометрический ряд Фурье, в ряд по различным системам ортогональных полиномов или функций, в ряд Тейлора и т.д.). В последнем случае имеют место *полиномиальные регрессионные модели* различного порядка (первого порядка, или линейная, задаваемая полиномом первого порядка по факторам; второго порядка, или квадратичная, задаваемая полиномом второго порядка по факторам, и т.д.).

Эксперимент – это система операций, воздействий и (или) наблюдений, направленных на получение информации об объекте при исследовательских испытаниях.

Опыт – это воспроизведение исследуемого явления в определенных условиях проведения эксперимента при возможности регистрации его результатов. Опыт – отдельная элементарная часть эксперимента.

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

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям. В более общем смысле под планированием эксперимента понимают всю совокупность действий, направленных на разработку стратегии экспериментирования от начальных до заключительных этапов изучения объекта исследования (от получения априорной информации до создания работоспособной математической модели или определения оптимальных условий). Иными словами, планирование эксперимента – это целенаправленное управление, которое реализуется в условиях неполного знания механизма изучаемого явления.

Принято подразделять эксперимент на пассивный и активный. При *пассивном эксперименте* существуют только факторы группы Z и экспериментатор находится в положении пассивного наблюдателя, фиксирующего изменения факторов Z и отклика Y (в таком положении пребывает, например, астроном, наблюдающий процессы в далеких галактиках, вмешаться в ход которых он не в состоянии). Особенности планирования пассивного эксперимента достаточно

подробно освещены в литературе. Однако основным предметом исследования теории планирования эксперимента является все-таки *активный эксперимент*, т.е. когда существуют лишь факторы группы X и экспериментатор в соответствии с планом эксперимента может их целенаправленно изменять. В дальнейшем будут обсуждаться только вопросы, связанные именно с активным экспериментом.

Для активного эксперимента *область действия* — область возможных значений факторов X при экспериментировании; *область планирования* — область значений факторов X , в которой находятся точки, отвечающие условиям проведения опытов используемого плана эксперимента. Область планирования составляет часть области действия или совпадает с ней. Если область планирования задается интервалами возможного изменения факторов $X_{i\min} \leq X_i \leq X_{i\max}$, $i = 1, 2, \dots, n$, то она представляет собой гиперпараллелепипед (при соответствующем масштабировании — гиперкуб). В ряде случаев область планирования удобнее задавать в виде гипершара. Тем не менее задание этой области во многом произвольно (говорят о «планировании на кубе», «планировании на шаре» и т.д.). Сравнительно редко встречаются сложные области планирования (действия) произвольной формы.

Точка плана — упорядоченная совокупность численных значений факторов, соответствующая условиям проведения опыта, точка факторного пространства, в которой проводится эксперимент. Точке плана с номером g отвечает вектор

$$X_g^T = \|X_{1g}, X_{2g}, \dots, X_{ng}\|,$$

где g — номер опыта; T — знак транспонирования матрицы.

Общая совокупность таких точек образует *план эксперимента*.

Фиксированное значение фактора называют *уровнем фактора*. Факторы могут различаться по числу уровней, на которых возможна фиксация в данной задаче. Понятие уровня фактора часто используются при описании характерных точек из области действия (планирования) фактора X_i : минимальный ($X_{i\min}$) и максимальный ($X_{i\max}$) уровни, основной уровень фактора (X_i^0 , $i = 1, 2, \dots, n$ (i — номер фактора)).

Основной уровень фактора указывает такие условия эксперимента, которые представляют наибольший интерес для исследования в данный момент.

Обычно вектор $X^{0T} = \|X_1^0, X_2^0, \dots, X_n^0\|$ задает в факторном пространстве точку, являющуюся в каком-то смысле центром области планирования, центром эксперимента. В ее окрестностях и располагаются все точки плана. Часто координаты X_i^0 выбираются с помощью соотношения

$$X_i^0 = \frac{X_{i\min} + X_{i\max}}{2}.$$

В ряде случаев полезно понятие *интервала (шага) варьирования фактора* X_i :

$$\Delta X_i = \frac{X_{i\max} - X_{i\min}}{2}, \quad (1.3)$$

так, что $X_{i\max} = X_i^0 + \Delta X_i$; $X_{i\min} = X_i^0 - \Delta X_i$.

Зная X_i^0 , ΔX_i , $i=1, 2, \dots, n$, можно реализовать операцию нормализации факторов (операцию перевода факторов в стандартизованный масштаб или к безразмерной шкале). Операция нормализации сводится к изменению начала отсчета координатных осей и масштаба в соответствии с соотношением

$$x_i = \frac{X_i - X_i^0}{\Delta X_i}, \quad i=1, 2, \dots, N. \quad (1.4)$$

Таким образом, для переменной x_i в стандартизованном масштабе начало координат совмещено с центром эксперимента, а в качестве единицы измерения используется шаг варьирования фактора. Применение безразмерных переменных (факторов) x_i часто существенно облегчает математические выкладки и запись конечных результатов, поэтому такое представление факторов будет широко употребляться в дальнейшем.

Матрица плана – стандартная форма записи условий проведения экспериментов в виде прямоугольной таблицы, строки которой отвечают опытам, а столбцы – факторам.

1.2. Основные принципы планирования эксперимента

Принципы, положенные в основу теории планирования эксперимента, направлены на повышение эффективности экспериментирования, т.е. на получение максимума информации при минимуме опытов.

Отказ от полного перебора возможных входных состояний. Для получения исчерпывающей информации о свойствах функции отклика в принципе необходимо проведение бесконечного числа опытов во всех точках области планирования эксперимента. В противном случае всегда существует теоретическая возможность пропустить некоторую особенность поверхности отклика. Указанную разновидность эксперимента можно назвать *экспериментом с полным перебором всех входных состояний*. Ясно, что он носит чисто умозрительный, гипотетический характер и нереализуем на практике. Если для

однофакторного объекта можно еще представить себе некий эксперимент, в какой-то степени близкий к полному перебору, когда экспериментатор, постепенно увеличивая или уменьшая значение фактора, непрерывно следит за откликом Y , то в случае, когда число факторов больше одного, подобный эксперимент уже становится принципиально нереализуемым. Экспериментатор просто вынужден задаться дискретной сеткой значений факторов, выбрать какое-то фиксированное число уровней каждого фактора. А коль скоро это так, то оказывается, что экспериментатор, желая того или не желая, фактически задается некоторыми свойствами поверхности отклика, постулирует определенную степень гладкости этой поверхности.

Если, например, для некоторого фактора выбрано k уровней варьирования и используется полиномиальная аппроксимация, ясно, что какие-то более быстрые изменения фактора, чем те, что могут быть описаны полиномом степени $k-1$, уже не будут обнаружены экспериментаторами. Опасность такого рода ошибок всегда существует при экспериментировании, и исследователь, интуитивно опасаясь их, как правило, стремится применить максимально большое число уровней варьирования. В многофакторных задачах это приводит к катастрофическому росту общего числа опытов. Тогда экспериментатор начинает уменьшать число уровней варьирования по каждому фактору с тем, чтобы общее число опытов стало приемлемым. В теории планирования эксперимента сознательно отказываются от полного перебора входных состояний или от эксперимента, близкого к нему по своей конструкции. Выбор числа уровней варьирования по каждому фактору непосредственно связывается с выбором вида функции отклика или, точнее, с выбором вида ее аппроксимации (в частности, в наиболее распространенных полиномиальных регрессионных моделях — со степенью аппроксимирующего полинома). Для линейной зависимости $\bar{y} = f(x_1, x_2, \dots, x_n)$ считается достаточным изменять факторы на двух уровнях (соответствующих максимальному и минимальному значениям каждого фактора в данном эксперименте — \max , \min или «+» и «-»). Для квадратичной зависимости (полинома 2-й степени) факторы изменяют на трех уровнях. Выбор вида функции отклика на стадии подготовки к планированию эксперимента можно осуществить, используя следующие принципы.

Принцип постепенного усложнения математической модели (принцип последовательного планирования). Этот принцип очень прост: в отсутствие априорной информации о свойствах функции отклика нет смысла сразу строить сложную модель объекта. Получение сложной модели требует большего числа опытов, чем простой модели, а может оказаться, что в сложной модели нет необходимости, что она вырождается в простую модель, поскольку таковы свойства объекта.

С точки зрения конечного результата это хорошо, но ведь напрасно затрачены время и средства на проведение лишних опытов уже не вернешь. Поэтому теория планирования эксперимента рекомендует, как правило, начинать с простейшей модели, соответствующей имеющейся априорной информации (например, с линейной модели, если нет никакой информации о свойствах объекта или есть информация, что объект не обладает ярко выраженными нелинейными свойствами, или с квадратичной модели, если известно, что функция отклика, по всей видимости, нелинейна).

Логика экспериментирования здесь такова: постановка небольшого числа опытов для получения простейшей модели, проверка ее пригодности. Если модель удовлетворяет исследователя, эксперимент заканчивается. Если же модель непригодна, необходим следующий этап (цикл) экспериментирования: постановка новых (дополнительных) опытов, позволяющих получить более сложную модель, ее проверка и т.д. до тех пор, пока не будет получена модель, которую исследователь признает достаточно хорошей.

Если обратиться к наиболее распространенным полиномиальным моделям, то подобная логика означает, что исследователь обычно начинает с построения простейшей линейной модели, для чего достаточно варьировать каждый фактор всего на двух уровнях. Затем в случае неудачи он переходит к построению квадратичных моделей; для этого нужно минимум три уровня варьирования по каждому фактору. Таким образом, если необходимо, постепенно увеличивается степень аппроксимирующего полинома или, что часто делается, усложняется модель с помощью подходящего нелинейного преобразования факторов. В итоге, если функция отклика очень сложная, можно оказаться в ситуации, близкой к полному перебору возможных входных состояний, но так практически никогда не происходит. Обычно исследователь довольно быстро определяет подходящую модель и экономит значительное число опытов по сравнению с вариантом, когда сразу ищется модель максимальной сложности.

Приведенная здесь логика экспериментирования, вообще говоря, составляет часть общей концепции последовательного эксперимента. Согласно этой концепции при проведении эксперимента необходимо использовать последовательную, шаговую стратегию. После каждого шага производится анализ результатов, затем принимается решение о дальнейшей деятельности. Исследователь отказывается от попытки заранее задать строго фиксированную схему проведения эксперимента. Используемая им стратегия предусматривает возможность принятия решений в зависимости от результатов, полученных на отдельных этапах исследования. Подобная концепция находит свое воплощение не только в упомянутых выше задачах построения

математических моделей, но и при проверке статистических гипотез (последовательные процедуры Вальда), контроле за спонтанно изменяющимся процессом (построение контрольных карт), в поисковых процедурах экспериментальной оптимизации, при построении последовательных планов эксперимента, когда каждый очередной опыт проводится после обработки ранее полученных данных в точке, дающей наибольшую информацию на этом шаге относительно интересующего исследователя свойства объекта, в различного рода последовательных процедурах обучения, распознавания образов, классификации и т.д.

В таких последовательных методах весьма важным является вопрос о проверке качества результатов на каждом этапе исследования (например, о проверке пригодности или адекватности найденной модели). В теории планирования эксперимента для проверки используются различные, математически строго обоснованные статистические процедуры, вытекающие из некоторых вероятностных свойств шума.

Принцип сопоставления с шумом. Предположим, что имеет место однофакторная ситуация (рис. 1.3). Не окажется ли в такой ситуации простейшая линейная модель вполне достаточной, например, для предсказания Y , хотя теоретическая функция отклика весьма сложной формы? Разумно ли при значительной случайной помехе тратить силы на получение весьма сложной модели? Если собираются использовать модель для единичного предсказания Y по X в ситуации, представленной на рис. 1.3, то ясно, что это делать нецелесообразно: все равно точность предсказания будет определяться в основном шумовой составляющей. Систематическая ошибка, связанная с неточностью модели, может проявиться только при многократном получении экспериментальных значений Y для фиксированного X

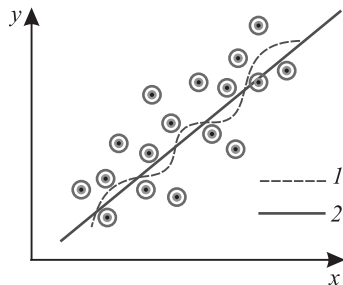


Рис. 1.3. Однофакторный объект (к принципу сопоставления с шумом):

1 — теоретическая функция отклика;
2 — линейная модель

в процессе последующего применения модели, но тогда эти значения при необходимости могут быть использованы для ее уточнения. Ясно, что при первоначальном построении модели ее точность должна соответствовать уровню (интенсивности) случайной помехи с тем, чтобы систематическая ошибка, связанная с неточностью аппроксимации функции отклика, оказалась бы сопоставимой с величиной шумового поля.

Фактически здесь можно провести известную аналогию с задачами теории измерений и измерительной техники.

Именно там рассматривается вопрос о случайной и неслучайной (систематической) составляющей суммарной ошибки измерения и об оптимальных соотношениях между ними. Считается, что резкое преобладание одной из этих компонент говорит об определенных недостатках в методе измерения и о целесообразности его изменения. Ясно, например, что применение прибора высокого класса точности для измерения переменной, отягощенной большой случайной ошибкой, есть расточительность.

Резюмируя сказанное, можно еще раз подчеркнуть: точность получаемой модели обязательно должна быть сопоставлена с интенсивностью случайной помехи, воздействующей на результат измерения отклика Y (конечно, с учетом той конкретной задачи, которая решается экспериментатором). При прочих равных условиях, чем меньше уровень помехи, тем более точной (и, как правило, сложной) должна быть модель; чем выше уровень помехи, тем в большей степени можно ожидать, что более простая (и обычно менее точная) модель окажется работоспособной.

Поскольку многие реальные объекты характеризуются высоким уровнем помех, при их описании получили наибольшее распространение полиномиальные регрессионные модели, причем в подавляющем числе случаев порядок такой модели равен 1 или 2. Подобные модели широко используются при создании различного рода инженерных методик расчета тех или иных устройств, схем и агрегатов, так как необходимая точность расчетов обычно весьма невелика (порядка 5–15%). Конечно, с теоретической точки зрения для выяснения физической сути механизма явления подобные полиномиальные модели менее содержательны, чем теоретические модели типа дифференциальных уравнений. Однако с практических позиций получение полиномиальных моделей является весьма эффективным, а часто и единственным средством изучения сложного объекта исследования.

Задачи определения пригодности (адекватности) той или иной построенной модели, как и задачи, связанные с уточнением вида модели, и проверки правильности посылок, при которых она была найдена, решаются с помощью стандартных статистических процедур (критериев). Использование их в известной степени формализует процедуру принятия решений после каждого этапа исследования, с одной стороны, облегчая работу экспериментатора, а с другой – уменьшая риск принятия необоснованных, «волевых» решений при подготовке результатов эксперимента под теоретическую концепцию исследователя.

Принцип рандомизации (принцип приведения к случайности). Этот принцип состоит в такой организации эксперимента, которая позволяет сделать случайным (рандомизировать) систематически

действующие переменные, не поддающиеся или поддающиеся с трудом учету и контролю, для того чтобы можно было рассматривать их как случайные величины и, следовательно, учитывать статистически. Иными словами, не в силах учесть действие неслучайных переменных, исследователь искусственно создает в эксперименте случайную ситуацию, т.е. переводит такие переменные в разряд случайных, избавляясь от возможных систематических ошибок в конечных результатах. При этом уровень шумового поля увеличивается, что, однако, особой роли не играет.

При различных статистических исследованиях принцип рандомизации предусматривает чисто случайный выбор элементов для последующего анализа из общей совокупности, подлежащей изучению. Тем самым обеспечивается представительность полученной выборки, т.е. гарантируется возможность с помощью измерения свойств конечного набора элементов из совокупности высказать обоснованное суждение о свойствах всей совокупности в целом.

Технически процедура рандомизации может быть организована следующим образом. Пронумеруем все элементы совокупности, подлежащей изучению, запишем эти номера на отдельные карточки, поместим их в урну и проведем своеобразную «лотерею», вытаскив из урны наугад нужное количество карточек. Номера, записанные на извлеченных карточках, укажут, какие элементы общей совокупности должны быть использованы для последующих измерений. Практически, конечно, подобные «лотереи» не проводятся, а для организации процедуры рандомизации используются таблицы равномерно распределенных случайных чисел (табл. 1.1).

Таблица 1.1. Пример таблицы равномерно распределенных случайных чисел

81	51	36	68	19	94	44	02	21	24
70	28	90	43	71	61	75	26	69	13
44	33	62	98	08	54	91	27	34	45
92	97	42	39	17	39	77	34	01	19
05	51	16	45	84	77	18	77	53	27
71	22	73	25	97	27	00	80	74	64
79	36	00	65	41	46	08	66	49	16
19	14	86	53	03	87	27	56	60	73
79	77	18	73	71	03	42	67	93	16
50	40	85	09	29	41	39	13	30	28

В подобных таблицах цифры можно группировать в многозначные числа в зависимости от общего количества элементов исследуемой совокупности (т.е. можно использовать по одной цифре, если эта совокупность содержит не более 10 элементов с номерами

от 0 до 9; образовать двухразрядные числа, если их количество не превосходит 100 — номера от 00 до 99, и т.д.). Двигаясь по таблице в любом направлении (по строке или столбцу), начиная с произвольного места в таблице и выписывая поочередно числа, соответствующие номерам элементов совокупности, можно выяснить, какие из них извлекаются для исследования.

При проведении различного рода экспериментов принцип рандомизации предусматривает случайный порядок реализации опытов, т.е. случайный порядок реализации строк матрицы плана. Использование такого, казалось бы, простейшего приема, как рандомизация опытов, позволяет устранить в математических моделях те или иные смещения, вызванные действием неконтролируемых систематических переменных.

Следует отметить, что при практической реализации процедуры рандомизации обязательно должен применяться некий объективный вероятностный механизм типа описанной «лотереи» или таблиц равномерно распределенных случайных чисел.

Принцип оптимальности планирования эксперимента. Этот принцип является центральным в теории планирования эксперимента. В соответствии с ним план эксперимента должен обладать некоторыми оптимальными свойствами с точки зрения определенного, заранее выбранного критерия оптимальности плана или совокупности подобных критериев.

Критерии оптимальности планов могут формулироваться по-разному. Фактически в таких критериях в строгой математической форме представлены, формализованы те или иные интуитивные соображения специалистов-экспериментаторов о хорошем, качественном эксперименте. При этом, конечно, общая направленность теории планирования — «меньше опытов — больше информации, выше качество результатов» — сохраняется. Конкретная форма критерия зависит прежде всего от типа решаемой задачи, назначения плана, хотя даже в рамках одного типа задач могут быть предложены различные критерии.

ЗАДАЧИ ПОДГОТОВКИ К ВЫПОЛНЕНИЮ ЭКСПЕРИМЕНТОВ. ОСОБЕННОСТИ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ

2.1. Однофакторный эксперимент

Однофакторный пассивный эксперимент проводится путем выполнения n пар измерений в дискретные моменты времени единственного входного параметра x (независимой переменной) и соответствующих значений выходного параметра y (зависимой переменной) (рис. 2.1).

Как известно, независимую переменную x называют фактором, а зависимую переменную y – параметром оптимизации.

Целью однофакторного пассивного эксперимента является построение регрессионной модели – установление зависимости

$$\bar{y} = f(x). \quad (2.1)$$

Вследствие случайной природы измеряемых величин x и результатов y , что характерно для исследований в области машиностроения, зависимость (2.1) устанавливает взаимосвязь математического ожидания параметра y (или его среднего арифметического значения \bar{y}) со значениями фактора x . Такая зависимость (модель) называется регрессионной.

Типы однофакторных экспериментов определяются природой фактора x . Факторы x могут быть качественными или количественными, внутри их области определения они могут изменяться непрерывно или дискретно.

Количественными факторами считаются факторы, поддающиеся измерению с помощью различных средств измерения. Примерами количественных факторов являются режимы резания при обработке деталей машин.



Рис. 2.1. Схема однофакторного эксперимента

Качественные факторы характеризуют условия эксперимента, которые не поддаются количественной оценке. Например, при изучении процессов резания металлов к качественным факторам можно отнести станок, оператора, партию заготовок, вид СОЖ и др.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	3
Глава 1. ПОНЯТИЯ И ПРИНЦИПЫ ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА	4
1.1. Основные понятия и определения теории планирования эксперимента.....	4
1.2. Основные принципы планирования эксперимента	9
Глава 2. ЗАДАЧИ ПОДГОТОВКИ К ВЫПОЛНЕНИЮ ЭКСПЕРИМЕНТОВ. ОСОБЕННОСТИ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ	16
2.1. Однофакторный эксперимент.....	16
2.2. Задачи выбора условий выполнения однофакторного эксперимента.....	19
2.3. Особенности планирования многофакторных экспериментов.....	19
2.4. Требования к планам экспериментов.....	21
2.5. Задачи выбора условий выполнения многофакторных экспериментов.....	23
2.6. Развитие методов планирования многофакторных экспериментов	24
2.7. Свойства и критерии планов эксперимента	27
2.8. Содержание и задачи эксперимента.....	29
Глава 3. МЕТОДЫ ВЫБОРА УСЛОВИЙ И ТИПОВ ОДНОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ	32
3.1. Выбор средств измерений	32
3.2. Выбор числа наблюдений. Запись результатов.....	33
3.3. Выбор интервалов между экспериментальными уровнями факторов и порядка проведения опытов.....	35
3.4. Методы рандомизации экспериментов.....	36
Глава 4. МЕТОДЫ ВЫБОРА УСЛОВИЙ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ	40
4.1. Требования к параметру оптимизации.....	40
4.2. Пути минимизации числа параметров оптимизации	42
4.2.1. Минимизация числа параметров оптимизации с помощью корреляционного анализа.....	43
4.2.1.1. Исключение параметров оптимизации, имеющих тесную связь с другими параметрами	43
4.2.1.2. Исключение параметров оптимизации путем решения «задачи о лидере»	44
4.2.2. Расчет обобщенного параметра оптимизации	46
4.2.2.1. Расчет на основе представлений о природе изучаемого объекта	46

4.2.2.2. Расчет на основе упрощенных представлений о качестве изучаемого объекта (процесса) при разных значениях параметров оптимизации (откликов)	47
4.2.2.3. Расчет на основе применения функции желательности	48
4.2.3. Экспертная оценка роли рассматриваемых факторов (параметров оптимизации) в обеспечении качества объекта (процесса)	55
4.3. Выбор факторов, влияющих на изучаемый процесс или объект, минимизация их числа	60
4.3.1. Методы выбора основных факторов, влияющих на изучаемый процесс (объект)	60
4.3.2. Требования к факторам	61
4.3.3. Отбор наиболее существенных факторов с помощью дисперсионного анализа	61
4.3.4. Методы планирования отсеивающих экспериментов	62
4.5. Предварительное определение области изменения каждого фактора, внутри которой функция отклика принимает экстремальное значение	71
Глава 5. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ИСХОДНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ, ПРАВИЛА ИХ ПРЕДСТАВЛЕНИЯ	78
5.1. Оценка результатов, содержащих резко выделяющиеся опытные данные	78
5.2. Основные понятия и задачи корреляционно-регрессионного анализа	79
5.3. Предпосылки применения корреляционно-регрессионного анализа, методы проверки их соблюдения	81
5.4. Правила округления результатов	86
5.5. Вычисления с приближенными числами	86
Глава 6. РАСЧЕТ И АНАЛИЗ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПО РЕЗУЛЬТАТАМ ОДНО- И МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ	88
6.1. Методы подбора эмпирических формул по результатам однофакторного эксперимента	88
6.2. Оценка точности прогнозирования с помощью однофакторной регрессионной модели	97
6.3. Полный факторный эксперимент типа 2^k	98
6.4. Расчет параметров математической модели по результатам полного факторного эксперимента	100
6.5. Порядок составления плана и обработки результатов полного факторного эксперимента	103
6.6. Оценка значимости коэффициентов регрессии	105
6.7. Оценка адекватности уравнения регрессии	107
ПРИЛОЖЕНИЯ	109
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	116

- Кане, М. М.**
К19 Анализ исходных данных при статистической обработке результатов научных исследований : учебное пособие / М. М. Кане. — Минск : Вышэйшая школа, 2024. — 118 с. : ил. ISBN 978-985-06-3600-3.

Приведены основные понятия и принципы планирования эксперимента, его виды и задачи, цели и методы выбора условий выполнения одно- и многофакторных экспериментов, пути сокращения числа параметров оптимизации в многофакторном эксперименте. Описаны методы выбора основных факторов, влияющих на изучаемый объект, с помощью экспертной оценки роли рассматриваемых факторов, дисперсионного анализа, путем применения насыщенных и сверхнасыщенных планов эксперимента. Представлены методы дихотомии, Фибоначчи и золотого сечения поиска области значений фактора, при которых функция приобретает экстремальные значения. Рассмотрены методы предварительного анализа и уточнения исходных данных; получения эмпирических зависимостей; основные понятия, задачи, предпосылки, способы реализации и проверки достоверности результатов корреляционно-регрессионного анализа и полного факторного эксперимента.

УДК 621:001.89(075.8)
ББК 34.4(я73)

Учебное издание

Кане Марк Моисеевич

**АНАЛИЗ ИСХОДНЫХ ДАННЫХ
ПРИ СТАТИСТИЧЕСКОЙ ОБРАБОТКЕ
РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ**

Учебное пособие

Редактор *Т.К. Хваль*

Художественный редактор *Т.Ю. Таран*

Компьютерная верстка *Ю.Н. Трусевич*

Корректор *Т.К. Хваль*

Подписано в печать 07.02.2024. Формат 60×90/16. Бумага офсетная. Печать офсетная.

Усл. печ. л. 7,5. Уч.-изд. л. 8,0. Тираж 200 экз. Заказ 125.

Республиканское унитарное предприятие «Издательство “Вышэйшая школа”». Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/3 от 08.07.2013. Пр. Победителей, 11, 220004, Минск. e-mail: market@vshph.com <http://vshph.com>

Республиканское унитарное предприятие «СтройМедиаПроект».

Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 2/42 от 13.02.2014. Ул. В. Хоружей, 13/61, 220123, Минск.